1. Introduction

1.1 Overview

Model-based decision analyses play a fundamental role in technology assessment, clinical decision making, comparative effectiveness research, benefit-risk assessment, cost-effectiveness analysis, and the development of health policy. This paper reports the recommendations of the ISPOR-SMDM Modeling Task Force for State-Transition Modeling for both cohort (Markov) and individual-level (microsimulation) implementations and is directed both to modelers and to “consumers” of modeling results such as clinicians, clinical guideline developers, manufacturers, or policy makers.

The goal of this paper is to provide guidance on methods for developing, analyzing, interpreting, reporting, and communicating about state-transition models (STMs).

1.2 Use of State-Transition Models

Decision-analytic models are used to systematically and explicitly synthesize the best available evidence that relate to the natural history of disease as well as to the effectiveness, risk, costs, and cost-effectiveness of possible interventions in population(s) of interest.

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 analyze the decision problem. STMs conceptualize a decision problem in terms of a set of states and transitions among these states. Several modeling dimensions fall within this broad model category. For example, some STMs can allow for interactions among groups, in that the transition probabilities depend on the states in which the other simulated individuals reside, while others assume no interactions. Some STMs allow transitions to occur only at specified time intervals, while other STMs use a continuous state-space process. STMs can be used to simulate a closed cohort over
time or to simulate a dynamic population (e.g., the US adult population between 2010 and 2030). STMs may also simulate the modeled individuals simultaneously or may simulate one individual at a time.

We focus on two of the more common modeling frameworks in healthcare: cohort models, commonly known as “Markov” models (Beck and Pauker, 1983; Sonnenberg and Beck, 1993); and individual-level models, commonly known as first-order Monte Carlo or microsimulation models (Spielauer, 2007; Sonnenberg and Beck, 1993; Groot Koerkamp et al., 2010; Weinstein, 2006). These two state-transition modeling frameworks do not capture interactions, model a single (closed) cohort, and allow transitions to occur only at specified time intervals. Several other variations are covered in the companion best practices papers of this series.

If a decision must be made between a simple decision tree or an STM, the STM should be used if one or more of the following conditions is true: the model contains time-dependent parameters such as the probability of recurrence after cancer treatment, the time to event is important, such as disease-free survival, or the model must capture repeated events, such as a second myocardial infarction (Hunink and Glasziou, 2001).

2. Key Concepts and Definitions

2.1 Terminology

The formal elements of a state transition-model 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. The terms used in this report are defined in Appendix 1.

2.2 Model Structure

STMs are structured around a set of mutually exclusive and collectively exhaustive health states. A modeled individual must be in one of the states, and only one state, in any given cycle. Events that can 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 simulated individuals reside in each state can be used in conjunction with state values (e.g., life years, health-related quality-of-life weight, cost) to estimate life expectancy, quality-adjusted life expectancy, and expected lifetime costs.

An STM can capture many of the features present in the course of a disease or clinical process, such as the risk of disease over time, changing states over time, or episodic events, although this is not the only modeling approach that can capture these features (Caro, 2005). The principal advantage of cohort STMs is that they are relatively simple to develop, debug, communicate, and analyze using user-friendly software. The primary disadvantage of the cohort STM (i.e., Markov model) is the underlying assumption that the transition probability from one state to another does not depend on past history—neither the states in the past nor time spent in the current state. This assumption, known as the Markovian property, can
be a very limiting assumption for clinical applications where prior states and time in a given state tend to be strong
determinants of what happens next. The way that a Markov model handles the “memoryless” property is to create state
descriptions that include past history, either past states or time spent in a state. This approach can greatly increase the
number of states, resulting in very large models that are difficult to manage. This is referred to as “state explosion.”

In contrast, individual-based STMs are not limited by the Markovian property. The distinguishing feature of such models is
that they simulate one individual at a time. Such models fall into the category of microsimulation models, which are
evaluated using first-order Monte Carlo simulation. Whereas cohort models are analyzed as single cohorts progressing
through the health states simultaneously (which does not allow one to distinguish one simulated individual from another
except by the state descriptions), individual-level STMs can keep track of each simulated individual’s history (“tracker
variables”). This can greatly reduce the number of health states. The main disadvantage of individual-based models are that
they are computationally intensive, often requiring millions of individuals to be simulated to obtain stable estimates of the
expected value of the outcomes of interest. These types of models are also more difficult to debug compared to cohort
models. Figure 1 displays the Markov trace of a cohort simulation and the possible paths of an individual-level Monte Carlo
microsimulation. Table 1 compares the two model types regarding different criteria.

Figure 1. Cohort vs. individual-level Monte Carlo microsimulation in a state-transition model

(a) Cohort Simulation in a State-Transition Model
(b) Monte Carlo Simulation in a State-Transition Model

In a cohort simulation (a), the entire cohort is (re-)distributed across states after each cycle. In an individual-based
microsimulation (b), a finite number of individuals are simulated using first-order Monte Carlo microsimulation.

Table 1. Cohort vs. individual-level state-transition models
<table>
<thead>
<tr>
<th></th>
<th>higher (if number of states is limited)</th>
<th>lower</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ease of model development</td>
<td>higher</td>
<td>lower</td>
</tr>
<tr>
<td>Ease of model debugging</td>
<td>higher</td>
<td>lower</td>
</tr>
<tr>
<td>Ease of communication to non-experts</td>
<td>higher</td>
<td>lower</td>
</tr>
<tr>
<td>Markov assumption, memoryless</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Ease of modeling many different subgroups</td>
<td>lower</td>
<td>higher</td>
</tr>
<tr>
<td>Danger of explosion in number of states</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Distribution of outcomes (as opposed to only means)</td>
<td>possible, but technically more difficult</td>
<td>yes</td>
</tr>
<tr>
<td>Report of individual patient histories</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Decision-analytic software available</td>
<td>yes</td>
<td>yes (need advanced knowledge)</td>
</tr>
</tbody>
</table>

A decision-analytic state-transition model must start with a decision node, from which the branches of the intervention strategies originate. These branches either lead directly to one Markov node per intervention strategy (followed by an STM) (Figure 2-a) or the model contains an up-front decision tree leading to multiple Markov nodes per intervention strategy (Figure 2-b). In principle, an STM following one branch can have a different structure than one following another branch (see Recommendation 9 on model symmetry).
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 up-front decision tree modeling the first events and leading to multiple Markov nodes per intervention strategy.

2.3 Types of Interventions

Because decision-analytic models are used to compare different interventions or technologies, it is helpful to classify the general types of interventions for which STMs are used. We categorized the types of intervention as prevention, screening, diagnosis, and treatment.

**Prevention.** STMs used to evaluate primary prevention strategies are primarily concerned with the reduction in the risk of developing a disease, for example, coronary artery disease or cancer after postmenopausal hormone replacement therapy (Col et al., 1997). Hence, their primary focus is typically on what happens prior to disease, such as the number and severity of risk factors. The starting cohort for interventions aimed at primary prevention would be individuals who are free of disease or complications.

**Screening.** STMs are well-suited to evaluate different screening strategies (Eddy, 1980; Russel, 1994). In general, there are two types of screening: one-time screening, such as newborn screening (e.g., Grill et al., 2005) or genetic screening (e.g., Rogowski, 2009; Gutierrez de Mesa, 2007; Meckley et al., 2010); or repeated (interval) screening, such as screening for cancer (e.g., Goldie et al., 2006; Lee et al.; 2010, Sroczynski et al., 2011; Zauber et al., 2008) or HIV (e.g., Sanders et al., 2005; Sanders et al., 2008). The evaluated screening strategies can differ in many respects, such as the type and sequence of screening tests used, the diagnostic work-up modes, the screening interval, and the 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 (Trikalinos et al., 2009). Testing options may involve the use of one test vs.
Another, one vs. multiple tests, different test combinations or sequences, or using one positivity criterion vs. another within a single test (e.g., Gould et al., 2003) or a diagnostic score of multiple tests, or focusing the development of new diagnostic technologies (Hunink et al., 1999).

**Treatment.** We broadly define treatment as any intervention that is available for an individual who already has a clinical condition or disease that ultimately affects his or her health consequences and/or prognosis of that condition. The disease process of an STM should reflect the natural history of disease, the expected prognostic pathways in the absence of any intervention, and the effect of treatment on natural history, for example, in chronic cardiovascular disease (e.g., Cohen et al., 1994) or in chronic infectious disease (e.g., Sanders et al., 2005; Siebert et al., 2005).

3. Recommendations

A STM is often a reasonable choice when the decision problem can be framed in terms of states, interactions between individuals are not relevant to the model, and the population of interest is a closed cohort and not a dynamic population. Note that the recommendations below pertain to the development of models for the purpose of addressing a single policy question, and do not address the development of multipurpose disease-specific models aiming to address a range of questions from prevention, diagnosis, and treatment, such as the Cost Effectiveness of Preventing AIDS Complications (CEPAC) model (Freedberg et al., 1998).

3.1 Cohort vs. Individual Model Recommendations

**Recommendation 1: Choosing the type of state-transition model.**

*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 a cohort simulation and an individual-level simulation model, the modeler must define the characteristics of the cohort or individuals that must be carried through the model. In many cases, the type of STM is not chosen until after the initial conceptualization of the decision problem, or a cohort model is initially chosen and later changed to an individual-based model due to state explosion. The characteristics considered in the model (i.e., state descriptors or tracker variables) must include all relevant states pertaining to the disease or clinical process and intervention(s), all relevant history (e.g., past states visited, past risk factors, time in a current state, time since last event)
that is a determinant of transition probabilities (e.g., determinants of disease incidence, progression, mortality) and/or state values (e.g., determinants of utilities and costs).

An additional advantage of using an individual-level STM is the ability to model parameter such as risk factors as continuous variables. In cohort STMs continuous variables (e.g., blood pressure) has to be categorized, though some guidance exists for determining how many states to create (Bentley et al., 2009). In addition, the ability to evaluate dynamic intervention strategies – ones in which future decisions depend on current and past patient characteristics (e.g., in personalized medicine) – are more straightforward with an individual-level STM.

Two examples of STMs that required microsimulation include one developed by Lee and colleagues (2008) to compare intermediate and long-term clinical outcomes of different imaging screening strategies for breast cancer in women with BRCA1 gene mutations, and the United Kingdom Prospective Diabetes Study (UKPDS) model, an individual-level STM developed to estimate the long-term impact of health interventions for people with type 2 diabetes (Clarke et al., 2004). Details of the models are provided in Appendix 2.

3.2 Model Structure Recommendations

Recommendation 2: Statement of the problem.

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.

Intervention types may be single-time interventions such as one-time vaccination or surgery, static interventions over time that do not depend on intervention outcomes or any other events, or dynamic interventions (i.e., strategies) that consist of a sequence of decision rules for how to start, stop, or change interventions over time (Robins et al., 2004). Dynamic strategies are ones in which the specific strategy changes over time according to the preceding events. Examples of dynamic intervention strategies are 1) the start of a preventive behavioral intervention if BMI increases, 2) increase of the screening interval in cervical cancer screening if a woman has a repeatedly negative screening result, 3) the repetition of a diagnostic test in individuals after an equivocal test result, or 4) change to second line drug after first-line treatment failure.

This recommendation refers to standard STMs. However, there are other methods such as Markov decision processes (MDPs) that generalize standard Markov models by allowing embedding of sequential decisions in the model, and thus, multiple decisions can be made in multiple time periods. (Alagoz et al., 2010; Puterman, 1994)
Recommendation 3: Starting cohort.

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 (e.g., individuals defined by a set of demographic and clinical characteristics) allow for the comparison of alternative strategies for that cohort, such as whether individuals are screened or not. The comparison of strategies can be reported for different cohorts if the optimal decision may vary by subgroup characteristics (e.g., defined by age, sex, and risk factors known to the decision maker at the time of the decision). If model outputs are desired for a population-based starting cohort (i.e., all individuals for whom the intervention applies; for example, across different age strata) then the model would need to be run multiple times, one for each stratum, and then aggregated across strata. For example, Dewilde and Anderson (2004) demonstrated that the cost-effectiveness of cervical cancer screening varies greatly depending on whether a single birth cohort is simulated or a multiple birth cohort simulation is performed.

Recommendation 4: Defining states.

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 and the effects of interventions. These should be specified as mutually exclusive, meaning that any individual can be in only one state during each cycle. The states must also be collectively exhaustive, meaning that every individual in the initial cohort must be in a state during each cycle. Ideally, the model should include a set of health states that describe the biological/theoretical process of the disease or condition being modeled, with transitions among the states that would be expected in the absence of intervention. The states should adequately capture the benefits or harms of any interventions. These effects can characterize the state values, such as differences in symptoms and quality of life, or reflect changes in the probabilities of transitioning among states.

At the start of a cohort simulation, the modeled population must be allocated among the states. Each state is homogeneous; every individual in that state has the same transition probabilities. This implies that any characteristics that determine those probabilities must not differ within the state.

If an individual’s history is important in determining transition probabilities, either in terms of prior states or time spent in a current state, then the relevant states should carry that history as part of their definition for cohort models. For example, if the risk of a myocardial infarction depends on whether an individual has had a prior myocardial infarction, then the model’s states would need to include this historical element. In an individual-level STM, the characteristics that determine the transition probabilities and other model parameters can be heterogeneous within a health state but must be tracked throughout the model, and transition probabilities must be defined as function of these characteristics.
When there are multiple alternatives for modeling natural history, such as defining the health states with biological but often unobservable measures of disease (such as FEV\(_1\) percent predicted for asthma) or by the symptomatic descriptions of the disease (such as Hoehn & Yahr “on treatment” stages for Parkinson’s disease), the analyst should clearly justify the approach used to define the states or compare alternative methods in the sensitivity analysis. Although it may be possible to describe the natural history of a disease solely on the basis of health care utilization, such a model would not provide direct insight into the health outcomes of interventions at the biological level and would be limited in its value for most decision problems.

If cause of death is an important outcome, or different causes of death have different costs associated with them, then competing causes of deaths should be modeled in an unbiased way. For example, the total probability of death could be modeled first, followed by a conditional distribution of cause-specific deaths.

Another important consideration in structuring an STM is the modeling of 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 the short-term events 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 Markov cohorts for the time that has elapsed. For example, an up-front decision tree could be used to represent the events that occur within the first 6 months of a simulation, such as results and subsequent outcomes from diagnostic test strategies (Gould et al., 2003), treatments with limited treatment duration (Siebert et al., 2005), or surgical or procedure-related short-term events Cohen and colleagues (1994) developed a model comparing different percutaneous coronary interventions in patients with coronary artery disease. This model consisted of two decision trees for the immediate outcomes of the initial procedures and short-term (6 months) events linked to a post-revascularization STM for the life-time consequences (Figure 3).

Figure 3. Elements of a model for coronary stenting

The model consists of two decision trees and one Markov model. From Siebert (2002), based on Cohen et al. (1994) (permission requested)
Recommendation 5: Intervention effects.

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 developed to evaluate primary prevention strategies, the different possible risk factor levels in the target population should be represented as pre-disease states or tracker variables. Current risk factors are often predictors of future risk factors, thus the state descriptions and tracker variables should capture the natural course of risk factors and their changes in sufficient detail. Models evaluating preventive strategies may not require as much detail for post-disease description as would models evaluating intervention strategies targeted towards patients with the disease, but they should still be sufficiently complete to capture the relevant elements of the disease. One useful approach is for analysts developing a model to evaluate preventive interventions to collaborate with investigators with well-developed disease-specific models to derive outcomes related to the diagnosis of the disease of interest, such as lifetime costs or quality-adjusted life expectancy (Lieu et al., 1997; Berrington de Gonzalez et al., 2011).

Models that evaluate screening strategies should define states that reflect 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 probability of disease. For cancer screening models, health 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-time bias. In screening models with repeated or interval screening, dynamic screening strategies may depend on prior screening history ("individualized screening"). For example, some algorithms for cervical cancer screening recommend extension of the screening interval after repeated negative tests. In this case, capturing the screening history in the Markov states can easily lead to state explosion. Therefore, it may be necessary to model these types of strategies with an individual-level state-transition model, in which the comprehensive screening history can be included as tracker variables (Goldie et al., 2006).

In models that evaluate alternative diagnostic strategies, it is typical to have the testing pathways represented by a decision tree prior to utilizing STMs to simulate outcomes associated with different diagnostic realities, such as true-positive individuals (i.e., persons with the disease of interest and a positive test and thus are treated for their disease) or false-positive individuals (i.e., persons without disease but get treated nonetheless due to a false-positive test result). If multiple diagnostic tests are performed in combination or in sequence, and some of these diagnostic markers are also prognostic factors that can change over time, the history of such diagnostic and prognostic markers must be incorporated into the model’s states or implemented as tracker variables.

In models that evaluate alternative interventions for individuals who already have disease ("treatment"), the modeler should be specific about the mechanism in which the treatment alters the course of disease (either the natural history of disease or the course of disease under “usual care”), whether it be a reduction in risk of events, reduction in mortality, slowing the progression of disease, etc. In addition, the modeler should specify how the harms of the intervention(s) affect the prognosis of the patients. STMs should incorporate realistic assumptions about adherence to interventions over time.
The effectiveness and costs of long-term treatment often depend on time-varying heterogeneous patient characteristics. In particular, many medical treatment regimes are “personalized” and follow dynamic rules. For example, the future 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 of the STM.

Recommendation 6: Heterogeneity.

States need to be homogeneous with respect to both the 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, such as baseline age, gender, presence of comorbidities, and stage of disease among newly diagnosed patients, are known at the time of the decision and can be used to define the starting cohorts (see Recommendation 3). 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 with a low number of comorbidities at the start may develop more comorbidities over time, and to capture this a modeler would need to incorporate this attribute within the state space. Variables that affect transition probabilities but that are not known at the time of decision making, such as genetic mutation or undiagnosed infection, have the potential to create “heterogeneity bias” (Kuntz & Goldie, 2002; Zaric 2003) and inclusion of such variables should be considered.

Recommendation 7: Time horizon.

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 of a STM relates to the number of cycles it runs or the duration for which the initial cohort is tracked. Common approaches to this issue include running a model up to a maximum age of 120 years or tracking the cohort until more than 99.9% of the modeled individuals are dead. In particular, if the intervention affects mortality, the time horizon should be lifetime in order to capture (quality-adjusted) life-years gained from prevented or delayed deaths.

Recommendation 8: Cycle length.

Cycle length should be short enough to represent the frequency of clinical events and interventions.
The choice of cycle length should be based on three factors: the clinical problem, the remaining life expectancy, and computational efficiency. The cycle length should allow transitions to occur in a manner consistent with the clinical problem and the intervention effects. For example, a model to assess monthly screening requires cycles no longer than one month. Cycle length should be chosen short enough so the likelihood that a specified event relative to the course of disease occurs at most once per cycle. STMs provide an approximation of life expectancy (area under the survival curve). Thus, the shorter the cycle length the better the approximation is. If the life expectancy is relatively short, such as for individuals with an acute disease or a cohort of older age, then a shorter cycle length should be considered for purposes of obtaining a less biased estimate, even if the nature of the clinical problem does not warrant the shorter cycle length. Although shorter cycles will always yield less biased estimates, the size of the bias gets very small when the number of cycles required to run the cohort increases.

**Recommendation 9: Model symmetry.**

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

Symmetrical 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 vs. initial medical therapy should both specify true underlying disease status (i.e., angiographic extent of coronary artery disease) even though the true disease status is not observed in the medical therapy model (Sonnenberg et al., 1994). If the underlying extent of disease is not specified in the medical therapy arm of the model, the model will produce erroneous results when conducting sensitivity analysis on the underlying probability of any particular anatomy (e.g., left main disease) in the population of interest.

### 3.3 Data Recommendations

STMs should provide clear justification for how the estimates and ranges for sensitivity analysis were derived for the transition probabilities and state values ("rewards"). Analyses should be based on the best available evidence. The quality of the evidence may vary among model inputs, and for some inputs, subjective estimates may be required.

**Recommendation 10: Data sources.**

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

Ideally, transition probabilities pertaining to the natural history of a disease process or condition being modeled should be derived from population-based epidemiological studies, as these data are likely to be the most representative. Baseline
transition probabilities may be derived from the control arms of intervention studies, with the recognition that these
probabilities may be less generalizable due to the selection criteria for the trial participants, or from other types of
observational studies.

If data from 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 the
quality of individual studies are available (Owens et al., 2010; Balshem et al., 2011; Guyatt et al., 2011). In the absence of
good systematic reviews, a detailed summary table of the evidence should be provided in an appendix with a description
and justification of how key parameters—including the ranges used for sensitivity analyses—were derived.

Recommendation 11: Derivation of parameters.

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

Transition probabilities and rates should be used appropriately (Miller and Homan, 1994). The conversion of transition
probabilities from one time unit to another (e.g., annual transition probabilities to monthly transition probabilities) should
be done through rates. Rates should never be presented as percentages. To avoid confusion, probabilities should never be
called rates.

The assumed functional relationship between disease-specific mortality and background (life table) mortality should be
stated explicitly. Because an assumption of additive rates can give very different results than the assumption of
multiplicative rates (Kuntz & Weinstein, 1995), modelers must be clear in their assumption and diligent in evaluating the
impact of this assumption.

Recommendation 12: Intervention effects.

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.

The manner in which an intervention affects the course of disease should be clearly described and justified. For example,
the analyst might assume that an intervention improves the quality of life for a given state or states by reducing the
transition probability to a progressed state or by decreasing the number of symptoms (i.e., increasing the utility assigned to
a state). If several plausible ways exist by which an intervention may affect the natural history component of the model,
then they should be evaluated in sensitivity analysis.

Efficacy parameters derived from randomized clinical trials may have to be adjusted for compliance in order to reflect real
world effectiveness in routine practice (Cox et al., 2009). Effectiveness parameters 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 the intervention and the outcome) require special statistical methods such as marginal structural methods or g-estimation (Robins et al., 2004; Johnson, 2009). When results from observational studies are used in the model, causal graphs can be used to explicitly state causal assumptions (Cox et al., 2009).

A reported reduction in all-cause mortality for an intervention should not be applied directly in a model as a reduction in all-cause mortality since there is a large increase in life-table ("background") mortality with age. If data for disease-specific mortality are not available, a relative risk reduction could be applied to disease-specific mortality and would be a conservative estimate of the treatment benefit. Alternatively, life-table mortality could be subtracted from total mortality to derive an estimate of the reduction in disease-specific mortality (Kuntz and Weinstein, 2001).

For primary or secondary prevention interventions, if evidence is available for the reduction in disease or event incidence and the reduction in mortality, the modeler must be careful not to double count when using both parameters (i.e., a reduction in disease/event risk and a reduction in mortality conditional on the getting the disease/event). If there is concern about double counting, calibration exercises should be conducted to evaluate the consistency of the model-generated reductions in disease and mortality with the estimates from clinical studies.

Recommendation 13: Valuation of states.

The valuation of intermediate outcomes/states should be justified. The expected outcomes derived from STMs depend on the values assigned to each state. For example, quality-adjusted life years can be derived if utilities are assigned to each state. State values assigned should be justified, preferably based on theory.

3.4. Recommendations for Analysis

This section, which focuses primarily on individual-level models, offers recommendations on best practices in analyzing a STM and the performance of appropriate uncertainty analyses.

Recommendation 14: Half-cycle correction.

A half-cycle correction factor should be applied to both costs and effectiveness and should be applied in the first cycle. A half-cycle correction should also be applied in the final cycle for analyses that do not use a lifetime horizon.
Common modeling convention is that transitions occur in the middle of each cycle. To properly account for this assumption, a half-cycle correction is generally performed, which assigns one-half of the state reward for simulated individuals starting in each state. Giving simulated individuals a full cycle at the start of the model (i.e., transitions occur at the beginning of the cycle) overestimates the expected values; assigning simulated individuals no value at the start of the model (i.e., transitions occur at the beginning of the cycle) underestimates expected values (Sonnenberg and Beck, 1993).

**Recommendation 15: Analyzing distributions.**

*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 to know whether a new treatment with a gain in life expectancy of one year extends life by one year for each person or it extends life by three life years in half of the people and reduces life by one year in the other half. Distributions of outcomes are straightforward to derive from individual-based STMs, but can also be derived with some effort from the Markov trace of a cohort model. Alternatively, cohort models can be run as individual-based models to generate distributions of outcomes.

**Recommendation 16: Performing a microsimulation.**

*The number of individuals modeled in an individual-based simulation should be large enough to generate stable estimates of the expected value of interest.*

In order to achieve stable modeling results in an individual-based Monte Carlo simulation, it is necessary to use a sufficiently large number of first-order trials (i.e., random walks simulating one individual at a time). The analyst should assess the stability of model results by calculating the variability in the results from multiple model runs with identical number of trials. The expected variance with a larger number of trials can be calculated (Kuntz and Weinstein, 2001). The variance in the expected outcomes should be much smaller than the smallest difference expected between strategies. Variance reduction techniques such as the use of common random numbers can be used to reduce the number of trials necessary (Stout and Goldie, 2008).

### 3.5 Recommendations for Communicating Model Results

The goal of a graphical representational of an STM is to communicate the key model structure and the most important assumptions that have been made regarding states and allowable transitions. Modelers should strive for transparency and a comprehensive—though not necessarily exhaustive—description of the model. For most reports, simple and accessible
model descriptions in the body of a manuscript can be enhanced with an appendix that provides more detail about modeling assumptions, parameterization of these assumptions, valuation of the parameters, and supplementary results. Since rigorous studies evaluating alternative methods of presenting data are lacking, these recommendations represent our best judgments based on our collective experiences as authors and editors.

Recommendation 17: Presenting the model.

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

Modeling parameters such as transition probabilities, costs, or utilities, and modeling assumptions such as the time horizon, are often best presented in tables or text rather than added to figures. A typical table includes:

- A description of the transition probabilities and state values plus a description of intervention variables that affect benefits and harms. These descriptions should be in plain language rather than reporting actual variable names.

- An estimate of the base-case value of the parameters.

- An indication of the uncertainty associated with the estimate. For deterministic models, the uncertainty is represented by a range of values indicating plausible estimates that will be tested in sensitivity analyses. For models that incorporate probabilistic sensitivity analyses, the uncertainty should be represented by a measure of uncertainty (usually the standard error or the mean) and an indication of which distributions were used for each parameter. When analysts use the method of moments to calculate distribution parameters, it is preferable to report the mean and standard error alongside the type of distribution rather than the parameters of the distribution, since the former will be more familiar to most audiences.

- References to the sources used to derive the parameter values. When parameter values are based on opinion or assumptions, this should be clearly stated in the table with justifications provided in the text.

It is often useful to organize parameters in a table according to their roles in the model. In general, parameters for an STM can be classified as describing the natural history of a disease, the effectiveness and risk of an intervention, and state values (e.g., utilities, costs). To facilitate communication, all parameters should incorporate a similar time frame (e.g., annual costs and annual transition probabilities). When the time frame is different from the model's cycle length, the text or table footnote should note that difference.

Parameters that represent efficacy estimates from clinical trials should be clear how they are using the reported efficacy measure (e.g., risk difference, risk ratio, odds ratio, hazard ratio). When effect measures are in formats different than natural history measures, the text should clarify how these were combined (e.g., multiplicative or additive models). The description of the parameter in the table should also indicate whether treatment effects are modeled as relative or absolute differences from the natural history model. For functions of other parameters such as the Framingham score for
cardiovascular outcomes or utility functions depending on component parameters, the complete formula with coefficients should be provided.

Figures should include plain language descriptions of states and avoid using model variable names or other shorthand notation. 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 represent transitions as arrows between these compartments. While 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 diagrams are desirable. Markov cycle trees, often stylized, can display the transitions between states, and probabilities and transitions that are conditional upon other events or parameters. For options to simplify the graphical representation of complex STMs see Appendix 3.

Recommendation 18: Presenting model results.

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 event risks), the average number of events per lifetime, the percent of the initial cohort that experienced two or more events in their lifetime, or the 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. For example, a model of atrial fibrillation for stroke prevention could report the average amount of time spent without a stroke and the average time from the first stroke to death.

In addition such measures can be used for validating model results with empirical data. As STMs allow deriving the time at which particular transitions occur, the model results can be represented as (modeled) probability or survival curves (Figure 4) and directly compared (e.g., superimposed) to Kaplan survival curves from empirical studies.
3.6 Validation and Consistency Recommendations

STMs are an abstract representation of a real-life problem under simplifying assumptions. Therefore, ensuring that the model provides a sufficiently accurate representation of the real system is an important part of any modeling study. This is achieved through model validation. General recommendations on model validation can be found in the paper “Model Transparency and Validation” of this series (Eddy et al., 2011). In this section, we provide validation examples relevant to STMs.

Internal validity

The first step in validation is to ensure that the model does not contain any errors (i.e., debugging the model). In addition to the output variables of interest, the analyst should generate various model outcomes that would help identify potential issues with the model. For example, a modeler could generate the mean transition time between two health states and then check whether the estimate is reasonable or not. A more advanced method for internal validation is tracking individual entities of the model (in particular in individual-level STMs) while recording detailed, time-stamped events that are experienced by the individual entity throughout the model run.

A useful method of identifying programming errors is to check whether helpful (but not necessarily mandatory) model-building rules such as the use of symmetric branches or states in STMs are followed in the model. In addition, an extreme conditions test should be used to determine how resistant the model is to extreme changes in the variables. For example, in an STM used to evaluate the decision between treatment and no treatment, increasing the immediate mortality rate of the treatment to 1 should lead the model to choose no treatment as the best therapy. Any conflicting result is a sign of an error in the model. An extreme conditions test can be conducted for all model parameters.
Another debugging technique is to set the effect parameters and intervention costs to the null value (e.g., relative risks = 1, risk differences = 0, drug costs = 0) in the intervention branch, and then compare results with the no intervention branch (there should be no difference). Inspection of the Markov trace can also help find errors. For example, the analyst can set the model parameters in such a way that it is easy to predict how the trace will look (e.g., that no modeled individual will visit a state during the simulation) and ensure that intuition holds.

Sensitivity analysis – primarily used for uncertainty analysis – also provides a powerful tool for testing the internal consistency/validity of an STM by challenging the model and investigating if it provides the expected outcome. Sensitivity analysis can be conducted at two levels in STMs: parametric sensitivity analysis and structural sensitivity analysis. Parametric sensitivity analysis involves running the model with different parameter values to determine how the changes affect the results. For example, in an STM considering the benefit of a screening program, as the cost of screening decreases, one would expect the screening option to become more cost-effective. A contradictory result suggests a modeling error. Structural sensitivity analysis involves investigating the effects of using a different structure than the existing one on model performance. For example, one could test the effects on model results of omitting specific (uncertain) transitions or collapsing certain states into one state. A less-aggregated model can also be used to ensure that the simplifications of model development do not change the conclusions of the study.

Face validity

Evaluating the face validity of a model ensures that its findings make sense and can be explained at an intuitive level. An STM should represent the sequence of events similar to an existing clinical process. Detailed self-documentation, including clear definitions of all variables used in the model, is essential for face validity. In addition, the labels for the model components should be intuitive. For example, a model for breast cancer screening representing cancer stages as part of the state space should use well-known stage definitions of cancer.

Validation of model assumptions

STMs include two types of assumptions: data assumptions and structural assumptions that involve the way real-life biologic and clinical processes are represented. Validation of model assumptions should include verification of both types of assumptions. Structural assumptions, such as the sequence of events in a clinical process, should follow real clinical processes. These assumptions can be confirmed by observing the actual process and through discussions with clinical experts in that field.

External validity
External validation (often referred to as external consistency) involves determining whether the model outputs are consistent with the findings from relevant research that was not used in estimating model parameters. For example, in STMs based on histological disease states in chronic hepatitis C, the incidence of liver cirrhosis derived from the modeled progression of mild hepatitis to moderate hepatitis to liver cirrhosis can be compared with the results of registry studies reporting the overall 20-year incidence of cirrhosis (Siebert et al., 2005) to check the validity of the model. In such validation exercises, it is important to mimic the population of the published empirical study in the STM. Specifically, the properties of the registry cohort (such as age, gender distribution, all individuals start in mild disease, etc.) must be used in the model during the evaluation, even if these properties are not the properties of the target population of the actual modeling study.

Predictive validity

Predictive validity is useful, but it is often not feasible when the model is being developed or initially used. Some authors (ISPOR Task Force, 2003, Weinstein et al., 2001) note that a model's ability to predict future events is valuable but not essential, whereas others (Hay and Jackson, 1999; Sendi et al., 1999; Soto, 2002) suggest that a valid model should be able to predict future scenarios. Since models are developed as aids to decision making at a particular point in time and not necessarily to predict the future, there is no empirical test of predictive validity (Sculpher et al., 2000; Sonnenberg et al., 1994; Phillips et al., 2006). Furthermore, it is important to acknowledge that the outputs of a model depend on the input assumptions that are mostly based on historical information that may not hold for the future cases.

4. Conclusions

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

Disclaimer: 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.
5. References


Appendix 1: Terms Used in this Report

**State.** A state is a condition that the modeled unit (e.g., a patient) can be in. It is characterized by aspects such as quality of life (e.g., utility), severity of disease (e.g., mild/moderate/severe, cancer stage, New York Heart Association classification), prognosis (e.g., risk of progression, mortality), use of resources (e.g., direct costs of medical treatment) and relevant history (e.g., occurrence of prior events, time spent in state).

**State homogeneity.** For cohort models, all individuals residing in a particular state are indistinguishable from the others in that state. This feature does not apply to states in individual-level models.

**Transitions.** A transition is the shift of people from one model state to another. The likelihood of a specific transition is known as the “transition probability.”

**Initial state vector.** Describes the probabilities of residing in each of the states at the beginning of the simulation (i.e., at time 0).

**Cycle/cycle length.** A cycle is a predefined interval of time (the “cycle length”) on which the transition probabilities are based.

**State values.** State values (“rewards”) are values assigned to the time spent in a particular state. Often rewards represent the health outcomes (e.g., life years, quality-adjusted life years) or costs for residing in a state for one cycle.

**Half-cycle correction.** An adjustment method that is made to implement the assumption that all transitions in an STM occur in the middle of each cycle.

**Termination criteria.** Stopping criteria are the rules applied to determine when an STM stops cycling.

**Natural history of disease.** Natural history describes the disease process (e.g., falling ill, progression, regression, symptom detection, dying) in the absence of an intervention.

**Hazard rates vs. probabilities.** Hazard rates represent the instantaneous risk of an event (i.e., incidence density, values range from zero to infinity) while transition probabilities (values range from zero to one) in an STM represent the chance of an event happening during a particular interval of time.

**Deterministic vs. stochastic.** A deterministic model provides the same results for a given set of inputs every time it is run because none of the elements that determine the results are subject to random variability. In contrast, in a stochastic model the results depend on random number draws and will not provide the same results for a given set of inputs despite maintaining exactly the same set of inputs.

**Cohort vs. individual-level simulation.** These terms refer to the manner in which STMs are evaluated. In a cohort simulation, the modeled cohort is progressively partitioned amongst the model states each cycle in a deterministic fashion.
In contrast, in an individual-level simulation, the modeled cohort is simulated one individual at a time, and an individual’s progression through the states is determined stochastically.

**Cycle-based vs. event-based models.** Cycle-based models refer to models for which transitions can only occur at specified time steps (i.e., cycles). Event-based models refer to models for which transitions occur at the time of an event.

**Interaction vs. no interaction.** Models for which interaction is allowed are those where the transition from one health state to another depends on the states in which the other individuals reside. For example, transitions to an infectious disease state might depend on the prevalence of the disease in the population (as in dynamic transmission models), or admission to a hospital might depend on the availability of a hospital bed (as in discrete-event simulation models).

**Cohort vs. population.** A single (closed) cohort is a group of subjects for whom membership is defined in a permanent fashion and no members can enter the cohort after it is defined. In contrast, a dynamic (open) population allows new members who were not the initial cohort to enter and existing members to exit the population. Subjects can enter an open population through various mechanisms (e.g., by birth, migration, by reaching a certain age, developing a symptom). Similarly, subjects can exit (e.g., by dying, aging out, emigrating, or developing a condition that disqualifies them from being a member of the population).

**Markov model.** A cohort STM.

**Markov trace.** Table that shows the distribution of the states in an STM at each cycle.

**Monte Carlo.** A term that refers to a random draw. A first-order Monte Carlo analysis simulates individuals one by one. In an individual-level STM, transition probabilities and a random number generator are used to produce a subject’s path through the states. This path is called a random walk or a “trial.” Counters (also known as tracker variables) can record the entire history of constant and time-varying patient characteristics and events along the subject’s path.

**Microsimulation.** A modeling technique that operates at the level of individual units (e.g., patients), where each unit is evaluated one at a time. The simulation results include both the total aggregate outcomes and the distributions of these outcomes. Individual-level STMs are one type of microsimulation model.
Appendix 2: Examples of State-Transition Microsimulations

Example 1: Imaging screening strategies for breast cancer in women with BRCA1 gene mutations

Lee and colleagues (2008) developed an individual-based STM to compare the intermediate and long-term clinical outcomes of different imaging screening strategies for breast cancer in women with BRCA1 gene mutations (Figure 5). In this model, the probabilities of clinical detection of invasive breast cancer and ductal carcinoma in situ were a function of the current tumor diameter, and the probabilities of lymph node involvement and distant metastases were a function of primary tumor diameter at diagnosis. The probability of each event after treatment was a function of breast cancer stage at diagnosis, estrogen receptor status, and age at diagnosis. Monte Carlo microsimulation made it possible to track tumor diameter and other characteristics of each individual throughout the model.

Figure 5. State-transition microsimulation model for breast cancer imaging screening

Overview of natural history and screening model.

Overview of a natural history and screening model for women with the BRCA1 gene. Ovals indicate health states for patients who are alive. Arrows indicate the state in which patients begin each cycle and point to the state into which they enter during a Markov cycle. After presenting with clinical symptoms of breast cancer, women entered the Staged and Treated Breast Cancer state. With the addition of screening (in this example, with combined annual mammography and MR imaging), asymptomatic women with breast cancer were either given diagnoses with screening or presented between screening events with clinical symptoms. (permission requested)

Example 2: United Kingdom Prospective Diabetes Study (UKPDS) type 2 diabetes model

The UKPDS model is an individual-level STM, which was developed to estimate the long-term impact of health interventions for people with type 2 diabetes (Clarke et al., 2004). The main aim of the model was to estimate the first occurrence of each of eight diabetes-related complications (fatal or non-fatal myocardial infarction, other ischemic heart disease, stroke, heart
failure, amputation, renal failure, eye disease measured in terms of blindness in one eye, and death) in order to estimate lifetime outcomes and quality-adjusted life expectancy (Figure 6). In particular, the impact of changing risk factors such as blood glucose level, blood pressure, lipid levels, and smoking status on life expectancy and quality-adjusted life expectancy was assessed for several risk interventions. The model is based on 14 risk functions predicting the absolute risk of any of the above complications based on a patient’s demographic characteristics, time-varying risk factors such as HbA1c, and history of complications. In this model, the states were alive and dead, and all other characteristics including past events were updated in annual cycles.

Figure 6. UKPDS microsimulation model.

From Clarke et al. (2004). (permission requested)
Appendix 3: Some Options for Producing Simplified Graphical Model Representations

- Transitions from multiple states into a single state can be represented by a single bracket rather than multiple arrows. For example, models that include mortality as an outcome can represent death as a single state at the edge of the figure and a large bracket indicating that all states can lead to death.

- States that are similar can be grouped together. For example, a model might include separate states for disease-specific mortality and for all-cause mortality. In a figure, these can be grouped into a single state of death. Such considerations might be particularly important for cohort models in which extra states have been added to incorporate “memory” into the model. For example, a model might have separate states for first, second, or third myocardial infarction, but a figure might group these together to enhance clarity.

- States that differ by only one characteristic can be represented by designations at the side of the figure. For example, a model may have two states that appear to be similar, but one includes toxicity and the other does not. A single state annotated with text such as “± toxicity” will often suffice.

- States might include information about multiple characteristics of the cohort, each of which is important when considering transitions. For example, states in a model of human immunodeficiency virus infection might include information about CD4 count, viral load, treatment regimen, and AIDS illness. Separate figures representing each transition is often more accessible than a single figure representing all transitions. Some straightforward transitions can also be described in the figure legend (such as from high to low viral load count) rather than graphically displayed.
Appendix 4: Table of Recommendations

Recommendation 1: Choosing the type of state-transition model. 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.

Recommendation 2: Statement of the problem. 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.

Recommendation 3: Starting cohort. 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).

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

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

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

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

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

Recommendation 9: Model symmetry. Components of state-transition models that reflect similar clinical courses should not be recreated but rather should be incorporated once and linked to that structure throughout the model.

Recommendation 10: Data sources. Transition probabilities and intervention effects should be derived from the most representative data sources for the decision problem.

Recommendation 11: Derivation of parameters. All methods and assumptions used to derive transition probabilities and intervention effects should be described.

Recommendation 12: Intervention effects. 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.

Recommendation 13: Valuation of states. The valuation of intermediate outcomes/states should be justified.

Recommendation 14: Half-cycle correction. A half-cycle correction factor should be applied to both costs and effectiveness and should be applied in the first cycle. A half-cycle correction should also be applied in the final cycle for analyses that do not use a lifetime horizon.

Recommendation 15: Analyzing distributions. For certain decision problems, it may be important to report not only the expected value but also the distribution of the outcomes of interest.
Recommendation 16: Performing a microsimulation. The number of individuals modeled in an individual-based simulation should be large enough to generate stable estimates of the expected value of interest.

Recommendation 17: Presenting the model. The report should use nontechnical language and clear figures and tables that enhance the understanding of the model to communicate its key structural elements, assumptions, and parameters.

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