# 9

# Cost-Effectiveness, Policy-Making and Regulation

## 9.1 INTRODUCTION

In this chapter we go beyond making inferences based on single or multiple studies in order to focus on the consequences of adopting particular health interventions. This broader perspective reflects the increasing attention given to the cost-effectiveness of new and existing treatments, leading to the development of technology-appraisal agencies, such as the National Institute of Clinical Excellence (NICE) in the UK, which are intended to give guidance to health providers and decide on treatments to be covered under relevant reimbursement schemes. We need, however, to take careful account of the context of the evaluation, particularly with regard to specification of prior distributions and loss functions, and a framework is outlined in Section 9.2.

As is clear from the name, cost-effectiveness analysis requires a focus on the dual outcomes of *costs* and *effectiveness*, and a typical formulation requires specification of a model for both, which will contain parameters whose plausible values will depend on both judgement and evidence. The 'standard' approach to cost-effectiveness analysis is outlined in Section 9.3, in which the value of concepts such as *incremental net benefit* and the *cost-effectiveness plane* are emphasised. In many circumstances randomised trial evidence may be lacking or limited to certain aspects of the model, leading naturally to the use of the generalised evidence synthesis techniques outlined in Chapter 8.4. In Section 9.4 we identify two alternative approaches to combining evidence synthesis with a cost-effectiveness model. The first approach is termed *two-stage*: in the first stage the evidence from multiple sources is synthesised and used as a basis for the distributions given to parameters; in the second stage, the effects of the

*Bayesian Approaches to Clinical Trials and Health-Care Evaluation* D. J. Spiegelhalter, K. R. Abrams and J. P. Myles
 ISBN: 0-471-49975-7

resulting uncertainty are propagated through the cost-effectiveness model. The second stage, in which distributions are placed on unknown parameters, has become known in the cost-effectiveness literature as *probabilistic sensitivity analysis*. The second, *integrated*, approach simultaneously carries out the synthesis and cost-effectiveness analysis. The two-stage approach is illustrated in Section 9.5, in which *cost-effectiveness acceptability curves* are introduced and shown to be easily handled in the Bayesian framework, illustrated using closed-form, Monte Carlo and MCMC approaches. The integrated approach is then demonstrated in Section 9.6.

In view of the potential complexity of the resulting models and analysis it is important that there is a clear description of the different components of uncertainty, and in Section 9.7 a taxonomy is provided. This is applicable to complex cost-effectiveness models, typically discrete-state, discrete-time Markov models, which are commonly used to make predictions of the longer-term consequences of a particular intervention. Section 9.8 describes their structure and the use of simulation methods both for micro-simulation of individual cases and probabilistic sensitivity analysis.

Since this chapter emphasises decisions as well as inferences, a strict decision-theoretic approach may be appropriate (see Sections 3.14 and 6.2). For example, Luce and Claxton (1999) point out that hypothesis testing is of limited relevance in economic studies, and when a cost-effectiveness analysis is being used as one of the inputs into a formal decision concerning drug regulation or health policy, they recommend a full decision-theoretic approach in which an explicit loss function of the decision-maker is assessed. Such a loss function can also be used as a basis for valuing the expected benefit from further evidence, and this *expected value of information* approach to deciding research priorities is discussed in Section 9.10; a brief critique of this approach is contained in Section 9.11. Finally, we briefly consider the role of regulatory authorities and the particular issues that arise in relation to Bayesian analysis (Section 9.12).

The combined literature on these topics is becoming large and only selected references will be provided: Briggs (2000) introduces many of these issues in a non-technical style, and we make extensive use of Spiegelhalter and Best (2003) although with some changes in notation. We also note a special issue on Bayesian methods of the *International Journal of Health Technology Assessment in Health Care* which features many relevant articles (Luce *et al.*, 2001), and the primer by O'Hagan and Luce (2003).

## 9.2 CONTEXTS

Throughout this book we have emphasised that it is vital to take into account the *context* in which a clinical trial is being either designed or analysed and interpreted, and more generally when evaluating any health-care intervention. The appropriate prior opinions, and the possibility of explicit loss functions,

depend crucially on whose behalf any analysis is being reported or a decision is being made.

This becomes particularly important when considering the 'end stage' of an evaluation – predicting the effects of actually getting the intervention into practice. We can address this issue using the broad categories of *stakeholders* introduced in Section 3.1:

- *Sponsors*, e.g. pharmaceutical industry, medical charities or granting agencies. In deciding whether to fund studies, they will be concerned with the potential 'payback' from research (Section 9.10), which in industry takes the form of a portfolio of drug development programmes. For such 'internal' analyses it will be quite reasonable for prior distributions to be based on subjective judgements and for loss functions to be based, in industry, on profitability. Very different considerations apply for 'external' analyses done on behalf of others – see below.
- *Investigators*, *i.e.* those responsible for the conduct of a study, whether funded by industry or publicly. In previous chapters we have focused primarily on those carrying out a single study, whose main concern is with the accuracy of the inferences to be drawn from their work, although again they may carry out a cost-effectiveness analysis on behalf of others.
- *Reviewers*, e.g. regulatory bodies (Section 9.12). They will be concerned with the appropriateness of the inferences drawn from the studies, and so may adopt their own prior opinions and reporting standards (Section 3.21). Regulatory bodies will generally only be concerned with safety and efficacy issues, and cost-effectiveness analyses will be dealt with by health-policy agencies.
- *Policy-makers*, e.g. agencies or clinicians setting health policy. Health-care organisations may be concerned with the cost-effectiveness of an intervention, although the sponsor or investigator may carry out this analysis on their behalf. Any analysis is likely to be open to external scrutiny, and hence any prior distributions used at this stage would need to be evidence-based or subject to careful justification and sensitivity analysis. Values would be societally based such as quality measures based on surveys, and future costs and benefits may be discounted according to accepted criteria.
- *Consumers*, e.g. individual patients or clinicians acting on their behalf. These would ideally demand individualised prognostic predictions under available alternative interventions, which could be combined with the patient's own utility function. We shall not deal with such individualised decision-making here, although it has been recommended that clinical trial results are presented in such a form as to help such judgements to be made (Simes, 1986).

There is a large literature on the appropriate means of dealing with values, whether concerning utility measures, quality adjustments, discount rates for costs and benefits, and so on, but these important issues are beyond the scope of this book. See Claxton *et al.* (2000) for a brief overview from a health-economic

perspective, including a contrast between the perspective of health-policy agencies and the wider society in general.

## 9.3 'STANDARD' COST-EFFECTIVENESS ANALYSIS WITHOUT UNCERTAINTY

Cost-effectiveness analyses aim to combine information regarding both clinical effectiveness and economic costs. Given known mean economic costs $m_{c1}$ and $m_{c2}$ under two different treatment options T1 and T2, and similar estimates of mean clinical effectiveness, $m_{e1}$ and $m_{e2}$, define $\theta_c = m_{c2} - m_{c1}$, $\theta_e = m_{e2} - m_{e1}$ as the incremental mean costs and effectiveness. Then the *incremental cost-effectiveness ratio* (ICER) is defined by

$$\text{ICER} = \frac{\theta_c}{\theta_e} = \frac{m_{c2} - m_{c1}}{m_{e2} - m_{e1}}. \tag{9.1}$$

The ICER can be considered as the cost per unit increase in effectiveness by adopting treatment option T2 rather than T1.

Until recently almost all cost-effectiveness analyses reported findings in terms of the ICER. Nevertheless, whilst the ICER appears appealing, difficulties arise in both the calculation of confidence intervals and its interpretation when the denominator is negative or zero. Figure 9.1 (O'Hagan *et al.*, 2000) shows a *cost-effectiveness plane* divided into four quadrants corresponding to different signs of $\theta_c$ and $\theta_e$, with the line $\theta_c = K\theta_e$ drawn, where $K$ represents a maximum acceptable cost per unit of effectiveness; we shall discuss the specification of $K$ at the end of this section.

A conceptual difficulty with the ICER is that its interpretation changes according to the sign of $\theta_e$. Quadrants II and IV correspond to the 'domination' of T1 and T2 respectively, in that one treatment is both less costly and more effective; in these quadrants the ICER is negative and the interpretation is clear. In quadrant I, T2 is more costly but more effective: in area IA, T2 is an acceptable choice as the additional benefit is achieved at a smaller unit cost than $K$ (here ICER $< K$), whereas in IB, T2 would be unacceptable. In quadrant III, T2 is less costly but less effective: in area IIIA, T2 would be considered unacceptable as insufficient gains in cost were being obtained for the effectiveness lost, the ICER being less than $K$, whereas in the area IIIB, where T2 is acceptable, the ICER is greater than $K$.

Thus, if there is any possibility that $\theta_e < 0$, it could be very misleading to base any conclusions on possible values of the ICER, since T2 is favoured by small values of the ICER when $\theta_e < 0$, and large values of the ICER when $\theta_e > 0$. In fact, the area where T2 is favoured corresponds to all the cost-effectiveness plane lying below the dashed line, which includes *all* possible values of the ICER. See O'Hagan *et al.* (2000) and Heitjan *et al.* (1999) for further discussion and illustrations.

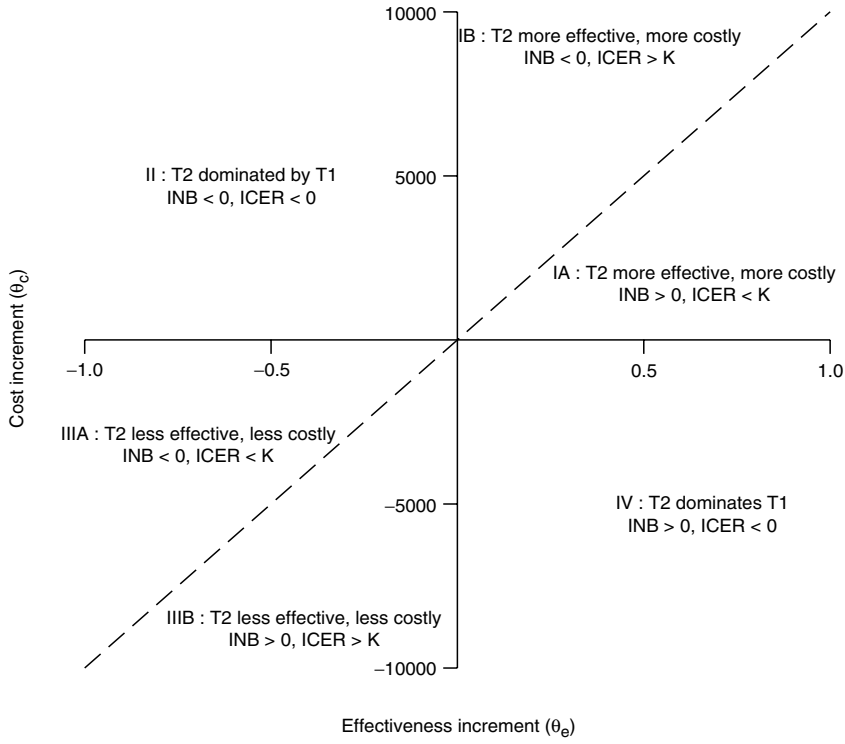


**Figure 9.1** Interpretation of different segments of the incremental cost-effectiveness plane. The dashed line represents $\theta_c = K\theta_e$, where $K$ is the willingness to pay for a unit of benefit. Since the incremental net benefit $\text{INB} = K\theta_e - \theta_c$, the dashed line represents $\text{INB} = 0$, the breakeven point. The incremental cost-effectiveness ratio $\text{ICER} = \theta_c/\theta_e$.

The *incremental net benefit* (INB) function has been proposed as an alternative means of interpretation of cost-effectiveness analyses which avoids the problems associated with the ICER, and is defined by

$$\text{INB}(K) = K\theta_e - \theta_c. \tag{9.2}$$

INB($K$) as defined by (9.2) represents the incremental net monetary benefit in terms of economic costs, and provides a connection to classical cost–benefit analysis. INB can also be transformed to the incremental net *health* benefit, in which case INB$^*$($K$) is given by

$$\text{INB}(K)/K = \text{INB}^*(K) = \theta_e - \theta_c/K. \tag{9.3}$$

It is straightforward to see that the regions in Figure 9.1 which correspond to $\text{INB} > 0$, *i.e.* acceptability of T2, represent all the regions below the dashed line, *i.e.* IA, IV and IIIB.

Setting INB = 0 yields the 'breakeven' cost per unit effectiveness $K_0 = \theta_c/\theta_e$ which is numerically equal to the ICER, and this value can be subject to deterministic sensitivity analysis of alternative assumptions.

The value $K$ must be handled with care. Taking the perspective of a health-care agency, it represents their 'willingness to pay' for the gain of a unit of effectiveness. Such a value would not usually be considered as fixed, nor as a random quantity. Instead it is natural to carry out an analysis of sensitivity to alternative values of $K$, with values of around \$50 000 perhaps being considered reasonable in the USA, and lower values such as £20 000 in the UK. See Claxton *et al.* (2000) for a recent discussion of this quantity.

## 9.4 'TWO-STAGE' AND INTEGRATED APPROACHES TO UNCERTAINTY IN COST-EFFECTIVENESS MODELLING

Let $\psi$ represent state-of-the-world parameters in a cost-effectiveness model, for example the true mean cost and benefit of an intervention, and let $X$ be a set of unknown generic outcomes of interest, both costs and benefits, taking on a value $x$. Suppose, for a specified value of $\psi$, we can specify a predictive distribution $p(x|\psi)$, the *chance variability* between outcomes on future patients. Our primary interest is in $\mathrm{E}(X|\psi) = \int x\, p(x|\psi)dx = m_\psi$, the expected outcome in a homogeneous population. $m_\psi$ will often be available in closed form, say when using discrete-time, discrete-state Markov models (Section 9.8).

Any uncertainty concerning $\psi$ may be expressed as a distribution $p(\psi)$, from which we can obtain a joint distribution for $m_\psi$, the expected costs and benefits of the intervention. By considering different interventions we can thus obtain a joint distribution over the incremental expected costs and effectiveness from a new intervention, denoted $\theta_c$ and $\theta_e$ respectively, the quantities of interest in a cost-effectiveness analysis (Section 9.3). In practice this will generally require simulation of a value of $\psi$ from $p(\psi)$, which is propagated through the cost-effectiveness model to obtain $m_\psi$, which in turn provides a value for $\theta_c, \theta_e$. Repeated simulations provide a joint distribution for $\theta_e$, $\theta_c$, and hence a distribution for any functions of $\theta_c$, $\theta_e$ such as the INB. The construction and analysis of this joint distribution has been termed *probabilistic sensitivity analysis* in the cost-effectiveness literature, to distinguish it from *deterministic sensitivity analysis* in which parameters are varied systematically across ranges.

Two approaches are possible. The *two-stage* approach proceeds as follows. First, $p(\psi)$ is constructed as a closed-form distribution, based on subjective judgements, data analysis or a combination of the two: $p(\psi)$ can be thought of as a prior distribution even though it may be partly based on evidence. Generally the elements of $\psi$ will be assumed independent and parametric distributions adopted. Values of $\psi$ are then simulated from $p(\psi)$ and the cost-effectiveness model provides the relevant outcomes $\theta_e, \theta_c$. This is a natural application of Monte Carlo methods (Section 3.19.1) in homogeneous populations, which has become a standard tool

in risk analysis to deal with 'second-order uncertainty', as opposed to first-order 'chance' uncertainty (Section 9.7). It is implementable as a Microsoft Excel® macro, either from commercial software such as @RISK (Palisade Europe, 2001) and Crystal Ball (Decisioneering, 2000), or self-written. Here, however, we use the freely available WinBUGS software (Section 3.19.3) in order to facilitate both approaches. A schematic representation is shown in Figure 9.2(a). Applications of the two-stage approach are demonstrated in Example 9.1 for the simple normal case, and Example 9.3 for a more complex model.

The *integrated* or *unified* approach unifies the two stages described above, in that $p(\psi)$ is taken to be a posterior distribution arising from a data analysis, which feeds directly into the cost-effectiveness model without an intermediate summary step. This corresponds to a full Bayesian probability model and

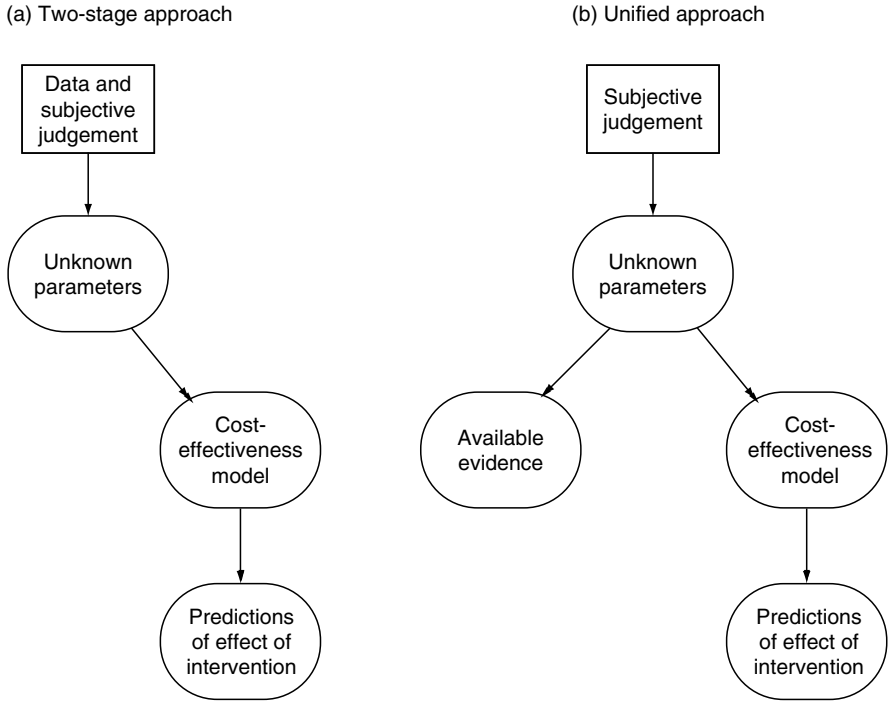


**Figure 9.2** Schematic graph showing the two approaches to incorporating uncertainty about parameters into a cost-effectiveness analysis. (a) The *two-stage approach* subjectively synthesises data and judgement to produce a prior distribution on the parameters which is then propagated through the cost-effectiveness model. (b) The *unified* or *integrated approach* adopts a fully Bayesian analysis: after taking into account the available evidence, initial prior opinions on the parameters are revised by Bayes theorem to posterior distributions, the effects of which are propagated through the cost-effectiveness model in order to make predictions. An integrated Bayesian approach ensures that the full joint uncertainty concerning the parameters is taken into account.

requires MCMC rather than simply Monte Carlo techniques, since in effect the evidence from the data has to be propagated 'against the arrow' in order to give the uncertainty on the parameters, and then 'forwards' through the cost-effectiveness model; a schematic representation is shown in Figure 9.2(b). Implementation will generally be in a full MCMC program such as WinBUGS: see Examples 9.2 and 9.4. The potential advantages and disadvantages of this integrated approach over the two-stage process are discussed in Section 9.9.2.

## 9.5 PROBABILISTIC ANALYSIS OF SENSITIVITY TO UNCERTAINTY ABOUT PARAMETERS: TWO-STAGE APPROACH

From a strict decision-theoretic approach, any uncertainty about the parameters $\theta_c, \theta_e$ is irrelevant to decision-making, and their expectations need only be placed in (9.2) for a specified $K$, and T2 chosen if INB $> 0$. Nevertheless, for reasons outlined in Sections 3.14 and 6.2, and discussed further in Section 9.11, it is generally considered appropriate to specify a measure of certainty that T2 is in fact an acceptable option. Confidence intervals for INB can be derived within the classical framework, but a Bayesian approach is natural and straightforward and allows the inclusion of additional prior information.

If we take the two-stage approach (Section 9.4) and assume that a joint prior distribution $(\theta_e, \theta_c)$ is available based on judgment, data, or a mixture of the two, then this can be plotted on the cost-effectiveness plane shown in Figure 9.1 and the probability of specific conclusions may be obtained by integrating over the appropriate areas (Grieve, 1998). As mentioned in Section 9.4, this has become known as probabilistic sensitivity analysis (Briggs and Gray, 1999). In addition, Heitjan *et al.* (1999) suggest obtaining the distribution of the ICER conditional on being in each quadrant of Figure 9.1.

A joint distribution on $(\theta_e, \theta_c)$ implies a distribution on INB. If we denote $\mathrm{E}[\theta_e] = \mu_e, \mathrm{V}[\theta_e] = \tau_e^2, \mathrm{E}[\theta_c] = \mu_c, \mathrm{V}[\theta_c] = \tau_c^2, \mathrm{Corr}[\theta_e, \theta_c] = \rho$, and similarly for costs, then without further distributional assumptions we have, for INB $= K\theta_e - \theta_c$, that

$$\mathrm{E}[\mathrm{INB}] = K\mu_e - \mu_c, \tag{9.4}$$

$$\mathrm{V}[\mathrm{INB}] = K^2\tau_e^2 - 2K\rho\tau_e\tau_c + \tau_c^2. \tag{9.5}$$

Thus we can plot E[INB] and, for example, its $\pm 2$ standard deviation interval for different values of $K$. The breakeven point $K_0$ occurs at $\mu_c/\mu_e$.

In terms of decision-making it is natural to consider the probability that INB($K$) in (9.2) is positive for any given value of $K$, *i.e.*

$$Q(K) = P(\mathrm{INB}(K) > 0). \tag{9.6}$$

$Q(K)$ is referred to as the *cost-effectiveness acceptability curve* (CEAC); see van Hout *et al*. (1994). Although $Q(K)$ has been interpreted in frequentist terms, the CEAC is most naturally handled within a Bayesian approach.

It may be reasonable to make a normal approximation to the distribution of INB, and then the CEAC is given by

$$Q(K) = P(\text{INB} > 0) = \Phi\left(\frac{K\mu_e - \mu_c}{\sqrt{K^2\tau_e^2 - 2K\rho\tau_e\tau_c + \tau_c^2}}\right), \tag{9.7}$$

and this expression is exact if we assume bivariate normality (Section 2.6.10) for $\theta_e, \theta_c$ – it is also possible to solve (9.7) explicitly to find the value $K$ at which, for example, $Q(K) = 0.95$ or some other desired level of 'significance'. O'Hagan *et al*. (2000) describe various closed-form approximations when normality is not assumed, but in this situation it seems preferable to move to the MCMC approaches as described in the next section.

Not all inferences of interest can be obtained in closed form even when assuming joint normality for $\theta_e, \theta_c$, and in this case it can be better computationally to model the joint distribution in two stages: from Section 2.6.10 we see that $\theta_e \sim \text{N}[\mu_e, \tau_e^2]$, and $\theta_c|\theta_e$ is normal with mean and variance

$$\begin{aligned} \text{E}[\theta_c|\theta_e] &= \mu_c + \frac{\rho\tau_c}{\tau_e}(\theta_e - \mu_e), \\ \text{V}[\theta_c|\theta_e] &= \tau_c^2(1 - \rho^2). \end{aligned} \tag{9.8}$$

Thus we can simulate $\theta_e$ followed by $\theta_c|\theta_e$. This is illustrated in Example 9.1.

---

**Example 9.1** *Anakinra: Two-stage approach to cost-effectiveness analysis*

*Reference:* van Hout *et al*. (1994).

*Intervention:* Human recombinant interleukin-1 receptor antagonist (anakinra) in the treatment of sepsis syndrome.

*Aim of study:* To assess the cost-effectiveness of anakinra compared to placebo.

*Study design:* RCT with 25 patients per arm.

*Outcome measure:* Effectiveness measured by survival (proportion surviving), and costs of treatment measured in Dutch guilders. The guilder, now replaced by the euro, was valued at around 2.2 to the US dollar.

*Statistical model and evidence from study:* Table 9.1 shows the data for one of the outcomes of the trial. There is clearly substantial evidence of a clinical benefit, but considerable uncertainty about increases in costs.

**Table 9.1** Available data from anakinra study.

| Quantity | Estimate | SD | Correlation |
|---|---|---|---|
| $\theta_e$: Increase in effectiveness (survival) | 0.28 | 0.123 | 0.34 |
| $\theta_c$: Increase in costs (guilders) | 1380 | 5657 | |

*Prior distribution:* We may approximate a joint prior as having the same properties as the sample data shown in Table 9.1, so that $\mu_e = 0.28,\ \tau_e = 0.123,\ \mu_c = 1380,\ \tau_c = 5657,\ \rho = 0.34$. By further assuming joint normality, the contours for $(\theta_c,\ \theta_e)$ may be plotted as in Figure 9.3.

*Computation/software:* The distribution of INB can be obtained exactly from (9.4) and (9.5), while the CEAC is given by (9.7). Other calculations, such as the distribution of the ICER and the probabilities of lying in each of the quadrants, are carried out by Monte Carlo methods implemented using WinBUGS, taking advantage of the conditional sampling scheme described in (9.8).

*Bayesian interpretation:* Figure 9.3(a) plots cost per extra survivor when $K = 5000$ and 35 000 guilders. The probabilities of lying in quadrants I, ..., IV are 59.3%, 0.3%, 0.9%, 39.6% respectively, so that there is around a 40% chance that anakinra dominates placebo in costs and benefits. The ICER has median 5146 and 95% interval $-79\,260$ to $+57\,990$. However, it is not clear whether the high values occur in quadrant I or III, which would have a completely different interpretation. Heitjan *et al*. (1999) report that *if* the ICER is in quadrant I, then it has an interval from 791 to 163 400 additional guilders per life saved, while *if* the ICER is in quadrant III, the interval is from 8400 to 4 580 000 guilders saved per life sacrificed. While these conditional statements reveal the different nature of the ICER in different quadrants, their interpretation is not straightforward.

Figure 9.3(b) plots the distribution of the incremental net benefit INB for $K = 5000$, 35 000, 100 000: for $K = 5000$ there appears to be almost complete indifference between the options, while the INB increases substantially as the willingness to pay per additional survivor increases. The mean and 95% intervals for the INB for a wide range of $K$ are shown in Figure 9.3(c), while Figure 9.3(d) plots $Q(K) = P(\text{INB} > 0)$ against $K$: the analysis suggests, on balance, that anakinra is cost-effective provided $K$ is greater than around 5000 guilders, and we can be 95% sure that anakinra is cost-effective provided $K$ is greater than around 45 000 guilders. Whether this would provide an appropriate basis for recommendation of the treatment depends on the decision-maker.

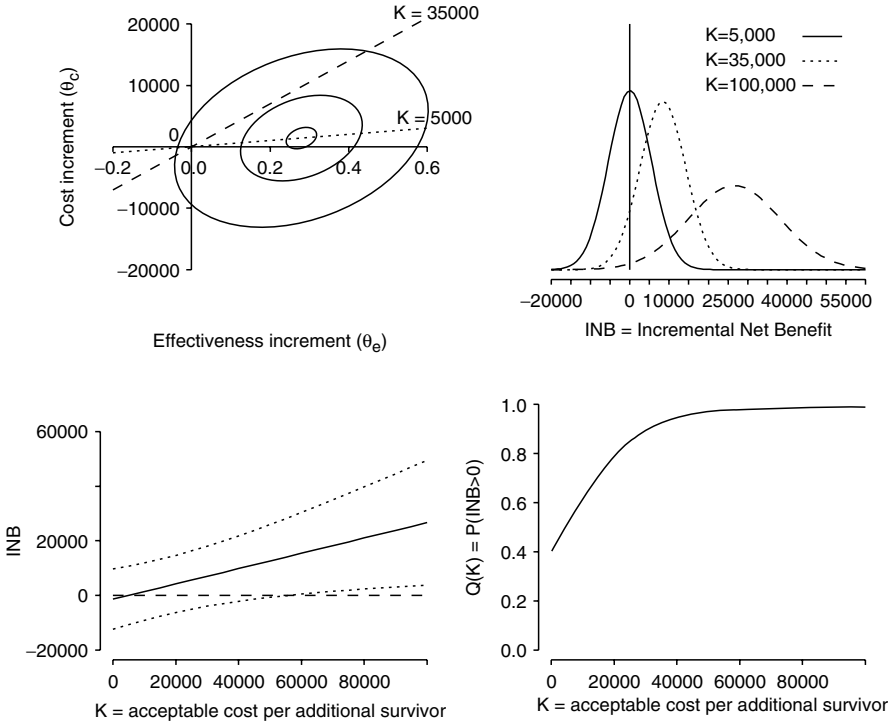


**Figure 9.3** Results for anakinra study. (a) Joint distribution of $(\theta_e, \theta_c)$, superimposed on lines representing maximum acceptable cost per additional survivor $K = 5000$, 35 000. (b) Distribution of incremental net benefit for $K = 5000$, 35 000, 100 000. (c) E[INB] and 95% intervals for a range of values of $K$. (d) Cost-effectiveness acceptability curve.

*Sensitivity analyses:* The primary sensitivity analysis concerns the specification of $K$.

## 9.6 COST-EFFECTIVENESS ANALYSES OF A SINGLE STUDY: INTEGRATED APPROACH

In the previous section we assumed $p(\theta_e, \theta_c)$ was a prior distribution based on a subjective synthesis of evidence and judgement. We now suppose we have data sources available from which to derive a posterior distribution $p(\theta_e, \theta_c|\,\text{data})$, and adopt the integrated approach outlined in Section 9.4. We emphasise that $\theta_e$ and $\theta_c$ must be population *mean* effectiveness and cost increments, in order to make measures additive across individuals. Hence, although cost data will generally have a highly skewed distribution, we must be careful to make inferences about their mean rather than some other measure of location.

Data sources available may include clinical trials, meta-analyses, observational studies and so on, and in later sections we shall consider how to exploit various sources of evidence. Here we shall only consider data from a single clinical trial, in which we assume we have observed pairs $(e_{ij}, c_{ij})$ representing the observed effect and cost when treatment $i$ is given to patient $j$. The process of modelling the joint sampling distribution of $(e_{ij}, c_{ij})$ within each treatment group requires care and statistical insights which are beyond the scope of this book – we refer to O'Hagan and Stevens (2002a) for a variety of approaches in this context. An obvious starting point is to assume bivariate normality (O'Hagan *et al.*, 2001), although the skewness of the cost data will generally make this unreasonable and log-costs might better be assumed normal. Cost data are frequently bimodal and a mixture of distributions may be appropriate (O'Hagan and Stevens, 2001; Cooper *et al.*, 2003*c*). It is also natural to consider a two-stage approach in which we model effectiveness and then costs conditional on effectiveness: this is the approach taken in Example 9.2. In any of these situation the complexity of the necessary inferences makes MCMC the computational procedure of choice; Fryback *et al.* (2001a) provide a further example of a posterior distribution being used as a direct input to probabilistic sensitivity analysis using WinBUGS.

---

**Example 9.2** *TACTIC: integrated cost-effectiveness analysis*

*References:* O'Hagan *et al*. (2001), O'Hagan and Stevens (2001, 2002a).

*Intervention:* Turbuhaler (treatment 2), a novel inhaler for asthmatics, compared to conventional CFC pressurised metered dose inhaler (pMDI, treatment 1).

*Aim of study:* To investigate whether asthmatic patients who were considered to be adequately treated using a conventional pMDI could be transferred to Turbuhaler without decrease in the effect of treatment, whilst reducing average costs.

*Study design:* RCT with prospective collection of costs: we use the data of O'Hagan *et al*. (2001) which comprise only the UK portion of the study.

*Outcome measure:* Number of days with exacerbation and total costs in pounds sterling.

*Planned sample size:* The original trial was designed to be able to detect a 10% improvement in the proportion of patients experiencing no exacerbations during the course of the trial, from 50% on pMDI to 60% on Turbuhaler.

*Evidence from study and statistical model:* The summary data are presented in Table 9.2. Turbuhaler patients suffered fewer exacerbations: the high proportion with no exacerbations suggests a normal distribution for

**Table 9.2** Results from UK portion of TACTIC trial of Turbuhaler compared to pMDI: log-costs are given separately for patients with and without exacerbations.

| | Treatment | *n* | No. exacerbations | Log-costs (mean and SD) | |
|---|---|---|---|---|---|
| | | | | With exac. | No exac. |
| T1 | pMDI | 58 | 26 (45%) | 6.02 (1.11) | 5.87 (1.47) |
| T2 | Turbuhaler | 62 | 36 (58%) | 6.37 (0.98) | 6.13 (0.85) |

clinical outcome is unreasonable and instead we follow O'Hagan and Stevens (2001) in adopting a binary outcome to measure benefit: $e_{ij} = 0$ if exacerbation occurred, 1 otherwise, with proportion $\phi_i$ in treatment group $i$.

Figure 9.4 shows the distribution of log-costs in the two treatment groups and according to whether exacerbations were experienced: it is important to note that there were two extremely high costs of 19 871 and 26 201 in the pMDI group who suffered no exacerbations, which are extremely influential in a normal model for costs (O'Hagan *et al*., 2001) and lead to a higher standard deviation for log-costs. Nevertheless, the empirical distributions in Figure 9.4 suggest adopting a dependent model in which log-costs are assumed normally distributed with mean and standard deviation dependent on treatment and exacerbation. We thus have a model

$$\begin{aligned} e_{ij} &\sim \text{Bern}[\phi_i], \\ \log(c_{ij})|e_{ij}=0 &\sim \text{N}[\lambda_{i0}, \sigma^2_{i0}], \\ \log(c_{ij})|e_{ij}=1 &\sim \text{N}[\lambda_{i1}, \sigma^2_{i1}]. \end{aligned}$$

The mean costs $m_{ci}$ in each treatment group are therefore a weighted average of the means in each exacerbation group and hence, from the known properties of the log-normal distribution (Section 2.6.8), are

$$m_{ci} = (1-\phi_i)e^{\lambda_{i0}+\sigma^2_{i0}/2} + \phi_i e^{\lambda_{i1}+\sigma^2_{i1}/2},$$

from which we can derive the mean cost and effectiveness differences

$$\begin{aligned} \theta_c &= m_{c2} - m_{c1}, \\ \theta_e &= \phi_2 - \phi_1, \end{aligned}$$

which are the inputs to the cost-effectiveness analysis.

*Prospective analysis?:* No.

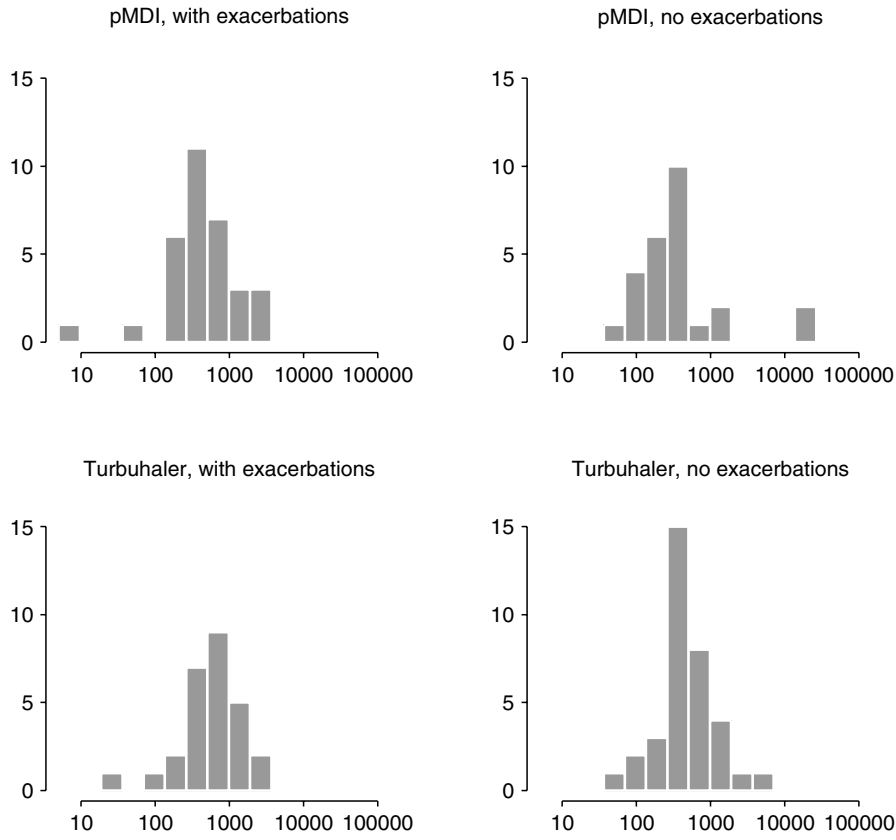


**Figure 9.4** Costs for TACTIC data, broken down by treatment (pMDI or Turbuhaler) and whether exacerbations occurred or not.

*Prior distribution:* O'Hagan and Stevens (2001) use an informative prior for the clinical effectiveness ($\phi_1$, $\phi_2$), with a mean of 0.1 on $\phi_2 - \phi_1$ which matches the difference used in the power calculations. This initial bias may be considered unreasonable by any regulatory body unless based on substantial evidence, and in any case the evidence from the trial is reasonably strong, and so we adopt independent uniform priors on $\phi_1$ and $\phi_2$ (an alternative might be uniform on logit($\phi_2$) and on $\theta_e = \phi_2 - \phi_1$, but this has negligible impact).

For the log-cost distributions, we assume independent uniform priors for the $\lambda_{i0}, \lambda_{i1}$. Partly in view of the potential influence of individual observations, and because we might expect the variability in costs to be similar, O'Hagan and Stevens (2001) suggest assuming $\sigma_{10}, \sigma_{11}, \sigma_{20}, \sigma_{21}$ exchangeable in order to 'smooth' the four observed

standard deviations towards a common value. We shall assume the $\log(\sigma)$s are normally distributed, such that

$$\log \sigma_{ij} \sim \text{N}[\mu_\sigma, \tau_\sigma^2]; \qquad i = 1,\ 2,\ j = 0,\ 1,$$

where $\mu_\sigma, \tau_\sigma$ are given uniform priors.

*Loss function or demands:* No.

*Computation/software:* MCMC using WinBUGS.

*Bayesian interpretation:* Figure 9.5(a) plots the joint posterior distribution of $\theta_e$ and $\theta_c$, showing they are reasonably independent: the posterior probability is 0.53 that Turbuhaler is cheaper, and 0.93 that it is more effective; the probability that it dominates pMDI is 0.51. Figure 9.5(b) shows the posterior distribution of the incremental net benefit assuming $K = £500$ per patient prevented from having exacerbations – a value at which there is approximate indifference as to the preferred treatment. The expected INB and 95% intervals are displayed in Figure 9.5(c), showing a steady preference for Turbuhaler as the willingness to pay for preventing exacerbations increases. The CEAC in Figure 9.5(d) suggests we can be 90% sure of the cost-effectiveness of Turbuhaler provided that $K$ exceeds £5000. Estimates and intervals for relevant quantities are given in Table 9.3; comparison of the estimates of the $\sigma$s with those shown in Table 9.3 reveals the shrinkage arising from the exchangeability assumption.

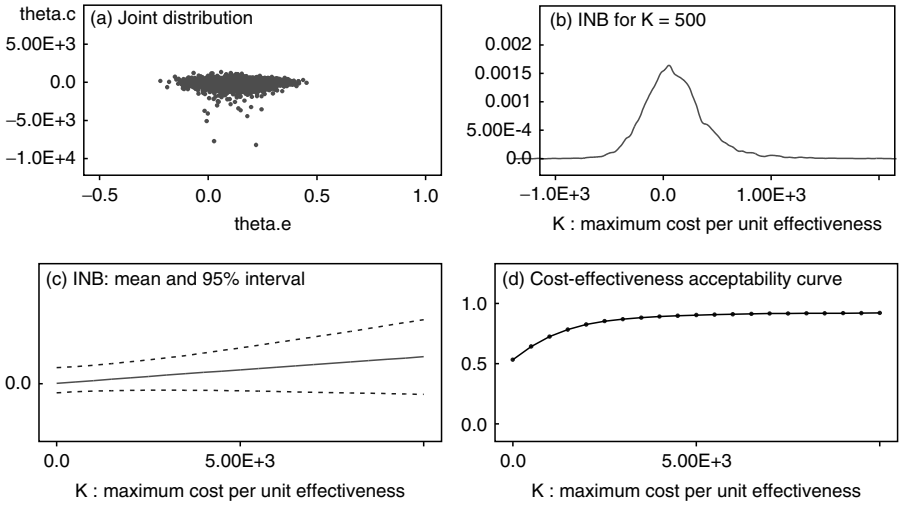


**Figure 9.5** Plots of (a) joint distribution of incremental mean benefits $\theta_e$ and mean costs $\theta_c$, (b) distribution of incremental net benefit assuming $K = £500$, (c) the expected INB and 95% interval, and (d) the CEAC for a range of $K$. These plots are direct output from WinBUGS.

**Table 9.3** Prior-to-posterior cost-effectiveness analysis of Turbuhaler compared to pMDI: results are given assuming that the standard deviations of the log-costs are either exchangeable or independent.

| Parameter | | Posterior (exch.) | | Posterior (indep.) | |
|---|---|---|---|---|---|
| | | Median | 95% interval | Median | 95% interval |
| Effect of pMDI | $\phi_1$ | 0.45 | 0.33 to 0.58 | 0.45 | 0.33 to 0.58 |
| Effect of Turbuhaler | $\phi_2$ | 0.58 | 0.45 to 0.70 | 0.58 | 0.45 to 0.70 |
| Excess effect of Turbuhaler | $\theta_e = \phi_2 - \phi_1$ | 0.13 | −0.04 to 0.30 | 0.13 | −0.04 to 0.30 |
| Mean cost of pMDI | $m_{c1}$ | 862 | 581 to 1620 | 983 | 625 to 2222 |
| Mean cost of Turbuhaler | $m_{c2}$ | 835 | 626 to 1235 | 817 | 620 to 1225 |
| Excess mean cost of Turbuhaler | $\theta_c = m_{c2} - m_{c1}$ | −21 | −801 to 455 | −161 | −1409 to 371 |
| SD of log-costs, pMDI, exac. | $\sigma_{10}$ | 1.12 | 0.89 to 1.41 | 1.14 | 0.89 to 1.51 |
| SD of log-costs, pMDI, no exac. | $\sigma_{11}$ | 1.37 | 1.08 to 1.84 | 1.52 | 1.17 to 2.08 |
| SD of log-costs, Turbuhaler, exac. | $\sigma_{20}$ | 1.02 | 0.80 to 1.34 | 1.01 | 0.78 to 1.39 |
| SD of log-costs, Turbuhaler, no exac. | $\sigma_{21}$ | 0.92 | 0.72 to 1.20 | 0.87 | 0.70 to 1.14 |
| INB(500) | | 89 | −394 to 851 | 438 | −238 to 1652 |
| INB(5000) | | 694 | −350 to 1783 | 2834 | −829 to 6455 |
| INB(10 000) | | 1349 | −528 to 3194 | 5423 | −1685 to 12380 |
| $Q$(500) | | 0.64 | | 0.90 | |
| $Q$(5000) | | 0.90 | | 0.94 | |
| $Q$(10 000) | | 0.92 | | 0.93 | |

*Sensitivity analysis:* The assumption of exchangeable $\sigma$s is the only form of informative prior that is currently being used. If we adopt independent uniform priors on the $\sigma$s we obtain the results shown in the final two columns of Table 9.3. The independence assumption allows the two outlying costs to exert a strong influence on $\sigma_{11}$, which in turn substantially increases the estimated mean cost of pMDI ($m_{c1}$). This increases the INB of Turbuhaler, which substantially increases the probability $Q(K)$ of cost-effectiveness even for low values of $K$. The posterior probability is 0.72 that Turbuhaler is cheaper, and 0.93 that it is more effective: the probability that it dominates pMDI is 0.68.

Given the extreme sensitivity to two outlying costs, it would be important to identify the precise reasons for these values, and ideally collect further cost information on additional patients.

## 9.7 LEVELS OF UNCERTAINTY IN COST-EFFECTIVENESS MODELS

Approaches to uncertainty in cost-effectiveness analysis have been extensively reviewed by Briggs and Gray (1999), who emphasise the distinction between conducting 'deterministic' sensitivity analysis in which inputs to a model are systematically varied within a reasonable range, and 'probabilistic' sensitivity analysis in which the relative plausibility of unknown parameters is taken into account.

We can relate these different approaches to analysis of sensitivity to different sources of uncertainty; similar taxonomies have been described by Briggs (2000) and the US Panel on Cost-Effectiveness (Manning *et al.*, 1996).

1. *Chance variability*. This is the unavoidable *within-individual* predictive uncertainty concerning specific outcomes, which will be empirically demonstrated by variability in outcomes between homogeneous individuals. We are usually not interested in this 'first-order' uncertainty (Briggs, 2000) since our focus is on the *expected* outcomes in homogeneous populations, but we shall illustrate its calculation in Section 9.8.
2. *Heterogeneity*. This source concerns *between-individual* variability in expected outcomes, due to either (a) identifiable subgroups of individuals with characteristics such as age, sex and other covariates, or (b) unmeasurable differences (latent variables). These are termed 'patient characteristics' by Briggs (2000). We shall generally want to use deterministic sensitivity analysis to see how expected outcomes vary between identifiable subgroups, possibly followed by probabilistic averaging over population subgroups according to their incidence.
3. *Parameter uncertainty*. This concerns *within-model* uncertainty as to the appropriate values for parameters. Parameters can be divided into two types:
   (a) *States-of-the-world*, which could, in theory, be measured precisely if sufficient evidence were available (e.g. risks, disease incidences): these have also been termed 'parameters that could be sampled' (Briggs, 2000). These can have distributions placed on them, corresponding to the 'second-order' uncertainty used in risk analysis (Burmaster and Wilson, 1996), and so be subject to probabilistic sensitivity analysis.
   (b) *Assumptions*, which are quantitative judgements placed in the model which can only be made precise through consensus agreement, for example discount rates for health benefits. These can be considered as one source of 'methodological uncertainty' (Briggs, 2000), and sensitivity to assumptions can only be carried out deterministically by rerunning analyses under different scenarios.

   The appropriate category for a quantity is not always clear. For example, whether values placed on quality-of-life scales are states-of-the-world or assumptions is a controversial point, and costs might also be placed in either category.
4. *'Ignorance'*. this *between-model* uncertainty describes our basic lack of knowledge concerning the appropriate qualitative structure of the model, for example, the dependence of hazard rates on background factors and history. This is also a component of 'methodological uncertainty' (Briggs, 2000). Deterministic sensitivity analysis takes the form of running through alternative models, although there is a Bayesian argument that model

structure can itself be considered as an unknown state-of-the-world and be subject to probabilistic sensitivity analysis (Draper, 1995).

In this chapter we shall primarily be concerned with probabilistic sensitivity analysis, although we will also illustrate deterministic sensitivity analysis with respect to parameter assumptions.

## 9.8 COMPLEX COST-EFFECTIVENESS MODELS

We have so far considered the situation in which the necessary estimates of effectiveness and costs are derived directly from clinical trial data. However, a clinical trial may neither address precisely the population of interest, nor last long enough for the rate of important long-term outcomes to be accurately assessed. In the former situation the trial results may need to be adjusted in order to generalise the cost-effectiveness analysis to other populations of interest (Rittenhouse, 1997), which may involve the type of adjustments used in cross-design synthesis (Section 8.4) and the explicit modelling of biases in observational studies (Section 7.3). In the latter case we will need a model for long-term outcomes, such as the Markov models that have been used extensively in cost-effectiveness analysis.

### 9.8.1 Discrete-time, discrete-state Markov models

These models are generally applied to the development of a disease process over time, and assume that in each 'cycle' an individual is in one of a finite set of states, and that there is a certain chance of transferring to a different state at the next cycle. The 'Markov' label refers to the assumption that the chance of entering a new state at the start of each cycle does not depend on the path the individual took to their current state (although the chance may depend on the cycle and other risk factors). There are obviously many extensions to this reasonably flexible framework (Briggs and Sculpher, 1997, 1998).

We shall first formally describe the generic structure of the model for a single homogeneous set of patients with common parameters. Assume a discrete-time model comprising $N$ cycles labelled $t = 1, \ldots, N$, and that within each cycle $t$ a patient remains in one of $R$ states, and that all transitions occur at the start of each cycle. The probability distribution at the start of the first cycle $t = 1$ is represented by the row vector $\boldsymbol{\pi}_1$, and we assume a transition matrix $\boldsymbol{\Lambda}_t$ whose $(i, j)$th element $\Lambda_{t,ij}$ is the probability of moving from state $i$ to state $j$ between cycle $t-1$ and $t$; thus the probability, for example, of being in state $j$ during the second cycle is $\sum_i \pi_{1i}\Lambda_{2,ij}$. Hence, the marginal probability distribution $\boldsymbol{\pi}_t$ during cycle $t > 1$ obeys the recursive relationship

$$\boldsymbol{\pi}_t = \boldsymbol{\pi}_{t-1}\boldsymbol{\Lambda}_t. \tag{9.9}$$

Suppose the cost, at current prices, of spending a cycle in state $r$ is $C_r$, $r = 1, \ldots, R$ and there is a fixed entry cost $C_0$. It is standard practice in economic evaluations to discount costs that occur in future years, at rate $\delta_c$ (say) per cycle. Then the total cost acquired by each patient in the population is expected to be

$$m_c = C_0 + \sum_{t=1}^{N} \frac{\pi_t \boldsymbol{C}'}{(1+\delta_c)^{t-1}}. \tag{9.10}$$

Similarly, if the benefits associated with spending one cycle in each state are given by a row vector $\boldsymbol{b}$, discounted at rate $\delta_b$ per cycle, the total expected benefit for each patient is

$$m_e = \sum_{t=1}^{N} \frac{\pi_t \boldsymbol{b}'}{(1+\delta_b)^{t-1}}. \tag{9.11}$$

We note that different types of benefit may be reported, for example both life-years ($\boldsymbol{b} = 1$) and quality-adjusted life-years (QALYs), in which case $\boldsymbol{b}$ comprises a row vector of quality adjustments. A range of discount rates may also be explored: for example, guidance from NICE in the UK currently recommends that costs should be discounted at $\delta_c = 6\%$ per annum, while benefits are discounted at $\delta_b = 1.5\%$ (NICE, 2001). However, they add that sensitivity analyses should include assumptions of $\delta_b = 0\%$ and 6%.

Suppose there are $S$ discrete subgroups labelled by $s$. The model described above can clearly be extended to allow, say, for different transition matrices within subgroups by extending the notation to $\boldsymbol{\Lambda}_{st}$: this possibility is explored in detail in Spiegelhalter and Best (2003).

### 9.8.2 Micro-simulation in cost-effectiveness models

If we are using a more complex model in which it is not possible to write a formula for the expected outcomes, then it may be necessary to perform a much more complex simulation involving the trajectories of individual patients – this is known as *micro-simulation*. The sample mean of the simulations can be used as an estimate of the expected outcome in the population, and this approach does have the side-effect of giving the whole distribution of outcomes and, in particular, the variance among the population. This 'first-order simulation' approach is illustrated by Briggs (2000) and has been extensively exploited in the context of evaluating screening interventions (Cronin *et al.*, 1998).

For example, if we wished to explore this approach for the model described in Section 9.8.1, then we could simulate a starting state $y_1$ from the distribution $\pi_1$. We then simulate this individual's next state $y_2$ from

the distribution comprising the $y_1^{\text{th}}$ row of $\mathbf{\Lambda}_2$, and so on. The discounted costs and benefits for the individual are then

$$C = C_0 + \sum_{t=1}^{N} \frac{\boldsymbol{C}_{y_t}}{(1+\delta_c)^{t-1}}, \tag{9.12}$$

$$B = \sum_{t=1}^{N} \frac{b_{y_t}}{(1+\delta_b)^{t-1}}. \tag{9.13}$$

Averaging over many simulated patients (iterations) gives Monte Carlo estimates of the required expectations and also the variability of each outcome due to chance; Example 9.3 illustrates this process.

Note that if we simulate a patient under two treatments, then the incremental net benefit for that patient is estimated as

$$\text{INB} = K(B_2 - B_1) - (C_2 - C_1).$$

We could therefore estimate the proportion of the population for which the INB $> 0$ – this has been termed the 'probability of net benefit' (Willan, 2001). O'Hagan and Stevens (2002b) emphasise that this estimated population proportion must be carefully distinguished from the probability plotted in a CEAC, which reflects our uncertainty about the expectation over the whole population, and does not in any way take into account heterogeneity in benefit.

### 9.8.3 Micro-simulation and probabilistic sensitivity analysis

The previous section has described micro-simulation of individual patients, but this is all carried out for fixed parameters value $\psi$. Performing a probabilistic sensitivity analysis to allow for uncertainty in parameters is considerably more difficult in this context, and care must be taken. It would be tempting, but potentially misleading, to carry out a double simulation, in which a parameter value $\psi^j$ is sampled from $p(\psi)$, followed by simulation of an outcome $X^j$ conditional on $\psi^j$. The problem is that the variability in the subsequent $X^j$s combines that due to parameter uncertainty and that due to chance variability; unfortunately the two cannot be easily disentangled.

We first note that the total variance of $X$ can be written, using the identity (2.14) for conditional variances, as

$$\text{V}[X] = \text{E}_\psi[\text{V}(X|\psi)] + \text{V}_\psi[\text{E}(X|\psi)], \tag{9.14}$$

*i.e.* the expectation with respect to $\psi$ of the conditional variance of $X$, plus the variance of the conditional expectations. For a probabilistic sensitivity analysis we are only really interested in the second term, since the first term is concerned with chance variability in the population of patients.

These two components may be separated using a time-consuming nested simulation procedure (Halpern *et al.*, 2000). We briefly discuss the necessary computations, when assuming a distribution $p(\psi)$ derived from either the two-stage or integrated approach. A value $\psi^j$ for $\psi$ is simulated from $p(\psi)$, followed by simulation of $N$ (where $N$ is large) values of the outcome $X_1^j, \ldots, X_N^j$ conditional on $\psi^j$. The sample mean $\overline{X}_N^j$ and variance $V_N^j$ are stored. Monitoring $\overline{X}_N$ and $V_N$ will allow estimation of the components of the overall variability shown in (9.14), since $\mathrm{V}_\psi[\overline{X}_N]$ will estimate variability due to parameter uncertainty, while $\mathrm{E}_\psi[V_N]$ gives that due to chance variability. This technique will be laborious, particularly when heterogeneity is present, although $\mathrm{E}_\psi[V_N]$ may perhaps be reasonably estimated using only a limited set of $\psi$. See Cronin *et al.* (1998) for an application.

---

**Example 9.3** *HIPS: Cost-effectiveness analysis using discrete-time Markov models*

*References:* Spiegelhalter and Best (2003) and Fitzpatrick *et al*. (1998).

*Intervention:* Prosthesis for total hip replacement (THR).

*Aim of study:* To model the costs and outcomes of THR in a specific subgroup, men aged 65–74, assuming a Charnley prosthesis as a baseline analysis.

*Study design:* Cost-effectiveness model.

*Outcome measure:* Effectiveness measured by life expectancy and QALYs, and costs of treatment measured in pounds sterling.

*Statistical model:* We assume a discrete-time, discrete-state Markov model with cycles of 1 year. Figure 9.6 illustrates the various states and possible transitions between states. Patients initially enter state 1 (primary THR) at time $t = 0$. The first cycle ($t = 1$) is assumed to start immediately following the primary operation; patients have either died at operation or post-operatively, in which case they enter state 5 (death), otherwise they remain in state 1. In each subsequent cycle, surviving patients remain in state 1 until they either die from other causes (progress to state 5) or their hip replacement fails and they require a revision THR operation. Since the need for revision and the operation are assumed simultaneous, patients undergoing a revision operation enter one of two states depending on whether they die at or post-operation (state 2) or survive (state 3). Surviving patients progress to state 4 (successful revision THR) in the following cycle, unless they die from other causes (progress to state 5). Patients in state 4 remain there until they either die from other causes (state 5) or require another revision THR operation, in which case they progress back to states 2 or 3 as before. We also assume a

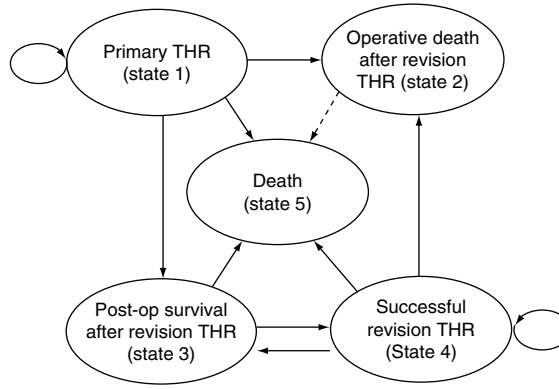


**Figure 9.6** Markov model for outcomes following primary total hip replacement.

transition from state 2 to state 5 in the cycle following operative death after a revision THR. This is slightly artificial but is necessary to avoid multiple counting of revision costs if patients were to remain in state 2.

We assume $\lambda_{\text{op}}$ is the operative mortality rate, $\gamma_t$ is the chance of revision in year $t$, $\lambda_t$ is the mortality rate $t$ years after primary operation, and $\rho$ is the re-revision rate which is assumed constant. The vector of state probabilities in cycle $t = 1$ is $\pi_1 = (1 - \lambda_{\text{op}},\ 0,\ 0,\ 0,\ \lambda_{\text{op}})$. We shall only consider one stratum, men between 65 and 74, and take 25 cycles of the model assumed to run between ages 70 and 95. The transition matrix $\mathbf{\Lambda}_{t,\ jk}$ is the probability of being in state $j$ in year $t - 1$ and moving to state $k$ at the start of year $t$; the transition probability matrix for $t = 2,\ \ldots,\ 25$ is given by

$$\begin{bmatrix} 1-\gamma_t-\lambda_t & \lambda_{\text{op}}\gamma_t & (1-\lambda_{\text{op}})\gamma_t & 0 & \lambda_t \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1-\lambda_t & \lambda_t \\ 0 & \rho\lambda_{\text{op}} & \rho(1-\lambda_{\text{op}}) & 1-\rho-\lambda_t & \lambda_t \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Baseline assumptions for the parameters of the model are given in Table 9.4; sources for these assumptions are provided in Fitzpatrick *et al*. (1998). Notable is the assumption that the revision risk increases linearly with time since operation, and constant re-revision risk. Health-related quality of life (HRQL) is measured in QALYs based on the degree of severity of pain patients would be likely to experience in different states of the model. Based on results from a Canadian study (Laupacis *et al*., 1993), Fitzpatrick *et al*. (1998) assign values $v_1 = 1,\ v_2 = 0.69,\ v_3 = 0.38$ and $v_4 = 0.19$ for the HRQL of patients experiencing no, mild, moderate and severe pain, respectively. They then assume that after a successful THR operation, 80% of patients experience no pain and 20% experience

**Table 9.4** Baseline parameters of total hip replacement model using a Charnley prosthesis: benefit weights $b$ are 1 for life expectancy, $b = q_k$ for QALYs.

| Parameter | | Value |
|---|---|---|
| Operative mortality rate | $\lambda_{op}$ | 0.01 |
| Revision rate | $\gamma_t = h(t-1)$ | $0.0016(t-1)$ |
| Re-revision rate | $\rho$ | 0.04 |
| Mortality rate | $\lambda_t$ | 0.038 (65–74) |
| | | 0.091 (75–84) |
| | | 0.196 (84+) |
| Primary cost | $C_0$ | £4052 |
| Revision cost | $C_2,\ C_3$ | £5290 |
| Cost discount rate | $\delta_c$ | 6% |
| Benefit discount rate | $\delta_b$ | 1.5% |
| Quality weights | $q_1$ | 0.938 |
| | $q_2$ | −0.622 |
| | $q_3$ | −0.337 |
| | $q_4$ | 0.938 |
| | $q_5$ | 0 |

mild pain. For patients whose hip replacements fail, they assume that 15% experience severe pain and 85% experience moderate pain in the year preceding the year of the revision operation, with a 50–50 split between those experiencing moderate pain and severe pain in the year of operation. We therefore calculate quality weights for each state in our Markov model as follows:

$$q_1 = 0.8v_1 + 0.2v_2 = 0.938,$$
$$q_2 = 0 + 1.06 \times (0.85v_3 + 0.15v_4 - 0.8v_1 - 0.2v_2) = -0.622,$$
$$q_3 = (v_3 + v_4)/2 + 1.06 \times (0.85v_3 + 0.15v_4 - 0.8v_1 - 0.2v_2) = -0.337,$$
$$q_4 = 0.8v_1 + 0.2v_2 = 0.938,$$
$$q_5 = 0.$$

We note that the rather odd negative weights arise from the need to essentially 'subtract' quality from preceding years.

*Prior distribution:* One relevant state-of-the-world parameter in our model for prognosis following THR is the revision 'hazard' parameter $h$. It may be reasonable to assume uncertainty of ±50% about our assumed revision hazard which we now denote $h_0$. This gives an approximate 95% interval of $(h_0/1.5,\ h_0 \times 1.5)$ for $h$, which corresponds to a prior standard deviation on the log scale of around 0.2 (Table 5.2). We therefore specify the prior distribution for the log-hazard parameter as

$$\log(h) \sim \mathrm{N}[\log(h_0),\ 0.2^2]. \tag{9.15}$$

*Computation/software:* MCMC methods implemented using WinBUGS.

*Bayesian interpretation:*

1. The closed-form calculation of expectations using (9.10) and (9.11) is shown in the 'closed-form' column of Table 9.5. Note that the expected life-years are around 10, and are not substantially reduced by quality adjustment.
2. The micro-simulation study showing variability among individuals is shown in 'population distribution' columns. The huge chance variability in the population is evident: however, as emphasised in Section 9.7, this between-individual variability is not of primary interest. The sampled means match the closed-form values up to Monte Carlo error – 100 000 iterations are used as the variability is so great, and even then the agreement for expected life-years is not good.
3. The final columns show the probabilistic sensitivity analysis by sampling from $p(\log(h))$ given in (9.15), and calculating the closed-form expectations at each iteration. This shows that the uncertainty about the revision hazard has a very limited effect on the expectations, particularly for life expectancy.

**Table 9.5** Predicted outcomes from hip replacement in men aged 65–74 years. The baseline expectation is obtained in closed form assuming known parameters. The population distribution is obtained by micro-simulation of individuals. The probabilistic sensitivity analysis summarises the predictive distribution of the expectation, allowing for a subjective prior distribution on the hazard rate.

| Parameter | Closed-form expectation | Population distribution | | Prob. sens. analysis | |
|---|---|---|---|---|---|
| | | Mean | SD | Median | 95% interval |
| Life-years | 9.939 | 9.954 | 5.426 | 9.939 | 9.936 to 9.941 |
| QALYs | 9.17 | 9.18 | 4.96 | 9.17 | 9.10 to 9.22 |
| Costs | 4458 | 4453 | 1220 | 4459 | 4334 to 4629 |

## 9.8.4 Comprehensive decision modelling

The primary advantage of a Bayesian approach is that it allows the synthesis of all available sources of evidence – whether from RCTs, databases, or expert judgement – into a single coherent and explicit model that can then be used to evaluate the cost-effectiveness of alternative policies. The approach has been termed 'comprehensive decision modelling', and can be thought of as extending the evidence synthesis methods described in Chapter 8 to allow for costs in

particular and for utilities in general, and possibly incorporating a predictive model for the natural history of a disease. Alternatively, it can be thought of as extending standard economic modelling techniques such as decision or Markov models so that they are probabilistic.

Parmigiani (2002) discusses such models in detail, pointing out that models should be 'requisite', in the sense of only being as complex as necessary. Ideally such models should allow a variety of viewpoints to be considered and incorporate the 'best possible' evidence, while encouraging analysis of sensitivity to both deterministic inputs and uncertain parameters. From a computational perspective, comprehensive decision models might be implemented in spreadsheets if a two-stage Monte Carlo approach is being adopted, or using MCMC software if integrated evidence synthesis and predictions are desired.

A number of case studies have been reported. Parmigiani and Kamlet (1993) and Parmigiani (1999) apply the idea to screening for breast cancer, and many sources of evidence are brought together in a single model that predicts the consequences of alternative screening policies, while Cronin *et al.* (1998) use micro-simulation at the level of the individual patient to predict the consequences of different policy decisions on lowering expected mortality from prostate cancer. Samsa *et al.* (1999) consider ischaemic stroke and construct a model for natural history using data from major epidemiological studies, and a model for the effect of interventions based on databases, meta-analysis of trials, and Medicare claim records. They also use micro-simulation of the long-term consequences of different stroke-prevention policies in order to compare their cost-effectiveness. Matchar *et al.* (1997), Parmigiani *et al.* (1996, 1997), and Parmigiani (2002) consider further use of their Stroke Prevention Policy Model. Fully integrated applications using WinBUGS have also been reported by Cooper *et al.* (2002, 2003a, 2003b).

## 9.9 SIMULTANEOUS EVIDENCE SYNTHESIS AND COMPLEX COST-EFFECTIVENESS MODELLING

The previous section has illustrated the two-stage approach to incorporating uncertainty into a complex cost-effectiveness model, and we now consider the full integration with Bayesian prior-to-posterior analysis.

### 9.9.1 Generalised meta-analysis of evidence

Example 9.2 provided a simple case for the integrated framework using the evidence from a single study and without a complex cost-effectiveness model, but the common situation in which evidence is available from a variety of sources demands a more challenging statistical analysis of the kind discussed

in detail in Chapter 8. If the evidence comprises a set of similar trials then a standard Bayesian random-effects meta-analysis may be sufficient. In more complex situations there may be multiple studies with relevance to the quantities in question but which may suffer from a range of potential inadequacies, such as being based on different populations, having non-randomised control groups, outcomes measured on different scales, and so on. As described in Section 8.4, it is natural to extend Bayesian random-effects modelling to allow variance components corresponding to different study designs (*i.e.* assuming study types are exchangeable), resulting in hierarchical models with a study type 'level'. There are clearly a number of issues in carrying out such potentially controversial modelling, such as when to judge studies or study types as 'exchangeable', how to put appropriate prior distributions on variance components, and how to carry out sensitivity analyses.

We shall consider as an illustration a somewhat simple formulation of such a model. Suppose we have a set of studies that are each intending to estimate a single parameter $\mu$ but, due to differences in populations studied and so on, any particular study (if carried out meticulously) would in fact be estimating a biased parameter $\theta_h$. Here $\theta_h - \mu$ is the 'external bias', and a standard random-effects formulation might then assume $\theta_h \sim \mathrm{N}[\mu,\ \tau^2]$ (note that the mean would not necessarily be $\mu$ if we suspected systematic bias in one direction). However, suppose that due to quality limitations there is additional 'internal bias' in the study, so that the true parameter being estimated is $\theta_h + \delta_h$. Then we might assume $\delta_h \sim \mathrm{N}[0,\ \sigma^2_{\delta h}]$ if we did not suspect that the internal bias would favour one or other treatment. If we assume all the studies have the same potential for external bias, then we are left with a random-effects model in which, for study $h$, the data are estimating a parameter

$$\begin{aligned}\theta_h &\sim \mathrm{N}[\mu,\ \tau^2 + \sigma^2_{\delta h}] \\ &\sim \mathrm{N}[\mu,\ \tau^2_h/q_h],\end{aligned}$$

where $q_h = \tau^2/(\tau^2 + \sigma^2_{\delta h})$ can be considered the 'quality weight' for each study, being the proportion of between-study variability unrelated to internal biasing factors. Thus a high-quality randomised trial might have $q = 1$, while a non-randomised study may be downweighted by assigning $q = 0.1$. Note that if we assume all studies are of equal 'quality', then we have the standard random-effects meta-analysis.

Estimates or prior distributions of the between-study variance $\tau^2$ and the quality weights $q_h$ might be obtained from a possible combination of empirical random-effects analyses of RCTs of this intervention, historical 'similar' case studies, and judgement. Of course, sensitivity analysis of a range of assumptions about the quality weights can be carried out.

This technique is illustrated in Example 9.4.

**Example 9.4** *HIPS (continued): Integrated generalised evidence synthesis and cost-effectiveness analysis*

*Reference:* Spiegelhalter and Best (2003).

*Available evidence:* In order to illustrate the trade-off between increased costs and benefits, we shall compare the cost-effectiveness of the Charnley prosthesis with a hypothetical alternative cemented prostheses costing an extra £350 but with some evidence for lower revision rates. We assume that all other costs (operating staff/theatre costs, length of hospital stay, X-rays etc.) are the same for both prosthesis types, and that the same method of QALY assessment is applicable for both types of prosthesis.

For illustration, we assume that the revision hazard for our hypothetical alternative is similar to that for the Stanmore prosthesis (a popular alternative to the Charnley in practice). Evidence on the relative revision hazards for the two prostheses is limited. The report by NICE on cost-effectiveness of different prostheses for THR (NICE Appraisal Group, 2000) cites three sources providing direct comparisons between Charnley and Stanmore revision rates:

1. The Swedish Hip Registry (Malchau and Herberts, 1998) provides non-randomised data submitted from all hospitals in Sweden from 1979, with record linkage to further procedures and death. Nine-year follow-up results are used for around 30 000 Charnley and 1000 Stanmore prostheses.
2. A British RCT (Marston *et al*., 1996) randomised around 400 patients to each of Charnley or Stanmore and reported a mean follow-up of 6.5 years.
3. A case series (Britton *et al*., 1996) of around 1200 patients in a single hospital with a mean follow-up of 8 years.

The available evidence from these three sources on revision hazards for Charnley and Stanmore prostheses is summarised in Table 9.6.

*Statistical model:* We assume the following model for pooling evidence on the revision hazard ratio for Stanmore versus Charnley prostheses. Let $n_{ik}$ and $r_{ik}$ denote the total number of patients receiving prosthesis $i$ ($1 =$ Charnley, $2 =$ Stanmore) in study $k$, and the number requiring a revision operation, respectively. We assume $r_{ik}$ is binomially distributed with proportion $p_{ik}$, although a little care is required in relating these cumulative failure rates to a hazard ratio. From Section 2.4.2 we know that, assuming proportional hazards, the hazard ratio $\mathrm{HR}_k$ for Stanmore versus Charnley prostheses obeys

**Table 9.6** Summary of evidence on revision hazards for Charnley and Stanmore prostheses: hazard ratios less than 1 are in favour of Stanmore.

| | **Charnley** | | **Stanmore** | | Estimated hazard ratio | |
|---|---|---|---|---|---|---|
| **Source** | Number of patients | Revision rate | Number of patients | Revision rate | HR | (95% int.) |
| | | | | | *Fixed-effects model* | |
| Registry | 28 525 | 5.9% | 865 | 3.2% | 0.55 | (0.37 to 0.77) |
| RCT | 200 | 3.5% | 213 | 4.0% | 1.34 | (0.45 to 3.46) |
| Case Series | 208 | 16.0% | 982 | 7.0% | 0.44 | (0.28 to 0.66) |
| | | | | | *Common-effect model* | |
| | | | | | 0.52 | (0.39 to 0.67) |
| | **Quality weights [Registry, RCT, Case Series]** | | | | *Random-effects model* | |
| | | | | [0.5, 1.0, 0.2] | 0.61 | (0.36 to 0.98) |
| | | | | [1.0, 1.0, 1.0] | 0.54 | (0.37 to 0.78) |
| | | | | [0.1, 1.0, 0.05] | 0.82 | (0.36 to 1.67) |

$$\mathrm{HR}_k = \frac{\log(1 - p_{2k})}{\log(1 - p_{1k})}$$

and hence

$$\log(\mathrm{HR}_k) = \log(-\log(1 - p_{2k})) - \log(-\log(1 - p_{1k})).$$

Denoting the 'complementary log–log' parameter by $\log(-\log(1 - p_{1k})) = \psi_k$ leads to the following likelihood:

$$r_{ik} \sim \mathrm{Bin}[p_{ik},\ n_{ik}], \quad i = 1, 2,$$

$$\log(-\log(1 - p_{1k})) = \psi_k,$$
$$\log(-\log(1 - p_{2k})) = \psi_k + \log \mathrm{HR}_k.$$

We consider three models: (a) fixed effects assuming independent intervention effects $\mathrm{HR}_k$; (b) common effect in which $\mathrm{HR}_k = \mathrm{HR}$; and (c) random effects. The random-effects analysis with quality weights described in Section 8.4 leads to the model

$$\log(\mathrm{HR}_k) \sim \mathrm{N}\left[\log(\overline{\mathrm{HR}}),\ \frac{\tau^2}{q_k}\right],$$

where $\overline{\mathrm{HR}}$ is the overall estimate of the revision hazard ratio pooled across studies.

*Prior distributions:* For the fixed and common effects, independent uniform prior distributions are placed over the study effects $\psi_k$ and $\log(\mathrm{HR}_k)$ or $\log(\mathrm{HR})$. For the random-effects model, three studies do not provide sufficient evidence to accurately estimate the between-study standard deviation $\tau$, and so substantial prior judgement is necessary. We would expect considerable heterogeneity in revision rates between studies, even if they are internally unbiased, and so assume $\tau$ has a normal distribution with mean 0.2 and standard deviation 0.05 (approximate 95% interval 0.1 to 0.3), corresponding to expecting $\pm$50% variability in true hazard ratios between studies, with 95% uncertainty limits of 20% to 80% variability (e.g. at the upper end of the interval, $e^{1.96\times 0.3} = 1.8$ or $\pm$ 80% variability in hazard). Our knowledge of the potential biases of registries and case series suggests downweighting the non-randomised evidence. As a baseline assumption for the quality weights we take $q_k$ equal to 0.5, 1.0 and 0.2 for the registry, RCT and case series studies, respectively. This corresponds to assuming that 'bias' in the registry and case series studies leads to a two- or fivefold increase in the revision rate variance, respectively, over and above the between-study variability expected for RCTs.

*Computation/software:* MCMC methods implemented using WinBUGS.

*Bayesian interpretation:* The results of the evidence synthesis are given in Table 9.6. The 'fixed-effects' estimates of the hazard ratio for each source are shown in the first three rows, revealing reasonable concordance between the non-randomised studies but with the randomised trial showing some evidence against the Stanmore. Forcing a common hazard ratio leads to the registry overwhelming the other sources (row 4 of Table 9.6). The results of a baseline random-effects analysis, with quality weights 0.5, 1, 0.2, are shown in row 5 of Table 9.6, with the hazard ratio estimated in favour of the Stanmore but with the 95% interval only just excluding 1.

Feeding these simulated parameter values into the cost-effectiveness model developed in Example 9.3 provides the estimated incremental changes in benefits and costs associated with a Stanmore rather than a Charnley prosthesis shown in Table 9.7. The estimated expected benefit is somewhat marginal, equivalent to 21 additional days ($0.0579 \times 365$) of discounted quality-adjusted survival, but the CEAC suggests reasonable confidence of cost-effectiveness provided one is willing to pay more than around £10 000 per QALY.

*Sensitivity analyses:* As a sensitivity analysis, we consider two other choices of quality weights. First, we can further downweight all non-randomised evidence by taking $q_k$ equal to 0.1, 1.0 and 0.05, respectively, which leads to an equivocal result with substantial uncertainty, as shown in Table 9.6. At the opposite extreme, setting all quality weights to 1 permits the domination of the registry data, leading to increased benefit.

The sensitivity of the final conclusions to the choice of quality weights is examined in Figure 9.7(a), which also illustrates the sensitivity to two different discount rates for health: 0% and 6%. It is clear that the choice of quality weights has a much stronger influence than the discount rates:

**Table 9.7** Incremental changes in expected benefits and costs associated with using Stanmore rather than Charnley prostheses in men aged 65–74, assuming a synthesis of evidence using quality weights (0.5, 1.0, 0.2) for registry, RCT and case series data, respectively. INB($K$) is the incremental net benefit per patient when the maximum acceptable cost per unit of effectiveness is $K$, and $Q(K) = \text{P}(\text{INB}(K) > 0)$ is the CEAC. Costs are discounted at 6% per annum, benefits at 1.5% per annum.

| Parameter | Median | Prediction 95% interval |
|---|---|---|
| Incremental change in expected life-years | 0.0026 | 0.0001 to 0.0049 |
| Incremental change in expected QALYs | 0.0579 | 0.0007 to 0.1078 |
| Incremental change in expected costs | 219 | 87 to 372 |
| INB(5 000) | 71 | −362 to 452 |
| INB(10 000) | 360 | −352 to 991 |
| INB(15 000) | 649 | −344 to 1529 |
| $Q$(5 000) | 0.66 | |
| $Q$(10 000) | 0.87 | |
| $Q$(15 000) | 0.92 | |

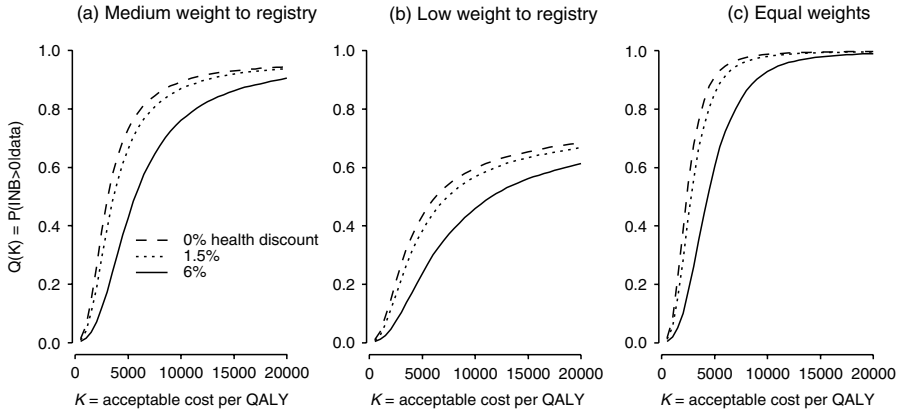


**Figure 9.7** CEACs for a Stanmore compared to a Charnley prosthesis. (a) corresponds to the baseline analysis with quality weights (0,5, 1.0, 0.1) for registry, RCT and case series data, respectively, showing limited sensitivity to the annual discount rate for health benefits. (b) uses quality weights of (0.1, 1.0, 0.05); substantial down-weighting the non-randomised evidence prevents a strong conclusion of cost-effectiveness. (c) weights all sources equally, and the increased role of the registry data leads to a high probability of cost-effectiveness.

if the non-randomised evidence is substantially downweighted (Figure 9.7(b)) the CEAC shows poor evidence for cost-effectiveness regardless of $K$, while equal weighting (Figure 9.7(c)) shows strong evidence for moderate $K$, even when discounting costs at 6%.

### 9.9.2 Comparison of integrated Bayesian and two-stage approach

To recap on Section 9.4, the integrated approach to evidence synthesis and cost-effectiveness analysis simultaneously derives the joint posterior distribution of all unknown parameters from a Bayesian probability model, and propagates the effects of the resulting uncertainty through the predictive model underlying the cost-effectiveness analysis. In contrast, the 'two-stage' approach would first carry out the evidence synthesis, summarising the joint posterior distribution parametrically, and then in a separate analysis use this as a prior distribution in a probabilistic sensitivity analysis in the cost-effectiveness model.

The advantages of the integrated approach include the following. First, there is no need to assume parametric distributional shapes for the posterior probability distributions, which may be important for inferences for smaller samples. Second, and perhaps more important, the appropriate probabilistic dependence between unknown quantities is propagated (Chessa *et al.*, 1999), rather than assuming either independence or being forced into, for example, multivariate normality. This can be particularly vital when propagating inferences which are likely to be strongly correlated, say when considering both baseline levels and treatment differences estimated from the same studies.

The disadvantages of the integrated approach are its additional complexity and the need for full MCMC software. The 'two-stage' approach, in contrast, might be implemented in a combination of standard statistical and spreadsheet programs. However, experience with such spreadsheets suggests that they might not be particularly transparent for complex problems, due to clumsy handling of arrays and opaque formula equations.

## 9.10 COST-EFFECTIVENESS OF CARRYING OUT RESEARCH: PAYBACK MODELS

### 9.10.1 Research planning in the public sector

Any organisation funding clinical trials must make decisions concerning the relative importance of alternative proposals, and hence there have been increased efforts to measure the potential 'payback' of expenditure on research. Buxton and Hanney (1998) review the issues and propose a staged

semi-quantitative structure, while Eddy (1989) suggested a fully quantitative model based on assessing the future numbers to benefit and the expected benefit, with a subjective probability distribution over the potential benefits to be shown by the research. However, Eddy's limited approach was not adopted by its sponsors, the US Institute of Medicine, who preferred a more informal method that employed weights.

It is clearly possible to extend this broad approach to increasingly sophisticated models within a Bayesian framework, and Hornberger and Eghtesady (1998) state that 'by explicitly taking into consideration the costs and benefits of a trial, Bayesian statistical methods permit estimation of the value to a health care organisation of conducting a randomised trial instead of continuing to treat patients in the absence of more information'. Clearly this is a particular example of a decision-theoretic Bayesian approach, applied at the planning stage of a trial (Section 6.5) rather than at interim analyses (Section 6.6.4). Examples include Detsky (1985), Hornberger *et al.* (1995) and Hornberger and Eghtesady (1998) and others who explicitly calculate the expected utility of a trial in order to select sample sizes; such calculations can also, in theory, be used to rank studies that are competing for resources, and hence to decide whether the trial is worth doing in the first place.

The early analysis by Detsky (1985) assumed that a trial would need to achieve statistical significance in order to have an impact on future treatments, but Claxton (1999b) strongly argues that dependence on such inferential methods, whether classical or Bayesian, will lead to sub-optimal use of health resources. He recommends a full decision-theoretic approach to both fixed (Claxton and Posnett, 1996) and sequential (Claxton, 1999b) trials, basing his analysis on quantifying the expected benefit of further experimentation. This *value of information* approach is outlined briefly in Section 9.10.3.

### 9.10.2 Research planning in the pharmaceutical industry

Given the 'bottom line' of profitability in the pharmaceutical industry, it is natural to attempt to apply a decision-theoretic approach to individual trial design, designing a research programme for a specified intervention, and for selecting among competing research opportunities. Many of these ideas have already been discussed in the context of individual clinical trials, but here we are concerned with the 'corporate' context: a whole research programme in which there are multiple competing projects at different stages of drug development. Bergman and Gittins (1985) review quantitative approaches to planning a pharmaceutical research programme. Many of the proposed methods are sophisticated uses of bandit theory (Section 6.10) in order to allocate resources in a dynamically changing environment, but Senn (1996, 1997b) suggests a fairly straightforward scheme based on the Pearson index, which is the expected net present value divided by expected net present costs. He discusses the difficulties

of eliciting suitable probabilities for the success of each stage of a drug development programme, conditional on the success of the previous stage, but suggests that formal Bayesian approaches involving subjective probability assessment and belief revision should be investigated in this context.

An integral part of this process is a realistic assessment of the chances of regulatory approval, and subsequent sales in the light of future competition and so on: although there must inevitable be a degree of speculation in these assessments, it still seems preferably to have explicit recognition of the relevant uncertainties when making decisions as to whether to pursue a particular development programme.

### 9.10.3 Value of information

Suppose we are deciding whether to adopt treatment 1 or treatment 2 as a policy, and wondering whether to fund further research to more accurately determine their relative advantages. The true costs and effectiveness are denoted by $\theta$. Based on current information, the incremental net benefit INB($\theta$) is positive for $\theta$ in a region $\Theta_2$, where treatment 2 would be preferred, and negative for $\theta$ in $\Theta_1$, where treatment 1 would be preferred. We do not know $\theta$, but suppose that we have a current posterior for which $\mathrm{E}[\mathrm{INB}(\theta)|\mathrm{data}] > 0$ and so, on balance, treatment 2 is preferred. If, in fact, $\theta$ is in $\Theta_2$ then we have made the right decision and there is no gain in knowing the exact value of $\theta$, whereas if $\theta$ is truly in $\Theta_1$ we have made the wrong decision and stand to lose $-\mathrm{INB}(\theta)$. The *value of perfect information*, VPI($\theta$), is defined as the amount we would gain by knowing $\theta$ exactly: VPI($\theta$) is 0 when $\mathrm{INB}(\theta) > 0$, and $-\mathrm{INB}(\theta)$ when $\mathrm{INB}(\theta) < 0$, which can be expressed as

$$\mathrm{VPI}(\theta) = \max(-\mathrm{INB}(\theta), 0).$$

Hence our expected value of perfect information, EVPI, is

$$\mathrm{EVPI}_2 = \mathrm{E}[\max(-\mathrm{INB}(\theta), 0)|\mathrm{data}], \tag{9.16}$$

where the subscript 2 indicates that treatment 2 is the currently preferred option. By symmetry, the EVPI when $\mathrm{E}[\mathrm{INB}(\theta)|\mathrm{data}] < 0$, *i.e.* when treatment 1 is the preferred option, is

$$\mathrm{EVPI}_1 = \mathrm{E}[\max(\mathrm{INB}(\theta), 0)|\mathrm{data}].$$

This quantity is easy to calculate using MCMC by simulating values of $\theta$, calculating INB($\theta$) and the VPI, and recording its Monte Carlo average over many iterations. However, we shall see in Example 9.5 that care must be taken with the Monte Carlo error.

We can obtain the EVPI in closed form if INB($\theta$) has a normal distribution, and this also sheds some light on the interpretation of this quantity. Suppose

$$\text{INB}(\theta) \sim \text{N}[\mu_I, \tau_I^2],$$

where the standardised statistic is denoted $z_I = \mu_I/\tau_I$; we assume $\mu_I > 0$ and hence treatment 2 is preferred. For simplicity of notation we shall temporarily drop the subscripts and denote INB by $Y$. Then EVPI $= \text{E}[\max(-Y, 0)]$, and therefore

$$\begin{aligned}
\text{EVPI} &= \int_{-\infty}^{0} -y \frac{e^{-(y-\mu)^2/(2\tau^2)}}{\sqrt{2\pi}\tau} dy \\
&= \int_{-\infty}^{-\mu/\tau} (-t\tau - \mu) \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \qquad \text{(substituting } t = (y-\mu)/\tau) \\
&= -\tau \int_{-\infty}^{-z} t \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt - \mu\Phi(-z) \\
&= \tau \left[ \frac{e^{-z^2/2}}{\sqrt{2\pi}} - z\Phi(-z) \right].
\end{aligned} \tag{9.17}$$

The expression in square brackets is denoted $L(z)$ and is known as the 'unit normal loss function' (Claxton *et al.*, 2000). Figure 9.8 shows $L(z)$ plotted against the 'tail area' $\Phi(-z)$: the latter is $P(\text{INB}(\theta) < 0|\text{data})$, the posterior probability that the wrong treatment is being preferred. The direct relationship in Figure 9.8 reveals that $L(z)$ is qualitatively equivalent to the tail area (being around 30–50% of its value in the region of interest), and hence EVPI in (9.17) is, approximately, proportional to the probability of making a wrong preference, weighted by $\tau$, which reflects the potential importance of drawing a wrong conclusion. We also note that when $z_I = 0$, which occurs when $K$ achieves its breakeven point, the EVPI reaches its maximum of $\tau/\sqrt{2\pi}$.

In terms of applying the EVPI to a population of current and future patients over the time horizon of a health-care intervention ($T$), the EVPI requires an adjustment to account for the incidence $I_t$ of patients in each time period $t$ and the discount rate $\delta_c$, so that

$$\text{EVPI}_{\text{POP}} = \text{EVPI} \times \sum_{t=1}^{T} \frac{I_t}{(1+\delta_c)^{t-1}}, \tag{9.18}$$

assuming no discounting in the first period.

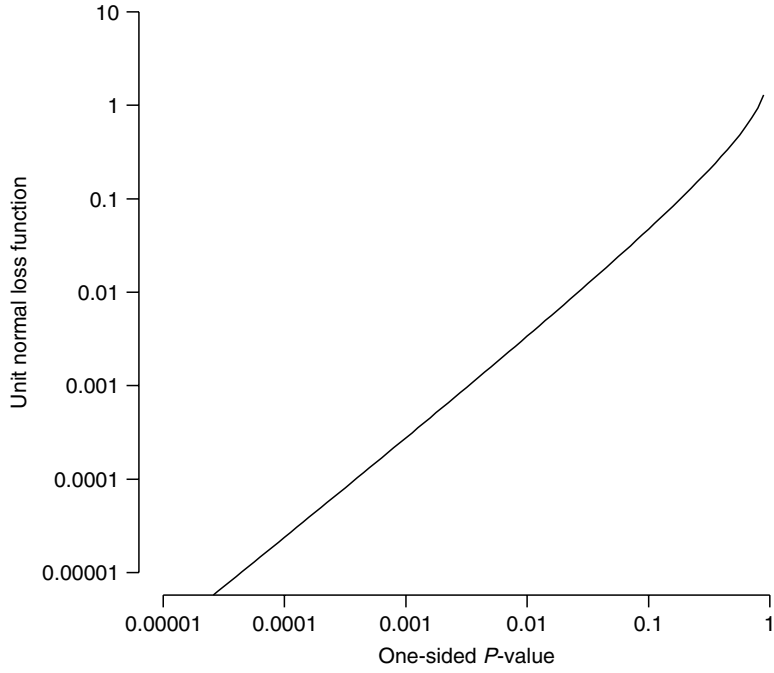


**Figure 9.8** Plot of 'unit normal loss function' against $P$: the EVPI is the unit normal loss function multiplied by the standard deviation of the incremental net benefit.

---

**Example 9.5** *HIV (continued): Calculating the expected value of perfect information*

*Reference:* Ades and Cliffe (2002) – see Example 8.7.

*Costs and utilities:* Ades and Cliffe (2002) specify the cost per test as $T = 3$, and the net benefit $K$ per maternal diagnosis is judged to be around £50 000, with a range of £12 000 to £60 000. In this instance there is explicit net monetary benefit from maternal diagnosis and so it may be reasonable to take $K$ as an unknown parameter, and Ades and Cliffe (2002) perform a *probabilistic* sensitivity analysis by giving $K$ a somewhat complex prior distribution. In contrast, we prefer to continue to treat $K$ as a willingness to pay for each unit of benefit, and therefore follow previous examples and conduct a *deterministic* sensitivity analysis in which $K$ is varied up to £60 000.

The prenatal population in London is $N = 105\,000$, and hence the annual incremental net benefit is

$$\text{INB} = N(1 - a - b)(Ke(1 - h) - T(1 - eh)).$$

We can also calculate the CEAC, given by $Q(K) = P(\text{INB} > 0|\text{data})$.

Finally, we consider the calculation of the EVPI, as defined by (9.16). This is calculated in two ways: first, using MCMC methods; and second, by assuming a normal approximation to the posterior distribution of INB($K$) and using (9.17). Taking a 10-year horizon and discounting at 6% per year gives a multiplier of 7.8 (not discounting the first year) in (9.18).

*Bayesian interpretation:* Following the findings in Example 8.7, the analysis is conducted without data source 4. Figure 9.9(a) shows the normal approximations to the posterior distributions of INB for different values of $K$. The expected INB and 95% limits are shown in Figure 9.9(b) for $K$ up to £60 000, indicating that the policy of universal testing is preferred on balance provided that the benefit $K$ from a maternal diagnosis is greater than around £10 000; $K$ is certainly judged to exceed this

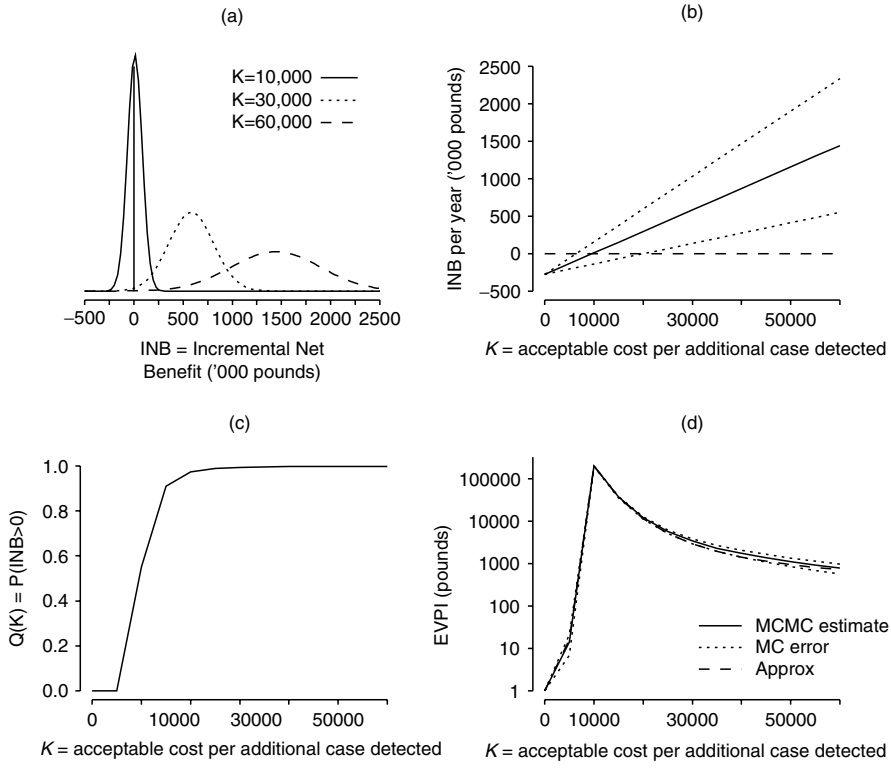


**Figure 9.9** (a) and (b) show incremental net benefits, (c) cost-effectiveness acceptability curve, and (d) expected value of perfect information for universal versus targeted prenatal testing for HIV. Note that the EVPI is maximised at the threshold value of $K$ at which the optimal decision changes.

value. The CEAC in Figure 9.9(c) points to a high probability of universal testing being cost-effective for reasonable values of $K$. Figure 9.9(d) shows the EVPI ($\pm 2$ Monte Carlo errors) calculated using 100 000 MCMC iterations and also using the normal approximation to the distribution of INB and (9.17). The Monte Carlo error is considerable even after 100 000 iterations and care must clearly be taken when using MCMC to calculate the EVPI. Nevertheless, (9.17) provides an adequate approximation. The EVPI is substantial for low values of $K$, but for values around £50 000 the EVPI is negligible. Hence, there appears to be little purpose in further research to determine the parameters more accurately.

---

The EVPI is intended for use in deciding whether to pursue a research programme, how to design it, and when to stop. First, the EVPI must be higher than the cost of research in order to pass the first 'hurdle' for a proposed programme to overcome, and this should continue to hold throughout the programme. Roughly, when the chance of making a wrong decision, weighted by its consequences, is sufficiently low then the programme can stop and a firm recommendation can be made. Another element of a value of information approach to research planning is that of partial expected value of perfect information (PEVPI), which considers each parameter in the cost-effectiveness analysis in turn, and thus informs the decision whether to conduct future research to yield more precise estimates of particular parameters. Claxton *et al*. (2001) provide a worked example.

In practice, no further research is going to lead to perfect information. Hence, the most relevant quantity may be the expected value of sample information (EVSI), which is essentially the EVPI allowing for the sampling error of a trial. This must exceed the sample costs to overcome the hurdle for a specific proposed trial, and the EVSI minus sample costs is known as the expected net benefit from sampling (ENBS). This model allows for unbalanced allocation of patients between arms, and the ability to revise design based on interim analyses (Claxton and Thompson, 2001; Claxton *et al*., 2001), in order to optimise the ENBS. Felli and Hazen (1998, 1999) extend this utility perspective to sensitivity analysis, suggesting that an analysis should be considered sensitive to a particular uncertain input if the expected gain in utility from eliminating the uncertainty about that input exceeds a certain specified threshold.

## 9.11 DECISION THEORY IN COST-EFFECTIVENESS ANALYSIS, REGULATION AND POLICY

The debate about the formal role of decision theory in policy-making is continuing, and here we briefly run through some arguments for and against. Claims for its use include the following:

- Decision theory and economic argument clearly state that maximised expected utility is the sole criterion on which to choose between two options. Therefore measures of 'significance', posterior tail areas of incremental net benefit, and high probabilities on a CEAC are all irrelevant. (Claxton and Posnett, 1996). Claxton *et al.* (2000) point out that 'Once a price per effectiveness unit has been determined, costs can be incorporated, and the decision can then be based on (posterior) mean incremental net benefit measured in either monetary or effectiveness terms'.
- To maximise the health return from the limited resources available from a health budget, health-care purchasers should use rational resource allocation procedures. Otherwise the resulting decisions could be considered as irrational, inefficient and unethical.
- Uncertainty is taken into account through evaluating the benefit of further experimentation, as measured by a value of information analysis.
- This framework provides a formal basis for designing trials, assessing whether to approve an intervention for use, deciding whether an intervention is cost-effective, and commissioning further research.
- Specifying all necessary values may be difficult, but it is necessary for rational decision-making. Claxton (1999b) suggests the first step should be to establish a normative framework that best meets the needs of a system, and separately to conduct studies to see how to get the research into practice.

Among the arguments against are the following:

- The standard criticisms of decision-theoretic approaches to trials apply (Section 6.2): in particular, it is not realistic to specify a full model for the possible impact of research results (which may not even be 'significant') on clinical practice.
- The idea of a null hypothesis (the status quo), which lies behind the use of 'statistical significance' or posterior tail areas, is fundamentally different from that of an alternative hypothesis (a novel intervention). The consequences and costs of the former are generally established, whereas the impact of the latter must contain a substantial amount of judgement. Often, therefore, a choice between two treatments is not a choice between two equal contenders to be decided solely on the balance of net benefit – some convincing evidence is required before changing policy.
- A change in policy carries with it many hidden penalties: for example, it may be difficult to reverse if later found to be erroneous, and may hinder the development of other, better innovations. It would be difficult to explicitly model these phenomena with any plausibility.
- Value of information analysis is dependent on having the 'correct' model, which is never known and generally cannot be empirically checked. Sensitivity analysis can only compensate to some extent for this basic ignorance.

## 9.12 REGULATION AND HEALTH POLICY

### 9.12.1 The regulatory context

Regulatory bodies have a duty to protect the public from unsafe or ineffective therapies. Opinions on the relevance of Bayesian methods to drug or device regulation cover a broad spectrum: Whitehead (1997b, p. 204) and Koch (1991) see any use of priors as being controversial and inappropriate, while on the other hand Matthews (1998) claims that the use of sceptical priors 'should not be optional but mandatory'. Keiding (1994) criticises the 'ritual dances' currently prescribed for regulation, but wonders whether Bayesian methods will allow anything less ridiculous. O'Neill (1994), as a senior US Food and Drug Administration (FDA) statistician, acknowledges the appropriate conservatism arising out of the use of sceptical priors, and considers that Bayesian methods should be investigated in parallel with other techniques.

The full decision-theoretic approach (Section 9.11) takes an even more radical perspective. Claxton (1999a) and Claxton *et al.* (2000) suggest that agencies use decision theory for regulation, and evaluate the expected value of further investigation in order to assess whether sufficient evidence is available to permit approval. The crucial idea is that current demands for statistical significance (e.g. two independent studies with $P < 0.05$) is an inadequate criterion as it takes no account of the potential population at risk, the potential consequences of inappropriate approval, and the costs of obtaining more evidence.

### 9.12.2 Regulation of pharmaceuticals

The website of the FDA allows one to search for references to Bayesian methods among their published literature (Section A.2), although much of the discussion concerns medical devices (see Section 9.12.3). Guidelines for population pharmacokinetics are provided (US Food and Drug Administration, 1999a), which can be thought of as an empirical Bayes procedure (Section 6.12). There is also an interesting use of a Bayesian argument in the approval of the drug enoxaparin (Lovenox). The transcript of the Cardiovascular and Renal Drugs Advisory Committee meeting on 26 June 1997 (US Food and Drug Administration, 1986, pp. 212–218) shows the pharmaceutical company had been asked to make a statement about the effectiveness of enoxaparin plus aspirin as compared to placebo (aspirin alone), whereas their clinical trial had used an active control of heparin plus aspirin. They therefore used meta-analysis data comparing heparin plus aspirin with aspirin alone in order to produce a posterior distribution on the treatment comparison of interest: an example of indirect-comparison inference (Section 8.3). Analyses were repeated using the meta-analysis data directly, but also expressing scepticism about its relevance and

reducing its influence, with results being expressed as posterior probabilities of treatment superiority over placebo. The committee welcomed this analysis and voted to approve the drug.

It is important to note that the latest international statistical guidelines for pharmaceutical submissions to regulatory agencies state that 'the use of Bayesian and other approaches may be considered when the reasons for their use are clear and when the resulting conclusions are sufficiently robust' (International Conference on Harmonisation E9 Expert Working Group, 1999). Unfortunately they do not go on to define what they mean by clear reasons and robust conclusions, and so it is still open as to what will constitute an appropriate Bayesian analysis for a pharmaceutical regulatory body.

### 9.12.3 Regulation of medical devices

The greatest enthusiasm for Bayesian methods appears to be in the FDA Center for Devices and Radiological Health (CDRH). They co-sponsored a workshop on Bayesian methods in November 1998, and have proposed a document *Statistical Guidance on Bayesian Methods in Medical Device Clinical Trials* (US Food and Drug Administration, 1998a).

Campbell (1999) described the potential for Bayesian methods in assessing medical devices, emphasising that devices differed from pharmaceuticals in having better-understood physical mechanisms, which meant that effectiveness was generally robust to small changes. Since devices tended to develop in incremental steps, a large body of relevant evidence existed and companies did not tend to follow established phases of drug development. The fact that an application for approval might include a variety of studies, including historical controls and registries, suggests that Bayesian methods for evidence synthesis might be appropriate. However, the standard conditions apply that the source and robustness of the prior information must be assessed, and that Bayesian analysis does not compensate for poor science and poor experimental design.

Campbell drew attention to the Transcan Breast Scanner, which was approved by the CDRH in April 1999 (US Food and Drug Administration, 1999b). A primary 'intended use' study on 72 women was supplemented by two additional studies of differing designs, using a hierarchical multinomial logistic regression model with study introduced as a random effect. MCMC simulation methods were used by means of the BUGS software. Searching the FDA website reveals a growing number of device submissions that exploit Bayesian reasoning.

## 9.13 CONCLUSIONS

In this chapter we have attempted to explore a range of concerns that arise in cost-effectiveness modelling, but acknowledge that there are a number of issues

that we have passed over. In particular, we have not explored the sensitivity of the conclusions to 'ignorance' (Section 9.7) about the structure of the appropriate model: alternative models that could be used in this context include survival-type models with competing risks. It is vital to admit that even a reasonably complex model, such as that investigated in our example, cannot be assumed to be realistic and must be subject to careful criticism (Russell, 1999; Sculpher *et al.*, 2000).

As attempts are made towards evidence-based health policy in both clinical and public health contexts, models will inevitably become more complex and, while the methods described in this chapter may appear complicated, we feel that techniques such as these may well become commonplace in the future. If decisions made with the help of such analyses are to be truly accountable, it is important that the models and methods are transparent, easily updatable, and can be run by many parties in order to check sensitivity. Models implemented in spreadsheet programs have some of these characteristics, but we feel that user-friendly Bayesian simulation programs could contribute substantially to the field.

## 9.14 KEY POINTS

1. A Bayesian approach allows explicit recognition of multiple perspectives from the stakeholders involved.
2. Cost-effectiveness analyses fall naturally into a Bayesian framework, whether or not the evidence synthesis is carried out separately (the two-stage approach) or integrated in with the cost-effectiveness analysis.
3. Comprehensive decision modelling is likely to become increasingly important in making both healthcare and policy decisions.
4. Increased attention to pharmacoeconomics may lead decision-theoretic models for research planning to be explored, although this will not be straightforward.
5. There appears to be great potential for formal methods for planning in the pharmaceutical industry.
6. The regulation of devices is leading the way in establishing the role of evidence synthesis.
7. We expect this to be a significant area of research activity over the coming years.

## EXERCISES

9.1. Consider the TACTIC study described in Example 9.2, and suppose we try to use the simple bivariate normal model of Section 9.5 to analyse this problem.

(a) Run the WinBUGS code for Example 9.2, and record the posterior correlation between $\theta_e$ and $\theta_c$ under the exchangeable model.
(b) Plot the joint posterior samples for $\theta_e$ $\theta_c$ and check whether bivariate normality might be a reasonable assumption.
(c) Making this assumption, use the methods of Section 9.5 to estimate the CEAC and INB, and hence check whether these analytical methods yield similar conclusions to those used in Example 9.2.

9.2. Gray *et al.* (2002) report the results of an economic analysis carried out alongside an RCT to evaluate the use of an intensive blood glocose control policy in patients with type 2 diabetes. Table 9.8 reports the results of the trial in terms of both costs and event-free years. They differentiate between the actual costs observed during the trial, and those adjusted for the fact that during the trial patients required additional clinical visits, and thus incurred additional costs above those seen in routine clinical practice. The latter estimate of costs is referred to as *non-trial*. Using the methods of Section 9.5, examine whether the policy of intensive glucose control is cost-effective for the different scenarios summarised in Table 9.8, *i.e.* whether to use trial costs or adjusted trial costs and/or whether to discount either costs or costs and life-years. Gray *et al.* (2002) did not report the correlation between costs and life-years, so consider assessing cost-effectiveness either (a) assuming specific values for the correlation $\rho$, or (b) placing a suitable prior distribution on $\rho$.

9.3. Consider the case of whether to use prophylactic antibiotics for women undergoing Caesarean sections described in Exercise 3.13. The problem may be formulated as a cost-effectiveness decision model and evaluated using WinBUGS, taking into account sources of uncertainty.

The odds ratio for infection (antibiotics vs. control) is estimated to be 0.40 (95% CI from 0.33 to 0.47) from a Cochrane systematic review, while the probability of wound infection without antibiotics is estimated to be

**Table 9.8** Mean costs (£ at 1997 prices) and event-free life-years for intensive and conventional blood glucose control in patients with type 2 diabetes.

| | Discount rate | Intervention ($n = 2729$) | | Control ($n = 1138$) | | Difference | |
|---|---|---|---|---|---|---|---|
| | | Mean | SD | Mean | SD | Mean | 95% CI |
| Costs (£) | | | | | | | |
| Total trial | 0% | 9608 | 8343 | 9869 | 120 222 | −261 | −1027 to +505 |
| | 6% | 6958 | 5774 | 7170 | 8 689 | −212 | −761 to +338 |
| Total non-trial | 0% | 8349 | 8153 | 7871 | 11 841 | +478 | −275 to +1232 |
| | 6% | 6027 | 5674 | 5689 | 8 615 | +338 | −207 to +882 |
| Event-free years | | | | | | | |
| Within trial | 0% | 14.89 | 6.93 | 14.29 | 7.06 | +0.60 | +0.12 to +1.10 |
| | 6% | 9.17 | 3.20 | 8.88 | 3.44 | +0.29 | +0.06 to +0.53 |

**Table 9.9** RCTs evaluating the effectiveness of using prophylactic antibiotics for women undergoing elective Caesarean sections in terms of infection rates. (Study quality: A=Good, B=OK, C=Poor.)

| | | Antibotics | | Control | | Study |
|---|---|---|---|---|---|---|
| Study | Year | Infections | Total | Infections | Total | quality |
| Dashow | 1986 | 3 | 100 | 0 | 33 | A |
| De Boer | 1989 | 1 | 11 | 5 | 17 | B |
| Duff | 1982 | 0 | 42 | 0 | 40 | B |
| Jakobi | 1994 | 4 | 167 | 5 | 140 | B |
| Lewis | 1990 | 1 | 36 | 1 | 25 | B |
| Mahomed | 1988 | 12 | 115 | 15 | 117 | A |
| Rothbard | 1975 | 0 | 16 | 1 | 16 | C |

0.08, based on observing 60 infections in 750 women. The costs of administering antibiotics include a fixed cost of £10 plus between 4 and 7 minutes of consultant's time at £1 per minute. The hospital costs for Caesarean section without infection are £173 per day, and the average length of stay is 6.7 days (SE 0.33). If there is infection, the average length of stay rises to 8.8 days (SE 0.55) and the daily cost to £262. Utilities are assumed known at 0.95 QALYs without infection and 0.80 QALYs with infection.

(a) Obtain an algebraic expression for the incremental net benefit of using antibiotics for various choices of $K$, the acceptable cost per QALY.

(b) Use the information provided above to obtain the posterior distributions for the INB, and hence plot the cost-effectiveness acceptability curve.

9.4 Extend the model in Exercise 9.3 to take account of the actual meta-analysis of RCTs considering only elective Caesarean sections presented in Table 9.9 (Cooper *et al.*, 2002). Explore the sensitivity to downweighting studies according to their assessed quality.

9.5 In Example 9.5, Ades and Cliffe (2002) carried out a probabilistic sensitivity analysis for $K$, the net benefit of a maternal diagnosis. They adopted a distribution representing an estimate of £50 000, with a range from £12 000 to £60 000.

(a) What might be a suitable functional form for a prior distribution with these qualities?

(b) With such a prior distribution, carry out a probabilistic sensitivity analysis and estimate the incremental net benefit, the probability of cost-effectiveness and the EVPI.

9.6 In Example 9.4, what would be the effect of including a (hypothetical) additional randomised trial in which 28/400 (7%) of Charnley prostheses had needed revision, compared to 16/400 (4%) of Stanmore?