1

Concepts

1.1 INTRODUCTION

There is a growing expectation from health care policymakers that evidence supporting the cost-effectiveness of new health care interventions, particularly pharmaceuticals, be provided along with the customary data on efficacy and safety. In Australia (Commonwealth of Australia, 1990) and Canada (Detsky, 1993) there are formal requirements that pharmaceutical companies present evidence of cost-effectiveness before a drug is granted reimbursement status on a formulary. In the United States there is demand for such economic data from third-party insurers, see Leaf (1989).

There are two general approaches to performing an economic evaluation of a health care intervention, see O'Brien (1996). One approach combines the efficacy and safety data from randomized clinical trials (RCTs) with cost data from secondary, non-trial sources in a decision analysis model. In such models the problem of inferential uncertainty is addressed using sensitivity analyses to determine what effect varying the model assumptions has on the results, see Briggs *et al.* (1994). The other approach uses health care utilization data collected on individual patients prospectively as part of an RCT. The health care utilization data combined with the appropriate price weights yield a measure of cost for each patient. Measuring effectiveness and cost at the patient level permits the use of more conventional methods of statistical inference to quantify the uncertainty due to sampling and measurement error. Since the early 1990s, when such data became

Statistical Analysis of Cost-effectiveness Data. A. Willan and A. Briggs

more common, numerous articles have been published in the area of the statistical analysis of cost-effectiveness data. Initially, efforts were concentrated on providing confidence intervals for incremental cost-effectiveness ratios, but more recently, due to concerns regarding ratio statistics, the concept of incremental net benefit has been proposed as an alternative.

The purpose of this book is to provide an illustrated summary of some of the key developments published in the last 10 years that deal with statistical issues related to the cost-effectiveness comparison of two groups when measures of effectiveness and cost are observed at the subject level. The context used throughout the book is that of patients in a two-arm RCT where patients are randomized to Treatment (T) or Standard (S), but the methods apply to the comparison of any two groups, subject to the concerns one might have regarding bias due to the lack of random group allocation.

1.2 COST-EFFECTIVENESS DATA AND THE PARAMETERS OF INTEREST

In a cost-effectiveness analysis (CEA), whether an incremental cost-effectiveness ratio (ICER) or an incremental net benefit approach is taken, five parameters need to be estimated. Two of the parameters are the differences between treatment arms of mean effectiveness and costs, denoted by Δ_e and Δ_c, respectively. The other three parameters are the variances and covariance of those estimators. With the estimators of these five parameters, a CEA, based on either the incremental cost-effectiveness ratio or incremental net benefit, can be performed. For non-censored data the estimators are simple functions of the sample means, variances and covariance. For censored data estimation procedures are decidedly more complex.

Typically, the measure of effectiveness in a CEA is associated with a clinical event experienced by the patient. Quite commonly the event is death, but it could be relapse or reaching a pre-specified level of symptom relief. For simplicity, unless otherwise noted, we assume that the event is death. The simplest measure of effectiveness based on event data is the probability of the event not occurring within a

specified period of time from randomization. The specified period of time is often referred to as the duration of interest and denoted by τ. Let the random variable D_{Ti} be the time from randomization to the event for the ith patient on arm T and let $S_T(t) = \Pr(D_{Ti} \geq t)$, then the measure of effectiveness is given by $S_T(\tau)$ and denoted as π_T. $S_T(t)$ is the survival function for patients on arm T. Defining D_{Si}, $S_S(t)$ and π_S similarly for patients on arm S, the parameter of interest for effectiveness, denoted Δ_e, is given by

$$\Delta_e = S_T(\tau) - S_S(\tau) = \pi_T - \pi_S \tag{1.1}$$

The quantity Δ_e is the absolute risk reduction, and $1/\Delta_e$ is the mean number of patients that need to be treated with T rather than S to prevent a death. This quantity is usually referred to informally as the 'number-needed-to-treat' or more simply as the NNT. If the probability of surviving 5 years is 0.6 for patients on arm T and only 0.5 for patients on arm S, then we say that 10 (i.e. 1/0.1) patients need to be treated with T rather than S to prevent one death, or more simply that the NNT is 10.

Another measure of effectiveness based on event data is the mean survival time over the duration of interest, otherwise referred to as the restricted mean survival time. The restricted mean survival time is sensitive to the entire survival curve from 0 to τ, and not just its value at τ. The restricted mean survival time for a particular arm, denoted as μ_j, $j = T, S$, is the area under the respective survival curve from 0 to τ, i.e.

$$\mu_j = \int_0^\tau S_j(t)\mathrm{d}t$$

and the parameter of interest to be estimated for effectiveness is given by

$$\Delta_e = \int_0^\tau S_T(t)\mathrm{d}t - \int_0^\tau S_S(t)\mathrm{d}t = \mu_T - \mu_S \tag{1.2}$$

For mean survival time quantity $1/\Delta_e$ is the NNT to gain one year of life over the duration of interest. If the restricted mean survival time over 5 years for a patient on arm T is 4 and only 3.75 for a patient on

arm S, then the NNT to gain one year of life is 4 (i.e. $1/0.25$). The mean total restricted survival time of 4 patients on arm T is $4 \times 4 = 16$, while on arm S the mean total restricted survival time of 4 patients is $4 \times 3.75 = 15$.

The third measure of effectiveness based on survival data is the mean quality-adjusted survival time over the duration of interest, otherwise referred to as the restricted mean quality-adjusted survival time. Quality-adjusted survival time is based on the concept that patients experience, at any given time, a certain quality of life based on a utility scale for which 1 is perfect health and 0 is death, see Torrance (1986). Negative values are used to allow for states of health considered worse than death. If the quality of life at time t for a patient on a particular treatment arm is given by $Q_j(t)$, $j = T, S$, then the restricted mean quality-adjusted survival time is given by

$$\varphi_j \equiv \int_0^\tau Q_j(t)\mathrm{d}t$$

and the parameter of interest to be estimated for effectiveness is given by

$$\Delta_e = \int_0^\tau Q_T(t)\mathrm{d}t - \int_0^\tau Q_S(t)\mathrm{d}t = \varphi_T - \varphi_S \tag{1.3}$$

For quality-adjusted survival time the quantity $1/\Delta_e$ is the NNT to gain one quality-adjusted life-year over the duration of interest.

The quantity Δ_e is the difference between treatment arms with respect to effectiveness, and is a different function of the survival curves, depending on which measure of effectiveness is of interest.

If we let ν_j be the mean cost over the duration of interest for a patient in arm j, $j = T, S$, then the parameter of interest to be estimated for cost is given by

$$\Delta_c = \nu_T - \nu_S \tag{1.4}$$

The observed cost for a given patient is simply the sum of the amounts of each resource consumed by the patient multiplied by the respective price weight. Which resources are included depends

on the perspective taken by the analysis. If the analysis takes the perspective of the health care system, only resources covered under the system would be included. However, if a broader societal perspective were taken, then costs not covered under the system and items such as time lost from work and care by a family member could also be included; for a fuller discussion the reader is referred to Drummond *et al*. (1997). Estimation methods for Δ_e, Δ_c and the corresponding variances and covariance are given in Chapter 2 for non-censored data and in Chapter 3 for censored data.

1.3 THE COST-EFFECTIVENESS PLANE, THE ICER AND INB

Researchers have long used the cost-effectiveness plane to explore the policy interpretation of cost-effectiveness analyses. The cost-effectiveness (CE) plane is a graph with Δ_c and Δ_e plotted on the vertical axis and horizontal axis, respectively, as illustrated in Figure 1.1. For more discussion on the cost-effectiveness plane the reader is referred to Black (1990). Let $\Delta = (\Delta_e, \Delta_c)^T$. If, for a particular Treatment/Standard comparison, the point Δ is located in the Southeast (SE) quadrant (i.e. $\Delta_e > 0$, $\Delta_c < 0$), Treatment is said to dominate Standard because it is more effective and less costly, and the argument to adopt it to replace Standard is self-evident. By contrast, if Δ lies in the Northwest (NW) quadrant (i.e. $\Delta_e < 0$, $\Delta_c > 0$) Treatment is dominated by Standard, and its rejection as a replacement for Standard is the rational policy choice. It is in the Northeast (NE) and Southwest (SW) quadrants, referred to as the trade-off quadrants, that the magnitudes of Δ_e and Δ_c need to be considered to determine if Treatment is cost-effective.

To assist in this determination researchers have traditionally used the incremental cost-effectiveness ratio. The ICER is defined as $R \equiv \Delta_c/\Delta_e$, but can be written as

$$\frac{1}{\Delta_e}\Delta_c = \text{NNT} \times \Delta_c$$

It is easy to see then that the ICER is the product of the number of patients that need to be given Treatment to achieve an extra unit

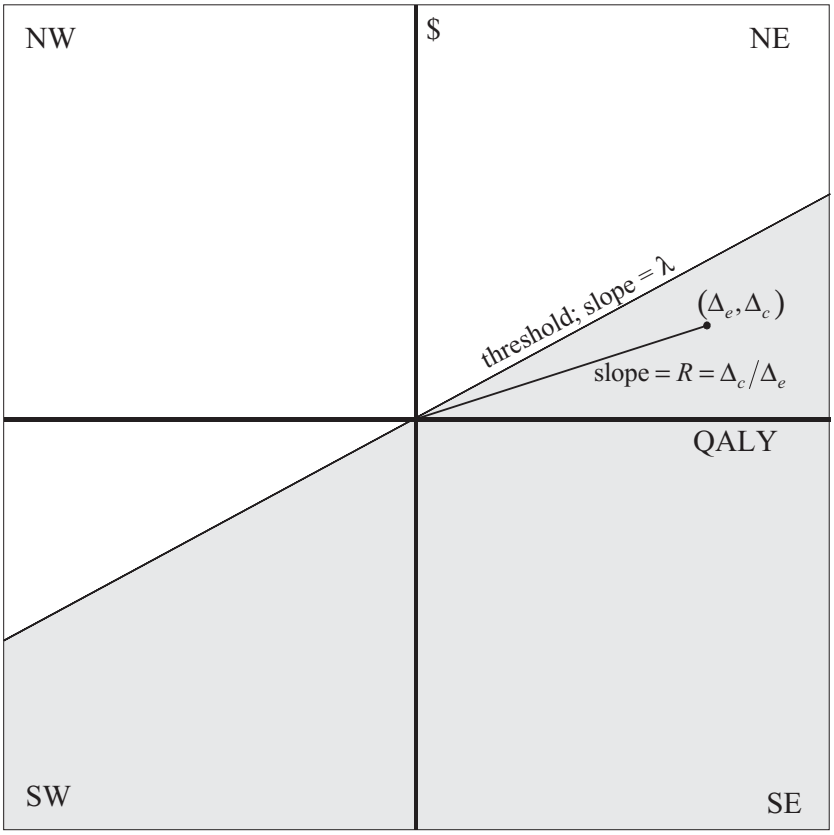

Figure 1.1 The cost-effectiveness plane

of effectiveness and the incremental cost of treating each of those patients, and is therefore the incremental cost of achieving a unit of effectiveness from using Treatment rather that Standard. On the CE plane the ICER is the slope of the line between the origin and the point Δ, see Figure 1.1. If the measure of effectiveness is the probability of surviving then the ICER is cost of saving a life (or preventing a death). If the measure of effectiveness is mean survival or mean quality-adjusted survival, then the ICER is the cost of achieving an extra year or quality-adjusted year of life (QALY), respectively. Essentially the ICER is the cost of an additional unit of effectiveness if Treatment is adopted over

Standard. This, as in any transaction, needs to be compared with what a policymaker is willing to pay.

The amount a policymaker is willing to pay is referred to as the willingness-to-pay (WTP), and is denoted by λ. The concept of WTP is discussed by Pauly (1995), and methods for quantifying it can be found in O'Brien and Gafni (1996), Johnson *et al.* (1998) and Hanley *et al.* (2003). By drawing a line through the origin with slope λ, the CE plane can be divided into two regions. For convenience this line will be referred to as the threshold. For points on the plane below and to the right of the threshold (the shaded area in Figure 1.1), Treatment is considered cost-effective, but for those above and to the left it is not. Since λ is positive, points in the SE quadrant are always below the threshold and therefore correspond to comparisons for which Treatment is cost-effective. On the other hand, points in the NW are always above the threshold and correspond to comparisons for which Treatment is not cost-effective. It is in the NE and SW quadrants that the concept of WTP allows for trade-off between effectiveness and costs. In the NE quadrant the slope of any point below the line is less than λ, i.e. $\Delta_c/\Delta_e < \lambda$ which implies that $\Delta_c < \Delta_e\lambda$. Therefore, the increase in value ($\Delta_e\lambda$) is greater than the increase in cost, making Treatment cost-effective. In the SW quadrant the slope of any point below the line is greater than λ, and since Δ_e and Δ_c are both negative (i.e. treatment is less effective and less costly), we have $\Delta_c/\Delta_e = |\Delta_c| \,/\, |\Delta_e| > \lambda$ which implies that $|\Delta_c| > |\Delta_e\lambda|$. Therefore, the value lost ($|\Delta_e\lambda|$) is less than the amount saved ($|\Delta_c|$), making Treatment cost-effective. In summary, Treatment is cost-effective if

$$\text{A}: \quad \frac{\Delta_c}{\Delta_e} < \lambda \qquad \text{if } \Delta_e > 0; \qquad \text{or} \quad \frac{\Delta_c}{\Delta_e} > \lambda \qquad \text{if } \Delta_e < 0 \quad (1.5)$$

Expression (1.5) (Hypothesis A) defines the region below the threshold and can be thought of as the alternative hypothesis for the null Hypothesis H, given by:

$$\text{H}: \quad \frac{\Delta_c}{\Delta_e} \geq \lambda \qquad \text{if } \Delta_e > 0; \qquad \text{or} \quad \frac{\Delta_c}{\Delta_e} \leq \lambda \qquad \text{if } \Delta_e < 0 \quad (1.6)$$

Rejecting H in favour of A would provide evidence to adopt Treatment. These expressions are somewhat awkward and can be simplified considerably by the introduction of incremental net benefit.

The incremental net benefit (INB) is a function of λ, and is defined as

$$b_\lambda \equiv \Delta_e \lambda - \Delta_c \tag{1.7}$$

b_λ is the incremental net benefit because it is the difference between incremental value ($\Delta_e \lambda$) and incremental cost (Δ_c). Treatment is cost-effective if, and only if, $b_\lambda > 0$, regardless of the sign of Δ_e. To see this, both inequalities involving the ICER in Expression (1.5) can be rearranged to the inequality $\Delta_e \lambda - \Delta_c > 0$. Similarly, both inequalities involving the ICER in Expression 1.6 can be rearranged to the inequality $\Delta_e \lambda - \Delta_c \leq 0$. Therefore, in terms of INB the null and alternative hypotheses become

$$\mathrm{H} : \Delta_e \lambda - \Delta_c \leq 0 \qquad \text{versus} \qquad \mathrm{A} : \Delta_e \lambda - \Delta_c > 0 \tag{1.8}$$

On the CE plane b_λ is the vertical distance from the point Δ to the threshold, being positive if it is below the line and negative otherwise. Because it has slope λ, the point on the threshold with abscissa equal to Δ_e is $(\Delta_e, \Delta_e \lambda)$ and so the vertical distance between it and Δ is $\Delta_e \lambda - \Delta_c$, see Figure 1.2.

The incremental net health benefit (INHB) is defined as $\Delta_e - \Delta_c / \lambda = b_\lambda / \lambda$ and measures net benefit in units of effectiveness. Since INHB is simply a positive constant times INB, statistical inference made on one will be identical to statistical inference made on the other. INB has the advantage of being linear in λ. Therefore in a sensitivity analysis of WTP, the plots of INB by λ, are straight lines. Another advantage of INB is that it generalizes to more than one outcome. In a trial of patients at risk of thrombosis, if Δ_{e1}, Δ_{e2} and Δ_{e3} are the differences of the probability of avoiding death, thrombosis and stroke, respectively, and if λ_1, λ_2 and λ_3 are the corresponding WTP values, then INB is defined as $\Delta_{e1}\lambda_1 + \Delta_{e2}\lambda_2 + \Delta_{e3}\lambda_3 - \Delta_c$. A corresponding formulation in INHB is not possible.

1.4 OUTLINE

The remainder of the book is organized as follows. Methods for estimating Δ_e and Δ_c and their variances and covariances for non-censored data are given in Chapter 2. The methods make use of simple statistics,

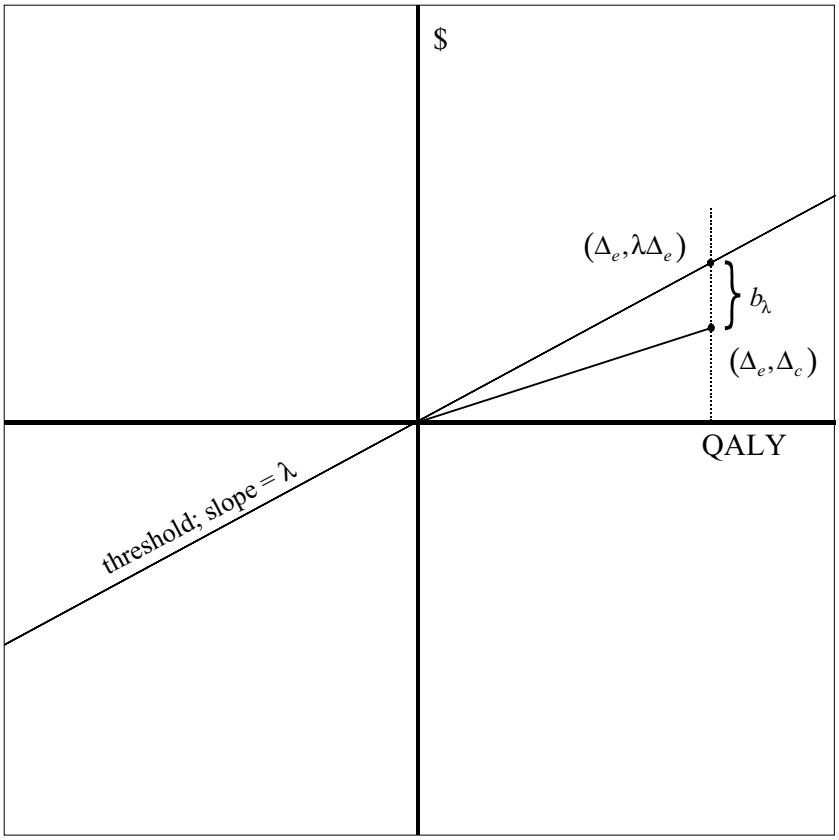

Figure 1.2 INB on the cost-effectiveness plane

such as proportions and sample means and variances. Estimation methods for censored data are given in Chapter 3. The methods include life-table procedures, the direct method of Lin *et al.* (1997) and inverse probability weighting. How the parameters are used in a cost-effectiveness analysis is described in Chapter 4. Emphasis is placed on estimating the ICER and INB, along with their confidence limits, and constructing cost-effectiveness acceptability curves. In Chapter 5 the methods of Chapters 2, 3 and 4 are illustrated with examples. Methods for determining sample sizes, using both classical and Bayesian

approaches, are given in Chapter 6. In Chapter 7 regression methods for covariate adjustment and testing for treatment by prognostic factor interactions are described, along with several examples. The issues regarding multicenter and multinational trials are the subject of Chapter 8. In Chapter 9 a more general framework of statistical modeling is proposed, which is based on modeling the separate components cost-effectiveness to build indirect estimates of incremental cost and effectiveness.