Chapter 9

Economic evaluation using decision-analytic modelling

9.1 Some basics

Chapter 8 considered economic evaluation based on a single study such as a randomized controlled trial (RCT), where data collected from a sample of patients or other study participants on resource use and outcomes facilitate estimates of differential mean costs and effects relating to different interventions. Although such studies are still undertaken and published, there is a growing use of decision-analytic modelling as an alternative vehicle for economic evaluation.

The ultimate purpose of economic evaluation is to inform different types of decision-maker about the efficient allocation of health care resources. In recent years, economic evaluation has been increasingly undertaken for specific decision-makers, who formally require economic evidence. For example, a number of health care systems now use economic evaluation to help them to decide whether new health technologies (particularly pharmaceuticals) represent sufficient value for money to be funded (see <http://www.ispor.org/HTARoadMaps/Default.asp> for details). The role of economic evaluation in decision-making is discussed in more detail in Chapter 2, but the greater use of these methods to inform particular decisions in specific jurisdictions has had implications for economic evaluation (Claxton et al. 2002). In particular, it has indicated that relying on a randomized trial as a single vehicle for economic evaluation has a number of limitations (Sculpher et al. 2006). As a result, economic evaluation for decision-making will usually need to draw on evidence from a range of sources. These could include clinical, resource use, and outcome data collected alongside randomized trials, but are also likely to include evidence from other types of study such as cohort studies and surveys. A decision-analytic model provides a means of bringing together this full range of evidence and directing it at a specific decision problem being addressed by a health system at a given point in time and in a particular jurisdiction.

Decision-analytic modelling has its theoretical foundations in statistical decision theory (Raiffa 1968), and shares common theoretical origins with expected utility theory (discussed in Chapter 5). It also has a close association with Bayesian statistics where statistical analysis is closely related to decision-making (Spiegelhalter et al. 2004). Decision analysis has been widely used outside health care, such as in business and engineering. It has an established basis as a framework for *clinical* decision-making—that is, decision-making relating to individual patients when costs are not necessarily a primary consideration—and several good texts exist introducing decision analysis in health care generally (Hunink et al. 2014; Weinstein and Fineberg 1980).

This chapter introduces the use of these methods in economic evaluation in particular. More advanced material on decision analysis for economic evaluation is available (Briggs et al. 2006).

Decision-analytic modelling provides a framework for decision-making under conditions of uncertainty. More specifically, a decision-analytic model defines a set of mathematical relationships between entities (usually health states or pathways) characterizing the range of possible disease prognoses and the impacts of alternative interventions. These entities themselves predict the quantities we are interested in for economic evaluation: costs and health effects. As a set of methods, decision analysis can satisfy five important objectives for any economic evaluation as discussed briefly here and more fully in subsequent sections.

- *Structure*. It can provide a structure that appropriately reflects the possible prognoses that individuals of interest may experience, and how the interventions being evaluated may impact on these prognoses. The structure will have to reflect the variability between apparently similar individuals in their prognoses and in the effect of interventions. The individuals will often be patients with a particular condition, but may be healthy or asymptomatic people in the context, for example, of a screening or public health programme.
- *Evidence*. It offers an analytic framework within which the full range of evidence relevant to the study question can be brought to bear. This is achieved partly through the structure of the model, but also in the estimates of the input parameters of the model.
- *Evaluation*. It provides a means of translating the full extent of relevant evidence into estimates of the cost and effects of the alternative options being compared. Using appropriate decision rules (see Chapter 4), the option that the analysis identifies as the best can be identified based on existing evidence.
- *Uncertainty, variability, and heterogeneity*. It facilitates an assessment of the various types of uncertainty relating to the evaluation. As described in detail in Chapter 11, this includes uncertainty relating to model structure and input parameters. Models also provide flexibility to characterize heterogeneity across different subgroups of individuals.
- *Future research*. Through the assessment of uncertainty, it can estimate the value of future research, and identify likely priorities for it. This can provide a more nuanced set of decision options relating to research development to be considered alongside a more standard 'adopt' or 'reject' (see Chapter 11).

9.2 **The role of decision-analytic models for economic evaluation**

In considering the role of decision models in economic evaluation, it is useful to contrast two different activities in health care evaluation—measurement and decision analysis—which are summarized in Box 9.1. In part, economic evaluation is concerned with the process of measurement through the collection of data relating to effectiveness, resource

Box 9.1 Contrasting activities in economic evaluation: measurement versus decision analysis

Health care evaluation in general, and economic evaluation in particular, involves two important but separate activities: measurement and decision analysis. Both are important in establishing the most economically appropriate form of management or intervention, but they have distinct roles. The process of measurement has the following features:

- A focus on estimating and testing hypotheses relating to particular parameters and relationships between parameters (e.g. the rate of clinical events, relative treatment effects, resource use, and HRQoL effects).
- A concentration on relatively few parameters.
- A primary interest in randomized trials as a vehicle for measurement, particularly of relative treatment effects.
- The focus of uncertainty analysis is on parameters, usually represented in terms of hypothesis tests.

The activity of decision analysis can be characterized as:

- The primacy of identifying an appropriate course of action from amongst a full range of alternatives for a specific recipient group in a particular location/jurisdiction.
- The process of informing decisions based on all relevant and currently available evidence.
- Identification of a preferred option based on the expected values of the alternatives (e.g. expected cost-effectiveness) rather than on individual parameters.
- An explicit acceptance that decisions will always be taken under conditions of uncertainty.

use, unit costs, and health-related quality of life (HRQoL) weights. Ultimately, though, economic evaluation is concerned with informing appropriate decisions in health care about resource allocation in specific jurisdictions under conditions of uncertainty. This will require the appropriate synthesis of all relevant evidence (see Chapter 10). Being clear about these different roles for economic evaluation emphasizes that decision models and randomized trials (and other clinical studies) are not competing alternatives. Rather, the latter are focused on measuring different effects of interventions on relevant costs and outcomes. Decision models, on the other hand, are concerned with informing specific decisions (Sculpher et al. 2006); they draw upon the measurements undertaken in clinical studies but also provide an explicit framework for the inevitable assumptions and judgements needed in decision-making.

This section considers the features of economic evaluation for decision-making that frequently necessitate the use of decision-analytic models rather than reliance on a

single data source such as an RCT. There are several different requirements for an economic evaluation that are relevant, and these are detailed in Sections 9.2.1–9.2.6.

9.2.1 The need to compare all options

As described in Chapters 2, 3, and 4, economic evaluation is about comparing the value for money of alternative courses of action (or options) for particular recipient groups. It is possible that decision-makers will be misled by a study that fails to compare all the relevant options, which might reflect the fact that more than one option is currently used in practice as well as the new option(s) available. Indeed, defining all relevant options will often involve specifying strategies rather than specific treatments. In the evaluation of pharmaceuticals this can take the form of the comparison of sequences of interventions. For example, Woolacott and colleagues compared the cost-effectiveness of alternative sequences of treatments for psoriasis where a treatment is typically used until it ceases to be effective or exhibits side effects (Woolacott et al. 2006). The alternative sequences consisted of several new biological treatments as well as older therapies and best supportive care. In the economic evaluation of diagnostics, strategies are specified in terms of sequences of tests and alternative decision rules about therapeutic choices conditional on test results. As a result, the number of options to compare can be very large. A particularly notable example is the 12^{14} (i.e. 1 283 918 464 548 860) options initially compared for prenatal screening and treatment to prevent group B streptococcal and other bacterial infections in early infancy (Colbourn et al. 2007). Clearly, in such circumstances, approaches need to be identified to make the number of options practical. In this study, uncertainty analysis was used to remove options from consideration that had a very limited probability of being cost-effective.

However, as discussed in Chapter 8, in randomized trials it is rarely the practice to compare all relevant options and a subset is typically studied. In some studies, moreover, the comparator is not an active intervention at all, but a placebo. To compare all relevant options, it is likely that effectiveness data will have to be taken from several trials. For example, in a cost-effectiveness analysis (CEA) of primary angioplasty compared to medical therapy in patients with myocardial infarction, 22 randomized trials were used as the basis of estimating the relative effectiveness for the analysis (Bravo Vergel et al. 2007). Meta-analysis and other forms of evidence synthesis will be necessary to synthesize this type of evidence (see Chapter 10), but the decision model will provide the framework to combine it with other types of evidence such as the underlying risk of clinical events (sometimes called natural history), resource use, and HRQoL weights. Models also provide a means of structuring the relationships between clinical variables and how their magnitudes change over time.

9.2.2 The need to reflect all relevant evidence

To offer a decision-maker guidance on the best course of action from an economic perspective, for a given patient group, it is important that all relevant evidence is brought to bear on the decision problem. This is consistent with the axioms of evidence-based medicine, where appropriate evidence is systematically and comprehensibly used to

make clinical decisions (Sackett et al. 1996). For economic evaluation, however, it is not just effectiveness evidence that is required. In addition, evidence relating to resource use, unit costs, and HRQoL is necessary, as well as the relationship between different parameters and how they change with time. This array of evidence is often not collected in trials and, if it is, it is unlikely to be the only source. Again, the decision model is used to combine all sources of evidence.

9.2.3 The need to link intermediate to final end points

As discussed in Section 8.2.1.2, many clinical studies measure effects in terms of end points which are clinically meaningful but which are only indirectly related to the ultimate measures of health that are central to most economic evaluations (e.g. changes in life expectancy and/or patients' HRQoL). Examples are the use of CD4 count and viral load in studies of the efficacy of therapies for HIV, time until progression in cancer, and cases detected in screening studies. Intermediate end points are often a challenge when assessing the cost-effectiveness of diagnostic strategies as clinical studies often focus on the accuracy of tests using, for example, estimates of sensitivity and specificity. A focus on these *intermediate* end points will limit the value of clinical studies as a vehicle for economic evaluation as they are unlikely to provide a reliable estimate of ultimate health outcomes. Decision analysis provides a means of linking intermediate and final outcomes. Evidence is still required on the relationships between the different types of outcomes, and these may be taken from observational studies. The clinical plausibility of these links and the associated uncertainty needs to be fully considered in developing the model. Box 9.2 contains some examples of studies which used models to link intermediate to final outcomes for economic evaluation.

9.2.4 The need to extrapolate over the appropriate time horizon of the evaluation

The issue of the appropriate time horizon for an economic evaluation was discussed in Chapter 3. In principle, the time horizon should be the period over which the costs and/or effects of the alternative options being compared might be expected to differ. Often the appropriate time horizon will need to be the patient's lifetime to capture these differences fully. For example, in the case of the treatment of chronic disease, the initiation of an intervention in a middle-aged patient may have cost and effect implications for the remainder of their life. Except in rare cases where palliative treatments are being compared (e.g. for advanced cancer), clinical studies will not follow all patients up until they die. An important role of decision models, then, is to bridge the gap between what has been observed in trials and what would be expected to happen, in terms of costs and effects, over a long-term time horizon.

It is frequently necessary to extrapolate when the options being compared differ in terms of mortality and this difference is to be expressed in terms of life-years or quality-adjusted life-years (QALYs) gained. This is illustrated in Figure 9.1 in the form of the survival curves of two interventions (treatment and control) being compared in a randomized trial. These show the proportion of patients surviving until particular

Box 9.2 Examples of studies where decision models have been used to link intermediate end points to measures of ultimate health outcome

Study	Intermediate endpoints	Method of extrapolation
A model to assess the cost-effectiveness of donepezil in mild to moderate Alzheimer's disease (AD) (Neumann et al. 1999)	Treatment effect from trial between baseline and 24 weeks in terms of transition between Clinical Dementia Rating (CDR) scale 2 (moderate) and CDR 0.5 or 1 (mild)	Markov model with states defined in terms of mild, moderate, and severe disease (in terms of CDR); also a dead state. Baseline disease progression between states was taken from a longitudinal cohort study. HRQoL weights were estimated for each Markov state using the Health Utilities Index (HUI)-2 (see Chapter 5) in a cross-sectional survey of caregivers of AD patients. The treatment effect was applied to the baseline transitions, with several alternative durations. After coming off the drug, patients return to baseline progression rate
A model to assess the cost-effectiveness of infliximab in rheumatoid arthritis (Kobelt et al. 2003)	Treatment effect in terms of change in Health Assessment Questionnaire (HAQ)- score, which is focused on functional disability—over a period of 54 weeks	A Markov model was developed with seven states—six relating to HAQ levels (functional disability) and one for death. HRQoL weights for each HAQ state were estimated using the EQ5D. The trial data were used to show short-term movements between the HAQ states on infliximab and for its comparator. Data from cohort studies were used to estimate longer-term transitions between states
A model to assess the cost-effectiveness of cardiovascular magnetic resonance in the diagnosis of coronary heart disease (Walker et al. 2013)	The sensitivity and specificity of a range of testing strategies was estimated in a non-randomized study with a reference standard of coronary angioplasty	A combination of a decision tree, to capture alternative diagnostic pathways, and a Markov model, to reflect treatments and their effect on outcomes, was used. The prevalence of disease requiring revascularization, together with the sensitivity and specificity of each test, were used to estimate the probability of a given patient being in one of four groups: true positive, false negative, true negative, and false positive. Based on group allocation, treatments received or delayed were modelled. Evidence from other studies linked treatments to outcomes in terms of rates of non-fatal cardiovascular events and mortality. The HRQoL effects, costs, and prognostic implications (in terms of mortality risk) of non-fatal events were included

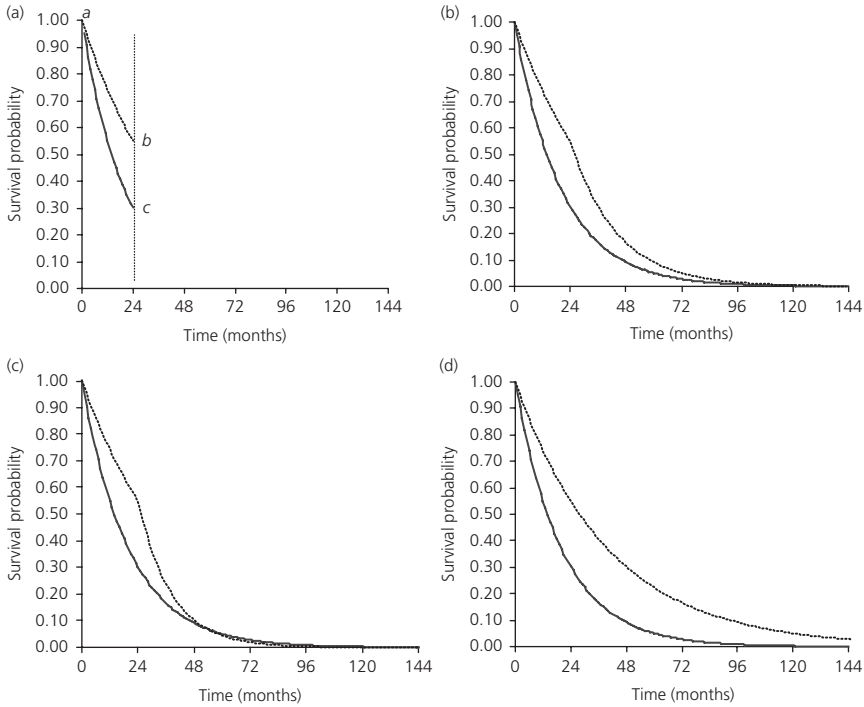


Fig. 9.1 Alternative extrapolation assumptions relating to survival data observed in a trial over a 24-month follow-up. Solid line, control group; dotted line, treatment group. (a) The curves as observed in the trial, with a 50% reduction in mortality rate with treatment compared with control. (b) The survival curves extrapolated over 144 months with the assumption of a 'one-time' benefit to patients. (c) Extrapolation with a rebound effect. (d) Assumption of a continuous treatment effect.

points of follow-up (measured in months). Figure 9.1a relates to the maximum follow-up during the trial of 24 months. It indicates that, in terms of mortality, treatment is more effective than control as it reduces the mortality rate by 50%. If the time horizon of the study is taken to be the same as the maximum follow-up in the trial, the measure of life-years gained from surgery is equivalent to the area between the two curves over 2 years (area abc in Figure 9.1a). This estimate of the *restricted* mean life-years over a 2 year time horizon assumes that patients who remain alive at the end of trial follow-up receive a maximum benefit of 2 years additional life expectancy—this effectively assumes that they die at the end of the trial! In reality, the patients living at the end of the trial will continue living afterwards, including the additional patients alive having received the treatment, so this 'within-trial' measure of life-years gained will inevitably be an underestimate.

The use of modelling to *extrapolate* beyond the follow-up period in the clinical study involves predicting what the survival curves will look like beyond what has been observed. A key question with extrapolation relates to appropriate assumptions about the shape of the survival curves after follow-up. For interventions that take place only

during the trial period, one assumption is that the more effective treatment during the trial confers a 'one-time' benefit to patients. As illustrated in Figure 9.1b, this means that, beyond the period of the trial, the rate of death per period of time, conditional on surviving until the end of the trial, is the same for patients originally allocated to treatment and to control. In other words, beyond the 24 month period, the 50% treatment effect (the reduction in the rate of mortality in the trial) is assumed to end and both groups become identical in terms of the mortality rate. The area between the two survival curves represents the gain in mean survival duration with treatment. This assumption of a one-time benefit has been used in several studies, particularly in the cardiovascular field. For example, in the base-case of their CEA of alternative thrombolytic therapies for acute myocardial infarction, Mark and colleagues used a model to extrapolate beyond the 1 year period of trial follow-up (Mark et al. 1995). To do this they used a separate source of data (a register of patients who had experienced acute myocardial infarction and survived the first year) to provide 15 year estimates of risk of mortality; beyond 15 years these risks were based on general population data.

In some contexts, it might be more appropriate to assume that the survival curves converge more rapidly after the trial follow-up period. That is, the conditional rate of death beyond trial follow-up becomes higher with treatment compared to control. This is illustrated in Figure 9.1c, where it is assumed that, beyond the trial follow-up period, the mortality rate increases by 40% in the treatment arm compared to control. This could happen, for example, if the more effective intervention delays the death of a high-risk subgroup of patients who, once treatment is ended (at the end of trial follow-up), die at a faster rate than those patients surviving in the other arm. This scenario is sometimes known as a 'rebound effect'. It can be seen that the area between the survival curves (the gain in mean survival duration with treatment) is less than when a 'one-time' benefit is assumed. The scenario was one considered in a CEA of endovascular aneurysm repair which extrapolated from a randomized trial (Epstein et al. 2008).

At the other extreme, it may be reasonable to assume that the treatment confers a continuous benefit beyond trial follow-up, and this is shown in Figure 9.1d. That is, the curves continue to diverge in the longer term, the 50% reduction in mortality of treatment continues, and patients randomized to that option continue to die at a slower rate. It can be seen that the area between the curves, assuming a continuous benefit, is larger than in Figures 9.1b and 9.1c. This may be a more appropriate assumption when treatment is still ongoing at the end of trial follow-up when, of course, costs are extrapolated as well as benefits. This was the base-case assumption of a CEA of ivabradine in heart failure, although alternative scenarios were used in sensitivity analysis (Griffiths et al. 2014).

The issue of the most reasonable assumption can be informed by the shape of the survival curves within the trial. For example, if they are ceasing to diverge in the latter period of follow-up, an assumption of continued divergence in the extrapolation is likely to be unwarranted. External non-trial data may also hold some clues. It is also important, however, to identify an appropriate assumption on the basis of what is known about the biology of the intervention—for example, in the case of a pharmaceutical, the length of time it remains active in the patient's body. Chapter 10 considers the use of evidence to model long-term extrapolation in more detail.

The choice of assumption made regarding extrapolation may have major implications on study results. For example, in an early cost-effectiveness study of therapy for patients with HIV, Schulman et al. (1991) estimated the incremental cost per life-year gained under two alternative assumptions about the effect of zidovudine on the development of AIDS and hence on mortality: a one-time effect and a continuous effect. The incremental cost per life-year gained of therapy ranged widely, from $6553 to $70 526 under those two scenarios. This shows that it is important to run alternative scenarios regarding plausible extrapolation assumptions. The judgements about plausibility are usually based on our knowledge about the epidemiology of the disease and the effects of other treatments that have been evaluated in the past.

It should be noted that extrapolating survival curves relates not just to mortality; rather, this can relate to any evidence on the time until a particular event and can include, for example, time until cancer progression and time until non-fatal cardiovascular event. Extrapolation can also relate to other assumptions regarding the effectiveness of treatments, for example what happens to outcomes once treatment is discontinued. An example is a CEA of biological therapies for psoriatic arthritis (Bojke et al. 2011). Evidence on efficacy was available for three biological therapies based on trial data with 3 month follow-up, and clinical treatment effects were mapped to HRQoL weights to estimate QALYs. It was assumed that HRQoL improvement continued while the patient was on therapy. When biological treatment was withdrawn, the effect on patients' HRQoL was modelled in terms of two alternative profiles. Firstly, that their HRQoL returned to what it was prior to treatment; or, alternatively, it rebounded to what it would have been had they never received treatment. The first of these was used in the base case, informed by a formal elicitation of opinion from clinical experts (see Chapter 10). Sensitivity analysis showed cost-effectiveness to be highly sensitive to this assumption.

9.2.5 The need to make results applicable to the decision-making context

Another situation where there may be a gap between the available evidence, particularly from randomized trials, and the requirements for a decision, relates to situations where the decision problem being addressed is inconsistent with the nature of the available clinical evidence. An example of the use of decision models to relate available evidence to a particular decision context is when some types of evidence are available from outside the jurisdiction where the decision is being taken and may not generalize to that location. In such a context, a model can be used to combine this outside evidence with other information and explicit assumptions to inform the relevant decision. An example of a study where the generalizability of clinical evidence to the decision-making jurisdiction, the United Kingdom, was subject to doubt was a CEA of glycoprotein IIb/IIIa antagonists for acute coronary syndrome where effectiveness was assessed in terms of the rate of fatal and non-fatal cardiovascular events (Palmer et al. 2005). As most of the trials of these therapies had randomized patients from outside the United Kingdom and, at the time, some aspects of cardiac care in the United Kingdom were considered to be at variance with those in the trials, the relevance of this evidence to the decision needed to be carefully assessed. The approach adopted was to distinguish the underlying (or baseline) rates of cardiovascular events under

usual practice (i.e. without the new drug therapy) from the *relative* effectiveness of the glycoprotein IIb/IIIa antagonists on that baseline risk. The former type of evidence was taken from a longitudinal observational study undertaken in the United Kingdom rather than from the randomized trials. Relative effectiveness, however, was based on a meta-analysis of the trials. Brought together within the model, these two types of evidence generated estimates of the absolute reduction in the risk of cardiovascular events generated by the new therapy. The explicit assumption with this approach was that the *relative* risk of events with glycoprotein IIb/IIIa antagonists (compared with usual care) from the trials generalizes to routine care in the United Kingdom; however, the absolute baseline risk does not generalize from the trials and needed to be estimated from UK sources. This example highlights a more general point with RCTs: that they are mainly designed to estimate relative effectiveness, and that the baseline rates of event may vary significantly between jurisdictions and indeed between patient subgroups. A key role for decision models, then, is often to bring together jurisdiction- or subgroup-specific estimates of baseline risk, often derived from non-randomized studies, with relative effectiveness estimates from a single RCT or meta-analysis.

Issues of generalizability in economic evaluation in general, and the use of models to facilitate the generalizability of evidence, have been considered (Sculpher et al. 2004). More recently, an International Society of Pharmacoeconomics and Outcomes Research (ISPOR) Task Force considered good practice in assessing and implementing transferability of economic evaluations across jurisdictions and made recommendations in areas including the interpretation of studies, economic evaluations based on individual patient studies (see Chapter 8), modelling studies, and further research activities (Drummond et al. 2009).

9.2.6 Using models to assess heterogeneity

Decision models can also be used to identify subgroups of a wider population of patients in whom an intervention is cost-effective, based on patients' clinical and socio-demographic characteristics. Understanding heterogeneity in cost-effectiveness between different types of patients is important in decision-making because, in principle, different decisions regarding the funding of interventions can be made for different subgroups (Sculpher 2008). Indeed, failing to reflect heterogeneity in economic evaluation (and hence in decisions) can impose costs on the health system in terms of opportunities for health gain forgone or resources wasted. This is because some types of patients will receive treatments which are not the most cost-effective given their characteristics (Coyle et al. 2003). The term *expected value of individualized care* has been developed (Basu and Meltzer 2007) to represent the added value of reflecting heterogeneity analytically and in decisions.

Chapter 8 considered how regression methods can be used to assess heterogeneity based on patient-specific estimates of total costs and benefits based on single studies such as trials. An advantage of the use of decision models in this context is that heterogeneity in different types of evidence can be more explicitly considered. Interest in subgroups for CEA extends beyond the clinical trialist's usual concerns regarding whether the relative effectiveness of an intervention is consistent across different types of patients (formally, whether any patient characteristics at baseline are treatment effect

modifiers). Economic evaluation includes consideration of possible heterogeneity in baseline risks, costs, and HRQoL. Heterogeneity in patients' preferences for different states of health has also been considered using modelling (Basu and Meltzer 2007; Owens and Nease 1997; Sculpher and Gafni 2001).

As discussed in Section 9.2.5, the methods used to estimate cost-effectiveness for specific subgroups often divides the absolute clinical benefit of a treatment upon which cost-effectiveness is based (e.g. the absolute reduction in the risk of an event such as the rate of cardiovascular events in the glycoprotein IIb/IIIa antagonists example in Section 9.2.5) into two elements: baseline risks and relative treatment effects. Baseline risks are the measure of events under (one of) the comparator intervention(s). The relative treatment effect is often a ratio (e.g. an odds ratio, relative risk, or hazard ratio) representing the effectiveness of the newer therapy *relative* to the comparator intervention, which is typically the main focus when the clinical results of a randomized trial are reported. Often the clinical report of a trial will indicate that there is no evidence of differences between subgroups in terms of *relative treatment effect*. However, cost-effectiveness is driven by *absolute* benefit, and there may still be important heterogeneity between subgroups in baseline event rates. Indeed, an assumption of constant relative effects being applied to subgroup-specific baseline event rates is common in cost-effectiveness models.

An example of a modelling study using this assumption compared lifetime costs and QALYs of two alternative hip prostheses used in primary hip replacement (Briggs et al. 2004). The effectiveness of the two prostheses, in terms of their failure rate over time, was taken from a large register developed in Sweden. Reflecting the register data, the model allowed the baseline failure rate (that for the 'usual care' prosthesis) to vary by the patient's age and sex. The relative reduction in failure rate with the newer prosthesis was, however, assumed constant over those subgroups. It is also important to note that the 'background' mortality rate (i.e. the population rate from all causes) is known to vary by age and sex. As death from reasons unrelated to hip replacement was also included in the model (as a 'competing risk'), this increases the differences between age and sex subgroups in terms of the cost-effectiveness of the newer prosthesis. Using the top-right and bottom-right quadrants of the cost-effectiveness plane (see Figure 3.2), Figure 9.2 shows how the cost-effectiveness of the newer prosthesis varied by age and sex. The top line relates to incremental costs and effects of the newer prosthesis relative to the older one for females, with various points on that line representing subgroups based on patients' age. The lower line shows the similar relationship for males. It can be seen that, for both males and females, the newer prosthesis is more cost-effective for the younger age groups: the newer device is dominant for men aged 70 or younger and for women aged 60 or younger. This is because the lifetime risk of failure with the older prosthesis is higher in younger patients because they impose greater wear and tear on their hips due to their greater activity, and because they would be expected to live longer. Therefore, a constant relative reduction in the failure rate with the newer prosthesis will confer a greater absolute benefit in these younger age groups. In addition to heterogeneity in the cost-effectiveness of the newer prosthesis by age, it is also less cost-effective in women (i.e. the line in Figure 9.2 is higher). Again, this is because men have a higher baseline event rate with the existing prosthesis as their activity levels place greater stress on their hips.

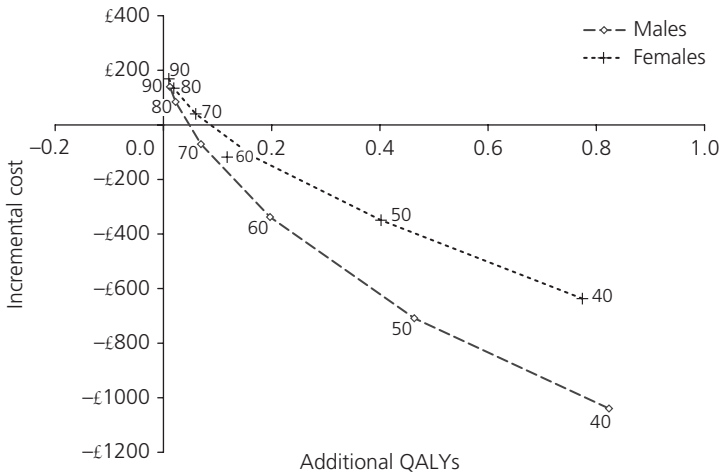


Fig. 9.2 Results of a cost-effectiveness study by Briggs et al. (2004) showing heterogeneity in the cost-effectiveness of a newer prosthesis by patient subgroup defined in terms of age and sex. QALY, quality-adjusted life-year.

In some clinical settings the cost-effectiveness of an intervention is very sensitive to a range of parameter values which are heterogeneous across a number of different patient-specific characteristics. For example, Henriksson and colleagues considered the cost-effectiveness of early intervention with angiography and possible revascularization, compared to pharmaceutical therapy alone, in patients with non-ST-elevation myocardial infarction (Henriksson et al. 2008). Based on evidence from an RCT, the study quantified the probability of non-fatal myocardial infarction or cardiovascular death as a function of the intervention received as well as a series of baseline patient characteristics including age, presence of diabetes, whether or not there had been a prior myocardial infarction, and smoking status. Mean costs during the initial hospitalization and 1 year afterwards were also estimated as a function of treatment received as well as covariables including age, sex, and severity of angina. Finally, patients' HRQoL at baseline and the change over follow-up was also estimated as a function of a similar set of covariables. Ultimately, the cost-effectiveness of early intervention was shown to be heterogeneous with respect to a patient's underlying risk of a further clinical event. Incremental cost-effectiveness ratios ranged from £12 750 per additional QALY in the most high-risk patients to £53 760 per additional QALY in the most low-risk patients. When heterogeneity in the relative effectiveness of early intervention—as well as the underlying risk of events—was allowed for, early intervention became dominated by pharmaceutical therapy in the most low-risk patients with an incremental cost-effectiveness ratio (ICER) of £10 476 in the most high-risk patients.

9.3 Key elements of decision-analytic modelling

There are some key elements to decision analysis that are common to all models. These are the use of probabilities to reflect the likelihood of events or changes in health, and the expected values to inform decisions. Different model types such as decision trees and Markov models could also be added to this list, but these are discussed more fully in Section 9.4.

9.3.1 Probabilities

Probabilities are used widely in quantitative methods in many fields, and have an important role in clinical decision-making (Weinstein and Fineberg 1980). A common way of thinking about probability is as the measured frequency of an event in a given sample or population. For example, if a sample of 200 patients is treated with a particular medicine over 1 year and 10 patients have an adverse event, the proportion of 0.05 can be taken as an estimate of the 1 year probability of a patient experiencing an adverse event with that therapy. Assuming this estimate generalizes to other patients, a future patient will either experience the adverse event or not; when the decision is being taken regarding whether to administer the therapy, the uncertain outcome for the individual can be expressed in terms of a probability estimated from the experiences of similar patients.

This concept of probability as a number indicating our 'state of knowledge' regarding whether an event will or will not take place is a feature of Bayesian statistics, and is not shared with the classical or 'frequentist' statistical methods that are widely used in the analysis of randomized trials (Spiegelhalter et al. 2004). This emphasizes the common origins of decision analysis and Bayesian statistics. This concept of probability can be generalized to represent strength of belief which, for a given individual, is based on their previous knowledge and experience. This view of probability is important in decision analysis as, in many analyses, especially when the probability of particular events may not have been informed by formal studies such as trials, estimates from relevant experts may need to be elicited. Given that decisions about the use of finite resources have to be taken regardless of the strength of the evidence available, on the basis of assumptions and judgements, decision analysis provides an analytical framework within which this can be done explicitly. Chapter 10 considers the use of elicitation of expert beliefs as an input into decision models in more detail. Bayesian statistical methods are also valuable in thinking about uncertainty in the context of decision-making (see Chapter 11).

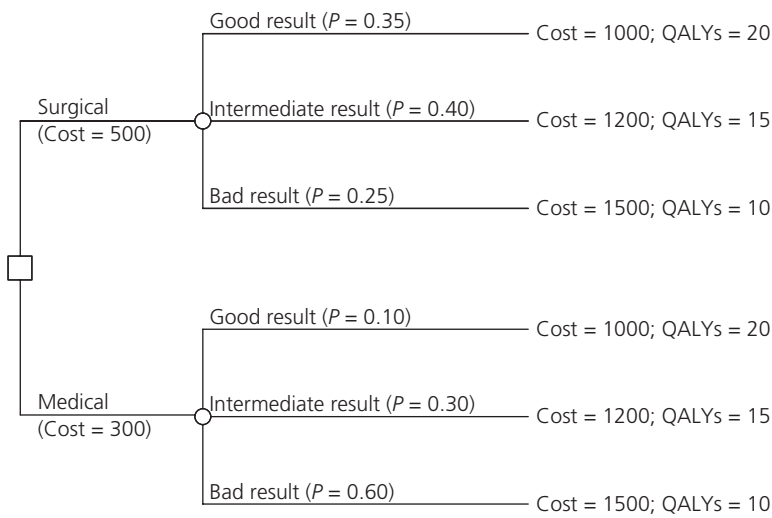
Some specific probability principles are also important in decision analysis, and these are summarized in Box 9.3.

9.3.2 Expected values

A key concept in decision analysis is the expected value of the costs or outcomes or a measure of cost-effectiveness of an option. This is illustrated in Figure 9.3, which compares two alternative interventions, medical and surgical. For each intervention, a given patient can follow one of three possible pathways which result, respectively, in a bad, intermediate, or good outcome. Before treatment, it is unknown which pathway a specific patient will follow, but probabilities are used to express the likelihood of each

Box 9.3 Probability concepts

Joint probability	The probability of two events occurring concomitantly. In terms of notation, the joint probability of events A and B: *P*[A and B]. When the events are independent, *P*[A and B] = *P*[A] × *P*[B]
Conditional probability	The probability of an event A given that an event B is known to have occurred. The notation is *P*[A|B]. Joint and conditional probabilities are related in the following equation: *P*[A and B] = *P*[A|B] × *P*[B]
Independence	Events A and B are independent if *P*[A] is the same as *P*[A|B]



Expected cost of surgery: 500 + (0.35 × 1000) + (0.40 × 1200) + (0.25 × 1500) = 1705
Expected QALYs of surgery: (0.35 × 20) + (0.40 × 15) + (0.25 × 10) = 15.5
Expected cost of medicine: 300 + (0.10 × 1000) + (0.30 × 1200) + (0.60×1500) = 1660
Expected QALYs of medicine: (0.10 × 20) + (0.30 × 15) + (0.60 × 10) = 12.5

Incremental cost per QALY gained of surgery: (1705 – 1660)/(15.5 – 12.5) = 15

Fig. 9.3 Simple decision tree showing example of the calculation of expected values. QALY, quality-adjusted life-year.

occurring. These are likely to differ by therapy. For the alternative therapies, each pathway has a cost and an outcome expressed in terms of QALYs; there is also a cost of the intervention itself which is incurred whatever pathway the patient follows. For each of the therapies, an expected cost and expected outcome can be calculated. The expected cost is the cost of the intervention plus the therapy-specific sum of the costs of the three pathways, weighted by the probability of a patient following each pathway with that

treatment. The same idea is applied to calculate expected outcome. On that basis, it is clear that surgery has both a higher expected cost and higher expected QALYs. Using the methods of incremental analysis introduced in Chapter 3, the incremental cost per additional QALY generated by surgery can be calculated.

The concept of expected value is clearly analogous to the mean value of an end point when sample data are available. In a trial-based CEA, for example, the mean costs and QALYs across patients in each of the randomized groups are used as the basis of the incremental analysis (see Chapter 8). As for mean values in studies based on a single trial, the expected value from a decision model represents the best estimate of the end-points of interest for decision-making. As decision analysis shares common theoretical origins with expected utility theory described in Chapter 5, the expected values calculated in decision models were originally seen as being strictly von Neumann–Morgenstern utilities. Given that expected utility theory is a normative framework for decision-making under conditions of uncertainty, the expected utilities from decision models would provide a clear indication of the preferred option from those being compared. However, decision analysis is widely used for situations when outcomes other than von Neumann–Morgenstern utilities are used. The expected value can still provide the key input to guide decision-making as long as the outcomes have been chosen appropriately.

9.4 Stages in the development of a decision-analytic model

The development of a decision-analytic model for economic evaluation involves a number of stages. This section considers each of the stages to provide a fuller understanding of the role of decision modelling in this field.

9.4.1 Defining the decision problem

One of the key stages in the development of the model is the specification of the question being addressed, sometimes called the decision problem. This process closely mirrors the specification of the study question for economic evaluation in general as discussed in Chapters 2 and 3. In particular, there is a need to define the recipient group (patients or others) and the relevant options being compared. It is important to emphasize that, in defining these options, this may include more than specific interventions. It may include, for instance, starting and stopping rules for treatments—for example, when to start and stop medical therapy for a particular chronic condition. As discussed in Section 9.2.1, for some evaluations the options will represent *clinical treatment strategies* or *pathways*, such as the sequence of therapies that might be used for the treatment of a condition characterized by treatment failure with some therapies. An example of such a study is a decision model looking at the cost-effectiveness of alternative therapies for epilepsy where assumptions were made about which therapies patients were placed on if they failed on initial treatment (Wilby et al. 2005).

9.4.2 Defining the boundaries of the model

All models are simplifications of reality so, in developing an analysis, decisions have to be taken about what to include. In part, this relates to general issues in economic

evaluation such as the choice of perspective, the appropriate measure of effect/benefit, and the time horizon (see Chapter 2). However, it is important to consider how far a model should go to cover all the possible implications of an intervention or programme. A simple example would be whether or not to include rare side effects in a model. A more complex example is that, in considering the cost-effectiveness of antibiotic treatment for a given condition, the issue of the cost and health effects of antibiotic resistance could be an important consideration for decision-making, but relatively few cost-effectiveness models consider this aspect—that is, they have drawn the boundaries of the model to exclude it from consideration. For decision-making, it would be important to consider whether this exclusion limits the value of the model. The CEA of alternative preventive strategies for group B streptococcal infection referred to in Section 9.2.1 assessed the extent to which resistance would have to impact negatively on population health to change the cost-effectiveness ordering of the options evaluated (Colbourn et al. 2007).

Another example is a decision model that was developed to assess the cost-effectiveness of routine antenatal HIV testing (Ades et al. 1999). These authors assessed the impact of different screening strategies on the extent to which a woman's HIV status was known during pregnancy. Through the use of interventions, this knowledge could have three beneficial health effects on: (1) the woman through earlier use of antiretroviral therapy; (2) the child through the use of interventions to reduce the mother-to-child (vertical) transmission rate; and (3) the child through the earlier use of antiretroviral therapy and prophylaxis if the child is born with HIV. The broader health benefit relating to reductions in infections to others through changes in sexual behaviour (horizontal transmission) was not, however, considered in the model, thus defining the model boundary.

Decisions about the boundaries in decision models will partly be based on the availability of data and complexity of the modelling task, but they should mainly be driven by the extent to which extending the boundaries (adding complexity) is considered likely to impact on the cost-effectiveness of the options being compared and hence the most appropriate decision.

9.4.3 Conceptualizing a decision model

A key stage in the development of a decision model is the process of deciding on a structure. Formally, this involves a series of decisions concerning how the input parameters in the model are to be related and, in particular, choices about how to characterize the clinical events and health states of interest (e.g. episodes of a disease, disease progression, case identification). Each economic evaluation brings with it different structural issues, but a few common ones are given below.

- Do the events of interest occur just once (e.g. death) or could they happen several times over the relevant time horizon (e.g. a non-fatal myocardial infarction)?
- Are patients at risk of several events over time (i.e. *competing* risks); for example, the risk of a heart attack but also of stroke?
- As discussed in Section 9.2.4, when extrapolating events over time, what is the durability of the effectiveness of an intervention relative to comparators?

- Do the probabilities of events change as time elapses or are they constant with respect to time?
- Are all important events included and has double counting of events been avoided?
- Does a patient's prognosis partly depend on the events they have already experienced in the model?
- For the management of a chronic disease, does the structure of the model allow for the costs and effects of subsequent therapies to be included?
- Is the clinical prognosis of a given patient partly dependent on the clinical status of other patients as might be the case, for example, with infectious diseases?

The way in which these issues are handled can essentially be defined as a series of mathematical relationships between parameters. Indeed, some studies present the structure of their decision models in terms of a series of equations (e.g. see Spiegelhalter and Best 2003). Most decision models used in economic evaluation, however, present the structure of their model schematically. The collaboration between the ISPOR and the Society of Medical Decision Making (SMDM) developed a number of useful guidance documents, one of which focused on model conceptualization (Roberts et al. 2012).

9.4.4 Implementing a model—the decision tree

The judgements that are taken by an analyst to characterize the nature of a disease and the impacts of alternative interventions are implemented in a specific decision model. Various types of model are used for economic evaluation, but one of the most widely used is the *decision tree*.

The decision tree represents individuals' possible prognoses, following some sort of intervention, by a series of pathways. Figure 9.3 used a simple decision tree to explain the concept of expected value. A clinical example is used to illustrate the use of decision models in more detail—the comparison of two antiemetic prophylactic therapies for patients undergoing chemotherapy for cancer. The example is based on a published study comparing ondansetron and metoclopramide which was undertaken before a price had been determined for ondansetron (Buxton and O'Brien 1992). The study considered the acquisition cost of the therapies, as well as the cost of the adverse events and of their treatment, and the cost of treatment failure (i.e. an episode of emesis). Effects were expressed in terms of the probability of a patient being successfully treated, which was defined as the absence of emesis and adverse events. The decision tree is shown in Figure 9.4, and this can be used to describe a series of general features with this sort of model structure.

9.4.4.1 Nodes

Decision nodes The square box at the start of the tree is a decision node and represents the decision being addressed in the model: here, which of ondansetron and metoclopramide is the more cost-effective in preventing episodes of emesis without adverse events.

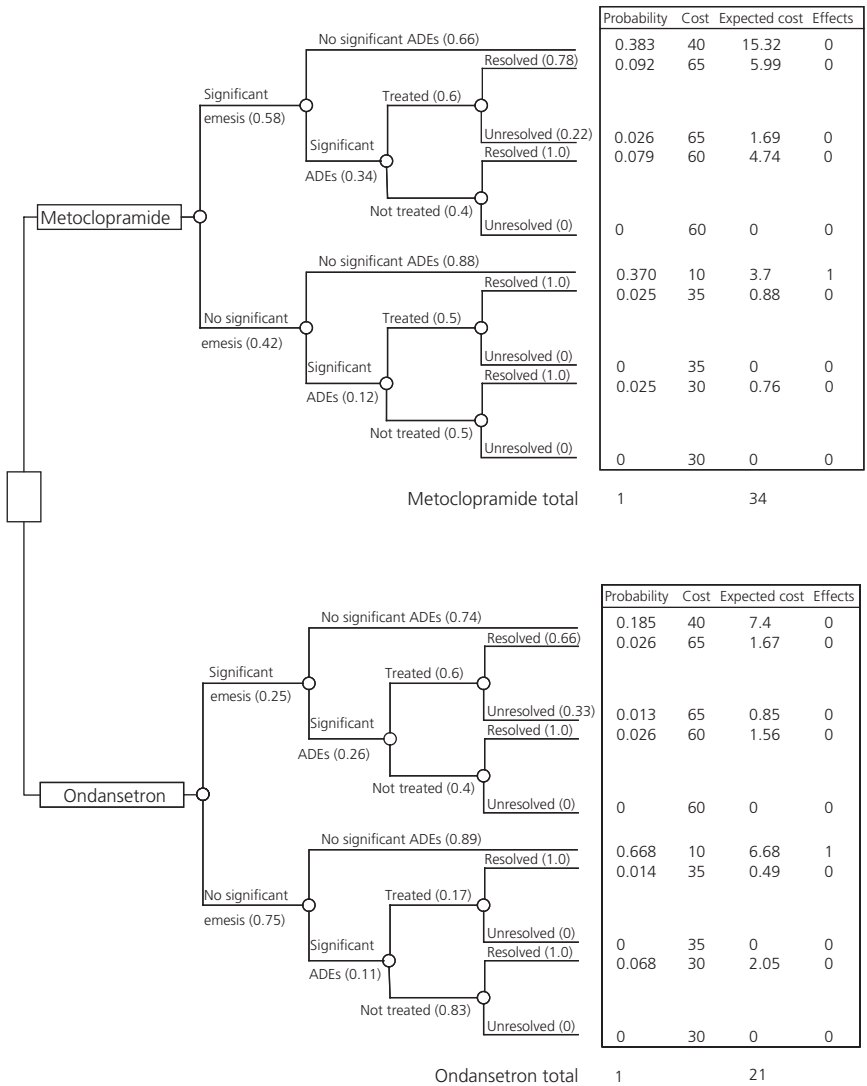


Fig. 9.4 Example of a decision tree taken from Buxton and O'Brien (1992). Other inputs into the model: price of both treatments, £10; cost of an episode of emesis, £30; cost of side effects, £20; cost of treating side effects, £5. ADE, adverse drug events.

Chance nodes Coming out of the decision node is the range of possible pathways that characterize the effects of the alternative therapies. The pathways are built up through a series of branches representing particular events. Here the events are significant emesis, significant adverse events, the treatment of adverse events and the resolution of adverse

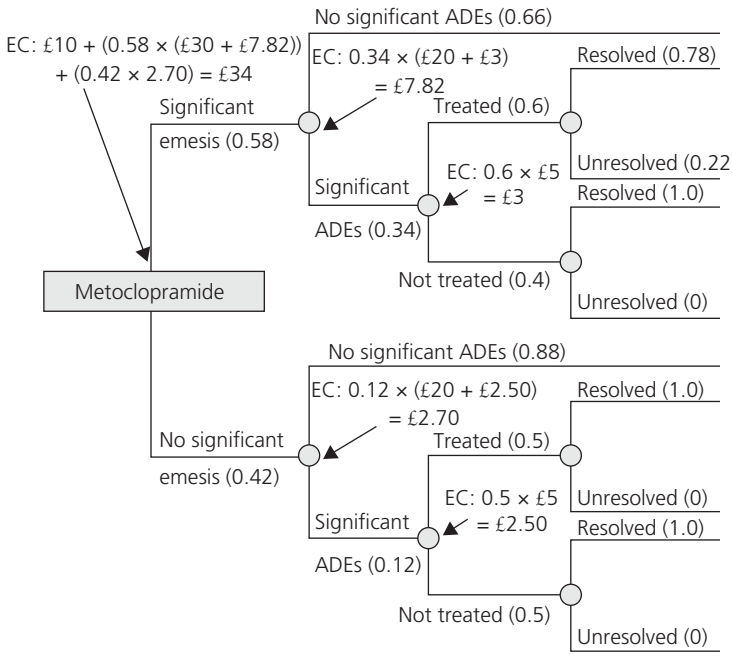


Fig. 9.5 The process of 'rolling back' a decision tree to calculate expected values. The example shows the calculation of expected cost for metoclopramide in the decision tree shown in Figure 9.4. ADE, adverse drug events; EC, expected costs.

Source: data from Buxton, M.J. and O'Brien, B.J., Economic evaluation of ondansetron: preliminary analysis using clinical trial data prior to price setting, *British Journal of Cancer*, Volume 66, Supplement XIX, pp. S64–S67, Copyright © 1992.

events. Given that, *ex ante*, it is not known whether a particular patient will experience a given event and follow a particular branch, the circular nodes (chance nodes) define points of uncertainty for an individual in the tree.

9.4.4.2 Probabilities

Branch probabilities The branches issuing from a chance node represent the possible events patients may experience at that point in the tree. The likelihood of the event is represented in terms of branch probabilities. For both treatments, the first chance node relates to whether or not a patient experiences an episode of emesis, and the probability of emesis and its complement (i.e. 1 – the probability of that event) are shown on the respective branches.

Conditional probabilities Moving from left to right, chance nodes show subsequent uncertain events. The probabilities of these events are *conditional probabilities* (see Box 9.3) because they can differ according to whether or not patients have experienced particular previous events. For example, for those patients who have experienced an episode of emesis on metoclopramide, the conditional probability of significant adverse events is 0.34; and the probability of treatment conditional on having experienced such an event is 0.6. It can be seen that, although the possible events are the same for the

two therapies, the probabilities in the two parts of the tree are not the same. Specifically, the efficacy of ondansetron was considered higher than that of metoclopramide, so there is a lower probability of emesis. Ondansetron was also considered less toxic, as the probabilities of adverse events (whether or not the drug had been efficacious) are lower than those for metoclopramide. There are also differences between the two therapies in the probabilities of an adverse event being treated and of it resolving.

9.4.4.3 Pathways

The combination of the different branches in the tree determines a series of pathways along which patients can pass in the tree. For both treatments, there are ten possible pathways. The top pathway for each treatment incorporates significant emesis and no adverse events; the second pathway contains significant emesis and a significant adverse event which is treated and resolves, and so on. The final pathway for each treatment relates to no significant emesis and an adverse event that is not treated and that does not resolve. These pathways are mutually exclusive (a given patient can only follow one of the pathways) and exhaustive (a given patient must follow one of the pathways).

Pathway probabilities To the right of the decision tree in Figure 9.4 is a series of columns of numbers. The first contains the probability of a given patient passing along each of the pathways. These probabilities are calculated by multiplying the initial branch probability by subsequent conditional probabilities. So the probability of the first pathway, with ondansetron, is the product of the probability of significant emesis (0.25) and the probability of no significant adverse events conditional on significant emesis (0.74), which equals 0.185. As the pathways are mutually exclusive, the probabilities for a given treatment must sum to 1.

Pathway costs Each pathway in the tree also has costs associated with it. These represent the sum of the costs of each of the events patients experience in that pathway. For the first pathway, for example, the relevant costs are the cost of the drug itself (£10) and the cost of significant emesis (£30), totalling £40. The second pathway cost is the sum of the drug cost (£10), significant emesis (£30), and significant adverse events (£20) which are treated (£5), equalling £65. The same principle is applied to the other pathways in the tree. It can be seen that the pathway costs are the same for metoclopramide and ondansetron as it is assumed that the two products have the same acquisition price and event costs.

9.4.4.4 Expected values

The expected cost for the two therapies can be calculated by weighting each pathway cost by its respective probability, and then summing across all the pathways. This can be seen in the expected cost column of Figure 9.4, and adding down this column generates expected costs for metoclopramide and ondansetron of £34 and £21, respectively.

This decision model used the probability of successful treatment as the relevant measure of effect in the CEA (i.e. the probability of no emesis and no adverse events). In terms of expected values, this is equivalent to giving the pathway of no emesis and no adverse events the value 1 and all other pathways the value 0. Assuming equal prices for the products, this version of the model indicates that ondansetron is dominant, as it has

a lower expected cost than metoclopramide and a higher expected effect. The original paper considered a number of sensitivity analyses involving alternative assumptions about the prices for the products (Buxton and O'Brien 1992). Another way of working out the expected costs and effectiveness for a given option in a decision tree is by 'rolling back' the tree. It will give exactly the same answer as the approach outlined above, but involves working from the right-hand side of the tree towards the left, calculating expected values at each chance node. Figure 9.5 shows how the decision tree for the alternative antiemetics is rolled back to estimate expected costs for metoclopramide. It could be argued that the use of the probability of successful treatment as the effectiveness measure is a weakness of this analysis as it assumes that all other pathways have the same (zero) value. Weighting each pathway using HRQoL and calculating expected QALYs for each treatment would probably have been a more informative analysis.

9.4.4.5 Limitations of the decision tree

The decision tree is widely used in economic evaluation, but has important limitations. The first is that events are implicitly taken as occurring over an instantaneous discrete period. In the antiemetic case study discussed above, for example, costs and effects over an undefined treatment period were considered. In other words, time is not explicitly defined in a decision tree unless the analyst does so in characterizing the different branches. Therefore, those elements of an economic evaluation that are time dependent can be difficult to implement. This is true of discounting, where the time at which costs and outcomes are accrued is very important. It also applies to the process of adjusting survival duration for HRQoL in calculating QALYs where it is necessary to know when a change in health status occurs (e.g. to reflect the impact of age on HRQoL).

The second, and related, limitation of decision trees is that they can become very complex when they are used to model complicated long-term prognoses, particularly related to chronic diseases. For example, to model the future prognosis of a woman with early-stage breast cancer, a decision tree would have to characterize a whole series of competing risks that a patient would have to face including adverse treatment effects, cancer recurrence (of various types), remission from cancer, and death (from various causes). Once an event is experienced in one time period (e.g. cancer recurrence), a series of new risks may present themselves for future time periods. In principle, these recurring event risks could be structured using a decision tree where a set of chance nodes and branches could be used to characterize events in a particular time period, and the same or similar ones could be used for subsequent time periods. However, for a long-term chronic disease, where a patient is at risk of events for many years, the tree could become very 'bushy', with many mutually exclusive pathways. A model of this type would probably be very time consuming to programme and analyse.

9.4.5 Implementing a model—the Markov model

The limitations of the decision tree are the main reason why another model structure—the Markov model—is also widely used in economic evaluation to handle particular decision problems (Briggs and Sculpher 1998; Sonnenberg and Beck 1993). Whereas decision trees characterize possible prognoses in terms of alternative branches, Markov models are based on a series of 'states' that a patient can occupy at a given point

in time. Time elapses explicitly with a Markov model, with the probability of a patient occupying a given state assessed over a series of discrete time periods, called cycles. The length of these cycles will depend on the disease and interventions being evaluated, but might be a month or a year. In judging the appropriate cycle length, a key consideration is to limit the probability that a given patient could experience more than one event the period of the cycle. Each state in the model generally has a cost associated with it and, when QALYs are used as the ultimate outcome measure, a HRQoL weight. The time duration during which the average patient occupies the various states in the model will, when weighted by the relevant cost or HRQoL weight, be used to calculate expected costs and outcomes. The speed with which patients move between the states in the model is determined by a set of transition probabilities.

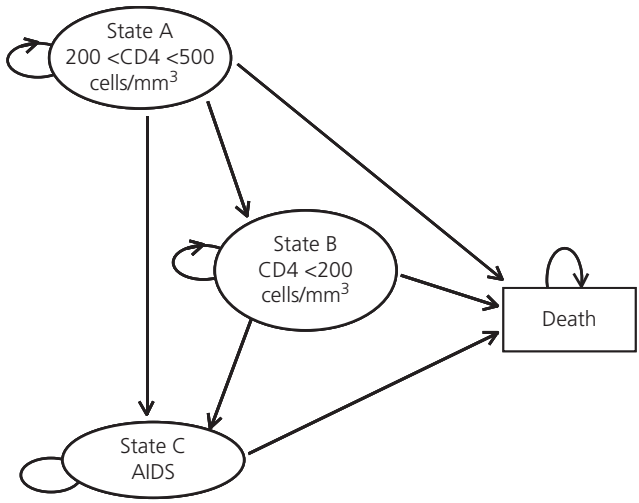
These concepts are described in more detail using an example from the published literature, which evaluated the cost-effectiveness of two antiretroviral therapies (zidovudine monotherapy versus zidovudine plus lamivudine combination therapy) for patients with HIV infection (Chancellor et al. 1997). Although more recent models for HIV are more complex, this study provides a good example to understand the key features of Markov models.

9.4.5.1 Markov states

Figure 9.6 shows a schematic of the Markov model used in the HIV example. The model is structured in terms of four Markov states. Two of these are related to a patient's CD4 count, which indicates the strength of their immune system. State A represents the healthiest patients with relatively high CD4 counts, and State B includes patients with lower CD4 counts. State C includes patients who have progressed to AIDS, and the patient moves to State D when they die. The arrows in the model show how patients can progress through the model over the cycles, which were taken to be 1 year. If a patient starts in State A in the first cycle, various transitions are possible in the second cycle: the patient can (1) remain in State A; (2) move to State B as their CD4 count drops; (3) move to State C if they suffer an AIDS-defining illness; or (4) move to State D if they die. Once a patient has moved to State B, in the next cycle they can remain in this state, or progress to State C or to State D. In this particular model, it is not possible for a patient's health to improve, so they cannot, for example, move from State B to State A. Once in State C, in the next cycle they can remain in that state or die, but not move back to States A or B. State D (death) is an absorbing state from which, sadly, there is no escape!

9.4.5.2 Transition probabilities

Figure 9.6 shows the transition probabilities that define the speed at which patients move between the Markov states under monotherapy, and the cycle length is 1 year. The matrix shows the state in which the patient starts the cycle, and the probabilities associated with the various transitions during one cycle conditional on a starting state (so these are conditional probabilities). For example, if a patient is in State A on monotherapy, there is a probability of 0.721 that they will remain in that state in the next cycle, of 0.202 that they will progress to State B, of 0.067 that they will progress to State C, and of 0.01 that they will die. The zeros in the matrix represent situations where backwards transitions are not considered feasible in this particular model. It can be seen that, because a patient always has to be in one of the states, the sum of the



Transition probabilities—monotherapy

	Transition to			
Transition from	State A	State B	State C	State D
State A	0.721	0.202	0.067	0.01
State B	0	0.581	0.407	0.012
State C	0	0	0.75	0.25
State D	0	0	0	1

Fig. 9.6 Markov diagram for a cost-effectiveness model in HIV taken from Chancellor et al. (1997). Below the diagram are the transition probabilities used for the monotherapy treatment.

probabilities across the lines must always equal 1. These 'baseline' probabilities, relating to what was then current practice, were taken from a longitudinal cohort study, which is a common type of evidence source for this type of parameter. Appropriate analytical methods are required to translate longitudinal data into transition probabilities over discrete cycles—see Chapter 10, also Briggs et al. (2006). A second set of transition probabilities was calculated for combination therapy in an attempt to reflect its effectiveness, compared to monotherapy, based on a meta-analysis of RCTs.

In this model the transition probabilities are the same for every cycle in the model. This implies, for example, that a patient with AIDS is at the same risk of death over the next year regardless of factors such as their age or the duration of time they have had AIDS. Markov models with fixed transition probabilities with respect to time are known as Markov chains. Some transition probabilities can also be allowed to vary over time depending on the structure of the model (Briggs et al. 2006).

9.4.5.3 Costs and outcomes

In the Markov model, costs are typically implemented each cycle according to the state a patient occupies (although they can also be applied to the proportion making a transition, e.g. to a 'dead' state). For the two therapies being evaluated in the HIV example, the cost of being in a given state is the same, and the only difference is in the acquisition price of the therapies. Hence, as for the decision tree example above, the only elements of the Markov model that differ between the two therapies are the acquisition cost of the therapies themselves and the probabilities that determine how a patient moves through the model. On the outcomes side, the HIV model used expected survival duration (life-years) as the measure of effectiveness, and this was evaluated over a lifetime time horizon (i.e. until the probability of being alive is very small).

9.4.5.4 Expected values

The process of calculating expected costs and effectiveness with a Markov model is very similar to that in a decision tree. This because both models are examples of *cohort models* which are set up to calculate the outcomes of interest in the *average patient*, and there is no consideration to how *individual patients* vary between each other. Instead of summing pathway costs and effects and weighting by their probabilities as with the decision tree (or rolling back the tree), the costs and values of each Markov state are weighted by the time a patient spends in that state. This is made up of two stages. In the first, the probability of a patient being in a given state for each cycle is calculated (this can also be understood as the proportion of the patient cohort in that state at a point in time). This is usually done in a spreadsheet or similar software using a method known as the cohort simulation method, which produces a 'Markov trace' showing the proportion of the cohort in each state over time.

This is illustrated in Figure 9.7 with respect to the monotherapy intervention in the HIV example, using a time horizon of 20 annual cycles. It is assumed that 1000 patients begin in the cohort, but the number is irrelevant in a cohort model as the focus is the average patient and only the proportions of the cohort in particular states at a given time point matter. One patient or one million patients could be used as the starting cohort, and the answer will be the same. For each cycle, the proportion of the cohort in each state is calculated on the basis of the proportions in the various states in the last cycle and the transition probabilities. Figure 9.7 shows the calculations for the first cycle. In a spreadsheet or other software, once the equations have been determined for the first cycle, it is normally a case of simply copying down the formulas for subsequent cycles. As more and more cycles are added, the proportion of the cohort in the absorbing state (here death) increases, and all but a very small proportion should have died once a cycle number that is consistent with the relevant life expectancy has been reached.

Once the proportion of patients (or, in other words, the probability of a given patient being) in each state for each cycle has been calculated, the second stage involves working out expected costs and effects. On the cost side, this involves calculating an expected cost per cycle by adding the cost of each state weighted by the proportion of the cohort in each state. The overall expected cost simply involves summing the expected cost of all cycles. Implementing discounting is straightforward, with the standard formula (see Chapter 4) used to adjust the expected cost of every individual cycle. Expected outcomes are calculated on a similar basis. In the case of survival duration, this simply involves

Cycle	State A	State B	State C	State D	Total
0	1000	0	0	0	1000
	1000 × 0.721	1000 × 0.202	1000 × 0.067	1000 × 0.01	
	↓	↓	↓	↓	
1	721	202	67	10	1000
2	520	263	181	36	1000
3	375	258	277	90	1000
4	270	226	338	166	1000
5	195	186	363	256	1000
6	140	147	361	351	1000
7	101	114	340	445	1000
8	73	87	308	532	1000
9	53	65	271	611	1000
10	38	48	234	680	1000
11	27	36	197	739	1000
12	20	26	164	789	1000
13	14	19	135	831	1000
14	10	14	110	865	1000
15	7	10	89	893	1000
16	5	7	72	916	1000
17	4	5	57	934	1000
18	3	4	45	948	1000
19	2	3	36	959	1000
20	1	2	28	968	1000

Fig. 9.7 The results of the Markov trace for the monotherapy group in the HIV example shown in Figure 9.6. The trace assumes a starting cohort of 1000 beginning in State A.

weighting the proportion of patients in each state per cycle by 1 if they are alive, and by 0 if they are dead. Adding up across the cycles (with discounting as necessary) will provide the expected number of life-years experienced by the cohort. In the case of QALYs, this is slightly different because the proportion of the cohort in each state is weighted by the HRQoL value associated with that state, and then summed across the cycles.

A cohort simulation is undertaken for each option being evaluated. In the case of the HIV model, cohort simulations were undertaken separately for monotherapy and combination therapy. On this basis, combination therapy was found to be more costly and more effective, with an incremental cost per life-year gained of £6276.

9.4.6 Implementing a model—other model types

This chapter has considered two popular model types used in economic evaluation. The Markov model is used in situations when the decision tree would become too unwieldy, typically when events can recur over a long time horizon. The HIV example in Section 9.4.5 used a Markov model in a conventional manner, where a cohort simulation is undertaken for each of the respective options being compared. The two interventions differed only in terms of the initial (acquisition) costs of the interventions and the

transition probabilities. In some situations, a decision analysis may involve the combination of both a decision tree and a Markov model. This was the case in the evaluation of glycoprotein IIb/IIIa antagonists example discussed in Section 9.2.5, for example (Palmer et al. 2005). A short-term decision tree was used to establish the proportion of patients with each therapy experiencing non-fatal myocardial infarction or death over an initial 6 month period reflecting trial results; a Markov model was used to calculate long-term expected costs and quality-adjusted survival duration conditional on which events had been experienced in the short-term model.

9.4.6.1 Time dependency and the memoryless property

A simple Markov model may be unsuitable to capture the key aspects of some prognoses. The HIV example described above, for example, assumed that transition probabilities did not vary over time. However, in some situations, this assumption may be difficult to sustain because of evidence suggesting that the probability increases or decreases with time. Some forms of 'time dependency' in transition probabilities can be handled quite easily in Markov models. This is the case, for instance, when the transition probability changes as the age of the patient increases. This can be implemented simply by having a different transition probability for each cycle in the cohort simulation. Equivalently, it is not difficult to implement transition probabilities that change as a function of the time a patient has been in a state, as long as all patients start in that state, and none returns to it once they have left it. The use of time-dependent transition probabilities in Markov models is dealt with in more detail in Briggs et al. (2006).

In other situations, it is less straightforward to implement time dependency in transition probabilities because of the key assumption underlying Markov models. This 'Markov assumption' is often described as the 'memoryless' feature of these models. It holds that the probability of a given transition in the model is independent of the nature or timing of earlier transitions. This can be illustrated using the HIV model shown in Figure 9.6. In this model, patients can enter the AIDS state (State C) from either State A or State B. However, once a patient has entered the AIDS state, the model cannot 'remember' where the patient came from; that is, it cannot distinguish the origin of the patients in the state at a given time point and treats them as homogenous. So, if the prognosis (cost and/or health impact) of being in the AIDS state was considered likely to differ according to the CD4 count a patient had at the point of entering that state, this particular Markov model would not be able to reflect this.

This assumption might be difficult to justify if evidence suggests, for example, that mortality risk is higher in patients who have experienced AIDS-defining events having previously had lower CD4 counts. If the memoryless feature represents an oversimplification of the epidemiological evidence, then one approach is to add additional states to the Markov model. In the example above, for example, two AIDS states could be used: one containing patients who moved there from State A and the other including patients who arrived from State B. These two AIDS states could then differ with respect to the risk of mortality and, if necessary, in terms of the cost per cycle of occupying that state.

Giving a Markov model additional 'memory' by adding states can, however, become unwieldy if numerous additional states have to be added. In this situation, one option is to move to a different modelling approach. The decision tree and Markov model are examples of cohort models. As shown above, this involves calculating the proportions of

a homogeneous cohort that would move along particular pathways or occupy specific Markov states. Using these proportions to weight the costs and outcomes associated with pathways or Markov states, this provides the route to calculating overall expected costs and outcomes for each of the options being compared.

9.4.6.2 Individual sampling models

The difficulty in incorporating 'memory' and time dependency are two of the limitations of cohort models. An alternative approach to decision modelling is to move away from the cohort model towards modelling individual patients moving through models. These *individual sampling models* (ISM) calculate the costs and effects of a large number of simulated patients and average across these patients to estimate expected values for the alternatives under evaluation. When ISM is used in the context of state transition models with discrete cycles, this is referred to as micro-simulation or first-order Monte-Carlo simulation (Siebert et al. 2012). This type of modelling literally tracks the process of individual simulated patients through particular states, and allows them to accumulate costs and benefit over time. They have the potential to offer greater flexibility than cohort models as the future prognosis of a given patient can vary according to their 'history'. In the case of the HIV example, the use of micro-simulation would mean that a patient who experiences an AIDS event with a CD4 count of 400 could have a different risk of future events than a patient who experienced such an event with a CD4 count of 200. Given the focus of economic evaluation on expected values, such a model has to simulate the costs and outcomes of a large number of patients and estimate the average over those simulations.

An alternative form of an ISM is *discrete event simulation* (DES) (Karnon et al. 2012). Whereas micro-simulation retains the concept of the state and discrete cycles as with a Markov model, DES simulates the time until the next event for a given simulated patient. The way time is handled in DES means that these models can advance to the next time a given simulated patient has an event, thus avoiding modelling time and effort in unnecessary interim computations. More detail about the use of ISM models in economic evaluation can be found elsewhere (Barton et al. 2004; Caro et al. 2010; Davies 1985; Standfield et al. 2014).

ISMs have some important limitations, however. The opportunity to incorporate patient history with such models may allow greater structural flexibility, but it will typically require additional evidence to populate such models. This is because parameters representing possible future prognoses for a given patient need to be conditional on history, thus increasing the number of parameters to be estimated. A second limitation is that the simulation requirements of these models can be time consuming, even with modern computers. This is particularly the case when probabilistic sensitivity analysis (PSA) is undertaken to quantify parameter uncertainty (see Chapter 11).

9.4.6.3 Dynamic transmission models

All of the types of models considered so far in this chapter assume that the individuals being modelled are independent from each other with respect to their health. That is, the health of one individual does not impact on the health of one or more others. This independence assumption may be untenable in the context of infectious disease where the incidence of new infections depends on the existing number of individuals who are infected. During an epidemic this number changes dynamically. Therefore, decision-analytic

models relating to infectious diseases may need to consider explicitly this dynamic feature of such diseases. Allowing for the interaction between individuals in the context of infectious disease may be necessary for interventions including vaccination, screening, and treatments where the transmissibility of the individual is affected. Many economic evaluations of these types of intervention for infectious disease often retain the standard 'static' decision model which does not allow for the non-linear interactions between individuals. For example, it has been found that most models used for the economic evaluation of screening for *Chlamydia trachomatis* used static methods and, therefore, were unlikely to have correctly estimated its cost-effectiveness (Roberts et al. 2006).

The importance of using dynamic transmission models for economic evaluation is now more fully understood. Useful introductions to the specifics of these models are available (Brisson and Edmunds 2003; Jit and Brisson 2011; Pitman 2014). There has also been a recent report from a task force on infectious disease modelling established by SMDM and ISPOR which, as well as providing another accessible introduction to the area, also offers a useful guide to good practice (Pitman et al. 2012). Further details of different model types can be found elsewhere (Barton et al. 2004; Brennan et al. 2006).

9.4.7 Selecting a model

Identifying an appropriate structure, and the type of model with which to implement it, is an extremely important stage of the decision modelling process. It is not possible to provide definitive guidelines for the selection of a particular model structure, as these have to depend on the overall objective of the economic evaluation as well as the nature of the disease process and impacts of the interventions. It has to be emphasized that all models are simplifications of reality, and the ultimate objective in selecting an appropriate structure for a decision model is to make the model no more complex than it has to be to address the policy questions appropriately. The value of a model is limited if it has been made highly complex in order to provide a more 'accurate' estimate of expected cost-effectiveness but its complexity precludes full uncertainty analysis.

9.5 Critical appraisal of decision-analytic models

Although decision-analytic modelling can provide a valuable framework for economic evaluation, its results are always conditional on the evidence used and on structural assumptions. In other words, there are good and bad decision models. As for any other form of evaluation, it is crucial that decision models are subject to careful critical review, and their results should not be used blindly in decision-making. There are a number of examples in the literature of papers that have discussed the characteristics of a 'good model' (Eddy et al. 2012; McCabe and Dixon 2000; Sculpher et al. 2000; Shemilt et al. 2013).

Methods guidelines in decision modelling were reviewed to compare and contrast their recommendations for good practice by Philips et al. (2004). There was a fair amount of consistency between the papers in their guidelines, but some areas of conflict. One example of this was the extent to which the availability of data should constrain the structure of a model as opposed to structure being determined based on the understanding of a condition and the effect of a treatment. The authors argued that, in principle, structure should not be influenced by the extent or quality of the data available to populate a model

but, in practice, this will not always be possible and more detailed guidance would be of value for analysts. Philips and colleagues went on to synthesize the available guidelines and, based on this, came up with a checklist to apply to specific decision models used in economic evaluation. It should be emphasized that this checklist relates to models and is not a substitute for those that relate to economic evaluation in general, including the one introduced in Chapter 3. This modelling checklist is reproduced in Annex 9.1.

9.6 Conclusions

Decision modelling is an important vehicle for economic evaluation, particularly where there is a specific resource allocation decision to be taken. The value of a formal analytic framework for decision-making is that it offers a means of synthesizing available evidence from a range of sources rather than relying on a *single study*, provides a way of relating the available evidence to the specific decision problem being posed, provides a framework within which the limitations of a single clinical study *as a vehicle for economic evaluation* can be addressed, helps decision-makers identify optimal interventions under conditions of uncertainty, and can contribute to the process of setting research priorities.

This chapter has provided an introduction to these methods, and further reading is available (Briggs et al. 2006). There remain important methods questions to address in decision modelling. These include how to develop efficient methods to identify evidence relating to all parameters in decision models and not just those relating to treatment effects, how to synthesize all available evidence in models and reflect the uncertainty and correlation in these data, and how to deal with uncertainty in the structure of decision models and reflect this in the value of information analysis. These issues are dealt with in Chapters 10 and 11. Although decision modelling has the potential to provide a powerful input for decision-making, there are good and bad applications of these methods, and critical appraisal is essential.

9.7 Exercise: developing a decision-analytic model

9.7.1 Background

Imagine that you have been asked to advise local decision-makers on the cost-effectiveness of antenatal HIV testing (i.e. testing pregnant women for HIV infection). You undertake a literature search to identify published economic evaluations, but find nothing to help you in your analysis. You quickly realize that you will have to undertake a decision analysis of your own using evidence from available sources.

9.7.2 The evidence

From a literature search you identify publications that provide you with the following information:

- If a woman has HIV and her infection is not known during pregnancy, the probability that she will transmit the infection to her child is 26%.
- If a woman's infection is known during pregnancy, however, it is possible to use risk-reduction interventions such as caesarean section, antiretroviral therapy, and bottle-feeding. These interventions cost £800 more than a normal delivery and

reduce the probability of vertical transmission to 7%, but only 95% of infected women accept them.

- Discussion with midwifery staff indicates that offering the test to women could be achieved at negligible additional cost, but your pathology laboratories suggest that each blood test will cost £10; they also indicate that the tests are 100% accurate (i.e. there are no false negatives or false positives).
- A published paper suggests that the prevalence of previously undetected HIV in the antenatal population in your area is 5%.

9.7.3 Assumptions

Discussions with professional staff indicate that the following assumptions can be justified:

- No woman will select to terminate on discovering she has HIV infection.
- All women who are tested positive will be offered risk-reduction interventions.

9.7.4 The task

1 Your task is the following:
   a To structure a decision tree characterizing the decision regarding whether or not to offer antenatal HIV testing.
   b To calculate the expected cost per true positive case detected.
2 What are likely to be the key sensitivity analyses to undertake?
3 What are the weaknesses of the analysis?

9.7.5 Solutions

1a The decision tree is shown in Figure 9.8.

1b Expected cost of testing

$= (810 \times 0.0033) + (810 \times 0.0441) + (10 \times 0.0007) + (10 \times 0.0019) + (10 \times 0.95)$

$=$ £47.92.

Probability of vertical transmission with testing $= 0.0033 + 0.0007 = 0.004$.

Expected cost of no testing $= 0$.

Probability of vertical transmission $= 0.013$.

Additional expected cost per HIV-infected child avoided:

- additional cost $=$ £47.92
- reduced vertical transmission $= 0.013 - 0.004 = 0.009$
- additional cost per HIV-infected birth avoided $= 47.92/0.009 =$ £5324.

2 The following sensitivity analyses would be warranted:
   - Due to parameter uncertainty:
     - probabilities of transmission
     - probability of acceptance of interventions
     - prevalence.

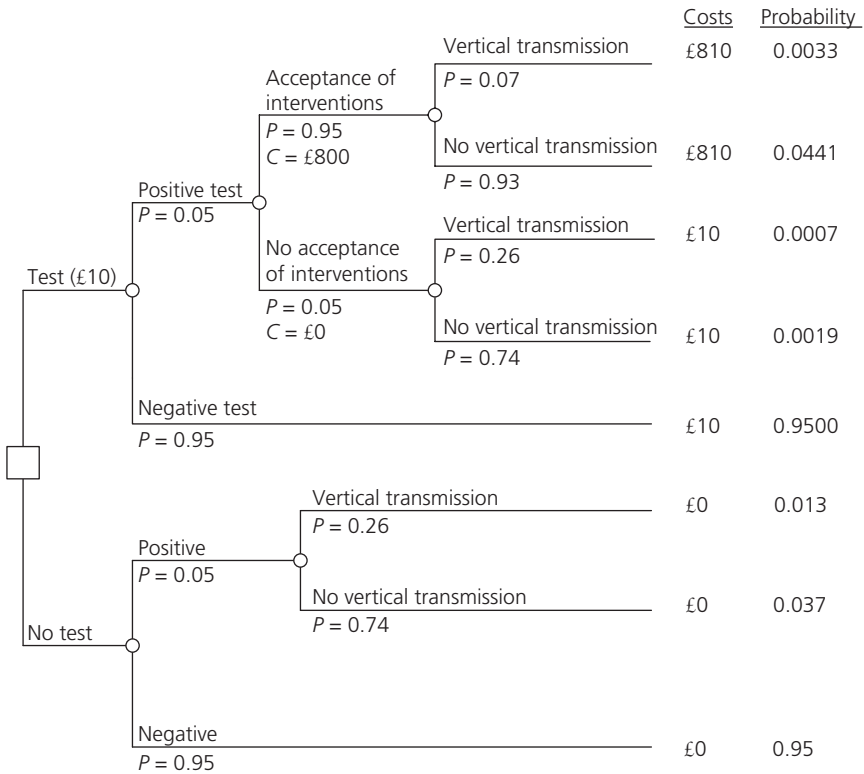


Fig. 9.8 Decision tree for antenatal HIV testing.

- Due to heterogeneity:
  - costs
  - probability of acceptance of interventions
  - prevalence.
- Due to structural uncertainty:
  - accuracy of test (need to restructure tree)
  - termination rate: in this model, termination may increase or decrease costs; there would be a difficulty in how this is dealt with on the outcomes side
  - uptake of test will not affect cost-effectiveness.

3 Weaknesses of the analysis:

- In reality, the relevant options to compare would be more likely to be universal testing versus high-risk group testing versus on-demand testing, rather than testing versus not.
- A full PSA would ideally be undertaken for to assess parameter uncertainty.
- The scope of analysis is limited:
  - It should have a longer-term model to include an assessment of the costs and (quality-adjusted) life-years conditional on HIV transmission.

- There should be a consideration of lifetime costs and outcomes.
- The model should consider the effect of testing on the women themselves in terms of the costs and outcomes of earlier treatment than would be expected without testing.
- Possible effects on horizontal transmission might be considered in a wider scope.

References

Ades, A.E., Sculpher, M.J., Gibb, D.M., et al. (1999). Cost effectiveness analysis of antenatal HIV screening in United Kingdom. *BMJ*, **319**, 1230–4.

Barton, P., Bryan, S., and Robinson, S. (2004). Modelling in the economic evaluation of health care: selecting the appropriate approach. *Journal of Health Services Research and Policy*, **9**, 110–18.

Basu, A. and Meltzer, D. (2007). Value of information on preference heterogeneity and individualized care. *Medical Decision Making*, **27**, 112–27.

Bojke, L., Epstein, D., Craig, D., et al. (2011). Modelling the cost-effectiveness of biologic treatments for psoriatic arthritis. *Rheumatology*, **50**(suppl 4), iv39–iv47.

Bravo Vergel, Y., Palmer, S., Asseburg, C., et al. (2007). Results of a comprehensive decision analysis. Is primary angioplasty cost effective in the UK? *Heart*, **93**, 1238–43.

Brennan, A., Chick, S.E., and Davies, R. (2006). A taxonomy of model structures for economic evaluation of health technologies. *Health Economics*, **15**, 1295–310.

Briggs, A. and Sculpher, M.J. (1998). An introduction to Markov modelling for economic evaluation. *PharmacoEconomics*, **13**, 397–409.

Briggs, A., Sculpher, M., Dawson, J., et al. (2004). The use of probabilistic decision models in technology assessment: the case of hip replacement. *Applied Health Economics and Policy*, **3**, 79–89.

Briggs, A., Claxton, K., and Sculpher, M. (2006). *Decision modelling for health economic evaluation*. Oxford: Oxford University Press.

Brisson, M. and Edmunds, W.J. (2003). Economic evaluation of vaccination programs: the impact of herd-immunity. *Medical Decision Making*, **23**, 76–82.

Buxton, M.J. and O'Brien, B.J. (1992). Economic evaluation of ondansetron: preliminary analysis using clinical trial data prior to price setting. *British Journal of Cancer*, **66**, S64–S67.

Caro, J.J., Moller, J., and Getsios, D. (2010). Discrete event simulation: the preferred technique for health economic evaluations? *Value in Health*, **13**, 1056–60.

Chancellor, J.V., Hill, A.M., Sabin, C.A., et al. (1997). Modelling the cost effectiveness of lamivudine/zidovudine combination therapy in HIV infection. *PharmacoEconomics*, **12**, 1–13.

Claxton, K., Sculpher, M.J., and Drummond, M.F. (2002). A rational framework for decision making by the National Institute for Clinical Excellence. *Lancet*, **360**, 711–15.

Colbourn, T., Asseburg, C., Bojke, L., et al. (2007). Prenatal screening and treatment strategies to prevent group B streptococcal and other bacterial infections in early infancy: cost-effectiveness and expected value of information analyses. *Health Technology Assessment*, **11**(29), 1–240.

Coyle, D., Buxton, M.J., and O'Brien, B.J. (2003). Stratified cost-effectiveness analysis: a framework for establishing efficient limited use criteria. *Health Economics*, **12**, 421–7.

Davies, R. (1985). An assessment of models of a health system. *Journal of the Operational Research Society*, **36**, 679–87.

Drummond, M.F., Barbieri, M., Cook, J., et al. (2009). Transferability of economic evaluations across jurisdictions: ISPOR Good Research Practices Task Force Report. *Value in Health*, **12**, 409–18.

Eddy, D.M., Hollingworth, W., Caro, J.J., et al. (2012). Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Value in Health*, **15**, 843–50.

Epstein, D.M., Sculpher, M.J., Manca, A., et al. (2008). Modelling the long-term cost-effectiveness of endovascular or open repair for abdominal aortic aneurysm. *British Journal of Surgery*, **95**, 183–90.

Griffiths, A., Paracha, N., Davies, A., et al. (2014). The cost effectiveness of ivabradine in the treatment of chronic heart failure from the UK National Health Service perspective. *Heart*, **100**, 1031–6.

Henriksson, M., Epstein, D.M., Palmer, S.J., et al. (2008). The cost-effectiveness of an early interventional strategy in non-ST-elevation acute coronary syndrome based on the RITA 3 trial. *Heart*, **94**, 717–23.

Hunink, M., Weinstein, M.C., Wittenberg, E., et al. (2014). *Decision making in health and medicine: integrating evidence and values*, 2nd edition. Cambridge: Cambridge University Press.

Jit, M. and Brisson, M. (2011). Modelling the epidemiology of infectious diseases for decision analysis. A primer. *PharmacoEconomics*, **29**, 371–86.

Karnon, J., Stahl, J., Brennan, A., et al. (2012). Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. *Value in Health*, **15**, 821–7.

Kobelt, G., Jonsson, B., Young, A., et al. (2003). The cost-effectiveness of infliximab (Remicade) in the treatment of rheumatoid arthritis in Sweden and the United Kingdom based on the ATTRACT study. *Rheumatology*, **42**, 326–35.

Mark, D.B., Hlatky, M.A., Califf, R.M., et al. (1995). Cost effectiveness of thrombolytic therapy with tissue plasminogen activator as compared with streptokinase for acute myocardial infarction. *New England Journal of Medicine*, **332**, 1418–24.

McCabe, C. and Dixon, S. (2000). Testing the validity of cost-effectiveness models. *PharmacoEconomics*, **17**, 501–13.

Neumann, P.J., Hermann, R.C., and Kuntz, K.M. (1999). Cost-effectiveness of donepezil in the treatment of mild or moderate Alzheimer's disease. *Neurology*, **52**, 1138–45.

Owens, D.K. and Nease, R.F. (1997). A normative analytic framework for development of practice guidelines for specific clinical populations. *Medical Decision Making*, **17**, 409–26.

Palmer, S., Sculpher, M., Philips, Z., et al. (2005). Management of non-ST-elevation acute coronary syndromes: how cost-effective are glycoprotein IIb/IIIa antagonists in the UK National Health Service? *International Journal of Cardiology*, **100**, 229–40.

Philips, Z., Ginnelly, L., Sculpher, M., et al. (2004). A review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technology Assessment*, **8**(36), 1–158.

Pitman, R. (2014). Infectious disease modelling, in A. J. Culyer (ed.), *Encyclopedia of health economics*. Amsterdam: Elsevier.

Pitman, R., Fisman, D., Zaric, G.S., et al. (2012). Dynamic transmission modeling: A report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-5. *Medical Decision Making*, **32**, 712–21.

Raiffa, H. (1968). *Decision analysis: introductory lectures on choices under uncertainty*. Reading, MA: Addison-Wesley.

Roberts, M., Russell, L.B., Paltiel, A.D., et al. (2012). Conceptualizing a model: A report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. *Value in Health*, **15**, 804–11.

Roberts, T.E., Robinson, S., Barton, P., et al. (2006). Screening for Chlamydia trachomatis: a systematic review of the economic evaluations and modelling. *Sexually Transmitted Infections*, **82**, 193–200.

Sackett, D.L., Rosenberg, W.M.C., Gray, J.A.M., Haynes, R.B., and Richardson, W.S. (1996). Evidence-based medicine: what it is and what it isn't. *BMJ*, **312**, 71–2.

Schulman, K.A., Glick, H.A., Rubin, H., and Eisenberg, J.M. (1991). Cost-effectiveness of HA-lA monoclonal antibody for gram-negative sepsis. *JAMA*, **266**, 3466–71.

Sculpher, M.J. (2008). Subgroups and heterogeneity in cost-effectiveness analysis. *PharmacoEconomics*, **26**, 799–806.

Sculpher, M.J. and Gafni, A. (2001). Recognizing diversity in public preferences: the use of preference sub-groups in cost-effectiveness analysis. *Health Economics*, **10**, 317–24.

Sculpher, M., Fenwick, E., and Claxton, K. (2000). Assessing quality in decision-analytic cost-effectiveness models. A suggested framework and example of application. *PharmacoEconomics*, **17**, 461–77.

Sculpher, M.J., Pang, F.S., Manca, A., et al. (2004). Generalisability in economic evaluation studies in health care: a review and case studies. *Health Technology Assessment*, **8**(49), 1–213.

Sculpher, M.J., Claxton, K.P., Drummond, M.F., et al. (2006). Whither trial-based economic evaluation for health care decision making? *Health Economics*, **15**, 677–87.

Shemilt, I., Wilson, E., and Vale, L. (2013). Quality assessment in modeling in decision-analytic models for economic evaluation, in A.J. Culyer (ed.), *Encyclopedia of health economics*. Amsterdam: Elsevier.

Siebert, U., Alagoz, O., Bayoumi, A.M., et al. (2012). State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-3. *Value in Health*, **15**, 812–20.

Sonnenberg, F.A. and Beck, J.R. (1993). Markov models in medical decision making. *Medical Decision Making*, **13**, 322–38.

Spiegelhalter, D.J., Abrams, K.R., and Myles, J.P. (2004). *Bayesian approaches to clinical trials and health-care evaluation*. Chichester: Wiley.

Spiegelhalter, D.J. and Best, N.G. (2003). Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Statistics in Medicine*, **22**, 3687–709.

Standfield, L., Comans, T., and Scuffham, P. (2014). Markov modelling and discrete event simulation in health care: a systematic comparison. *International Journal of Technology Assessment in Health Care*, **30**(2), 1–8.

Walker, S., Girardin, F., McKenna, C., et al. (2013). Cost-effectiveness of cardiovascular magnetic resonance in the diagnosis of coronary heart disease: an economic evaluation using data from the CE-MARC study. *Heart*, **99**, 873–81.

Weinstein, M.C. and Fineberg, H.V. (1980). *Clinical decision analysis*. Philadelphia: W.B. Saunders.

Wilby, J., Kainth, K., Hawkins, N., et al. (2005). A rapid and systematic review of the clinical effectiveness, tolerability and cost effectiveness of newer drugs for epilepsy in adults. *Health Technology Assessment*, **9**(15), 1–157.

Woolacott, N., Hawkins, N., Mason, A., et al. (2006). Efalizumab and etanercept for the treatment of psoriasis. A systematic review. *Health Technology Assessment*, **10**(46), 1–258.

Annex 9.1 Checklist for assessing quality in decision-analytic models

Table A9.1, from Philips et al. (2004), provides a suggested checklist for assessing quality in decision-analytic models.

Table A9.1 A suggested checklist for assessing quality in decision-analytic models

Dimension of quality		Attributes of good practice	Questions for critical appraisal
Structure			
S1	Statement of decision problem/objective	There should be a clear statement of the decision problem prompting the analysis. The objective of the evaluation and of the model should be defined. The primary decision-maker should be stated clearly.	Is there a clear statement of the decision problem? Is the objective of the evaluation and model specified and consistent with the stated decision problem? Is the primary decision-maker specified?
S2	Statement of scope / perspective	The perspective of the model (relevant costs and consequences) should be stated clearly, and the model inputs should be consistent with the stated perspective and overall objective of the model. The scope of the decision model should be specified and justified. The outcomes of the model should reflect the perspective and scope of the model and should be consistent with the objective of the evaluation.	Is the perspective of the model stated clearly? Are the model inputs consistent with the stated perspective? Has the scope of the model been stated and justified? Are the outcomes of the model consistent with the perspective, scope, and overall objective of the model?

(*continued*)

Table A9.1 (continued) A suggested checklist for assessing quality in decision-analytic models

Dimension of quality		Attributes of good practice	Questions for critical appraisal
S3	Rationale for structure	The structure of the model should be consistent with a coherent theory of the health condition under evaluation and the treatment pathways (disease states or branches) should be chosen to reflect the underlying biological process of the disease in question and the impact of the intervention. The structure should not be dictated by current patterns of service provision. All sources of evidence used to develop and inform the structure of the model (i.e. the theory of disease) should be described. The structure should be consistent with this evidence.	Is the structure of the model consistent with a coherent theory of the health condition under evaluation? Are the sources of data used to develop the structure of the model specified? Are the causal relationships described by the model structure justified appropriately?
S4	Structural assumptions	All structural assumptions should be transparent and justified. They should be reasonable in the light of the needs and purposes of the decision-maker.	Are the structural assumptions transparent and justified? Are the structural assumptions reasonable given the overall objective, perspective, and scope of the model?
S5	Strategies/comparators	There should be a clear definition of the options under evaluation. All feasible and practical options relating to the stated decision problem should be evaluated. Options should not be constrained by the immediate concerns of the decision-maker, or data availability, nor limited to current clinical practice.	Is there a clear definition of the options under evaluation? Have all feasible and practical options been evaluated? Is there justification for the exclusion of feasible options?

Table A9.1 (continued) A suggested checklist for assessing quality in decision-analytic models

Dimension of quality		Attributes of good practice	Questions for critical appraisal
S6	Model type	The appropriate model type will be dictated by the stated decision problem and the choices made regarding the causal relationships within the model.	Is the chosen model type appropriate given the decision problem and specified causal relationships within the model?
S7	Time horizon	A model's time horizon should extend far enough into the future in order for it to reflect important differences between options. It is important to distinguish between the time horizon of the model, the duration of treatment and the duration of treatment effect.	Is the time horizon of the model sufficient to reflect all important differences between options? Are the time horizon of the model, the duration of treatment, and the duration of treatment effect described and justified?
S8	Disease states/pathways	Disease states/pathways should reflect the underlying biological process of the disease in question and the impact of interventions.	Do the disease states (state transition model) or the pathways (decision tree model) reflect the underlying biological process of the disease in question and the impact of interventions?
S9	Cycle length	For discrete time models, the cycle length should be dictated by the natural history of disease. It should be the minimum interval over which the pathology or symptoms are expected to alter.	Is the cycle length defined and justified in terms of the natural history of disease?

(*continued*)

Table A9.1 (continued) A suggested checklist for assessing quality in decision-analytic models

Dimension of quality		Attributes of good practice	Questions for critical appraisal
Data			
D1	Data identification	Methods for identifying data should be transparent and it should be clear that the data identified are appropriate given the objectives of the model. There should be justification of any choices that have been made about which specific data inputs are included in a model. It should be clear that particular attention has been paid to identifying data for those parameters to which the results of the model are particularly sensitive. Where expert opinion has been used to estimate particular parameters, sources and methods of elicitation should be described.	Are the data identification methods transparent and appropriate given the objectives of the model? Where choices have been made between data sources, are these justified appropriately? Has particular attention been paid to identifying data for the important parameters in the model? Has the quality of the data been assessed appropriately? Where expert opinion has been used, are the methods described and justified?
D2	Data modelling	All data modelling methodology should be described and based on justifiable statistical and epidemiological methods. Specific issues to consider include those below.	Is the data modelling methodology based on justifiable statistical and epidemiological techniques?
D2a	Baseline data	Baseline probabilities may be based on natural history data derived from epidemiological/observational studies or relate to the control group of an experimental study. Rates and interval probabilities should be transformed into transition probabilities appropriately. If there is evidence that time is an important factor in the calculation of transition probabilities in state transition models, this should be incorporated. If a half-cycle correction has not been used on all transitions in state transition model (costs and outcomes), this should be justified.	Is the choice of baseline data described and justified? Are transition probabilities calculated appropriately? Has a half-cycle correction been applied to both cost and outcome? If not, has this omission been justified?

Table A9.1 (continued) A suggested checklist for assessing quality in decision-analytic models

Dimension of quality		Attributes of good practice	Questions for critical appraisal
D2b	Treatment effects	Relative treatment effects derived from trial data should be synthesized using recognized meta-analytic techniques. The methods and assumptions that are used to extrapolate short-term results to final outcomes should be documented and justified. This should include justification of the choice of survival function (e.g. exponential or Weibull forms). Alternative assumptions should be explored through sensitivity analysis. Assumptions regarding the continuing effect of treatment once treatment is complete should be documented and justified. If evidence regarding the long-term effect of treatment is lacking, alternative assumptions should be explored through sensitivity analysis.	If relative treatment effects have been derived from trial data, have they been synthesized using appropriate techniques? Have the methods and assumptions used to extrapolate short-term results to final outcomes been documented and justified? Have alternative assumptions been explored through sensitivity analysis? Have assumptions regarding the continuing effect of treatment once treatment is complete been documented and justified? Have alternative assumptions been explored through sensitivity analysis?
D2c	Costs	Costing and discounting methods should accord with standard guidelines for economic evaluation.	Are the costs incorporated into the model justified? Has the source for all costs been described? Have discount rates been described andjustified given the target decision-maker?
D2d	Quality of life weights (utilities)	Utilities incorporated into the model should be appropriate for the specified decision problem.	Are the utilities incorporated into the model appropriate? Is the source for the utility weights referenced? Are the methods of derivation for the utility weights justified?

(*continued*)

Table A9.1 (continued) A suggested checklist for assessing quality in decision-analytic models

Dimension of quality		Attributes of good practice	Questions for critical appraisal
D3	Data incorporation	All data incorporated into the model should be described and the sources of all data should be given and reported in sufficient detail to allow the reader to be aware of the type of data that have been incorporated. Where data are not mutually consistent in the model, the choices and assumptions that have been made should be explicit and justified. The process of data incorporation should be transparent. It should be clear whether data are incorporated as a point estimate or as a distribution. If data have been incorporated as distributions as part of probabilistic analysis, the choice of distribution and its parameters should be described and justified.	Have all data incorporated into the model been described and referenced in sufficient detail? Has the use of mutually inconsistent data been justified (i.e. are assumptions and choices appropriate)? Is the process of data incorporation transparent? If data have been incorporated as distributions, has the choice of distribution for each parameter been described and justified? If data have been incorporated as distributions, is it clear that second-order uncertainty is reflected?
D4	Assessment of uncertainty	In assessing uncertainty, modellers should distinguish between the four principal types of uncertainty.	Have the four principal types of uncertainty been addressed? If not, has the omission of particular forms of uncertainty been justified?
D4a	Methodological	Methodological uncertainty relates to whether particular analytic steps taken in the analysis are the most appropriate.	Have methodological uncertainties been addressed by running alternative versions of the model with different methodological assumptions?
D4b	Structural	There should be evidence that structural uncertainties have been evaluated using sensitivity analysis.	Is there evidence that structural uncertainties have been addressed via sensitivity analysis?
D4c	Heterogeneity	It is important to distinguish between uncertainty resulting from the process of sampling from a population and variability due to heterogeneity (i.e. systematic differences between patient subgroups).	Has heterogeneity been dealt with by running the model separately for different subgroups?

Table A9.1 (continued) A suggested checklist for assessing quality in decision-analytic models

Dimension of quality		Attributes of good practice	Questions for critical appraisal
D4d	Parameter	Where data have been incorporated into the model as point estimates, the ranges used for sensitivity analysis should be stated and justified. Probabilistic analysis is the most appropriate method of handling parameter uncertainty because it facilitates assessment of the joint effect of uncertainty over all parameters (see data incorporation).	Are the methods of assessment of parameter uncertainty appropriate? If data are incorporated as point estimates, are the ranges used for sensitivity analysis stated clearly and justified?
Consistency			
C1	Internal consistency	There should be evidence that the internal consistency of the model has been evaluated in terms of its mathematical logic.	Is there evidence that the mathematical logic of the model has been tested thoroughly before use?
C2	External consistency	The results of a model should be explicable. Results should either make intuitive sense or counter-intuitive results should be fully explained. All relevant available data should be incorporated into a model. Data should not be withheld for purposes of assessing external consistency. The results of a model should be compared with those of previous models and any differences should be explained.	Are any counter-intuitive results from the model explained and justified? If the model has been calibrated against independent data, have any differences been explained and justified? Have the results of the model been compared with those of previous models and any differences in results explained?