VALUE IN HEALTH 15 (2012) 843–850

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Model Transparency and Validation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7

David M. Eddy, PhD, MD[1,], William Hollingworth, PhD[2], J. Jaime Caro, MDCM[3,4], Joel Tsevat, MD, MPH[5], Kathryn M. McDonald, MM[6], John B. Wong, MD[7], on Behalf of the ISPOR–SMDM Modeling Good Research Practices Task Force*

[1]*Archimedes, Inc., San Francisco, CA, USA;* [2]*School of Social and Community Medicine, University of Bristol, Bristol, UK;* [3]*United BioSource Corporation, Lexington, MA, USA;* [4]*Faculty of Medicine, McGill University, Montreal, QC, Canada;* [5]*College of Medicine University of Cincinnati, and Cincinnati Veterans Affairs Medical Center, Cincinnati, OH, USA;* [6]*Center for Health Policy/Center for Primary Care and Outcomes Research, Stanford, CA, USA;* [7]*Division of Clinical Decision Making, Informatics and Telemedicine, Tufts Medical Center, and Tufts University School of Medicine, Boston, MA, USA*

A B S T R A C T

Trust and confidence are critical to the success of health care models. There are two main methods for achieving this: transparency (people can see how the model is built) and validation (how well the model reproduces reality). This report describes recommendations for achieving transparency and validation developed by a taskforce appointed by the International Society for Pharmacoeconomics and Outcomes Research and the Society for Medical Decision Making. Recommendations were developed iteratively by the authors. A nontechnical description—including model type, intended applications, funding sources, structure, intended uses, inputs, outputs, other components that determine function, and their relationships, data sources, validation methods, results, and limitations—should be made available to anyone. Technical documentation, written in sufficient detail to enable a reader with necessary expertise to evaluate the model and potentially reproduce it, should be made available openly or under agreements that protect intellectual property, at the discretion of the modelers. Validation involves face validity (wherein experts evaluate model structure, data sources, assumptions, and results), verification or internal validity (check accuracy of coding), cross validity (comparison of results with other models analyzing the same problem), external validity (comparing model results with real-world results), and predictive validity (comparing model results with prospectively observed events). The last two are the strongest form of validation. Each section of this article contains a number of recommendations that were iterated among the authors, as well as among the wider modeling taskforce, jointly set up by the International Society for Pharmacoeconomics and Outcomes Research and the Society for Medical Decision Making.

Keywords: decision making, best practices, modeling, simulation, transparency, validation.

Background to the Task Force

A new Good Research Practices in Modeling Task Force was approved by the ISPOR Board of Directors in 2010, and the Society for Medical Decision Making was invited to join the effort. The Task Force cochairs and members are expert developers and experienced model users from academia, industry, and government, with representation from many countries. Several teleconferences and hosted information sessions during scientific meetings of the Societies culminated in an in-person meeting of the Task Force as a whole, held in Boston in March 2011. Draft recommendations were discussed 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 each recommendation, and if the latter, to provide the reasons. Each group received the results of the survey and endeavored to address all issues. The final drafts of the seven articles were available on the ISPOR and Society for Medical Decision Making Web sites for general comment. A second group of experts was invited to formally review the articles. The comments received were addressed, and the final version of each article was prepared. (A copy of the original draft article, as well as the reviewer comments and author responses, is available at the ISPOR Web site: http://www.ispor.org/workpaper/Model-Transparency-and-Validation.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*. Other articles in this series [1–6] describe best practices for conceptualizing models, building and applying particular types of models, and addressing uncertainty. This article addresses best practices for transparency and validations and is intended to apply to all types of models. Examples are cited throughout, without implying endorsement or preeminence of the articles referenced.

* *Address correspondence to:* David M. Eddy, Archimedes, Inc., 201 Mission Street, Suite 2900, San Francisco, CA 94105.
E-mail: david.eddy@archimedesmodel.com.

1098-3015/$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2012.04.012

Introduction

The purpose of health care models is to provide decision makers with quantitative information about the consequences of the options being considered. For a model to be useful for this purpose, decision makers need confidence in the model's results. Specifically, they need to know how accurately the model predicts the outcomes of interest and account for that information when deciding how to use the model results.

Modelers can impart such confidence and enhance model credibility in two main ways: 1) transparency—clearly describing the model structure, equations, parameter values, and assumptions to enable interested parties to understand the model and 2) validation—subjecting the model to tests such as comparing the model's results with events observed in reality [7–14].

Some health care models are intended to be "general" or "multiapplication" in the sense that with appropriate modifications they can address a range of problems [15,16]. For example, an "HIV model" could be used repeatedly to address different questions relating to that condition [17–21]. Other models are built for single, specific applications and are not intended to be reused [22]. For instance, a model may be built with the sole purpose of extrapolating the results of a trial of an implantable cardioverter-defibrillator to determine whether it becomes cost-effective over the lifetime of patients [23]. Some models that are initially built for a single application may later be expanded to address others. The methods described in this article apply to both types of models. For a multiapplication model, transparency, validation, and reporting are ongoing processes. The multiapplication model is described (transparency) [24] and validated [25], and the descriptions and validations are continuously updated as science and the model evolve [26]. In addition, each instantiation of the model is described, validated, and reported as each application is done [27]. For a single–application model, its description and validation, and the reporting of its application, are typically conducted at one time, although there may be additional validations after initial use, particularly if problems are found.

Our objective was to describe practices that we consider to be "best" in the sense of providing potential users of a model with the information necessary to determine their confidence in the results, and hence their application of the model's results. Every model today should be able to achieve the best practices we recommend for transparency. We recognize, however, that not all models will be able to achieve all the recommended best practices for validation. Rather than establish minimum quality standards, we have described optimal practices that all models should strive toward. For all models, their developers should describe their process for conducting validations and the level of validation their model achieved. These recommendations are particularly important in light of high-profile examples of scientific misconduct and fraudulent research published in leading scientific journals, leading to increasing emphasis on transparency and "shining a light on black boxes" [28–34].

Transparency

Transparency refers to the extent to which interested parties can review a model's structure, equations, parameter values, and assumptions. It does not refer to the formulation, conduct, or results of a particular analysis. Transparency serves two purposes: 1) to provide a non–quantitative description of the model to readers who want to understand in a general way how a model works and 2) to provide technical information to readers who want to evaluate a model at higher level of mathematical and programming detail, and possibly replicate it (the term "reader" describes anyone who needs to evaluate a model, including journal reviewers, journal readers, and users of a model's results). Taken together, the intention is to provide sufficient information to enable the full spectrum of readers to understand a model's accuracy, limitations, and potential applications at a level appropriate to their expertise and needs.

Nontechnical documentation

Nontechnical documentation should be accessible to any interested reader [35–37]. It should include descriptions of the following:

1. The model and its purpose
2. Types of applications it is designed to address (e.g., forecasting of short–term costs, cost–effectiveness analysis)
3. Sources of funding and their role
4. Structure (e.g., graphical representation of the variables and their relationships)
5. Components that define it and determine its performance
6. Inputs, outputs, and other parameters
7. Equations and their sources
8. How the data sources were identified and selected
9. Model validation and summary of results
10. Methods for customizing to specific applications and settings
11. Effects of uncertainty
12. Main limitations for its intended applications
13. Examples of actual equations (optional)
14. Reference to the model's technical documentation

The nontechnical documentation provides an overview of the model and what it does, but it may not contain sufficient information to enable readers to replicate it.

Technical documentation

Full technical transparency is achieved by providing documents that detail the model, including its structure, components, equations, and computer code. The documentation should be sufficiently detailed to enable those with the necessary expertise and resources to reproduce the model. Provision of technical documentation is subject to some conditions and limitations:

1. Access should be provided in a way that enables protection of intellectual property. Building a model can require a significant investment in time and money; if those who make such investments had to give their models away without restriction, the incentives and resources to build and maintain complex models could disappear.
2. While not mandatory, an increasing number of journals request that authors state whether full technical documentation is available to readers, and if so, under what terms [28,38]. Technical documents may be placed in appendices or made accessible by other means [28,29,31,39]. Provision of such documentation is not without concerns that the context of the original analysis may be missing [40].
3. Because most multiapplication models change over time—expanded and updated to incorporate new information and advances in health care technologies—technical documents should be updated periodically.
4. Equations and detailed structure will mean little to readers without the necessary technical background. Even with such information, reviewing a model can take considerable time. Furthermore, it is very difficult to understand how accurate a model is simply by examining its equations. Even if the equations appear to be valid in a mathematical sense and the parameters appear to be estimated using appropriate sources and methods, it is virtually impossible for anyone to determine a model's accuracy by "running" it in one's head. Providing the code does not solve this problem unless the reader has the time and resources to actually implement it, which for large models