1

Introduction

This book is at heart about a fusion of ideas from medical statistics, clinical epidemiology, decision analysis, and health economics. These are distinct academic disciplines, each with its own history and perspectives. Recently, though, developments within these separate fields, together with political changes across the developed world in attitudes to funding public services, have conspired to bring particular strands of thought from each of these disciplines together in a fruitful and compelling way. Before describing the contents of the book, we shall therefore start with a look at these wider developments. This contextual background has two purposes. It will give readers from each discipline an essential introduction to the others, but it will also explain why this book has come about.

1.1 The rise of health economics

The last 20 years has witnessed an enormous change in how new medical technology is deployed. Previously, clinicians were free to treat their patients in whatever way seemed best. The perceived efficacy or treatments, taking account of course the risk of adverse side effects, was the dominant factor. In countries with a centralised state-funded health service there were clearly budget constraints, which could be realised through limits on the numbers of hospitals, doctors or nurses. But the idea that health care itself should be *rationed* would not only have been regarded as politically infeasible, but was rejected by the medical professions as a threat to 'clinical freedom'.

Clinical freedom, placing no limit on the costs of health care, is an open invitation to pharmaceutical and medical device manufacturers to enter the market place with a stream of genuinely new products and 'me-too' products. Eventually, both nationalised and insurance-based health systems have had to find ways of curtailing

Evidence Synthesis for Decision Making in Healthcare, First Edition. Nicky J. Welton, Alexander J. Sutton, Nicola J. Cooper, Keith R. Abrams and A.E. Ades.

clinical freedom by considering not only the clinical efficacy of new procedures, but also their economic implications. The numbers of economic evaluations has mushroomed accordingly, and in many countries including the UK, Netherlands, France and Germany, the introduction of new clinical treatments as well as new screening programmes is routinely accompanied by some form of economic analysis.

Economic analyses have been subdivided into *cost-benefit, cost-effectiveness* and *cost-utility analyses*. All compare the costs of an intervention with the 'outcomes'. Cost-benefit considers benefits as seen by the recipients of the intervention, and as they would be evaluated in cash terms. Cost-effectiveness studies compare costs and 'effects' measured for example as the number of patients cured, or the average weight-loss measured in kilograms. Cost utility is similar, but the 'effects' are measured in Quality Adjusted Life Years (QALYs) or Disability Adjusted Life Years (DALYs). The output of all these forms of analysis tends to be a ratio measure of the additional costs of the intervention divided by the additional benefits, effects, or utilities. In cost-effectiveness analysis this is called the Incremental Cost-Effectiveness Ratio (ICER).

There is one particular form of economic analysis, net benefit analysis [1, 2], that has had a major influence on the field. Its advent has opened the way for many of the developments described in this book, but it also reveals a great deal about how public attitudes have changed. Net Benefit analysis unifies the three approaches by translating the utilities into their money 'values' using an exchange rate λ, which is simply the money value attributed to the unit of health gain. If U is the lifetime utility of some intervention and C its lifetime cost, the Net Benefit is: $U\lambda - C$. If we now agree to always use QALYs to represent the measure U, Net Benefit immediately becomes a very powerful decision-making tool:

- First, health planners can, in principle, evaluate any intervention on the same level playing field: dialysis for end-stage renal disease, herceptin for breast cancer, group counselling for smoking cessation, or newborn screening for sickle cell disease.

- Secondly, planners can choose rationally between alternative treatment strategies for the same condition by selecting the one with the highest net benefit, S^*:

$$\begin{aligned} NB(s) &= U(s)\lambda - C(s) \\ S^* &= ArgMax_s(NB(s)) \end{aligned} \tag{1.1}$$

  Here, the function $ArgMax_s$ picks out the strategy s which yields the highest Net Benefit.

- Finally there is its mathematical convenience. We avoid the difficulties associated with ratios of costs and benefits/effects/utilities (for example, how does one interpret a negative ICER?), and instead can deal with a single quantity, Net Benefit, on a well-understood scale. This will become particularly useful when we consider uncertainty (Section 1.2).

Under this regime, health economic evaluation becomes a *modelling* exercise. Based on the evidence available, a model of the disease process under each of the interventions of interest, s, is constructed in order to arrive at values of $U(s)$ and $C(s)$. We are left, in effect, with a *decision analysis* and we will shortly turn to see how the methods routinely used in *operations research* can be used to carry this out.

However, the legitimacy of this entire exercise (and indeed much of this book!) rests on whether defensible values of λ and U can be arrived at. We would be misleading readers who are unfamiliar with health economics if we were anything but frank about the foundations that this book rests on. We shall summarise the situation briefly.

To set λ a decision maker must answer a question that sounds very simple, but which is in fact profoundly difficult: how much am I willing to pay to gain one QALY for a single patient? One way to justify a range of figures is to consider a wide selection of noncontroversial, accepted interventions, and to look at the QALY gain, net of costs, that they appear to confer. Exercises of this sort [3] offer a benchmark range of figures against which proposed new interventions can be assessed. For many years, officials at the Department of Health (London, UK) have considered that interventions that gain one QALY at a cost of £10 000 would be 'a bargain', while those that cost over £30 000 would be 'too expensive', but these figures were never made officially public. However, it is public knowledge that the National Institute for Health and Clinical Excellence (NICE) in the UK, uses a range of $\lambda = £20\,000–30\,000$ as a benchmark to decide whether new technologies should be adopted by the National Health Service. The fact that this is openly known illustrates that, even if the British public are not yet aware that their health is being valued this way, at least British policy makers have owned up to what they are doing. Ten years ago, if these ideas were presented to a British audience there would still be angry challenges from clinicians. Now, the reality of health care rationing in this form is unlikely to raise any complaints. At the same time, it is only fair to acknowledge that the entire basis for decision making that this assumes is far from being accepted by all health economists. To some extent the opposition is based on philosophical considerations underpinning health economic theory [4]. The controversy is also fuelled by the considerable doubts about how QALYs are measured [5]. It is important to recognise that patients, clinicians, decision analysts, decision makers are, quite rightly, openly sceptical about whether existing measures and methods really allow us to put treatments for Alzheimer's disease on the same basis as cancer drugs, or whether cancer treatment can be in any way equated to newborn screening. But, in spite of these doubts, decisions still have to be made, and if they cannot be made on a perfect basis, they must be made on the best basis that current methods will permit. The attempt to create the 'level playing field' may not have succeeded in flattening all the dips and lumps in the grass – it may never do so. However, the commitment to the aspiration may be more important than the success – or even the feasibility – of its execution.

1.2 Decision making under uncertainty

1.2.1 Deterministic models

Once the concept of Net Benefit is accepted, decision modelling *in principle* becomes an essentially mechanical exercise. (Only in principle, as the question of where the parameter values come from – a major focus of this book – is far from simple.) The Net Benefit of each alternative strategy is calculated and the decision maker simply chooses the strategy with the highest net benefit. The nature of the task can be made plain by a *decision tree*. The example in Figure 1.1 shows a *choice node* between two strategies for managing a particular condition. The choice node is depicted as a small square. Each strategy corresponds to a treatment. Standard Treatment 1 has a probability p_1 of success, new Treatment 2 has probability p_2. These probabilities would hopefully be based on evidence from Randomised Controlled Trials (RCTs) (see Section 1.3). To complete the tree we need only add the (lifetime) costs C_1, C_2 and QALYs attaching to treatment success U^+ and treatment failure U^-.

Note the assumption that the QALYs attaching to each strategy are based solely on the proportion of patients experiencing 'success', not on the treatment itself. This assumption is not required.

$$\begin{aligned} NB_1 &= \lambda[p_1U^+ + (1-p_1)U^-] - C_1 \\ NB_2 &= \lambda[p_2U^+ + (1-p_2)U^-] - C_2 \end{aligned} \tag{1.2}$$

Economic evaluation (Section 1.1) is concerned with a comparative analysis of the available options. The critical measure is therefore the Incremental Net Benefit (INB). This leads to an important simplification:

$$\begin{aligned} INB &= NB_2 - NB_1 \\ &= \lambda[(p_2 - p_1)(U^+ - U^-)] - (C_2 - C_1) \\ &= \lambda\Delta_p\Delta_U - \Delta_C \end{aligned} \tag{1.3}$$

The INB of new Treatment 2 compared with standard Treatment 1 can be expressed in terms of the *difference* Δ_p in the probabilities of successful treatment, the QALY *difference* Δ_U between the success and failure outcomes, and the cost difference Δ_C. The ambitious data collection project required to identify the absolute lifetime costs and lifetime QALYs under each treatment is reduced to the far simpler – though still difficult – task of finding differences between the treatments. The epidemiologist is therefore still in familiar territory. All the parameters can be informed by controlled, i.e. comparative, studies. Δ_p in particular can be informed by an RCT. The decision now depends on whether INB is positive (NB_2 is greater than NB_1, choose new Treatment 2) or not (choose standard Treatment 1).

It is worth noting that the results of the economic evaluation will depend on the treatment difference on the absolute probability scale. This contrasts with the ratio measures – Odds Ratio, Risk Ratio or Hazard Ratio – that are typically reported

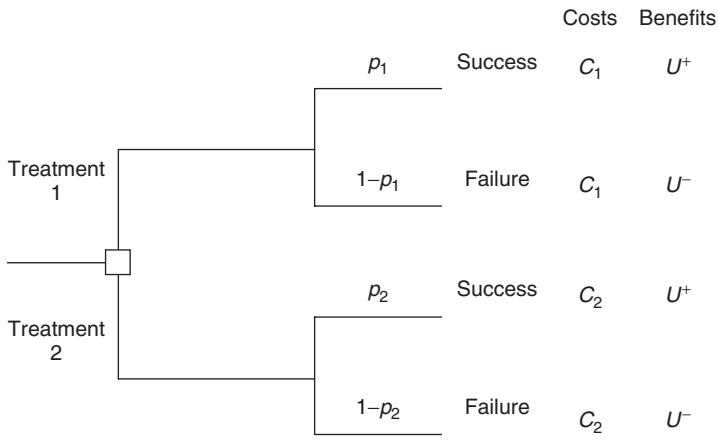

Figure 1.1 A simple decision tree with two treatment options, a binary (success/failure) outcome, and costs C for each treatment and utilities U for each outcome.

in RCTs, and the Log Odds Ratios (LORs), Log Risk Ratios and Log Hazard Ratios commonly used in meta-analysis (see Section 1.3). The decision analyst is therefore likely to need to map from the ratio quantities familiar from clinical epidemiology into the absolute quantities required for economic analysis. For example, one will need to transform information on a LOR δ into information on p_2 at a specified p_1:

$$\begin{aligned} \delta &= \frac{\text{logit}(p_2)}{\text{logit}(p_1)} \\ p_2 &= \frac{\exp(\text{logit}(p_1) + \delta)}{1 + \exp(\text{logit}(p_1) + \delta)} \\ \Delta_p &= \frac{\exp(\text{logit}(p_1) + \delta)}{1 + \exp(\text{logit}(p_1) + \delta)} - p_1 \end{aligned} \tag{1.4}$$

Finally, then, substituting this expression for Δ_p back into Equation (1.3), INB emerges as a function of $p_1, \delta, \Delta_U, \Delta_C$. Given values for these inputs, we can derive values for p_2, and then for Δ_p, and finally for *INB*. This is illustrated in Table 1.1. Given the inputs provided, the INB of Treatment 2 compared with Treatment 1 is £4795. This is above zero, so Treatment 2 emerges as the optimal strategy.

Figure 1.1 represents a particularly simple situation. In applications, the decision trees downstream and upstream of the treatment can become far more extensive. Frequently they may include a Markov model of disease progression (Chapter 10). These consist of a series of states representing different stages of illness. Probability or rate parameters describe the expected numbers of transitions from one state to another in a given time period.

Table 1.1 Deterministic decision analysis.

Parameter	Symbol	Value
Pr(success) on Treatment 1	p_1	0.25
Log Odds Ratio	δ	0.8
Pr(success) on Treatment 2	$p_2 = \frac{\exp(\text{logit}(p_1) + \delta)}{1 + \exp(\text{logit}(p_1) + \delta)}$	0.425897
Difference in Pr(success)	$\Delta_p = p_2 - p_1$	0.175897
QALY gain for successful outcome	Δ_U	2
Incremental cost Treatment 2	Δ_C	£4000
Monetary value of one QALY	λ	£25 000
Incremental Net Benefit	$INB = \lambda\Delta_p\Delta_U - \Delta_C$	£4794.838

1.2.2 Probabilistic decision modelling

The previous section laid out what has been called a *deterministic* decision model. It takes no account of uncertainty in the parameter values. Of course, the analyst is free to vary the input parameters to see whether this will reverse the sign of the INB and thus change the decision. Perhaps the confidence intervals on the original published parameter estimates, $p_1, \delta, \Delta_U, \Delta_C$, would be used to guide the choice of input parameters in a *deterministic sensitivity analysis*. There are a number of difficulties with the deterministic approach to analysing uncertainty [6]. First, there are usually far too many parameters for one to try out all the combinations. Secondly, the decision maker would usually like to know, in view of the uncertainty in the input parameters, whether a decision based on INB is secure. A deterministic procedure does not actually deliver a clear, interpretable metric. Finally, once we admit uncertainty in parameters, a deterministic procedure may not even guarantee to deliver the correct decision, as we will see below.

To take account of parameter uncertainty, each parameter is characterised by a probability distribution. For example, the probability of success on standard treatment p_1 might be characterised as a Beta distribution (see Appendix 2), which is the natural choice for probabilities. If it is based on a particular study where 10 successful outcomes were observed in 40 patients, a Beta(10,30) distribution would be appropriate. Typically, information on p_2 is introduced indirectly, via the LOR δ, as shown in Equation (1.4), δ would be assigned a normal distribution with a mean and variance taken from the meta-analysis. Table 1.2 sets out all the parameters needed to define INB, together with their probability distributions.

Because INB is a function of these parameters, it now also has a probability distribution. It can be shown that, under conditions of uncertainty, the optimal decision is based on the *expected net benefit*. The expectation, of course, is over the parameters $p_1, \delta, \Delta_U, \Delta_C$; in other words, we must average the INB over all the values of the input parameters. Writing INB as a function of parameters,

the decision will be based on the expectation:

$$E[INB(p_1, \delta, \Delta_U, \Delta_C)] \tag{1.5}$$

The simplest way to carry out this integration is via Monte Carlo simulation. Values are drawn from the distributions of each of the input parameters, and INB is evaluated for each set of input parameter values. This is done thousands of times, creating in numerical form a *distribution* of INB (Figure 1.2). The analyst can then compute the mean of all these samples, to be used as an estimate of expected INB (Equation (1.5)), and can also examine the variance and centiles.

Monte Carlo simulation has formed the basis for probabilistic decision analysis [7, 8] for many years, as it has for modelling in general throughout the physical, biological and social sciences. It is important to note, however, that besides providing us with an analysis of *uncertainty*, probabilistic analysis has the further advantage that it gives a correct computation of expected INB under uncertainty, namely Equation (1.5). A deterministic analysis is almost invariably based on the mean values of the input parameters.

$$INB(E[p_1], E[\delta], E[\Delta_U], \Delta_C) \tag{1.6}$$

But the mean of a function of parameters (Equation (1.5)) is not necessarily equal to the same function applied to the parameter means (Equation (1.6)). This 'short-cut' computation of the mean of a function of parameters is correct under certain special circumstances: the function must be *linear* in all parameters (or multi-linear), and the parameters must not be correlated. In this particular example these conditions are not met. INB is linear in p_1 and Δ_U, but it is *not* linear in δ, as in evident from Equation (1.4). In this example the expected INB from the probabilistic analysis is £4871, nearly 2% higher than the deterministic analysis – not a huge amount, but enough to illustrate the point.

We may also compute a probability that the optimal decision based on the evidence inputs is the correct decision: this is simply the area to the right of the point of neutrality, $INB=0$. In our example, this probability is 0.64. There is a 36% chance therefore that the optimal decision is the wrong decision. This statement may sound nonsensical to some: if so, imagine that if all uncertainty was eliminated, the

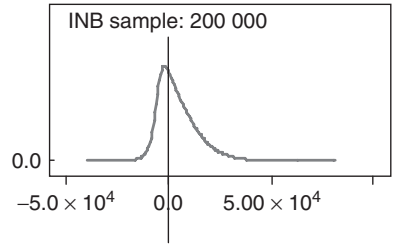

Figure 1.2 Distribution of Incremental Net Benefit, based on 200 000 samples from a Monte Carlo simulation.

Table 1.2 Probabilistic decision analysis, based on 100 000 Monte Carlo simulations. We distinguish between the true mean of the input distribution and the estimate based on the simulation.

Parameter	Symbol	Distribution	Mean	Estimated mean (95% CI) from simulation
Pr(success) on Treatment 1	p_1	Beta(10,30)	0.25	0.2501 (0.13, 0.39)
Log Odds Ratio	δ	Normal(0.8,0.7^2)	0.8	0.8002 (−0.5766, 2.171)
Pr(success) on Treatment 2	$p_2 = \frac{\exp(\text{logit}(p_1) + \delta)}{1 + \exp(\text{logit}(p_1) + \delta)}$			0.4278 (0.1308, 0.7710)
Difference in Pr(success)	$\Delta_p = p_2 - p_1$			0.1776 (−0.08876, 0.4836)
QALY gain for successful outcome	Δ_U	Normal(2,0.5)	2	0.1998 (0.6171, 3.378)
Incremental cost Treatment 2	Δ_C	Constant	£4000	
Monetary value of one QALY	λ	Constant	£25000	
Incremental Net Benefit	$INB = \lambda\Delta_p\Delta_U - \Delta_C$			4871 (−8442, 25120)

variance of INB would be zero, and the probability of making a 'wrong' decision would be zero.

Deterministic sensitivity analysis tends to be used to answer the question 'what if this (these) parameter(s) took this (these) value(s)?', but it fails to pay much attention to how likely the various scenarios are. The probabilistic approach allows us to take the extent of parameter uncertainty much more seriously. The uncertainty in the input parameters is *propagated* faithfully through the model, to be faithfully reflected in the uncertainty in the decision.

1.3 Evidence-based medicine

Evidence-Based Medicine (EBM) has taken the form of a 'movement'. While it would seem to be almost a truism that clinical practice should be based on 'evidence', the EBM movement has grown up alongside the ideas of *systematic review*. The essential insight is that any review of the literature on, for example, the effect of Treatment 2 compared with Treatment 1, must have a defined *protocol*, with a clear rationale for inclusion and exclusion of studies. The aim is to prevent arbitrary selection of studies and put the process of deriving a summary estimate of the treatment effect on an objective and repeatable basis. At the same time, careful and thorough searching of literature would strengthen the summary estimate, reducing uncertainty [9].

The EBM movement has also been concerned with grading the quality of evidence. Direct RCT evidence on clinical end-points is graded as superior to evidence from observational studies, or evidence of surrogate end-points, or indirect evidence of various kinds. A 'hierarchy of evidence' has been established [10]. Further quality assessment is undertaken within each the category of studies: for example, studies of disease prevalence, or studies of diagnostic tests, are also graded on their own specific quality criteria. In general, evidence of the highest grade available is used.

EBM has been a response to the massive increase in the sheer volume of evidence. Its insistence that every systematic review must have a clear *method* governing study inclusion and exclusion, translates directly into the purposes of this book. Putting it into our terms, there must be a clear method underlying the selection of evidence to inform every model parameter. There can be no arbitrary selection or exclusion of evidence. But, having said what we would share with EBM, we must also be explicit about where we differ.

First, the position taken in this book tends to be more inclusive about evidence. We would generally argue that evidence of different sorts should be 'synthesised', that is statistically combined within a single coherent model. Methods for carrying out this kind of synthesis are discussed in Chapters 7–10. While certain studies or types of evidence may be definitely or potentially biased, we would see it as wrong to treat them as totally irrelevant. Our preference would therefore be to include them, but to explicitly account for and adjust biases (see Chapter 11). However, the broader question of what evidence to include or exclude deserves much greater discussion than we can give it in this book.

Secondly, we would probably go further in the analysis of study-specific bias than is typical in a Cochrane review. The desire to make systematic reviews repeatable and objective is worthy, but it has given rise to the unrealistic belief that the review process and the construction of the summary measure can be mechanised to the extent that any computer-literate person can do it. Our view is that no economic analysis can be undertaken without considerable input from individuals with clinical and epidemiological expertise in the relevant areas. The subjective element cannot be eliminated. Indeed, there is no reason to eliminate it: instead – and in fact very much in line with EBM core procedures – the analysis of study-specific bias, and the elicitation of expert opinion could be put on a far more systematic basis [11].

A third, related, point of difference has to do with the concept of 'summary measure'. In Cochrane reviews, the summary measure tends to be seen as a 'summary of the literature', and meta-analysis is simply seen as a process that produces it. Our view is, again, a little different because 'summary of the literature' does not exactly capture what an economic assessment requires. The decision maker has a specific target population and a specific protocol in mind. What the decision analyst therefore requires is an estimate of, for example, the treatment effect in this particular group with this specific protocol. Very possibly, there may be no evidence that addresses this precise question, but there may be quite a lot of evidence from possibly imperfect studies using *similar* protocols in *similar* patients. The evidence synthesis task is then to take account of both internal biases, and the possibility that the target parameter may differ from the parameters estimated in the studies available. Input from knowledgeable experts and clinicians is obviously essential. At the same time of course, meta-analysis is not just a procedure that is applied to a bunch of studies to derive a summary: it is a set of statistical models. Statisticians have a large toolbox for evaluating models, and we will be opening this box regularly in Chapters 6 and 8–10. This introduces a statistical element to decision modelling that most decision modellers will be unfamiliar with.

Finally, EBM has spawned evidence-based decision making and evidence-based policy. In the UK, even government ministers make frequent reference to the 'evidence base' in their response to each public health emergency that confronts them. At first sight this seems wholly laudable. However, closer examination reveals a belief that one can go directly from 'evidence' to decision. Our view would be that the role of evidence is to inform a *model*. Even when the evidence appears to arise from studies that seem to replicate exactly the decision question at hand, there is still an implicit model that relates the historical evidence to what would be expected to happen following a decision to be made in the future. In fact the expression 'evidence-based modelling' would not be out of place in the title of this book.

1.4 Bayesian statistics

The first defining feature of the Bayesian approach is the focus on parameter uncertainty, or more generally how we can draw inferences about uncertain parameters given the data, rather than what we can say about the likelihood of the data at

specific values of the parameters. It is evident from the discussion of decision making under uncertainty (Section 1.2) that the exercise this book is embarked on is necessarily going to be a Bayesian exercise, and this has been widely recognised [12, 13].

However, the way the example in Section 1.2 was presented is somewhat ambiguous regarding the second major defining feature of Bayesian data analysis, the updating of prior distributions with data to form posterior distributions (see Chapter 2). The approach adopted in Section 1.2, which corresponds to what is done in the vast majority of probabilistic decision models, makes no explicit mention of prior distributions being updated to form posteriors. What has happened is this: the evidence on each parameter has been summarised and characterised as a distribution. This is followed by 'forward' Monte Carlo simulation from what could be considered 'informative priors' (see Sections 3.5 and 8.1).

There is nothing inherently wrong with this, except that we are forced to treat each parameter as an independent item. This severely limits the way evidence can be used, and eventually would even limit *which* evidence can be used. A full use of Bayesian methodology, providing us with posterior distributions, will allow us to be much more flexible in our use of data. In particular, we will be able to combine evidence on parameters, such as in Table 1.1, with indirect evidence, and with evidence on complex functions of the parameters (Chapter 8–10). The decision context actually *requires* a Bayesian posterior analysis just because this is the only way to make sure that all available evidence will be included.

Until recently, Bayesian analysis, except in particularly simple circumstances, was accessible only to those with considerable mathematical and computing skills. The situation has dramatically changed with the recent the arrival of flexible Markov Chain Monte Carlo (MCMC) software in the form of WinBUGS [14]. Models that previously would have presented intractable difficulties can be estimated conveniently by users with relatively little mathematical knowledge. Even so, we must warn readers that Bayesian estimation with WinBUGS requires a degree of care and thought that goes beyond the use of standard frequentist software. In the decision-making context of this book, however, MCMC methods have the very obvious attraction of being simulation based. MCMC therefore allows us to replace simulation from unrelated prior distributions by simulation from a joint posterior parameter distribution. Immediately, we can now integrate the statistical estimation step with the calculation of net benefit in a single, one-step process (Chapter 7). Furthermore rather than propagate evidence uncertainty 'forward' from parameter distributions, we propagate it 'back' from data, onto parameters, then forward through the decision model.

1.5 NICE

In England and Wales the National Institute for Health and Clinical Excellence (NICE), has the responsibility to decide whether new interventions will be reimbursed by the National Health Service. NICE is one of many similar bodies

worldwide though it was regarded by the World Health Organisation as pre-eminent in 2003. The role of NICE is to ensure that new products and procedures are properly appraised before being introduced. The purpose is both to limit expenditure, but, equally importantly to stop what had come to be called the 'postcode lottery', by which the treatments that patients received simply depended on which hospital they attended.

NICE's procedures emphasise transparency. Because its decisions are scrutinised by manufacturers, clinicians' professional organisations, and the Department of Health – as well as being open to public inspection – transparency of method is essential to maintain credibility and to create a process in which stakeholders who may on occasions have fundamentally different objectives can nevertheless participate and reach a degree of consensus. Once the specific decision question is set, specifying the patient group, the new technology and the comparator technologies, manufacturers are invited to make submissions which include analyses of the efficacy and cost-effectiveness of the products in question. Or, if multiple new technologies are to be compared, a Technical Assessment Report is produced by one of the academic units contracted to undertake this kind of work. This includes an independent assessment of efficacy and cost-effectiveness, as well as a review of previous cost-effectiveness analyses and of the manufacturers' submission.

NICE methods guidelines [15] bring together all the key elements described above: economic analyses are generally based on net benefit analysis; probabilistic decision analysis is the expected primary analysis, and clear reasons must be advanced if submissions depart from this; submissions adhere to the general principles of 'hierarchy of evidence', with explicit protocols for the literature search and study inclusion or exclusion. Many submissions include Bayesian evidence synthesis with simulation from a posterior distribution in WinBUGS.

1.6 Structure of the book

Each chapter includes running worked examples, and ends with suggestions for further reading, and exercises. The further reading suggestions are carefully selected to lead readers into slightly more complex syntheses, as well as to provide a broader exposure to the important issues which must be addressed in cost-effectiveness modelling, but which we cannot cover in this book. The full WinBUGS code for all the examples and exercises, as well as solutions, are available from the publisher's website www.wiley.com/go/decision_making_healthcare. Sections with more advanced material, and more difficult exercises, are asterisked.

Chapter 2 introduces Bayesian reasoning and explains why it is suited to decision modelling. It also provides an introduction into the practical use of the WinBUGS software. Chapter 3 introduces deterministic and stochastic decision models and how cost-effectiveness acceptability curves can be constructed. Chapter 4 describes a Bayesian approach to fixed and then random effects meta-analysis for pairs of treatments, which is then extended to meta-regression and the use of baseline as a covariate in Chapter 5. Chapter 6 asks how users can evaluate

meta-analysis models, and what issues to consider when choosing between Fixed and Random Effects models. Chapter 7 shows how the synthesis methods developed this far can be embedded within cost-effectiveness analyses.

Chapters 8 and 9 take two different evidence structures in turn, epidemiological models, then mixed and indirect treatment comparisons, and introduce multi-parameter evidence synthesis techniques. These all involve simultaneous estimation using 'direct' evidence on model parameters *and* 'indirect' evidence on functions of model parameters. Chapter 10 introduces WinBUGS programming for Markov models of disease progression. Chapter 11 takes on the challenge represented by data that is relevant, but potentially biased – such as observational studies of treatment efficacy. Chapter 12 briefly describes Expected Value of Information analysis, and gives some guidance on how this can be computed in the presence of the complex correlations between parameters that are induced by multi-parameter evidence synthesis.

The book includes a listing of all abbreviations used (Appendix 1), and a guide to common distributions (Appendix 2).

1.7 Summary key points

Developments in several distinct fields have coalesced to create an environment for the development of the methods described in this book.

- In health economics Net Benefit analysis has put health gains and costs on the same basis, and creating an environment where Bayesian Expected Value decision making can be carried out.
- Probabilistic decision modelling became the preferred way to evaluate cost-effectiveness analyses, as it faithfully maps parameter uncertainty into decision uncertainty, and correctly delivers Expected Net Benefit.
- EBM insists on a protocol for inclusion and exclusion of evidence in reviews and summary measures, and has acted to generally raise awareness of issues around the quality of evidence, and thus the uncertainty in estimates derived from it.
- Simultaneously, the development of Bayesian MCMC methods has made Bayesian data analysis routinely accessible, and dovetails exactly with the growing popularity of probabilistic cost-effectiveness models.

1.8 Further reading

Statisticians seeking a more rounded idea of health economics could consult Hunink *et al*. [16] or Briggs *et al*. [17] could be used as a good introduction to medical decision modelling.

Health economists and decision modellers interested in Bayesian statistics and its implications are recommended Spiegelhalter *et al*. [18]. Standard textbooks on Bayesian methods are: Berry [19], O'Hagan and Luce [20] and Lee [21].

The philosophy behind the Evidence-Based Movement is well captured in a series of papers 'Users Guide to the Medical Literature' [10, 22–27]. The dominant force in EBM methodology is the Cochrane Collaboration handbook [28]. For some recent literature that raises more open questions about evidence inclusion, Cooper *et al*. [29] can be consulted. A completely different approach to grading studies, which is fully compatible with the approach taken in this book, can be found in Turner *et al*. [11].

In line with its policy of complete transparency, all NICE guidelines and guidance, as well as the vast majority of the submitted evidence, can be found on the NICE website. Readers may find the methods guidelines of particular interest [15].

References

1. Claxton K., Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Economics* 1996;**5**:513–524.
2. Stinnett A., Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analyses. *Medical Decision Making* 1998;**18**: S68–S80.
3. Laupacis A., Feeny D., Detsky A.S., Tugwell P.X. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaulations. *Canadian Medical Association Journal* 1992;**146**: 473–481.
4. Coast J. Is economic evaluation in touch with society's health values ? *British Medical Journal* 2004;**329**: 1236.
5. Dolan P., Cookson R. A qualitative study of the extent to which health gain matters when choosing between groups of patients. *Health Policy* 2000;**51**:19–30.
6. Ades A.E., Claxton K., Sculpher M. Evidence synthesis, parameter correlation and probabilistic sensitivity analysis. *Health Economics* 2005;**14**:373–381.
7. Doubilet P., Begg C.B., Weinstein M.C., Braun P., McNeill B.J. Probabilistic sensitivity analysis using Monte Carlo simulation: a practical approach. *Medical Decision Making* 1985;**5**:157–177.
8. Critchfield G.C., Willard K.E. Probabilistic analysis of decision trees using Monte Carlo simulation. *Medical Decision Making* 1986;**6**:85–92.
9. Mulrow C.D. Rationale for systematic reviews. In: Chalmer I., Altman D.G., eds. *Systematic Reviews*. London: BMJ Publishing Group, 1995;1–8.
10. McAlister F., Laupacis A., Wells G.A., Sackett D.L. Users' guides to the medical literature. XIX Applying clinical trial results. B. Guidelines for determining whether a drug is exerting (more than) a class effect. *Journal of the American Medical Association* 1999;**282**:1371–1377.

11. Turner R.M., Spiegelhalter D.J., Smith G.C.S., Thompson S.G. Bias modelling in evidence synthesis. *Journal of the Royal Statistical Society, Series A* 2009;**172**:21–47.

12. Felli J.C., Hazen G. A Bayesian approach to sensitivity analysis. *Health Economics* 1999;**8**:263–268.

13. Briggs A.H. A Bayesian approach to stochastic cost-effectiveness analysis. *Health Economics* 1999;**8**:257–261.

14. Lunn D.J., Thomas A., Best N., Spiegelhalter D. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 2000;**10**: 325–337.

15. National Institute for health and Clinical Excellence. *Guide to the Methods of Technology Appraisal*. London: NICE, 2008.

16. Hunink M., Glasziou P., Siegel J., *et al*. *Decision Making in Health and Medicine: Integrating Evidence and Values*. Cambridge: Cambridge University Press, 2001.

17. Briggs A., Sculpher M., Claxton K. *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press, 2006.

18. Spiegelhalter D.J., Abrams K.R., Myles J. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. New York: John Wiley & Sons, Ltd, 2004.

19. Berry D.A. *Statistics: A Bayesian Perspective*. London: Duxbury, 1996.

20. O'Hagan A., Luce B. *A Primer on Bayesian Statistics in Health Economics and Outcome Research*. www.bayesian-initiative.com: Bayesian Initiative in Health Economics and Outcomes Research, MedTap Intl, 2003.

21. Lee P.M. *Bayesian Statistics: An Introduction*. London: Arnold, 1997.

22. Drummond M.F., Richardson W.S., O'Brien B.J., Levine M., Heyland D. Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *Journal of the American Medical Association* 1997;**277**:1552–1557.

23. Guyatt G.H., Sinclair J., Cook D.J., Glasziou P. Users' guides to the medical literature. XVI. How to use a treatment recommendation. *Journal of the American Medical Association* 1999;**281**:1836–1843.

24. Guyatt G.H., Sackett D.L., Cook D.J. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *Journal of the American Medical Association* 1993;**270**:2598–2601.

25. Guyatt G.H., Haynes R.B., Jaeschke R.Z., *et al*. Users' guides to the medical literature: XXV. Evidence-based medicine: principles for applying the users' guides to patient care. *Journal of the American Medical Association* 2000;**284**:1290–1296.

26. Oxman A.D., Cook D.J., Guyatt G.H. Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. *Journal of the American Medical Association* 1994;**272**:1367–1371.

27. Oxman A.D., SACKETT D.L., Guyatt G.H. Users' guides to the medical literature. I. How to get started. The Evidence-Based Medicine Working Group. *Journal of the American Medical Association* 1993;**270**:2093–2095.

28. Higgins J.P.T., Green S. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.0 [updated February 2008]*. Chichester: The Cochrane Collaboration, John Wiley & Sons, Ltd, 2008.

29. Cooper N.J., Sutton A.J., Abrams K.R., Turner D., Wailoo A. Comprehensive decision analytical modelling in economic evaluation: a Bayesian approach. *Health Economics* 2003;**13**:203–226.