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

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.

Stavros Petrou professor of health economics1, Alastair Gray professor of health economics2

1Clinical Trials Unit, Warwick Medical School, University of Warwick, Coventry CV4 7AL, UK; 2Health Economics Research Centre, Department of Public Health, University of Oxford, Oxford, UK

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.

Decision analytical modelling compares the expected costs and consequences of decision options by synthesising information from multiple sources and applying mathematical techniques, usually with computer software. The aim is to provide decision makers with the best available evidence to reach a decision—for example, should a new drug be adopted? 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, and they can be an excellent way of clarifying the options of interest. 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.5 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. Unlike decision trees, which represent sequences of events as a large number of potentially complex pathways, Markov 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. 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-oesophageal reflux disease used one month cycles to capture treatment switches and side effects, whereas an analysis of cervical cancer screening used six monthly cycles to model lifetime outcomes.

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. In order for the Markov process to end, some termination condition must be set. This could be a specified number of cycles, a proportion passing through or accumulating in a particular state, or the entire population reaching a state that cannot be left (in our example, dead); this is called an absorbing state.

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). In our example, the probability of a person dying from breast cancer is independent of the number of past recurrences and also independent of how long the person spent in the well state before moving to the recurrent state. 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.

The final stage is to assign values to each health state, typically costs and health utilities. 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 across cycles. 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.

### 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. 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.

*Discrete event simulations* describe the progress of individuals through healthcare processes or systems, affecting their characteristics and outcomes over unrestricted time periods. 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—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 on the likelihood of infection over time, and their results can differ substantially from those obtained from static models.

The choice of modelling approach will depend on various factors, including the decision maker’s requirements.

### 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. 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.
Probabilistic sensitivity analysis, in which all variables are
of joint uncertainty and correlation between variables.

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 or might not be
generalisable to the jurisdiction of interest. These
methodological considerations have increased interest in
multiparameter evidence synthesis (box) in decision analytical
modelling. These techniques acknowledge the importance of
trying to incorporate correlations between variables in models,
which may have an important influence on the resulting
estimates of cost effectiveness. 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—for example, interval
probabilities reported in the literature may have to be
transformed into instantaneous rates and then into transition
probabilities corresponding to the cycle length used in a Markov
model.

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.2 3

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, sometimes referred to as Monte Carlo uncertainty,
is not informative and needs to be eliminated by running the
model repeatedly until a stable estimate of the central tendency
has been obtained.2 3 There is little evidence or agreement on
how many model runs are needed to eliminate such variability,
but it may be many thousands.

Parameter uncertainty reflects the uncertainty and imprecision
surrounding the value of model variables such as transition
probabilities, costs, and health utilities. Standard sensitivity
analysis, in which each variable is varied separately and
independently, does not give a complete picture of the effects
of joint uncertainty and correlation between variables.4

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 therefore the preferred way of
assessing parameter uncertainty.5 Probabilistic sensitivity
analysis is usually executed by running the model several
thousand times, each time varying the parameter values across
the specified distributions and recording the outputs—for
example, costs and effects—until a distribution has been built
up and confidence intervals can be estimated. 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.6 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.7

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.8 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.9 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.10

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.11 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. For
example, Mihaylova et al, recognising the absence of evidence
of heterogeneity in treatment effect across subgroups in the
Heart Protection Study, applied the same relative risk reduction
to different subgroups defined in terms of absolute risk levels
at baseline, resulting in large but reliable differences in cost
effectiveness.12 13

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. 1 This may also require experiments to assess whether setting some variables at null or extreme values generates predictable effects on model outputs.

**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.2–3 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.22 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.2 This approach has been criticised as an ad hoc search for values that makes it impossible to characterise the uncertainty in the model correctly.30

**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.22 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. We have established that the analyst will never be certain that the value placed on each variable is correct. As a result, there are distributions surrounding the outputs of decision analytical models that can be estimated using probabilistic sensitivity analysis and synthesised using cost effectiveness acceptability curves.4 These techniques indicate the probability that the decision to adopt an intervention on grounds of cost effectiveness is correct. The techniques 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.5

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.32 This can then be compared with the potential costs of further research to determine whether further studies are economically worthwhile.33 34 The approach has been extended in the form of expected value of partial perfect information (EVPPPI), which estimates the value of obtaining perfect information on a subset of parameters in the model, and the expected value of sample information (EVSII), which focuses on optimal study design issues such as the optimal sample size of further studies.3

### Conclusions

Further detail on the design, conduct, analysis, and reporting of economic evaluations using decision analytical modelling is available elsewhere.4–6 This article and our accompanying article7 show that there is considerable overlap between modelling based and trial based economic evaluations, not only in their objectives but, for example in dealing with heterogeneity and presenting results, and in both cases we have argued the benefits of using individual patient data. These two broad approaches should be viewed as complements rather than as competing alternatives.

**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/doi_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 funds 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.

---

**References**

15. NHS Centre for Reviews and Dissemination. Underpinning systematic reviews of research on effectiveness: CRD’s guidance for those carrying out or commissioning reviews. CRD, University of York, 2001.
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.

References:


Accepted: 8 February 2011

Cite this as: BMJ 2011;342:d1766

Reprints: http://journals.bmj.com/cgi/reprintform
Subscribe: http://resources.bmj.com/bmj/subscribers/how-to-subscribe
Figures

**Fig 1** Decision tree for breast cancer screening options. The decision tree illustrates the screening policy decisions: screening every 2 years, no screening, test positive, test negative, have cancer, no cancer.

**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, 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.

<table>
<thead>
<tr>
<th>State at time t</th>
<th>State at t + 1</th>
<th>Well</th>
<th>Recurrence</th>
<th>Dead</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well</td>
<td>Well</td>
<td>1-(0.3 + 0.1)</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>Recurrence</td>
<td>Recurrence</td>
<td>0.3</td>
<td>1-(0.1 + 0.2)</td>
<td>0.2</td>
</tr>
<tr>
<td>Dead</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>