

Recent Developments in Decision-Analytic Modelling for Economic Evaluation

Milton C. Weinstein

Department of Health Policy and Management, Harvard School of Public Health, Boston, Massachusetts, USA

Abstract

The past few years have seen rapid changes in the methods of decision-analytic modelling of healthcare programmes for the purposes of economic evaluation. This paper focuses on four developments in modelling that have emerged over the past few years or have become more widely used.

First, no one optimal method for extrapolating outcomes from clinical trials has yet been established. Modellers may draw from a set of varied assumptions about survival extrapolation that encompass a range of possibilities from highly optimistic to extremely cautious.

Secondly, the practicality and appeal of microsimulation as a method for analysing healthcare decision problems has increased dramatically with the speed of computing technology. Individual instantiations of a system are generated by using a random process to draw from probability distributions a large number of times (also known as Monte Carlo or probabilistic simulation). Microsimulation is moving in new directions, such as discrete-event simulations that simulate sequences of events by drawing directly from probability distributions of event times; this approach is now being broadly applied to model situations where populations of patients interact with healthcare delivery systems. Microsimulation modelling of transmission systems at the population level is also rapidly developing.

Thirdly, model calibration is emerging as a new tool that may offer health scientists a means of generating important fundamental knowledge about disease processes. Model calibration allows evidence synthesis in which observations on observable quantities are used to draw inferences about unobservable quantities. The methodology of model calibration has advanced considerably, drawing on theories of numerical analysis and mathematical programming such as gradient methods, intelligent grid search algorithms, and many more.

As a fourth issue, an area of extraordinary activity is in the use of transmission models to analyse interventions for infectious diseases, including population-wide effects of vaccination. Transmission models use differential equations to simulate, deterministically for the most part, transitions among infection-related health states. Only recently have modelling methodologies been combined so that cost-effectiveness analyses can consider explicitly not only the patient-level

benefits of interventions but also the secondary benefits through transmission dynamics.

Advances in technology allow more realistic and complex healthcare models to be simulated more rapidly. However, decision makers will not readily accept results from models unless they can understand them intuitively and explain them to others in relatively simple terms. The challenge for the next generation of modellers is not only to harness the power available from these newly accessible methods, but also to extract from the new generation of models the insights that will have the power to influence decision makers.

1. Background

Cost-effectiveness analysis was originally conceived as an extension of clinical decision analysis – a method intended to help healthcare decision makers arrive at the best choices under conditions of uncertainty, conflicting objectives and resource constraints.^[1] When viewed as decision analyses, healthcare cost-effectiveness analyses are intended to help decision makers use the available evidence to reach a decision that they cannot avoid; these decision-focused analyses are not intended to reveal scientific truth. Unlike a scientific experiment, such as a clinical trial finding that not enough evidence exists on which to base a conclusion as to whether or not a hypothesis is true, a decision-analytic cost-effectiveness analysis cannot fail to reach a conclusion. In other words, a decision maker who uses such an analysis cannot fail to act, even if the decision is to postpone action until more evidence is obtained.

Because the evidence on which healthcare decisions must be based is always incomplete, cost-effectiveness analyses that are viewed as decision analyses under resource constraints must inevitably rely on what is commonly called ‘modelling’. A model is a representation of reality itself, based on interrelationships derived from theory or observation. In cost-effectiveness modelling, the model is a mathematical one, with probabilities interacting to produce simulated patients, or fractions of a population, in different health states over time. Attached to these simulated outcomes are utilities and costs that combine to form the denominators and numerators of cost-effectiveness ratios. By their nature as sim-

plified representations, models cannot be perfect reflections of past reality, let alone predictors of future reality. It is wrong to insist that models be ‘validated’ by events that have not yet occurred; after all, the modeller cannot anticipate advances in technology, or changes in human behaviour or biology. All that can be expected is that the model reflects the current state of knowledge in a reasonable way, and that it is free of logical errors (‘bugs’). Previous papers have discussed the types of validation that may be reasonably applied to models,^[2] but the ability to predict the future is not one of them.

Even cost-effectiveness analyses that are conducted concurrently with, or ‘piggybacked upon’, clinical trials rely on modelling – to extrapolate the time horizon, to generalise the patient population, or to evaluate composite strategies made up of treatment and testing components. Although trial-based cost-effectiveness studies may have the objective of testing a hypothesis about the true value of a cost-effectiveness ratio, or a net economic benefit, in the end they are used by decision makers to select healthcare strategies and to allocate resources.

The art and science of decision modelling in healthcare have become much more complex, and even arcane, since the early decision-tree and state-transition models that we pioneers were able to run on pocket calculators or program ourselves in Fortran or Basic. With complexity comes loss of transparency, which is unfortunate. One of the values of a good model is that it can yield qualitative insights that cause a decision maker to say ‘Aha!’ when a counterintuitive strategy proves to be optimal *and for an understandable reason*. A complex model that yields a result that is surprising but not under-

standable in simple terms is unlikely to change the behaviour of decision makers until the result is verified in a clinical trial or until it can be explained in logically understandable terms.

Guidelines for the conduct of healthcare decision models have been developed by expert panels, including a task force convened by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), which I chaired and on which Bernie O'Brien served and participated actively.^[3] These guidelines, and others like them, have been widely cited. However, new developments in modelling methodology, or existing methods that have come to be more widely used as computing power increases, make the art and science of healthcare decision modelling an evolving process.

In the present paper, I comment on four developments in modelling that have emerged over the past few years or have become more widely used. They are as follows: (i) methods for extrapolating outcomes from clinical trials, particularly survival or disease-free survival; (ii) use of random microsimulation as opposed to deterministic cohorts to project population outcomes; (iii) the method of model calibration, both as a method of estimating model parameters consistent with observed population data and as a method of retrospective validation; and (iv) incorporation of transmission dynamics in modelling treatment and prevention programmes for infectious diseases. I will not comment on a fifth important area of methods development, namely probabilistic forms of sensitivity analysis that report probability distributions of outcomes in addition to, or instead of, ranges and decision thresholds, because this topic is the focal point of another paper in this issue.^[4] I do not represent this article as a comprehensive or systematic review of these topics. On the contrary, what follows contains personal observations on their strengths and limitations, and on their appropriate role in cost-effectiveness modelling. I also apologise for over-representing my own work and the work of my colleagues and students among the examples of the methods discussed, but these are the models I know best.

2. Extrapolating from Clinical Trials

One of the first cost-effectiveness analyses I was involved in concerned coronary artery bypass surgery, back in the 1970s, before angioplasty or stents.^[5] At the time, the only evidence from randomised clinical trials came from a handful of studies that followed patients with multiple-vessel coronary artery disease for periods of up to 4 years. Because of the relatively high surgical mortality rates at the time, short-term data tended to show a survival disadvantage for surgery up to about 2 years, after which time the survival curves for surgery and medical management crossed and the surgery curve became superior. Because follow-up was short – at most 4 years in the longest study at the time – truncating the calculation of life expectancy to the trial observation period would have yielded almost no difference between the treatment arms. However, the survival curve at the end of follow-up was higher for surgery than for medical management, and the survival curves seemed to be diverging, reflecting an apparently lower annual mortality rate with surgery after the initial perioperative period. Of course, this situation is now very commonly encountered in trial-based economic evaluation, in which survival curves are either separated at the end of follow-up, or they seem to be heading in different directions with different slopes, or both. With a lifetime horizon, what should the analyst assume about the subsequent trajectories of the survival curves?

The same question arises, of course, when the outcome is disease-free survival, event-free survival, or any other outcome represented by a failure-time analysis. The outcome for cost-effectiveness analysis could be life expectancy, disease-free life expectancy, or quality-adjusted life expectancy. Analogous questions arise for extrapolation of costs that are observed in clinical trials.

As we all know, the field has embraced a set of standard assumptions about survival extrapolation that encompass a range of possibilities from highly optimistic to extremely cautious.^[6] They include the following:

1. Assume that the treatment-specific mortality rates or event rates observed in the latest period or time interval will continue indefinitely into the future. Geometrically, this assumption means that the survival curves continue to diverge in favour of the superior treatment, and that life-year gains will accrue at an ever-increasing rate, dampened only by the age-related background mortality that eventually kills off both groups.

2. Assume that the accrued survival advantage will be retained, but that subsequent mortality rates are equal in the treatment groups. The common mortality rate may be determined, for example, by pooling the experiences of the two groups in the latter years of observation. Geometrically, this assumption means that the survival curves will remain separated as they were at the end of follow-up, but that they will decline at equal rates. Even under this assumption, the treatment that was superior at the end of follow-up continues to accumulate life expectancy gains after the trial ends. Therefore, it is not the most cautious form of extrapolation.

3. Assume that the accrued survival advantage is immediately lost at the moment the trial ends, so that the survival in the groups becomes identical. Geometrically, this assumption means that the higher survival curve drops precipitously to the level of the lower curve, and that they remain superimposed thereafter. This assumption, referred to affectionately by some of us as the 'stop-and-drop' assumption in survival modelling, is equivalent to truncating the analysis at the end of the trial follow-up period.

There are many intermediate assumptions possible. For example, one might assume that the annual mortality or event rates remain different, as in assumption 1, for some period of time after the trial, possibly an interval equal in duration to the trial itself, and then become equal as in assumption 2. Or, the survival curves might immediately stop diverging as in assumption 2, and then gradually converge so that they become identical after some period of time, as in assumption 3. In this latter scenario, the mortality or event rates in the superior treatment would actually be higher for a period of time, until

the cumulative amount of mortality or events 'catches up' with the inferior treatment.

Further variations on these themes would make different assumptions depending on whether or not treatment was continued and, if so, for how long. However, these variations would, of course, apply to drugs or long-term therapies, and not to discrete interventions such as surgeries or device implantations.

A notable example of how models have differed in their approaches to these types of assumptions based on short-term trials can be found in the literature on economic evaluation of drugs for multiple sclerosis.^[7-9] In the context of multiple sclerosis, the issue is not survival or incidence of events, but the reductions in rates of disease progression and relapse, which affect the trajectories of health-related utility with and without treatment. Some of the most prominent models in the literature have obtained stunningly disparate results and cost-effectiveness ratios, largely driven by their disparate assumptions about the durability of treatment effects. The contingent reimbursement scheme for multiple sclerosis drugs in the UK is a unique policy response to this situation, whereby payment will be retroactively adjusted in response to observed outcomes that could only be modelled from short-term trial data at the time the drugs were introduced.^[10]

3. Microsimulation

The practicality of Monte Carlo, or probabilistic, simulation of individual patients as a method for analysing healthcare decision problems has increased dramatically with the speed of computing technology. I use the term 'microsimulation' to describe any analysis in which individual instantiations of a system – such as a patient's lifetime or the course of an epidemic – are generated by using a random process to 'draw' from probability distributions a large number of times, in order to examine the central tendency and, possibly, the distribution of outcomes. This procedure is also known as 'first-order Monte Carlo simulation'. For policy purposes, the means of the empirical distributions – in other words, the expected health outcomes and costs – are

the objects of interest, although decision makers responsible for small populations may wish to examine the small-sample variability of outcomes.

An important distinction is between the variance across individual replicates of a microsimulation and the variance across the mean values that are generated by varying the parameters of the model according to specified parameter or empirical distributions.^[11] The latter, which is sometimes called 'second-order' uncertainty because it relates to uncertainty in the probability parameters, and not in individual outcomes, is what drives probabilistic sensitivity analysis. The former, the variation between actual instantiations at the individual level, is really only random noise from a decision maker's point of view, and can be overcome by increasing the sample size of the microsimulation. Unlike clinical trials, of course, there is no limit on the sample size of a microsimulation other than computing time, so that the notion of p-values or significance tests to compare decision strategies in microsimulation runs is prone to misinterpretation. Such significance tests, when used to compare the mean outcomes of two strategies simulated in multiple replications of a microsimulation model, address only the question of whether the means of the simulated outcomes are truly different, given the model's parameters, but they are irrelevant for evaluating uncertainty about the parameter values themselves.

Random number generators play a key role in this technology. Because 'random' numbers are not truly random, care must be taken, both in debugging and in doing sensitivity analyses, that either the sequence of random numbers is purposefully kept constant, or that the sequence is reshuffled in subsequent iterations.

Historically, microsimulation began to supplant cohort analyses of state-transition models because of the desire to avoid an unmanageable number of health states. In order to reflect the dependence of transition probabilities on multiple risk factors and patient histories that evolve over time within a model, health states need to be created so that the model 'remembers' those patient histories. The alternative is to simulate one patient at a time, record-

ing the values of so-called 'tracker' variables that describe the current patient's history but without the need to keep track simultaneously of the numbers of patients who have each possible history. Computing time increases, but model complexity and computer memory requirements are contained. This trade-off, plus the impracticality of performing probabilistic sensitivity analyses in the context of patient-level microsimulations, remains to this day a concern for modellers who must choose between deterministic cohort analyses and probabilistic microsimulations.^[12]

Although most early applications of microsimulation in healthcare decision models were applied to Markov models or other state-transition processes that unfold in lock-step with a time clock and that apply to groups of individual patients, recent developments in microsimulation have moved in different directions. Discrete-event simulations, instead of modelling events one cycle at a time, simulate sequences of events by drawing directly from probability distributions of event times. As an example, consider the difference between a cycle-driven microsimulation of a disease and a discrete-event simulation. In the cycle-based state-transition model, the probability of disease progression, response to treatment, or mortality would be applied in each model cycle, such as a month, and each simulated patient would progress or not, respond to treatment or not, die or not, and, if the patient survives, age by one cycle. At the end of each cycle, each patient would transition to a new health state according to which events had occurred. In a discrete-event simulation, probability distributions would govern the time of disease progression, the time of mortality, and the like, so that once these draws were made from their parent distributions, the patient's trajectory would unfold as a deterministic case history. At some level, the two approaches are mathematically equivalent when applied to patient-level models; however, from a computational point of view and in terms of graphic displays of individual case histories, there may be advantages in one approach or the other. Discrete-event simulation has captured a large share of the modelling market in situations where

populations of patients interact with healthcare delivery systems. For example, discrete event simulations have been used to guide the scheduling of elective surgery in hospitals that also admit many surgical patients from emergency departments, the design and management of operating rooms, and processes for allocating resources such as transplantable organs. In fact, queueing systems in which both supply and demand unfold probabilistically over time, as in the organ-allocation setting, are fertile areas for application of discrete-event microsimulation models.

Another rapidly developing area of microsimulation modelling is in the area of the population dynamics of infectious diseases. While the transmission dynamics of infections can be modelled deterministically by using differential equation models – discussed further in section 5 – there may be substantial heterogeneity in the population that affects not only the transmission rates per contact, but also the propensities of different people in different geographical and social segments of the population to interact. This type of heterogeneity, combined with the inherent uncertainty regarding individual events such as contacts or infections, contributes to a substantial amount of uncertainty as to how rapidly an epidemic may spread, or even whether it will be sustained. Microsimulation modelling of transmission systems at the population level is now becoming more practical with the availability of powerful computing technology. Each random simulation represents not an individual person, but an entire population – an ecosystem, so to speak. In this situation, not only the average outcomes, but also the first-order probabilities of different population outcomes – i.e. even assuming certainty about the underlying parameters – are of interest and accessible with microsimulation techniques. For example, what is the probability that an outbreak will become a pandemic? In environmental ecology, by analogy, this type of model can be used to estimate the probability that a species will survive or become extinct under different assumptions.

4. Model Calibration

Modellers rely on diverse sources of data for estimating the parameters of their models. Relative effects of different treatments are inferred from clinical trials. Baseline health event rates come from natural history studies. Relative risks across patient subgroups defined by age and other risk factors come from cohort or case-control studies. Often the process of synthesising and reconciling diverse sources of evidence is challenging, as in the situation where one has evidence of treatment A versus placebo, and treatment B versus placebo, but in different populations. Modellers have resorted to creative assumptions that are based very indirectly on observed data, in order to assign values to parameters. This phenomenon is illustrated by an historical synopsis of how certain assumptions evolved in the specification of the Cost-Effectiveness of Preventing AIDS Complication (CEPAC) model, a microsimulation model of HIV-AIDS that has evolved over more than a decade.^[13-16] The original model incorporated the assumption that the risk of an opportunistic infection for a patient infected with HIV and with a CD4+ cell count of 200/mm³ doesn't depend on the patient's viral load or on whether the patient is receiving antiretroviral therapy. As evidence emerged with the use of sophisticated statistical methods, it became clear that an alternative assumption was more strongly supported by the evidence, i.e. that the risk of an opportunistic infection for a patient infected with HIV and with a CD4+ cell count of 200/mm³ is lower if a patient is receiving antiretrovirals than if he or she is not. In either case, there is the additional question of whether it matters if the patient is male or female, and our research group is currently investigating that question using a newly available dataset.

Unfortunately, there may be model parameters for which direct observation of evidence is impossible. Consider the case of cervical cancer. It is known that cervical cancer emerges as the result of an initial infection with the human papilloma virus (HPV). HPV infection can result in the formation of lesions that may or may not progress to cancer. In countries where screening is widespread, women with lesions

receive interventions that preclude direct observation of how they would have progressed.

Even for models of conditions with substantial evidence directly applicable to individual model parameters, there is always ‘wobble room’ in the estimates of each parameter, more for some than for others. And there is the question of whether the assumptions that modellers use to glue the individual bits of evidence together are valid. For all of these reasons, modellers are devoting an increasing amount of attention to evaluating whether models reflect evidence that relates not to individual parameters, such as CD4+ cell-specific rates of opportunistic infections or mortality, but to aggregate and observable outcomes, such as overall survival with HIV. Age-specific mortality is an output, not an input, in most complex disease models, but there is often direct, albeit imperfect, evidence bearing on such outputs. Should a modeller ignore evidence simply because it relates to an output rather than to an input? If not, then the modeller is faced with the task of fitting the model’s parameters so that the modelled outputs match the observed evidence on the outputs. This is model calibration.

Model calibration can be thought of as the solution of a system of simultaneous equations, in which the model parameters, or some of them, are the unknowns. However, since the model’s equations aren’t linear, and they don’t lend themselves to closed-form solution, the only approach is to search, either systematically or by trial and error, for the combination of parameter values that best fits the evidence. But this is more complicated than it may seem. There is evidence about some, but not all, of the outputs, and on some, but not all, of the parameters. The evidence is itself subject to uncertainty, owing to issues of internal validity including sampling error, and to issues of external validity including the question of whether data from one population or population sample are applicable to the conditions being modelled. Despite the remarkable advances in Bayesian evidence synthesis,^[17] the task of synthesising evidence on a complex space of input parameters, intermediate outcomes, and end outputs, can be daunting.

Almost 2 decades ago, my colleagues and I performed a primitive calibration of the Coronary Heart Disease Policy Analysis at two points in its development, once when it was first launched^[1,18] and again a decade later when more evidence became available on trends in coronary heart disease mortality – an output of the model.^[19] The methods we used might now seem embarrassingly primitive; however, I am not embarrassed because few healthcare models up to that time attempted to claim that their models fitted the evidence on outputs. We first identified a set of parameters that we regarded as the most uncertain, not in any formal sense, but either because there were several studies that diverged widely, or because there was no good study at all. We simply varied the most uncertain parameters over their plausible ranges, and simulated all the combinations within this hypercube within the parameter space. Then we needed a metric for judging the calibration. We used the very simple criterion that all of the outputs for which there was evidence had to be within 10% of the target values. We chose the combination of parameters that satisfied the 10% criterion for all targeted outputs, and also fell within 5% for most of the targeted outputs.

The methodology of model calibration has advanced considerably since those pioneer days. Methods of searching for best-fitting parameter sets have drawn on theories of numerical analysis and mathematical programming and include gradient methods, intelligent grid search algorithms, and many more. Criteria for goodness of fit can be based on maximum likelihood, or even Bayesian posterior probability. For example, consider how calibration was done in one model of treatment for hepatitis C viral infection (HCV).^[20] A systematic grid search was performed over a parameter hypercube bounded by upper and lower limits on individual parameters, as inferred from a systematic review of literature. The criterion for goodness of fit was maximum likelihood. Actually, the criterion was a weighted average of the likelihoods for the two outcome targets – HCV incidence and liver cancer mortality in the US. The 50 best-fitting parameter sets were selected, and the results were averaged over these, in

effect assigning a 2% probability to each one. This approach to model calibration yielded a probabilistic sensitivity analysis as a natural byproduct. An alternative approach to probabilistic sensitivity analysis might have been to assign weights to each parameter set in proportion to its goodness of fit, or some function of its goodness of fit.

Calibration is a form of evidence synthesis in which observations on observable quantities are used to draw inferences about unobservable quantities, such as latent rates of disease progression, remission, or mutation. It has become clear, as this approach has been introduced by decision modelers, that it may be capable of generating important fundamental knowledge about disease processes. In effect, model calibration, when viewed as evidence synthesis, offers health scientists a new tool for learning about how diseases progress.

An exciting, albeit controversial, example of this phenomenon has emerged from the breast cancer modelling group at the University of Wisconsin, a part of the National Cancer Institute's CISNET programme.^[21] Through a painstaking and lengthy process of calibration, the modelling team was at first unable to simultaneously achieve calibration targets for breast cancer incidence, prevalence, and mortality by stage at detection. No matter how hard they tried, and no matter how widely they varied the parameters, the model wouldn't match the data. Then they tried something seemingly bizarre: they incorporated the possibility that some tumours might actually disappear! This opened up a new range of possibilities, and excellent calibration was achieved. Does this prove that breast tumours regress? No. But it is evidence in favour of that hypothesis, and scientists would do well to pay attention to the possibility, no matter how preposterous it may seem. After all, the best available empirical evidence, when assembled in the form of a model, does not fit with reality – at least not unless the possibility of tumour regression is entertained.

5. Transmission Dynamics

An area of extraordinary activity in methods development for economic evaluation has been in the

analysis of interventions for infectious diseases. Cost-benefit and cost-effectiveness analyses of vaccination and treatment programmes have until recently considered only the benefits and costs for individuals, without explicitly evaluating the external benefits and cost offsets resulting from reduced transmission by immunised members of a population – the phenomenon known as 'herd immunity'. Recent interest in modelling the costs and health benefits of detection and treatment programmes for HIV-AIDS and tuberculosis (TB) has also had to deal with other kinds of external effects. In HIV modelling, one might wish to consider the effects of successful antiretroviral treatment on the probability that a patient will transmit the virus. In TB modelling, one might want to capture the effects of improved treatment on the spread of drug-resistant infections.

Concurrently with the emergence of cost-effectiveness analysis from the disciplines of decision science and health economics, the field of transmission modelling emerged as a sub-discipline of epidemiology, largely influenced by the work of Anderson and May.^[22] Transmission models use differential equations to simulate, deterministically for the most part, transitions among health states such as susceptible, latently infected, actively infected, and recovered. Population-wide effects of interventions, such as vaccination, can be evaluated by transmission models. However, the focus among transmission modellers has been mainly on the long-term or steady-state behaviours of epidemics rather than on the accumulation of health benefits and costs among members of a population. Only recently have these two modelling methodologies been joined so that cost-effectiveness analyses can consider explicitly not only the patient-level benefits of interventions but also the secondary benefits through transmission dynamics.

The earliest economic evaluations for infectious disease programmes date back to the late 1960s, when the US Public Health Service published a cost-benefit analysis of measles vaccination.^[23] This analysis compared scenarios with and without measles vaccination in the US population. By taking a

population approach based on empirical data concerning disease incidence over time, rather than by modelling patient-level outcomes, the analysis implicitly captured the external benefits owing to herd immunity by projecting incidence at the population level with and without vaccination. A decade later, a cost-benefit analysis of pertussis vaccination used a decision tree model structure, but was nonetheless able to capture the effects of herd immunity by comparing population-wide vaccination strategies in which the incidence of pertussis infection was assumed to be reduced more than proportionately to the level of vaccine coverage.^[24] Subsequent decision-analytic models of vaccination programmes that have used decision tree or Markov structures, however, have tended to focus on individual patient outcomes and, in doing so, have not considered the external effects of herd immunity.

Recently, with increased interest in the population-wide effects, not only of vaccination programmes but also and especially of HIV-AIDS and TB treatment programmes, the models used in economic evaluations have begun to incorporate explicitly the transmission dynamics of infectious diseases. There have been two distinct methodological approaches to this challenge.

The first approach is to retain a Markov model structure, but to simulate the numbers of secondary cases of transmitted disease in relation to the time patients are infectious. These secondary cases, in turn, can be reported simply as numbers of cases and not included formally in the cost-effectiveness ratio or net-benefit calculation, or they can be associated with QALYs lost and costs induced per case, discounted appropriately to present value, and combined with the patient-level outcomes and costs.

This was the approach used in two recent cost-effectiveness analyses of HIV testing programmes.^[15,25] In one of those analyses, which is based on a probabilistic simulation of a state-transition model, the number of secondary infections each patient transmits over the course of a lifetime is assumed to depend on the stage at which the infection is detected.^[15] Specifically, the analysis used the epidemiological concept of R_0 , the average

number of infections transmitted by an infectious individual over a lifetime in a susceptible population. The key assumption in the analysis is that R_0 is lower for patients whose infection is detected early by screening and who enter treatment at an earlier stage, compared with patients whose infectious status is not discovered until either an opportunistic infection has occurred, or until the infection is identified as a result of testing at an incidental clinical encounter. This method of attaching a 'multiplier' to reflect secondary cases is plainly an approximation of the transmission dynamics. For example, assuming that the number of transmissions of infections is the same whether the population is fully susceptible or mostly infected or immune does not take into account the fact that the rate of transmission by an infectious patient depends on the evolving infectious or immune status of the rest of the population. As another example of the limitations of this approach, it does not consider tertiary or higher-level cases, i.e. infections transmitted by the secondary cases. Fortunately, these two errors tend to work in opposite directions, although there is no reason to expect that they are similar in magnitude.

An alternative approach to incorporating external or secondary consequences of interventions for infectious diseases is to embed the economic evaluation directly into a population transmission model of susceptible, infected, and recovered persons. This approach was pioneered by Edmunds and colleagues^[26] in a study of chickenpox (varicella zoster virus) vaccine. As in Markov models, the numbers of persons in each health state contribute costs and QALYs in each time period, and these are finally summed up with appropriate time discounting to yield cost-effectiveness ratios or net-benefit calculations. This approach provides a more realistic representation of the transmission dynamics, but at the possible loss of clinical reality at the individual level because the number of health states possible in a transmission model is limited by computing capabilities.

This fully-dynamic approach was recently used to evaluate alternative strategies to identify and treat cases of multiple-drug-resistant (MDR) TB^[27] in

countries with a high prevalence of MDR-TB. By including health states for drug-susceptible and MDR-TB patients, and by modelling the transmission of MDR-TB across health states, both the direct and indirect benefits of aggressive MDR-TB treatment strategies could be reflected in calculations of the incremental cost per death averted and the incremental cost per QALY gained.

6. Concluding Remarks

The past few years have seen rapid changes in the methods of modelling healthcare programmes for the purposes of economic evaluation. Few of these changes represent true innovations, since they can be found in textbooks and scientific papers up to half a century ago. However, because they do not lend themselves to either closed-form mathematical solutions or primitive computing technology, they have been slow to penetrate the market for applications of decision modelling. The microprocessor revolution has itself revolutionised the way health economists and decision scientists pursue their work. For the most part, this has been for the good, as more realistic models can be simulated more rapidly. However, there are disadvantages associated with complexity in models. Transparency may be lost. Insight into the reasons why a result holds may also be lost in the details of equations and complex multidimensional graphics.

Decision makers will not readily accept results from models unless they can understand them intuitively and explain them to others in relatively simple terms. The challenge for the next generation of modellers is not only to harness the power available from these newly accessible methods, but also to extract from the new generation of models the insights that will have the power to influence decision makers.

Acknowledgements

No sources of funding were used to assist in the preparation of this article. The author has no conflicts of interest that are directly relevant to the content of this article.

References

- Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med* 1977; 296: 716-21
- Weinstein MC, Toy EL, Sandberg EA, et al. Modeling for health care and other policy decisions: uses, roles, and validity. *Value Health* 2001; 4: 348-61
- Weinstein MC, O'Brien B, Hornberger J, et al. Principles of good research practice for decision analytic modeling in health-care evaluation: report of the ISPOR task force on good research practices: modeling studies. *Value Health* 2003; 6: 9-17
- Claxton KP, Sculpher MJ. Using value of information analysis to prioritise health research. *Pharmacoeconomics* 2006; 24 (11): 1055-68
- Weinstein MC, Stason WB. Cost-effectiveness of coronary artery bypass surgery. *Circulation* 1982; 66 (5 Pt 2): III56-66
- Kuntz K, Weinstein M. Modelling in economic evaluation. In: Drummond M, McGuire A, editors. *Economic evaluation in health care: merging theory with practice*. Oxford: Oxford University Press, 2001
- Chilcott J, McCabe C, Tappenden P, et al. Modelling the cost effectiveness of interferon beta and glatiramer acetate in the management of multiple sclerosis. *BMJ* 2003; 326: 522-6
- Kobelt G, Jonsson L, Fredrikson S. Cost-utility of interferon beta-1b in the treatment of patients with active relapsing-remitting or secondary progressive multiple sclerosis. *Eur J Health Econ* 2003; 4: 50-9
- Prosser LA, Kuntz KM, Bar-Or A, et al. Cost-effectiveness of interferon beta-1a, interferon beta-1b, and glatiramer acetate in newly diagnosed non-primary progressive multiple sclerosis. *Value Health* 2004; 7: 554-68
- Buxton M. Economic evaluation and decision making in the UK. *Pharmacoeconomics* 2006; 24 (11): 1133-43
- Hunink M, Glasziou P, Siegel J, et al. *Decision making in health and medicine: integrating evidence and values*. Cambridge: Cambridge University Press, 2001: Chap. 11
- Griffin S, Claxton K, Hawkins N, et al. Probabilistic analysis and computationally expensive models: necessary and required? *Value Health* 2006; 9: 244-52
- Weinstein MC, Goldie SJ, Losina E, et al. Genotypic resistance testing to guide the choice of therapy in HIV: clinical impact and cost-effectiveness. *Ann Intern Med* 2001; 134: 440-50
- Freedberg KA, Losina E, Weinstein MC, et al. The cost-effectiveness of combination antiretroviral therapy in HIV. *N Engl J Med* 2001; 344: 824-31
- Paltiel AD, Weinstein MC, Kimmel AD, et al. Expanded screening for HIV in the United States: an analysis of cost-effectiveness. *New Engl J Med* 2005; 352: 586-95
- Sax PE, Islam R, Walensky RP, et al. Should resistance testing be performed in treatment-naive HIV-infected patients? A cost-effectiveness analysis. *Clin Infect Dis* 2005; 41: 1316-23
- Ades AE, Cliffe S. Markov chain Monte Carlo estimation of a multiparameter decision model: consistency of evidence and the accurate assessment of uncertainty. *Med Decis Making* 2002; 22: 359-71
- Weinstein MC, Coxson PG, Williams LW, et al. Forecasting coronary heart disease incidence, mortality, and cost: the Coronary Heart Disease Policy Model. *Am J Public Health* 1987; 77: 1417-26
- Hunink MGM, Goldman L, Tosteson ANA, et al. The recent decline in mortality from coronary heart disease, 1980-1990:

- the effect of secular trends in risk factors and treatment. *JAMA* 1997; 277: 535-42
20. Salomon JA, Weinstein MC, Hammitt JK, et al. Empirically calibrated model of hepatitis C virus infection in the United States. *Am J Epidemiol* 2002; 156: 761-73
 21. Fryback DG, Stout NK, Rosenberg MA, et al. Chapter 7: the Wisconsin Breast Cancer Epidemiology Simulation Model. *J Natl Cancer Inst Monogr* 2006; (36): 37-47
 22. Anderson RM, May RM. *Infectious diseases of humans: dynamics and control*. Oxford: Oxford University Press, 1992
 23. Axnick NW, Shavell SM, Witte JJ. Benefits due to immunization against measles. *Public Health Reports* 1969; 84: 673-80
 24. Koplan JP, Schoenbaum SC, Weinstein MC, et al. Pertussis vaccine: an analysis of benefits, risks, and costs. *N Engl J Med* 1979; 301: 906-11
 25. 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
 26. Edmunds WJ, Medley GF, Nokes DJ. Evaluating the cost-effectiveness of vaccination programmes: a dynamic perspective. *Stat Med* 1999; 18: 3263-82
 27. Resch SC, Salomon JA, Murray M, et al. Cost-effectiveness of treating multidrug-resistant tuberculosis. *PLoS Med* 2006 Jul 4; 3 (7): e241

Correspondence and offprints: Dr *Milton C. Weinstein*, Department of Health Policy and Management, Harvard School of Public Health, Boston, MA 02115, USA.