Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting

Stavros Petrou,[1] Alastair Gray[2]

[1]Clinical Trials Unit, Warwick Medical School, University of Warwick, Coventry CV4 7AL, UK
[2]Health Economics Research Centre, Department of Public Health, University of Oxford, Oxford, UK
Correspondence to: S Petrou
S.Petrou@warwick.ac.uk
Accepted: 8 February 2011

Cite this as: *BMJ* 2011;342:d1766
doi: 10.1136/bmj.d1766

Evidence relating to healthcare decisions often comes from more than one study. Decision analytical modelling can be used as a basis for economic evaluations in these situations

Economic evaluations are increasingly conducted alongside randomised controlled trials, providing researchers with individual patient data to estimate cost effectiveness.[1] However, randomised trials do not always provide a sufficient basis for economic evaluations used to inform regulatory and reimbursement decisions. For example, a single trial might not compare all the available options, provide evidence on all relevant inputs, or be conducted over a long enough time to capture differences in economic outcomes (or even measure those outcomes).[2] In addition, reliance on a single trial may mean ignoring evidence from other trials, meta-analyses, and observational studies. Under these circumstances, decision analytical modelling provides an alternative framework for economic evaluation.

Following on from our article on trial based economic evaluations,[1] we outline issues relating to the design, conduct, analysis, and reporting of economic evaluations using decision analytical modelling.

Defining the question

The first stage in the development of any model is to specify the question or decision problem. It is important to define all relevant options available for evaluation, the recipient population, and the geographical location and setting in which the options are being delivered.[3] The requirements of the decision makers should have a crucial role in identifying the appropriate perspective of the analysis, the time horizon, the relevant outcome measures, and, more broadly, the scope or boundaries of the model.[4] If these factors are unclear, or different decision makers have conflicting requirements, the perspective and scope should be broad enough to allow the results to be disaggregated in different ways.[5]

Decision trees

The simplest form of decision analytical modelling in economic evaluation is the decision tree. Alternative options are represented by a series of pathways or branches as in figure 1, which examines whether it is cost effective to screen for breast cancer every two years compared with not screening. The first point in the tree, the decision node (drawn as a square) represents this decision question. In this instance only two options are represented, but additional options could easily be added. The pathways that follow each option represent a series of logically ordered alternative events, denoted by branches emanating from chance nodes (circular symbols). The alternatives at each chance node must be mutually exclusive and their probabilities should sum exactly to one. The end points of each pathway are denoted by terminal nodes (triangular symbols) to which values or pay-offs, such as costs, life years, or quality adjusted life years (QALYs), are assigned. Once the probabilities and pay-offs have been entered, the decision tree is "averaged out" and "folded back" (or rolled back), allowing the expected values of each option to be calculated.[4]

Decision trees are valued for their simplicity and transparency. However, they are limited by the lack of any explicit time variable, making it difficult to deal with time dependent elements of an economic evaluation.[6] Recursion or looping within the decision tree is also not allowed, so that trees representing chronic diseases with recurring events can be complex with numerous lengthy pathways.

Markov models

An alternative form of modelling is the Markov model. These models permit a more straightforward and flexible sequencing of outcomes, including recurring outcomes, through time. Patients are assumed to reside in one of a finite number of health states at any point in time and make transitions between those health states over a series of discrete time intervals or cycles.[3 6] The probability of staying in a state or moving to another one in each cycle is determined by a set of defined transition probabilities. The definition and number of health states and the duration of the cycles will be governed by the decision problem: one study of treatment for gastro-

SUMMARY POINTS

- Decision analytical modelling for economic evaluation uses mathematical techniques to determine the expected costs and consequences of alternative options
- Methods of modelling include decision trees, Markov models, patient level simulation models, discrete event simulations, and system dynamic models
- The process of identifying and synthesising evidence for a model should be transparent and appropriate to decision makers' objectives
- The results of decision analytical models are subject to the influences of variability, uncertainty, and heterogeneity, and these must be handled appropriately
- Validation of model based economic evaluations strengthens the credibility of their results

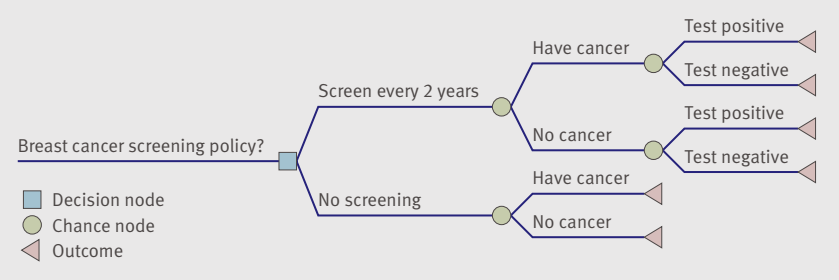


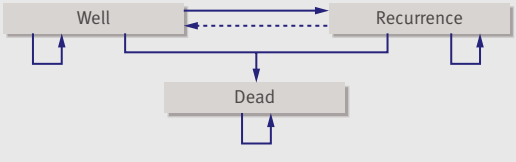
Fig 1 | Decision tree for breast cancer screening options[4]

oesophageal reflux disease used one month cycles to capture treatment switches and side effects,[7] whereas an analysis of cervical cancer screening used six monthly cycles to model lifetime outcomes.[8]

Figure 2 presents a state transition diagram and matrix of transition probabilities for a Markov model of a hypothetical breast cancer intervention. There are three health states: well, recurrence of breast cancer, and dead. In this example, the probability of moving from the well state at time *t* to the recurrence state at time *t+1* is 0.3, while the probability of moving from well to dead is 0.1. At each cycle the sum of the transition probabilities out of a health state (the row probabilities) must equal 1.

An important limitation of Markov models is the assumption that the transition probabilities depend only on the current health state, independent of historical experience (the Markovian assumption). This limitation can be overcome by introducing temporary states that patients can only enter for one cycle or by a series of temporary states that must be visited in a fixed sequence.[4]

The final stage is to assign values to each health state, typically costs and health utilities.[6] Most commonly, such models simulate the transition of a hypothetical cohort of individuals through the Markov model over time, allowing the analyst to estimate expected costs and outcomes. This simply involves, for each cycle, summing costs and outcomes across health states, weighted by the proportion of the cohort expected to be in each state, and then summing



State at time t	State at t + 1		
	Well	Recurrence	Dead
Well	1-(0.3 + 0.1)	0.3	0.1
Recurrence	0.1	1-(0.1 + 0.2)	0.2
Dead	0	0	1

Fig 2 | Markov state diagram and transition probability matrix for hypothetical breast cancer intervention. The arrows represent possible transitions between the three health states (well, recurrence, and dead), loops indicate the possibility of remaining in a health state in successive cycles, and the dashed line indicates the possibility of backwards transition from recurrence of breast cancer to the well state after successful treatment. The cycle length is set at one year

across cycles.[3] If the time horizon of the model is over one year, discounting is usually applied to generate the present values of expected costs and outcomes.[1]

Alternative modelling approaches

Although Markov models alone or in combination with decision trees are the most common models used in economic evaluations, other approaches are available.

Patient level simulation (or microsimulation) models the progression of individuals rather than hypothetical cohorts. The models track the progression of potentially heterogeneous individuals with the accumulating history of each individual determining transitions, costs, and health outcomes.[3 9] Unlike Markov models, they can simulate the time to next event rather than requiring equal length cycles and can also simulate multiple events occurring in parallel.[9]

Discrete event simulations describe the progress of individuals through healthcare processes or systems, affecting their characteristics and outcomes over unrestricted time periods.[9] Discrete event simulations are not restricted to the use of equal time periods or the Markovian assumption and, unlike patient level simulation models, also allow individuals to interact with each other[10]—for example, in a transplant programme where organs are scarce and transplant decisions and outcomes for any individual affect everyone else in the queue.

Dynamic models allow internal feedback loops and time delays that affect the behaviour of the entire health system or population being studied. They are particularly valuable in studies of infectious diseases, where analysts may need to account for the evolving effects of factors such as herd immunity.[11]

Identifying, synthesising, and transforming data inputs

The process of identifying and synthesising evidence to populate a decision analytical model should be consistent with the general principles of evidence based medicine.[3] These principles are broadly established for clinical evidence. Less clear is the strategy that should be adopted to identify and synthesise evidence on other variables, such as costs and health utilities, other than it should be transparent and appropriate given the objectives of the model.[12]

If evidence is not available from randomised trials, it has to be drawn from other sources, such as epidemiological or observational studies, medical records, or, more controversially, expert opinion. And sometimes the evidence from randomised trials may not be appropriate for use in the model—for example, cost data drawn from a trial might reflect protocol driven resource use rather than usual practice[13] or might not be generalisable to the jurisdiction of interest.[5] These methodological considerations have increased interest in multiparameter evidence synthesis (box).[14] These techniques acknowledge the importance of trying to incorporate correlations between variables in models.[2] However, accurately assessing the correlation between different clinical events, or between events and costs or health utilities, may be difficult without patient level data from a single source. Another complication is that evidence may have to be transformed in complex ways to meet the requirements of the model.[3 4]

Quantifying and reporting cost effectiveness

Once data on all variables required by the model have been assembled, the model is run for each intervention being evaluated in order to estimate its expected costs and expected outcomes (or effects). The results are typically compared in terms of incremental cost effectiveness ratios and depicted on the cost effectiveness plane (box).[1]

Handling variability, uncertainty, and heterogeneity

The results of a decision analytical model are subject to the influences of variability, uncertainty, and heterogeneity, and these must be handled appropriately if decision makers are to be confident about the estimates of cost effectiveness.[3 15]

Variability reflects the randomness arising from the modelling process itself—that is, the fact that models typically use random numbers when determining whether an event with a given probability of occurring happens or not in any given cycle or model run, so that an identical patient will experience different outcomes each time they proceed through the model. This variability needs to be eliminated by running the model repeatedly until a stable estimate of the central tendency has been obtained.[16]

Parameter uncertainty reflects the uncertainty and imprecision surrounding the value of model variables such as transition probabilities, costs, and health utilities. Probabilistic sensitivity analysis, in which all variables are varied simultaneously using probability distributions informed by estimates of the sample mean and sampling error from the best available evidence, is the preferred way of assessing parameter uncertainty.[15] Probabilistic sensitivity analysis also allows the analyst to present cost effectiveness acceptability curves, which show the probability that each intervention is cost effective at an assumed maximum willingness to pay for health gains.[17] If a model has been derived from a single dataset, bootstrapping can be used to model uncertainty—that is, repeatedly re-estimating the model using random subsamples drawn with replacement from the full sample.[18]

Structural or model uncertainty reflects the uncertainty surrounding the structure of the model and the assumptions underpinning it—for example, the way a disease pathway is modelled. Such model uncertainty is usually examined with a sensitivity analysis, re-running the model with alternative structural assumptions.[6] Alternatively, several research groups could model the same decision problem in different ways and then compare their results in an agreed way. This approach has been used extensively in fields such as climate change but less commonly in health economics. However, one example is provided by the Mount Hood Challenge, which invited eight diabetes modelling groups to independently predict clinical trial outcomes on the basis of changes in risk factors and then compare their predictions.[19] How the results from different models can be reconciled in the absence of a gold standard is unclear; however, Bojke and colleagues have recommended some form of model averaging, whereby each model's results could be weighted by a measure of model adequacy.[20]

Finally, heterogeneity should be clearly differentiated from variability because it reflects differences in outcomes or in cost effectiveness that can in principle be explained by variations between subgroups of patients, either in terms of baseline characteristics such as age, risk level, or disease severity or in terms of both baseline characteristics and relative treatment effects. As in the analysis of clinical trials, subgroups should be predefined and carefully justified in terms of their clinical and economic relevance. A model can then be re-run for different subgroups of patients.

Alternatively, heterogeneity can be addressed by making model variables functions of other variables—for example, transition probabilities between events or health states might be transformed into functions of age or disease severity. As with subgroup analysis in clinical trials, care must be taken to avoid generating apparently large differences in cost effectiveness that are not based on genuine evidence of heterogeneity.

Model evaluation

Evaluation is an important, and often overlooked, step in the development of a decision analytical model. Well evaluated models are more likely to be believed by decision makers. Three steps in model validation of escalating difficulty are face validation, internal validation, and external validation:

Face or descriptive validation entails checking whether the assumptions and structure of a model are reliable, sensible, and can be explained intuitively. This may also require experiments to assess whether setting some variables at null or extreme values generates predictable effects on model outputs.

GLOSSARY OF TERMS

- *Cost effectiveness acceptability curve*—Graphical depiction of the probability that a health intervention is cost effective across a range of willingness to pay thresholds held by decision makers for the health outcome of interest
- *Cost effectiveness plane*—Graphical depiction of difference in effectiveness between the new treatment and the comparator against the difference in cost
- *Discounting*—The practice of reducing future costs and health outcomes to present values
- *Health utilities*—Preference based outcomes normally represented on a scale where 0 represents death and 1 represents perfect health
- *Incremental cost effectiveness ratio*—A measure of cost effectiveness of a health intervention compared with an alternative, defined as the difference in costs divided by the difference in effects
- *Multiparameter evidence synthesis*—A generalisation of meta-analysis in which multiple variables are estimated jointly
- *Quality adjusted life year* (QALY)—Preference-based measure of health outcome that combines length of life and health related quality of life (utility scores) in a single metric
- *Time horizon*—The start and end points (in time) over which the costs and consequences of a health intervention will be measured and valued
- *Value of information analysis*—An approach for estimating the monetary value associated with collecting additional information within economic evaluation

bmj.com Previous articles in this series
- Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting (*BMJ* 2011;342:d1548)
- Strengthening the reporting of genetic risk prediction studies: the GRIPS statement (*BMJ* 2011;342:d631)
- Interpretation of random effects meta-analyses (*BMJ* 2011;342:d549)
- Strategy for intention to treat analysis in randomised trials with missing outcome data (*BMJ* 2011;342:d40)
- Three techniques for integrating data in mixed methods studies (*BMJ* 2010; 341:c4587)

Internal validation requires thorough internal testing of the model—for example, by getting an independent researcher or using different software to construct a replicate of the model and assess whether the results are consistent.[21] Internal validation of a model derived from a single data source, for example a Markov model being used to simulate long term outcomes beyond the end of a clinical trial, may involve proving that the model's predicted results also fit the observed data used in the estimation.[18] In these circumstances some analysts also favour splitting the initial data in two and using one set to "train" or estimate the model and the other to test or validate the model. Some analysts also calibrate the model, adjusting variables to ensure that the results accord with aggregate and observable outcomes, such as overall survival.[22] This approach has been criticised as an ad hoc search for values that makes it impossible to characterise the uncertainty in the model correctly.[23]

External validation assesses whether the model's predictions match the observed results in a population or over a time period that was not used to construct the model. This might entail assessing whether the model can accurately predict future events. For example, the Mount Hood Challenge compared the predictions of the diabetes models with each other and the reported trial outcomes.[19] External validation might also be appropriate for calibrated models.

Value of additional research

Decision analytical models are increasingly used as a framework for indicating the need for and value of additional research. These techniques indicate the probability that the decision to adopt an intervention on grounds of cost effectiveness is correct. They also allow a quantification of the cost of making an incorrect decision, which when combined with the probability of making an incorrect decision generates the expected cost of uncertainty. This has become synonymous with the expected value of perfect information (EVPI)—that is, the monetary value associated with eliminating the possibility of making an incorrect decision by eliminating parameter uncertainty in the model.[24] A population-wide EVPI can be estimated by multiplying the EVPI estimate produced by a decision analytical model by the number of decisions expected to be made on the basis of the additional information.[25] This can then be compared with the potential costs of further research to determine whether further studies are economically worthwhile.

Contributors: SP conceived the idea for this article. Both authors contributed to the review of the published material in this area, as well as the writing and revising of the article. SP is the guarantor.

Competing interests: All authors have completed the unified competing interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare no support from any organisation for the submitted work; The Warwick Clinical Trials Unit benefited from facilities funded through the Birmingham Science City Translational Medicine Clinical Research and Infrastructure Trials Platform, with support from Advantage West Midlands. The Health Economics Research Centre receives funding from the National Institute of Health Research. SP started working on this article while employed by the National Perinatal Epidemiology Unit, University of Oxford, and the Health Economics Research Centre, University of Oxford, and funded by a UK Medical Research Council senior non-clinical research fellowship. AG is an NIHR senior investigator. They have no other relationships or activities that could appear to have influenced the submitted work.

Provenance and peer review: Commissioned; externally peer reviewed.

1 Petrou S, Gray A. Economic evaluation alongside randomised clinical trials: design, conduct, analysis and reporting. *BMJ* 2011;342:d1548.
2 Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? *Health Econ* 2006;15:677-87.
3 Briggs A, Claxton C, Sculpher M. *Decision modelling for health economic evaluation*. Oxford University Press, 2006.
4 Gray A, Clarke P, Wolstenholme J, Wordsworth S. *Applied methods of cost-effectiveness analysis in health care*. Oxford University Press, 2010.
5 Drummond M, Manca A, Sculpher M. Increasing the generalizability of economic evaluations: recommendations for the design, analysis, and reporting of studies. *Int J Technol Assess* 2005;21:165-71.
6 Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart G. *Methods for the economic evaluation of health care programmes*. 3rd ed. Oxford University Press, 2005.
7 Bojke L, Hornby E, Sculpher M. A comparison of the cost-effectiveness of pharmacotherapy or surgery (laparoscopic fundoplication) in the treatment of GORD. *Pharmacoeconomics* 2007;25:829-41.
8 Legood R, Gray A, Wolstenholme J, Moss S, LBC/HPV Cervical Screening Pilot Studies Group. The lifetime effects, costs and cost-effectiveness of using HPV testing to manage low-grade cytological abnormalities: results of the NHS pilot studies. *BMJ* 2006;332:79-83.
9 Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ* 2006;15:1295-310.
10 Cooper K, Brailsford SC, Davies R. Choice of modelling technique for evaluating health care interventions. *J Oper Res Soc* 2007;58:168-76.
11 Brisson M, Edmunds WJ. Economic evaluation of vaccination programs: the impact of herd-immunity. *Med Decis Making* 2003;23:76-82.
12 Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess* 2004;8:iii-xi,1.
13 Coyle D, Lee LM. The problem of protocol driven costs in pharmacoeconomic analysis. *Pharmacoeconomics* 1998;14:357-63.
14 Ades AE, Sutton A. Multiparameter evidence synthesis in epidemiology and medical decision-making: current approaches. *J R Stat Soc* 2006;169:5-35.
15 Barton P, Bryan S, Robinson S. Modelling in the economic evaluation of health care: selecting the appropriate approach. *J Health Serv Res Policy* 2004;9:110-8.
16 Weinstein MC. Recent developments in decision-analytic modelling for economic evaluation. *Pharmacoeconomics* 2006;24:1043-53.
17 Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves: facts, fallacies and frequently asked questions. *Health Econ* 2004;13:405-15.
18 Clarke PM, Gray AM, Briggs A, Farmer A, Fenn P, Stevens R, et al. A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) outcomes model. *Diabetologia* 2004;47:1747-59.
19 Mount Hood. Computer modeling of diabetes and its complications: a report on the fourth Mount Hood challenge meeting. *Diabetes Care* 2007;30:1638-46.
20 Bojke L, Claxton K, Sculpher M, Palmer S. Characterizing structural uncertainty in decision analytic models: a review and application of methods. *Value Health* 2009;12:739-49.
21 Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. *Pharmacoeconomics* 2006;24:355-71.
22 Stout NK, Knudsen AB, Kong CK, McMahon PM, Gazelle GS. Calibration methods used in cancer simulation models and suggested reporting guidelines. *Pharmacoeconomics* 2009;27:533-45.
23 Ades AE, Cliffe S. Markov chain Monte Carlo estimation of a multi-parameter decision model: consistency of evidence and the accurate assessment of uncertainty. *Med Decis Making* 2002;22:359-71.
24 Claxton K, Ginnelly L, Sculpher M, Philips Z, Palmer S. A pilot study on the use of decision theory and value of information analysis as part of the NHS health technology assessment programme. *Health Technol Assess* 2004;8:1-103,iii.
25 Philips Z, Claxton K, Palmer S. The half-life of truth: appropriate time horizons for research decisions? *Med Decis Making* 2008;28:287-99.