



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

ScienceDirect

journal homepage: [www.elsevier.com/locate/jval](http://www.elsevier.com/locate/jval)

## Questionnaire to Assess Relevance and Credibility of Modeling Studies for Informing Health Care Decision Making: An ISPOR-AMCP-NPC Good Practice Task Force Report

J. Jaime Caro, MDCM, FRCPC, FACP<sup>1,2,\*</sup>, David M. Eddy, MD, PhD<sup>3</sup>, Hong Kan, PhD<sup>4</sup>, Cheryl Kaltz, RPh, MBA<sup>5</sup>, Bimal Patel, BS, MS, PharmD<sup>6</sup>, Randa Eldessouki, MBBCH, MSc, MD<sup>7</sup>, Andrew H. Briggs, DPhil<sup>8</sup>

<sup>1</sup>Faculty of Medicine, McGill University, Montreal, Canada; <sup>2</sup>Evidera, Lexington, MA, USA; <sup>3</sup>Archimedes, Inc., San Francisco, CA, USA; <sup>4</sup>Glaxo Smith Kline, Research Triangle Park, NC, USA; <sup>5</sup>Prescription Drug Plan, University of Michigan, Northville, MI, USA; <sup>6</sup>Outcomes and PE Clinical Research Department, MedImpact Healthcare Systems, Inc., San Diego, CA, USA; <sup>7</sup>Scientific & Health Policy Initiatives, ISPOR, Lawrenceville, NJ, USA; <sup>8</sup>William R. Lindsay Chair of Health Economics, Health Economics & Health Technology Assessment, Institute of Health & Wellbeing, University of Glasgow, Glasgow, UK

### ABSTRACT

The evaluation of the cost and health implications of agreeing to cover a new health technology is best accomplished using a model that mathematically combines inputs from various sources, together with assumptions about how these fit together and what might happen in reality. This need to make assumptions, the complexity of the resulting framework, the technical knowledge required, as well as funding by interested parties have led many decision makers to distrust the results of models. To assist stakeholders reviewing a model's report, questions pertaining to the credibility of a model were developed. Because credibility is insufficient, questions regarding relevance of the model results were also created. The questions are formulated such that they are readily answered and they are

supplemented by helper questions that provide additional detail. Some responses indicate strongly that a model should not be used for decision making: these trigger a "fatal flaw" indicator. It is hoped that the use of this questionnaire, along with the three others in the series, will help disseminate what to look for in comparative effectiveness evidence, improve practices by researchers supplying these data, and ultimately facilitate their use by health care decision makers.

**Keywords:** credibility, good practices, modeling, quality assurance, validity.

Copyright © 2014, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

### Introduction

Four Good Practices task forces developed a consensus-based set of questionnaires to help decision makers evaluate 1) prospective and 2) retrospective observational studies, 3) network meta-analysis (indirect treatment comparison), and 4) decision analytic modeling studies with greater uniformity and transparency [1,2]. The primary audiences for these questionnaires are assessors of research studies for health technology assessment, formulary placement, and health care services decisions. They are intended for use without requiring deep levels of knowledge and expertise. This report focuses on the questionnaire to assess the relevance and credibility of modeling.

Decision analytic models consist of equations that are used to estimate the results of a decision, such as health outcomes and costs, for specified scenarios involving particular interventions in

specified populations [3]. They can help decision makers anticipate the outcomes they can expect if they implement the interventions in their settings [4]. Mathematical models can be very useful when there are no empirical studies that adequately address the decision maker's problem, and when it is not feasible to conduct such studies within the time limit of the decision [5,6]. In such cases, a model may be the only tool available to give decision makers the information they need for an informed decision [7,8].

Like any other source of information, mathematical models have limitations that decision makers should understand before they use the models' results [9]. All models involve assumptions about the clinical condition and its course, possible interventions and their effects, the behavior of people involved (patients, clinicians, caregivers, etc.), and other determinants of what may happen [10]. Some of these assumptions are encoded mathematically in equations that relate the change in one parameter

**Conflict of Interest Statement:** Some 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. Some of the points made in this article 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 this article.

\*Address correspondence to: J. Jaime Caro, Evidera, 430 Bedford Street, Suite 300, Lexington, MA 02420.

E-mail: [jaimo.caro@mcgill.ca](mailto:jaimo.caro@mcgill.ca).

1098-3015/\$36.00 – see front matter Copyright © 2014, International Society for Pharmacoeconomics and Outcomes Research (ISPOR).

Published by Elsevier Inc.

<http://dx.doi.org/10.1016/j.jval.2014.01.003>

## Background to the Task Force

On May 21, 2011, the Board of Directors approved, in principle, ISPOR's participation with the Academy of Managed Care Pharmacy (AMCP) and the National Pharmaceutical Council (NPC) in the Comparative Effectiveness Research Collaborative Initiative (CER-CI) for advancing appropriate use of outcomes research evidence to improve patient health outcomes. ISPOR's contribution to the CER-CI was to develop articles on how to assess prospective and retrospective observational studies, indirect treatment comparison (network meta-analysis), and decision analytic modeling studies to inform health care decision making. Four Good Practice task forces were created to develop these articles. Task Force Chairs were identified from leaders of ISPOR Good Research Practices task forces. Each Task Force consisted of two members from the AMCP, the NPC, and the ISPOR.

Each Task Force met independently via teleconference. In addition, the Task Force Chairs met via teleconferences and face-to-face meetings held on April 20, 2012 (San Francisco,

CA, USA), June 3, 2012 (Washington, DC, USA), June 28–29, 2012 (Boston, MA, USA), November 4, 2012 (Berlin, Germany), and May 21, 2013 (New Orleans, LA, USA), to coordinate a common outline and format for these articles. A focus group representing the US formulary decision-making community (22 participants) was convened April 20, 2012, at the AMCP Meeting, San Francisco, CA, USA, for feedback on the draft outline, format, and content of the assessment articles. The content of these reports was presented for comment at the ISPOR Annual International Meetings held June 4, 2012, and May 22, 2013, and the European Congress held November 5 and 6, 2012.

Draft modeling studies Task Force reports were sent for comment to their respective review group. Comments for each group were considered, and final draft reports were sent to the ISPOR membership for comment on September 23, 2013. Overall, there were 48 written comments for the modeling Task Force. All written comments are published on the ISPOR Web site, which can be accessed via the Research menu on ISPOR's home page: <http://www.ispor.org>. The final report was submitted to *Value in Health*.

(e.g., the risk of death) to changes in others (e.g., dose taken) [11]. The model is implemented in software used to compute the expected results given the particular inputs and assumptions made [12].

In using a modeling study, decision makers must determine how much confidence to place in the results [13]. Some may not have sufficient technical proficiency to fully understand the workings of the model, the equations that power it, or even the way data were analyzed to supply the inputs [14,15]. Moreover, there is always uncertainty about the results of a modeling analysis because the inputs are often derived from analyses of other data ("parameter" uncertainty) [16] and because of the assumptions made in constructing the model, including about the system at issue [17]. Thus, decision makers need to know the extent to which they should trust the results of the model [18,19]. Is it sufficiently relevant and credible to inform the decision at hand?

Few tools have been developed to aid in the assessment of decision analytic models [20,21], and most of these are aimed at technically adept users [22]. Moreover, new guidelines concerning the development of models have recently been issued [3,10,15,16,23–25]. Thus, it is timely to provide a questionnaire to help users assess the extent to which the model follows good practices and whether it can help them make an informed choice.

## Questionnaire Development

One issue in creating questionnaires for decision makers is whether they should be linked to a scorecard. Concerns were raised by the task forces that a scorecard with an accompanying scoring system may be misleading; it would not have adequate validity and measurement properties. Scoring systems may also provide users with a false sense of precision and have been shown to be problematic in the interpretation of randomized trials [26].

An alternative to a scorecard is a checklist. It was felt by the Task Force that checklists might also mislead users in their assessments because a study may satisfy all the elements of a checklist and still harbor "fatal flaws" in the methods applied in the publication. Moreover, users might have the tendency to count up the number of elements present converting it into an implicit score, and then apply that implicit scoring to their overall assessment of the evidence. In addition, the applicability of a study may depend on whether there is any other evidence that addresses the specific issue or the decision being made. In general, an evidence evaluator needs to be aware of the strengths

and weaknesses of each piece of evidence and apply his or her own reasoning. Indeed, it was felt that in addition to the potential for implicit or explicit scoring to be misleading, it would undermine the educational goal that the questionnaires provide.

## Questionnaire Items

When evaluating a modeling study, two main questions should be considered. First, how closely does the problem analyzed by the modelers apply to the problem faced by the decision maker ("relevance")? Second, how credible is the modeling?

Every analysis is conducted for a specific population, interventions, comparison, outcomes, and time horizon (the "model setting"). For the modeling results to accurately predict the outcomes that will occur in the "decision setting," two conditions must be met. First, the two settings should be as similar as possible. Any discrepancies between the two will affect how well the model predicts what will occur in the decision setting. Second, the model should be as credible as possible in predicting the outcomes in the model setting.

Accordingly, the modeling assessment questionnaire is divided into two sections. The first section helps assess the "relevance" of the model in the decision setting; the second helps determine the "credibility." Each section consists of a number of "main questions" that cover the principal topics (Table 1). To assist users in responding to these main questions, optional "helper questions" have been provided as well. Possible responses to each question are "Yes," "No," or "Can't Answer." If the user selects "Can't Answer," four additional responses are available: "Not Applicable," "Not Reported," "Not Enough Information," or "Not Enough Training." Because there will inevitably be degrees of sufficiency, users may also want to keep appropriate notes (e.g., "spot on" vs. "similar enough to consider the results, though they can't be applied literally"). A space for notes is provided for every question and the overall assessments. After responding to individual questions within each section, the user will determine whether the modeling study provides information that is "sufficiently" or "insufficiently" relevant and credible to consider in making the decision at hand.

## Relevance

Relevance addresses the extent to which the results of the model apply to the setting of interest to the decision maker. Stated another way, relevance asks how closely the model setting

**Table 1 – Questionnaire\* to assess the relevance and credibility of a modeling study.**

S. no.	Question	Helper questions to consider
<b>Relevance</b>		
1	Is the population relevant?	Are the demographics similar? Are risk factors similar? Are behaviors similar? Is the medical condition similar? Are comorbidities similar?
2	Are any critical interventions missing?	Does the intervention analyzed in the model match the intervention you are interested in? Have all relevant comparators been considered?
3	Are any relevant outcomes missing?	Does the background care in the model match yours? Are the health outcomes relevant to you considered?
4	Is the context (settings and circumstances) applicable?	Are the economic end points relevant to you considered? Is the geographic location similar? Is the health care system similar? Is the time horizon applicable to your decision? Is the analytic perspective appropriate to your decision problem?
<b>Credibility Validation</b>		
1	Is external validation of the model sufficient to make its results credible for your decision?	Has the model been shown to accurately reproduce what was observed in the data used to create the model? Has the model been shown to accurately estimate what actually happened in one or more separate studies? Has the model been shown to accurately forecast what eventually happens in reality?
2	Is internal verification of the model sufficient to make its results credible for your decision?	Have the process of internal verification and its results been documented in detail? Has the testing been performed systematically? Does the testing indicate that all the equations are consistent with their data sources? Does the testing indicate that the coding has been correctly implemented?
3	Does the model have sufficient face validity to make its results credible for your decision?	Does the model contain all the aspects considered relevant to the decision? Are all the relevant aspects represented and linked according to the best understanding of their characteristics? Have the best available data sources been used to inform the various aspects? Is the time horizon sufficiently long to account for all relevant aspects of the decision problem? Are the results plausible? If others have rated the face validity, did they have a stake in the results?
<b>Design</b>		
4	Is the design of the model adequate for your decision problem?	Was there a clear, written statement of the decision problem, modeling objective, and scope of the model? Was there a formal process for developing the model design (e.g. influence diagram, concept map)? Is the model concept and structure consistent with, and adequate to address, the decision problem/objective and the policy context? Have any assumptions implied by the design of the model been described, and are they reasonable for your decision problem? Is the choice of model type appropriate? Were key uncertainties in model structure identified and their implications discussed?
<b>Data</b>		
5	Are the data used in populating the model suitable for your decision problem?	All things considered, do you agree with the values used for the inputs? Did the approaches to obtaining and processing the data inputs meet the criteria from their corresponding questionnaires?

Table 1 – continued

S. no.	Question	Helper questions to consider
Analysis		
6	Were the analyses performed using the model adequate to inform your decision problem?	
7	Was there an adequate assessment of the effects of uncertainty?	
Reporting		
8	Was the reporting of the model adequate to inform your decision problem?	Did the report of the analyses provide the results needed for your decision problem? Was adequate nontechnical documentation freely accessible to any interested reader? Was technical documentation, in sufficient detail to allow (potentially) for replication, made available openly or under agreements that protect intellectual property?
Interpretation		
9	Was the interpretation of results fair and balanced?	
Conflict of Interest		
10	Were there any potential conflicts of interest?	
11	If there were potential conflicts of interest, were steps taken to address these?	

Note. **Relevance** questions relate to the usefulness of the modeling study to inform the particular health care decision at issue. Based on responses to the individual questions, the overall relevance of the modeling study for that decision is assessed as *Sufficient* or *Insufficient*. If the modeling study is considered sufficiently relevant, its **credibility** is assessed. The **credibility** is captured with questions in the following seven domains, *Validation, Design, Data, Analysis, Reporting, Interpretation, and Conflict of interest*. Based on responses to the individual questions, each domain is assessed as a *Strength, Neutral, Weakness, or Fatal flaw*. If a response triggers a “fatal flaw” indicator, the domain is marked as a fatal flaw, indicating that the modeling study has serious credibility issues. Based on the domain assessments, the overall credibility of the modeling study is judged as *Sufficient* or *Insufficient*.

\* The questionnaire consists of 15 questions related to the **relevance** and **credibility** of a modeling study. Each question is answered with *Yes/No/Can't Answer*.

matches the decision setting [27]. There is no single rating for relevance. Because each decision maker is interested in a different setting, the relevance of the model analyses may vary for each setting. Relevance has to be determined by each decision maker, and the relevance conclusion determined by one will not necessarily apply to other decision makers. At the end of the Relevance section, the user determines whether the relevance of the study is “Sufficient” or “Insufficient” to use in making a decision.

#### 1. Is the population relevant?

Does the population modeled sufficiently match the decision maker's population? Population characteristics to consider include *demographic characteristics* such as age and sex, nationality, race/ethnicity; *risk factors* such as average blood pressure, cholesterol, and body mass index levels; *behaviors* such as smoking and adherence to treatment; details of the *condition*, such as stage and severity and past and current treatments; and *comorbidities*. These aspects are covered in five helper questions.

#### 2. Are any critical interventions missing?

Do the modeled *interventions match those of interest* to the decision maker? Were all *relevant comparators* considered; How well does *background care* in the model match the modes of care in the decision setting? Intervention characteristics to consider include the following: technology used; administration technique; dose; duration of treatment (or protection for a vaccine); mode of administration; skill level and behavior of provider; posttreatment monitoring and care; and duration of follow-up. These aspects are covered in three helper questions.

#### 3. Are any relevant outcomes missing?

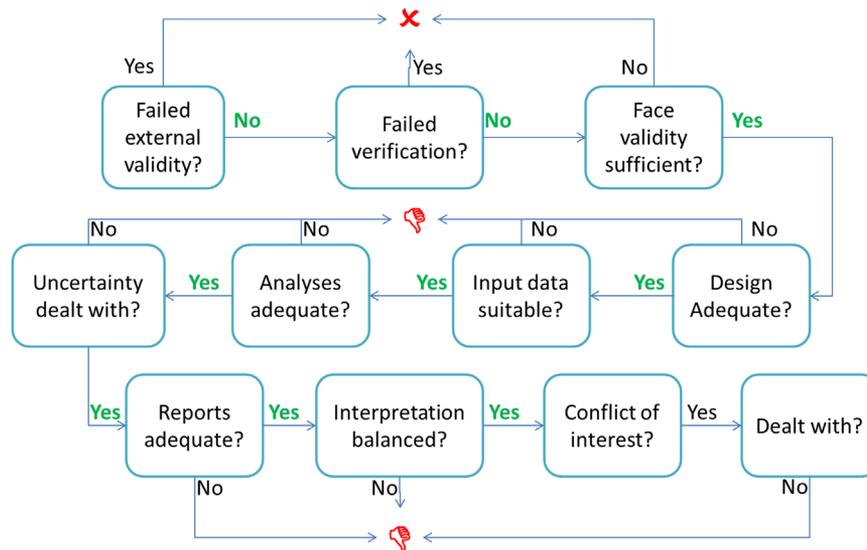
Are the clinical and patient-centered *health outcomes* assessed in the study meaningful to the decision maker? Have all relevant *economic end points* been addressed? Outcomes that matter to patients or health care systems are increasingly emphasized instead of surrogate end points (e.g., cholesterol levels). These aspects are addressed in two helper questions.

#### 4. Is the context (settings and circumstances) applicable?

Is the context of the study applicable to the decision setting? Factors to consider are the study *time horizon*; characteristics of the *health care system*, including the *cost structure*; *analytic perspective*; and practices related to the *geographic location*. These are covered in five helper questions.

The time horizon refers to the period of time for which the simulation is run. Typically, this is defined in terms of the patients simulated rather than calendar time (e.g., everyone was simulated until death—a lifetime horizon). The time horizon should match the time horizon of interest to decision makers, but, ideally, should not be so short that important outcomes cannot manifest.

The perspective is the point of view taken in the analysis. Although a strict economic principle holds that all analyses should be from the so-called societal perspective (include all economic consequences across the community), nearly all analyses focus on factors relevant to stakeholders. For example, from a health care plan's perspective, the cost of care and expected patient outcomes are considered but other economic consequences such as worker productivity, absenteeism, or presenteeism may not be taken into account. A narrower perspective should be



**Fig. 1 – Modeling study assessment questionnaire flowchart. The cross-out mark indicates a fatal flaw and the thumbs down an unfavorable response. Of note, all questions provide a “can’t answer” option that is not reflected in the diagram. Thus, insufficient information does not necessarily trigger an unfavorable response or fatal flaw.**

clearly described and justified, so that users can judge how relevant it is in the decision setting.

### Credibility

The credibility of a modeling analysis should be assessed at several levels: validation, design, data, analyses, reporting, interpretation, and conflicts of interest. Validation assesses how well the model accords with reality. The design should follow accepted standards for conceptualizing and framing the model. The data used in building the model should be suitable for the purpose, properly analyzed, and incorporated in the model. Analyses should provide the information required to support the decision maker. Although not specifically pertaining to a model’s credibility, its reports should be sufficient to assess relevance and credibility. To evaluate the degree to which credibility has been properly assessed (and what the level of accuracy is), adequate documentation of the model and of the validity assessments must be provided. Although users are free to reach their own conclusions, it is helpful if interpretations are balanced and consistent with the results, keeping in mind the limitations of the model and data. Finally, any potential conflicts of interest that might have affected the modeling study and what the researchers have done about them should be described.

The flow of questions in this section is shown in [Figure 1](#).

### Validation

Validation is a process for judging a model’s accuracy in making relevant predictions [15]. That information can be used by decision makers to determine to what extent they should trust the results. While transparency can help users understand what a model does and how it does it, validation is the only way for readers to determine how well it does it. It is very difficult for anyone other than the modelers to assess validity fully. Because models use mathematical structures to make their predictions, directly making the validity assessments requires technical expertise and full access to the model and external data. A more feasible task for users of models is to evaluate the extent to which the modelers have assessed validity (bearing in mind that validity pertains to a particular application of the model—it is not

a property of the model per se, and the degree of validity required depends on the question posed). It should be noted that peer review, as currently practiced, is insufficient to determine that a model’s results are fully validated. Just because a modeling analysis has been published does not mean that it is credible—much depends on the quality of the peer review process. If a model has not been sufficiently validated, or the decision maker cannot tell this information, then the results of the model should not be trusted (i.e., this is a fatal flaw).

Points to think about in general:

- Did the model builders have a formal process for validating their model? [10]
- Has a report of the validation been made available?
- Is the validation process well described?
- Have the types of validations performed been detailed?
- Were the approaches to finding data sources for the validation reasonably comprehensive?
- Were the data sources used for validation appropriate for the proposed uses of the model?
- Were the methods for setting up the simulation of each source adequately described? Do they seem reasonable? For example, how well was the simulated population matched to the validation one?
- Were those performing the validations blind to the results of the model?
- Were results of the validations provided in sufficient detail?
- Were the implications of the validations discussed adequately?
- Were there quantitative measures of how well the model’s results match the outcomes observed in the data source?

There are three main categories: external validity, internal verification, and face validity. The questionnaire includes one main question and several helper questions for each of these.

### External Validation

1. *Is external validation of the model sufficient to make its results credible for your decision?*

In external validation, a model is used to simulate a real data source, such as a clinical trial, and the calculated outcomes are

compared with the real-world ones [28]. External validation can be applied to the model as a whole, or to parts of the model that address such things as creation of simulated populations, disease incidence (including effects of patient characteristics, risk factors, and behaviors), disease progression, care processes and behaviors, occurrence of clinical outcomes, and effects of interventions. External validation is essential in establishing the credibility of the model, and a “No” answer to this question should be considered a fatal flaw. Helper questions address types of external validation. There are three main types: dependent (including partially dependent), independent, and predictive.

When reviewing external validations, important questions to ask include such things as how many were performed; were landmark epidemiological studies and clinical trials included; were all relevant arms of a trial included; were all relevant outcomes included; and was there an attempt to “cherry pick” the studies, treatment arms, or outcomes. It is important that the sources used for validation provide a good check of the model’s credibility. For example, although it may appear impressive that a model reproduces each one of the data sources used to build it, this only reflects that the data sources were correctly implemented. It does not assure you that the model can accurately estimate results beyond that.

**Dependent validation.** This type of validation involves simulating a data source that was used to build the model or parts of the model that are being validated [29]. It is considered “dependent” because the same source is used to both estimate and validate the model’s equations. For example, a model might use data from a clinical trial to estimate disease progression. The modelers would simulate the same trial to confirm that their results are in accord. Every model should, at a minimum, be able to reproduce the data sources that were used to create it. Models validated only by dependent validations are not as credible as those validated by independent or predictive validations but can still be useful.

In partially dependent validations, the model is used to simulate a data source used for one component (the dependent part), but this simulation covers other components as well (the “independent” part) [30]. For example, a clinical trial might be used to estimate a single equation for mortality but other simulation components are derived from other sources. If simulating this trial yields results that are in accord, then a fairly strong partially dependent validation has been attained.

**Independent validation.** In an independent validation, the model is used to simulate a study that was not used in any way to build the model [31]. Ideally, the independent sources are chosen by an independent panel on the basis of the intended use of the model, not on convenience or likelihood of a successful validation. Data sources for independent validations should be identified through a formal search to involve settings, populations, interventions, and outcomes similar to those in the decision setting. They should also have strong designs (e.g., large size, representative population, formal protocol, and detailed reporting) and be as recent as possible.

**Predictive validation.** Predictive validations involve identifying a data source, such as a clinical trial, that has not yet been completed, and predicting the results before they are known [32]. This type of validation is very convincing because it eliminates any opportunity for the modelers to be affected by the results. Opportunities for this type of validation are relatively rare because there may not be any suitable studies in progress or the model’s results are needed before the studies will be completed. While predictive validation enhances the credibility of a model, absence of predictive validation should not negatively affect its credibility.

## Verification

### 2. Is internal verification of the model sufficient to make its results credible for your decision?

Verification (also called internal validity, internal consistency, or technical validation) examines the extent to which the mathematical calculations are performed correctly in the model and are consistent with the model’s specifications [29]. The methods will depend on the model’s complexity. There are two main steps: verifying the individual equations, and verifying their implementation in code. Equations and parameters should be validated against their sources. Verification [15] helps to ensure that there are no unintentional computational errors, but it does not evaluate the accuracy of the model’s structure or predictions. Parameters for the equations might be fitted using good data sources and technique, and the equations might be accurately coded, but the resulting model might still be inaccurate if the structure is poorly chosen. Helper questions address the following: Has the process of internal verification and its results been documented? Has the testing been performed systematically? Does the testing indicate that the equations are consistent with their data sources? Does the testing indicate that the coding has been correctly implemented?

## Face Validity

### 3. Does the model have sufficient face validity to make its results credible for your decision?

Face validity of the model addresses how plausibly the model represents the diseases, settings, populations, interventions, and outcomes it is intended to analyze [33]. This is the easiest aspect of credibility for a user to check because it does not require in-depth technical knowhow. Helper questions address whether the model contains all the aspects of the population, diseases, interventions, outcomes, and setting appropriate to its intended use; all the aspects are represented and linked according to current knowledge; assumptions made about the occurrence and progression of diseases are sufficiently realistic for the intended uses; the best available data sources were used; the time horizon was sufficiently long to account for all relevant aspects of the decision problem; and the results were plausible or explainable. If others rated the face validity, did they have a stake in the results that might bias their conclusions?

If parts of the model fail face validity, the effect on credibility depends on the user’s judgment about whether the questionable parts are so unrealistic or inappropriate that they affect the accuracy of the results. For example, a state-transition model might assume that everyone with, say, osteoporosis will respond equally to treatment, no matter how long they have had the condition, how many previous fractures they have had, their bone mineral density, and so forth. If that model is to be used to evaluate osteoporosis treatment, those assumptions would have a strong negative effect on face validity. If the only purpose of the model is to estimate the age-specific incidence of osteoporosis, however, then assumptions about treatment, no matter how inaccurate they may be, will not materially affect the results.

It should be noted that the model may contain additional aspects beyond the ones relevant to the problem at hand. Users may evaluate the face validity of only those parts of the model that are pertinent to their decision problem.

## Design

### 4. Is the design of the model adequate for your decision problem?

The *design* of the model refers to the aspects of the decision problem that are included, the relationships between them, and

the mathematical structure used to represent them [10]. “Aspects” is used very broadly to include such things as physiological parameters, treatments, cure processes, behaviors, health outcomes, costs, environmental factors, and so forth. The mathematical structure determines how these aspects and their relationships are represented mathematically. The structure strongly affects such things as whether the model is continuous in time [34], whether variables that are continuous in reality are represented as such in the model [35] (e.g., is blood pressure measured continuously or is it represented through categories such as “normal,” mild hypertension,” etc.); whether diseases are assumed to be either absent or present, or modeled as progressing continuously in severity; whether the effects of multiple treatments are multiplied or their interactions modeled in a more physiologically realistic way; how comorbidities are handled; how the effects of timing of treatments are handled; and so forth.

There are several things to consider in assessing the credibility of the model’s design. The first is a clear statement of the decision problem(s) the model is intended to address. This must be provided in sufficient detail to enable the user to evaluate not only the credibility of the design but also its relevance to the decision.

The second is the process used to determine the design (parameters, interactions, mathematical structure). There should be a formal process, and it should be well described. An informal process (i.e., mostly in the modelers’ heads), is not only more difficult to evaluate but also more subject to errors such as omission of important aspects and/or relationships, and use of mathematical structures that have unrecognized limitations. The process should begin with a description of the desired qualities of the model—what the model should be able to do. It is important that these qualities be identified before the structure or data sources are chosen [10]. If the mathematical structure or data sources are chosen first, there is a high risk that the resulting model will be distorted to fit the preselected structure or data sources. When building models, it is always necessary to compromise on the ideal list of qualities because of deficiencies in the available data, but it is very important to begin with a list of qualities. This helps ensure that there is a rigorous search for the best possible data sources and the model comes as close as possible to having the desired qualities. The process for designing a model should include graphical aids such as influence diagrams, concept maps, and flow diagrams.

The third is to examine the aspects and relationships and determine whether they are sufficient to address the intended problems. It is not necessary that the model include every physiological variable and other factors that might possibly affect an outcome. The issue is whether it includes the most important determinants of the outcome. Practical questions might take the form “Would I believe a model’s estimate of the risk of neuropathy complications if the model includes ‘diabetes,’ but does not include its duration, severity, treatment, or response to treatment?” Or “Would I believe a model’s predictions of the spread of infection in a population if it assumes the population is evenly spread across a geographical area, and there are no areas of crowding?” The choices made regarding what to include and how create considerable uncertainty in the results. This structural uncertainty is difficult to address but, at a minimum, the modelers should identify what these elements are and address how they might affect the results [36].

The fourth is to review the strengths and limitations of the chosen modeling technique [37–40]. It is important that the model be designed at an appropriate level of detail to represent the intended decision problem. All models are simplified representations of reality, but the simplifications must not be so severe that they cause the model to produce misleading results.

It is also important that the model’s design be based on medical science and the decision problem, rather than a prechosen modeling technique (e.g., regression equation, state-transition model, and event-driven simulation). All modeling techniques are subject to some inherent assumptions. As with most aspects of model evaluations, the importance of any particular assumption will depend on the intended uses—the key is whether the chosen technique is appropriate for the intended uses. For example, the assumption that time jumps forward in annual intervals will have little detrimental effect on a model’s ability to calculate annual incidence rates, but would critically affect its ability to analyze issues such as frequency of screening, delay in diagnosis, optimal follow-up times for monitoring cancer posttreatment, and so forth. Several of the more common types of modeling techniques and their strengths, assumptions, and limitations are described in a previous series of articles [23–25].

### Data

5. *Are the data used in populating the model suitable for your decision problem?*

The data used to estimate equations for the model should be suitable for the decision problem. The user can assess this by asking whether the values used for the inputs seem reasonable. Users with quantitative training can assess whether adequate methods were used for point estimate and interval estimation, such as 95% CIs, or distributions [41]; extrapolating beyond the observed data [42]; developing the model equations; incorporating correlation among parameters [16].

### Analysis

6. *Were the analyses adequate to inform your decision problem?*

7. *Was there an adequate assessment of the effects of uncertainty?* The analyses performed using the model must be adequate to inform the decision problem. Relevant factors include the specification of the simulated population, interventions, and outcomes; number of simulated individuals included in the calculation (addressing stochastic uncertainty); and time horizon. Analysis of the results should also include calculating uncertainty due to imprecision in parameter estimates [16] and structural uncertainty generated by the assumptions made in the model [43]. These uncertainty analyses are often known as “sensitivity” analyses.

### Reporting

8. *Was the reporting of the model adequate to inform your decision problem?*

Adequate documentation of the model and its validation must be provided [15,44]. While inadequate documentation does not necessarily mean that the model is inaccurate, poorly documented models should not be considered credible. Documentation should include a freely available nontechnical description of the model that covers the type of model and intended applications; its structure, including the aspects considered and their relationships; modeling technique; assumptions; data sources; validation methods and results; limitations; and funding sources.

In addition, modelers should make available (under agreements that protect intellectual property, if they wish) full technical documentation in sufficient detail to allow for replication of the model.

### Interpretation

9. *Was the interpretation of results fair and balanced?*

The interpretation of the model’s results should be fair and justified by the results. Limitations should be discussed. Results

should be placed in context of any applicable empirical studies (e.g., clinical trials) or other models. Specific questions include the following: Were the results consistent with previous known information? Are the results (differences demonstrated) considered clinically meaningful? Have the implications of the validation results been adequately considered?

### Conflict of Interest

10. Were there any potential conflicts of interest?
11. If there were potential conflicts of interest, were steps taken to address these? Finally, the user should assess the potential effects of any conflicts of interest. Specific questions include the following: Did the authors describe how the study was funded? Did the sponsor have a role in the design, execution, or interpretation of the study? If there were potential conflicts of interest, were steps taken to address these?

### Final Assessment

Upon completion of the questions, each of the domains is rated by the user as a “strength,” a “weakness,” or “neutral.” On the basis of these evaluations, the user would then rate the credibility of the research study as either “sufficient” or “insufficient” to inform decision making.

Some questions may be identified as a “fatal flaw.” This means that there is a significant risk that the findings are misleading and the decision maker should use caution in applying the results to inform decisions. The occurrence of a fatal flaw does not prevent a user from completing the questionnaire, however. The questionnaire also does not force the user to judge the evidence as “insufficient” but the presence of a fatal flaw should raise a strong caution and should be carefully considered when the overall body of evidence is reviewed.

Once the user has completed the questionnaire, the assessments are summarized as follows:

*In evaluating this study, I made the following judgments:*

- I found the study (relevant/not relevant) for decision making because I considered that the population/interventions/outcomes/setting (applied/did not apply) to the decision I am informing.
- I found the study (credible/not credible) for decision making because:
  - There (were/were not any) fatal flaws—that is, critical elements that call into question the validity of the findings.
    - The presence of a fatal flaw suggests significant opportunities for the findings to be misleading and misinterpreted; extreme caution should be used in applying the findings to inform decisions.
    - The following domains contained fatal flaws:
  - There are strengths and weakness in the model:
    - The following domains were evaluated as strengths:
    - The following domains were evaluated as weaknesses:

### Meeting Educational Needs

Across jurisdictions, the resources and expertise available to inform health care decision makers vary widely. While there is broad experience in evaluating evidence from randomized controlled clinical trials, there is less experience with modeling studies. ISPOR has provided Good Research Practice recommendations on modeling studies [3,10,15,16,23–25]. This questionnaire is an extension of those recommendations and serves as a platform to assist the decision maker in understanding what a systematic evaluation of modeling study requires. It is hoped that this will lead to a general increase in the understanding and

application of evidence generated by modeling studies. To that end, we anticipate additional educational efforts and promotion of this questionnaire and that it will be made broadly available. In addition, an interactive Web-based tool has been developed at <https://www.healthstudyassessment.org/> to aid and facilitate uptake and support the educational goal that the questionnaire provides.

### Conclusions

The Task Force developed a consensus-based questionnaire to help decision makers assess the relevance and credibility of modeling studies. The questionnaire aims to provide a guide for assessing the degree of confidence that should be placed in a modeling analysis, and increases awareness among decision makers of the subtleties involved in evaluating these kinds of studies. It is anticipated that user feedback will permit periodic evaluation and modification to the questionnaire. The goal is to make all the questionnaires in the series as useful as possible to the health care decision-making community.

### Acknowledgments

The members of the Task Force thank the three societies (ISPOR, AMPC, and NPC) for coming together to support this endeavor. Maria Swift, Rebecca Corey, and Danielle Mroz from ISPOR were instrumental in keeping the Task Force moving forward. We are also very appreciative of the efforts of the many people who agreed to review the article and who contributed greatly with their thoughtful comments.

Source of financial support: The authors have no other financial relationships to disclose.

### REFERENCES

- [1] Berger M, Martin B, Husereau D, et al, A questionnaire to assess the relevance and credibility of observational studies to inform health care decision making: an ISPOR-AMCP-NPC good practice task force report. *Value Health* 2014;17:143–156.
- [2] Jansen J, Trikalinos TA, Cappelleri JP, et al, Indirect treatment comparison/network meta-analysis study questionnaire to assess study relevance and credibility to inform healthcare decision-making: an ISPOR-AMCP-NPC Good Practice Task Force report. *Value Health* 2014;17:157–173.
- [3] 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;15:796–803.
- [4] Petrou S, Grey A. Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. *BMJ* 2011;342:1766–72.
- [5] Caro JJ, Payne K. Editorial: disease simulation models and health care decisions. *CMAJ* 2000;162:1001–2.
- [6] Watkins JB, Minshall ME, Sullivan SD. Application of economic analyses in U.S. managed care formulary decisions: a private payer's experience. *J Manag Care Pharm* 2006;12:726–35.
- [7] Siebert U. When should decision-analytic modeling be used in the economic evaluation of health care? *Eur J Health Econ* 2003;4:143–50.
- [8] Buxton M, Drummond M, Van Hout B, et al, Modelling in economic evaluation: an unavoidable fact of life. *Health Econ* 1997;6:217–27.
- [9] Evans C, Dukes EM, Crawford B. The role of pharmacoeconomic information in the formulary decision-making process. *J Managed Care Pharm* 2000;6:112–21.
- [10] 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;15:804–11.
- [11] Johnson M, Crown W, Martin B, et al, Good research practices for comparative effectiveness research: analytic methods to improve causal inference from non-randomized studies of treatment effects using secondary data sources. Report of the ISPOR Retrospective Database Analysis Task Force: part III. *Value Health* 2009;12:1062–73.

- [12] Menn P, Holle R. Comparing three software tools for implementing Markov models for health economic evaluations. *Pharmacoeconomics* 2009;27:745–53.
- [13] 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. *Value Health* 2010;A192.
- [14] Eddy DM. Accuracy versus transparency in pharmacoeconomic modelling: finding the right balance. *Pharmacoeconomics* 2006;24:837–44.
- [15] Eddy DM, Hollingworth W, Caro JJ, et al, Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. *Value Health* 2012;15:843–50.
- [16] Briggs AH, Weinstein MC, Fenwick E, et al, Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. *Value Health* 2012;15:835–42.
- [17] 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.
- [18] Wang A, Halbert RJ, Baerwaldt T, Nordyke RJ. Payer perspectives on evidence for formulary decision-making. *Am J Manag Care* 2012;18:SP71–6.
- [19] Olson BM, Armstrong EP, Grizzle AJ, Nichter MA. Industry's perception of presenting pharmacoeconomic models to managed care organizations. *J Managed Care Pharm* 2003;2:159–67.
- [20] Lee RC, Donaldson C, Cook LS. The need for evolution in healthcare decision modeling. *Med Care* 2003;41:1024–33.
- [21] 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.
- [22] Philips Z, Bojke L, Sculpher M, et al, Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. *Pharmacoeconomics* 2006;24:355–71.
- [23] 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;15:812–20.
- [24] Karnon J, Stahl JE, 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;15:821–7.
- [25] Pitman RJ, 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;15:828–34.
- [26] Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282:1054–60.
- [27] Sculpher MJ, Pang FS, Manca A, et al, Generalisability in economic evaluation studies in healthcare: a review and case studies. *Health Technol Assess* 2004;49:1–192.
- [28] Pagano E, Gray A, Rosato R, et al, Prediction of mortality and macrovascular complications in type 2 diabetes: validation of the UKPDS Outcomes Model in the Casale Monferrato Survey, Italy. *Diabetologia* 2013;56:1726–34.
- [29] van Kempen BJ, Ferket BS, Hofman A, et al, Validation of a model to investigate the effects of modifying cardiovascular disease (CVD) risk factors on the burden of CVD: the Rotterdam ischemic heart disease and stroke computer simulation (RISC) model. *BMC Med* 2012;10:158–71.
- [30] Simpson KN, Pei PP, Möller J, et al, Lopinavir/ritonavir versus darunavir plus ritonavir for HIV infection: a cost-effectiveness analysis for the United States. *Pharmacoeconomics* 2013;31:427–44.
- [31] 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;16:1007–21.
- [32] Eddy DM, Schlessinger L. Validation of the Archimedes diabetes model. *Diabetes Care* 2003;26:3102–10.
- [33] Hammerschmidt T, Goertz A, Wagenpfeil S, et al, Validation of health economic models: the example of EVITA. *Value Health* 2003;6:551–9.
- [34] Soares MO, Canto E, Castro L. Continuous time simulation and discretized models for cost-effectiveness analysis. *Pharmacoeconomics* 2012;30:1101–17.
- [35] Bentley TGK, Weinstein MC, Kuntz KM. Effects of categorizing continuous variables in decision-analytic models. *Med Decis Making* 2009;29:549–56.
- [36] Kim LG, Thompson SG. Uncertainty and validation of health economic decision models. *Health Econ* 2010;19:43–55.
- [37] Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ* 2006;15:1295–310.
- [38] Stahl JE. Modelling methods for pharmacoeconomics and health technology assessment: an overview and guide. *Pharmacoeconomics* 2008;26:131–48.
- [39] Caro JJ. Pharmacoeconomic analyses using discrete event simulation. *Pharmacoeconomics* 2005;23:323–32.
- [40] Borshchev A, Filippov A. From system dynamics and discrete event to practical agent based modeling: reasons, techniques, tools. *Simulation* 2004. Available from: <http://www2.econ.iastate.edu/tesfatsi/systemdyndiscreteeventabmcompared.borshchevfilippov04.pdf>. [Accessed February 12, 2014].
- [41] Law AM. How to select simulation input probability distributions. Presented at Proceedings of the Winter Simulation Conference, Phoenix, AZ, December 11–14, 2011. Available from: <http://www.proceedings.com/13970.html>. [Accessed February 12, 2014].
- [42] Davies C, Briggs A, Lorgelly P, et al, The “hazards” of extrapolating survival curves. *Med Decis Making* 2013;33:369–80.
- [43] Bojke L, Claxton K, Sculpher M, Palmer S. Characterizing structural uncertainty in decision analytic models: a review and application of methods. *Value Health* 2009;12:739–49.
- [44] Husereau D, Drummond M, Petrou S, et al, ISPOR Health Economic Evaluation Publication Guidelines-CHEERS Good Reporting Practices Task Force. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)—explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. *Value Health* 2013;16:231–50.