Volume 6 • Number 1 • 2003
VALUE IN HEALTH

The ISPOR Good Practice Modeling Principles—A Sensible Approach: Be Transparent, Be Reasonable

Louis P. Garrison, Jr., PhD

Pharmaceuticals Division, F. Hoffmann-La Roche Ltd.

> Decisions will be made, with or without the model.

Keeping this fact firmly in focus, the ISPOR Task Force on Good Research Practices in Modeling Studies [1] has produced a welcome and needed addition to the numerous guidelines and checklists that already abound in our field. In a nutshell, models should aim to aid decision makers—not to prove or disprove cost-effectiveness for all time or in all situations. It is generally better to base decisions on imperfect data synthesized in transparent and reasonable models than on no information at all. The good practice principles detailed by this Task Force, under the leadership of its chair Professor Milton Weinstein, provide sound guidance to the producers and users of modeling studies. And, from the perspective of one who produces economic models directly or indirectly (working with consultants) to inform and influence decision makers—the principles strike just the right balance between sensible practice and scientific rigor, while wisely avoiding the more contentious issues of sponsorship and bias. Although the aim here was clearly to produce general principles, the authors also provide some useful hints and tips along the way, thereby advancing the state of the art.

This contribution comes from one of four Good Research Practices Task Forces established by ISPOR. Their aim is scientific—to promote "good practice in health outcomes research and its use in decision making" and "to provide guidance to the researcher." The overlapping missions of the four Task Forces—on modeling studies, prospective studies, retrospective database studies, and quality-of-life outcomes assessment—were recognized from the outset. This task force has clearly met its aim, identifying good current practice while maintaining a focus on the scientific aspects of modeling.

When it comes to decision-analytic modeling, the message here is be transparent, be reasonable, be practical, and be flexible. Of these, transparency is probably the most important because shortcomings are likely and errors are possible. With transparency, the users of models have a much better chance of understanding intuitively the key factors and variables in a model and of identifying and understanding the limitations. The need for and value of transparency is widely recognized and is cited in many other guidelines.

How do these principles differ from other guidelines and checklists that have been produced? As one who has struggled to define the term "model," we should all be grateful for their useful definition of a health-care evaluation model:

> An analytic methodology that accounts for events over time and across populations, that is based on data drawn from primary and/or secondary sources, and whose purpose is to estimate the effects of an intervention on valued health consequences and costs.

Compared with other statements written by committees, they have clearly hit a home run. Indeed, the entire document is well written and clear.

One aspect of their approach that stands out is the emphasis on the "expected value of information" and its role in evaluating the reasonable bounds of models. The principles laid out here recognize the costs of gathering and assembling information in a way that has generally not been reflected in most pharmacoeconomic guidelines and checklists. The reality is that we in industry are increasingly asked by payers to justify prices, reimbursement, and formulary access at launch. For reasons of cost and incentives to reach the market quickly, our randomized clinical trials (RCTs) for licensing are limited in size to demonstrate clinical efficacy and a reasonable risk-benefit trade-off between this efficacy benefit and safety. Even the major definitive RCTs are not necessarily powered to detect clinically significant safety differences, and also generally, they are underpowered to detect resource use differences, especially those owing to safety differences. Regulatory authorities recognize (implicitly, if not explicitly) that it would be unwise to limit access to many new drugs until all side effects are fully known. It is a reasonable approach to gather this information through postmarketing surveillance. This reveals at least an intuitive under-

 1098-3015/03/$15.00/6 6–8

standing by regulators of the value of information principle. A decision can be made conditional on the information available at the time and then modified or even reversed in light of additional information, as recent drug withdrawals from the market illustrate. Pricing and reimbursement decisions for drugs also increasingly seem to be subject to a kind of conditional approval. If it is not reasonable to make the final definitive assessment of product safety at launch, is it reasonable to make a once-and-for-all decision on reimbursement and funding? These principles underscore the point that model conclusions are, at best, conditional on the input data.

The Gold et al. [2] volume first enunciated the useful "rule of reason" criterion in this field: "Any cost or outcome that is not appreciable in the context of the analysis need not be included in the analysis." In other words, if it will not affect the analysis, do not worry about collecting any additional information on it. (This reminds one of the comment on perfectionism often attributed to Warren Buffett: "If something is not worth doing, it's not worth doing well.") The value of information criterion cited by the Task Force in several places is a kind of generalization of this, but with an interesting and new twist: "A case should be made that reasonable opportunities to obtain new additional data prior to modeling have been considered." Being sensible, the Task Force says that it is sufficient to give a "heuristic argument." This is a valuable and new suggestion, to my knowledge, that reviewers should apply in evaluating the models they receive. It should not be too burdensome or costly for model developers to comment on this explicitly.

Another underappreciated feature of modeling noted by the Task Force is: "Data availability may affect choices regarding model structure." They provide a useful example in the area of disease staging, where it might be better to not use a superior prognostic index in the model if more data are available for the somewhat inferior predictor. This observation also pertains to the tendency of some modelers to begin by describing a detailed, exhaustive pathway of clinical consequences for which there are no data to make the estimates (and, more frequently in the past, to use expert panels to estimate everything). Clearly, these kinds of models are difficult to estimate and validate. Well-meaning modeling novices, especially those with a mostly clinical orientation, have a natural tendency to head off initially in this direction. Heeding the wisdom of the Task Force will, hopefully, prevent some of them from making this common error. At an early stage, modeling is an art, as some practical judgment about the trade-off between clinical realism and economic significance comes into play.

At a practical level, the Task Force provides some useful advice and tips, but, more importantly, they differ from other guidelines in their clear recognition of some of the constraints that we face and their practical implications. First, they do not necessarily assume that the latest techniques should become the standard. For example, there is still lots of room for deterministic models and sensitivity analyses despite the growing usefulness of first- and second-order Monte Carlo simulations and the like. The Task Force also rightly emphasizes the need for cross-validation and corroboration. Too frequently, we can be faulted for not attempting to interpret the results of our model in relation to existing models in the literature. Also, their suggested use of more "structural sensitivity analyses" is a good one in this context: how sensitive are the results to different structural models—not just alternative parameter estimates?

As thoughtful as this document is, it is not just broad philosophy: there are also some useful statements on the best current practice in several technical areas. For example, under data modeling there are 10 tips on issues that beginners almost immediately encounter but about which they might have trouble finding clear guidance in the literature, such as using on-treatment data for effectiveness measurement, using life tables for all-cause mortality, combining domain-specific utilities into multiattribute utility, and adjusting for inflation in alternative ways. Useful tips in other areas are that memory should not be ignored in health state transitions and that it is reasonable to use expert consensus methods in certain situations.

Although this Task Force and the other three Good Practice Task Forces are focusing on the scientific aspects of these issues, it should be clear that their proposals might have important policy implications. At this stage, it is certainly safe to say that modeling is an accepted fact of life in our field and that its inevitability and usefulness seems to be widely recognized by public and private payers. Although the evidence on the impact of Section 114 of the FDA Modernization Act of 1997 is not yet in, managed care formulary committees are increasingly requesting the kinds of evidence synthesized in the types of models consistent with these principles.

It is important to recognize that the key principle of transparency can provide some measure of

protection against the bad effects of bias and error, but it is clearly not a perfect solution. One person's weighted average is another person's black box: in other words, models do not have to be mathematically sophisticated to be hard to follow. The widespread emphasis on the systematic review and summarization of the clinical evidence is well founded. The clinical evidence is generally the driver in these models, and understanding its strength is essential for intuitively grasping the robustness of the model.

From a purely personal perspective, after 10 years of working in industry, I do not have the sense that the researchers in our shops or the consultants we work with will to go to any extreme to get a favorable result. Admittedly, the incentives are there to produce what one might call the "best plausible case" (or to be silent if the case is not strong). But plausibility is defined within the scientific standards and methods that the vast majority of us attempt to follow. By better defining and improving those standards, the Task Force will help all of us to build better models, so that decision makers can make better decisions. (As an aside, it is worth mentioning that we are users as well as producers of models: these same economic models support strategic product development decisions from the earliest stage—when only a target product profile is available—through launch. Clearly, we find them sufficiently reliable to be helpful to support major investment decisions.)

Mark Pauly [3] stated years ago ". . . as long as efforts are compensated, bias cannot be avoided." He also noted that some systems for dealing with this reality might be better than others. Transparency, adherence to professional standards and guidance, peer review, skeptical buyers, and market forces may actually provide a reasonable level protection. Decisions will be made: are we better off with limited and potentially biased information or no information?

One important policy question that the value of information question raises is: are we doing enough (as a society—or as a developed world) to gather the information and build the models to address the health-care resource allocation questions we face? Information is a public good, and the private market provision alone would tend to undersupply the optimal amount, other things being equal. Michael Drummond [4] has recently suggested that we should consider raising the standards for the clinical evidence that underlies our models—using more head-to-head trials and longer follow-up. This certainly deserves more discussion and analysis. But those of us in the pharmaceutical industry, sometimes wonder if the considerable amount of scrutiny and analysis focused on new drugs is truly efficient—when older products and practices are seldom reviewed. We all know that incentives matter, but also we are already working in a "second-best world" where the signals we receive are not always clear and well structured. It is really very difficult to demonstrate that we are spending too little or too much on these evaluations at this time given the distortions to incentives that are embedded in our health-care financing and delivery systems. Nonetheless, public payers around the world and now private payers in the United States are requesting more evaluations (and modeling). It appears that this trend will continue.

As the Task Force emphasizes, these principles are not written in stone and there are no last words on these matters. Likewise, there are many other related issues that deserve further exploration: sponsorship and bias, incentives to invest in information gathering, the role the "fourth hurdle," and so on. All in all, Professor Weinstein and his colleagues on the Task Force should be commended and thanked graciously for their substantial contribution and service to ISPOR and to all of us working in this field.

References

1 Weinstein M, 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.

2 Gold MR, Siegel JE, Russell LB, et al., editors. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996.

3 Pauly M. Guidelines for pharmacoeconomics studies. Conference paper: Policy Issues in Pharmaceutical Cost-Effectiveness Research. American Enterprise Institute, Washington, D.C., 1996;Nov 20.

4 Drummond M. Time for a change in drug licensing requirements? Eur J Health Econ 2002;3:137–8.