

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

The View of a Model user on the ISPOR-SMDM Modeling Good Research Task Force Report

The ISPOR Good Modeling Task Force, like other ISPOR task forces, was created to advance the field of health care outcomes research and to promote the utilization of outcomes research in making health care decisions. It is, however, unique in being the first task force to represent joint recommendations of ISPOR and the Society for Medical Decision Making. This collaboration represents an important milestone in the development of good practice recommendations. Currently, another such collaboration is ongoing—the Comparative Effectiveness Research-Collaborative Initiative—involving ISPOR, the Academy of Managed Care Pharmacy, and the National Pharmaceutical Council—aiming to develop greater uniformity and transparency in the evaluation and use of outcomes research in health care coverage and decision making through the development of a user-friendly toolkit. Such partnerships not only make use of talent residing in multiple organizations but also enlarge the credibility and practical applications of task force recommendations.

The use of models in the support of scientific endeavors is a powerful tool, especially when the phenomena under scrutiny elude direct observation. George Box in this context observed that all models are wrong, but some are useful [1]. Similarly Niels Bohr, the great builder of models within physics, quipped that it is hard to make predictions, especially about the future [2]. Within clinical or health care policy decision making, peering into the future (e.g., predicting the lifetime consequences of health care decisions) is of paramount importance. Thus, models are working hypotheses about reality that should and must be revised as more information is obtained. The Task Force is alert to the limitations of models as simplifications representing “some aspects of reality at a sufficient level of detail to inform a clinical or policy decision.” As such, all models carry an unstated caveat: that their predictions hold true only so long as their key assumptions are correct and that the models are free of important unknown confounders. We need only look back to the Office of Technology Assessment’s 1995 report, “The Effectiveness and Costs of Osteoporosis Screening and Hormone Replacement Therapy [HRT]” [3], to illustrate this point. This report’s results were very sensitive to the presumed salutary effects of HRT on cardiovascular risk. We now know that the impact of HRT on cardiovascular risk is by no means as straightforward as explored in this model. The HRT episode highlights one of the key virtues of models: they help us focus on the critical assumptions that drive the models’ results. A model is credible only to the degree that its assumptions hold water—especially the assumption that no important unknown confounders are at play.

The Task Force places a central focus on model credibility as opposed to model robustness. Model credibility can be most simply put in Bayesian terms: if the results of the model are counter-intuitive, are the results sufficiently compelling to alter one’s subsequent actions? When models confirm our intuitions, they

stimulate less curiosity; simpler models will generally be preferred in these circumstances because of their greater transparency.

Indeed, complexity always carries the potential to undermine model credibility as decision makers do not like “black boxes.” As the field has moved from the predominant use of decision-tree models to the routine application of state-transition models to the more recent adoption of discrete event simulation models, we observe that the ability of many decision makers to understand the inner working of models has fallen dramatically. The Task Force appropriately notes that “model simplicity is desirable for transparency, ease of validation and description,” but also that the model must be “sufficiently complex to answer the question at the level of detail consistent with the problem being modeled.”

While the use of progressively more complex models over the past 25 years has proved useful in many cases, this complexity has not always been justified. In numerous instances, complex models have generated the same results as “back of the envelope” calculations. One piece of evidence that supports this observation is the study demonstrating that quality-adjusting life-years mattered only in a minority of cost-effectiveness analyses [4]. It is far from clear that modelers in general investigate systematically whether or not more complex models provide significantly improved insights to decision makers, as would be necessary to justify the loss of transparency as well as the additional effort expended in developing these complex models. While complexity may be demanded by some decision makers, many others would opt for greater transparency if the sacrifice with respect to accuracy were tolerable.

This Task Force should be credited with providing a set of guidelines that are probably the best and certainly the most current among those available, to assist model users and model reviewers in the evaluation of models. As noted, these guidelines reflect current best practice and will require updating as methodological development moves forward. The Task Force’s recommendations are not a simple checklist; a level of sophistication both about the clinical issue under examination and about how models function is required to appropriately evaluate models and their results.

The Task Force appropriately recommends that models include reasonable estimates of parameters and parameter uncertainty. This issue may bear greater scrutiny than discussed in the Task Force report. While documenting credible sources for parameter estimates and the use of sensitivity analyses are *de rigueur*, the Task Force could have commented more on how sensitivity analyses could be employed to highlight to end users of models the importance of critical assumptions through an examination of how results vary when the sensitivity analysis uses a maximal range of parameter estimates. End users—especially a non-modeler reader—would benefit from seeing a model pushed to its limits. Indeed, I have often found that published models employ

ranges of parameter estimates in sensitivity analyses that, while plausible, do not explore all potential scenarios of interest. And while one-way and two-way sensitivity analyses have given way to probabilistic sensitivity analyses, the causes of transparency and reducing decision-maker uncertainty might be better served by a “best-case, worst-case” sensitivity analysis. But perhaps the most challenging conundrum for the credibility of complex models—as the Task Force notes—is the difficulty in providing an adequate sensitivity analysis for “structural uncertainty”: how are we to understand the sometimes conflicting results from different models that appear well designed and that address the same issue?

Model users are left with proxy measures for model credibility, such as the reputation of the model’s creators and the potential for conflicts of interest. In addition, model users gain confidence in a model’s output when they can manipulate the model itself through changing the input parameters and/or aspects of the model’s structure to gain a feel for how it operates. In this regard, the discussion of transparency and intellectual property within the Task Force’s report is particularly salient.

Finally, it is crucial that the community of modelers effectively educate end users—including researchers, health care decision makers, health care providers, and patients—regarding the appropriate use and limitations of models. A model is one important source of information to heed within a well-designed, deliberative, and transparent decision process. No model directly provides the

ultimate answer to our fundamental question, “What is the right thing to do?”—nor would experienced modelers claim that.

Marc L. Berger, MD
OptumInsight, New York, NY, USA

1098-3015/\$36.00 – see front matter
Copyright © 2012, International Society for
Pharmacoeconomics and Outcomes Research (ISPOR).
Published by Elsevier Inc.
<http://dx.doi.org/10.1016/j.jval.2012.04.003>

REFERENCES

- [1] Box GEP. In: Launer RL, Wilkinson, GN, eds. *Robustness in the Strategy of Scientific Model Building*. San Diego, CA: Academic Press, San Diego, CA; 1979.
- [2] Bohr N. Brainy quote. Available from: <http://www.brainyquote.com/quotes/quotes/n/nielsbohr130288.html>, [Accessed February 13, 2012].
- [3] Office of Technology Assessment. Effectiveness and costs of osteoporosis screening and hormone replacement therapy. TA1995; OTA-BP-H-144, GPO stock #052-003-01424-6. Available from: <http://www.fas.org/ota/reports/9550.pdf>. [Accessed February 13, 2011].
- [4] Chapman R, Berger M, Weinstein M, et al. When does quality-adjusting life years matter in cost-effectiveness analysis? *Health Econ* 2004;13: 429–36.