State-Transition Modeling : A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force –3

Uwe Siebert, Oguzhan Alagoz, Ahmed M. Bayoumi, Beate Jahn, Douglas K. Owens, David J. Cohen and Karen M. Kuntz

Med Decis Making 2012 32: 690
DOI: 10.1177/0272989X12455463

State-Transition Modeling: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force–3

Uwe Siebert, MD, MPH, MSc, ScD, Oguzhan Alagoz, PhD, MS, Ahmed M. Bayoumi, MD, MSc, Beate Jahn, PhD, Douglas K. Owens, MD, MS, David J. Cohen, MD, MS, Karen M. Kuntz, ScD

State-transition modeling (STM) is an intuitive, flexible, and transparent approach of computer-based decision-analytic modeling, including both Markov model cohort simulation as well as individual-based (first-order Monte Carlo) microsimulation. Conceptualizing a decision problem in terms of a set of (health) states and transitions among these states, STM is one of the most widespread modeling techniques in clinical decision analysis, health technology assessment, and health-economic evaluation. STMs have been used in many different populations and diseases, and their applications range from personalized health care strategies to public health programs. Most frequently, state-transition models are used in the evaluation of risk factor interventions, screening, diagnostic procedures, treatment strategies, and disease management programs. ***Key words:*** *decision-analytic modeling; guidelines; Markov models; state-transition modeling.* ***(Med Decis Making 2012;32:690–700)***

Received 2 March 2012 from UMIT–University for Health Sciences, Medical Informatics and Technology,Hall/Tyrol, Austria (US); Departments of Industrial and Systems Engineering and Population Health Sciences, University of Wisconsin-Madison, Madison, WI, USA (OA); Department of Medicine and Institute of Health Policy, Management and Evaluation, University of Toronto, and St. Michael's Hospital, Toronto, ON, Canada (AMB); UMIT–University for Health Sciences, Medical Informatics and Technology, Hall i.T., and Oncotyrol Center for Personalized Cancer Medicine, Innsbruck, Austria (BJ); VA Palo Alto Health Care System, Palo Alto, CA, and Stanford University, Stanford, CA, USA (DKO); Saint Luke's Mid America Heart Institute, University of Missouri-Kansas City School of Medicine, Kansas City, MO, USA (DJC); and University of Minnesota, School of Public Health, Minneapolis, MN, USA (KMK). The authors thank the many people who reviewed this article and contributed greatly with their thoughtful and critical comments. US and BJ were supported in part by the Oncotyrol Center for Personalized Cancer Medicine (COMET center). The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government.

Supplementary material for this article is available on the *Medical Decision Making* Web site at http://mdm.sagepub.com/supplemental.
Revision accepted for publication 15 June 2012.

Address correspondence to Uwe Siebert, MD, MPH, MSc, ScD, Professor of Public Health Dept. of Public Health and Health Technology Assessment UMIT - University for Health Sciences, Medical Informatics and Technology, Eduard Wallnoefer Center I, A-6060 Hall i.T., Austria, Tel.: +43(0)50-8648-3930, Fax: +43(0)50-8648-673931; Email: uwe.siebert@umit.at

DOI: 10.1177/0272989X12455463

A new Good Research Practices in Modeling Task Force was constituted by the ISPOR Board of Directors in 2010, and the Society for Medical Decision Making was invited to join the effort. This paper, along with six others,[1-6] is part of a series commissioned by the Task Force.

The goal of this article is to provide consensus-based guidelines for the application of state-transition modeling (STM) in the context of health care. We structured the best practice recommendations in the following sections: choice of model type (cohort v. individual-level model), model structure, model parameters, analysis, reporting, and communication. In each of these sections, we give a brief description, address the issues that are of particular relevance to the application of STMs, give specific examples from the literature, and provide best practice recommendations for STM. These recommendations are directed

Related Materials

For more information on the ISPOR-SMDM Task Force, visit the website at http://www.ohsu.edu/epc/mdm/modeling.cfm. See ''Modeling Good Research Practices—Overview, Issues, and Preferred Practices: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force,'' by J. Jaime Caro, Andrew H. Briggs, Uwe Siebert, and Karen M. Kuntz, published in this issue on pages 667-677, for an overview of the series.

both to modelers and to users of modeling results such as clinicians, clinical guideline developers, manufacturers, or policy makers.

USE OF STATE-TRANSITION MODELS

Many clinical situations can be described in terms of the conditions that individuals can be in (''states''), how they can move among such states (''transitions''), and how likely such moves are (''transition probabilities''). In these situations, STMs are often well suited to the decision problem, as they conceptualize it in terms of a set of states and transitions among these states. Several dimensions fall within this broad category. For example, some STMs allow for interactions among groups (i.e., the transition probabilities depend on the states of other individuals), whereas others assume no interactions. Some allow transitions to occur only at specified time intervals, whereas others use a continuous state-space process. STMs can be used to simulate a closed cohort over time or a dynamic population (e.g., the US adult population). They may simulate all individuals simultaneously or one at a time.

We focus on two common frameworks in health care: cohort, or ''Markov,'' models[7,8] and individual-based models, commonly known as ''first-order Monte Carlo'' or ''microsimulation'' models.[9–11] These frameworks do not capture interactions, model a single (closed) cohort, and allow transitions to occur only at specified time intervals.

An STM should be used, rather than a simpler model with limited ability to reflect time (e.g., decision tree), if it requires time-dependent parameters (e.g., recurrence probability after cancer treatment), time to an event (e.g., disease-free survival), or repeated events (e.g., second myocardial infarction).[12] Other modeling techniques are also suitable for these situations (e.g., discrete event simulation).

KEY CONCEPTS AND DEFINITIONS

The formal elements of an STM are states, transitions, initial state vector, transition probabilities, cycle length, state values (''rewards''), logical tests performed at the beginning of each cycle to determine the transitions, and termination criteria.

Model Structure

STMs are structured around a set of mutually exclusive and collectively exhaustive health states. A modeled individual must be in only one state in any cycle. Events that occur within a cycle can be modeled with a Markov cycle tree—a series of chance nodes representing the events. The average number of cycles that individuals reside in each state can be used in conjunction with state values (e.g., life years, health-related quality of life, cost) to estimate life expectancy, quality-adjusted life expectancy, and expected costs.

An STM can capture many features present in the course of a disease or clinical process (e.g., disease risk over time, changing states, episodic events), although this is not the only approach that can capture these features.[13] The principal advantage of cohort STMs is that they are relatively simple to develop, debug, communicate, and analyze using user-friendly software, if the number of states is not too large. The primary disadvantage is the underlying assumption that transition probabilities do not depend on history— neither on past states nor time spent in the current state. This assumption (the ''Markovian'' property) can be very limiting for clinical applications where these aspects tend to be strong determinants of what happens next. A Markov model can handle memory by creating states that include history, but this can greatly increase the number of states, resulting in very large models that are difficult to manage (i.e., ''state explosion'').

Individual-level STMs (Table 1) are not limited by the Markovian property as they simulate one individual at a time. These microsimulations are evaluated using first-order Monte Carlo simulation: Whether an individual facing a certain transition probability makes this transition depends on a random number.

Whereas cohort models are analyzed as single cohorts progressing through the states simultaneously (which does not allow distinguishing one individual from another except by state descriptions), individual-level STMs keep track of each individual's history (''tracker variables''). This can greatly reduce the number of states. The main disadvantages are that they are computationally intensive, often requiring simulation of millions of individuals to obtain stable values for the outcomes of interest, and they are more difficult to debug. Figure 1 displays the Markov trace of a cohort simulation and the possible paths of a microsimulation.

An STM must start with a decision node from which the intervention branches originate. Each branch either leads directly to one Markov node (followed by an STM) (Figure 2a), or there is a decision tree leading to multiple Markov nodes per intervention (Figure 2b). An STM following one branch can have a different structure than one following another branch.

Table 1 Cohort v. Individual-Level State-Transition Models

	Cohort State-Transition Models	Individual-Level State-Transition Models
Ease of model development	Higher (if number of states is limited)	Lower
Ease of model debugging	Higher (if number of states is limited)	Lower
Ease of communication to nonexperts	Higher	Lower
Markov assumption, memoryless	Yes	No
Ease of modeling many different subgroups	Lower	Higher
Danger of explosion in number of states	Yes	No
Distribution of outcomes (as opposed to only means)	Possible, but technically more difficult	Yes
Report of individual patient histories	No	Yes
Decision-analytic software available	Yes	Yes (need advanced knowledge)

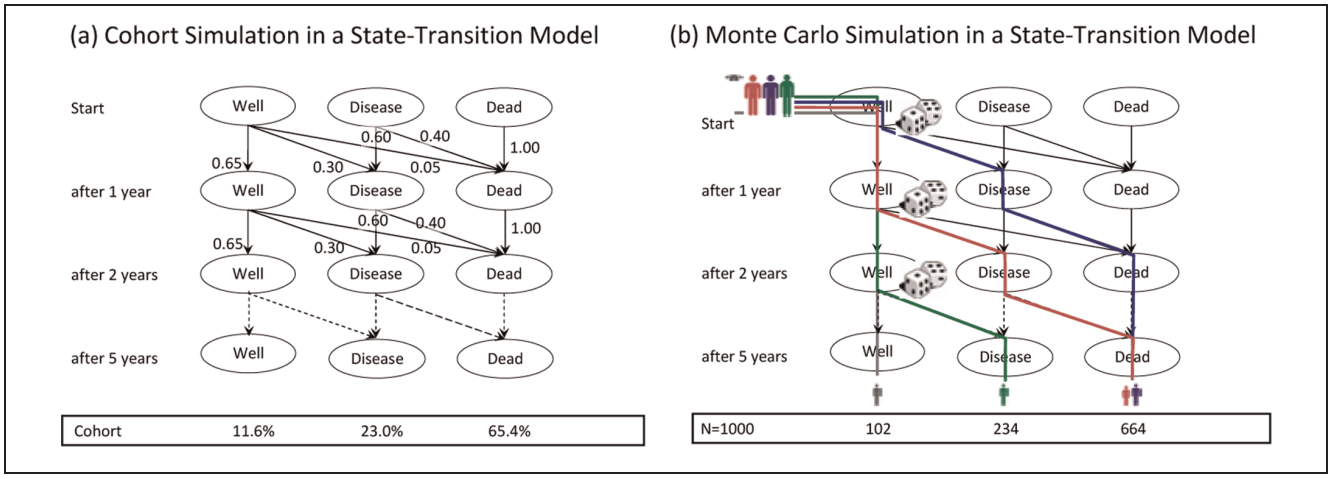

Figure 1 *In a cohort simulation (a), the entire cohort is (re)distributed across states after each cycle. In an individual-level microsimulation (b), a finite number of individuals are simulated using first-order Monte Carlo microsimulation. In this simple example, all individuals start in the state ''Well'' and the disease is chronic (i.e., there is no regression from ''Disease'' to ''Well''). In principle, individuals can start in different states and they can regress to states they have already been in.*

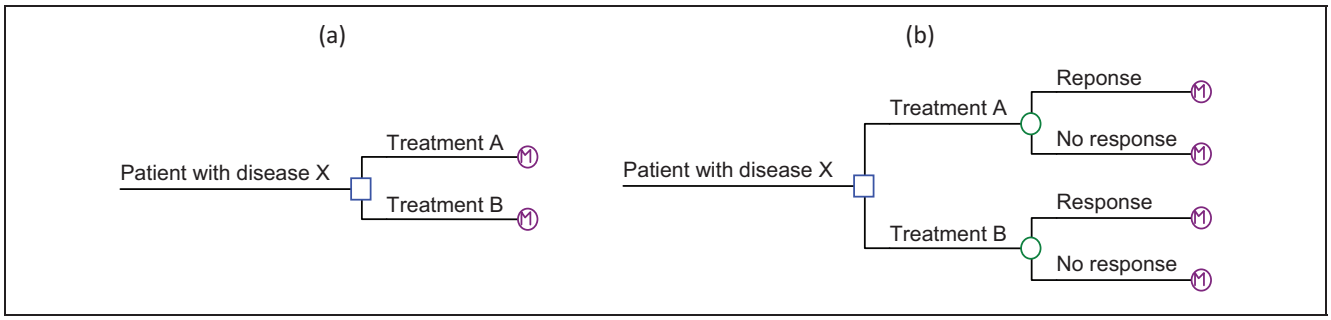

Figure 2 *In model (a), decision branches lead directly to one Markov node per intervention strategy, and the first events are modeled within the Markov cycle tree. Model (b) contains an upfront decision tree modeling the first events and leading to multiple Markov nodes per intervention strategy.*

Types of Interventions

STMs can be used to compare various types of interventions.[14]

Primary prevention. STMs used to evaluate primary prevention strategies are concerned with reduction in risk of developing a disease (e.g., Col and others[15]). Hence, their focus is on what happens prior to disease, such as number and severity of risk factors. The starting cohort is individuals free of disease or complications being modeled.

Screening. STMs used to evaluate screening strategies[16,17] consider two types: one-time screening (e.g., of newborns[18] or genetic screening[19–21]) or repeated (interval) screening (e.g., for cancer[22–25] or HIV[26,27]). The evaluated screening strategies can differ in many respects, such as type and sequence of tests used, diagnostic workup modes, screening interval, and ages at which screening begins and ends.

Diagnosis. Diagnostic models are used to identify optimal diagnostic strategies among individuals who present with signs or symptoms of one or more suspected diseases.[28] Testing options may involve use of one test v. another, one v. multiple tests, different combinations or sequences, one positivity criterion v. another (e.g., Gould and others[29]), or a multiple-test diagnostic score, or focusing the development of new diagnostic technologies.[30]

Treatment. This is defined as any intervention available for someone who already has a clinical condition that affects health consequences or prognosis. An STM disease process should reflect the disease's natural history, expected prognostic pathways in the absence of intervention, and treatment effects.[31–33]

RECOMMENDATIONS

An STM is a reasonable choice when the decision problem can be framed in terms of states, interactions between individuals are not relevant, and the population of interest is a closed cohort. Multipurpose disease-specific models (e.g., Freedberg and others[34]) are not addressed here.

Choice of Model Type

Best practices

III-1 If the decision problem can be represented with a manageable number of health states that incorporate all characteristics relevant to the decision problem, including the relevant history, a cohort simulation should be chosen because of its transparency, efficiency, ease of debugging, and ability to conduct specific value-of-information analyses. If, however, a valid representation of any aspect of the decision problem would lead to an unmanageable number of states, then an individual-level state-transition model is recommended. Validity should not be sacrificed for simplicity.

Before choosing between cohort or individual-level simulation, the characteristics of the population that must be carried through the model (i.e., state descriptors or tracker variables) must be specified. These must include all relevant states pertaining to the disease or clinical process and intervention(s), all relevant history (e.g., past states, risk factors, time in state, time since last event) that are determinants of transition probabilities (e.g., determinants of disease incidence, progression, mortality), or state values (e.g., determinants of utilities and costs). An advantage of using an individual-level STM is the ability to model individual characteristics as continuous variables and to evaluate dynamic intervention strategies—ones in which future decisions depend on current and past patient characteristics. In cohort STMs, continuous variables (e.g., blood pressure) have to be categorized; some guidance exists for determining how many states to create.[35] Individual-level STMs,[36,37] however, require more computation time, which may be important if probabilistic sensitivity analyses or value-of-information analyses are performed.

Two examples of STMs that required microsimulation include one comparing intermediate and long-term clinical outcomes of different imaging screening strategies for breast cancer in women with *BRCA1* gene mutations[36] and one developed to estimate the long-term impact of interventions for people with type 2 diabetes.[37]

Model Structure

Problem Statement

Best practices

III-2 The strategies being evaluated should be clearly defined. In particular, sequential decisions should not be modeled within the Markov cycle tree but rather be part of the specification of the alternative intervention strategies that precede the Markov tree.

Interventions may be single-time (e.g., one-time vaccination or surgery), static over time (i.e., not

depending on intervention outcomes or other events), or dynamic (consisting of decision rules for how to start, stop, or change interventions over time[38]). Examples of dynamic strategies are 1) the start of a preventive behavioral intervention if body mass index (BMI) increases, 2) increase of the screening interval for cervical cancer if a woman has repeatedly tested negative, 3) repetition of a diagnostic test after an equivocal result, or 4) change to another drug after first-line treatment failure.

This recommendation refers to standard STMs. Other methods such as Markov decision processes generalize STMs by allowing embedding of sequential decisions, and thus, multiple decisions can be made in multiple time periods.[39,40]

Starting Cohort

Best practices

III-3 The starting cohort should be defined by the demographic and clinical characteristics that affect the transition probabilities or state values (e.g., quality of life and cost).

The model outputs for a single cohort allow for the comparison of alternative strategies for that cohort. If the optimal decision varies by subgroup (e.g., defined by age, sex, and risk factors known to the decision maker at the time of the decision), the comparison can be reported for different cohorts. If outputs are desired for a population-based starting cohort, then the model must be run multiple times, one for each stratum, and then aggregated across strata (e.g., Dewilde and Anderson[41]).

Defining States

Best practices

III-4 Specification of states and transitions should generally reflect the biological/theoretical understanding of the disease or condition being modeled.

Conceptualization of an STM should begin by identifying states that reflect the disease/health process, with transitions among the states that would be expected in the absence of intervention, as well as the interventions' effects on these. The states should be specified as mutually exclusive (any individual can be in only one state during each cycle) and collectively exhaustive (every individual in the initial cohort must be in a state during each cycle), and they should adequately capture the benefits or harms of any interventions. These effects can characterize the state values (e.g., differences in symptoms and quality of life) or reflect changes in transition probabilities.

At the start of a cohort simulation, the modeled population is allocated among the states. Each state is homogeneous—every individual in that state has the same transition probabilities—implying that any characteristics that determine those probabilities must not differ within the state. If history (prior states or time spent in a current state) is important in determining transition probabilities, then the relevant states should carry that history in their definition (e.g., if risk of myocardial infarction depends on prior myocardial infarction [MI], then the states would need to include this historical element by using states such as ''disease-free, no prior MI'' and ''disease-free, prior MI''). In an individual-level STM, these characteristics and other parameters can be heterogeneous within a state but must be tracked throughout, and transition probabilities must be defined as a function of these characteristics.

When there are alternatives for modeling natural history (e.g., defining states with biological but often unobservable disease measures such as spirometry in asthma or with symptomatic descriptions such as ''on treatment'' stages for Parkinson disease), the analyst should justify the approach used or compare alternatives in sensitivity analysis. Although it may be possible to describe natural history solely on the basis of health care utilization, this does not provide direct insight into health outcomes at the biological level, and its value is limited for most decision problems. If cause of death is an important outcome or different causes have different costs, then competing causes should be modeled in an unbiased way (e.g., probability of death modeled first, followed by a conditional distribution of cause-specific deaths).

Another important consideration in structuring an STM is initial immediate or short-term events. An efficient and transparent way to model such events is as a decision tree preceding the STM, unless there are justified reasons for representing them within the states. When events are modeled preceding an STM, the time spent before entering the STM should be captured appropriately by giving credit to the starting cohorts for the time elapsed (e.g., an upfront decision tree could be used to represent results and subsequent outcomes from diagnostic test strategies,[29] or treatments with limited duration,[33] or initial coronary interventions[31]; Figure 3).

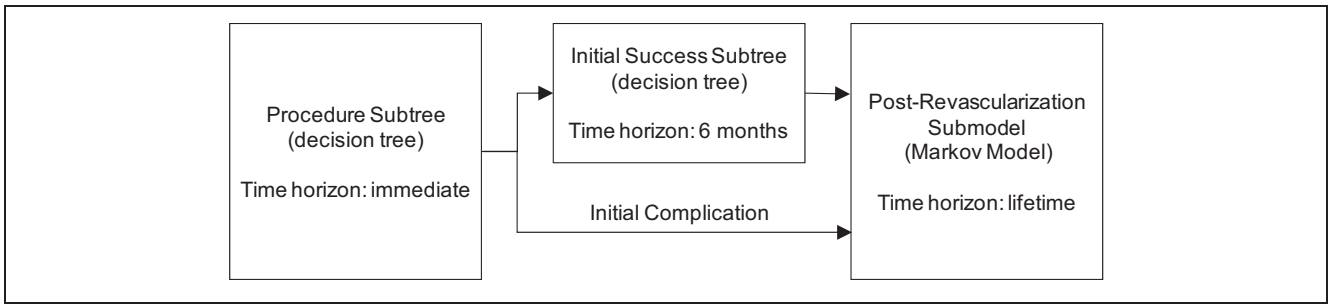

Figure 3 The model consists of two decision trees and one Markov model. Source: Siebert U. The role of decision-analytic models in the prevention, diagnosis and treatment of coronary heart disease. Z Kardiol. 2002;91(Suppl. 3):144–51. Reprinted with kind permission from Springer Science+Business Media.

Intervention Effects

Best practices

III-5 States should adequately capture the type of intervention (i.e., prevention, screening, diagnostics, treatment) as well as the intervention's benefits and harms.

For models evaluating primary prevention, possible risk factor levels in the target population should be represented as predisease states or tracker variables. As current risk factors are often predictors of future ones, the state descriptions or tracker variables should capture their course and changes in sufficient detail. These models may not require as much detail postdisease as would those evaluating interventions for the disease, but they should sufficiently capture the relevant disease elements. One useful approach is to collaborate with investigators with well-developed disease-specific treatment models to derive relevant eventual outcomes (e.g., lifetime costs or survival) that can be used in the prevention model.[42,43]

Models evaluating screening should define states reflecting the underlying disease process, especially for interval screening programs. It is not appropriate to take an empirical estimate of the probability of a positive screen from a study because this does not explicitly incorporate the underlying disease probability. For cancer screening, states should distinguish between cases detected by screening and incidental findings (e.g., through other diagnostic tests) or cases detected by symptoms. Modelers should describe how they have controlled for lead-time and length bias.[44] Dynamic interval screening strategies (''individualized screening'') may depend on prior screening history (e.g., some algorithms for cervical cancer screening recommend extension of the interval after repeated negative tests). As capturing screening history in the Markov states can lead to state explosion, it may be necessary to use an individual-level STM, with screening history included as tracker variables.[22]

In models evaluating diagnostic strategies, it is typical to represent the testing pathways and their outcomes (e.g., true positive, false positive) in a pre-STM decision tree. If multiple tests are performed in combination or sequence, and some results are prognostic factors that can change over time, their history must be incorporated in the states or implemented as tracker variables.

In models evaluating therapeutic strategies, the mechanism by which the treatment alters the disease course should be explicit (e.g., reduction in event risk or mortality, slowing disease progression). In addition, how harms of the intervention(s) affect prognosis should be specified. STMs should incorporate realistic assumptions about adherence over time. Long-term treatment effectiveness and costs often depend on time-varying heterogeneous patient characteristics, and many treatments are ''personalized'' and follow dynamic rules (e.g., dose and second-line treatments, as well as compliance and treatment success, depend on current treatment response and side effects). These dynamic characteristics should be considered in the states or tracker variables.

Heterogeneity

Best practices

III-6 States need to be homogeneous with respect to both observed and unobserved (i.e., not known by the decision maker) characteristics that affect transition probabilities.

In a cohort STM, all individuals in a given state are indistinguishable in terms of their transition probabilities. Many characteristics that affect transition probabilities (e.g., age, sex, comorbidities, disease stage) are known at the time of the decision and can

be used to define the starting cohorts. These characteristics do not need to be incorporated into state definitions or tracker variables unless they are expected to change over time in a meaningful way. For example, a cohort starting with few comorbidities may develop more over time, and to capture this requires incorporating this attribute in the states. Variables that affect transition probabilities but are not known at decision time (e.g., genetic mutation, undiagnosed infection) can create ''heterogeneity bias,''[45,46] and inclusion of such variables should be considered.

Time Horizon

Best practices

III-7 The time horizon for the model should be sufficiently large to capture all health effects and costs relevant to the decision problem.

The time horizon relates to the number of cycles or duration for which the cohort is tracked. Common approaches include modeling to an age of 120 years or tracking the cohort until more than 99.9% of individuals are dead. If the intervention affects mortality, the time horizon should be lifetime to capture (quality-adjusted) life years gained from delayed deaths.

Cycle Length

Best practices

III-8 Cycle length should be short enough to represent the frequency of clinical events and interventions.

Choice of cycle length should be based on the clinical problem, remaining life expectancy, and computational efficiency. It should allow transitions to occur consistent with the clinical problem and intervention effects (e.g., a model assessing monthly screening requires cycles no longer than 1 month). Cycle length should be short enough that an event occurs at most once per cycle. Shorter cycle lengths provide better approximations of life expectancy. If life expectancy is relatively short (e.g., with an acute disease or at older ages), then a shorter cycle length should be considered, even if the clinical problem does not warrant it. Although shorter cycles will always yield more precise estimates, the error gets very small when the number of cycles required increases.

Model Symmetry

Best practices

III-9 Components of state-transition models that reflect similar clinical courses should not be re-created but rather should be incorporated once and linked to that structure throughout the model.

Symmetric models ensure that the disease process is represented consistently across strategies. For example, STMs used to compare cardiac catheterization and subsequent treatment dictated by its results v. initial medical therapy should specify true underlying disease status even though it is not observed in the medical therapy strategy.[47] Otherwise, errors result when conducting sensitivity analysis on the underlying probability of any particular anatomy (e.g., left main disease).

Data

STMs should provide clear justification for estimates of transition probabilities and state values and their ranges for sensitivity analysis.

Data Sources

Best practices

III-10 Transition probabilities and intervention effects should be derived from the most representative data sources for the decision problem.

Ideally, transition probabilities pertaining to natural history are derived from population-based epidemiological studies, as these are most likely to be representative. Transition probabilities may be derived from the control arms of trials, recognizing that these may be less generalizable due to selection criteria for participants. If multiple sources are available, summarized data from a systematic review or meta-analysis are best for informing transition probabilities or state values. Methods assessing the quality of a body of evidence rather than quality of individual studies are available.[48–50] In the absence of good systematic reviews, a detailed evidence table should be provided in an appendix with a description and justification of how key parameters—including the ranges used for sensitivity analyses—were derived.

Parameter Derivation

Best practices

III-11 All methods and assumptions used to derive transition probabilities and intervention effects should be described.

Transition probabilities and rates should be used appropriately.[51] The conversion of transition probabilities from one time unit to another should be done through rates, which should never be presented

as percentages. To avoid confusion, probabilities should never be called rates.

The assumed functional relationship between disease-specific and background mortality should be stated. Because an assumption of additive rates can give very different results than a multiplicative one,[52] the impact of this assumption should be assessed.

Intervention Effects

Best practices

III-12 All parameters relating to the effectiveness of interventions derived from observational studies should be correctly controlled for confounding. Time-varying confounding is of particular concern in estimating intervention effects.

Efficacy derived from randomized clinical trials may have to be adjusted for compliance to reflect real-world effectiveness.[53] Effectiveness derived from observational studies must be adjusted for confounding (e.g., using multivariate regression techniques or propensity scoring). Adjustment for time-varying confounding (i.e., confounders that simultaneously act as intermediate steps in the pathway between intervention and outcome) require special methods such as marginal structural analysis or g-estimation.[38,54] When results from observational studies are used in the model, causal graphs can be used to explicitly state causal assumptions.[53]

When extrapolating beyond a trial's duration, reductions in all-cause mortality should not be applied directly since background mortality (from other causes) increases with age. If disease-specific mortality is not available, a relative reduction can be applied to disease-specific mortality, providing a conservative estimate of treatment benefit. Alternatively, life-table mortality could be subtracted from total mortality to estimate reduction in disease-specific mortality.[55]

For preventive and therapeutic interventions, if evidence is available for reduction in disease incidence, events or progression, and also for mortality, using both may double count. If this is a concern, the consistency of the model-generated reductions should be validated with estimates from clinical studies.

State Valuation

Best practices

III-13 The valuation of intermediate outcomes/states should be justified.

Expected outcomes depend on values assigned to each state (e.g., quality-adjusted life years can be derived if utilities are assigned). State values should be justified, preferably based on theory.

Analysis

Half-Cycle Correction

Best practices

III-14 A half-cycle correction should be applied to costs and effectiveness in the first cycle and in the final cycle if not using a lifetime horizon.

When it is not known when the transitions occur within the cycle, we expect that, on average, they will occur about halfway through the cycle. To account for this, a half-cycle correction is made by assigning half of the reward in each state. Giving a full reward at the start (i.e., transitions occur at cycle end) overestimates expected values; assigning no reward (i.e., transitions occur at cycle start) underestimates them.[8]

Analyzing Distributions

Best practices

III-15 For certain decision problems, it may be important to report not only the expected value but also the distribution of the outcomes of interest.

It may be important for the decision maker (e.g., for equity reasons) to know whether a treatment with a 1-year life expectancy gain extends life by 1 year for each person or by 3 years in 50% but reduces life by 1 year in the other 50%. Distributions are derived easily from individual-level STMs but can also be derived from cohort models, either by analyzing the Markov trace or by running the model individually (but without tracker variables).

Performing Microsimulation

Best practices

III-16 The number of individuals simulated should be large enough to generate stable estimates of the expected values.

To achieve stable results in an individual-based simulation, sufficient individuals must be modeled. Stability of model results is assessed by calculating variance from multiple runs with identical number of individuals,[55] which should be much smaller than the smallest difference expected between strategies. Variance reduction techniques (e.g., using common random numbers) can decrease the required numbers.[56]

Communicating Results

Graphical representation of an STM helps communicate key structure and most important assumptions made regarding states and allowable transitions. Since rigorous studies evaluating alternative presentation methods are lacking, these recommendations represent best judgments based on experience.

Presenting the Model

Best practices

III-17 The report should use nontechnical language and clear figures and tables that enhance understanding of the STM to communicate its key structural elements, assumptions, and parameters.

STMs are often represented by two types of diagrams: a state-transition diagram, also known as a ''bubble diagram,'' and a Markov cycle tree (a set of probability nodes that describes the progression from one state to the next). State-transition diagrams represent states as discrete compartments (''bubbles'') and transitions as arrows between them. Although a relatively simple model with few states may be fully represented in a single diagram, complicated models consisting of multiple states cannot feasibly be represented by displaying all states and transitions. Such models are invariably cluttered, and the resultant tangle of arrows and states often impairs communication, rather than enhancing it. In such circumstances, simplified or partial diagrams are desirable. Markov cycle trees, often stylized, can display the transitions between states, as well as probabilities and transitions that are conditional upon other events or parameters.

Presenting Results

Best practices

III-18 In addition to final outcomes, intermediate outcomes that enhance the understanding and transparency of the model results should also be presented.

Presenting intermediate results can be helpful for demonstrating face validity for clinical experts, epidemiologists, and decision makers. Useful measures include incidences related to a fixed time horizon (e.g., 10-year risks), average number of events per lifetime, percent of initial cohort that experienced two or more events in their lifetime, or mean age at which the first event occurred. In addition, it can be useful to generate summary data from STMs to indicate how much time is spent in certain states (e.g., a model of stroke prevention in atrial fibrillation could report the average amount of time spent without a stroke and the average time from first stroke to death). Such measures can also be used for debugging the model or for validating model results with empirical data or for internal debugging. As STMs allow deriving the time at which particular transitions occur, the results can be represented as (modeled) probability or survival curves and directly compared with survival curves from empirical studies.

VALIDATION AND CONSISTENCY

Ensuring that the STM provides a sufficiently accurate representation of the real system is important. This is covered in detail elsewhere in this series.[6] A useful method of identifying programming errors in an STM is to check whether model-building rules such as the use of symmetric branches or states in STMs are followed. Inspection of the Markov trace can also help find errors, by setting parameters in such a way that how the trace will look can be predicted (e.g., so that no modeled individual will visit a given state during the simulation).

CONCLUSIONS

STMs can provide a comprehensive and powerful tool to guide decisions in health care. Best practices, for cohort and individual-based STMs, are recommended for the development, analysis, validation, and reporting of STMs. Although many more aspects than those described in this article may have to be considered for good modeling practice, and not all models may be able to comply with all the recommendations, following these recommendations should help to make STMs more valid, transparent, and useful in guiding health care decisions.

REFERENCES

1. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling good research practices—overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–1. Med Decis Making. 2012;32(5):667–677.

2. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–2. Med Decis Making. 2012;32(5):678–689.

3. Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Möller J. Modeling using discrete event simulation: a report of the ISPOR-SMDM

Modeling Good Research Practices Task Force–4. Med Decis Making. 2012;32(5):701–711.

4. Pitman R, Fisman D, Zaric GS, et al. Dynamic transmission modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group–5. Med Decis Making. 2012;32(5): 712–721.

5. Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group–6. Med Decis Making. 2012; 32(5):722–732.

6. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–7. Med Decis Making. 2012;32(5):733–743.

7. Beck JR, Pauker SG. The Markov process in medical prognosis. Med Decis Making. 1983;3:419–58.

8. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. Med Decis Making. 1993;13:322–38.

9. Spielauer M. Dynamic microsimulation of health care demand, health care finance and the economic impact of health behaviours: survey and review. Int J Microsimuation. 2007;1:35–53.

10. Groot Koerkamp B, Weinstein MC, Stijnen T, Heijenbrok-Kal MH, Hunink MG. Uncertainty and patient heterogeneity in medical decision models. Med Decis Making. 2010;30:194–205.

11. Weinstein MC. Recent developments in decision-analytic modelling for economic evaluation. Pharmacoeconomics. 2006; 24:1043–53.

12. Hunink MGM, Glasziou P, Siegel JE, et al. Decision Making in Health and Medicine: Integrating Evidence and Values. Cambridge, UK: Cambridge University Press; 2001.

13. Caro JJ. Pharmacoeconomic analysis using discrete event simulation. Pharmacoeconomics. 2005;23:323–32.

14. Siebert U. When should decision-analytic modeling be used in the economic evaluation of health care? Eur J Health Econ. 2003;4:143–50.

15. Col NF, Eckman MH, Karas RH, et al. Patient-specific decisions about hormone replacement therapy in postmenopausal women. JAMA. 1997;277:1140–7.

16. Eddy DM. Screening for Cancer: Theory, Analysis and Design. Englewood Cliffs, NJ: Prentice-Hall; 1980.

17. Russell LB. Educated Guesses: Making Policy about Medical Screening Tests. Los Angeles: University of California Press; 1994.

18. Grill E, Hessel F, Siebert U, et al. Comparing the clinical effectiveness of different new-born hearing screening strategies: a decision analysis. BMC Public Health. 2005;5:12.

19. Gutierrez de Mesa E, Hidalgo I, Christidis P, Ciscar JC, Vegas E, Ibarreta D. Modeling the impact of genetic screening technologies on healthcare: theoretical model for asthma in children. Mol Diagn Ther. 2007;11:313–20.

20. Meckley LM, Gudgeon JM, Anderson JL, Williams MS, Veenstra DL. A policy model to evaluate the benefits, risks and costs of warfarin pharmacogenomic testing. Pharmacoeconomics. 2010;28:61–74.

21. Rogowski WH. The cost-effectiveness of screening for hereditary hemochromatosis in Germany: a remodeling study. Med Decis Making. 2009;29:224–38.

22. Goldie SJ, Kim JJ, Myers E. Cost-effectiveness of cervical cancer screening. Vaccine. 2006;24:S164–70.

23. Lee JM, McMahon PM, Kong CY, et al. Cost-effectiveness of breast MR imaging and screen-film mammography for screening BRCA1 gene mutation carriers. Radiology. 2010;254:793–800.

24. Sroczynski G, Schnell-Inderst P, Muhlberger N, et al. Cost-effectiveness of primary HPV screening for cervical cancer in Germany—a decision analysis. Eur J Cancer. 2011;47: 1633–46.

25. Zauber AG, Lansdorp-Vogelaar I, Knudsen AB, Wilschut J, van Ballegooijen M, Kuntz KM. Evaluating test strategies for colorectal cancer screening: a decision analysis for the US Preventive Services Task Force. Ann Intern Med. 2008;149:659–69.

26. Sanders GD, Bayoumi AM, Sundaram V, et al. Cost effectiveness of screening for HIV in the era of highly active antiretroviral therapy. N Engl J Med. 2005;352:570–85.

27. Sanders GD, Bayoumi AM, Holodniy M, Owens DK. Cost effectiveness of HIV screening in patients older than 55 years of age. Ann Intern Med. 2008;148:889–903.

28. Trikalinos TA, Siebert U, Lau J. Decision-analytic modeling to evaluate benefits and harms of medical tests: uses and limitations. Med Decis Making. 2009;29:E22–9.

29. Gould MK, Sanders GD, Barnett PG, et al. Cost effectiveness of alternative management strategies for patients with solitary pulmonary nodules. Ann Intern Med. 2003;138:724–35.

30. Hunink MGM, Kuntz KM, Fleischmann KE, Brady TJ. Noninvasive imaging for the diagnosis of coronary artery disease: focusing the development of new diagnostic technology. Ann Intern Med. 1999;131:673–80.

31. Cohen DJ, Bakhai A, Shi C, et al. Cost-effectiveness of sirolimus-eluting stents for treatment of complex coronary stenoses: results from the Sirolimus-Eluting Balloon Expandable Stent in the Treatment of Patients With De Novo Native Coronary Artery Lesions (SIRIUS) trial. Circulation. 2004;110:508–14.

32. Sanders GD, Hlatky MA, Owens DK. Cost effectiveness of the implantable cardioverter defibrillator (ICD) in primary prevention of sudden death. N Engl J Med. 2005;353:1471–8.

33. Siebert U, Sroczynski G, German Hepatitis C Model (GEHMO) Group, HTA Expert Panel on Hepatitis C. Effectiveness and cost-effectiveness of initial combination therapy with interferon/peginterferon plus ribavirin in patients with chronic hepatitis C in Germany: a health technology assessment commissioned by the German Federal Ministry of Health and Social Security. Int J Tech Assess Health Care. 2005;21:55–65.

34. Freedberg KA, Scharfstein JA, Seage GR III, et al. The cost-effectiveness of preventing AIDS-related opportunistic infections. JAMA. 1998;279:130–6.

35. Bentley TGK, Weinstein MC, Kuntz KM. Effects of categorizing continuous variables in decision-analytic models. Med Decis Making. 2009;29:549–56.

36. Lee JM et al. Breast cancer screening in BRCA1 mutation carriers: effectiveness of MR imaging—Markov Monte Carlo decision analysis. Radiology. 2008;246:763–71.

37. Clarke PM, Gray AM, Briggs A, et al. A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model. Diabetologia. 2004;47:1747–59.

38. Robins JM, Hernán MA, Siebert U. Estimations of the effects of multiple interventions. In: Ezzati M, Lopez AD, Rodgers A, Murray CJL, eds. Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors. Vol. 1. Geneva, Switzerland: World Health Organization; 2004. p 2191–230.

39. Puterman ML. Markov Decision Processes: Discrete Stochastic Dynamic Programming (Wiley Series in Probability and Statistics). New York: John Wiley; 1994.

40. Alagoz O, Hsu H, Schaefer AJ, Roberts MS. Markov decision processes: a tool for sequential decision making under uncertainty. Med Decis Making. 2010;30:474–83.

41. Dewilde S, Anderson R. The cost-effectiveness of screening programs using single and multiple birth cohort simulations: a comparison using a model of cervical cancer. Med Decis Making. 2004;24:486–92.

42. Lieu TA, Gurley RJ, Lundstrom RJ, et al. Projected cost-effectiveness of primary angioplasty for acute myocardial infarction. J Am Coll Cardiol. 1997;30:1741–50.

43. Berrington de Gonzalez, Kim KP, Knudsen AB, et al. Radiation-related cancer risks from CT colonography screening: a risk-benefit analysis. AJR Am J Roentgenol. 2011;196:816–23.

44. Primer on Lead Time, Length and Overdiagnosis Bias. Effective Clinical Practice 1999 (March/April 1999). Available from: http://www.acponline.org/clinical_information/journals_publications/ecp/primers/marapr99.htm. Accessed 17 January 2012.

45. Kuntz KM, Goldie SJ. Assessing the sensitivity of decision-analytic results to unobserved markers of risk: defining the effects of heterogeneity bias. Med Decis Making. 2002;22:218–27.

46. Zaric GS. The impact of ignoring population heterogeneity when Markov models are used in cost-effectiveness analysis. Med Decis Making. 2003;23:379–86.

47. Sonnenberg FA, Roberts MS, Tsevat J, et al. Toward a peer review process for medical decision analysis models. Med Care. 1994;32(7, Suppl.):JS52–64.

48. Owens DK, Lohr KN, Helfand M, et al. Grading the strength of a body of evidence when comparing medical interventions—AHRQ and the effective health care program. J Clin Epidemiol. 2010;63:513–23.

49. Guyatt GH, Oxman AD, Schunemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. J Clin Epidemiol. 2011;64:380–2.

50. Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011;64(4):401–6.

51. Miller DK, Homan SM. Determining transition probabilities: confusion and suggestions. Med Decis Making. 1994;14:52–8.

52. Kuntz KM, Weinstein MC. Life expectancy biases in clinical decision modeling. Med Decis Making. 1995;15:158–69.

53. Cox E, Martin BC, Van Staa T, Garbe E, Siebert U, Johnson ML. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report—Part II. Value Health. 2009;12:1053–61.

54. Johnson ML, Crown W, Martin BC, Dormuth CR, Siebert U. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part III. Value Health. 2009;12:1062–73.

55. Kuntz KM, Weinstein MC. Modelling in economic evaluation. In: Drummond M, McGuire A, eds. Economic Evaluation in Health Care. Merging Theory with Practice. Oxford, UK: Oxford University Press; 2001. p 141–71.

56. Stout NK, Goldie SJ. Keeping the noise down: common random numbers for disease simulation modeling. Health Care Manage Sci. 2008;11:399–406.