Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices—Modeling Studies

Milton C. Weinstein, PhD,[1] Bernie O'Brien, PhD,[2] John Hornberger, MD, MS,[3] Joseph Jackson, PhD,[4] Magnus Johannesson, PhD,[5] Chris McCabe, MSc,[6] Bryan R. Luce, PhD[7]

[1]Center for Risk Analysis, Harvard School of Public Health, Boston, MA, USA, and Innovus Research, Inc., Medford, MA, USA; [2]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada; [3]Acumen, LLC, and Stanford University School of Medicine, Stanford, CA, USA; [4]Pharmaceutical Research Institute, Bristol-Myers Squibb, Princeton, NJ, USA; [5]Centre for Health Economics, Stockholm School of Economics, Stockholm, Sweden; [6]Trent Institute for Health Services Research, University of Sheffield, Sheffield, UK; [7]MEDTAP International, Bethesda, MD, USA

ABSTRACT

Objectives: Mathematical modeling is used widely in economic evaluations of pharmaceuticals and other health-care technologies. Users of models in government and the private sector need to be able to evaluate the quality of models according to scientific criteria of good practice. This report describes the consensus of a task force convened to provide modelers with guidelines for conducting and reporting modeling studies.
Methods: The task force was appointed with the advice and consent of the Board of Directors of ISPOR. Members were experienced developers or users of models, worked in academia and industry, and came from several countries in North America and Europe. The task force met on three occasions, conducted frequent correspondence and exchanges of drafts by electronic mail, and solicited comments on three drafts from a core group of external reviewers and more broadly from the membership of ISPOR.
Results: Criteria for assessing the quality of models fell into three areas: model structure, data used as inputs to models, and model validation. Several major themes cut across these areas. Models and their results should be represented as aids to decision making, not as statements of scientific fact; therefore, it is inappropriate to demand that models be validated prospectively before use. However, model assumptions regarding causal structure and parameter estimates should be continually assessed against data, and models should be revised accordingly. Structural assumptions and parameter estimates should be reported clearly and explicitly, and opportunities for users to appreciate the conditional relationship between inputs and outputs should be provided through sensitivity analyses.
Conclusions: Model-based evaluations are a valuable resource for health-care decision makers. It is the responsibility of model developers to conduct modeling studies according to the best practicable standards of quality and to communicate results with adequate disclosure of assumptions and with the caveat that conclusions are conditional upon the assumptions and data on which the model is built.

Introduction

Mathematical modeling is used widely in economic evaluations of pharmaceuticals and other health care technologies. The purpose of modeling is to structure evidence on clinical and economic outcomes in a form that can help to inform decisions about clinical practices and health-care resource allocations.

Address correspondence to: Milton Weinstein, Center for Risk Analysis, Harvard School of Public Health, 718 Huntington Avenue, Boston, MA 02115. E-mail: mcw@hsph.harvard.edu

Models synthesize evidence on health consequences and costs from many different sources, including data from clinical trials, observational studies, insurance claim databases, case registries, public health statistics, and preference surveys. A model is a logical mathematical framework that permits the integration of facts and values and that links these data to outcomes that are of interest to health-care decision makers. For decisions about resource allocation, the end result of a model is often an estimate of cost per quality-adjusted life year (QALY) gained or other measure of value-for-money.

Although evidence from randomized clinical trials (RCTs) remains central to efficacy testing, taken alone it can be misleading if endpoints are not translated into measures that are valued by patients, providers, insurers, and the general public. For example, suppose that an RCT demonstrates that a treatment reduces the risk of a rare sequela of a chronic disease by 50%. Further, suppose that another trial shows that a different treatment reduces the risk of a different, more common, sequela by 10%. The latter intervention may well be more effective, and cost-effective, than the former, but a simple comparison of the trial results would not suffice. However, a model could be helpful in revealing that fact to decision makers. The comparison between the two interventions would depend on a synthesis of evidence on the incidence of the sequelae in the target population, the relative risk reductions offered by treatment, survival and quality of life with and without the sequelae, and the costs of the interventions and the medical care required to diagnose and treat the sequelae.

The value of a model lies not only in the results it generates, but also in its ability to reveal the logical connection between inputs (i.e., data and assumptions) and outputs in the form of valued consequences and costs. For this reason, a model should not be a "black box" for the end-user but be as transparent as possible, so that the logic behind its results can be grasped at an intuitive level. Also for this reason, model results should never be presented as point estimates or as unconditional claims of effectiveness or cost. Instead, the outputs of models should be represented as conditional upon the input data and assumptions, and they should include extensive sensitivity analysis to explore the effects of alternative data and assumptions on the results.

The purpose of this article is to state a consensus position of the ISPOR Task Force on Good Research Practices—Modeling Studies. Like models themselves, this position represents the best judgment of the Task Force at this time, and is subject to change as new technologies for modeling emerge, through advances in computing and analysis, and as fundamentally new dimensions of health-care technology and the environment, such as genomic or microbial resistance to drugs, become more pervasive.

Task Force Process

The Chair of the ISPOR Task Force on Good Research Practices—Modeling Studies, Milton C. Weinstein, was appointed in 2000 by the Chairman of the ISPOR Health Sciences Committee, Bryan R. Luce. The members of the Task Force were invited to participate by the Chair, with advice and consent from the ISPOR Board of Directors. We sought individuals who were experienced as developers or users of pharmacoeconomic models; who were recognized as scientific leaders in the field; who worked in academia, industry, and as advisors to governments; and who came from several countries. A reference group of ISPOR members was also identified as individuals from whom comments would be sought. The Task Force held its first meeting at the Annual North American Scientific Meeting of ISPOR in Arlington, Virginia, in May 2000. The Task Force utilized electronic mail to exchange outlines and ideas during the subsequent months. A draft report was prepared by the Chair and circulated to the Task Force members for revision and additional comment. The revised draft was circulated to the reference group, and after receiving their comments, another draft was prepared. A summary of this draft was presented at a plenary session of the Annual North American Scientific Meeting of ISPOR in Arlington, Virginia, in May 2001. Comments from the audience were incorporated into a newly revised draft, which was posted on the ISPOR Web site for general comment. The next draft was presented at the Annual European Scientific Meeting of ISPOR in Cannes, France, in November 2001, and a revised draft was posted for further comment on the ISPOR Web site. This report reflects the input from all of these sources of comment.

Model Defined

The National Research Council, in its report on the uses of microsimulation modeling for social policy, offered this definition of a simulation model: "... a replicable, objective sequence of computations used for generating estimates of quantities of concern..." [1]. We define a health-care evaluation model as 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.

As part of our working definition, we assume that cost-effectiveness models are meant to be aids to decision making. This means that their purpose is not to make unconditional claims about the consequences of interventions, but to reveal the relation between assumptions and outcomes. These assumptions include structural assumptions about causal linkages between variables; quantitative parameters

such as disease incidence and prevalence, treatment efficacy and effectiveness, survival rates, health-state utilities, utilization rates, and unit costs; and value judgments such as the nature of the consequences that are valued by decision makers. A good study based on a model makes all of these assumptions explicit and transparent and states its conclusions conditionally upon them.

Model Evaluation

Models should be used only after careful testing to ensure that the mathematical calculations are accurate and consistent with the specifications of the model (internal validity), to ensure that their inputs and outputs are consistent with available data (calibration), and to ensure that their results make sense and can be explained at an intuitive level (face validity). To the extent that different models of the same decision come to different conclusions, modelers should also be expected to explain the sources of the differences (cross-validation). The description of the model should be sufficiently detailed that the model can be replicated mathematically.

Tests of predictive validity—the ability of the model to make accurate predictions of future events—are valuable, but not absolutely essential. Since future events convey information that is not available at the time the model is developed and calibrated, a model should not be criticized for failing to predict the future. However, a good model should be susceptible to recalibration or respecification to adapt to new evidence as it becomes available. The criterion for determining whether, and to what degree, tests of predictive validity are required before model use depends on the benefits in terms of improving the model for decision making and the costs of delaying the flow of information while obtaining the additional data [2].

Assessing the Quality of Models

The remainder of this statement describes the consensus of the Task Force regarding the attributes that define a good health-care decision model. We borrow heavily from several excellent papers that propose criteria for assessing the quality of models [3–6]. The attributes are organized under the major headings of structure, data, and validation.

Structure

1. The model should be structured so that its inputs and outputs are relevant to the decision-making perspective of the economic evaluation. Both costs and health consequences should reflect the chosen decision-

making perspective. For example, if the study is meant to assist decision makers in allocating resources across a broad range of health interventions at the societal level, then the outputs of the model should be broadly applicable, and important costs and consequences for all members of the affected population should be included. If a perspective narrower than societal is used, then the report should discuss, at least qualitatively, the implications of broadening the perspective to the societal perspective.

2. The structure of the model should be consistent both with a coherent theory of the health condition being modeled and with available evidence regarding causal linkages between variables. This does not mean that all causal linkages must have been proven, as is commonly understood in tests of hypotheses by showing that the effect size is statistically significant at a generally accepted level of significance (e.g., $P < .05$). Instead, it does mean that the linkages assumed are not contradicted by available evidence and are consistent with widely accepted theories.

3. If evidence regarding structural assumptions is incomplete, and there is no universally accepted theory of disease process, then the limitations of the evidence supporting the chosen model structure should be acknowledged. If possible, sensitivity analyses using alternative model structures—for example, using alternative surrogate markers or intermediate variables—should be performed.

Items 4–8 relate to state-transition (or compartmental or Markov) models:

4. Health states may be defined to correspond to the underlying disease process, which may be unobserved or unobservable, to observed health status, or to a combination of both. For example, screening models may define health states based on underlying pathology, on clinical status, or both. However, care should be taken to avoid structural bias when interventions modify both the underlying disease and the clinical presentation, as, for example, in models of cancer screening in which cases of detected cancer may have different prognoses depending on the method or frequency of screening. In general, structural bias is avoided by modeling underlying disease states and then by calibrating outputs to data on observed clinical status.

5. When transition rates or probabilities depend

on events or states that may have been experienced in prior time periods, this dependence, or "memory," should be reflected in the model. This may be done either by incorporating clinical or treatment history in the definition of health states or by including history as a covariate in specifying the transition probabilities.

6. States should not be omitted because of lack of data. Examples might be chronic health states corresponding to uncommon adverse events or disease sequelae that are not observed within clinical trials. However, inclusion of a health state should be based on evidence consistent with recommendation 2 above.

7. Reasons to include additional subdivisions of health states may be based on their clinical importance, their relation to mortality, their relation to quality of life or patient preferences, their relation to resource costs, or any combination. Disease states that may not be considered clinically important may well be important to include separately in the model for these other reasons. Conversely, health states that are regarded as having clinical importance may be included to enhance face validity, even if they do not materially affect the model's results.

8. The cycle length of the model should be short enough so that multiple changes in pathology, symptoms, treatment decisions, or costs within a single cycle are unlikely. The choice of cycle length should be justified.

9. The structure of the model should be as simple as possible, while capturing underlying essentials of the disease process and interventions. It is not necessary to model the full complexity of a disease if the decision can be informed by a more aggregated structure, in terms of disease states or population subgroups. If simplifications are made, these should be justified on grounds that they would be unlikely to materially affect the results of the analysis. Sometimes a structural sensitivity analysis that uses a less aggregated model can provide reassurance that the simplifications do not materially affect the results.

10. Options and strategies should not be strictly limited by the availability of direct evidence from clinical trials. Neither should the range of modeled options and strategies be limited by currently accepted clinical practice. There should be a balance between including a broad range of feasible options and the need to keep the model manageable, interpretable, and evidence-based.

11. While the structure of the model should reflect the essential features of the disease and its interventions irrespective of data availability, it is expected that data availability may affect choices regarding model structure. For example, if a particular staging system has been used most frequently in clinical studies, then health states might well be defined according to that staging system even if other staging systems perform better in terms of predicting outcomes or in terms of differentiating quality of life and cost.

12. Failure to account for heterogeneity within the modeled population can lead to errors in model results. When appropriate, modeled populations should be disaggregated according to strata that have different event probabilities, quality of life, and costs. This is particularly important when recurrent event rates over time are correlated within subpopulations that have different event rates, since failure to do so can lead to biased estimates of long-term outcomes.

13. The time horizon of the model should be long enough to reflect important and valued differences between the long-run consequences and costs of alternative options and strategies. Lifetime horizons are appropriate for many models and are almost always required for models in which options have different time-varying survival rates. Shorter horizons may be justified if survival and long-term chronic sequelae do not differ among options or based on an understanding of the disease process and the effect of interventions. In any case, the lack of long-term follow-up data should not be used as a rationale for failing to extend the time horizon as long as is relevant to the decision under analysis.

Data

Our recommendations on data inputs to models are grouped into three categories: data identification, data modeling, and data incorporation.

Data identification

1. A model should not be faulted because existing data fall short of ideal standards of scientific rigor. Decisions will be made, with or without the model. To reject the model because of incomplete evidence would imply that a deci-

sion with neither the data nor the model is better than a decision with the model but without the data. With the model, the available evidence can be used in a logical way to inform the decision; without the model, an opportunity to utilize the available evidence within the logical framework will have been forgone.

2. Systematic reviews of the literature should be conducted on key model inputs. Evidence that such reviews have been done, or a justification for failing to do so based on the adequacy and generalizability of readily obtained data, should accompany the model.

3. Ranges (i.e., upper and lower bounds) should accompany base-case estimates of all input parameters for which sensitivity analyzes are performed. The choice of parameters for sensitivity analysis is a matter of judgment by the analyst, but failure to perform sensitivity analysis on a parameter whose value could be disputed leaves the conclusions open to question.

4. Specification of probability distributions for input parameters based on sampling uncertainty and/or between-study variations may be incorporated into formal probabilistic sensitivity analysis. This is not always necessary or cost-effective, however. For purposes of assessing input distributions, the preferred methodology is to use posterior distributions obtained from formal meta-analyses and Bayesian analysis, but practical considerations may lead to the use of expert judgment (see item 7 below).

5. If known data sources are excluded from consideration in estimating parameters, the exclusion should be justified.

6. Data sources and results should not be rejected solely because they do not reach generally accepted probability thresholds defining "statistical significance" (e.g., $P > .05$). All evidence, even if insufficient to rule out randomness as a cause, may be legitimately incorporated into models. This is subject to the proviso that uncertainty about the estimates is disclosed and tested in sensitivity analyses and that conclusions are clearly framed as conditional upon the input estimates used.

7. Expert opinion is a legitimate method for assessing parameters, provided either that these parameters are shown not to affect the results importantly or that a sensitivity analysis is reported on these parameters with a clear statement that results are conditional upon this (these) subjective estimate(s). If expert opinion is elicited, and the results are sensitive to the elicitations, then the process of elicitation should be disclosed in detail. Expert estimates derived from formal methods such as Delphi or Nominal Group techniques are preferred.

8. A case should be made that reasonable opportunities to obtain new additional data prior to modeling have been considered. "Reasonable" in this context means that the cost and delay inherent in obtaining the data are justified by the expected value of the new information in the analysis. While formal methods of assessing value of information exist, it is sufficient to give a heuristic argument as to why the current body of evidence was optimal from the point of view of informing current decisions. This can often be accomplished using sensitivity analysis, to show that reasonable ranges of data would lead to qualitatively similar findings, or by arguing that the cost and delay in obtaining the data are not worth the forgone benefits of acting on current evidence.

Data modeling

1. Data modeling refers to the mathematical steps that are taken to transform empirical observations into a form that is useful for decision modeling. Examples include:

    a. The method for incorporating estimates of treatment effectiveness from clinical trials with estimates of baseline outcomes from epidemiologic or public health data. Effectiveness estimates may be based either on intention-to-treat or on-treatment data, depending on the objectives of the analysis. Often, an appropriate approach is to derive estimates of relative risk (or odds ratios) between treatment options from clinical trials and to superimpose these on estimates of baseline (e.g., untreated or with conventional treatment) probabilities of survival or other endpoints from population-based sources.

    b. The method for transforming interval probabilities from the literature or from a clinical trial into an instantaneous rate and then into a transition probability or event probability corresponding to the time interval used in the model.

    c. The method for combining disease-specific and all-cause mortality into the model. In general, it is acceptable to derive all-cause mortality probabilities from national life tables, unless an alternative source can be justified. In general, it is not necessary to

correct for the fact that all-cause mortality includes disease-specific mortality in the general population, unless the disease represents a major cause of death in the demographic groups being modeled.

d. The method for modeling survival (e.g., as an exponential, gamma, Weibull, or Gompertz distribution). The choice of functional form for disease-specific mortality should be specified and justified. In general, all-cause mortality should be modeled nonparametrically based on life table data.

e. Modeling risk factors or interventions as having an additive or multiplicative effect on baseline probabilities or rates of disease incidence or mortality. Evidence supporting either the additive or the multiplicative form should be sought from studies that examine the effect of the risk factor or intervention in a population stratified by base risk.

f. The method for combining domain-specific utilities into a multiattribute utility function. It is preferable to use validated health-related quality-of-life instruments with prespecified scoring systems based on "forced-choice" methods (standard gamble, time trade-off).

g. The method for transforming health status values (such as rating scales or health-state classifications) into quality-of-life weights.

h. The method for transforming charges to costs.

i. The method for adjusting for inflation or purchasing power across time and among countries. Adjustment for inflation should be based on the Consumer Price Index (CPI), its health-care components, or one or more of its subcomponents such as medical care services or equipment. The choice between the general CPI and its health-care component or subcomponents depends on whether the resources being priced are better represented by the general "market basket" in the CPI or by the health-care market basket. A limitation of the health-care CPI is that it reflects not only the prices but also to some degree the quantities of input resources used to produce health-care services. The method of choice for making adjustments across countries is to use purchasing power parity. However, a simple currency conversion would be appropriate if there is an international market for an input at a fixed price.

j. The method for discounting costs and health effects to present value.

2. Data modeling assumptions should be disclosed and supported by evidence of their general acceptance and, preferably, of their empirical validity. Key steps taken in developing the model should be carefully documented and recorded. Model credibility may be enhanced by showing how a model was conceived, for example, before or during a phase III or IV clinical trial, and how its structure and data inputs evolved in light of new evidence (e.g., after completion of a clinical trial) in response to subsequent discussions with clinical, regulatory, and policy experts.

3. When alternative, but equally defensible, data modeling approaches may lead to materially different results, sensitivity analyses should be performed to assess the implications of these alternatives. For example, if a model predicts smaller gains in life expectancy at older ages, but the model uses a multiplicative specification of the effect of an intervention of baseline mortality, then the alternative of an additive model should be tested. If there is stronger empirical evidence in support of one functional form, then that form should be the base case and the alternative form(s) should be tested in sensitivity analysis.

4. Data modeling methods should follow generally accepted methods of biostatistics and epidemiology. For modeling, meta-analysis is a valid and desirable approach, provided that care is taken to recognize heterogeneity among data sources. Heterogenity can be considered either by segregating estimates based on different groupings of primary studies or by estimating formal hierarchical models to combine information from heterogeneous studies.

Data incorporation

1. Measurement units, time intervals, and population characteristics should be mutually consistent throughout the model.

2. Either probabilistic (Monte Carlo, first-order) simulation or deterministic (cohort) simulation is acceptable.

3. If first-order, Monte Carlo simulation is used, evidence should be provided that the random simulation error (e.g., the standard deviation of output values per run) is appreciably smaller than the effect sizes of interest.

4. All modeling studies should include extensive sensitivity analyses of key parameters. Either deterministic (one-way and multiway) or probabilistic sensitivity analyses are appropriate.

5. When possible, sensitivity analyses within models that use Monte Carlo simulations should use fixed random number "seeds" within each sensitivity analysis, to minimize random simulation error.

6. If cohort simulation is used, sensitivity analysis may be done using probabilistic (Monte Carlo, second-order) simulation, using the specified probability distributions of parameter inputs. In specifying those parameter distributions, care should be taken to ensure that interdependence among parameters is reflected properly in the joint distribution of parameters.

7. When appropriate, and if the differences in quality-adjusted survival between alternatives are less than one cycle length, the half-cycle correction should be used to adjust time-related estimates in the model.

Validation

Our recommendations on validation of models are grouped into three categories: internal validation, between-model validation, and external validation.

Internal validation

1. Models should be subjected to thorough internal testing and "debugging." Evidence that this has been done should be provided. This process should include using null or extreme input values to test whether they produce the expected outputs. It may also include examination of the program code for syntactical errors, and tests of replication using equivalent input values.

2. Models should be calibrated against data when possible. Calibration is possible when there exist data on both model outputs and model inputs, over the time frame being modeled. Calibration data can come from national health statistics, such as aggregate and age–sex-specific numbers of deaths, hospitalizations, procedures, or resource costs. The calibration data should be from sources independent of the data used to estimate input parameters in the model. A model should not be criticized if independent calibration data do not exist. However, a model is subject to criticism if independent data suitable for validation do exist and either the model fails to produce outputs consistent with those data (or discrepancies cannot be explained) or the modeler has not examined the concordance between model outputs and such data.

3. While the source code should generally remain the property of the modeler, reasonable requests for copies of models with adequate user interface should be made available for peer review purposes, under conditions of strict security and protection of property rights.

Between-model validation

1. Models should be developed independently from one another, to permit tests of between-model corroboration (convergent validity).

2. If a model's outputs differ appreciably from published or publicly available results based on other models, the modeler should make a serious effort to explain the discrepancies. Are the discrepancies due to differences in model structure or input values?

3. Modelers should cooperate with other modelers in comparing results and articulating the reasons for discrepancies. (We applaud funding agencies that support this type of collaboration, e.g., the CISNET program of cancer modeling supported by the US National Cancer Institute.)

External and predictive validation. Models should be based on the best evidence available at the time they are built. In areas such as HIV and hyperlipidemia, early models assumed that health consequences are mediated by risk factors (CD4 cell counts, serum cholesterol). Subsequent data from some clinical trials have been found to be at variance with the estimates from initial models, while others are consistent with the model assumptions. Insights from clinical trials have led to a second generation of models in both HIV and hyperlipidemia, the estimates from which track more closely with those of the clinical trials. In HIV, this has been accomplished by incorporating antiretroviral drug resistance into treatment efficacy estimates and HIV RNA as a marker of disease virulence; in hyperlipidemia, this has been accomplished by modeling the lipid fractions LDL and HDL as risk factors. Remaining discrepancies between direct empirical evidence and model results are unexplained. Whether these relate to artifacts of clinical trial design (e.g., patient selection, treatment crossovers) or underlying biological factors (e.g., C-reactive protein and statins, immunologic recovery and antiretroviral therapy) is still unknown. Models therefore not only capture the understanding of the science at the time the model is constructed (at a time when there still might be limited long-term data on new treatment), but they can also provide a basis for contrasting and interpreting information from new studies. The ability of models to adapt to new evidence and scientific understanding should be regarded as a strength, not as a weakness, of the modeling approach.

1. Since models are intended as aids to current decision making, and since their outputs should be reported as conditional upon the input assumptions, it is not necessary that every data estimate or structural assumption be tested in prospective studies, in advance of model use.

2. The decision to obtain additional data to inform a model should be based on a balance between the expected value of the additional information and the cost of the information.
   a. The "expected value of information" refers to the decision-theoretic concept which values information in terms of its expected (or average) effect on the consequences of decisions. For example, the expected value of information would be zero for a study of a model parameter whose prior range does not include the threshold for the choice among decision options. Judgment concerning prior probabilities of possible study results is inevitably part of the assessment of "expected value of information."
   b. The "cost of the information" includes the resource cost of performing an empirical study or trial, as well as the expected forgone benefits of delaying decisions until the study or trial is completed. Judgment concerning prior probabilities of treatment effects is inevitably part of the assessment of cost of information.
   c. Recommendations for the conduct or design of research investigations to guide future decision making can be based on formal analysis of the value of information or on informal interpretation of the implications of sensitivity analyses.

3. Models should never be regarded as complete or immutable. They should be repeatedly updated, and sometimes abandoned and replaced, as new evidence becomes available to inform their structure or input values. As a corollary, models that have been shown to be inconsistent with subsequent evidence, but that have not been revised to calibrate against or incorporate this new evidence, should be abandoned until such recalibration has been accomplished.

Concluding Comments

While these guidelines represent the views of this Task Force at this time, they should not be regarded as rigid or cast in stone. This is not a "rule book." Different circumstances will lead to deviations from these guidelines, depending on resources available to the modeler (time, money, and data) and on the purpose of the model.

In our view, the most important thing to keep in mind in evaluating a health-care evaluation model is that its outputs must not be regarded as claims about the facts or as predictions about the future. Rather, its purpose is to synthesize evidence and assumptions in a way that allows end users to gain insight into the implications of those inputs for valued consequences and costs. Its outputs are always contingent on its inputs, which is why it is so important that its inputs be as transparent and accessible as is practical.

Further Reading on Modeling Methodology

The purpose of this report is not to provide an overview of modeling methodology, but rather to identify those aspects of methodology that the Task Force regards as good research practice. We recommend the following sources for readers who wish to acquaint themselves with the basics of modeling methods. For an introductory textbook on decision analysis, including decision trees and Markov models, see Hunink et al. [7]. For contemporary methods of modeling in economic evaluations, including an overview of methods for modeling survival from trial data, and an overview of deterministic and stochastic approaches to modeling, see Kuntz and Weinstein [8]. For an overview of methods for handling uncertainty in models, see Briggs [9] and Chapter 11 of Hunink et al. [7].

References

1 National Research Council. Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling, Vol. 1, Review and Recommendations. Washington: National Academy Press; 1991.
2 Weinstein MC, Toy EL, Sandberg EA, et al. Modeling for health care and other policy decisions:

uses, roles, and validity. Value Health 2001;4: 348–61.

3 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.

4 Hay J, Jackson J, Luce B, et al. Methodological issues in conducting pharmacoeconomic evaluations—modeling studies. Value Health 1999;2:78–81.

5 Akehurst R, Anderson P, Brazier J, et al. Decision analytic modeling in the economic evaluation of health technologies. Pharmacoeconomics 2000;17: 443–4.

6 Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-Effectiveness in Health and Medicine. Report of the Panel on Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996.

7 Hunink M, Glasziou P, Siegel J, et al. Decision Making in Health and Medicine: Integrating Evidence and Values. Cambridge: Cambridge University Press; 2001.

8 Kuntz K, Weinstein M. Modelling in economic evaluation. In: Drummond M, McGuire A, editors. Economic Evaluation in Health Care: Merging Theory with Practice. Oxford: Oxford University Press; 2001.

9 Briggs A. Handling uncertainty in economic evaluation and presenting the results. In: Drummond M, McGuire A, editors. Economic Evaluation in Health Care: Merging Theory with Practice. Oxford: Oxford University Press; 2001.