Discrete Event Simulation: The Preferred Technique for Health Economic Evaluations?

Jaime J. Caro, MD,1,2 Jörgen Möller, MSc,3 Denis Getsios, BA4

1Department of Epidemiology, Biostatistics and Occupational Health, Division of General Internal Medicine, McGill University, Montreal, QC, Canada; 2United BioSource Corporation, Lexington, MA, USA; 3United BioSource Corporation, Eslov, Sweden; 4United BioSource Corporation, Halifax, NS, Canada

ABSTRACT

Objectives: To argue that discrete event simulation should be preferred to cohort Markov models for economic evaluations in health care.

Methods: The basis for the modeling techniques is reviewed. For many health-care decisions, existing data are insufficient to fully inform them, necessitating the use of modeling to estimate the consequences that are relevant to decision-makers. These models must reflect what is known about the problem at a level of detail sufficient to inform the questions. Oversimplification will result in estimates that are not only inaccurate, but potentially misleading.

Results: Markov cohort models, though currently popular, have so many limitations and inherent assumptions that they are inadequate to inform most health-care decisions. An event-based individual simulation offers an alternative much better suited to the problem. A properly designed discrete event simulation provides more accurate, relevant estimates without being computationally prohibitive. It does require more data and may be a challenge to convey transparently, but these are necessary trade-offs to provide meaningful and valid results.

Conclusion: In our opinion, discrete event simulation should be the preferred technique for health economic evaluations today.

Keywords: decision tree, discrete event simulation, economic evaluation, Markov, modeling.

Introduction

Despite enormous investments in research on the effects of new health technologies, the resulting data will almost always be insufficient to inform health-care decisions. Information coming from clinical trials may have strong internal validity, but it will suffer from poor applicability to the messier world of actual practice. Moreover, it will tend to be short term and cover only an extremely limited range of available options. Cost data from clinical trials are even less usable in most cases because the study environment so modifies practice that the resulting resource use is a poor reflection of reality. These data are unlikely to be specific or adequately transferable to a particular country and, in any case, will not provide information on the longer-term consequences of adopting the new technology [1,2]. Indeed, economic data from clinical trials may not be available at all. Thus, modeling of the economic outcomes is an essential component of the evaluations [3].

Models inform decisions when relevant, real-world data are not yet available [4]. They can be used to test a wide range of scenarios and strategies to identify the most efficient and equitable allocation of resources and allow extrapolation to other countries or regions and other populations. To be useful tools, models must reflect reasonably well what is understood about the illness and its management that is relevant to the problem at hand. This face validity requires that the chosen technique be able to include all the pertinent components and that it be capable of handling the known relationships in the data accurately. Although models always involve some simplification of reality, it is important that these assumptions not distort the picture to the extent that the result is misinformation [5].

In this article, we argue that discrete event simulation is the preferred modeling technique for health economic evaluations, if these are to be sufficiently accurate to be taken seriously when informing health-care decisions, and that this must take precedence over familiarity and the ease of conveying the methods. It is also argued that the prevailing Markov approach is rarely adequate for this purpose.

Elements of the Choice

Modeling techniques can be characterized according to many features [6,7], but two particularly important ones for defining the choice for health economic evaluations are the nature of the object that is modeled and the basis for conceptualizing what happens to those objects. The basic options for objects are to model a population as a whole (often referred to as cohort or aggregate modeling) or to model the individuals in the population (also known as patient-level simulation). The main alternative conceptualizations are to represent the problem as a series of states that the objects can be in or to think instead of the events the objects can experience. Multiple variations of these approaches exist, including hybrids, but it is common practice in health economics today to use a cohort, state-transition technique—the typical Markov model. By contrast, an event-driven individual method—discrete event simulation—has been proposed as a better choice [8].

Cohort versus Individual

There are two major problems with a cohort approach, both readily illustrated with the simplest of examples. Suppose we are evaluating a treatment that prevents the transition from being healthy to being sick—for example, having a myocardial infarction. In the cohort approach, the entire population is healthy at the start; and after a given time period, some proportion will have had the myocardial infarction (Fig. 1).

One major problem is in determining that proportion at each relevant time point. That proportion is given by the risk the
population is exposed to more than a given period, and that risk is affected, presumably, by treatment. It is not only treatment that modifies the risk, however. It will also depend on patient characteristics, such as age, sex, smoking, and other risk factors. That is why we are interested in characterizing the population and examining these factors. So the risk depends on the characteristics of the population, but how do we express these and compute the transition probability? The common approach of using the “mean” profile given by the average value of each characteristic will not yield the correct answer because the factors never distribute perfectly normally with no correlations among them. But there is an even bigger problem in determining the proportion. The flow out of healthy to sick is not random: It is the higher-risk people who tend to move out earlier—the older ones, those who smoke, and so on. This phenomenon, known in epidemiology as depletion of susceptibles [9], means that characterizing the cohort population over time becomes extremely difficult because the distributions of risk factors are being altered by the transitions. The typical approach today ignores this and simply adds the cycle time to the age, assumes the proportions of males and smokers are constant, and so on. This is incorrect (Fig. 2) and will yield inaccurate results.

A similar problem occurs with the portion of the cohort that has become sick. It is very difficult to characterize these patients in terms of features that may be determinants of further risk, mortality, quality of life, resource use, and other elements relevant to the model because the arriving fractions mix into the ones already there. Thus, it is difficult to accurately estimate these important items. By the same token, if any of them depend on the duration of illness (for example, how long it has been since the myocardial infarction occurred), the estimates will be inaccurate given that the arriving portions mix into a single group and do not retain any “memory” of when they became sick.

An attempt at solving these problems is to define the cohort as homogeneous—for example, only 65-year-old male smokers. Given that everyone in the population is the same in terms of the modeled risk factors, it is hoped that the issue goes away. For most applications, however, the number of homogeneous populations required to reasonably reflect the profiles defined by combinations of modeled risk factors would be vast, forcing the analyst to limit these to a feasible few. Even then, the depletion of susceptibles still occurs, and use of a constant probability would be inappropriate.

A solution to the problem of losing memory of the time when sickness started is to keep the arriving subgroups separate according to the time the sickness started (so-called “tunnel states” [7]), but this greatly increases complexity and is only feasible for a very limited number of items that need to be remembered [10].

The second major problem with the cohort approach is in the application of competing risks over time. If we add another condition to our example, say, death (Fig. 3), then at each cycle one must determine the proportion of the population transition-
myocardial infarction would be the transition from healthy to healthy and one for sick—and a third state for dead (Fig. 3). The in our simple example, there would be two states—one for of snapshots of the states that the population may be in. These competition means that some people will not be affected by one risk because the other one “got them” first—people who die are no longer exposed to the risk of myocardial infarction, for example. So the population will manifest a lower probability than would be the case if it was subject to only the risk of myocardial infarction, but deriving that value is very difficult, as has long been recognized in epidemiology and biostatistics.

All the serious problems with the cohort approach are readily solved by modeling individuals instead of the entire population in the aggregate. For each individual (known as entities in discrete event simulation), the risk can be computed based on their characteristics, the values of these factors can be easily updated over time as appropriate, the risks can be recalculated when necessary, and any number of competing risks can be correctly applied by deriving the implied time to each event. Individuals can face any number of risks simultaneously. When an event happens, their characteristics can be updated without problem. If there are time dependencies, these can be considered, and of course, what happened before can be remembered and used appropriately. There is no need to work at the mean or restrict the analyses to homogeneous populations or to deploy any of the variety of work-arounds, such as applying one risk before the other or randomly ordering the risks, common in today’s Markov models.

To be sure, individual modeling requires more calculations and is somewhat more difficult to implement, especially if the implementation is required to be in a spreadsheet, but cohort models will not produce accurate estimates and will force untenable assumptions. Choosing a cohort approach would only be appropriate if population characteristics do not affect transition probabilities, there are no competing risks, and there is no need to reflect dependence on time or prior history and other such features of the course of illness. Clearly, these conditions never hold in health economic evaluations.

### State versus Event

In a state-transition model, the world is conceptualized as a series of snapshots of the states that the population may be in. These snapshots occur at fixed, discrete time points called cycles. Thus, in our simple example, there would be two states—one for healthy and one for sick—and a third state for dead (Fig. 3). The myocardial infarction would be the transition from healthy to sick, and the model would derive the distribution of the population into these three states after each cycle.

This is an awkward way to conceptualize medical conditions because they tend not to fall into discrete states, especially ones that must be mutually exclusive. Trying to represent the condition in this way leads to either a vast number of states (in our simple problem, one would have to consider things like diabetes, hypertension, smoking, hypercholesterolemia, and so on) corresponding to all the possible combinations or to inaccuracies resulting from the required simplification. Moreover, at least as commonly implemented in health economics, the snapshots are taken at fixed, discrete intervals, and only one transition is allowed in between—both limitations that lead to further inaccuracy.

Not only does the state approach lead to considerable unnecessary complexity, it also leaves no room for important elements of the problem that cannot be represented as states. For example, the physician may make several decisions when the patient suffers a myocardial infarction, and even beforehand, treatment may be altered according to results of tests and the patient’s experience. Occurrence of the infarction can lead to an emergency room visit where a series of actions take place, perhaps culminating in hospitalization. Depending on what is happening, various risks, including that of death, may be changing. There is no clear way to represent any of this in a state-transition model.

It is much more natural to conceptualize the world in terms of the events that can happen. By thinking in terms of events rather than states, the problems are solved. It is very easy to design the model in terms of what can happen. The myocardial infarction is an obvious event, but so is the visit to the doctor where various tests happen, decisions are taken, and treatment is altered. Arrival in the emergency visit and any number of actions taken there including hospitalization, and of course, at any point, death is similarly of the event form. The resulting design is much more compact and transparent, yet of greatly increased accuracy, than any comparable state-based design. Indeed, it is difficult to think of any reason, other than familiarity with the prevalent approach, for choosing a state-based representation. The original one, having to do with the easy mathematical solution to a matrix reflecting the Markov chain transitions, does not hold in the vast majority of situations faced in health economics.

### Discrete Event Simulation versus Markov Model

In a discrete event simulation, the experience of individuals is modeled over time in terms of the events that occur and the consequences of those events. This approach is far preferable to the typical Markov one, which tries to shoehorn the world into a series of states more than which a cohort is successively distributed. All of the many limitations and inaccuracies of Markov models are easily avoided with discrete event simulation. There are some problems that arise, however.

Modeling at the individual level requires a much greater number of calculations. This is rarely a problem with the computing power readily available in laptops today, provided that two aspects are attended to. One is that the model should not be forced to process one individual at a time. Doing so is very inefficient because much of the time nothing relevant is happening to that individual, yet the model has to process the entity anyway. This problem is compounded if the simulation is set up to consider only one time unit (for example, a day) at a time. Instead, the model can and should be allowed to consider many individuals simultaneously. This allows the model to efficiently process events as they happen throughout the population and
substantially reduces the number of calculations, particularly if
time is permitted to jump to the occurrence of the next event
rather than proceed in fixed, even units.

The other aspect that needs to be taken care of to reduce the
number of calculations has to do with the stochastic nature of
individual simulations. In order to represent the varying experi-
cences of individuals and apply the various risks, decisions,
resource use, and other elements that are variable, the models use
random numbers (for example, if 10% of patients with myocardial
infarction are discharged home from the emergency room,
then a number drawn randomly between zero and one will
indicate discharge if it is below 0.1). As with any stochastic
process, accurate representation requires enough draws to pro-
perly reflect the variability, and this implies that the more variabil-
ity, the higher the number of draws. In a simulation, this
translates to requiring added numbers of individuals to be
modeled and thus many more calculations. Lessening this burden
is a matter of reducing the nuisance variance, and various tech-
niques, including cloning the individuals, exist to accomplish this
[11,12]. When these aspects are addressed, simulations run very
efficiently, and it is not a problem to carry out probabilistic or
other sensitivity analyses [11].

The quality of models also hinges on the supporting
data—the larger the gaps, the less certain in the estimates pro-
duced by the model. To be sure, a reasonably informative discrete
event simulation requires more detailed data than a typical
Markov cohort model. Nevertheless, as the paper reporting on
discrete event simulation of glaucoma treatments [13] points out,
the data required to inform an individual Markov model that
would approach the level of accuracy of a discrete event simul-
ations are just as burdensome. In fact, fitting the state transition
framework can require more calculations and processing than a
discrete event simulation where events can be predicted by mul-
tivariate time-to-event functions. Furthermore, the position that
lack of readily available data justifies a simple cohort model is
untenable when the purpose is informing real health-care deci-
sions. An inaccurate model is not preferable to nothing and
should become an increasingly unacceptable approach as
decision-makers demand appropriate estimates. We have found
that in most cases, the data are available, though greater efforts
may be required to acquire and analyze them so they can inform
a discrete event simulation.

If data are truly limited, a discrete event simulation provides
a substantial advantage because the inadequacy of the data is not
built into the structure of the model. The simulation can be
designed to properly reflect the problem and carry out explor-
atory analyses with the limited data and best guesses; it can then
incorporate additional data should these become available. A
cohort Markov model that only applies an average treatment
effect to the risk given by a mean patient profile cannot, for
example, be easily modified to account for compliance that is
affected by patient characteristics and treatment response. Such
an addition would force a complete reprogramming of the
model. In a discrete event simulation, compliance can be built in
from the outset (or, with a small effort, added at a later stage),
and when data become available, it is relatively straightforward
to incorporate treatment- and patient-specific effects.

Like most computer-intensive activities, building a discrete
event simulation is best done using appropriate software.
Although it is possible to construct one using spreadsheets, this is
not a good choice because their “calculate everything every time”
nature is precisely the opposite of the required “calculate sequen-
tially and only when relevant.” Moreover, structuring the simu-
lation and displaying it transparently is difficult to do with
spreadsheets. Software designed for decision trees is not much
better. Although linear programming languages, such as C++ can
be used, they too present drawbacks in terms of transparency and
the degree of programming skills required. Fortunately, there is
abundant software specifically designed for discrete event simu-
lations that greatly facilitates the development of these
models, their efficient calculation, and transparent presentation.
Of these, ARENA (Rockwell Automation, Warrendale, PA) is the
most widely used. The main limitation of these packages is that
they are not designed for health-care decision problems and,
thus, require the modeler to adapt tools and concepts from other
fields. The specialized software is also much more expensive than
typical spreadsheet software.

This brings us to perhaps the greatest challenge faced by
discrete event simulation in our field today: the familiarity of
analysts throughout academia, regulatory agencies, and the
private sector with cohort Markov models. It is understandable
that when faced with the need to develop, assess, or use complex
models, there is a reluctance to step out of the comfort zone and
enter unfamiliar territory. Hopefully, as the severe limitations of
the cohort Markov technique become more evident, those who
must have the expertise to develop reliable and valid models or to
evaluate them will adopt the much better discrete event simula-
tion approach and help educate nontechnical audiences on the
need for this more advanced modeling technique.

Transparency

The authors of the article on discrete event simulation of glau-
coma treatments [13] ably demonstrate the advantages of this
approach but lament that it is difficult to convey transparently.
Nevertheless, they have included an appendix where they have
been able to provide a concise and understandable description of
the model and data sources and a full technical description of the
data and methods. Although the technical appendix may seem
lengthy, a full description of any reasonably accurate model
requires this level of detail. Journals willing to include the tech-
nical details behind models enable transparency, and with the
option to post these online, space restrictions should no longer be
a limiting. Indeed, such appendices should be the norm, rather
than the exception.

That said, it may not be the lengthy reporting that some may
view as inhibiting transparency, but rather the sophisticated cal-
culations underlying the model and the number of relationships
considered. This can, indeed, be a challenge to convey, and even
more to review competently; but if it is the level of detail required
to adequately represent the problem, then this needs trumps any
transparency issues [5]. Moreover, as the glaucoma researchers
point out [13], the transition matrices required to approach the
level of detail necessary for their analysis would have been pro-
hibitive. A model based on thousands of transition matrices,
countless tunnel states, and hundreds of disease states would be
enormously difficult to navigate, and certainly not any more
transparent, particularly if programmed using spreadsheets, than
a discrete event simulation.

The complexity of the equations underlying the model calcula-
tions is another matter. It is undoubtedly true that an acceler-
ated Weibull function is more difficult to understand than a
simple constant probability. It is easier to describe the “average
patient” and an “average treatment effect” than it is to explain
the correlations required to properly simulate multiple individual
characteristics and how these relate to risk and treatment effect.
Again, accuracy must outweigh transparency, especially for those
without the background to evaluate the mathematics underlying
the model. Decision-makers must ensure availability of the ex-
pertise capable of evaluating accurate simulations, rather than
insisting on simplicity for its own sake, even it leads to unreliable models.

**Conclusion**

It is remarkable that an article arguing that discrete event simulation should be preferred more than Markov models is still required, given that in all other fields of human endeavor the techniques described here are commonplace and taken to be the standard [14]. Although several publications have proposed guidelines for choosing a modeling technique, we do not believe there is any substantive basis for selecting a cohort Markov approach to inform health care decisions. Our field would do well to relinquish its commitment to this rudimentary technique and focus instead on gaining experience with discrete event simulation, or indeed other more appropriate techniques such as agent-based simulation or dynamic transmission models, in order to supply decision-makers with models that are sufficiently accurate to provide credible information.

Source of financial support: None.

**References**