University of Oxford

A Bayesian Cost-Benefit Approach to Sample Size Determination and Evaluation in Clinical Trials

by

Takashi Kikuchi
BSci, MMedSci, MMedStat, DocMed & D.Phil.

St. Hugh's College

Thesis submitted to the Mathematical, Physical and Life Sciences Division for the degree of Doctor of Philosophy of the University of Oxford.

Department of Statistics, 1 South Parks Road,
Oxford OX1 3TG, United Kingdom

Trinity Term 2010

This work is dedicated to my parents and the progress of health service research in the world.

My viva was successfully completed on 11 November 2010 and the Division granted me leave to supplicate for the degree of Doctor of Philosophy approved on 3 December 2010. In the history of the Department of Statistics more than 100 years, I am the first Japanese student whom D.Phil in Statistics was awarded.

Takashi Kikuchi, BSci, MMedSci, MMedStat, DocMed & D.Phil.

Signed:.....

Date: 24 December 2010

Version: 19.07.2010 (approved).

Abstract

Current practice for sample size computations in clinical trials is largely based on frequentist or classical methods. These methods have the drawback of requiring a point estimate of the variance of treatment effect and are based on arbitrary settings of type I and II errors. They also do not directly address the question of achieving the best balance between the costs of the trial and the possible benefits by using a new medical treatment, and fail to consider the important fact that the number of users depends on evidence for improvement compared with the current treatment.

A novel Bayesian approach, Behavioral Bayes (or BeBay for short) (Gittins and Pezeshk, 2000a,b, 2002a,b; Pezeshk, 2003), assumes that the number of patients switching to the new treatment depends on the strength of the evidence which is provided by clinical trials, and takes a value between zero and the number of potential patients in the country. The better a new treatment, the more patients switch to it and the more the resulting benefit. The model defines the optimal sample size to be the sample size that maximises the expected net benefit resulting from a clinical trial. Gittins and Pezeshk use a simple form of benefit function for paired comparisons between two medical treatments and assume that the variance of the efficacy is known.

The research in this thesis generalises these original conditions by introducing a logistic benefit function to take account of differences in efficacy and safety between two drugs. The model is also extended to the more general cases of unpaired comparisons and unknown variance. The expected net benefit defined by Gittins and Pezeshk is based on the efficacy of the new drug only. It does not consider the incidence of adverse reactions and their effect on patients' preferences. Here we include the costs of treating adverse reactions and calculate the total benefit in terms of how much the new drug can reduce societal expenditure.

We describe how our model may be used for the design of phase III clinical trials, cluster randomised clinical trials and bridging studies. This is done in some detail and using illustrative examples based on published studies. For phase III trials we allow the possibility of unequal treatment group sizes, which often occur in practice. Bridging studies are those carried out to extend the range of applicability of an established drug, for example to new ethnic groups. Throughout the objective of our procedures is to optimise the cost-benefit in terms of national health-care. BeBay is the leading methodology for determining sample sizes on this basis. It explicitly takes account of the

roles of three decision makers, namely patients and doctors, pharmaceutical companies and the health authority.

Acknowledgements

I feel as if several decades have already passed since I started my research in the University of Oxford. I began my research from almost nothing and my academic battle over 9 years was too hard to describe in a few sentences. I was often so depressed with my slow progress and desperate efforts. However, I have never given it up because I strongly believe that my research can help practical problems in current health service research and affect the current practices of clinical trials.

This thesis looks five or ten years ahead in health service research. I expect that cost-benefit evaluation will be a common practice and decision making by using null hypothesis testing will lose popularity within a decade. This belief was fuel to drive my research motivation. I hereby acknowledge that this research could not have been completed without assistance from many people and would like to sincerely express thanks.

First of all, I sincerely would like to thank my supervisor, Professor John Gittins, for the uncountable number of suggestions about my research, and for his time, patience and encouragement to me. He gave valuable advice and accomplished professional supervision throughout the course of this work, which was essential to the success of my project. Dr Mary Lunn also supervised my study during the first year and helped me in the later years as an academic advisor.

I thank supporting members of the Department of Statistics, Oxford, for their willingness to help. I would like to give special thanks to IT managers, Ms Susan Hutchinson, Mr David del Campo Hill and Ms Saffron Greenwood, for their professional maintenance of our computer system to an excellent standard and their quick responses to my enquiries. I am grateful for the help of Mrs Vivien Stchedroff, Mrs Jane Hornsby and Mrs Sue Wood in providing efficient and friendly working environments. The sudden death of Vivien shocked us all very much, and we will never forget her kindness to us.

I also acknowledge that I have received financial support: a hardship fund award and a departmental bursary from St Hugh's College and The Department of Statistics, respectively. I made presentations at two international academic conferences: 22nd annual conference of ISCB in Stockholm (2001) and the international conference of risk management for preventive medicine in Tokyo (2003), at the expense of St Hugh's College and the Department of Statistics.

I would also like to thank the two examiners who pointed out shortcomings in the previous two versions of the thesis, and the editors and anonymous

reviewers for the statistical journals in which the work is being published (Kikuchi et al., 2008; Kikuchi and Gittins, 2009, 2010a,b, 2011). These comments have led to substantial improvements.

The thesis and the five papers on which it is based were written under the supervision of Professor Gittins. His contribution to these joint papers and to the thesis itself has been the normal one of a research supervisor in providing comments and criticism. I included the name of Dr. Pezeshk as an author in my first paper (Kikuchi et al., 2008) in recognition of his earlier work on Behavioural Bayes theory at the department of statistics of the University of Oxford. He was not involved in writing any of these joint papers, nor in my D.Phil research. The series of extensions of Behavioural Bayes set out in this thesis are based on my ideas and research plan.

This research is not funded and there are no conflicts of interest. For the calculations needed for the example we used S-Plus 6.1. This was provided by Mathematical Systems, Inc, Tokyo (http://www.msi.co.jp/splus/).

I am very glad to have finally completed all the research that I intended to carry out in Oxford. It has turned out exactly as I had pictured in my mind. I finally would like to express my great thanks to my parents. Needless to say, my parents are the best contributors to my research. Without their eternal love and support to me, this research would never exist on earth.

Takashi Kikuchi, Spring 2010

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 THE TESTING AND EVALUATION OF NEW DRUGS

1.1.1 Evaluation

The classical approach to statistical hypothesis testing is not well designed for decision making. This is because any null hypothesis will be rejected if the sample size is large enough. On the other hand, if the hypothesis is not rejected all we know is that the sample size is too small to reject it. If the sample size is large, even a small difference causes the hypothesis to be rejected. Furthermore, the rejection depends on arbitrary settings of type I and type II errors. This framework is not what is needed for decision making in health policy but even so currently health service research is largely based on this approach.

In practice budgetary restrictions limit the size of the sample. We define the optimal sample size to be one which attains the right balance between the cost and benefit of a clinical trial. This chapter briefly reviews the current practices for clinical trials and points out several limitations. We shall present a novel Bayesian approach, Behavioral Bayes, which avoids these limitations. As we shall see our approach develops further the notion of the expected value of sample information, which is playing an increasing role in the evaluation of clinical trials.

The following chapters continue in this vein and are very close in content to a series of published papers: Kikuchi et al. (2008); Kikuchi and Gittins

(2009, 2010a,b, 2011). These include some further review material.

Statistical methods for the analysis of data on cost-effectiveness have experienced rapid development over the past two decades. We shall review the current standard practice of health service research for decision making on the approval of new drugs. The main approaches are the incremental cost-effectiveness ratio (ICER) (Black, 1990), the cost-effectiveness acceptability curve (CEAC) (Van Hout et al., 1994) and the incremental net benefit (INB). Early development focused on the incremental cost-effectiveness ratio, but more recently, due to the mathematical problems associated with ratio statistics, health service researchers' attention has shifted to INB.

Suppose we compare two treatments T1 and T2. If the mean economic costs m_{c1} and m_{c2} and the mean clinical effectiveness m_{e1} and m_{e2} are given, we define the incremental means of costs and effectivenesses by $\theta_c = m_{c2} - m_{c1}$ and $\theta_e = m_{e2} - m_{e1}$, respectively. The ICER is defined as

$$\text{ICER} = \theta_c/\theta_e.$$

O'Hagan et al. (2000) proposed a cost-effectiveness plane as shown in Figure 1.1. The plane is divided into four quadrants depending on four possible combinations of signs of θ_c and θ_e. The slope K of the dashed line $\theta_c = K\theta_e$ is the maximum acceptable cost per unit of effectiveness and represents a health-care agency's "willingness to pay" (WTP) for a quality-adjusted life year (QALY). The QALY value of a year in perfect health is assigned the value of 1.0, with lower values for different levels of imperfect health.

Whilst the ICER is intuitively understandable, some difficulties arise in the calculation and the interpretation of confidence intervals when the denominator is negative or zero. Furthermore, the interpretation of ICER itself can be changed according to the sign of θ_e; T2 is favored by small values of ICER if $\theta_e < 0$, and by large values of ICER if $\theta_e > 0$. Therefore, we cannot judge the cost-effectiveness only by the values of ICER and there is a risk of being misled unless the sign of θ_e is clearly specified.

The incremental net benefit (INB) was proposed as an alternative to the ICER. It avoids the mathematical difficulties associated with ICER as above, and is defined as

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

Thus INB = 0 when ICER= K, which in both cases means that the cost effectiveness is just acceptable. For given values of θ_e, θ_c, the function $Q(K) = \text{P}(\text{INB}(K) > 0)$ gives the cost-effectiveness acceptability curve (CEAC) that is depicted by plotting $(K, Q(K))$.

Figure 1.1: Interpretation of different segments of the ICER plane.

Let us consider a joint distribution (a bivariate normal distribution) for $\theta = (\theta_c, \theta_e)$ and denote $\mathrm{E}(\theta_e) = \mu_e, \mathrm{V}(\theta_e) = \tau_e^2, \mathrm{E}(\theta_c) = \mu_c, \mathrm{V}(\theta_c) = \tau_c^2$, and $\mathrm{Corr}(\theta_c, \theta_e) = \rho$. For $\mathrm{INB}(\theta) = K\theta_e - \theta_c$ if K is fixed we have

$$\mathrm{E}(\mathrm{INB}) = K\mu_e - \mu_c \text{ and} \tag{1.1}$$

$$\mathrm{V}(\mathrm{INB}) = K^2\tau_e^2 - 2K\rho\tau_e\tau_c + \tau_c^2. \tag{1.2}$$

Suppose $\theta_e \sim \mathrm{N}(\mu_e, \tau_e^2)$ and θ_c given θ_e follows a normal distribution with the mean $\mathrm{E}[\theta_c] = \mu_c + \rho\tau_c(\theta_e - \mu_e)/\tau_e$ and variance $\mathrm{V}[\theta_c] = \tau_c^2(1-\rho^2)$. If we carry out Monte Carlo simulation for θ_e followed by θ_c given θ_e, we can simulate a large number of points for $\theta = (\theta_e, \theta_c)$ and draw dots as shown in Figure 1.2. Note that each dot in the figure shows a sample for θ from the distributions. In the figure dots above the WTP line have INB> 0 and those below the line have INB< 0. Thus if we count the number of dots locating above and below the line, we can approximately estimate Q(K)=P(INB> 0)=Number of dots above the WTP line/Total number of dots.

Koerkamp et al. (2007) point out that the CEAC is unable to distinguish

dramatically different joint distributions of incremental cost and effect. This limitation is demonstrated graphically in Figure 1.2. The cause of the limitation is that the CEAC is not sensitive to any change of the incremental joint distribution in the upper left and lower right quadrants of the cost-effectiveness plane, nor is it sensitive to radial shift of the incremental joint distribution in the upper right and lower left quadrants.

Consider the cost-effectiveness plane (CE-plane) shown in Figure 1.2A. It reflects a dense distribution for small values of the expected incremental costs and effects and some uncertainty. The arrows represent (hypothetical) changes in the distribution that would not modify the cost-effectiveness acceptability curve (CEAC). Points in the upper left or lower right quadrant can move anywhere else within the same quadrant, without changing the CEAC. Points in the two other quadrants can move radially toward or away from the origin, again without changing the CEAC. Figure 1.2B shows a plot which might result from moves of this kind. The result is a sparse distribution with large expected incremental costs and effects and more uncertainty. However, the CEACs of the two plots are identical. As Koerkamp et al. (2007) say "It is not clear why CEACs are so popular in light of these limitations".

Figure 1.2: Limitations of CEACs.

The CEAC provides the posterior probability that incremental cost effectiveness is positive (see equations (1.1) and (1.2)). However, if there are several options for treatments we cannot make the best choice by considering this probability only. Instead, the expected net benefit should be used for evaluation and the uncertainties of parameters values need to be incorporated.

The value of information is emerging as the preferred alternative for CEACs. This is shown by, for example, Pratt et al. (1994); Claxton (1999a); Claxton

et al. (2002); Ades et al. (2004); Assessment Health Technology (2004). The value of information is the amount that a decision maker would be willing to pay for information prior to making a decision. Suppose that we are deciding between treatments T1 or T2 and the true incremental cost and effectiveness are denoted by $\theta = (\theta_c, \theta_e)$. Recall that $\theta_c = m_{c2} - m_{c1}$ and $\theta_e = m_{e2} - m_{e1}$. Based on perfect information, if $\text{INB}(\theta) = K\theta_e - \theta_c > 0$ for given K then T2 would be preferred, or if $\text{INB}(\theta) < 0$ then T1 would be preferred. If θ is unknown, but we know its probability distribution, then we can calculate $\text{E}[\text{INB}(\theta)]$. If $\text{E}[\text{INB}(\theta)] > 0$ then we choose T2, and if it is also true that $\text{INB}(\theta) > 0$, it means that we made the right decision and there is no gain in knowing the exact value of θ. However, if in fact $\text{INB}(\theta) < 0$ and we choose T2, it means that we made the wrong decision and lose $-\text{INB}(\theta)$. The value of perfect information $\text{VOI}(\theta)$ is the value that we would gain by knowing θ exactly and may be written as

$$\text{VOI}(\theta) = \max(-\text{INB}(\theta), 0).$$

The expected value of perfect information, EVPI if T2 is currently preferred may be written as

$$\text{EVPI}_2 = \text{E}[\max(-\text{INB}(\theta), 0)].$$

This expresses that $\text{EVPI}_2 = 0$ if $\text{INB}(\theta) > 0$ or $\text{EVPI}_2 = -\text{INB}(\theta)$ if $\text{INB}(\theta) < 0$. Similarly, if T1 is preferred, we have $\text{E}[\text{INB}(\theta)] < 0$ and the EVPI is written as

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

The EVPI is valuable to policy makers as a means of evaluating the need for more research. The same principle can be extended to define the expected net-benefit of sample information (EVSI), which may be compared with the expected cost of sampling. This is equivalent to maximising the expected net benefit, which is the approach developed in this thesis, both for evaluating treatments and for determining sample sizes.

1.1.2 Sample size determination, earlier approaches

Current practice for sample size computations in Phase III clinical trials is largely based on frequentist or classical methods to test the null hypothesis that a new treatment is no better than a standard treatment. Classical methods define the sample size of clinical trials by controlling both the size and the power of the test: typical requirements are, at most a 5% chance of incorrectly accepting that the new treatment is better, and at least an 80%

chance of detecting a clinically relevant difference when it is actually present Desu and Raghavarao (1990); Lemeshow et al. (1990).

The determination of sample size is an important aspect of the design of clinical trials. There is an extensive literature from both the frequentist and Bayesian perspectives and in this section we briefly review some representative examples.

Frequentist method

The classical frequentist method poses the question of what sample size is needed to detect a particular difference with an hypothesis test with specified probabilities of type I and type II errors. The approach is very straightforward. Let Δ, α and β be the difference of effect to be detected, and the required probabilities of types I and II errors, respectively. If the variance σ^2 is known, the sample size n is given by

$$n = \{\frac{\sigma(z_\alpha + z_{1-\beta})}{\Delta}\}^2,$$

where $\Phi(z) = \int_{-\infty}^{z} \exp(-w^2/2)/\sqrt{2\pi}\, dw$ and the values for Δ, α and β are arbitrary. Note that $\Phi(z_\alpha) = \alpha$ and $\Phi(z_{1-\beta}) = 1-\beta$ and we usually use 0.05 and 0.8 for α and $1-\beta$, respectively. This indicates that the smaller the difference Δ we want to detect, the larger the sample size we need, and that the size depends on these arbitrary values.

Bayesian methods

This subsection is based on Pezeshk (2003) and reviews the main Bayesian methods for sample size determination. Bayesian methods may be classified into two groups: inferential and decision theoretic rules (or fully Bayesian rules). Adcock (1987), Spiegelhalter et al. (1994), and Joseph et al. (1997) have all written papers on Bayesian inference. Adcock (1997) reviews both the inferential and the fully Bayesian approaches. The first paper on sample size determination by the fully Bayesian approach was written by Grundy et al. (1956). Raiffa and Schlaifer (1961) describe decision theoretic rules for sample size determination in detail.

Let $\theta \in \Theta$ denote the parameter of interest and $\pi(\theta)$ the prior density of θ. Suppose that the experimental data are $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$, where n is the sample size and the components of $\boldsymbol{x}$ are exchangeable and belong to the data

space χ. Adcock (1987) introduced the average coverage criterion (ACC) that allows the coverage probability $1-\alpha$ to vary with $\boldsymbol{x}$ whilst holding the length l of highest posterior density (HPD) region. If the parameter θ of interest is contained in the interval of length l with a probability $1-\alpha$, we write

$$\int_{a(\boldsymbol{x},n)}^{a(\boldsymbol{x},n)+l} \pi(\theta|\boldsymbol{x})d\theta = 1-\alpha, \tag{1.3}$$

where $a(\boldsymbol{x},n)$ is the lower limit of the HPD interval. The posterior density $\pi(\theta|\boldsymbol{x})$ depends on $\boldsymbol{x}$ and n and before sampling equation (1.3) is required to hold on average over all possible samples. Therefore, we need to choose n to be such that

$$\int_{\chi}\{\int_{a(\boldsymbol{x},n)}^{a(\boldsymbol{x},n)+l} \pi(\theta|\boldsymbol{x})d\theta\}f(\boldsymbol{x})d\boldsymbol{x} = 1-\alpha,$$

where $f(\boldsymbol{x}) = \int_{\Theta} l(\boldsymbol{x}|\theta)\pi(\theta)d\theta$ and $l(\boldsymbol{x}|\theta)$ is the likelihood of the data.

Joseph et al. (1995) introduced the average length criterion (ALC). For this the coverage probability is fixed and the HPD interval length varies depending on the data. For $\boldsymbol{x} \in \chi$ and a HPD length $l'(\boldsymbol{x},n)$, we write

$$\int_{a(\boldsymbol{x},n)}^{a(\boldsymbol{x},n)+l'(\boldsymbol{x},n)} \pi(\theta|\boldsymbol{x})d\theta = 1-\alpha.$$

The average length $\overline{L}$ of fixed coverage HPD intervals by the predictive density $f(\boldsymbol{x})$ is

$$\overline{L} = \int_{\chi} l'(\boldsymbol{x},n)f(\boldsymbol{x})\,d\boldsymbol{x}.$$

The sample size is the minimum integer n that satisfies $\overline{L} \leq l$, where l is a prespecified length.

ACC and ALC are based on averages over all possible samples. However, inferences are conditional on the observed sample and the ACC and ALC can lead to lengths that are undesirably small or large, respectively. We need to ensure a maximum length l and a minimum coverage probability $1-\alpha$, regardless of the data $\boldsymbol{x}$. Thus the sample size is the minimum value of n to satisfy

$$\inf_{\boldsymbol{x}\in\chi} \int_{a(\boldsymbol{x},n)}^{a(\boldsymbol{x},n)+l} \pi(\theta|\boldsymbol{x})d\theta \geq 1-\alpha.$$

This is the worst outcome criterion that was described by Pham-Gia and Turkkan (1992) and developed by Joseph et al. (1995).

If we have a utility function for the cost of a clinical trial and the potential benefit of the treatment, the optimal sample size is the size which maximises the expected utility. This standpoint is fully Bayesian and is strongly advocated by Lindley (1997). We define the utility $u(n, \boldsymbol{x}, a, \theta)$ to be the benefit of choosing n with the outcome $\boldsymbol{x}$ followed by the action a when the parameter has the value θ. Taking expectations over the random quantities θ and $\boldsymbol{x}$, and maximizing over the deterministic quantities a and n, the expected utility is

$$\max_n [\int_\chi \max_a \{\int_\theta u(n, \boldsymbol{x}, a, \theta)\pi(\theta|a, \boldsymbol{x})\, d\theta\} f(\boldsymbol{x}|n)\, d\boldsymbol{x}]. \tag{1.4}$$

If we assume that the utility does not depend on $\boldsymbol{x}$ and includes an additive cost of experimentation proportional to sample size we have $u(n, \boldsymbol{x}, a, \theta) = u(a, \theta) - cn$. The expected utility (1.4) becomes

$$\max_n r(n) = \max_n [\int_\chi \max_a \{\int_\theta u(n, a)\pi(\theta|\boldsymbol{x})\, d\theta\} f(\boldsymbol{x}|n)\, d\boldsymbol{x} - cn].$$

Thus, we need the optimal sample size to maximise $r(n)$. Note that the inference for θ depends on $\boldsymbol{x}$ and not on a and thus we have $\pi(\theta|a, \boldsymbol{x}) = \pi(\theta|\boldsymbol{x})$. This approach is called maximization of expected utility (MEU).

1.1.3 Bayesian decision theory in clinical trials

A major advantage of Bayesian, as opposed to classical frequentist, inference is that it provides a natural process for including information from different sources. This is just what is needed in clinical trials. Bayesian inference also lies at the heart of statistical decision theory, which provides the methodology needed for taking decisions which maximise expected net benefit. This is an important consideration in view of the inevitable pressure on health-care budgets.

Bayesian methods of inference are of course firmly established in clinical trials. The book by Spiegelhalter et al. (2003) is a valuable source in this area. It includes:

- An account of how to select suitable prior distributions and of the importance of doing this carefully, as stressed also by Lambert et al. (2005) and Sambucini (2010).
- A discussion of meta-analysis, which brings together results from different study centres, and is therefore relevant to bridging studies, see also, for example Ades et al. (2005).

- A section on the determination of sample size, though not from a decision theoretic standpoint.

Even so classical methods are still the norm in determining the size of clinical trials.

The use of statistical decision theory for this purpose was advocated by Anscombe (1963). This methodology was based on earlier work by Grundy et al. (1956) and by Raiffa and Schlaifer (1961). Anscombe's case was that in this way the expected net benefit could be maximised, thus avoiding the artificiality of a classical hypothesis-testing criterion based on fairly arbitrary conventional probabilities of type 1 and type 2 errors and an ill-defined 'smallest clinically important difference'. The case was and remains strong, so it is a bit surprising that 47 years later the classical methodology remains by far the most widely used. It is true that the data and computing capacity requirements are greater in the decision theoretic approach, but neither issue is any longer prohibitive.

In fact a number of authors have described applications of statistical decision theory to sample size determination in clinical trials. These include Berry and Ho (1988); Cheng et al. (2003), Claxton et al (Claxton and Posnett, 1996a; Claxton, 1999b; Claxton and Thompson, 2001; Claxton et al., 2002), Eckermann and Willan (Eckermann and Willan, 2007, 2008a,b, 2009; Willan and Eckermann, 2010), Hornberger et al (Hornberger et al., 1995, 1998; Halpern et al., 2001), Lindley (Lindley, 1997), and Willan et al (Willan, 2001; Willan and Pinto, 2005; Willan, 2007; Willan and Kowgier, 2008). A key concept in this development (see in particular Willan and Pinto (2005)) is the expected value of sample information (EVSI). The optimal sample size maximises the difference between EVSI and expected total cost. However, unlike BeBay, there is no explicit modelling of the number of subsequent users as a function of the efficacy and the frequency of adverse reactions revealed by the trial.

The implications of this criterion are worked out in the following contexts: possible delays with associated option values (Eckermann and Willan, 2007, 2008a,b); applications across multiple jurisdictions (Eckermann and Willan, 2009); multi-stage trials (Willan and Kowgier, 2008); and trials followed by decisions depending on the strength of the evidence (Willan and Eckermann, 2010). These papers include instructive practical examples.

Recent years have also seen substantial developments in the design of adaptive clinical trials, mainly not from a decision theoretic standpoint. This work is reviewed by Chang (2007). The aim is to allow modifications to some aspects of the trial as it proceeds without prejudice to its validity. A good

example is provided by Müller and Schäfer (2001) who discuss how group sequential trials may be carried out adaptively. Sample size is noted as being one of the main features in phase III clinical trials which it is desirable to treat adaptively.

1.1.4 Sample size determination, a behavioral Bayes approach

Gittins and Pezeshk (2000a,b, 2002a,b) proposed a decision theoretic approach which they describe as Behavioral Bayes (BeBay) for deciding the optimal sample size in clinical trials. The way BeBay works is as follows. Suppose that the new drug is expected to perform better in terms of efficacy and safety than the current drug. If the sample size is increased, the precision of the estimation of performance after the trial will increase. Thus the evidence for an improved performance is increased in expectation with a large sample size. This means that doctors and their patients will be more likely to be convinced that the new drug is better than the current one. Consequently more patients will want to use the new drug. If more patients switch to the better drug more of them will be cured and the costs of treating those patients are consequently saved. Thus as sample size is increased the expected benefit also increases, as does the cost of the trial. For some sample size the incremental costs and benefits are equal. This is the sample size which maximises the expected net benefit. BeBay can find this sample size and the corresponding expected net benefit.

1.2 THE SCOPE OF THIS RESEARCH

The original BeBay made the restrictive assumptions of paired comparison between two equally sized treatment groups and known effectiveness variance, with no consideration of safety. The weight function expressing patients' preference was crude, and expectations were evaluated by numerical integration, which led to computational difficulties. There was also no detailed account of how to choose the sample size for a bridging trial for extending application of an established drug to a new class of patients.

The motivation for this thesis is to relax these limitations in the original BeBay and to produce a more flexible and practical tool for planning clinical trials. The structure of the thesis is as follows. Following this chapter

of introduction, Chapter 2 describes the extension of the behavioral Bayes approach for finding the optimal sample size to allow unknown variance and a more flexible weight function. Chapter 3 takes account of the fact that patients' choices for their treatments depend not only on efficacy but also on safety, with a suitable redefinition of the weight function. Chapter 4 discusses optimal sample size determination, allowing unequally sized treatment groups. Chapter 5 illustrates the application of BeBay to determine sample size for a cluster randomised clinical trial with the interventions allocated to groups rather than to individual patients. Chapter 6 presents a methodology for the evaluation of a new drug in a bridging trial, with procedures to synthesise efficacy and safety data observed in foreign studies into those of the domestic study. Chapter 7 concludes this thesis with a brief summary of the progress reported in Chapters 2, 3, 4, 5 and 6. These chapters are based, respectively, on the following publications: Kikuchi et al. (2008); Kikuchi and Gittins (2009, 2010b, 2011, 2010a).

Chapter 2

A cost-benefit approach to the determination of sample size

2.1 ABSTRACT

Current practice for sample size computations in clinical trials is largely based on frequentist or classical methods. These methods have the drawback of requiring a point estimate of the variance of the treatment effect and are based on arbitrary settings of type I and II errors. They also do not directly address the question of achieving the best balance between the cost of the trial and the possible benefits from using the new treatment, and fail to consider the important fact that the number of users depends on the evidence for improvement compared with the current treatment. Our approach, Behavioral Bayes (or BeBay for short), assumes that the number of patients switching to the new medical treatment depends on the strength of the evidence which is provided by clinical trials, and takes a value between zero and the number of potential patients. The better a new treatment, the more patients want to switch to it and the more benefit is obtained. We define the optimal sample size to be the sample size that maximises the expected net benefit resulting from a clinical trial. Gittins and Pezeshk (2000a,b) used a simple form of benefit function and assumed paired comparisons between two medical treatments and that the variance of the treatment effect is known. We generalise this setting, by introducing a logistic benefit function, and by extending to the more usual unpaired case, without assuming the variance to be known.

This chapter is based on Kikuchi et al. (2008).

2.2 INTRODUCTION

Current practice for sample size computations in Phase III clinical trials is largely based on frequentist or classical methods to test the null hypothesis that a new treatment is no better than a standard treatment. Classical methods define the sample size of clinical trials by controlling both the size and the power of the test: typical requirements are, at most a 5% chance of incorrectly accepting (not rejecting) that the new treatment is better, and at least an 80% chance of detecting a clinically relevant difference when it is actually present (Desu and Raghavarao, 1990; Lemeshow et al., 1990).

We take our stand with Lindley, who insists that the choice of sample size is clearly a decision and the key ideas of decision analysis are therefore relevant (Lindley, 1997). Although a trial based on frequentist methods has some scientific validity, there is no reason other than conventional practice for believing that a clinical trial should actually be carried out at the calculated sample size. This is because the frequentist methodology does not explicitly balance the cost of the trial against the possible benefits of a new treatment. For example, the condition that a new drug is intended to test may be so rare that there is no realistic prospect of sales revenue covering the cost of further development; in that case it is not in the pharmaceutical company's interests to carry out any trial. On the other hand, if the cost per patient in the trial is low, a trial large enough to achieve more than an 80% chance of detecting a difference which is too small to be normally thought of as clinically relevant may be appropriate; the large meta-analyses which led to the acceptance of aspirin as helpful to the survival chances of stroke victims provide an example (van Gijn, 1999). The frequentist methods also fail to consider the important fact that the number of users, by which we mean primarily medical practitioners, depends on the extent to which they are convinced by the evidence of improvement compared with a standard drug. The increase over several years in the routine use of aspirin to treat actual and potential stroke victims as the evidence for its beneficial effect accumulated provides a striking example of this effect (Gaspoz et al., 2002). In contrast, Bayesian approaches can overcome these defects of the frequentist methodology and also take account of the costs and benefits of medical treatments.

There is a considerable literature on Bayesian methods for sample size determination in clinical trials (Spiegelhalter and Freedman, 1986; Hutton and Owens, 1993; Claxton and Posnett, 1996b; Joseph et al., 1997; Lee and Zelen, 2000). The first Bayesian analysis of the choice of sample size for a

clinical trial to assess cost-effectiveness may be a paper by O'Hagan and Stevens (2001). There is also Bayesian literature based on a more traditional hypothesis testing point of view. For example Spiegelhalter et al. (2003) calculate the sample size for which the posterior probability of efficacy exceeds a decision value, 0.80 say. In another recent similar paper, Tan and Machin (2002) calculate sample sizes by using non-informative prior distributions for phase II clinical trials, and show that the methodology has similar properties to the classic Simon design (Simon, 1989). The approach of Tan and Machin (2002) is extended by Gajewski and Mayo (2006), using a range of more or less optimistic informative prior distributions.

The assessment of costs and benefits is an important question, the answer to which depends on the agency for whom the sample size calculation is being carried out. Pharmaceutical companies are mainly interested in sales profits and development costs and it is therefore natural to use these definitions for cost and benefit to determine sample size. However, health authorities demand that pharmaceutical companies provide evidence that their new drugs are beneficial to public health as a whole. Since the purpose of clinical trials for pharmaceutical companies is to obtain licences, it is also necessary to plan clinical trials to meet the health authorities' requirements. O'Hagan and Stevens (1991) present a novel Bayesian procedure from this standpoint and this type of analysis is increasingly important. The health authorities in Australia (Department of Health and Aging (Australian Government)., 2002) and Canada (Canadian Coordinating Office for Health Technology Assessment., 1997) actually demand information about the costs and benefits of a new treatment. Norway stresses the importance of this kind of information (Norwegian Medicines Agency., 2004). In the UK, the National Institute for Clinical Excellence (NICE) carries out technology appraisals that include analysis of the cost-benefit of interventions including medicines, medical devices, and medical procedures (National Institute for Clinical Excellence, 2004). These cost-benefit appraisals have been used to issue guidance to the UK national health service.

We shall call our approach Behavioral Bayes (or BeBay for short). It involves making plausible assumptions about the possible decisions of the regulator and of potential users or patients, and the expected net benefit resulting from the clinical trial is maximised. Some of the key principles of BeBay are as follows.

- The total expected benefit of a clinical trial minus the cost is maximised.
- Benefits can be realised only if potential users are convinced that the

new treatment has sufficient efficacy.

- The number of patients switching to the new treatment depends on the strength of the evidence provided by clinical trials, and takes a value between zero and the total number of sufferers from the condition to be treated.
- The total benefit is assessed as a function of the number of patients who switch to the new treatment.
- If the expected net benefit is non-positive for all sample sizes, it is concluded that it is not worth carrying out the clinical trial.

The structure of this chapter is as follows. In section 2.3, the mathematical preliminaries are explained for the sample size determination in cases where the mean and variance of the efficacy difference between two treatments are unknown. In section 2.4, the benefits and their dependence on the behaviour of subsequent users are modelled. Section 2.5 describes the calculation of expected net benefit using Monte Carlo simulation. Section 2.6 describes and comments on some examples of such calculations. Section 2.7 concludes this chapter with some discussion.

The purpose of the chapter is to set out an extended and much more flexible version of the new methodology which we have described in earlier papers (Gittins and Pezeshk, 2000a,b). It is widely applicable in phase III trials: in the main trials for a new drug, or in extending the range of a drug, for example to the extrapolation of adult results to juveniles, or in a bridging study designed to achieve registration in a new region. For simplicity our discussion refers to sample size determination in a single trial. However the methodology may also be applied when the total sample is split over several trials, for example in a multi-centre clinical trial.

2.3 DISTRIBUTION THEORY

Suppose that X and Y are a new treatment and a standard treatment, respectively, and that the treatment groups have the same sample size $n(>1)$. Let two continuous variables, X_i and Y_i for $i = 1, 2, \cdots, n$, be the clinical outcomes on some appropriate scale such as blood pressure. The subscript i refers to patient i in each treatment group, and X_i and Y_i are unpaired.

If $X_i \sim \mathrm{N}(\theta + \delta, \sigma^2)$ and $Y_i \sim \mathrm{N}(\theta, \sigma^2)$, for $i = 1, 2, \cdots, n$, then, writing

$\overline{X_n} = (\sum_{i=1}^{i=n} X_i)/n$, $\overline{Y_n} = (\sum_{i=1}^{i=n} Y_i)/n$, and $\overline{Z_n} = \overline{X_n} - \overline{Y_n}$, it follows that

$$\overline{Z_n} \sim \mathrm{N}(\delta, 2\sigma^2/n). \tag{2.1}$$

By standard statistical theory, S_x^2/σ^2 and S_y^2/σ^2, where $S_x^2 = \sum_{i=1}^{i=n}(X_i - \overline{X_n})^2$ and $S_y^2 = \sum_{i=1}^{i=n}(Y_i - \overline{Y_n})^2$, both have chi-squared distributions with $n-1$ degrees of freedom, and thus $S^2/\sigma^2 = S_x^2/\sigma^2 + S_y^2/\sigma^2$ follows χ^2_{2n-2}. Let $f(s_x^2, s_y^2, \overline{z_n}|\delta, \sigma^2)$ denote the likelihood function for S_x^2, S_y^2 and $\overline{Z_n}$, which is proportional to

$$(\sigma^2)^{-\frac{(2n-1)}{2}} \exp[-\frac{1}{2\sigma^2}\{s_x^2 + s_y^2 + \frac{n}{2}(\overline{z_n} - \delta)^2\}]. \tag{2.2}$$

Following O'Hagan and Forster (1994), the conjugate joint prior density function for δ and σ^2 has the form

$$\pi(\delta, \sigma^2) = k(a, g, \omega)(\sigma^2)^{-\frac{g+3}{2}} \exp^{-\frac{1}{2\sigma^2}\left\{a + (\delta-\mu)^2/\omega\right\}}, \tag{2.3}$$

where

$$k(a, g, \omega) = a^{\frac{g}{2}} 2^{-\frac{g+1}{2}} (\pi\omega)^{-\frac{1}{2}} \left\{\Gamma(g/2)\right\}^{-1},$$

and a, g and ω are hyper-parameters assigned on the basis of prior information. Note that $\pi(\delta, \sigma^2) = \pi(\sigma^2)\pi(\delta|\sigma^2)$, where $\pi(\sigma^2)$ is an inverse chi-squared density function with parameters $g/2$ and $a/2$, and $\pi(\delta|\sigma^2)$ is $\mathrm{N}(\mu, \sigma^2\omega)$. In terms of the mean and variance of the prior distribution for σ^2,

$$g = 4 + 2\mathrm{E}(\sigma^2)^2/\mathrm{var}(\sigma^2), \text{ and} \tag{2.4}$$
$$a = \mathrm{E}(\sigma^2)(g-2), \text{ and} \tag{2.5}$$
$$\omega = \tau^2/\mathrm{E}(\sigma^2), \tag{2.6}$$

where τ^2 is the variance of the prior distribution for δ. Note that we must have $g > 4$ since $\mathrm{Var}(\sigma^2) = 2\mathrm{E}(\sigma^2)^2/(g-4)$.

Applying Bayes theorem and using (2.2) and (2.3), the posterior density for δ and σ^2 may be written,

$$\pi^{(n)}(\delta, \sigma^2 \mid x, y) = k(a', g', \omega')(\sigma^2)^{-\frac{g'+3}{2}} \exp^{-\frac{1}{2\sigma^2}\left\{a' + (\delta-\mu')^2/\omega'\right\}}.$$

Here the function $k(a', g', \omega')$ is as in (2.3), $x = (x_1, x_2, \cdots, x_n)$, $y = (y_1, y_2, \cdots, y_n)$, $s^2 = s_x^2 + s_y^2$,

$$\omega' = \frac{2\omega}{2 + n\omega}, \qquad \mu' = \frac{2\mu + n\omega\overline{z_n}}{2 + n\omega},$$
$$g' = g + 2n - 1, \qquad a' = a + s^2 + \frac{n(\overline{z_n} - \mu)^2}{2 + n\omega}.$$

The mean and variance of the prior distribution for δ are μ and $\omega a/(g-2)\,(=\tau^2)$, respectively. The marginal posterior density function of δ can be obtained by integrating the joint posterior density over σ^2. This gives

$$\pi^{(n)}(\delta \mid x, y) = \frac{1}{B(g'/2, 1/2)}(a'\omega')^{-\frac{1}{2}} \left\{1 + \frac{(\delta - \mu')^2}{a'\omega'}\right\}^{-\frac{g'+1}{2}}.$$

Therefore, $(\delta - \mu')/\sqrt{\omega' a'/g'}$ has a t-distribution with g' degrees of freedom. Thus, the mean of the posterior distribution for δ is μ' and its variance is $\tau'^2 = \omega' a'/(g'-2)$. We can also make inferences about σ^2 using

$$\pi^{(n)}(\sigma^2 \mid x, y) = \frac{(a'/2)^{\frac{g'}{2}}}{\Gamma(g'/2)}(\sigma^2)^{-\frac{g'+2}{2}} \exp(-\frac{a'}{2\sigma^2}).$$

These expressions provide the basis for the calculation of expected net benefit functions in this chapter. Minor changes suffice to deal with the case of paired comparisons.

The values of the parameters for the prior distributions are based on previous experience. This always includes the results of phase II trials, For a range extension study results from widespread earlier use of the drug will also be available.

2.4 COST AND BENEFIT FUNCTIONS

2.4.1 The net benefit function

An essential feature in our model is that the number of new users depends, among other things, on the posterior distribution for δ, the mean efficacy difference between two treatments. The model assumes that the number of patients who want to use the new treatment is greater if their medical advisers are convinced that it is more effective. The net benefit of a new treatment is estimated by subtraction of the cost for a clinical trial from the total benefit. The cost and the benefit may be defined as follows.

- The cost of a trial with n patients is $cn + l$.
- The benefit per patient switching to the new treatment is $b\delta + k$ in monetary terms.

Notes.

- The cost c includes the expected cost per patient of adverse reactions during the trial.
- The cost l includes an estimate for all the subsequent costs of bringing the drug to market, as well as the fixed cost incurred by running the trial.
- The benefit $b\delta + k$ is the appropriately discounted aggregate net benefit for one patient over the period for which the new treatment remains in use.

The constants b and c are supposed to be known, or capable of being estimated. To keep our discussion simple we shall set $k = 0$ and $l = 0$, so that the benefit and the cost per patient are $b\delta$ and c, respectively.

Let us turn our attention to the question of how big a clinical trial should be. We define the optimal sample size to be the one for which the expected net benefit from a clinical trial is maximised. It is necessary for the optimisation to formulate the expected net benefit function $r(n)$ and then to maximise it. Suppose the unit benefit $b\delta$ per patient is weighted by a weight function m. A weight function estimates the number of subsequent users of the new treatment and is a function of the sufficient statistics for δ and σ^2, $\overline{z_n}$ and s^2. For the calculation of $r(n)$, the joint density function $f(\delta, \sigma^2, \overline{z_n}, s^2)$ is used. The expected net benefit may be written as

$$\begin{aligned} r(n) &= \int_{s^2}\int_{\overline{z_n}}\int_{\sigma^2}\int_{\delta} b\, m\, \delta\, f(\delta, \sigma^2, \overline{z_n}, s^2)\, d\delta\, d\sigma^2\, d\overline{z_n}\, ds^2 - cn \\ &= \int_{s^2}\int_{\overline{Z_n}} b\, m\, \mu' f(\overline{z_n}, s^2)\, d\overline{z_n}\, ds^2 - cn \\ &= b\mathrm{E}[m\mu'] - cn, \end{aligned} \tag{2.7}$$

the expectation being with respect to the prior distribution for $\overline{z_n}$ and s^2. To evaluate this expression, the following identity can be used,

$$f(\overline{z_n}, s^2) = \int_{\sigma^2}\int_{\delta} \pi(\delta, \sigma^2) f(\overline{z_n} \mid \delta, \sigma^2) f(s^2 \mid \delta, \sigma^2)\, d\delta\, d\sigma^2,$$

noting that $\overline{z_n}$ and s^2 are conditionally independent if σ^2 and δ are given.

If the costs and benefits under consideration are all direct financial consequences for the company carrying out the trial, then the expected net benefit is the expected net present value to the company of the cash flows associated with the trial and subsequent production and marketing of the drug. The

need for regulatory approval can be modelled by assuming that a minimum performance level is required for approval, expressed either in Bayesian or in frequentist terms. Several specific examples are discussed in paper (Gittins and Pezeshk, 2000a).

There may also be regulatory requirements for the conduct of the trial, in particular for its size. If these mean that the trial has to be larger than BeBay suggests, compliance with the requirements and therefore a larger trial will normally be necessary. If the expected net benefit from the larger trial is negative then no trial should be carried out. On the other hand it is less likely that the regulator will insist on reducing the suggested trial size.

A different perspective arises if the trial is being commissioned by a national health authority. The price of the drug is now part of the costs. The benefit per patient requires some monetary measure for different states of well-being. And the change, which could be positive or negative, in the cost of treatment per patient resulting from switching to the new drug needs to be evaluated. This information can be input to BeBay, which again produces an expected net benefit function and a suggested sample size.

2.4.2 Simple weight function

We assume that the number of subsequent users of the new treatment depends on the posterior distribution for δ. The function defining this relationship is called a weight function and has many plausible forms. One of the simplest weight functions is proposed in Gittins and Pezeshk (2000a):

$$m = \begin{cases} 0 & \mu' < A' \ , \\ \frac{M}{B-A}\left[\mu' - A'\right] & A' \leq \mu' \leq B', \\ M & B' < \mu', \end{cases} \tag{2.8}$$

where $A' = A + 1.5\tau'$, $B' = B + 1.5\tau'$, μ' and τ' are the posterior mean and standard deviation for δ, and M is the total number of potential users. Let us call it the Pezeshk-Gittins weight function. The two parameters, A and B, are the lower and the upper critical points between which the number of subsequent users would vary linearly with δ if its value was known.

This function is based on the assumptions that after the clinical trial

- some patients will switch to the new treatment from the current treatment if the difference in efficacy exceeds the lower threshold A with a probability of at least 0.87, and

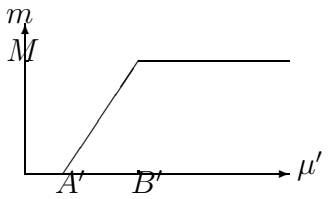

Figure 2.1: Pezeshk-Gittins weight function

- all patients switch to the new therapy if the difference exceeds the upper threshold B with a probability of at least 0.87.

The values of A and B need to be estimated. They depend on the severity of the condition to be treated, and on any differences in convenience in use or in expected cost between the new treatment and the current treatment. The principles involved are discussed in Gittins and Pezeshk (2000b), and there are several practical examples in Gittins and Pezeshk (2000a). The choice of the multiplier 1.5 is of course arbitrary. For example, we could replace it with 1.96, in which case the required probability is at least 0.95.

Using this weight function and (2.7), the expected net benefit function (for $n \geq 2$) may be written as

$$r(n) = \int\int_{A' \leq \mu' \leq B'} \frac{Mb}{B-A} \left(\mu' - A - 1.5\sqrt{\frac{\omega' a'}{g'-2}} \right) \mu' f(\overline{z_n}, s^2)\, d\overline{z_n}\, ds^2 + Mb \int\int_{\mu' > B'} \mu' f(\overline{z_n}, s^2)\, d\overline{z_n}\, ds^2 - cn.$$

2.4.3 Logistic weight function

The Pezeshk-Gittins weight function is rational and easy to interpret. However, the derivative does not exist at the points A' and B', which is unrealistic. Fortunately, Monte Carlo simulation allows any form of weight function to be used to calculate the expected net benefit. Let us introduce a smooth weight function m_l as follows. This function is a simple logistic function,

$$m_l = \frac{M}{1 + \phi \exp\{\frac{2\lambda(1.5\tau' - \mu')}{B-A}\}}, \tag{2.9}$$

where $\phi = \exp\{\frac{\lambda(A+B)}{(B-A)}\}$, $\lambda > 0$, and the number of subsequent users is $M/2$ if $\mu' = (A+B)/2 + 1.5\tau'$. This logistic function approximates the Pezeshk-Gittins weight function for the same values of A and B. Figure 2.2 is for the

case $\lambda = 1$. The two weight functions would agree more closely with a larger value of λ. With a benefit per patient of $b\delta$, the total expected benefit is thus

$$bm_l\mu' = \frac{Mb\mu'}{1 + \phi \exp\{\frac{2\lambda(1.5\tau' - \mu')}{B-A}\}}.$$

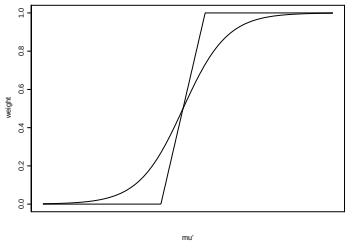

Figure 2.2: Pezeshk-Gittins and logistic weight functions

2.5 MONTE CARLO SIMULATION

Monte Carlo simulation is a more efficient and flexible method of calculation than the quadrature procedures used by Gittins and Pezeshk (2000a,b); these run into problems with accuracy, and are difficult to generalise to non-normal efficacy distributions and other weight functions. The main idea of Monte Carlo simulation is that the integral $\int_a^b g(x)p(x)\,dx$ may be approximated by, or $\cong$, $\frac{1}{k}\sum_{j=1}^{k} g(x_j)$, where the points $x_1, x_2 \cdots, x_k$ are chosen as independent pseudo-random samples with density $p(x)$ on the interval (a, b). Expectations over more complex sample spaces may be evaluated in similar fashion.

If Monte Carlo simulation is used in our model, any form of weight functions can be used to calculate the optimal sample size and the maximised expected net benefit. This section discusses the calculation procedure. The posterior expected benefit $bm\mu'$ is a function of s^2 and $\overline{z_n}$. Thus, Monte Carlo calculation of $b\mathrm{E}[m\mu']$ (see (2.7)) may be carried out by using a large number of random samples for s^2 and $\overline{z_n}$.

From (2.3), we have

$$\begin{aligned}\pi(\delta, \sigma^2) &= k(a, g, \omega)(\sigma^2)^{-\frac{g+3}{2}} \exp^{-\frac{1}{2\sigma^2}\left\{a + (\delta-\mu)^2/\omega\right\}} \\ &= \pi(\sigma^2)\pi(\delta|\sigma^2),\end{aligned}$$

where $\pi(\sigma^2)$ is an inverse chi-squared density function with parameters $g/2$ and $a/2$,

$$\pi(\sigma^2) = (\frac{a}{2})^{\frac{g}{2}} \frac{1}{\Gamma(g/2)} (\sigma^2)^{-\frac{g+2}{2}} \exp(-\frac{a}{2\sigma^2}) \tag{2.10}$$

and $\pi(\delta|\sigma^2)$ is $\mathrm{N}(\mu, \sigma^2\omega)$, so that

$$\pi(\delta|\sigma^2) = \frac{1}{\sigma\omega^{\frac{1}{2}}\sqrt{2\pi}} \exp^{-\frac{1}{2}\frac{(\delta-\mu)^2}{\sigma^2\omega}} . \tag{2.11}$$

Thus, given σ^2, δ can be sampled from a normal distribution. For $\overline{z_n}$ and s^2, we use (2.1) and a chi-squared distribution with $2n-2$ degrees of freedom, respectively, noting that $\overline{z_n}$ and s^2 are conditionally independent given σ^2 and δ. Note that s^2 is conditionally independent of δ given σ^2. Figure 2.3 summarises the mutual dependence of these parameters and statistics and indicates a possible algorithm for generating Monte Carlo random samples of $\overline{z_n}$ and s^2. The stepwise procedure is as follows; n is the number of patients

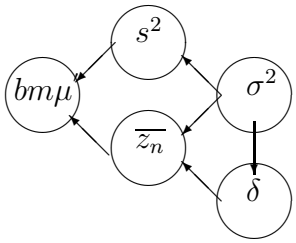

Figure 2.3: Monte Carlo sampling

in each group and N is the number of repeat simulations.

1. Set $n = 2$.
2. Draw samples $\sigma^2_{(j)}$ for $j = 1, 2 \cdots, N$ from the probability density function (2.10).
3. Draw a sequence of samples for $(\delta_{(1)}, \delta_{(2)} \cdots, \delta_{(N)})$ for δ from $\mathrm{N}(\mu, \sigma^2\omega)$, (see (2.11)), using the values $(\sigma^2_{(1)}, \sigma^2_{(2)} \cdots, \sigma^2_{(N)})$ for σ^2.
4. Sample N corresponding values $(s^2_{(n,1)}, s^2_{(n,2)}, \cdots, s^2_{(n,N)})$ of $s^2_{(n)}$ with these sample values of σ^2, using the fact that $\frac{s^2_{(n)}}{\sigma^2}$ has a chi-squared distribution with $2n-2$ degrees of freedom.
5. Sample values $(\overline{z_{(n,1)}}, \overline{z_{(n,2)}}, \cdots \overline{z_{(n,N)}})$ of $\overline{z_n}$ by sampling from $\mathrm{N}(\delta, 2\sigma^2/n)$, (see(2.1)), with values $(\delta_{(1)}, \delta_{(2)} \cdots, \delta_{(N)})$ for δ and $(\sigma^2_{(1)}, \sigma^2_{(2)} \cdots, \sigma^2_{(N)})$ for σ^2.

6. Calculate N samples $l_{(n,j)}$, $j = 1, 2, \cdots, N$, of the total benefit, using the formula

$$l_{(n,j)} = m(\overline{z_{(n,j)}}, s^2_{(n,j)})\mu'(\overline{z_{(n,j)}}).$$

The estimated expected net benefit is given by

$$r(n) \cong b\frac{\sum_{j=1}^{N} l_{(n,j)}}{N} - cn.$$

7. Set $n = n + 1$ and repeat steps 2 to 7.

8. Find the optimal sample size and the maximised expected net benefit from the estimated values for $r(n)$.

2.6 SIMULATION RESULTS

2.6.1 Behaviour of the expected net benefit function

The mean and variance of the marginal prior distribution for δ are

$$\mu \text{ and } \omega a/(g-2).$$

respectively. The mean, variance and mode for the marginal prior density of σ^2 are $a/(g-2)$, $\frac{2a^2}{(g-2)^2(g-4)}$ and $a/(g+2)$, respectively. All the available information about the treatment effect is thus expressed via a, g, ω and μ. Note that inverse chi-squared distributions are positively skewed, so that the mean is larger than the mode. This seems a reasonable property to require for $\pi(\sigma^2)$.

Let us consider as an illustration the following artificial parameter values for which the mean and variance of the prior distribution for σ^2 are 4 and 1.07^2, respectively, and the prior variance for δ is equal to 1.

$$A = 1.5,\ B = 2.5,\ a = 120,\ g = 32,\ \omega = 0.25, Mb = 5 \times 10^6 \text{ and } c = 6000. \tag{2.12}$$

These values are within the range of the case studies discussed in Gittins and Pezeshk (2000b). The number N of Monte Carlo samples is 2.5×10^4 throughout this section, and we shall write $R(n)$ for $r(n)/Mb$.

The values of the hyper-parameters a, g and ω are fixed by the prior variance τ^2 of δ and the prior mean $\mathrm{E}(\sigma^2)$ and variance $\mathrm{Var}(\sigma^2)$ of σ^2. Equations (2.4), (2.5) and (2.6) are used. These quantities are themselves determined by the

extent of the availability of data from previous similar trials. There would be a lot of relevant data, for example, if the trial is for a new class of patients with an established drug. For the values in our example the prior uncertainty about δ is comparable to $B - A$, so there is considerable uncertainty as to the number of patients who would use the new treatment. The high prior probability of a value of σ^2 close to 4 means that we expect the responses of individual patients to vary very significantly.

The logistic weight function (2.9) is used in this section with $\lambda = 1$. Simulated values of $R(n)$ plotted against n are shown for $\mu = 0.0, 0.5, 1.0, 2.0, 3.0$ and 4.0 in Figure 2.4. Estimates of the maximising values of n and the maximised $R(n)$ are shown in Table 2.1. The curves in the figures are smoothed by using the S-Plus generic function, *smooth.spline* with the default degree of freedom. The properties of the expected net benefit function which emerge from these

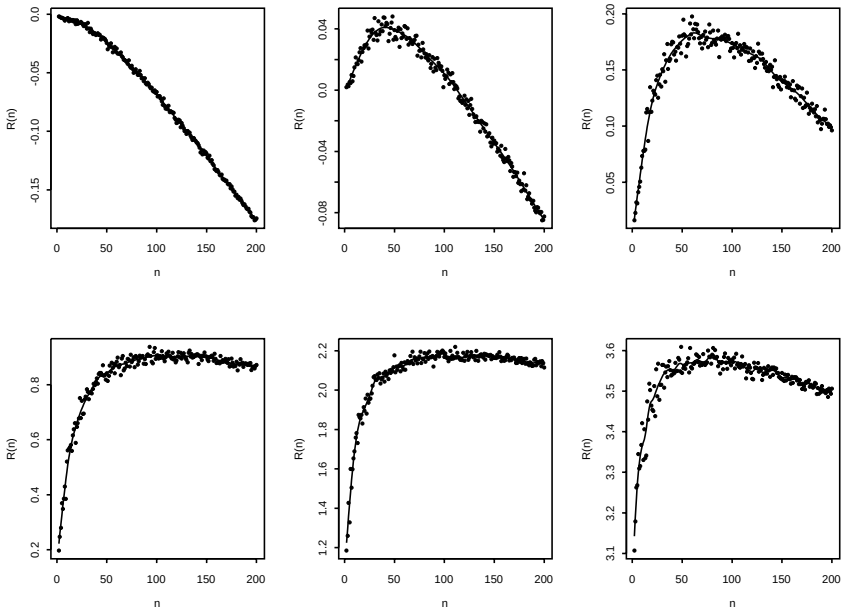

Figure 2.4: The expected net benefit by simulation (top panels from left to right, $\mu = 0.0, 0.5,$ and 1.0; bottom panels from left to right, $\mu = 2.0, 3.0,$ and 4.0.)

calculations may be summarised as follows.

- If the prospects of the new drug are poor, which is true, for example, for large negative values of μ, $R(n)$ is negative for all values of n. It

follows that no clinical trial is worth carrying out.

- If the value of μ is large, any trial simply confirms that the new drug is very promising, and only a small clinical trial is necessary.
- The optimal sample size is greatest for intermediate values of μ.
- If a drug is promising, the range of sample sizes which give positive expected net benefit is broad.

Table 2.1: Optimal sample sizes and $R(n)$ for various μ

μ	n	$R(n)$
0.0	0	0.00
0.5	41	0.04
1.0	62	0.18
2.0	98	0.91
3.0	97	2.18
4.0	81	3.59

Now let us suppose that the regulator takes the standard classical viewpoint and requires a trial of a size which ensures that the probabilities of type I and type II errors are at most α and β respectively, for a clinically relevant difference in outcome of Δ. This means (see for example Dobson (1984)), using the conventional known variance normal theory, that

$$n = 2\sigma^2(z_{\alpha/2} + z_\beta)^2/\Delta^2,$$

where z_ω denotes the $100(1-\omega)\%$ point of the standard normal distribution. For our example we might estimate σ^2 to be 4.0, follow convention by putting $\alpha = 0.05$ and $\beta = 0.2$, and set $\Delta = 1.0\,(= B - A)$, leading to a required sample size of at least 63. For a pharmaceutical company using BeBay this would suggest a sample size of

- 0 for $\mu < 0.5$ (since $R(63) < 0$),
- 63 for $0.5 \leq \mu \leq 1.0$ and for $\mu > \mu^*$ (to satisfy the regulator),
- the value derived from BeBay for $1.0 < \mu \leq \mu^*$.

Here, $\mu^* > 4.0$ and is such that the optimal n calculated by BeBay is 63.

2.6.2 Sensitivity analysis

It is of interest to explore the dependence of the expected net benefit and optimal sample size on the prior variance τ^2 for δ and the prior variance $\mathrm{Var}(\sigma^2)$ for σ^2. To increase τ^2 we need to increase ω, and to increase $\mathrm{Var}(\sigma^2)$ we need to decrease a and g without changing the value of $a/(g-2)$, so that τ^2 and $\mathrm{E}(\sigma^2)$ do not change. Tables 2.2 and 2.3 show how the optimal sample size n and the maximised expected net benefit $R(n)$ change when each of these variances is increased. The subscript b indicates values calculated for the base parameter values listed in (2.12) with $\mu = 2.0$. For the other tabulated values of n, $R(n)$ etc. the parameter values remain the same except for the indicated changes in ω and in a and g. Note that $\tau^2 \to \infty$ as $\omega \to \infty$ and $\mathrm{Var}(\sigma^2) \to \infty$ as $g \downarrow 4$, so that both tables show what happens as the prior distribution approximates more closely to an ignorance prior distribution.

From Table 2.2 we can see that the maximal $R(n)$ increases strongly as τ^2 increases, while the optimising value of n decreases sharply. This is because a vague prior allows the possibility of very large values of δ, and therefore very large benefits, while on the other hand only a small sample is required to determine whether the new treatment should be used or not, except in the unlikely event that δ is close to A and B. This means that it is important to take realistic account of prior information in determining the value of τ^2.

Table 2.2: Effect on n and $R(n)$ of increasing values of τ^2

ω	τ^2/τ_b^2	τ/μ	n/n_b	$R(n)/R(n_b)$
0.25	1	0.5	1.0	1.0
0.75	3	0.87	0.87	1.4
1.25	6	1.12	0.70	1.7
2.5	10	1.58	0.51	2.2
25	100	5.00	0.60	5.3
250	1,000	15.8	0.02	14.7
2500	10,000	50	0.02	44.5

In contrast, Table 2.3 shows that the optimising n and maximal $R(n)$, while they do vary with $\mathrm{Var}(\sigma^2)$, do not do so very markedly.

Table 2.3: Effect on n and $R(n)$ of increasing values of $\text{Var}(\sigma^2)$

a	g	$\text{var}(\sigma^2)/\text{var}(\sigma_b^2)$	n/n_v	$R(n)/R(n_b)$
120	32	1.0	1.0	1.0
8.4	4.1	280	0.824	0.991
8.04	4.01	2800	0.765	0.991
8.004	4.001	2,8000	0.866	0.995
8.0004	4.0001	280,000	0.992	0.977

2.6.3 Comparisons between the two weight functions

To compare the behaviour of a logistic weight function ($\lambda = 1$) with the Pezeshk-Gittins weight function (sec Section 2.4.2), let us consider the following values,

$$A = 1.7,\ B = 2.5,\ a = 480,\ g = 122,\ \omega = 0.3, Mb = 5 \times 10^6 \text{ and } c = 4000.$$

These values of a, g and ω mean that the prior variance τ^2 for δ is 1.2 and the prior mean $\text{E}(\sigma^2)$ and the variance $\text{Var}(\sigma^2)$ for σ^2 are 4 and 0.521^2, respectively. Figure 2.5 shows the results of calculations with $\mu = 0.5, 1.0, 2.0, 3.0$ and 4.0. Note that the horizontal and vertical scales are all different in these figures. The optimal sample sizes and the maximised expected net benefits are summarised in Table 2.4.

The behaviour of the expected net benefit function is similar for the two weight functions (Figure 2.5). Optimal sample sizes are almost identical (Table 2.4) but the values of $R(n)$ diverge as μ is decreased toward zero. In general, since we have the possibility of using the more plausible logistic weight function it is sensible to do so, and it is likely that the Pezeshk-Gittins weight function may provide an adequate alternative in some cases.

Table 2.4: Comparison of the optimal values of n and $R(n)$

μ	Logistic		Pezeshk-Gittins	
	n	$R(n)$	n	$R(n)$
0.5	69	0.057	69	0.019
1.0	102	0.21	104	0.16
2.0	143	0.96	143	0.96
3.0	169	2.22	170	2.35
4.0	93	3.62	92	3.75

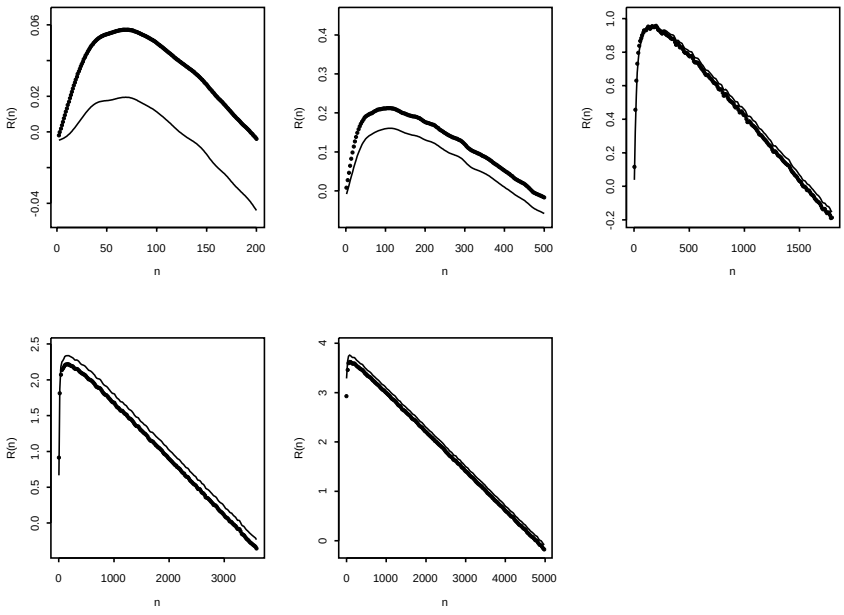

Figure 2.5: The behaviour of the Pezeshk-Gittins weight function (thin lines) and logistic weight function (thick lines) : top panels from left to right, $\mu = 0.5$, $\mu = 1.0$ and $\mu = 2.0$; bottom panels from left to right, $\mu = 3.0$ and $\mu = 4.0$.

2.7 CONCLUDING REMARKS

Frequentist methods for sample size calculation have the drawbacks of being based on a point estimate for the variance of the treatment effect, and on arbitrary settings of type I and II errors. They also fail to balance cost against the benefit expected from the trials, and take no explicit account of previous knowledge. They do not in particular provide direct guidance as to whether a clinical trial is actually worth carrying out. In contrast, BeBay uses a Bayesian approach which remedies these deficiencies. BeBay methodology is applicable under a great variety of distributional assumptions (Gittins and Pezeshk, 2000b), and with any weight function. It does, of course, require all costs and benefits to be converted into monetary values. This could mean straight-forward income and expenditure for the company carrying out the trial.

We note that our results support the view that the simple Pezeshk-Gittins weight function may give sufficient accuracy to determine the optimal sample

size in a clinical trial. This tentative conclusion warrants further investigation.

There are also potential applications in which monetary costs and benefits are assigned to health outcomes, including the costs of negative adverse reactions. A more detailed consideration of how to allow for the costs of adverse reactions, both during the trial and in later use, is desirable. In this extended model, the costs of adverse reactions express the degree of safety for a new drug. The costs of adverse reactions are considered in chapters 3, 4 and 6.

Another important benefit to consider is the saving to the national health budget in the case of a new treatment for a chronic condition. Calculations along these lines influence the decisions of regulatory authorities on whether to grant licences to new drugs. Chapters 3, 4, 5 and 6 take this standpoint.

Chapter 3

Sample size considering efficacy and safety

3.1 ABSTRACT

It is necessary for the calculation of sample size to achieve the best balance between the cost of a clinical trial and the possible benefits from a new treatment. Gittins and Pezeshk (2000b) have developed a decision theoretic approach which they describe as behavioral Bayes. This assumes that the number of users is an increasing function of the difference in performance between the new treatment and the standard treatment. The better a new treatment, the more the number of patients who want to switch to it. The optimal sample size is calculated in this framework. This BeBay approach takes account of three decision-makers, a pharmaceutical company, the health authority and medical advisers. Kikuchi et al. (2008) generalised this approach by introducing a logistic benefit function, and by extending to the more usual unpaired case, and with unknown variance. The expected net benefit in this model is based on the efficacy of the new drug but does not take account of the incidence of adverse reactions. The present chapter extends the model to include the costs of treating adverse reactions and focuses on societal cost-effectiveness as the criterion for determining sample size. The main application is likely to be to phase III clinical trials, for which the primary outcome is to compare the costs and benefits of a new drug with a standard drug in relation to national health-care.

This chapter is based on Kikuchi and Gittins (2009).

3.2 INTRODUCTION

Current practice for sample size computations in clinical trials is largely based on frequentist or classical methods. These methods have the drawbacks that they require a point estimate of the variance of the treatment effect, are based on arbitrary settings of type I and II errors and make no explicit use of prior information. They also do not directly address the question of achieving the best balance between the cost of the trial and the possible benefits from using the new treatment, and fail to consider the important fact that the number of subsequent users depends on the evidence for improvement compared with the current treatment. These issues are addressed by Gittins and Pezeshk (2000a,b, 2002a,b) in their behavioral Bayes (BeBay) methodology for sample size determination. The classical methodology is set out, for example, in references (Briggs and Gray, 1998; Laska et al., 1999; Willan and O'Brien, 1999; Garduber et al., 2000).

BeBay considers the roles of three distinct decision-makers. These are the agency carrying out the trial, often a pharmaceutical company, the regulatory authority, which decides whether to give a licence for any new drug, and the ultimate consumer, or more realistically his or her medical adviser, who chooses whether or not to use the new drug if it is ultimately marketed. The examples discussed in a paper (Gittins and Pezeshk, 2000b) show that BeBay can lead to very different sample sizes, either higher or lower, compared with those derived from classical frequentist criteria. These criteria are not always appropriate and can lead to seriously sub-optimal outcomes.

In a recent paper (Kikuchi et al., 2008), Kikuchi et al. describe extensions to BeBay which allow the treatment effect to have unknown variance and a wider range of functions for the number of subsequent users. This paper also includes an account of the substantial contributions to Bayesian methods by other authors (Spiegelhalter and Freedman, 1986; Hutton and Owens, 1993; Claxton and Posnett, 1996b; Joseph et al., 1997; Lee and Zelen, 2000).

Two further interesting papers are those by Brutti et al. (2008) and by Wang and Gelfand (2002). They address the problems of how to select a sample size to achieve specified performance with regard to one or more features of a model, and how to select a sample size to achieve specified separation of two models. However, neither paper considers the optimisation of net expected benefit, an approach which may be described as decision analytic. The same is true of papers (Spiegelhalter and Freedman, 1986; Hutton and Owens, 1993; Joseph et al., 1997; Lee and Zelen, 2000).

There have been relatively few papers setting out a decision analytical framework for sample size determination for clinical trials. Two which do are those by Claxton and Posnett (1996b) and by Patel and Ankolekar (2007). Their models incorporate relevant economic factors in interesting ways. However, unlike BeBay, there is no explicit modelling of the number of subsequent users as a function of the efficacy and the frequency of adverse reactions revealed by the trial.

In paper (Kikuchi et al., 2008) our model assumed that patients would decide whether to switch to a new drug on the basis of efficacy alone. This leaves out of account the important influence of safety considerations. The present chapter further expands the scope of BeBay with a discussion of the effect of adverse reactions, and of national medical-care costs and benefits. This chapter is structured as follows. First we describe the distribution theory needed for a Bayesian analysis of the evidence for efficacy and for adverse reactions provided both before and during a phase III trial, assuming that there is exchangeability among patients in phases II and III, and proceed to explain how BeBay calculates expected costs and benefits per patient. Then a weight function is introduced to calculate the number of subsequent users and the resulting benefit, and an algorithm is described which carries out Monte Carlo simulation to evaluate the total expected net benefit, and hence the optimal sample size. All this is illustrated by a practical example.

The agency carrying out the trial and deciding the sample size is likely to be the pharmaceutical company which is planning to produce the new drug. In this chapter the costs and benefits which we consider are those which affect a national health authority, excluding the cost of the drug itself in use after the trial. This means that the expected net benefit, which our model allows us to calculate, may in turn be used to calculate the maximum price per patient year which it would be cost-effective for the authority to pay for the drug. The drug company must decide on this basis whether it is worth conducting the trial. This means assessing likely production costs, and making a judgment on the price which it is likely to be able to negotiate with the health authority. The chapter concludes with a discussion of applications of the BeBay approach to issues other than sample size determination, followed by final comments.

3.3 THE MODEL

3.3.1 Assumptions

Throughout our discussion we assume that prior information on the parameters of probability distributions is expressible by means of conjugate prior distributions. Frequently this is a realistic assumption. The reasons for making it are that it simplifies what is in any case a complex presentation and ensures that the calculations which we need to make are computationally feasible (see for example Bernardo and Smith (1994)).

We shall also assume that the prior distributions for the difference in efficacy between the new and standard treatments, and for the incidence of adverse reactions with each of the treatments, are independent. Again this is to simplify our presentation. In general there may be a tendency for more active drugs to be both more efficacious and to cause more adverse reactions, leading to dependent prior distributions. However, there are also many cases for which an assumption of independence is reasonable; published examples include calcineurin inhibitors for immunosuppression in liver transplantation (Perry and Neuberger, 2005), rosuvastatin to reduce low-density lipoprotein cholesterol (Olsson et al., 2001; Davidson et al., 2002; Saito et al., 2003), and infliximab (a monoclonal antibody against Tumor Necrosis Factor) for Crohn's disease (Targan et al., 1997).

The values of the prior distribution hyperparameters, both for efficacy and for the incidence of adverse reactions, are based on previous experience. This always includes the results of phase II trials. For a range extension study results from widespread earlier use of the drug will also be available. Exactly how to estimate these prior hyperparameters is a matter for case by case judgment. In the example in this chapter we estimate the mean value for efficacy by the mean value of phase II efficacy, but assume substantial uncertainty about this estimate to reflect the differences between the conditions for phases II and III. For adverse reactions, and to simplify the presentation, we throughout make the assumption of exchangeability between patients in phases II and III.

3.3.2 Distribution theory (efficacy)

Suppose that X and Y are a new treatment and a standard treatment, respectively, and that the treatment groups have the same sample size $n\,(>1)$.

Let two independent continuous variables, X_i and Y_i for $i = 1, 2, \cdots, n$, be the clinical outcomes on some appropriate scale. The subscript i refers to patient i in each treatment group, and X_i and Y_i are unpaired and independent. If $X_i \sim \mathrm{N}(\theta+\delta, \sigma^2)$ and $Y_i \sim \mathrm{N}(\theta, \sigma^2)$, for $i = 1, 2, \cdots, n$, then, writing $\overline{X_n} = (\sum_{i=1}^{i=n} X_i)/n$, $\overline{Y_n} = (\sum_{i=1}^{i=n} Y_i)/n$, and $\overline{Z_n} = \overline{X_n} - \overline{Y_n}$, it follows that $\overline{Z_n} \sim \mathrm{N}(\delta, 2\sigma^2/n)$. Thus δ is the mean improvement in efficacy achieved by the new treatment.

Also, writing $S_x^2 = \sum_{i=1}^{i=n}(X_i - \overline{X})^2$, $S_y^2 = \sum_{i=1}^{i=n}(Y_i - \overline{Y})^2$, and $S^2 = S_x^2 + S_y^2$, since $S_x^2/\sigma^2 \sim \chi_{n-1}^2$ and $S_y^2/\sigma^2 \sim \chi_{n-1}^2$ we have, by standard statistical theory, $S^2/\sigma^2 \sim \chi_{2n-2}^2$. Let $f(s_x^2, s_y^2, \overline{z_n}|\delta, \sigma^2)$ denote the likelihood function for S_x^2, S_y^2 and $\overline{Z_n}$, which is proportional to

$$(\sigma^2)^{-\frac{(2n-1)}{2}} \exp[-\frac{1}{2\sigma^2}\{s_x^2 + s_y^2 + \frac{n}{2}(\overline{z_n} - \delta)^2\}]. \tag{3.1}$$

Following O'Hagan and Forster (1994), the conjugate joint prior density function for δ and σ^2 has the form

$$\pi(\delta, \sigma^2) = k(a, g, \omega)(\sigma^2)^{-\frac{g+3}{2}} \exp^{-\frac{1}{2\sigma^2}\left\{a+(\delta-\mu)^2/\omega\right\}}, \tag{3.2}$$

where $k(a, g, \omega) = a^{\frac{g}{2}} 2^{-\frac{g+1}{2}} (\pi\omega)^{-\frac{1}{2}} \left\{\Gamma(g/2)\right\}^{-1}$, and a, g and ω are hyperparameters assigned on the basis of prior information. Note that $\pi(\delta, \sigma^2) = \pi(\sigma^2)\pi(\delta|\sigma^2)$, where the distribution of a/σ^2 is chi-squared with g degrees of freedom, and $\pi(\delta|\sigma^2)$ is $\mathrm{N}(\mu, \sigma^2\omega)$.

In terms of the Expectation and Variance of the prior distribution for σ^2 we have

$$g = 4 + 2\mathrm{E}(\sigma^2)^2/\mathrm{Var}(\sigma^2), \tag{3.3}$$

$$a = \mathrm{E}(\sigma^2)(g - 2), \text{ and} \tag{3.4}$$

$$\omega = \tau^2/\mathrm{E}(\sigma^2), \tag{3.5}$$

where τ^2 is the variance of the prior distribution for δ. Applying Bayes theorem and using (3.1) and (3.2), the posterior density for δ and σ^2 may be written,

$$\pi^{(n)}(\delta, \sigma^2 \mid x, y) = k(a', g', \omega')(\sigma^2)^{-\frac{g'+3}{2}} \exp^{-\frac{1}{2\sigma^2}\left\{a'+(\delta-\mu')^2/\omega'\right\}}. \tag{3.6}$$

Here

$$\omega' = \frac{2\omega}{2 + n\omega}, \qquad \mu' = \frac{2\mu + n\omega\overline{z_n}}{2 + n\omega},$$

$$g' = g + 2n - 1, \qquad a' = a + s^2 + \frac{n(\overline{z_n} - \mu)^2}{2 + n\omega}.$$

The mean and variance of the prior distribution for δ are μ and $\omega a/(g-2)$ $(=\tau^2)$, respectively.

The marginal posterior density function of δ can be obtained by integrating the joint posterior density (3.6) over σ^2 and we have

$$\pi^{(n)}(\delta \mid x, y) = \frac{1}{B(g'/2, 1/2)}(a'\omega')^{-\frac{1}{2}} \left\{1 + \frac{(\delta - \mu')^2}{a'\omega'}\right\}^{-\frac{g'+1}{2}},$$

where $1/B(g'/2, 1/2) = \frac{[(g'-1)/2]!}{\sqrt{g'\pi}[(g'-1/2)/2]!}$. Therefore, $(\delta - \mu')/\sqrt{\omega' a'/g'}$ has a t-distribution with g' degrees of freedom and the mean and variance of the posterior distribution for δ are μ' and $\tau'^2 = \omega' a'/(g'-2)$, respectively.

If the outcomes are binary responses (success or failure) for each patient, with success probabilities P_{X_i} and P_{Y_i} for $i = 1, 2, \cdots, n$, we may convert the outcome to a continuous scale by assuming $X_i = \log\frac{P_{X_i}}{1-P_{X_i}} \sim \mathrm{N}(\theta + \delta, \sigma^2)$ and $Y_i = \log\frac{P_{Y_i}}{1-P_{Y_i}} \sim \mathrm{N}(\theta, \sigma^2)$, for $i = 1, 2, \cdots, n$.

3.3.3 Distribution theory (adverse reactions)

There are costs associated with adverse reactions both within the clinical trial and in later use, and for users of both the current and the new drugs. Their frequency also influences the take-up of the new drug. We shall assume that these frequencies for both drugs are estimated from experience during phase II trials, and that patients are exchangeable across phases II and III as regards the incidence of adverse reactions.

Assume that adverse reactions occur at the unknown Poisson rate (incidence rate) ξ for each patient. We distinguish the new and the current drugs by the subscripts $i = 1$ and 2, respectively. Suppose that before the phase II trials ξ_i has a prior distribution which is $\Gamma(k_{i0}, x_{i0})$, and that during phase II there are r_{i0} adverse reactions over a total of t_{i0} patient-years. By Bayes theorem it follows that the posterior density for ξ_i after phase II is proportional to

$$\begin{aligned}
&\Gamma(k_{i0}, x_{i0}) \text{ density} \times P(\text{a Poisson}(\xi_i t_{i0}) \text{ random variable} = r_{i0}) \\
&= \frac{\xi_i^{k_{i0}-1} x_{i0}^{k_{i0}}}{(k_{i0}-1)!} \exp^{-\xi_i x_{i0}} \frac{(\xi_i t_{i0})^{r_{i0}}}{r_{i0}!} \exp^{-\xi_i t_{i0}} \\
&\propto \xi^{k_{i0}+r_{i0}-1} \exp^{-\xi(x_{i0}+t_{i0})}.
\end{aligned}$$

Thus the posterior distribution for ξ_i after phase II is $\Gamma(k_{i0} + r_{i0}, x_{i0} + t_{i0})$. This can be used as the prior distribution for ξ before phase III, and we write $k_i = k_{i0} + r_{i0}$ and $x_i = x_{i0} + t_{i0}$.

Let r_1 and r_2 denote the numbers of adverse reactions in phase III, for the new and current drugs, respectively, during a total of t patient-years. It follows in similar fashion that the posterior distribution of ξ_i $(i = 1, 2)$ is

$$\Gamma(k_i^{'}, x_i^{'}) \text{ where } k_i^{'} = k_i + r_i \text{ and } x_i^{'} = x_i + t. \tag{3.7}$$

Our posterior distributions for ξ_i remain within the gamma family because this is the family of conjugate prior distributions for the parameter of a Poisson process. Note that a $\Gamma(k, x)$ distribution has mean kx^{-1} and variance kx^{-2}.

We shall denote by α and β^2 the prior (before phase III) mean and variance for $\xi_1 - \xi_2$, the difference in the rates of occurrence of adverse reactions, and the posterior mean and variance by $\alpha^{'}$ and $\beta^{'2}$, respectively. Thus we have

$$\alpha = k_1/x_1 - k_2/x_2 \text{ and } \beta^2 = k_1/x_1^2 + k_2/x_2^2. \tag{3.8}$$

Before phase III the joint (continuous/discrete) density for ξ_i and r_i may be written as

$$f(\xi_i, r_i) = \frac{\xi_i^{k_i - 1} x_i^{k_i}}{(k_i - 1)!} \exp^{-\xi_i x_i} \frac{(\xi_i t)^{r_i}}{r_i!} \exp^{-\xi_i t}. \tag{3.9}$$

To obtain the marginal, or predictive, density for r_i we integrate this joint density over ξ_i giving

$$p(r_i) = \frac{(k_i + r_i - 1)!}{(k_i - 1)! r_i!} \frac{x_i^{k_i} t^{r_i}}{(x_i + t)^{k_i + r_i}}. \tag{3.10}$$

It follows that r_i has the negative binomial distribution

$$r_i \sim \text{Neg-bin}(k_i, t/(x_i + t)), \tag{3.11}$$

and therefore $\mathrm{E}(r_i) = k_i(1 - p_i)/p_i$ and $\mathrm{Var}(r_i) = k_i(1 - p_i)/p_i^2$, where $p_i = t/(x_i + t)$; r_i can be interpreted as the number of Bernoulli failures obtained before the k_ith success, where the probability of success is p_i. Thus following (3.7) we can write the posterior mean α' and variance β' for $\xi_1 - \xi_2$ as

$$\alpha^{'} = (r_1 + k_1)/(x_1 + t) - (r_2 + k_2)/(x_2 + t) \text{ and} \tag{3.12}$$

$$\beta^{'2} = (r_1 + k_1)/(x_1 + t)^2 + (r_2 + k_2)/(x_2 + t)^2, \text{ respectively.} \tag{3.13}$$

3.3.4 Expected net benefit function

The expected net benefit is calculated by subtracting the costs for the clinical trial and the expected incremental costs for subsequent users from the

expected total benefit. The costs for the clinical trial consist of administration costs and expenditure for the treatment of adverse reactions within the trial. The expected incremental costs include the change in the cost of adverse reactions when a patient switches to the new drug, which could be positive or negative depending on the particular case.

We define b to be a monetary benefit per unit improvement of efficacy, and the optimal sample size to be the sample size n for which the expected net benefit $w(n)$ from the clinical trial is maximised. The benefit $b\delta$ per patient is multiplied by m, the number of users who switch to the new treatment. The definition of b depends on the purpose of the trial, for example a reduction of direct and indirect societal losses. In this chapter, we define the benefit to be reduced expenditure on the treatment. Note that the financial saving per patient and per unit improvement in efficacy is aggregated over the period T for which the drug will be used.

The expected net benefit $w(n)$ for a sample size n includes items proportional to n, which are the administration cost of the trial and the incremental cost of adverse reactions within the trial, and items proportional to T, the number of patient-years of subsequent use, namely the benefit of improved efficacy and the incremental cost of adverse reactions, in both cases in subsequent use of the drug. If T is large, it may be necessary to use discounting. With a discount parameter of γ, a short interval Δ at time t becomes $\Delta \exp^{-\gamma t} + o(\Delta)$, and the total discounted time T' is $\int_0^T \exp^{-\gamma s}\, ds = (1 - \exp^{-\gamma T})/\gamma$.

We shall assume that m depends on the posterior distributions of the parameters δ, σ^2 and $\xi_1 - \xi_2$, which measure the difference between the new and the current drugs. For each subsequent patient switching to the new drug we assume an expected net benefit rate $\mathrm{E}[b\delta - d(\xi_1 - \xi_2)]$, where b is suitably chosen and d is the expected cost of the treatment for an adverse event. Thus the m-dependent contribution to $w(n)$ is the expectation of $mT'\{b\delta - d(\xi_1 - \xi_2)\}$ with respect to the joint distributions of δ, σ^2, ξ_1 and ξ_2, and of the relevant statistics, namely $\overline{z_n}$, s^2, r_1 and r_2. If they are assumed to be independent we consider separately the joint densities of δ, $\sigma^2, \overline{z_n}$ and s^2, and of ξ_1, ξ_2 , r_1 and r_2.

We have $f(\delta,\, \sigma^2, \overline{z_n}, s^2) = \pi(\delta,\, \sigma^2) f(\overline{z_n}|\delta,\, \sigma^2) f(s^2|\delta,\, \sigma^2)$, which we can also write as $f(\overline{z_n}, s^2)\pi^{(n)}(\delta,\, \sigma^2|\overline{z_n}, s^2)$, the product of a predictive (or marginal) density and a posterior density. For the parameters and statistics associated with adverse reactions we have the joint density $f(\xi_1, r_1, \xi_2, r_2) = f(\xi_1, r_1) f(\xi_2, r_2)$. As we did for the efficacy parameters and statistics we write $f(\xi_i, r_i)$ as the product of a predictive density and a posterior density.

Here we use (3.7) and (3.10) giving

$$\begin{aligned} f(\xi_i, r_i) &= \frac{x_i^{k_i} t^{r_i}}{r_i!(k_i'-1)!} \xi_i^{k_i'-1} \exp^{-\xi_i x_i'} \\ &= p(r_i) \xi_i^{k_i'-1} \frac{x^{k_i'}}{(k_i'-1)!} \exp^{-\xi_i x_i'} . \end{aligned} \tag{3.14}$$

It follows that the m-dependent contribution to $w(n)$ may be written as

$$\begin{aligned} \mathrm{E}[mT'(b\delta - d(\xi_1 - \xi_2))] &= T'\mathrm{E}\{\mathrm{E}[m(b\delta - d(\xi_1 - \xi_2))|\overline{z_n}, s^2, r_1, r_2]\} \\ &= T'\mathrm{E}\{m(b\mu' - d\alpha')\} \\ &= T' \int_{s^2} \int_{\overline{z_n}} \sum_{r_1} \sum_{r_2} m(b\mu' - d\alpha')\, p(r_1, r_2) \\ &\times\ f(\overline{z_n}, s^2)\, d\overline{z_n}\, ds^2, \end{aligned} \tag{3.15}$$

where $p(r_1, r_2) = p_1(r_1)p_2(r_2)$, the joint predictive distribution for r_1 and r_2.

The expected net benefit is calculated by subtracting the cost for the clinical trial. Thus

$$w(n) = T'\mathrm{E}\{m(b\mu' - d\alpha')\} - n(2c + d\alpha\eta). \tag{3.16}$$

Here c is the administration cost per patient, $d\alpha\eta$ is the expected incremental cost of the treatment of adverse reactions for each patient who takes the new drug in the trial, and η is the duration in years of the trial. In the second term the incremental cost for adverse reactions is needed if we reckon that without the trial all $2n$ patients would be on the conventional treatment. The n patients in the trial on the conventional treatment may be assumed to suffer the same adverse reactions as they would have suffered with no trial. Note that if $w(n) < 0$ for all $n > 0$ the trial is not worth carrying out.

3.3.5 Logistic weight function

The more effective and safe a new treatment, the more the number of patients who want to switch to it, and the higher the resulting benefit. Let M denote the total number of patients who receive the current treatment and who might switch to the new treatment. The number of subsequent users m is a fraction of M and our model assumes that m is calculated by using a logistic weight function as follows.

$$m = M/[1 + \exp\{L\}], \text{ where } L = u(v - \mu') + z\alpha' + 1.5\lambda', \lambda'^2 = u^2\tau'^2 + z^2\beta'^2, \tag{3.17}$$

and u, v and z are suitably chosen constants.

Our rationale for this expression is as follows. For known δ, and ξ_1 and ξ_2 known to be equal, $[1+\exp\{u(v-\delta)\}]^{-1}$ is the probability that a patient switches to the new treatment. We define v to be the value of δ for which half of all users switch; it is likely to represent a fairly small increment in the treatment effect, and to be positive because of the reluctance of users to try something new except for good reason. The value of u should be chosen so that $[1+\exp\{u(v-\delta)\}]^{-1}=0.75$ for the value of δ for which 75% of users would switch to the new treatment if δ was known and ξ_1 and ξ_2 were known to be equal. The value of $z\,(>0)$ should be selected so that $[1+\exp\{z(\xi_1-\xi_2)\}]^{-1}=0.75$ for the value (<0) of $\xi_1-\xi_2$ for which 75% of users would switch to the new treatment if δ was known to equal v and ξ_1 and ξ_2 were known.

Thus when δ, ξ_1 and ξ_2 are all known $L=u(v-\delta)+z(\xi_1-\xi_2)$. Note that $K=u\delta-z(\xi_1-\xi_2)$ is a measure of users' preference for the new treatment and λ'^2 is the posterior variance of K.

In practice δ, ξ_1 and ξ_2 are not known and are replaced by their posterior expectations in the expression (3.17) for L. The additional term $1.5\lambda'$ represents the fact that in order to switch a user is likely to require a statistically significant estimated increase in K. Other similar multiples of λ', such as 1.96, are equally plausible.

3.3.6 Monte Carlo simulation

The purpose of this Monte Carlo simulation is to generate a large number of sample values of $m(b\mu'-d\alpha')$, and then to calculate the sample mean as an estimate of $\mathrm{E}\{m(b\mu'-d\alpha')\}$ given n. Note that m, μ', α' and β' are all functions of $\overline{z_n}, s^2, r_1$ and r_2. Figure 3.1 summarises the mutual dependence of the key parameters and statistics and also indicates our algorithm for generating Monte Carlo random samples of $\overline{z_n}$, s^2, r_1 and r_2.

From (3.2), we have $\pi(\delta,\sigma^2)=\pi(\sigma^2)\pi(\delta|\sigma^2)$ and $\pi(\sigma^2)$ is an inverse chi-squared density function with parameters $g/2$ and $a/2$,

$$\pi(\sigma^2)=(\frac{a}{2})^{\frac{g}{2}}\frac{1}{\Gamma(g/2)}(\sigma^2)^{-\frac{g+2}{2}}\exp(-\frac{a}{2\sigma^2}) \tag{3.18}$$

and $\pi(\delta|\sigma^2)$ is $\mathrm{N}(\mu,\sigma^2\omega)$,

$$\pi(\delta|\sigma^2)=\frac{1}{\sigma\omega^{\frac{1}{2}}\sqrt{2\pi}}\exp^{-\frac{1}{2}\frac{(\delta-\mu)^2}{\sigma^2\omega}}. \tag{3.19}$$

Thus δ can be sampled from a normal distribution given σ^2. Note that $\overline{z_n}$ and s^2 are conditionally independent given σ^2 and δ, and that s^2 is conditionally independent of δ given σ^2.

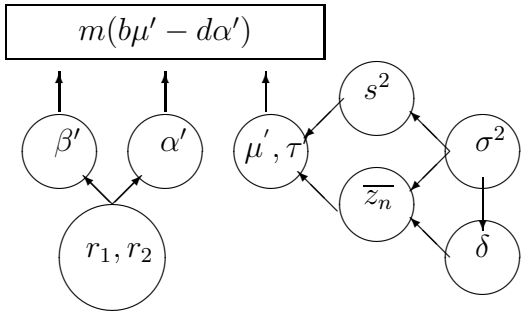

Figure 3.1: Monte Carlo sampling

To find the optimal sample size and the maximised expected net benefit we proceed as follows; n is the number of patients in each group and N is the iteration number of the simulation.

1. Set $n = 2$.
2. Draw samples $\sigma^2_{(j)}$ for $j = 1, 2 \cdots, N$ from the probability density function (3.18).
3. Draw a sequence of samples for $(\delta_{(1)}, \delta_{(2)} \cdots, \delta_{(N)})$ for δ from $\mathrm{N}(\mu, \sigma^2\omega)$, (see (3.19)), given the values $(\sigma^2_{(1)}, \sigma^2_{(2)} \cdots, \sigma^2_{(N)})$ for σ^2.
4. Sample Ncorresponding values $(s^2_{(n,1)}, s^2_{(n,2)}, \cdots, s^2_{(n,N)})$ of $s^2_{(n)}$ given these samples of σ^2, using the fact that $\frac{s^2_{(n)}}{\sigma^2}$ has a chi-squared distribution with $2n - 2$ degrees of freedom.
5. Sample values $(\overline{z_{(n,1)}}, \overline{z_{(n,2)}}, \cdots \overline{z_{(n,N)}})$ of $\overline{z_n}$ by sampling from $\mathrm{N}(\delta, 2\sigma^2/n)$, with values $(\delta_{(1)}, \delta_{(2)} \cdots, \delta_{(N)})$ for δ and $(\sigma^2_{(1)}, \sigma^2_{(2)} \cdots, \sigma^2_{(N)})$ for σ^2.
6. Draw samples $r_{1(j)}$ and $r_{2(j)}$ for $j = 1, 2 \cdots, N$ by using the probability density function (3.11) and calculate $\alpha'_{(j)}$ and $\beta'_{(j)}$ by (3.12) and (3.13), respectively.
7. Calculate N samples $h_{(n,j)}$, $j = 1, 2, \cdots, N$, of the total benefit, using the formula

$$h_{(n,j)} = m(\overline{z_{(n,j)}}, s^2_{(n,j)}, \alpha'_{(j)}, \beta'_{(j)})(b\mu'(\overline{z_{(n,j)}}) - d\alpha'_{(j)}),$$

   where m is defined by (3.17). The estimated expected net benefit (3.16)

is given by

$$w(n) \cong T' \frac{\sum_{j=1}^{N} h_{(n,j)}}{N} - n(2c + d\alpha\eta),$$

8. Set $n = n + 1$ and repeat steps 2 to 7.

9. Find the optimal sample size and the maximised expected net benefit from the estimated values for $w(n)$.

3.4 AN ILLUSTRATIVE EXAMPLE

3.4.1 Background

Cost-benefit analysis is not a standard procedure in clinical trials, though our view is that it should become a standard part of the phase III study. The aim of this example is solely to provide an illustration of how the calculations might proceed if the primary outcome is the evaluation of costs and benefits for a new drug. In our example we consider the situation that multi-centre phase II trials provide prior distributions and $w(n)$ is calculated for a phase III trial. The following example is based on published data. This, together with some further assumptions, gives us the necessary inputs. Note that the definition of benefit depends on the purpose of the study and our model can incorporate any kind of reasonable definition for benefit. The reduction of health-care costs by the new treatment is considered as the benefit in the following example.

Crohn's disease was discovered in 1932 by Crohn et al. and is a chronic inflammatory disease causing stomach pains, diarrhoea, and weight loss. The lower part of the small intestine (ileum) or the large intestine (colon) are typically affected and any part of the digestive system can also be affected. The disease is characterised by periods of activity and remissions and has no cure, though the symptoms can be controlled. A cytokine, tumor necrosis factor (TNF), is a key mediator of the inflammation associated with Crohn's disease, and an anti-TNF monoclonal antibody, infliximab, shows efficacy and tolerability in managing the symptoms of patients not responding to conventional treatments (Rutgeerts et al., 1999).

3.4.2 Application of the method

Targan et al. (1997) conducted a short term phase II study of infliximab with a small number of Crohn's disease patients who had scores on the Crohn's Disease Activity Index (CDAI) between 220 and 400. The CDAI scores can range from zero to 600. Higher scores indicate more severe illness and efficacy is measured by the change of the CDAI score. Patients were randomly assigned to a single dose of either placebo or infliximab at 5 mg/kg, 10 mg/kg or 20 mg/kg, and the sample sizes were 25, 27, 28 and 28, respectively in their study. Targan et al. defined the primary endpoint (a clinical response) as the percentage of patients with a reduction of 70 points or more in the CDAI scores at week four, as compared with the baseline scores. At four weeks, 81 percent of the patients given 5 mg/kg of infliximab, 50 percent of those given 10 mg/kg, and 64 percent of those given 20 mg/kg had a clinical response, as compared with 17 percent of patients in the placebo group. The difference in the clinical response between infliximab and the placebo group remained significant through the 12 weeks of follow-up. In the following calculations the CDAI scores for the group with 5 mg/kg of infliximab and the placebo group are converted into a utility between 0 and 1 and the difference of utility is evaluated on a continuous scale.

The means $\pm$ standard deviation of the CDAI scores at week 4 for the placebo group and the group with 5 mg/kg of infliximab are 271 ± 82 and 166 ± 76, respectively. On the basis of Targan's result we choose the hyperparameters for the prior distributions for δ and σ^2 so that $\mu = \mathrm{E}(\delta) = 271 - 166 = 105$ and $\mathrm{E}(\sigma^2) = (82^2 + 76^2)/2 = 6250$. The prior variances of δ and σ^2 we assume to be fairly large, with $\tau^2 = \mathrm{Var}(\delta) = 2\mu^2$ and $\mathrm{Var}(\sigma^2) = 2[\mathrm{E}(\sigma^2)]^2$. It follows that $a = 18750$, $g = 5$ and $\omega = 3.5$, using equations (3.3), (3.4) and (3.5). Since there is no specific information about the values of the parameters in the weight function, we assume that $u = 0.0366$ and $v = 40$. These figures imply that 75% of patients would choose infliximab if δ was known to be 70 and ξ_1 and ξ_2 were known to be equal.

Bassi et al. (2004) estimate that the cost is 20 pounds per medical consultation for primary care, which gives $d = 20$ in equation (3.16). Clark et al. (2001) estimate the number of patients with Crohn's disease to be 3.3×10^4 in England and Wales. According to a report by Abbot, over the course of their disease, as many as 75 percent of patients with Crohn's disease will undergo surgery at least once for complications or disease resistant to treatment (Press Release, 2007). Following these references we use the size of the target population $M = 3.3 \times 10^4 \times 0.75$ in the expected net benefit function.

Now we suppose that the next clinical trial (phase III) for infliximab examines the efficacy and safety at week 12 after a single dose of either 5 mg/kg of infliximab or placebo at week 0, and the evaluation of cost and benefit is the primary endpoint, and thus we set $\eta = 12/52$ in equation (3.16).

It is necessary for our calculations to convert the CDAI scores into a continuous scale for which each year in perfect health is assigned the value 1.0, and a year in imperfect health has a value between 0 and 1. Suppose that the health-care budget is 30,000 pounds for every lost unit of the scale for a patient with a chronic disease. For a patient with Crohn's disease and quality of life ϵ $(0 \leq \epsilon \leq 1)$, the annual health care cost is $(1-\epsilon)30,000$ pounds, and in general we may use $\epsilon = 1 - \text{CDAI}/600$. Thus, if there is a reduction of 1 in CDAI the annual health care cost is reduced by $b = 30,000/600 = 50$ pounds, assuming that CDAI, $(1-\epsilon)$ and the annual budget are all proportional to each other. Note that to evaluate $w(n)$ we need to multiply b by T', the expected total discounted time for infliximab usage. Suppose the annual discount rate γ and the time for use T are 3.5% and 10 years, respectively. This leads to $T' = 8.44$ years, which is used in this example.

The cost per patient of running Phase III clinical studies of new pharmaceutical products exceeds $ 17,167 (£12,350), on average, according to a recent benchmarking report (Cutting Edge Information, 2006). On this basis we have used alternative values of the cost c per patient in the clinical trial of 12,350, 18,525 and 24,700 pounds in our calculations.

As a guide to our treatment of adverse reactions we use articles by Targan et al. (1997) and Present et al. (1999). Targan et al. report that adverse reactions were experienced by 76 of 102 patients and 15 of 25 patients who received 5 mg/kg of infliximab and placebo, respectively, over 12 weeks. Present et al. report that adverse reactions were experienced by 10 and 12 out of 31 patients who received 5mg/kg of infliximab and placebo, respectively, over 18 weeks. Thus using ignorance prior distributions with $k_{i0} = x_{i0} = 0\,(i = 1, 2)$, and using these two sets of data to fill the role of phase II trials as described in section 3.3.3 of providing a prior distribution for a phase III trial, we have $k_1 = r_{10} = 10 + 76 = 86$, $x_1 = t_{10} = (31 \times 18 + 102 \times 12)/52 = 34.3$, $k_2 = r_{20} = 12 + 15 = 27$, $x_2 = t_{20} = (31 \times 18 + 25 \times 12)/52 = 16.5$. For phase III we have set $t = n \times 12/52$, assuming that the trial lasts for 12 weeks. It follows that $\alpha = 86/34.2 - 27/16.5 = 0.88$, $r_1 \sim \text{Neg-bin}(86, n/(148.3 + n))$ and $r_2 \sim \text{Neg-bin}(27, n/(71.5 + n))$. To estimate z we assume that 75% of potential users would switch to the new drug with δ known to equal $v = 40$, and $\xi_2 - \xi_1$ known to equal the fraction ζ of our prior estimate 27/16.5 for ξ_2. The value of z may then be calculated as described in section 3.3.5. For

example $z = 1.02$ when $\zeta = 2/3$, which means that adverse reactions occur with the new drug at approximately 1/3 of the rate at which they occur with the placebo.

3.4.3 Results

Figure 3.2 shows the result of calculations with $N = 1,000$ (see section 3.3.6). For each of the three curves $w(n)$ was evaluated for $n = 2, 3, 4, \cdots, 3000$. The curves were then smoothed using the S-Plus generic function, *spline*. The optimal sample sizes and the maximised expected net societal benefits are calculated to be 1,207, 1,144 and 1,091 patients and 1.20×10^9, 1.19×1.0^9 and 1.17×10^9, respectively, for the three alternative values of c. The power of a classical test to detect an improvement of 105 (the prior expectation of δ) in the CDAI scores is very close to 1.0 for these sample sizes, with a type I error of 0.001. This very high power and the insensitivity of the maximised expected net benefit to the cost c per patient included in the trial are both due to the fact that the total cost of the trial is much smaller than the expected net benefit. It is worth noting that these results are very different from what we would have found by classical frequentist methods for determining sample size. Note too that these very low probabilities of type I and type II error mean that the samples are almost certainly large enough for regulatory requirements. When this is not true the sponsor is likely to need to increase the size of the trial, as we have described in an earlier paper (Kikuchi et al., 2008). The value of ζ is a measure of users' sensitivity to the occurrence of adverse reactions, a low value indicating high sensitivity. Values of optimal sample size and maximised expected net benefit were calculated with $c = 18,525$ and different value of ζ, as shown in Table 3.1. The optimal sample size decreases as ζ increases, while the maximised expected net benefit is almost constant.

Table 3.1: Sensitivity to adverse reactions ($= 1/\zeta$)

ζ	z	Optimal sample size	Maximised expected net benefit (pounds)
0.90	0.75	1,096	1.19×10^9
0.67	1.02	1,144	1.19×10^9
0.50	1.36	1,185	1.19×10^9
0.33	1.83	1,247	1.19×10^9
0.25	2.71	1,319	1.18×10^9
0.12	5.42	1,560	1.15×10^9

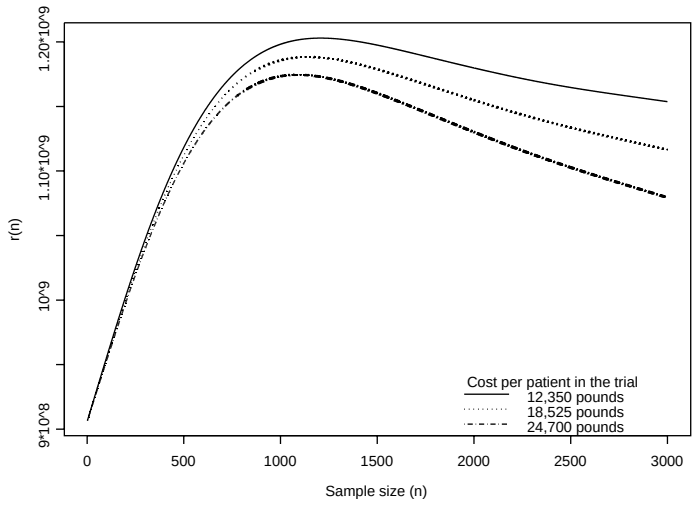

Figure 3.2: The expected net benefit against size of trial

To explore the sensitivity of our calculations to changes in the prior variance τ^2 for δ we now investigate the consequences of varying the prior hyperparameter ω, while leaving $\mathrm{E}(\sigma^2)$ unchanged. As we can see from equation (3.5), τ^2 is proportional to ω. Table 3.2 shows the optimal sample size n and the maximised expected net benefit $w(n)$ for different values of ω. Here $\zeta = 0.9$, so that $z = 0.75, c = 18,525$, and all other parameter values are the same as for the calculations illustrated in Figure 3.2.

Clearly the optimal sample size is a decreasing function of the prior variance and the optimised expected net benefit is an increasing function of the prior variance. Moreover the relationships are in both cases strong ones. The reason for this is that with vague prior knowledge about the efficacy of a new drug there is a real possibility that it is very much better than the existing treatment. If this is so it will not take a large sample to establish an improvement, and the resulting profitability will also be very large. It follows that it is important to take care in estimating the prior hyperparameters, particularly those relating to δ. It is advisable to carry out a sensitivity analysis with different prior distributions covering the range of plausible possibilities.

Table 3.2: Effect on n and $w(n)$ of increasing values of τ^2.

w	Optimal sample size	Maximised expected net benefit (pounds)
3.5	1,096	1.19×10^9
35	977	2.35×10^9
350	753	6.19×10^9
3,500	387	18.52×10^9

3.5 FURTHER APPLICATIONS

Cost-benefit analysis along the lines of this chapter is applicable to determination of the best medical treatment for a given condition, and to finding the optimal therapeutic dose for a new drug. It is also likely that governments will increasingly require pharmaceutical companies to evaluate the societal costs and benefits for subsequent users when a clinical trial has been completed. If the BeBay procedure is used, it is straightforward to evaluate the posterior expected net benefit conditional upon the results of the trial and hence to comply with this demand. The data of the completed clinical trial provide values for s^2, $\overline{z_n}$, α' and β'. The expected net benefit Q for subsequent users may be written

$$Q(n, \overline{z_n}, s^2, \alpha', \beta') = mT'(b\mu' - d\alpha').$$

If the national health authority can purchase the drug for substantially less than Q then it may well decide that it should do so, and the value of Q can be used as a criterion for the health authority in deciding a suitable price so that a positive net benefit can be assured.

The BeBay procedure may also be used to define a stopping rule for sequential clinical trials, as follows. Suppose that an interim analysis has been conducted with a sample size n_1. If the clinical trial is continued, then BeBay can be used to determine the optimal sample size n_2 starting from the point already reached. Let us define the expected value of sample information using BeBay (EVSIB) to be $w(n_2) - Q(n_1, \overline{z_{n_1}}, s_1^2, \alpha_1^{'}, \beta_1^{'})$, where the second term is the expected benefit calculated without a further clinical trial, which for brevity we shall write as Q_1. We propose the stopping rule that if either EVSIB or $w(n_2)$ is negative we stop the clinical trial, and otherwise we continue it. If $Q_1 > 0$ and EVSIB < 0 we can conclude that the new drug has positive net societal benefit and further sampling is unnecessary. If $Q_1 > 0$ and EVSIB > 0, the pharmaceutical company may be willing to continue the clinical trial to attain the higher expected net societal benefit, and possibly justify a higher price for the new drug. The other decisions are summarised

in Table 3.3, where n_2 is the optimal sample size for the next stage of the clinical trial.

Table 3.3: Decision table

Q_1	EVSIB	Decision
-	-	Stop. Reject the new drug.
+	-	Stop. Accept the new drug.
-	+	If $w(n_2) < 0$, stop and reject the new drug. Otherwise continue.
+	+	Continue.

3.6 FINAL COMMENTS

Current practice is for cost-effectiveness analysis to be carried out as a secondary analysis in clinical trials. It is our firm conviction that instead cost-effectiveness needs to be a primary endpoint if sample sizes are to achieve the appropriate balance between information gained and cost.

If national health expenditures are to be planned on the basis of their expected net benefit, there needs to be a careful accounting of the various costs and benefits. This need has been marked by the growth of the relatively new subject of pharmacoeconomics. BeBay is a contribution to the methodology of this field. Some of the data requirements will need to be supplied by imprecise estimates. However it is surely better to use imprecise estimates of the relevant quantities than to ignore them.

If clinical trials are carried out in more than one country, it is sometimes possible to incorporate the results of foreign studies in the evaluation of domestic studies. This issue is linked with the design of bridging studies, and is discussed in chapter 6 based on Kikuchi and Gittins (2010a).

Chapter 4

Sample size considering efficacy and safety: imbalanced sample size in treatment groups

4.1 ABSTRACT

The purpose of this chapter is to provide a rationale for experimental designs which allocate more patients to the new treatment than to the control group. The model uses a logistic weight function, including an interaction term linking efficacy and safety, which determines the number of patients choosing the new drug, and hence the resulting benefit. A Monte Carlo simulation is employed for the calculation. Having a larger group of patients on the new drug in general makes it easier to recruit patients to the trial and may also be ethically desirable. Our results show that this can be done with very little if any reduction in expected net benefit.

This chapter is based on Kikuchi and Gittins (2010b).

4.2 INTRODUCTION

The purpose of BeBay research is to provide a statistical methodology which may be used for phase III clinical trials and bridging studies. Phase III studies are randomized controlled multicentre trials on large patient groups and are aimed at assessment of the effectiveness of the new drug, in comparison

with the current treatment. They follow phase I and phase II trials which are designed to establish safe and effective dose levels. Bridging studies extrapolate clinical trial data arising in one region to a new region. A suitable bridging study reduces the need for additional clinical trials in the new region, thus reducing costs and making the new treatment available earlier.

The better the drug, the more patients want to use it, and the more the resulting benefit. This approach has been extended to take account of unpaired comparisons, unknown variance, safety as well as efficacy, together with a more flexible function relating efficacy and safety and the number of users (Kikuchi et al., 2008; Kikuchi and Gittins, 2009). These papers provided a methodology to calculate sample size with equal numbers in the active and control arms.

There are several proposals for sample size determination by Bayesian approaches. For example Brutti et al. (2008) say an experiment is successful if it yields a large posterior probability that an unknown parameter of interest is greater than a chosen threshold. In this context, a straightforward sample size criterion is to select the minimal number of observations so that the predictive probability of a successful trial is sufficiently large. However, their paper does not consider the optimisation of expected net benefit, an approach which may be described as decision analytic. The same is true of papers (Spiegelhalter and Freedman, 1986; Hutton and Owens, 1993; Joseph et al., 1997; Lee and Zelen, 2000).

There have been relatively few papers setting out a decision analytical framework for sample size determination for clinical trials. Three which do are those by Claxton and Posnett (1996b), Patel and Ankolekar (2007) and Willan (2008). Their models incorporate relevant economic factors in interesting ways. However, unlike BeBay, there is no explicit modelling of the number of subsequent users as a function of the efficacy and the frequency of adverse reactions revealed by the trial.

Frequently there is evidence from the Phase II trial that the new drug is superior to the current standard drug. This creates an incentive for choosing a sample size for the new drug in Phase III which is greater than for the control group taking the standard drug. In the first place the higher the chance of being given the new drug the more likely it is that a patient will wish to join the trial, and the less likely it becomes that enrolment rates are inadequate. There is also an ethical argument for allocating more patients to the drug which is likely to prove superior. However, currently imbalanced randomisation designs are rarely employed (see Dumville et al. (2006)), and there appears to be no published methodology on this subject which takes a

decision analytic approach.

Relevant papers taking the classical frequentist approach are those by Peto et al. (1976) and Peto (1978), who propose assignment in a 3:2 or 2:1 ratio for Phase II trials, and by Pocock (1979) who showed that the power of the hypothesis test in a Phase III trial of a given size is not very sensitive to the proportions of the total sample allocated to the new drug and to the control.

The purpose of the present chapter is to show how the BeBay approach may be used to explore the cost-benefit consequences of having imbalanced sample sizes in a Phase III clinical trial. Our illustrative example shows that very different sample sizes are compatible with relatively small reductions in the expected net benefit from a trial. This conclusion is in line with Pocock's results and is likely to apply more generally.

The structure of this chapter is as follows. First we describe the distribution theory needed for a Bayesian analysis of the evidence for efficacy and safety, and proceed to explain how BeBay calculates expected costs and benefits per patient. Then a weight function is introduced to calculate the number of subsequent users and the resulting benefit. An illustrative example shows how Monte Carlo simulation may be used to explore the dependence of expected net benefit on the sample sizes for the new treatment group and the control group. Finally we conclude with a short discussion.

Our illustrative example uses realistic values and supplemental assumptions due to a lack of individual patient data. It is not intended to be used directly for planning the clinical trials for the new statin for hypercholesterolemia which we discuss.

Throughout our discussion we assume that prior information on the parameters of probability distributions is expressible by means of conjugate prior distributions. Frequently this is a realistic assumption. The reasons for making it are that it simplifies what is in any case a complex presentation and ensures that the calculations which we need to make are computationally feasible (see for example Bernardo and Smith (1994)).

4.3 THE MODEL

4.3.1 Distribution theory(efficacy)

Let X and Y be a new drug and a standard drug or placebo, respectively, which is to be used in the trial. Suppose that the new and standard drug groups have sample sizes $n\,(>1)$ and $\gamma n\,(0<\gamma\leq 1, \gamma n>1)$, respectively. Let X_i $(i=1,2,\cdots,n)$ and Y_i $(i=1,2,\cdots,\gamma n)$ be the clinical outcomes on some continuous scale. The subscript i refers to patient i. Note that X_i and Y_i are unpaired and independent.

Let $X_i \sim \mathrm{N}(\theta+\delta,\sigma^2)$ and $Y_i \sim \mathrm{N}(\theta,\sigma^2)$. Writing $\overline{X_n}=(\sum_{i=1}^{i=n} X_i)/n$, $\overline{Y_{\gamma n}}=(\sum_{i=1}^{i=\gamma n} Y_i)/\gamma n$ and $\overline{Z_{n,\gamma}}=\overline{X_n}-\overline{Y_{\gamma n}}$, it follows that

$$\overline{Z_{n,\gamma}} \sim \mathrm{N}(\delta, \frac{\sigma^2}{n}(1+\frac{1}{\gamma})). \tag{4.1}$$

Thus δ is the mean incremental efficacy achieved by the new drug. To simplify our exposition σ^2 is assumed to be known. We assume that δ has the prior density $\mathrm{N}\,(\mu,\tau^2)$ where μ and τ^2 are based on previous clinical trials. Note that if n is sufficiently large then $\overline{Z_{n,\gamma}}$ follows a normal distribution for any distributions for X_i and Y_i by the Central Limit Theorem.

By Bayes theorem, the mean μ' and variance τ'^2 of the posterior distribution for δ are as follows.

$$\mu'(n,\gamma)=\frac{\tau^2\overline{Z_{n,\gamma}}+\frac{\sigma^2}{n}(1+\frac{1}{\gamma})\mu}{\tau^2+\frac{\sigma^2}{n}(1+\frac{1}{\gamma})}, \tag{4.2}$$

$$\tau'^2(n,\gamma)=\frac{\tau^2\frac{\sigma^2}{n}(1+\frac{1}{\gamma})}{\tau^2+\frac{\sigma^2}{n}(1+\frac{1}{\gamma})}. \tag{4.3}$$

4.3.2 Distribution theory (safety)

The incidence of adverse reactions can be modelled in a similar way. Let X_i^a and Y_i^a denote the numbers of adverse reactions for the ith patient in groups X and Y, respectively, and t be the treatment duration in the trial. Suppose X_i^a and Y_i^a have independent Poisson distributions with parameters λ_{Xi} and λ_{Yi}, and that these Poisson parameters are independently drawn from distributions with mean values $\theta^a+\delta^a$ and θ^a, respectively, both with variance ρ^2. Writing $\overline{X_n^a}=\sum X_i^a/n$ and $\overline{Y_{\gamma n}^a}=\sum Y_i^a/(\gamma n)$, by standard

theory it follows that $\overline{X_n^a}$ and $\overline{Y_{\gamma n}^a}$ have means $\theta^a + \delta^a$ and θ^a, and variances $(\theta^a + \delta^a + \rho^2)/n, (\theta^a + \rho^2)/(\gamma n)$. Unless γn is small, it follows from the central limit theorem that $\overline{X_n^a}$ and $\overline{Y_{\gamma n}^a}$ are approximately normally distributed. Thus writing $\overline{Z_{n,\gamma}^a} = \overline{X_n^a} - \overline{Y_{\gamma n}^a}$ it follows that

$$\overline{Z_{n,\gamma}^a} \sim \mathrm{N}\left(\delta^a, \frac{\theta^a + \delta^a + \rho^2}{n} + \frac{\theta^a + \rho^2}{\gamma n}\right).$$

The unbiased estimators

$$s_x^{a2}/(n-1) = \sum_i (X_i^a - \overline{X_n^a})^2/(n-1) \text{ and } s_y^{a2}/(\gamma n - 1) = \sum_i (Y_i^a - \overline{Y_{\gamma n}^a})^2/(\gamma n - 1) \tag{4.4}$$

are available for $(\theta^a + \delta^a + \rho^2)$ and $(\theta^a + \rho^2)$, respectively. We shall simplify our discussion by replacing these variances by their estimates. Thus if we write $\eta^{a2} = (\theta^a + \delta^a + \rho^2)/n + (\theta^a + \rho^2)/\gamma n$, then η^{a2} is regarded as known and equal to $\frac{s_x^{a2}}{n(n-1)} + \frac{s_y^{a2}}{\gamma n(\gamma n - 1)}$.

Suppose that δ^a has a $\mathrm{N}(\mu^a, \tau^{a2})$ prior distribution based on previous clinical trials. Using the above approximation and Bayes theorem for the normal distribution with known variance, it follows that the posterior distribution for δ^a after the clinical trial is normal with the mean and variance as follows.

$$\mu^{a\prime}(n, \gamma) = \frac{\tau^{a2}\overline{Z_{n,\gamma}^a} + \eta^{a2}\mu^a}{\tau^{a2} + \eta^{a2}} \quad \text{and} \tag{4.5}$$

$$\tau^{a\prime 2}(n, \gamma) = \frac{\tau^{a2}\eta^{a2}}{\tau^{a2} + \eta^{a2}}. \tag{4.6}$$

4.3.3 Expected net benefit function

The expected net benefit is calculated by subtracting the costs for the clinical trial from the expected total benefit. The costs for the clinical trial consist of administration costs and expenditure for the treatment of adverse reactions.

Let b be the monetary benefit per unit improvement of efficacy. The definition of b depends on the purpose of the trial. In this chapter we define the benefit to be reduced societal expenditure on the disease in the country. Let M denote the number of patients who receive the current treatment in the country and might switch to the new treatment by considering the results of the clinical trial. The total benefit from improved efficacy is calculated by multiplying the benefit per patient $b\delta$ by the number $m\,(0 \leq m \leq M)$ of patients switching to the new treatment.

The expected net benefit $w(n, \gamma)$ includes items proportional to n and γ, the main one being the administration cost of the trial. It also includes items proportional to T, the number of patient-years of subsequent use, which relates with the calculation of the benefit of improved efficacy and the incremental cost of adverse reactions in subsequent use of the drug. Note that the financial saving per patient and per unit improvement in efficacy is aggregated over the period T for which the drug will be used. If T is large, it may be necessary to use discounting. With a discount parameter of ξ, a short interval Δ at time t becomes $\Delta \exp^{-\xi t} + o(\Delta)$, and the total discounted time T' is $\int_0^T \exp^{-\xi s}\, ds = (1 - \exp^{-\xi T})/\xi$.

Note that the financial saving per patient depends on the time at which the drug becomes available, which in turn depends upon the size of the trial. Delays in availability attributable to the length of a bridging study can cause a serious social problem and are therefore an important factor in the calculation of the total benefit. These effects may be modelled by subtracting a suitable amount from T', depending on the particular case.

For each patient who switches to the new drug after the trial, there is a contribution $\mathrm{E}(b\delta - e\delta^a/t)$ to the expected net benefit rate per patient, where e is the expected cost of the treatment for an adverse event and t is the duration of the clinical trial. We must also estimate the change in prescription costs. Let k_1 and k_2 be the expected costs per patient per day, for the new and standard drugs, respectively. Note that the exact value of k_1 is unknown until the clinical trials have been completed and the health authority has approved the price.

The posterior expectation of this contribution after the trial is

$$T'\mathrm{E}[m(b\delta - e\delta^a/t))|\overline{z_{n,\gamma}}, \overline{z^a_{n,\gamma}}] = T'm(b\mu' - e\mu'^a/t),$$

where μ' and μ'^a are the posterior expectations of δ and δ^a.

We shall assume that m depends on the posterior distributions of δ and δ^a, the mean differences in efficacy and safety between the new and comparative drugs. Thus the m-dependent contribution to $w(n, \gamma)$ is the expectation of

$$T'm(b\delta - e\delta^a/t - 365(k_1 - k_2)). \tag{4.7}$$

To calculate $w(n, \gamma)$ we need to take the expectation with respect to the predictive distributions for $\overline{Z_{n,\gamma}}$ and $\overline{Z^a_{n,\gamma}}$ and to subtract the expected cost of the clinical trial and the incremental drug costs. Thus,

$$w(n, \gamma) = T'\mathrm{E}[m\{b\mu' - e\mu'^a/t - 365(k_1 - k_2)\}] - n\{(1 + \gamma)c + e\mu^a\}. \tag{4.8}$$

Here c is the administration cost per patient, $e\mu^a$ is the expected incremental cost of the treatment of adverse reactions for each patient in the clinical trial and receiving the new drug, and the duration of the trial is t. We assume that without the trial all $n(\gamma + 1)$ patients would be on the conventional treatment. The γn patients in the trial on the conventional treatment may be assumed to suffer the same adverse reactions as they would have suffered with no trial.

Note that as well as helping to decide on an appropriate sample size the calculations leading to equation (4.7) provide useful information when the trial has been completed. The health authority should negotiate the price which it pays for the new drug to make $T'm(b\mu' - e\mu'^a/t - 365(k_1 - k_2))$ positive. This will ensure that the new drug produces a positive net benefit compared with the current treatment.

4.3.4 Logistic weight function

The more effective and safe a new treatment, the more patients want to switch to it, and the higher the resulting total benefit. Let M denote the number of patients who receive the current treatment in the country and might switch to the new treatment by considering the results of the clinical trial. There may well be more than one current treatment. We can allow for this by accounting separately for the patients receiving each treatment, and by adding the corresponding contributions to the net benefit.

Note that the preferences of patients in the trial do not form part of the model. The preferences of subsequent patients when the drug is approved and available, on the other hand, are modelled to estimate the number of patients switching to the new drug, with the assumption that they vary from patient to patient.

Also note that the preferences of patients are strongly influenced by the advice which they receive from their medical adviser. There is also the possibility that health care authorities, such as NICE in the UK, may restrict prescription of a drug. When we refer to patients' preferences it should be understood that this refers to the choices made after taking account of these influences.

The number m of subsequent users who want to switch from the current treatment to the new treatment is a fraction of M and is calculated by using

a logistic weight function as follows.

$$m = M/[1 + \exp\{L\}], \text{ where} \tag{4.9}$$
$$L = u(v - \mu') + z\mu'^{a} + \psi\mu'\mu'^{a} + 1.5\lambda, \tag{4.10}$$
$$\lambda^2 = (u^2 + \psi^2\mu'^{a2})\tau'^2 + (z^2 + \psi^2\mu'^2)\tau'^{a2} + \psi^2\tau'^2\tau'^{a2}. \tag{4.11}$$

Here $\psi\mu'\mu'^{a}$ is an interaction term linking efficacy and safety, and u, v, z and ψ are suitably chosen constants.

If δ, δ^a were known, L could be replaced with $u(v - \delta) + z\delta^a + \psi\delta\delta^a$. Note that $K = u\delta - z\delta^a - \psi\delta\delta^a$ is a measure of patients' preference for the new treatment, $\mathrm{E}[K] = u\mu - z\mu^a - \psi\mu\mu^a$, and λ^2 is the posterior variance of K. In practice δ and δ^a are not known and are replaced by their posterior expectations in the expression (4.10) for L. The additional term 1.5λ represents the fact that in order to switch a user is likely to require a statistically significant estimated increase in K. Other similar multiples of λ, such as 1.96, are equally plausible.

Our rationale for this logistic weight function is as follows. If δ was known and δ^a was known to be zero, then $[1+\exp\{u(v-\delta)\}]^{-1}$ would be the probability p_s that a patient would switch to the new treatment. We define v to be the value of δ for which half of all users would switch to the new drug. The value of u should be chosen so that

$$[1 + \exp\{u(v - \delta)\}]^{-1} = 0.75, \tag{4.12}$$

for which the value of δ is such that 75% of users would switch to the new treatment if $\delta^a = 0$. The value of z should be selected so that $[1 + \exp\{z\delta^a\}]^{-1} = 0.75\,(z > 0)$ for which the value of δ^a is such that 75% of users would switch to the new treatment.

If we write $p_s = 1/[1 + \exp\{u(v - \delta) + z\delta^a + \psi\delta\delta^a\}]$ conditional on u, v and z as above, then for $\delta \neq 0$ and $\delta^a \neq 0$ we have

$$\psi = \frac{\log(1/p_s - 1) - u(v - \delta) - z\delta^a}{\delta\delta^a}. \tag{4.13}$$

The values of u, v, z and ψ need to be estimated on the basis of empirical data. Since our interest is in the preferences of patients after they have received medical advice, it would be best to gather this from a survey of the views of the patients and their medical practitioners by using previous trial results and simulated results, and then to estimate the relationship between the drug performance and preference by using standard statistical methods.

If the rate of adverse reactions of the new drug is similar to that of the comparative drug, it may be sufficient to calculate m in terms of efficacy only, leading to a simplified weight function with $L = u(v - \mu') + 1.5\lambda$ and $\lambda^2 = u^2\tau'^2$.

4.4 AN ILLUSTRATIVE EXAMPLE

4.4.1 Background

Our example is a real one but we have made plausible assumptions to fill gaps since some of the required data have not been published. Therefore, our example illustrates the statistical methodology and the calculation procedures rather than offering practical guidance for the case of rosuvastatin.

Cholesterol is transported into the cells of arteries in the form of low-density lipoprotein cholesterol (LDL-C). This can build up in the inner walls of the arteries that feed the heart and brain to form plaque, a thick, hard deposit that can narrow the arteries and make them less flexible resulting in coronary heart disease (CHD), leading to a heart attack or stroke. For this reason low-density lipoprotein cholesterol (LDL-C) is called "bad cholesterol".

4.4.2 Application of the method

Rosuvastatin is a new, synthetic, oral drug that reduces LDL-C and is intended for the treatment of patients suffering from hypercholesterolemia. Atorvastatin is a statin that was already known to be more efficacious than pravastatin, simvastatin, fluvastatin, or lovastatin in reducing LDL-C (Andrews et al., 2001). Schwartz et al. (2004) carried out a randomised controlled trial to compare efficacy and safety for rosuvastatin and atorvastatin in patients with hypercholesterolemia. The patients were randomly assigned to take 24 week-treatment in one of the following three groups, each with forced dose titration at 12 and 18 weeks; the starting and titrated doses for each group were rosuvastatin 5, 20 and 80 mg (group 1, 127 patients enrolled); rosuvastatin 10, 40 and 80 mg (group 2, 129 patients enrolled); and atorvastatin 10, 40 and 80 mg (group 3, 128 patients enrolled).

This study showed that LDL-C at 24 weeks was reduced significantly more with rosuvastatin (groups 1 and 2) than with atorvastatin. In this illustrative example, we assume that we wish to calculate the optimal sample size for a

further clinical trial (phase III) based on these results for combined groups 1 and 2 versus group 3.

The mean (standard error, SE) percent decreases from baseline in LDL-C at 24 weeks were 59.6 (1.0) and 52.0 (1.4) with rosuvastatin (combining groups 1 and 2, with data missing for one of 256 patients) and atorvastatin (group 3, with data missing for one of the 128 patients), respectively. The concentrations of LDL-C at baseline were 187 mg/dl and 188 mg/dl for rosuvastatin and atorvastatin, respectively, giving $\mu = 187 \times 0.596 - 188 \times 0.520 = 13.7$ mg/dl. To give a vague prior distribution we assume $\tau^2 = (\mu/2)^2 = 46.9$. The standard errors of 1.0 % and 1.4% are converted to 1.87 mg/dl and 2.63 mg/dl, so the usual unbiased esitmate for σ^2 is $(255 \times 254 \times 1.87^2 + 127 \times 126 \times 2.63^2)/(254 + 126) = 887.3$.

Based on the occurrence of drug-related adverse reactions caused by rosuvastatin (256 patients evaluated for safety) and atorvastatin (126 out of 127 patients evaluated due to the data missing) in this study, we calculate $\mu^a = 74/256 - 35/126 = 0.0113$ and a vague prior with $\tau^{a2} = (\mu^a/2)^2 = 3.18 \times 10^{-5}$ is used. Individual patient data for adverse reactions are not reported by Schwartz et al. (2004) and any other papers, so the calculations (4.4) for sufficient statistics are not feasible without a further assumption. In order to be able to proceed we shall assume that the adverse reactions which occurred all involved different patients. Thus our estimates for the variances $(\theta^a + \delta^a + \rho^2)$ and $(\theta^a + \rho^2)$ are $\{74(1 - 74/256)^2 + 182(74/256)^2\}/255 = 0.206$ and $\{35(1 - 35/125)^2 + 90(35/125)^2\}/124 = 0.203$, respectively.

Adult Treatment Panel (ATP) II and III guidelines by the National Cholesterol Education Program (NCEP) specified that the LDL-C concentration should be ≤ 100 mg/dl (Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults, 2002). Liu et al. (2002) estimated that the cost of illness associated with CHD is £7.06 billion per annum and the total patient population is 2.55 million ($= M$) in the UK, which means that £2,780 per patient was expended on average. To provide a first approximation to the benefit b per unit reduction in LDL-C concentration we assume that this cost is entirely attributable to patients with more than the target maximum of 100 mg/dl. We also assume that without treatment the mean concentration of LDL-C is 187.3 mg/dl, which is the mean of baseline concentration for patients entering the trial, and that the mean cost of treatment is proportional to the extent to which the concentration exceeds 100 mg/dl. This gives $b = 2780/87.3 = £31.8$ [mg/dl $\cdot$ patient^{-1}]. Thus our assumption means that a reduction of 1.0 mg/dl in the concentration of LDL-C for a patient reduces the required expenditure on medicare for hypercholesterolemia

by £31.8 on average.

Since no survey on patients' preferences for rosuvastatin and atorvastatin has been carried out, we need to set values for u, v, z and ψ on general grounds. If patients and their medical consultants are told that the new drug is more effective and equally safe compared with the current drug, it is likely that they will be reluctant to switch to the new drug, especially if the current drug is reasonably effective. This behavior shows a kind of *status quo bias*, a phenomenon discussed by two behavioural economists, Samuelson and Zeckhauser (1988). With this in mind we assume that 50% of patients would switch to the new drug if δ was known to be 15 mg/dl, and 75% would switch with $\delta = 30$ mg/dl. From our discussion in section 4.3.4 it follows that $v = 15$ and $u = 0.073$. We assume that $z = 32.4$ and $\psi = -0.001$.

The cost e for the treatment of an adverse event is estimated to be £12.74 by Liu et al. (2002). Schwartz et al. (2004) enrolled 1,233 patients with hypercholesterolemia in about 11 months, which means 3.7 patients per day on average. Assuming this rate of enrolment and a time horizon of 50 years for the new drug we have $T = 50 - n(1+\gamma)/(3.7 \times 365)$. This means that a longer recruitment period causes loss of the total benefit.

The cost per patient of running Phase III clinical studies of new pharmaceutical products is $ 26,000 (£16,242), on average, according to a new benchmarking report (Cutting Edge Information, 2006), so we set $c = 16,242$. According to Bazalo (2001), the mean and standard deviation of the cost per dose for statin treatment in the 10 largest health maintenance organizations (HMO) in the USA was $ 2.31 and $ 0.17, respectively, and the cost k_2 per dose of atrovastatin was $2.31, all in year 2000. It is reasonable to suppose that a new drug will be priced near the top of the current range, so we estimate k_1 by adding two standard deviations to the mean, giving an estimated value of $ 2.65. The incremental cost was thus $ 0.34 (£0.21) per dose. Note that this illustration assumes that M patients use atorvastatin and we calculate the benefit on the basis of those who switch to rosuvastatin.

4.4.3 Results

Values of the expected net benefit for different values of n and γ were calculated and are shown in Figure 4.1. Table 4.1 shows the values of n which produce the maximised expected net benefit (MENB) for fixed values of γ. There is a small difference in MENB and a gradual decrease as γ decreases, which of course means that the sample is becoming more imbalanced.

Table 4.1: Optimal n and maximised expected net benefits for fixed γ.

γ	n	$(1+\gamma)n$	MENB ($£ \times 10^{10}$)	T (years)
1.0	964	1,928.0	1.722	48.6
0.9	859	1,632.1	1.721	48.8
0.7	1,090	1,853.0	1.719	48.7
0.5	1,275	1,912.3	1.717	48.6
0.3	1,526	1,983.8	1.705	48.5
0.1	2,601	2,861.1	1.651	47.9

The largest calculated value of MENB is $£1.722 \times 10^{10}$ for T years (see Table 4.1) and with $\gamma = 1.0$. However, the variation with γ is not strong, with very little reduction of MENB when $\gamma = 0.3$. A research planner might very well regard this small loss in the calculated expected net benefit as being more than offset by an improved patient recruitment rate as a result of choosing γ less than 1.0.

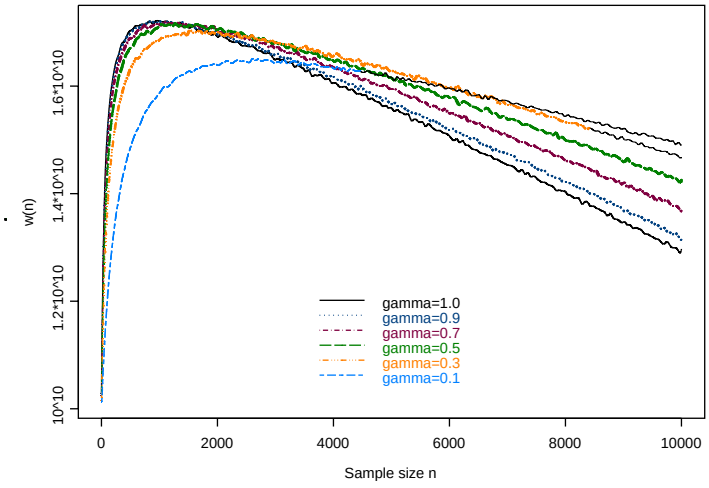

Figure 4.1: Expected net benefit for different n and γ.

It is interesting to compare these results with those which follow from the classical frequentist procedure for determining sample size. With equal numbers on each arm of the trial the standard formula for the sample size n which ensures that the probabilities of type I and type II errors are at most α and β respectively, for a clinically relevant difference in outcome of Δ is (see for example Dobson (1984)) $n = 2\sigma^2(z_{\alpha/2} + z_\beta)^2/\Delta^2$. This is based on a normal

approximation; σ^2 is the estimated variance, and z_ω denotes the $100(1-z_\omega)\%$ point of the standard normal distribution. Let us now follow convention by putting $\alpha = 0.05$ and $\beta = 0.2$, and set σ^2 equal to the unbiased estimate 887.3. Reasonable values for Δ are 15 or 30 mg/dl, which lead to required sample sizes of at least 14.08 or 35.2 per group, respectively. These sizes are far smaller than those calculated by BeBay. They also lead to very much lower values for the expected net benefit, namely 1.16×10^{10} and 1.31×10^{10}, which is only 67% and 76%, respectively, of the MENB for $\gamma = 1$ from Table 4.1.

4.5 CONCLUDING REMARKS

Our BeBay methodology has the distinct advantage as compared with current practice that it provides a means for taking patient preferences into account in modelling the likely usage of a new drug. There is an important need for empirical studies on what these preferences are.

In this chapter we have shown how the BeBay approach may be used in the determination of sample size for studies with a control group smaller than the group taking the new drug. It may also be used as a basis for calculating the maximum price that a health authority should be prepared to pay for a drug.

Our modelling of potential responses has been restricted to efficacy and safety and their interaction. In our example we have not modelled the relationship between recruitment rate and experimental design, nor have we modelled the ethical acceptability of a design. The quantification of the ethical acceptability is a topic which needs further extensive discussion. For the moment it seems best to deal with these issues separately from the formal model. Further modelling at some stage would be possible, given the necessary data on which to build it.

Chapter 5

Sample size determination for cluster randomised clinical trials

5.1 ABSTRACT

Cluster randomised clinical trials (CRCTs) are increasingly popular to evaluate disease control interventions for communities. In these trials health interventions are allocated randomly to complete clusters or groups rather than to individual subjects. Sample size calculation for CRCTs has been largely based on classical theory, taking account of between-cluster variation, and of type I and II errors. It is desirable to use an approach which maximises the expected net benefit, but there is as yet no established methodology along these lines. Gittins and Pezeshk (2000b) present an expected net benefit approach to sample size determination. In this chapter we extend that approach to CRCTs.

This chapter is based on Kikuchi and Gittins (2011).

5.2 INTRODUCTION

Randomised controlled trials are the gold standard for the evaluation of new medical interventions. The interventions are generally allocated randomly to individual subjects. In some trials, however, the medical interventions

are randomised not to individuals but to complete groups or clusters that might be clinics, families, villages, cities or other geographical areas. Trials like this are called "cluster randomised clinical trials (CRCTs)." They have recently been used in sociological and epidemiological studies such as the evaluation of nutrition interventions for infants in Nepal (Christian et al., 2003). Frequently the subjects in CRCTs are not patients and the purpose of the study is to evaluate the effect of preventive interventions.

The possible reasons for adopting CRCTs were cited by Smith and Morrow (1996) as follows: (1) the evaluation of interventions have to be implemented on a community basis because of the nature of the intervention, e.g. sanitation schemes, some educational, environmental interventions and group therapies; (2) logistical convenience, or to avoid contamination between interventions that might happen if they are allocated to individual subjects; (3) it is desired to capture the mass effect on the prevention of disease in a large proportion of the community members, e.g. vaccination; (4) although efficacy occurs at the individual level, it is measured on a community-basis, e.g. a decrease of mortality rate or of incidence rate.

Statistical methods for design and analysis of CRCTs are not as fully developed as those for individually randomised clinical trials (IRCTs). The majority of existing work on CRCTs depends on classical theory. There is the complication of the likely existence of between-cluster variability in addition to the usual between individual variability. This means that for a given sample size a CRCT may provide less information than an IRCT.

Sample size calculations using classical theory are based on the definition of a minimum clinical difference between two interventions which is to be detected with given type I and II errors. A number of papers discuss sample size calculations for CRCTs using classical theory, for example Donner et al. (1981), Shipley et al. (1989), Kocpsell et al. (1991), Shoukri and Martin (1992), Gail et al. (1992), Donner (1992) and Hayes and Bennett (1999). We will follow the terminology of Thompson et al. (1997) who derive a sample size formula for the matched-pair design that takes into account two components of variance: the within-cluster variance and the between-clusters-within strata variance. There are also articles describing Bayesian methods for analysis of binary outcomes for CRCTs, for example by Turner et al. (2001) and Thompson et al. (2004), and of continuous outcomes by Spiegelhalter (2001). None of these papers uses cost-benefit analysis for sample size determination.

Financial resources are inevitably restricted in public health. Thus decision makers have to trade off the anticipated benefits against the costs for introducing a particular implementation strategy. Therefore, decision makers and

funding bodies increasingly need information not only on the effectiveness of the intervention but also on its economic efficiency relative to other interventions. This means that economic evaluation needs to be standard practice alongside the evaluation of effectiveness for each intervention. However, there is not yet an established methodology for sample size determination and evaluation from a cost-benefit point of view in the literature on CRCTs. The closest approach, using a classical framework, is an article by Connelly (2003) that illustrates how researchers can minimise the cost of producing a given level of statistical power, other things being equal, or how power can be maximised given a specified budget.

The behavioural Bayes (BeBay) methodology was developed by Gittins and Pezeshk (2000a,b, 2002a,b). This approach was developed for individually randomised clinical trials and has been extended by Kikuchi et al. (2008) to take account of unpaired comparisons and unknown variance. Kikuchi and Gittins have further extended the approach: allowing the number of subsequent users to depend on both efficacy and safety (2009), to match the specific requirements of a bridging trial (2010a), and to allow different sample sizes for the new and control treatments (2010b).

The purpose of this chapter is to extend the BeBay approach to sample size determination for CRCTs. The structure of the chapter is as follows. First we describe the distribution theory needed for a Bayesian analysis of the efficacy in CRCTs and proceed to explain how BeBay calculates expected costs and benefits. Then a weight function is introduced to calculate the number of subsequent users and the resulting benefit. An illustrative example follows for a nurse-led lifestyle intervention in general practice to reduce cholesterol levels. The calculation uses Monte Carlo simulation to explore the dependence of expected net benefit on the sample sizes for the new intervention cluster and the control cluster. Graphical results are given in the last section of this chapter. We conclude with a short discussion.

Note that our purpose is to demonstrate a statistical methodology rather than to recommend a sample size in this case. Our illustrative example uses realistic values and supplemental assumptions due to a lack of individual patient data.

Throughout our discussion we assume that prior information on the parameters of probability distributions is expressible by means of conjugate prior distributions. Frequently this is a realistic assumption. The reasons for making it are that it simplifies what is in any case a complex presentation and ensures that the calculations which we need to make are computationally feasible (see for example Bernardo and Smith (1994)).

5.3 THE MODEL

5.3.1 Distribution theory

We here follow the notation used by Thompson et al. (1997). Let Y_{ijk} be the continuous outcome variable for subject $k = 1, \cdots, n_{ij}$ in stratum $j = 1, \cdots, N$, and for intervention i. Within each stratum there are separate clusters for the new intervention ($i = 1$) and control intervention ($i = 2$). To illustrate this notation suppose that N towns (strata) are randomly chosen from England, and two general practitioners (GPs) and corresponding clusters per town are selected. One of the two GPs in the town is assigned to the new intervention and the other GP to the control intervention. The number of subjects is n_{ij} for intervention i in town j.

In this imaginary example many patient factors may affect the results of the interventions and the patient factors are not identically distributed across England. We therefore need to include a town (stratum) effect. We also have to take account of a possible GP effect.

In this chapter we determine the optimal values of N and n_{ij} to maximise the expected net benefit of the clinical trial. Following Thompson et al. (1997), for a cluster randomised clinical trial we write

$$Y_{ijk} = \alpha_i + \gamma_j + \eta_{ij} + \epsilon_{ijk}, \tag{5.1}$$

where $\alpha_1 - \alpha_2 = \delta$ is the incremental effect of the new intervention, and γ_j is the stratum effect, for example a regional effect. For an experimental design with no strata the term γ_j is removed from equation (5.1) and the analysis is otherwise unchanged; this would be the situation, for example, if in our above illustration every GP selected for inclusion were to be from a different town. We assume that η_{ij} and ϵ_{ijk} are the random effects of cluster within strata and of the individual, respectively, and that these effects are normally distributed with $\mathrm{E}(\eta_{ij}) = 0$ and $\mathrm{E}(\epsilon_{ijk}) = 0$. With no strata η_{ij} is simply a random cluster effect - in our illustration a random GP effect, including any random variation between towns. We write $\mathrm{var}(\eta_{ij}) = \sigma_B^2$ and $\mathrm{var}(\epsilon_{ijk}) = \sigma_{ij}^2$.

Let $\overline{Z_j} = \overline{Y_{1j\cdot}} - \overline{Y_{2j\cdot}}$, where $\overline{Y_{1j\cdot}} = \sum_k Y_{1jk}/n_{1j}$ and $\overline{Y_{2j\cdot}} = \sum_k Y_{2jk}/n_{2j}$. Note that γ_j in equation (5.1) is canceled out for the calculation of $\overline{Z_j}$. We have $\mathrm{E}(\overline{Z_j}) = \delta$ and $\mathrm{var}(\overline{Z_j}) = 2\sigma_B^2 + (\sigma_{1j}^2/n_{1j} + \sigma_{2j}^2/n_{2j})$.

For our sample size calculation we suppose for simplicity that the between-individual variability is homogeneous across strata and clusters, so that σ_{ij}^2

can be represented by a single variance σ^2. We also assume that equal numbers n and nr of subjects are recruited in each cluster of subjects receiving intervention 1 and intervention 2, respectively. Thus a simple average over strata $\overline{Z} = \sum_{j=1}^{N} \overline{Z_j}/N$ has a normal distribution with the mean δ and the variance

$$\rho^2 = \frac{1}{N}\{2\sigma_B^2 + \frac{(1+1/r)\sigma^2}{n}\} \text{ for } N \geq 1 \text{ and } n \geq 1. \tag{5.2}$$

We assume a $\mathrm{N}(\mu, \tau^2)$ prior distribution for δ and that reasonable estimates are available for σ_B^2 and σ^2, so that ρ^2 is effectively known. Thompson et al. (1997) explain that whereas sufficient information is likely to provide a good estimate for σ^2, it is usually much more difficult to estimate σ_B^2 accurately. However, they managed to find rough estimates of σ_B^2 for a number of coronary risk factors by using a previous study by Shaper et al. (1981). From standard Bayesian theory it follows that the posterior distribution for δ is $\mathrm{N}(\mu', \tau'^2)$, where

$$\mu' = \frac{\tau^{-2}\mu + \rho^{-2}\overline{Z}}{\tau^{-2} + \rho^{-2}}, \tag{5.3}$$

$$\tau'^2 = (\tau^{-2} + \rho^{-2})^{-1}. \tag{5.4}$$

Notes:

1. The random effects may not in practice be normally distributed. However, our normal theory calculations remain approximately valid for reasonably large values of N, e.g. $N \geq 10$.

2. Our analysis could be carried out without assuming the variance ρ^2 to be known, along the lines described by Kikuchi et al. (2008). This would mean having a joint prior distribution for δ and ρ^2, leading to a posterior t-distribution for δ. This is often a more realistic model, although it does lead to more complex calculations.

3. If necessary, it is of course possible to use separate estimates σ_1^2 and σ_2^2 for σ_{1j}^2 and σ_{2j}^2, respectively, instead of the single value σ^2. In this case the equation (5.2) becomes

$$\rho^2 = \frac{1}{N}\{2\sigma_B^2 + \sigma_1^2/n + \sigma_2^2/(nr)\}.$$

   No other changes are required.

5.3.2 Expected net benefit function

The expected net benefit is calculated by subtracting the costs for the clinical trial from the expected total benefit. Let b be the monetary benefit per unit improvement of efficacy. The definition of b depends on the purpose of the trial. In this chapter we define the benefit to be reduced expenditure on treatment of the disease under investigation in the clinical trial. To calculate the total benefit, the benefit per person $b\delta$ is multiplied by the number m of subsequent users receiving the new intervention after the trial.

The expected net benefit $w(n, N, r)$ includes items proportional to n, N and r, the main one being the administration cost of the trial. It also includes the benefit of improved efficacy, which is proportional to T, the number of person-years of subsequent use of the new intervention.

The number m of subsequent users of the new intervention is assumed to depend on the posterior distribution of δ, the mean difference in effect between the new and comparative interventions. The expected net benefit rate is $\mathrm{E}(b\delta)$ per person who receives the new intervention and the m-dependent contribution to $w(n, N, r)$ is the expectation of $mTb\delta$. The posterior expectation of this contribution after the trial is

$$Tbm\mathrm{E}[\delta|\overline{z_n}] = Tbm\mu'. \tag{5.5}$$

To calculate $w(n, N, r)$ we need to take the expectation with respect to the predictive distribution for $\overline{Z}$ and to subtract the expected cost of the clinical trial. If the intervention and control costs are c_1 and c_2 per subject, respectively, we have

$$w(n, N, r) = Tb\mathrm{E}[m\mu'] - Nn(c_1 + rc_2), \tag{5.6}$$

where m is a function of μ' and τ'^2. The expectation in equation (5.6) is evaluated by Monte Carlo simulation, as no explicit compact formula is available.

5.3.3 Logistic weight function

The more effective a new intervention, the more GPs will want to use it for their patients, and the higher the resulting total benefit. Let M denote the total number of potential patients in the country. The number m of patients who will receive the new intervention as a result of their GPs' choice is a

fraction of M. It is calculated by using a logistic weight function as follows.

$$m = M/[1 + \exp\{L\}], \text{ where} \tag{5.7}$$

$$L = u(v - \mu') + 1.5\lambda', \tag{5.8}$$

$$\lambda'^2 = u^2\tau'^2, \tag{5.9}$$

and u and v are suitably chosen constants.

If δ was known, L could be replaced with $u(v - \delta)$. Note that $K = u\delta$ is a measure of the GPs' preference for the new intervention and $\mathrm{E}[K] = u\mu$, and that λ'^2 is the posterior variance of K. In practice δ is not known and is replaced by its posterior expectation μ' in the expression (5.8) for L. The additional term $1.5\lambda'$ represents the fact that in order to switch a user is likely to require a statistically significant estimated increase in K. Other similar multiples of λ', such as 1.96, are equally plausible.

Our rationale for this logistic weight function is as follows. For known δ, $[1 + \exp\{u(v - \delta)\}]^{-1}$ would be the probability that a GP switches from the current intervention to the new intervention. We define v to be the value of δ for which half of all GPs would choose the new intervention. As the result of their decision half of the potential patients in the country would take the intervention, assuming that half of GPs in the country are responsible for half of the potential patients. The value of u should be chosen so that

$$[1 + \exp\{u(v - \delta)\}]^{-1} = 0.75,$$

for the value of δ for which 75% of GPs would choose the new intervention and 75% of potential patients therefore receive it.

If we do not carry out a new clinical trial, the expected net benefit (ENB) Q is

$$Tmb\mu, \text{ where } L = u(v - \mu) + 1.5\lambda \text{ and } \lambda^2 = u^2\tau^2 \tag{5.10}$$

in (5.7). The values of μ and τ^2 are calculated from the currently available data. It is not worth carrying out a new trial unless $w(n, N, r) > \max[Q, 0]$.

5.4 AN ILLUSTRATIVE EXAMPLE

5.4.1 Background

Our example is a real one but we have made plausible assumptions to fill gaps since some of the required data are not available. For simplicity we have also

focused on just one endpoint, whereas several are in practice relevant. This means that our example illustrates the statistical methodology and the calculation procedures rather than offering practical guidance for the proposed study.

The British Family Heart Study (BFHS) (Wood et al., 1994a,b) was a cluster randomized clinical trial to assess the extent to which coronary risk factors, such as blood pressure, serum cholesterol levels and cigarette smoking, could be reduced by nurse-led lifestyle intervention in general practice over a period of one year. This illustrative example supposes that a new cluster randomised trial is planned for the same interventions for subjects recruited by the same inclusion criteria in England. The aims of the proposed new trial are to obtain more definite evidence of the benefits of nurse-led intervention, and if the evidence remains positive to encourage its regular use.

In this chapter we show how the BeBay methodology may be used to determine an appropriate sample size for the new trial. We shall make a number of simplifying assumptions. The main reason for this is that the more detailed information which would be needed for a full analysis is not available. Thus, the main outcome in the BFHS was the average difference between the levels of coronary risk factors at one year between the intervention group and the external control group. However, we shall use the difference in blood cholesterol concentration as the primary endpoint for the purpose of sample size determination.

In the BFHS study two GPs in each of 13 towns across the UK were chosen, and in each town one GP was randomly allocated to the intervention while the other served as the external comparison arm. Each pair of GPs was chosen on the basis of the similarity of the patient populations, taking account of a number of demographic and social characteristics, and their willingness to be randomized to either the intervention or comparison arm of the trial.

12,472 subjects (7460 men and 5012 women) aged 40-59 years and their partners were identified by household. The subjects were randomised to either the intervention arm or an internal comparison arm. Since the intervention was a nurse-led lifestyle intervention in general practice, the partners of the subjects were also invited to the educational sessions but the outcome was measured only for the 12,472 primary subjects. At baseline, families within the intervention arm were invited for screening and appropriate lifestyle advice and an overall coronary risk score (Tunstall-Pedoe, 1991) was calculated for each subject. Each subject and his or her partner were followed up individually over a year, with frequency of follow up dependent on their risk score at the initial screen. At one year the subjects in the intervention group were

invited for rescreening and the comparison group was screened for the first time. Thus there was no baseline measurement for the comparison group.

5.4.2 Application of the method

The mean difference in serum cholesterol (mmol/l) in the BFHS between intervention and comparison clusters was $0.127\,(=\mu)$. The 95% confidence interval is $0.127 \pm 1.96 \times 0.032$ and $\tau^2 = 0.032^2 = 0.001$. Thompson et al. (1997) estimate that $\sigma_B^2 = 0.015$ and $\sigma^2 = 1.10$. The value of T depends on the strength of the improvement and many other factors; we use $T = 15$ years in our calculations. Other values for μ will be investigated as a sensitivity analysis.

Hypercholesterolemia is a disease state with high levels of cholesterol in the blood. Cholesterol is transported into the cells of arteries in the form of low density lipoprotein-cholesterol (LDL-C). This can build up in the inner walls of the arteries that feed the heart and brain to form plaque, a thick, hard deposit that can narrow the arteries and make them less flexible, so establishing coronary heart disease (CHD), which can lead to a heart attack or stroke. For this reason low density lipoprotein-cholesterol (LDL-C) is called "bad cholesterol". Therefore, the reduction of LDL-C is beneficial for reducing risks for death and myocardial infarction among patients with established CHD as well as for decreasing the risk of establishing CHD for people with hypercholesterolemia. If LDL-C is higher than 140 mg/dl (3.61 mmol/l), the state is diagnosed as hypercholesterolemia. Pickin et al. (1999) estimated that 930,000 $(= M)$ potential patients are at high risk for CHD in England. This risk will be reduced for those people who manage to lower their level of LDL-C.

Pickin et al. (1999) do not give details of the distribution of LDL-C concentrations among these 930,000 potential patients. However Ebrahim et al. (1998) found that over a 15 year observation period almost half of the British men who suffered a CHD event had an LDL-C concentration of more than 6.25 mmol/l at the start of the observation period. We shall use this figure as an estimate for the mean LDL-C concentration level among the 930,000 potential patients.

If the LDL-C concentration is decreased below 2.58 mmol/l, the risk of a CHD event is negligible, see Wood et al. (1994b). Thus our estimate is that the primary target of the medical intervention is to reduce the mean concentration of LDL-C for the M potential patients at high risk from 6.25

mmol/l to 2.58 mmol/l, a reduction of 3.67 mmol/l.

To provide a first approximation to the benefit b per potential patient for CHD and per unit reduction in LDL-C we shall make the following further simplifying, and admittedly unrealistic, assumptions.

- All potential patients have the same probability of accepting the intervention, independent of their cholesterol level.
- The expected financial benefit of a reduction in LDL-C is independent of the pre-intervention cholesterol level and proportional to the extent of the reduction, up to the following limit.
- The annual cost of treating a potential patient whose LDL-C concentration is reduced by 3.67 mmol/l is zero.

The benefit in this example is the reduction in the cost of treatment as a result of preventing the potential patient from establishing CHD.

The prevalence of CHD in the UK in 1998 was 2.68 million patients, and the total cost was estimated to be £9.0 billion a year, of which around 35% was due to direct health care costs, 43% to productivity losses and 21% to the informal care of people with CHD (Petersen et al., 2004). Thus with the above assumptions our estimate for b, the expected financial benefit per person of a reduction of 1 mmol/l in LDL-C concentration, is $£9.0 \times 10^9/(2.68 \times 10^6 \times 3.67) = £915$.

Since no survey on GPs' preferences for the interventions is available, we need to set values for u and v on general grounds. The mean concentration of cholesterol for the control intervention in the BFHS was estimated to be 5.7 mmol/dl, which is the mean of baseline concentration for subjects with the new intervention. We assume that 50% of potential patients would follow the lifestyle advice on their GP's recommendation if $\delta = 5.70 - 3.36 = 2.34$ mmol/l, and 75% would do so if $\delta = 5.70 - 2.58 = 3.12$ mmol/l. From our discussion in section 5.3.3 it follows that $v = 2.34$ and $u = 1.41$. The mean intervention and control costs per subject were £63.14 and £37.30, respectively. The travel cost per subject-year was £42 in BFHS (Wonderling et al., 1996), and we assume that £10 per subject are spent for data analysis and reporting. Thus from these figures $c_1 = 63 + 42 + 10 = £115$ and $c_2 = 37 + 42 + 10 = £89$, and these values are used in our calculations.

It should here be noted that to find accurate estimates for M, b, u, v, c_1, c_2 and T requires further data and careful analysis. This issue is discussed in the following section. The values used in our illustrative example are less securely based and are intended simply to provide an illustration of the methodology.

We can now use equation (5.6) to calculate the expected net benefit for different values of n, N and r. Calculations were carried out for a range of values of n for $N = 5, 10, 20$ and 40, $r = 1.0, 0.75, 0.5$ and 0.25, and for $\mu = 0.127, 0.5$ and 1.5. Tables 5.1 and 5.2 show values of MENB, by which we mean ENB maximised with respect to r and n, for the chosen values of μ and N. For comparison, the ENB if no further trial is carried out, see equation (5.10), is $Q = £6.43 \times 10^7, £41.67 \times 10^7$ and $£425.96 \times 10^7$ for $\mu = 0.127, 0.50$ and 1.50, respectively. If MENB $< Q$ for all n, no further trial is warranted.

5.4.3 Results

Figures 5.1, 5.2 and 5.3 show the behaviour of the expected net benefit function. Each curve is based on Monte Carlo integration. We generated 100,000 random samples of $m\mu'$ to calculate $\mathrm{E}[m\mu']$ for each value of n ($n = 1, 2, 3, \cdots, 5000$). A smooth curve was then plotted using the S-plus generic function *spline*. To investigate the convergence of $w(n)$ at $n = 0$, we calculated $w(0.0001, N, r)$. For every curve this produced a value close to the corresponding Q, as expected. The calculation of each of the 36 curves took approximately two hours on a computer with mobile AMD Sempron 3000+, 1.79 GHz, 896 Mb of RAM. In practice it would normally be sufficient to calculate ENB for 1,000 samples of $m\mu'$ rather than for 100,000, which would take about 2 minutes of computer time per curve. For the case of unknown variance the time taken is likely to rise by a factor of up to about 5.

We can summarise our findings as follows.

- (MENB - Q) is clearly positive for all three values of the prior expected efficacy μ, suggesting, within the limits of our model, that a further trial is worthwhile.
- Both MENB and the optimal trial size, not surprisingly, increase strongly with μ.
- For all three values of μ, MENB is greatest for $N = 40$ and is insensitive to the value of r. Given the general difficulty of trial recruitment when the chance of being allocated to the control arm is high, we might tentatively conclude that $r = 1/3$ would be a reasonable choice, giving each patient recruited a 75% chance of being allocated to the active arm. It is clear that for $r = 1/3$ MENB is very close to its maximum value.

Table 5.1: MENB and the optimal r and n for $\mu = 0.127$

μ	N	r	n	$Nn(r+1)$	MENB ($£ \times 10^7$)	MENB - Q ($£ \times 10^7$)
0.127	5	1.00	93	930	6.44	0.01
$Q = 6.43$		0.75	121	1,059	6.44	0.01
		0.50	104	780	6.44	0.01
		0.25	149	931	6.44	0.01
	10	1.00	96	1,920	6.45	0.02
		0.75	83	1,453	6.45	0.02
		0.50	118	1,770	6.45	0.02
		0.25	116	1,450	6.44	0.01
	20	1.00	76	3,040	6.47	0.04
		0.75	80	2,800	6.47	0.04
		0.50	94	2,820	6.46	0.03
		0.25	103	2,575	6.45	0.02
	40	1.00	54	4,320	6.49	0.06
		0.75	65	4,550	6.49	0.06
		0.50	70	4,200	6.48	0.05
		0.25	87	4,350	6.46	0.03

5.5 FURTHER COMMENTS

5.5.1 Desirable refinements in parameter estimation

To calculate sample size on the basis of the expected net benefit as described in this chapter requires more investigation of epidemiology, preferences, costs and expected therapeutic effect, and more detailed computation, than a conventional calculation does. To proceed in this way, however, does focus on what really matters, so our strong recommendation is that, even though there will inevitably be some shortcomings in the accuracy and completeness of the available data, an analysis along these lines should be carried out. In this section we briefly review some of the ways in which it would be desirable to improve on the data sources used in the case-study discussed in the previous section.

Instead of a single figure for M, the number of potential patients at high risk for CHD, it would be preferable to use the distribution of cholesterol concentration over those M potential patients, and to model the expected benefit per patient as a function of initial concentration. For example it would be useful to survey the distribution of cholesterol concentrations after

Table 5.2: MENB and the optimal r and n for $\mu = 0.5, 1.5$

μ	N	r	n	$Nn(r+1)$	MENB ($£ \times 10^7$)	MENB - Q ($£ \times 10^7$)
0.50	5	1.00	329	3,290	41.82	0.15
$Q = 41.67$		0.75	334	2,923	41.82	0.15
		0.50	397	2,978	41.82	0.15
		0.25	545	3,406	41.80	0.13
	10	1.00	295	5,900	41.94	0.27
		0.75	325	5,688	41.93	0.26
		0.50	442	6,630	41.93	0.26
		0.25	552	6,900	41.90	0.23
	20	1.00	266	10,640	42.11	0.44
		0.75	233	8,155	42.11	0.44
		0.50	302	9,060	42.09	0.42
		0.25	414	10,350	42.05	0.38
	40	1.00	193	15,440	42.34	0.67
		0.75	176	12,320	42.34	0.67
		0.50	264	15,840	42.31	0.64
		0.25	333	16,650	42.24	0.57
1.50	5	1.00	965	9,650	427.50	1.54
$Q = 425.96$		0.75	981	8,584	427.51	1.55
		0.50	1,570	11,775	427.48	1.52
		0.25	1,845	11,531	427.44	1.48
	10	1.00	1,003	20,060	428.73	2.77
		0.75	1,033	18,078	428.74	2.75
		0.50	1,501	22,515	428.71	2.75
		0.25	2,259	28,238	428.61	2.65
	20	1.00	813	32,520	430.60	4.64
		0.75	1,004	35,140	430.59	4.63
		0.50	1,218	36,540	430.54	4.58
		0.25	1,478	36,950	430.40	4.44
	40	1.00	628	50,240	433.02	7.06
		0.75	617	43,190	433.03	7.07
		0.50	820	49,200	432.93	6.97
		0.25	1,209	60,450	432.71	6.75

one year for patients with different initial concentration levels and to combine these data with information on the relationship between cholesterol level and the frequency of cardiac events. The expected benefit per patient may well be different for different patients. This calculation of ENB still assumes that the achieved reduction in cholesterol concentration for each patient is proportional to the initial excess concentration over 2.58 mmol/l and the required annual expenditure per person is proportional to the excess over 2.58 mmol/l.

The parameters u and v, which indicate the distribution of GPs' responses to different levels of improved efficacy compared with the current intervention, should be based on data from a survey of GPs' attitudes.

The procedure for calculating the costs c_1 and c_2 per patient admitted to the trial in our illustrative example seems broadly satisfactory. There should however also be a setup cost for the trial as a whole. This would influence the decision as to whether or not to carry out a trial.

The 15 year estimate for T, the time for which the new intervention will be in force if it is adopted, if possible should be replaced by a probability distribution based on historical data for interventions in the same therapeutic area.

5.5.2 Final Remarks

In this chapter we show how the BeBay approach can be used in the determination of sample size for cluster randomised clinical trials. It also (Kikuchi and Gittins, 2010b) allows the possibility of recruiting a smaller number of subjects to the comparative intervention than to the new intervention. Trials which are unbalanced in this way are often preferable because they tend to have higher recruitment rates.

Note that as well as helping to decide on an appropriate sample size the calculations leading to equation (5.5) provide useful information when the trial actually has been completed. The health authority should seek to ensure that the price which it pays for the new intervention is less than Tmb. If the price is very high and/or ENB is very low the health authority should not implement the intervention.

Our analysis has been for a new drug entering a market with an established therapy. A more complete analysis would take account of the evolution of patients' preferences among several existing drugs as evidence for their rela-

tive effectiveness accumulates, and take into account the tendency to prefer more established drugs even if they actually perform less well.

5.6 GRAPHICAL RESULTS

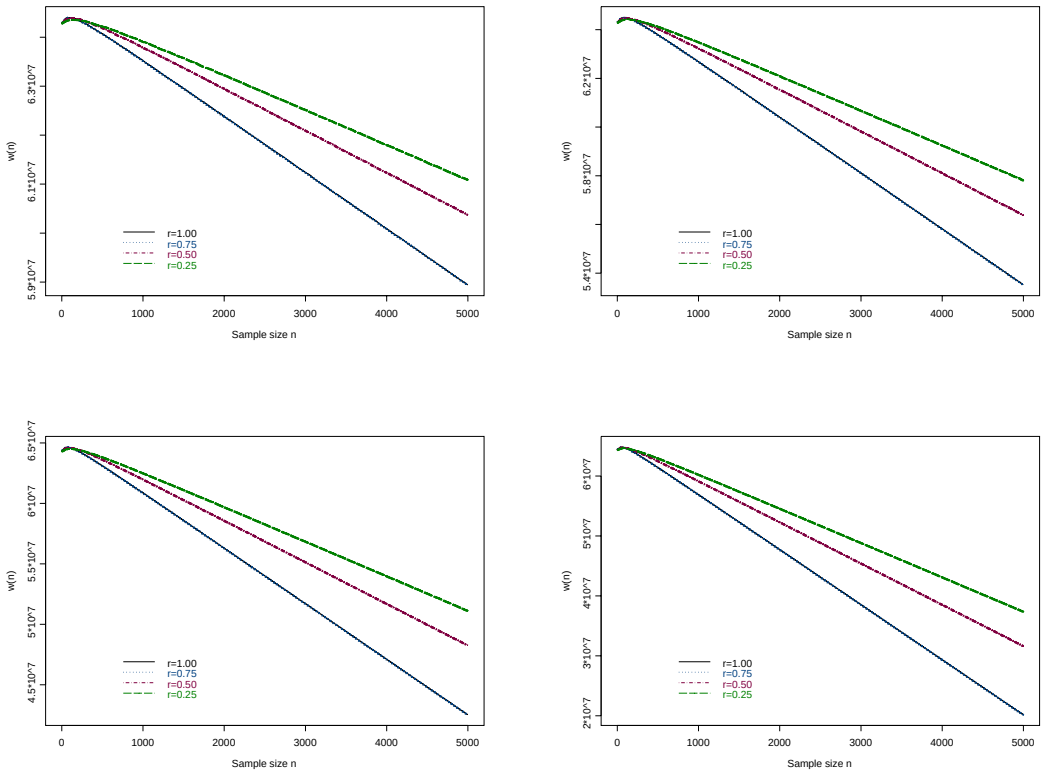

Figure 5.1: Expected net benefit functions ($\mu = 0.127, n \geq 1$) are shown for $N = 5, 10, 20$ and, 40 in the upper left, upper right, lower left and lower right figures, respectively.

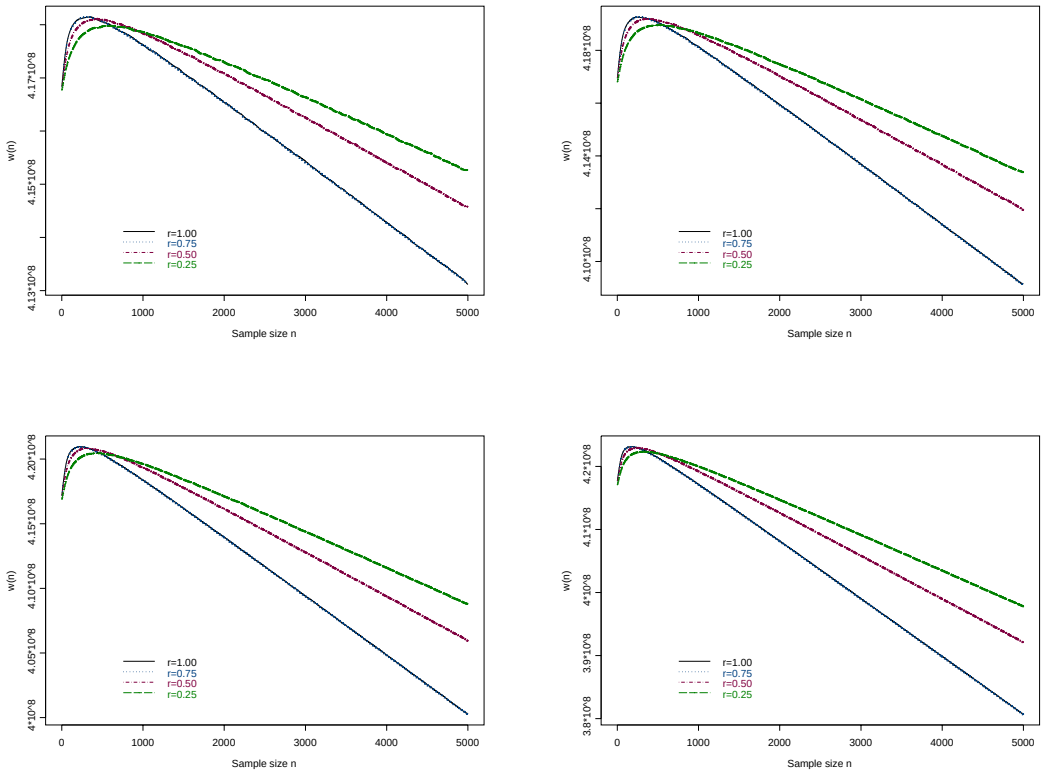

Figure 5.2: Expected net benefit functions ($\mu = 0.5, n \geq 1$) are shown for $N = 5, 10, 20$ and, 40 in the upper left, upper right, lower left and lower right figures, respectively.

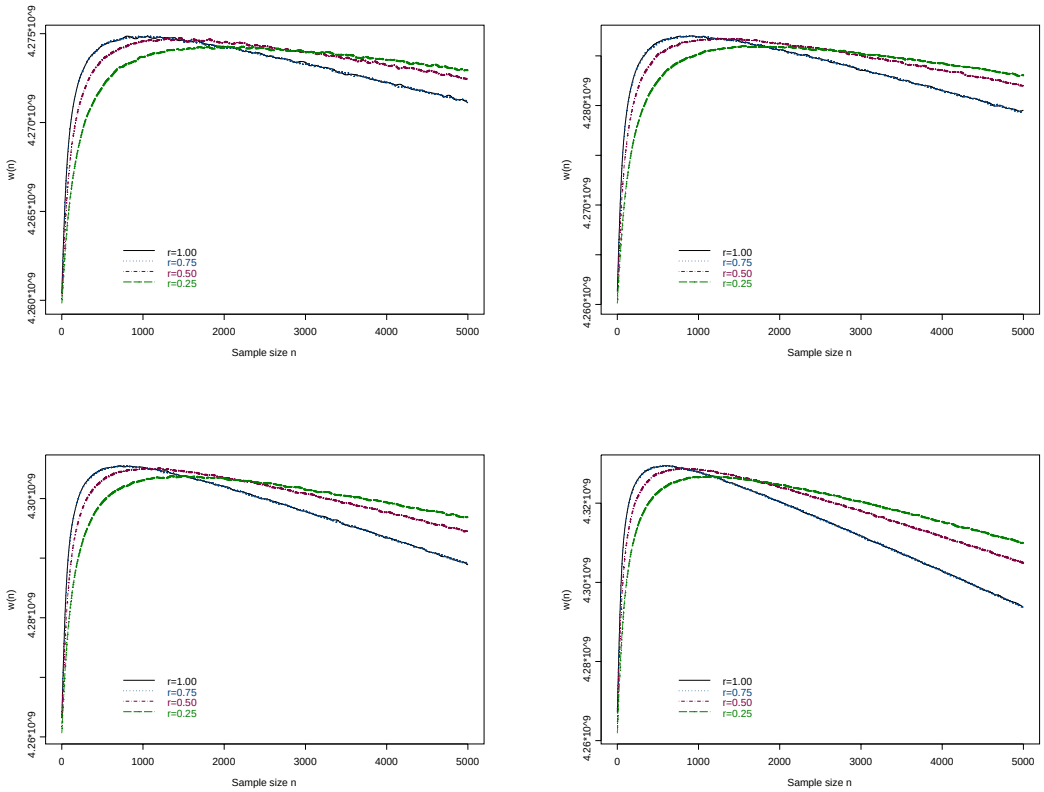

Figure 5.3: Expected net benefit functions ($\mu = 1.5, n \geq 1$) are shown for $N = 5, 10, 20,$ and, 40 in the upper left, upper right, lower left and lower right figures, respectively.

Chapter 6

A Bayesian adaptive design for the evaluation of a new drug in a bridging study

6.1 ABSTRACT

When an already established drug is considered for use in a new country the typical procedure is to carry out a relatively small trial, known as a bridging trial, in the new country, as a supplement to the trials which have already been carried out, and to make a case for introduction of the drug to the new country on the basis of both sources of information. There is no generally accepted methodology for allowing for heterogeneity between these two sources. This chapter proposes a mixed behavioural and Bayesian approach, based on maximising the expected societal net benefit in the new country. We show how this approach may be used in the design of a multi-stage bridging study.

This chapter is based on Kikuchi and Gittins (2010a).

6.2 INTRODUCTION

Throughout the world there are many national and regional health authorities, and until recently it was frequently necessary for new drugs to go through separate clinical trials for each jurisdiction before being approved for use.

This practice caused delay in new drug developments and increased drug prices resulting in the loss of benefit for patients. Patients were frustrated at not being able for many years to use new drugs that had already been licensed in other countries. To deal with this problem, the international conference on harmonisation of technical requirements for registration of pharmaceuticals for human use (ICH) was organised in 1990 and the standardisation of criteria for clinical trials is progressing among Japan, the US and the EU countries. The objective of the ICH is to explore the more economical use of human, animal and material resources, and to eliminate unnecessary delays in global development (see http://www.ich.org/).

The ICH guideline E5, Ethnic Factors in Acceptability of Foreign Clinical Data (ICH steering committee, 1998), released a guidance for pharmaceutical companies in dealing with ethnic factors influencing the safety and efficacy of medicines in different populations. According to the ICH E5, foreign clinical data on safety and efficacy may be accepted in the new country by extrapolation. To extrapolate foreign data to new countries bridging studies are necessary. Typically these are expected to be smaller studies which validate and build on the foreign data. However, the ICH E5 merely provides general policies and does not indicate any specific methodology or criteria showing how foreign data should be extrapolated. Furthermore, there is no written criterion on the assessment of similarity of clinical results between countries. If a new drug is investigated in different clinical trials, there are many potential causes of heterogeneous results, for example differences of the study design, protocol, medical practices and race. One of the problems in the ICH E5 is that decision making on the similarity of efficacy and safety between countries currently relies on the regulator's subjective opinions and experiences. It is likely that this situation will cause inconsistent decisions on new drug registration. We therefore need to develop statistical models to evaluate the results of bridging studies whilst taking account of heterogeneity between countries.

In this chapter our approach is a development of sample size determination set out by Gittins and Pezeshk (2000b), which they describe as behavioural Bayes, or BeBay. It is behavioural because it takes into account the relationship between evidence regarding efficacy and the number of patients who will wish to use the new drug. Detailed modelling of this relationship is a distinctive feature of BeBay, a simplified version of which has been adopted by Willan and Eckermann (2010). It is Bayesian (or decision theoretic) because it assumes that decisions on sample size and on whether a new drug should be officially recommended are taken so as to maximise the expected societal net benefit. The resulting sample sizes are sometimes very different

from those derived from standard classical methods.

The present chapter has two purposes. Firstly it extends BeBay to the design of a bridging study. For ethical and cost effectiveness reasons this is formulated as a sequential process. Secondly by focusing on societal costs and benefits it also shows how this approach may be used by a national licensing authority as a guide in deciding how much it should be prepared to pay for a new drug.

In this chapter we shall assume that we are dealing with a chronic condition for which different drug treatment regimes are to be compared. The influence of a drug in terms of efficacy and of the incidence of adverse events is assumed to continue only while the drug is being administered.

The construction of the chapter is as follows. The second section discusses a Bayesian approach to incorporating differences in the impact of a new drug on efficacy and safety between a foreign clinical study and a domestic bridging study. The third and fourth sections introduce a methodology for the calculation of the expected net benefit for patients in the home country and a sequential procedure for deciding whether a licence should be given to the new drug. In the fifth section an example of the calculation procedure is described using published articles on a bridging study and some supplementary assumptions. Note that this chapter is designed to demonstrate a statistical methodology. We are not able to offer specific guidance for the drug discussed in our example because the necessary data was not all available.

6.3 INCORPORATION OF HETEROGENEITY BETWEEN TRIALS

6.3.1 Efficacy

Efficacy is assumed to be measured on a uni-variate scale.

In this chapter we focus on two clinical trials, a phase III clinical trial in one country followed by a bridging study in a second country. Let A and B denote the original and new countries, respectively. Suppose that the trials in countries A and B are carried out under essentially the same protocol, including the following conditions.

1. The comparisons of drug efficacy between a new drug X and a comparative drug Y are unpaired. The mean and variance of the efficacy

of the new drug are unknown.

2. The sample sizes for the X and Y groups are the same, and are l and n in countries A and B, respectively.

Let X_{Ai}, Y_{Ai}, X_{Bi} and Y_{Bi} denote the clinical outcomes for the ith patient in the groups and countries indicated. We assume all these quantities to be statistically independent, and that $X_{Ai} \sim \mathrm{N}(\theta_A + \delta, \sigma^2)$, $Y_{Ai} \sim \mathrm{N}(\theta_A, \sigma^2)$, $i = 1, 2, \cdots, l$, and that $X_{Bi} \sim \mathrm{N}(\theta_B + \Delta, \sigma^2)$, $Y_{Bi} \sim \mathrm{N}(\theta_B, \sigma^2)$, $i = 1, 2, \cdots, n$. This assumes that the value of σ^2 is identical in countries A and B. Often this is a reasonable assumption, see for example Sakai et al. (2002), who consider a Japanese bridging study. It is a straightforward matter to develop a model in which this assumption is removed.

It follows that for country A the mean difference in observed efficacy between the two drugs $\overline{Z_A} = \overline{X_A} - \overline{Y_A}$ has a $\mathrm{N}(\delta, 2\sigma^2/l)$ distribution. Following standard statistical theory, S_{Ax}^2/σ^2 and S_{Ay}^2/σ^2, where $S_{Ax}^2 = \sum_{i=1}^{i=l}(X_{Ai} - \overline{X_A})^2$ and $S_{Ay}^2 = \sum_{i=1}^{i=l}(Y_{Ai} - \overline{Y_A})^2$, have chi-squared distributions with $l - 1$ degrees of freedom, and thus $S_A^2/\sigma^2 = S_{Ax}^2/\sigma^2 + S_{Ay}^2/\sigma^2 \sim \chi_{2l-2}^2$. The likelihood function $f(s_A^2, \overline{z_A}|\delta, \sigma^2)$ is proportional to

$$(\sigma^2)^{\frac{-(2l-1)}{2}} \exp[-\frac{1}{2\sigma^2}\{s_A^2 + l(\overline{z_A} - \delta)^2/2\}]. \tag{6.1}$$

Following O'Hagan and Forster (1994), the conjugate joint prior density function for δ and σ^2 has the form

$$f(\delta, \sigma^2) = k(a, g, \omega)(\sigma^2)^{-\frac{g+3}{2}} \exp^{-\frac{1}{2\sigma^2}\left\{a + (\delta-\mu)^2/\omega\right\}}, \tag{6.2}$$

where μ is the mean of the prior distribution for δ,

$$k(a, g, \omega) = a^{\frac{g}{2}} 2^{-\frac{g+1}{2}} (\pi\omega)^{-\frac{1}{2}} \left\{\Gamma(g/2)\right\}^{-1},$$

and a, g and ω are hyper-parameters. As usual the prior distribution expresses our prior knowledge about the unknown parameters, in this case δ and σ^2. This includes the information from phase II clinical trials in country A. Note that $f(\delta, \sigma^2) = f(\delta|\sigma^2)f(\sigma^2)$, and the prior distributions for σ^2 and δ are defined to be such that

$$\frac{a}{\sigma^2} \sim \chi_g^2 \quad \text{and}$$

$$\delta|\sigma^2 \sim \mathrm{N}(\mu, \sigma^2\omega).$$

In terms of the mean and variance of the prior distribution for σ^2 we have

$$g = 4 + 2\mathrm{E}(\sigma^2)^2/\mathrm{var}(\sigma^2), \tag{6.3}$$

$$a = \mathrm{E}(\sigma^2)(g-2), \quad \text{and} \tag{6.4}$$

$$\omega = \tau^2/\mathrm{E}(\sigma^2), \tag{6.5}$$

where τ^2 is the variance of the prior distribution for δ.

Applying Bayes theorem and using equations (6.1) and (6.2), the posterior density for δ and σ^2 after the country A trial may be written,

$$f^{(n)}(\delta, \sigma^2 \mid x, y) = k(a', g', \omega')(\sigma^2)^{-\frac{g'+3}{2}} \exp^{-\frac{1}{2\sigma^2}\left\{a' + (\delta-\mu')^2/\omega'\right\}}. \tag{6.6}$$

Here

$$\omega' = \frac{2\omega}{2+l\omega}, \qquad \mu' = \frac{2\mu + l\omega\overline{z_A}}{2+l\omega}, \tag{6.7}$$

$$g' = g + 2l - 1, \qquad a' = a + s_A^2 + \frac{l(\overline{z_A} - \mu)^2}{2+l\omega}. \tag{6.8}$$

The marginal posterior density function for δ can be obtained by integrating the joint posterior density (6.6) over σ^2, which shows that $(\delta - \mu')/\sqrt{\omega' a'/g'}$ has a t-distribution with g' degrees of freedom. The mean and variance of the posterior distribution for δ are μ' and $\omega' a'/(g'-2)$ $(= \tau'^2)$, respectively.

Integrating the joint posterior density (6.6) over δ, we see that the marginal posterior distribution for σ^2 is

$$\frac{(a'/2)^{g'/2}}{\Gamma(g'/2)}(\sigma^2)^{-(g'+2)/2} \exp\left(\frac{-a'}{2\sigma^2}\right), \tag{6.9}$$

which is inverse chi-squared with g' degrees of freedom and mean value $a'/(g'-2)$.

For a bridging study in country B we approximate the joint posterior distribution for δ and σ^2 after phase III trials in country A by assuming that σ^2 is known and equal to $a'/(g'-2)$ and that $\delta \sim \mathrm{N}(\mu', \tau'^2)$. This is a good approximation for $g' \geq 10$, which is true in most clinical trials. Calculation without making this approximation is feasible, though rather more complex. Thus for the bridging study we write $\overline{Z_B} = \overline{X_B} - \overline{Y_B}$, the mean difference in observed efficacy, and assume that $\overline{Z_B} \sim \mathrm{N}(\Delta, 2\sigma^2/n)$, where σ^2 is effectively known.

Let us now write $\xi = \Delta - \delta$, the mean difference between countries A and B in the efficacy improvement caused by the new drug. Our information

about the possible range of differences in efficacy between countries A and B is expressed by the prior distribution for ξ, which is the main measure of heterogeneity between regions. It is convenient and plausible to assume a normal distribution. In many cases if the new drug has not been tested in country B there is no a priori reason to expect either that the efficacy will be generally higher in region B than in region A, or that it will be generally lower. This suggests a zero prior mean; alternative assumptions will sometimes be needed.

With a prior mean of zero the prior variance of ξ is the main measure of the prior expectation of heterogeneity between countries A and B. A low variance means that not much heterogeneity is expected, which in turn means that a relatively small bridging study will be needed. It is important to make the choice carefully, taking account of all the relevant evidence. For example in many cases phase I trials reveal no significant differences between countries in the pharmacokinetics of absorption, distribution, metabolism and excretion (ADME) of the new drug. This would suggest a fairly low prior variance (see for example Yuh et al. (1994)). A thorough analysis of previous cases is needed to establish guidelines for translating the evidence to date for a new drug into an appropriate prior distribution. This has yet to be done in a sufficiently methodical way; the work of Pinto et al. (2005) is an example of what is needed.

In this chapter we assume that ξ has a prior variance of $(\mu'/2)^2$. This means that all values of ξ between $\pm\mu'$ are plausible, which in turn is likely to mean quite a wide and uninformative prior distribution, and therefore no great saving in the required size of the bridging study as compared with what would have been necessary with no evidence from country A.

These prior distributions for δ and ξ combine to give a $\mathrm{N}(\mu', \tau''^2)$ prior distribution for Δ, the average efficacy improvement in country B, where

$$\tau''^2 = \tau'^2 + (\mu'/2)^2. \tag{6.10}$$

By the Bayesian theory for the normal distribution with known variance, after a bridging study on population B the posterior distribution for Δ is normal and the mean μ^* and variance τ^{*2} are as follows.

$$\mu^* = \frac{2\sigma^2\mu' + n\tau''^2\overline{z_B}}{2\sigma^2 + n\tau''^2} \quad \text{and} \quad \tau^{*2} = \frac{2\sigma^2\tau''^2}{2\sigma^2 + n\tau''^2}. \tag{6.11}$$

6.3.2 Safety

As we now proceed to show, we can model the incidence of adverse events in a similar way to our treatment of efficacy. Let $X^a_{Ai}, Y^a_{Ai}, X^a_{Bi}$ and Y^a_{Bi} denote the numbers of adverse events during the trial for the ith patient in the groups and countries indicated. We assume that these have independent Poisson distributions with parameters λ_{XAi}, λ_{YAi}, λ_{XBi} and λ_{YBi}, and that these Poisson parameters are independently drawn from distributions with mean values $\theta^a_A+\delta^a, \theta^a_A, \theta^a_B+\Delta^a$ and θ^a_B, all with variance ρ^2. Note that the variance of a Poisson distribution is equal to its mean. Writing $\overline{X^a_A} = \sum X^a_{Ai}/l, \overline{Y^a_A} = \sum Y^a_{Ai}/l, \overline{X^a_B} = \sum X^a_{Bi}/n, \overline{Y^a_B} = \sum Y^a_{Bi}/n$, it follows from standard theory that $\overline{X^a_A}, \overline{Y^a_A}, \overline{X^a_B}$ and $\overline{Y^a_B}$ have means $\theta^a_A+\delta^a, \theta^a_A, \theta^a_B+\Delta^a$ and θ^a_B, and variances $(\theta^a_A+\delta^a+\rho^2)/l, (\theta^a_A+\rho^2)/l, (\theta^a_B+\Delta^a+\rho^2)/n$ and $(\theta^a_B+\rho^2)/n$, respectively. Provided the observed number of adverse events for each of the groups is not less than about eight, it also follows from the central limit theorem that these four quantities $\overline{X^a_A}, \overline{Y^a_A}, \overline{X^a_B}$ and $\overline{Y^a_B}$ are approximately normally distributed. Thus writing $\overline{Z^a_A} = \overline{X^a_A} - \overline{Y^a_A}$ and $\overline{Z^a_B} = \overline{X^a_B} - \overline{Y^a_B}$ it follows that $\overline{Z^a_A} \sim \mathrm{N}(\delta^a, \{2(\theta^a_A+\rho^2)+\delta^a\}/l)$ and $\overline{Z^a_B} \sim \mathrm{N}(\Delta^a, \{2(\theta^a_B+\rho^2)+\Delta^a\}/n)$.

After phase III trials in country A the unbiased estimators

$$s^{a2}_x/(l-1) = \sum (X^a_{Ai} - \overline{X^a_A})^2/(l-1) \text{ and } s^{a2}_y/(l-1) = \sum (Y^a_{Ai} - \overline{Y^a_A})^2/(l-1) \tag{6.12}$$

are available for the variances $(\theta^a_A+\delta^a+\rho^2)$ and $(\theta^a_A+\rho^2)$, respectively. We shall simplify our discussion by replacing these variances by their estimates, and treating these as known values. Thus we write $2\sigma^{a2} = 2(\theta^a_A+\rho^2)+\delta^a$, and regard $2\sigma^{a2}$ as known and equal to $(s^{a2}_x + s^{a2}_y)/(l-1)$, and also as a good approximation to the unknown quantity $2(\theta^a_B+\rho^2)+\Delta^a$, so that $2\sigma^{a2}/n$ is our approximation to the variance of $\overline{Z^a_B}$ before the bridging study in country B.

Before the phase III trials in country A suppose that δ^a has a $\mathrm{N}(\mu^a, \tau^{a2})$ prior distribution based on phase II data from country A. Using the above approximation and standard Bayesian theory for the normal distribution with known variance, it follows that the posterior distribution for δ^a after the country A trial is normal with mean and variance

$$\mu^{a\prime} = \frac{2\sigma^{a2}\mu^a + l\tau^{a2}\overline{z^a_A}}{2\sigma^{a2} + l\tau^{a2}} \quad \text{and} \quad \tau^{a\prime 2} = \frac{2\sigma^{a2}\tau^{a2}}{2\sigma^{a2} + l\tau^{a2}}. \tag{6.13}$$

This is now the prior distribution for δ^a for the bridging study in country B. Recall that $\overline{Z^a_B} \sim \mathrm{N}(\Delta^a, 2\sigma^{a2}/n)$, where σ^{a2} is assumed to be known.

Proceeding as in the case of efficacy, we define $\xi^a = \Delta^a - \delta^a$, and the prior distribution for ξ^a is assumed to be

$$\mathrm{N}(0, (\mu^{a\prime}/2)^2). \tag{6.14}$$

Thus the prior distribution for Δ^a is $\mathrm{N}(\mu^{a\prime}, \tau^{a\prime\prime 2})$, where $\tau^{a\prime\prime 2} = \tau^{a\prime 2} + (\mu^{a\prime}/2)^2$ and, proceeding as in the case of efficacy, the mean μ^{a*} and variance τ^{a*2} of the posterior distribution for Δ^a after the bridging study are as in equation (6.11), with the addition of the superscript a. Thus

$$\mu^{a*} = \frac{2\sigma^{a2}\mu^{a\prime} + n\tau^{a\prime\prime 2}\overline{z_B^a}}{2\sigma^{a2} + n\tau^{a\prime\prime 2}} \quad \text{and} \quad \tau^{a*2} = \frac{2\sigma^{a2}\tau^{a\prime\prime 2}}{2\sigma^{a2} + n\tau^{a\prime\prime 2}}. \tag{6.15}$$

6.4 THE EXPECTED NET BENEFIT OF THE NEW DRUG IN COUNTRY B

6.4.1 The calculation of the expected net benefit after the bridging study

The expected net benefit for the patients in country B is calculated in general by subtracting the costs for the clinical trial and the expected incremental costs for subsequent users in country B from the expected total benefit. This calculation is based on the data concerning efficacy and safety in the phase III trial in country A and the bridging study in country B. The costs for the clinical trial consist of administration costs and expenditure for the treatment of adverse events within the trial in country B. The expected incremental costs include the differences between the costs of the new and current drugs and between the costs of adverse events with each of the two drugs, which could be positive or negative.

To avoid further complexity all costs and benefits are evaluated from the standpoint of the health authority in country B. Thus in this chapter no value is attached to the well being of patients per se. This comes into the calculation only indirectly via the impact on the health authority's costs. In practice well being would usually also be included explicitly, by estimating and valuing quality adjusted life-years, and by assigning a well-being cost to adverse events.

We define b to be the annual monetary benefit per patient per unit improvement of efficacy, and Q to be the expected net benefit conditional on the

results of the bridging study for the patients in country B if the new drug is made available generally. The benefit per patient is multiplied by m, the number of regular users who switch to the new treatment in country B after the new drug is licenced by the health-authority. The definition of b depends on the purpose of the study, for example a reduction of direct and indirect societal losses caused by the disease. In this chapter, we define the benefit to be reduced societal expenditure on the new treatment. Note that this is a measure which gives licensing authorities a basis on which to decide whether or not to grant a licence, and if a licence is granted how much the health-care authorities should be willing to pay for the new drug.

The financial saving per patient and per unit improvement in efficacy is aggregated over the estimated period T for which the drug will be used. Note that the financial saving per patient depends on the time at which the drug becomes available, which in turn depends upon the size of the trial, and for multi-stage trials also upon the number of stages. Delays in availability attributable to the length of a bridging study can cause a serious social problem and are therefore an important factor in the calculation of the total benefit. These effects may be modelled by adjustments to T.

We shall assume that m depends on the posterior means and variances of the parameters Δ and Δ^a, the differences of efficacy and safety between the new and comparative drugs. Let k_1 and k_2 be the expected costs per patient per day, for the new and standard drugs, respectively. Normally these costs would include indirect costs, such as the costs of working days lost or productivity loss. The default assumption, in the absence of relevant data, is that the indirect costs are the same for both drugs, and therefore cancel each other in the calculation of $k_1 - k_2$. The exact value of k_1 is of course unknown until the clinical trials have been completed and the health authority has approved the price. For each patient switching to the new drug after the bridging study we assume an expected annual net benefit rate

$$\mathrm{E}(b\Delta - e\Delta^a/t_0 - 365(k_1 - k_2)).$$

Here b is suitably chosen, e is the expected cost of the treatment for an adverse event, and t_0 is the duration of treatment for each patient in the phase III and bridging study trials, which for simplicity we assume to be the same. Thus, making explicit some of the quantities on which Q depends, we have

$$\begin{aligned} Q(\overline{z_B}, \overline{z_B^a}, n) &= T\mathrm{E}[m\{b\Delta - e\Delta^a/t_0 - 365(k_1 - k_2)\}|\overline{z_B}, \overline{z_B^a}] \\ &= Tm\{b\mu^* - e\mu^{a*}/t_0 - 365(k_1 - k_2)\}. \end{aligned}$$

To allow for the time taken to enroll $2n$ patients in the bridging study we replace T by $T - 2n/(365r)$, where r is the number of patients per day enrolling for the bridging study. Thus, we have

$$Q(\overline{z_B}, \overline{z_B^a}, n) = \{T - 2n/(365r)\}m\{b\mu^* - e\mu^{a*}/t_0 - 365(k_1 - k_2)\}. \quad (6.16)$$

This means that if the enrolment rate r is very low it significantly reduces the maximised expected net benefit for patients switching from their current drug to the new drug.

As well as helping to decide on an appropriate sample size the calculations leading to equation (6.16) provide useful information when the trial has been completed. The health authority should negotiate the price k_1 which it pays for the new drug to make (6.16) positive. This will ensure that the new drug produces a positive net benefit compared with the current treatment.

6.4.2 Expected number of patients choosing the new treatment

We assume that the prior distributions for the difference in efficacy between the new and standard treatments, and for the incidence of adverse reactions with each of the treatments, are independent. Again this is to simplify our presentation. In general there may be a tendency for more active drugs to be both more efficacious and to cause more adverse reactions, leading to dependent prior distributions. However, there are also many cases for which an assumption of independence is reasonable. Published examples include: infliximab (a monoclonal antibody against Tumor Necrosis Factor) for Crohn's disease (Targan et al., 1997); rosuvastatin to reduce low-density lipoprotein cholesterol, Olsson et al. (2001), Davidson et al. (2002), Saito et al. (2003); and calcineurin inhibitors for immunosuppression in liver transplantation, Perry and Neuberger (2005).

The more effective and safe a new treatment, the larger the number of patients who want to switch to it, and the higher the resulting benefit. Let M denote the potential total number of patients in country B who regularly receive the current treatment and who might switch to the new treatment. The number of subsequent users m is a fraction of M, and our model assumes that m is determined by a logistic weight function as follows.

$$m = M/[1 + \exp\{L\}], \text{ where} \quad (6.17)$$

$$L = u(v - \mu^*) + z\mu^{a*} + \psi\mu^*\mu^{a*} + 1.5\varphi, \quad (6.18)$$

$$\varphi^2 = (u^2 + \psi^2\mu^{a*2})\tau^{*2} + (z^2 + \psi^2\mu^{*2})\tau^{a*2} + \psi^2\tau^{*2}\tau^{a*2}. \quad (6.19)$$

Here $\psi\mu^*\mu^{a*}$ is an interaction term linking efficacy and safety, and u, v, z and ψ are suitably chosen constants.

Our rationale for this expression is as follows. If Δ was known and Δ^a was known to be zero, $[1+\exp\{u(v-\Delta)\}]^{-1}$ would be the probability p_s that a patient would switch to the new drug. We define v to be the value of Δ for which half of all patients would switch to the new drug. The value of u should be chosen so that

$$[1+\exp\{u(v-\Delta)\}]^{-1} = 0.75, \tag{6.20}$$

where the value of Δ is such that 75% of patients would switch to the new drug if Δ^a was known to be zero. The value of $z(>0)$ should be selected so that $[1+\exp\{z\Delta^a\}]^{-1} = 0.75$ for the value of Δ^a for which 75% of patients would switch to the new treatment if Δ was known to be v. Note that other values than 0.75 could equally well be used, either instead of 0.75 or, which some form of least-squares fitting may help, together with 0.75, in calculating u.

If we write $p_s = 1/[1+\exp\{u(v-\Delta)+z\Delta^a+\psi\Delta\Delta^a\}]$, then for $\Delta \neq 0$ and $\Delta^a \neq 0$ we have

$$\psi = \frac{\log(1/p_s - 1) - u(v-\Delta) - z\Delta^a}{\Delta\Delta^a}. \tag{6.21}$$

The values used for u, v, z and ψ need to be estimated on the basis of empirical data. Since our interest is in the preference of patients after they have received medical advice it would be best to gather this data from a survey of the views of medical practitioners, and then to make estimates using standard statistical methods.

If Δ and Δ^a were known, L could be replaced with $u(v-\Delta)+z\Delta^a+\psi\Delta\Delta^a$. If we write $K = u\Delta - z\Delta^a - \psi\Delta\Delta^a$, K is a measure of patients' preference for the new treatment and φ^2 is the posterior variance of K. In practice Δ and Δ^a are not known and are replaced by their posterior expectations in the expression (6.18) for L. The additional term 1.5φ represents the fact that in order to switch a user is likely to require a statistically significant estimated increase in K. Other similar multiples of φ, such as 1.96, are equally plausible.

6.5 DESIGNING A SEQUENTIAL BRIDGING STUDY

Bridging studies can be regarded as part of a sequential clinical trial starting with phase III trials in country A. This allows decisions on whether to continue with a further stage to be taken repeatedly on the basis of the strength of current evidence for the efficacy and safety of the new drug. This may be done along the lines described by Kikuchi and Gittins (2009) for a sequential phase III trial. Normally bridging studies have just one stage. However, as we shall see, efficiency improvements are possible if we allow more than one stage. The basic principle of our proposed procedure is that after each stage a further stage is carried out in the bridging study if and only if the expected net benefit from doing so is positive. To set out this procedure in detail we first need some further notation.

Let $n_i (i = 1, 2, \cdots)$ be the number of patients on each treatment in stage i of the bridging trials. Let $(\mu_{i-1}, \tau_{i-1}^2)$ and $(\mu_{i-1}^a, \tau_{i-1}^{a2})$ denote the means and variances of Δ and Δ^a, respectively, before stage i. It follows that these quantities satisfy equations (6.11) and (6.15) if we replace $(\mu', \tau'', \mu^*, \tau^*, \mu^{a\prime}, \tau^{a\prime\prime}, \mu^{a*}, \tau^{a*}, n)$ by $(\mu_{i-1}, \tau_{i-1}, \mu_i, \tau_i, \mu_{i-1}^a, \tau_{i-1}^a, \mu_i^a, \tau_i^a, n_i)$ respectively. Also $\mu_0 = \mu'$, $\tau_0 = \tau''$, $\mu_0^a = \mu^{a\prime}$, $\tau_0^a = \tau^{a\prime\prime}$.

Let Q_i and R_i both be expected net benefits for patients in country B conditional upon the results of the first i stages of the bridging study, and further conditions as follows. For Q_i the further conditions are that the bridging study is then terminated and the new drug is made available in country B. For R_i the further conditions are that a further stage of the bridging study takes place of a target sample size n_{i+1} which would be optimal if this was the final stage, and the new drug is then made available in country B provided that the posterior expected net benefit after the further sampling is positive. It follows that Q_i satisfies equation (6.16) with $(\mu^*, \mu^{a*}) = (\mu_i, \mu_i^a)$, $i = 1, 2, \cdots$, and that

$$R_i = \max_n \{ \int \int Q_{i+1}^+(\overline{z_B}, \overline{z_B^a}, n) f(\overline{z_B}) f(\overline{z_B^a})\, d\overline{z_B}\, d\overline{z_B^a} - n(2c + e\mu_i^a t_i/t_0) \}. \tag{6.22}$$

Here $Q_{i+1}^+ = \max[Q_{i+1}, 0]$, and the predictive densities $f(\overline{z_B})$ and $f(\overline{z_B^a})$ are $\mathrm{N}(\mu_i, \tau_i^2 + 2\sigma^2/n)$ and $\mathrm{N}(\mu_i^a, \tau_i^{a2} + 2\sigma^{a2}/n)$, by standard Bayesian theory for the normal distribution with known variance. The expected trial cost per patient is c. The expected incremental cost of adverse events for a patient taking the new drug during the bridging study is $e\mu_i^a t_i/t_0$, where t_i is the duration of

stage i of the bridging study. The incremental cost for n patients is required if we reckon that without the trial all $2n$ patients would be receiving the standard treatment.

Let us also define $Q_0 = 0$.

Note that the double integral in (6.22) is an expectation which may be calculated by taking the average of a large number of values of Q_{i+1}^+, each obtained by randomly sampling values of $\overline{z_B}$ and $\overline{z_B^a}$ and using equation (6.16).

Our proposed decision rules are as follows.

After i stages $(i = 0, 1, 2, \cdots)$
IF $R_i > Q_i^+$ continue the study with a further n_{i+1} patients receiving each of the two drugs,
ELSE IF $Q_i > 0$ THEN STOP, and ACCEPT the new drug,
ELSE IF $Q_i \leq 0$ THEN STOP, and REJECT the new drug.

Thus, for example, if $R_0 < 0$ no bridging study is carried out and we reject the new drug.

6.6 AN ILLUSTRATIVE EXAMPLE

6.6.1 Background

This illustrative example treats a real case. However some of the required data have not been published and plausible assumptions are made to fill those gaps. Further work by epidemiologists and health economists is needed to make this process more reliable. Further approximations are mentioned later. This means that this section illustrates the methodology rather than providing practical guidance for the case of rosuvastatin.

Cholesterol cannot dissolve in blood and has to be transported to cells by low-density lipoprotein (LDL). As more LDL circulates in the blood, it transports more lipid into the cells of arteries as molecules of low-density lipoprotein cholesterol (LDL-C, low-density lipoprotein combined with cholesterol) and can slowly build up in the inner walls of the arteries that feed the heart and brain. Then, it can form plaque, a thick, hard deposit that can narrow the arteries and make them less flexible resulting in heart attack or stroke. For this reason low-density lipoprotein cholesterol (LDL-C) is called "bad cholesterol".

Rosuvastatin is a new, synthetic, oral drug that reduces LDL-C and is intended for use by patients suffering from hypercholesterolemia. This illustrative example is based on three randomised double-blind clinical trials, one in the UK, one in North America and another in Japan. To illustrate our methodology we treat the UK and North American studies (Olsson et al., 2001; Davidson et al., 2002) as phase II and phase III trials respectively in country A (Caucasian patients). This seems reasonable since the patients' background and the design of these studies are very similar. We treat the Japanese study (Saito et al., 2003) as the first stage of a sequential bridging study in country B. Our methodology investigates whether the study should be continued or concluded. All three studies used a placebo as the comparative drug, so we shall compare the efficacy and safety of rosuvastatin with those of a placebo. To estimate the number of patients m who switch to the new drug the comparison should of course be with currently used drugs, so we note that our estimate will be an overestimate for this reason.

6.6.2 Application of the model

In the UK (phase II) and North American (phase III) studies the patients received rosuvastatin at 0 (placebo), 1, 2.5, 5, 10, 20, 40 or 80 mg per day and at 0, 5 or 10 mg per day, respectively. In both studies treatment was for a period of six weeks. The Japanese (bridging) study assessed rosuvastatin at 0, 1, 2.5, 5, 10, 20 or 40 mg. Thus $t_0 = t_1 = 0.12$ years. The studies were published by Olsson et al. (2001), Davidson et al. (2002) and Saito et al. (2003), respectively. For the sake of simplicity we shall measure both efficacy and safety in terms of what happens with a dose of 10 mg once daily for six weeks.

The UK study by Olsson et al. (2001) had two sessions. They recruited 142 patients in total and randomised them to receive placebo or rosuvastatin at 1, 2.5, 5, 10, 20, 40 or 80 mg per day for six weeks in the first session. In the second session 64 of the 142 patients were re-randomised to receive placebo, 40 mg or 80 mg. No record was published of which patients received which doses, and we shall treat the 206 records as though they were for 206 different patients. On this basis 16 patients received a dose of 10 mg and 29 patients received the placebo. Decrease in LDL-C concentration from the baseline value is our measure of efficacy and for the two groups was recorded as 95.5 ± 4.5 mg/dl (mean $\pm$ standard error) and 7.1 ± 3.4 mg/dl, respectively. With this information the mean value μ of δ prior to the US (phase III) study was set at $95.5 - 7.1 = 88.4$ mg/dl, and the variance τ^2 at $(\mu/2)^2$, expressing

considerable uncertainty as to the true value of δ.

The phase III prior for σ^2 was estimated from the UK records on the assumption that it is the same for the UK and US studies, and for all doses in the UK study. The numbers of patient records for placebo and each of the seven dose levels were 29, 13, 13, 17, 16, 13, 34 and 31 (in total fewer than 206 as some data was missing), with standard errors of 3.4, 4.9, 4.9, 4.6, 4.5, 4,9, 2.8, and 3.8. The usual unbiased estimate for σ^2 is $(29\times28\times3.4^2+13\times12\times4.9^2+13\times12\times4.9^2+17\times16\times4.6^2+16\times15\times4.5^2+13\times12\times4.9^2+34\times33\times2.8^2+31\times30\times3.8^2)/(28+12+12+16+15+12+33+30)$. We write this as $s^2/k = 53,464.6/158 = 338.4$.

We assume the conventional improper prior distribution proportional to σ^{-2} for σ^2 before taking account of the phase II study. Since s^2/σ^2 has a χ^2_k distribution it follows from Bayes theorem that with the UK data the posterior distribution for σ^2 is as shown in (6.9) with $g' = k$ and $a' = s^2$. These are then the values of the parameters g and a for the country A phase III prior distribution. With $\tau^2 = (\mu/2)^2 = 44.2^2$ it follows from equation (6.5) that $\omega = 5.77$.

The numbers of patients receiving 10 mg and placebo for whom data on adverse reactions were recorded in the UK study were 17 and 31, respectively. The number of adverse reactions was 6 for each of the two groups. With this information the mean value μ^a of δ^a prior to the Davidson study was set at 6/17 - 6/31 = 0.16, and the variance τ^{a2} at $(\mu^a/2)^2$.

In the North American study reported by Davidson et al. (2002), the changes in LDL-C from the baseline values were 0 ± 2.2 mg/dl (mean $\pm$ SE) and 79.6 ± 2.4 mg/dl with placebo (132 patients) and 10 mg of rosuvastatin (129 patients), respectively. Thus $\overline{z_A} = 79.6$ mg/dl, and the standard errors and sample sizes yield $s_A^2 = 132 \times 131 \times 2.2^2 + 129 \times 128 \times 2.4^2 = 178,802.4$. Since the BeBay model assumes identical sample sizes for the two treatment groups, we adjust these imbalanced sizes in the study by setting $l = (129+132)/2 = 130.5$. The parameters in (6.7) and (6.8) are accordingly calculated as follows;

$$\omega' = 0.015, \mu' = 79.6, g' = 418, a' = 232,280.4 \text{ and } \tau'^2 = \omega' a'/(g'-2) = 8.4.$$

The variance σ^2 is treated as known and equal to its mean value $a'/(g'-2) = 558.4$.

Individual patient data for adverse reactions are not published by Davidson et al. (2002), so the calculations (6.12) for sufficient statistics require a further assumption, namely that the adverse reactions which occurred all involved

different patients. During the trial five drug-related adverse reactions were observed in both the rosuvastatin and placebo groups. Thus our estimate for $2\sigma^{a2}$ is $\{5(1-5/129)^2+124(5/129)^2+5(1-5/132)^2+127(5/132)^2\}/(128+131) = 0.037$. Using equations (6.13), and since $\overline{z_A^a} = 5/129 - 5/132 = 0.00088$, we have $\mu^{a\prime} = 0.0076$ and $\tau^{a\prime 2} = 0.00027$. Note that the frequencies of adverse reactions in all the trials in our example are too low for a normal approximation to their distributions to be accurate. However to avoid further notational complication, and since our example is only illustrative, we continue with the normal approximation.

The study of rosuvastatin in Japanese patients (Saito et al., 2003) showed very similar changes of LDL-C to those observed in the country A study. The decreases from the baselines were 6.1 ± 6.1 mg/dl (mean $\pm$ SE) and 90.6 ± 5.1 mg/dl for placebo ($n = 15$) and 10 mg of rosuvastatin ($n = 15$), respectively, for six weeks of treatment. These results give that $\overline{z_B} = 90.6 - 6.1 = 84.5$ mg/dl so that, using equation (6.10), $\tau''^2 = 8.4 + (79.6/2)^2 = 1592.4$, $\mu^* = 84.3$ mg/dl and $\tau^{*2} = 71.2$.

Based on the occurrence of drug-related adverse events in this Japanese study, $\overline{z_B^a} = 2/15 - 3/15 = -1/15$, and using the adverse events version of equation (6.10) we have $\tau^{a\prime\prime 2} = \tau^{a\prime 2} + (\mu^{a\prime}/2)^2 = 0.00027 + (0.0076/2)^2 = 0.00028$, giving $\mu^{a*} = -0.00093$ and $\tau^{a*2} = 0.00024$.

According to the 2005 annual report of the Ministry of Health, Labour and Welfare in Japan, 31 million people are estimated to receive medical treatment for hypercholesterolemia. This serves as an estimate for M, the number of potential users of rosuvastatin. There are no official figures for the unit benefit b, the cost e for the treatment of an adverse event, and the subsequent period of use T in Japan, so we constructed estimates for these quantities as follows. An industrial & marketing research company, Fuji-Keizai Group, estimated that 320 billion Japanese Yen (JPY) was spent on the treatment in 2007 (see https://www.fujikeizai.co.jp/). Also, according to the new diagnosis guideline recommended by the Japan Atherosclerosis Society, the normal standard range of LDL-C is (70, 139) mg/dl (Teramoto et al., 2007). The patients in the Japanese study had 190 ± 13.8 mg/dl (mean$\pm$SD) before treatment. The unit benefit is calculated proportionally up to a maximum change of $190 - (70 + 139)/2 = 85.5$ mg/dl. This gives $b = 3.2 \times 10^{11}/(85.5 \times M) = 120.7$ JPY per mg/dl per patient. Thus our assumption is that if there is a one mg/dl decrease in LDL-C for one patient this reduces the medical-care cost for hypercholesterolemia in Japan by 120.7 JPY. We also assume that the cost for the treatment of an adverse event is $e = 3.0 \times 10^3$ JPY (this follows Bassi et al. (2004) who estimate that the primary care cost in the

UK is 20 pounds per medical consultation), the cost per patient of a clinical trial is 3.0×10^6 JPY according to a benchmarking report (Cutting Edge Information, 2006), and $T = 50$ (years) in equation (6.16).

The mean ($\pm$ SD) of the currently approved prices for the 12 statins available in Japan is 73.0 ($\pm$ 48.4) JPY for the recommended daily dose (see http://www.okusuri110.com). We shall investigate the consequences of incremental costs $k_1 - k_2$ of 15, 20 and 25 JPY per day. Note that these figures do not include any explicit estimate for the difference in indirect costs between the two drugs, as no relevant data are available.

Since survey data on patients' preferences is not available, judgmental estimates are required for u, v, z and ψ. Suppose that $v = 50$ mg/dl and that 75% of patients would switch to the new drug if $\Delta = 100$ mg/dl. From (6.20) it follows that $u = 0.022$. We also assume that $z = 3.75$ and $\psi = 0.001$.

6.6.3 Results

With these assumptions Monte Carlo integration was used to investigate the relationship between expected net benefit and sample size for one further stage of the bridging study and with enrolment rates r varying from one patient every 40 days up to 125 patients per day. The results for $k_1 - k_2 = 15$ JPY are shown in Figures 6.1, 6.2 and 6.3, together with Q_1, the expected net benefit with no further stage of the bridging study. The values of Q_1 (called Q in the figures) are 4.469×10^{12}, 2.734×10^{12} and 0.999×10^{12} for daily incremental costs $k_1 - k_2$ of 15, 20 and 25 JPY respectively, all with $T = 50$. Maximised values of the expected net benefit R_1 and of the sample size for $k_1 - k_2 = 15$ are shown in Table 6.1.

If the health authority works on the basis that the drug price should be an increasing function of societal benefit, the drug company might expect on average to achieve a higher drug price by extending the bridging study in Japan to a further stage.

If the enrolment rate r is larger than 0.025, then the maximised expected net benefit R_1 is larger than Q_1. It follows from our discussion in section 4 that it would be cost effective in these cases to extend the bridging study in Japan to a further stage. On the other hand, for an enrolment rate less than or equal to 0.025 of a patient per day, $R_1 - Q_1$ is negative for all n and it is not worth proceeding to a further stage in the study. Even for rather higher enrolment rates R_1 does not improve greatly on Q_1 and the company might well decide not to extend the trial so that they can obtain the licence earlier,

Table 6.1: Optimal n and maximised expected net benefits for different enrolment rates($k_1 - k_2 = 15$ JPY)

r	n	R_1 ($\times 10^{12}$ JPY)	$(R_1 - Q_1)$ ($\times 10^{12}$ JPY)	$(R_1 - Q_1)/Q_1$
125	2,214	4.943	0.474	0.106
25	1,161	4.915	0.446	0.100
5	416	4.857	0.388	0.087
1.0	140	4.761	0.292	0.065
0.5	88	4.705	0.236	0.053
0.25	42	4.646	0.177	0.040
0.10	18	4.554	0.085	0.019
0.050	7	4.497	0.028	0.006
0.025	N.A.	N.A.	Negative	N.A.

Table 6.2: Optimal n and maximised expected net benefits for different incremental costs

$k_1 - k_2$	r	n	R_1 ($\times 10^{12}$ JPY)	$(R_1 - Q_1)$ ($\times 10^{12}$ JPY)
15.0	1.0	140	4.761	0.292
	0.5	88	4.705	0.236
	0.25	42	4.646	0.177
20.0	1.0	140	2.935	0.201
	0.5	79	2.900	0.166
	0.25	425	2.860	0.126
25.0	1.0	197	1.112	0.113
	0.5	91	1.097	0.098
	0.25	70	1.077	0.078

and to accept the price for the drug which they are able to negotiate on the basis of the existing evidence. The recent paper by Claxton et al. (2008) on value based pricing of NHS drugs gives a useful comment on the relevant issues.

Table 6.2 extends Table 6.1 to cover incremental daily costs 15, 20 and 25 JPY for moderate enrolment rates. The values of $(R_1 - Q_1)$ decrease as the incremental cost increases. If $k_1 - k_2 \geq 30.0$ JPY, then Q_1 is less than zero and we stop the trial and reject the new drug according to the decision rule discussed in section 6.5.

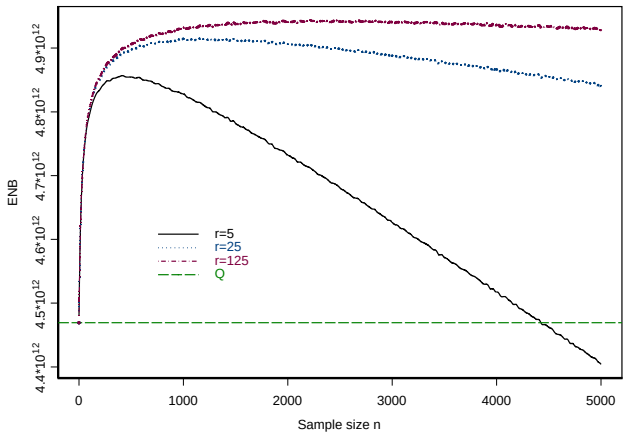

Figure 6.1: Expected net benefit with high enrolment rates

6.7 FINAL COMMENTS

The behavioural Bayes methodology is based on the maximisation of expected net benefit. As we have seen it is an approach which lends itself to the design of bridging studies. Apart from that application it is a more rational approach than that of much current practice for determining sample sizes in clinical trials. In particular it avoids the potential absurdities of the most widely used current criteria, which are still based on conventional, and rather arbitrary, requirements for achieving specified values for the size and power of classical hypothesis tests.

It also takes account of patients' preferences for their treatments, by means of the weight function. This is an important feature, and distinguishes it from the methodology of most publications and practice in health economics. It is not current practice to investigate patients' preferences as part of the assessment of a drug. In the authors' opinion it is very desirable that this should become standard practice, not least because patients are also taxpayers.

It is also not current practice for the health authorities to decide drug prices by considering how much the new drug can reduce the current social expenditure for treatment. Again this is a feature which we recommend for inclusion in the official guidelines.

To calculate the expected net benefit as described in this chapter requires

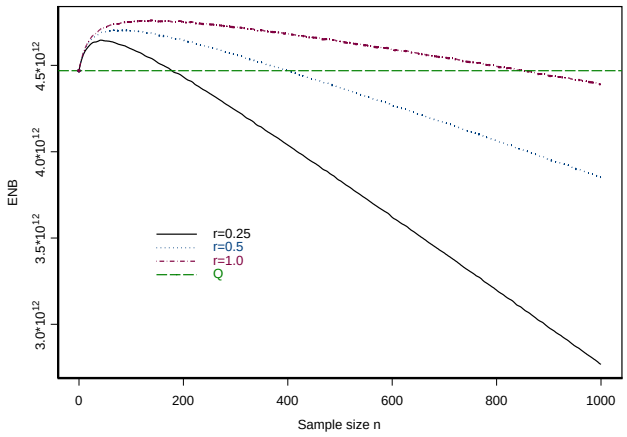

Figure 6.2: Expected net benefit with moderate enrolment rates

more investigation of epidemiology, preferences, costs and expected therapeutic effect, and more detailed computation, than a conventional calculation does. To proceed in this way, however, does focus on what really matters, so our strong recommendation is that, even though there will inevitably be some shortcomings in the accuracy and completeness of the available data, an analysis along these lines should be carried out. As computing capacity continually increases the increased demand is no longer prohibitive, but it needs to be kept in mind as the details of our methodology are worked out.

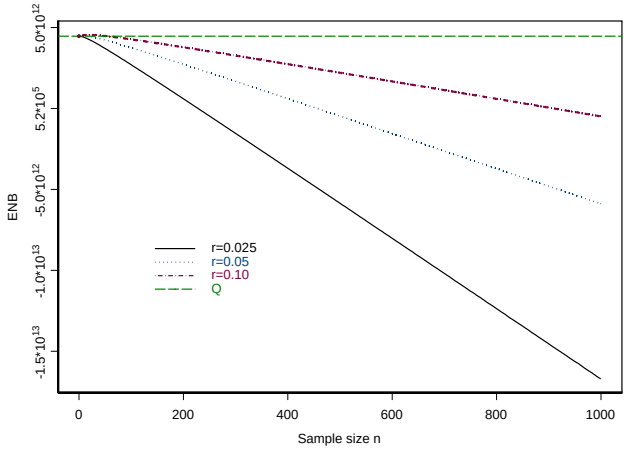

Figure 6.3: Expected net benefit with low enrolment rates

Chapter 7

Conclusion

In this thesis a methodology has been developed in some detail for determining the sample size for phase III clinical trials so as to maximise the expected net benefit. The same procedure can be used to assess how much a national health authority should be willing to pay for a new drug.

The calculations involve potentially lengthy computer simulations and require access to substantial data banks to be fully effective. However the methodology is more soundly based than most current practice. I look forward to seeing practice move in this direction over the next few years.

Bibliography

Adcock, C. (1987). A bayesian approach to calculating sample size for multinomial sampling. *The Statisticians 36*, 155–159.

Adcock, C. (1997). Sample size determination: a review. *The Statisticians 46*, 261–283.

Ades, A., G. Lu, and K. Claxton (2004). Expected value of sample information calculations in medical decision modeling. *Medical Decision Making 24*, 207–227.

Ades, A. E., G. Lu, and J. P. T. Higgins (2005). The Interpretation of Random-Effects Meta-Analysis in Decision Models. *Med Decis Making 25*(6), 646–654.

Andrews, T., C. Ballantyne, J. Hsia, and J. Kramer (2001). Achieving and maintaining National Cholesterol Education Program low-density lipoprotein cholesterol goals with five statins. *The American Journal of Medicine 111*, 185–191.

Anscombe, F. (1963). Sequential medical trials. *Journal of the American Statistical Association 58*, 365–383.

Assessment Health Technology (2004). *A pilot study on the use of decision theory and value of information analysis as part of the NHS Health Technology Assessment Programme*. Assessment Health Technology.

Bassi, A., S. Dodd, and K. Bodger (2004). Cost of illness of inflammatory bowel disease in the UK: a single centre retrospective study. *Gut 53*, 1471–1478.

Bazalo, G. (2001). Managed care trends in statin usage. *Managed Care 10*(10), 48–59.

Bernardo, J. and A. Smith (1994). *Bayesian Theory (Wiley Series in Probability and Statistics)*. Chichester: Wiley.

Berry, D. A. and C.-H. Ho (1988). One-sided sequential stopping boundaries for clinical trials: A decision- theoretic approach. *Biometrics 44*(1), 219–227.

Black, W. (1990). The CE plane: a graphic representation of cost-effectiveness. *Medical Decision Making 10*, 212–214.

Briggs, A. and A. Gray (1998). Sample size and power calculation for stochastic cost-effectiveness analysis. *Medical Decision Making 18*, S81–S92.

Brutti, P., F. De Santis, and S. Gubbiotti (2008). Robust Bayesian sample size determination in clinical trials. *Statistics in Medicine 27*, 2290–2306.

Canadian Coordinating Office for Health Technology Assessment. (1997, November). Guidelines for Economic Evaluation of Pharmaceuticals (2nd edition). http://www.farmacoeconomia.com/Directrices/canada.pdf.

Chang, M. (2007). *Theory and Application of Adaptive Design*. London: Chapman and Hall/CRC Biostatistics.

Cheng, Y., F. Su, and D. A. Berry (2003). Choosing sample size for a clinical trial using decision analysis. *Biometrika 90*(4), 923–936.

Christian, P., J. West, K. P., S. K. Khatry, S. C. LeClerq, E. K. Pradhan, J. Katz, and S. R. Shrestha (2003). Effects of maternal micronutrient supplementation on fetal loss and infant mortality: a cluster-randomized trial in Nepal. *American Society for Clinical Nutrition 78*, 1194–1202.

Clark, W., A. Burls, F. Song, J. Raftery, P. Barton, C. Cummins, and A. Fry-Smith (2001, July). Infliximab for the treatment of Crohns disease. http://www.nice.org.uk/pdf/FinalreportrevisedEconomicAlalysissection.pdf.

Claxton, K. (1999a). The irrelevance of inference: a decision making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics 18*, 341–364.

Claxton, K. (1999b). The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics 18*(3), 341 – 364.

Claxton, K., A. Briggs, M. Buxton, A. Culyer, C. McCabe, S. Walker, and M. Sculpher (2008). Value based pricing for nhs drugs: an opportunity not to be missed? *BMJ 336*, 251–254.

Claxton, K. and J. Posnett (1996a). An economic approach to clinical trial design and research priority-setting. *Health Economics 5*(6), 513–524.

Claxton, K. and J. Posnett (1996b). An economic approach to clinical trial design and research priority-setting. *Health Economics 5*, 513–24.

Claxton, K., M. Sculpher, and M. Drummond (2002). A rational framework for decision making by the National Institute For Clinical Excellence (NICE. *Lancet 360*, 711–715.

Claxton, K. and K. Thompson (2001). A dynamic programming approach to the effcient design of clinical trials. *Journal of Health Economics 20*, 797–822.

Connelly, L. (2003). Balancing the number and size of sites: an economic approach to the optimal design of cluster samples. *Controlled Clinical Trials 24*, 544–559.

Cutting Edge Information (2006). Clinical operations: Accelerating trials, allocating resources and measuring performance. http://www.ClinicalTrialBenchmarking.com.

Davidson, M., P. Ma, E. Stein, A. Gotto, A. Raza, R. Chitra, and H. Hutchinson (2002). Comparison of effect on low-density lipoprotein cholesterol and high-density lipoprotein cholesterol with Rosuvastatin versus Atorvastatin in patients with type IIa or IIb hypercholesterolemia. *The American Journal of Cardiology 89*, 268–275.

Department of Health and Aging (Australian Government). (2002, September). Guidelines for the Pharmaceutical Industry on Preparation of Submissions to the Pharmaceutical Benefits Advisory Committee; Including Submissions Involving Economic Analysis. http://www.health.gov.au/internet/wcms/publishing.nsf/Content/health-pbs-general-pubs-guidelines-index.htm.

Desu, M. and D. Raghavarao (1990). *Sample Size Methodology*. Boston, MA: Academic Press.

Dobson, A. (1984). Calculating sample size. *Trans. Menzies Foundation 7*, 75–79.

Donner, A. (1992). Sample size requirements for stratified cluster randomization designs. *Statistics in Medicine 11*, 743–750.

Donner, A., N. Brikett, and C. Buck (1981). Randomization by cluster: sample size requirements and analysis. *American Journal of Epidemiology 114*, 906–914.

Dumville, J., S. Hahn, and D. Torgerson (2006). The use of unequal randomisation ratios in clinical trials: A review. *Contemporary Clinical Trials 27*, 1–12.

Ebrahim, S., F. Lampe, G. Wannamthee, and G. Smith (1998). Cholesterol and coronary heart disease: screening and treatment. *Effective Health Care 4*(1), 1–16.

Eckermann, S. and A. Willan (2007). Expected value of information and decision making in hta. *Health Economics 16*(2), 195–209.

Eckermann, S. and A. Willan (2008a). Time and expected value of sample information wait for no patient. *Value in Health 11*(3), 522–526.

Eckermann, S. and A. Willan (2009). Globally optimal trial design for local decision making. *Health Economics 18*(2), 203–216.

Eckermann, S. and A. R. Willan (2008b). The Option Value of Delay in Health Technology Assessment. *Med Decis Making 28*(3), 300–305.

Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults (2002). *Third Report of the National Cholesterol Education-Program (NCEP) Expert Panel on Detection, Evaluation andTreatment of High Blood Cholesterol in Adults (Adult TreatmentPanel III) final report*, Volume 106 of *31*. Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults.

Gail, M., D. Byar, T. Pcchacck, and D. Corle (1992). Aspects of statistical design for the Community Intervention Trial for Smoking Cessation. *Controlled Clinical Trials 13*, 6–21. for the COMMIT study Group.

Gajewski, B. and M. Mayo (2006). Bayesian sample size calculation in phase II clinical trials using a mixture of informative priors. *Statistics in Medicine 25*(15), 2554–2566.

Garduber, J., M. Huebner, and J. Jetton (2000). Power and sample size assessments for tests of hypotheses on cost-effectiveness ratios. *Health Economics 9*, 227–34.

Gaspoz, J.-M., P. G. Coxson, P. A. Goldman, L. W. Williams, K. M. Kuntz, M. M. Hunink, and L. Goldman (2002). Cost effectiveness of Aspirin, Clopidogrel, or both for secondary prevention of coronary hart disease. *New England Journal of Medicine 346*(23), 1800–1806.

Gittins, J. and H. Pezeshk (2000a). A behavioral Bayes method for determining the size of a clinical trial. *Drug Information Journal 34*, 355–363.

Gittins, J. and H. Pezeshk (2000b). How large should a clinical trial be? *The Statistician 49*(2), 177–187.

Gittins, J. and H. Pezeshk (2002a). A decision theoretic approach to sample size determination in clinical trials. *Journal of Biopharmaceutical Statistics 12*(4), 535–551.

Gittins, J. and H. Pezeshk (2002b). A fully Bayesian approach to calculating sample sizes for clinical trials with binary responses. *Drug Information Journal 36*, 143–150.

Grundy, P., M. Healy, and D. Rees (1956). Economic choice of the amount of experimentation. *Journal of the Royal Statistical Society, Series A 18*, 32–48.

Halpern, J., B. Brown Jr., and J. Hornberger (2001). The sample size for a clinical trial: A bayesian-decision theoretic approach. *Statistics in Medicine 20*(6), 841–858.

Hayes, R. and S. Bennett (1999). Simple sample size calculation for cluster-randomized trials. *International Journal of Epidemiology 28*, 319–326.

Hornberger, J., J. Brown, and J. Halpern (1995). Designing a cost-effective clinical trial. *Statistics in Medicine 14*(20), 2249–2259.

Hornberger, J., MD, MS, P. Eghtesady, MD, and PhD (1998). The cost-benefit of a randomized trial to a health care organization. *Controlled Clinical Trials 19*(2), 198 – 211.

Hutton, J. and R. Owens (1993). Bayesian sample size calculation and prior beliefs about child sexual abuse. *The Statistician 42*, 399–404.

ICH steering committee (Ed.) (1998). *E5: Ethnic Factors Information to Support Drug Registration*, Volume 63. International Conference on Harmonization.

Joseph, L., R. du Berger, and P. Belisle (1997). Bayesian and mixed Bayesian likelihood criteria for sample size determination. *Statistics in Medicine 16*, 769–81.

Joseph, L., R. du Berger, and P. Bélisle (1997). Bayesian and mixed Bayesian/likelihood criteria for sample size determination. *Statistics in Medicine 16*, 769–781.

Joseph, L., D. Wolfson, and R. du Berger (1995). Sample size calculations for binomial proportions via a highest posterior density intervals. *Statistics in Medicine 44*, 143–154.

Kikuchi, T. and J. Gittins (2009). A behavioral bayes method to determine the sample size of a clinical trial considering efficacy and safety. *Statistics in Medicine 28*(18), 2293–2306.

Kikuchi, T. and J. Gittins (2010a). A Bayesian adaptive design for the evaluation of a new drug in a bridging study. *Biostatistics, Bioinformatics and Biomathematics 1*(2), 73–100. Mili Publications (India).

Kikuchi, T. and J. Gittins (2010b). A behavioral Bayes approach to the determination of sample size for clinical trials considering efficacy and safety: imbalanced sample size in treatment groups. *Statistical Methods in Medical Research 00*, 1–12. Published online 11th March, 2010.

Kikuchi, T. and J. Gittins (2011). A behavioral Bayes approach for sample size determination in cluster randomised clinical trials. *Journal of Royal Statistical Society C 60*, 1–14. accepted for publication.

Kikuchi, T., H. Pezeshk, and J. Gittins (2008). A Bayesian cost-benefit approach to the determination of sample size in clinical trials. *Statistics in Medicine 27*(1), 68–82. Published online 13th June, 2007.

Kocpsell, T., D. Martin, and P. Dichr (1991). Data analysis and sample size issues in evaluations of community-based health promotion and disease prevention programs: a mixed-model analysis of variance approach. *Journal of Clinical Epidemiology 44*, 701–713.

Koerkamp, G., M. Hunink, T. Stijnen, J. Hammitt, K. Kuntz, and M. Weinstein (2007). Limitations of acceptability curves for presenting uncertainty in cost-effectiveness analysis. *Medical Decision Making 27*, 101–111.

Lambert, P., A. Sutton, K. Abrams, and D. Jones (2005). How vague is vague? a simulation study of the impact of the use of vague prior distributions in mcmc using winbugs. *Statistics in Medicine 24*(15), 2401–2428.

Laska, E., M. Meisner, and C. Siegel (1999). Power and sample size in cost-effectiveness analysis. *Medical Decision Making 19*, 339–43.

Lee, S. and M. Zelen (2000). Clinical trials and sample size considerations: another perspective. *Statistical Science 15*, 95–103.

Lemeshow, S., J. D. Homser, J. Klar, and S. Lwanga (1990). *Adequacy of Sample Size in Health Studies*. Chichester: John Wiley and Sons.

Lindley, D. (1997). The choice of sample size. *The Statistician 46*, 129–138.

Liu, J., N. Maniadakis1, A. Gray, and M. Rayner (2002). The economic burden of coronary heart disease in the UK. *Heart 88*, 597–603.

Müller, H. and H. Schäfer (2001). Adaptive group sequential designs for clinical trials: combining the advantages of adaptive and of classical group sequential approaches. *Biometrics 57*, 886–891.

National Institute for Clinical Excellence (2004, October). The Institute's Interim Guidance to Manufacturers and Sponsors.

Norwegian Medicines Agency. (2004, November). Norwegian Guidelines for Pharmacoeconomic Analysis in Connection with Applications for Reimbursement. http://www.legemiddelverket.no/eng/reg/pharmacoeconomic-analysis-guidelines.htm.

O'Hagan, A. and J. Forster (1994). *Kendall's Advanced Theory of Statistics: Bayesian Inference*. London: Arnold.

O'Hagan, A., J. Steens, and J. Montmartin (2000). Inferences for the cost-effectiveness acceptability curve and cost-effective ratio. *Pharmacoeconomics 17*, 339–349.

O'Hagan, A. and J. Stevens (1991). Bayesian assessment of sample size for clinical trials of cost effectiveness. *Medical Decision Making 21*, 219–230.

O'Hagan, A. and J. Stevens (2001). Bayesian assessment of sample size for clinical trials of cost-effectiveness. *Medical Decision Making 21*, 219–30.

Olsson, A., J. Pears, J. Mckellar, J. Mizan, and A. Raza (2001). Effect of rosuvastatin on low-density lipoprotein cholesterol in patients with hypercholesterolemia. *The American Journal of Cardiology 88*, 504–508.

Patel, N. and S. Ankolekar (2007). A Bayesian approach for incorporating economic factors in sample size design for clinical trials of individual drug and portfolios of drugs. *Statistics in Medicine 26*, 4976–4988.

Perry, I. and J. Neuberger (2005). Immunosuppression: towards a logical approach in liver transplantation. *Clinical and Experimental Immunology 139*, 2–10.

Petersen, S., V. Peto, and M. Rayner (2004). *Coronary heart disease statistics 2004*. London: BHF.

Peto, R.and Pike, M. C., P. Armitage, N. E. Breslow, D. R. Cox, S. V. Howard, N. Martels, K. McPherson, J. Peto, and P. G. Smith (1976). Design and analysis of randomized clinical trials requiring prolonged observation of each patient, 1: Introduction and design. *British Journal of Cancer 34*, 585–612.

Peto, R. (1978). Clinical trial methodology. *Biomedicine 28*, 24–36.

Pezeshk, H. (2003). Bayesian techniques for sample size determination in clinical trials: a short review. *Statistical Methods in Medical Research 7*(12), 489–504.

Pham-Gia, T. and N. Turkkan (1992). Sample size determination in bayesian analysis. *The Statisticians 41*, 389–392.

Pickin, D., C. McCabe, L. Ramsay, N. Payne, I. Haq, W. Yeo, and P. Jackson (1999). Cost effectiveness of HMG-CoA reductase inhibitor (statin) treatment related to the risk of coronary heart disease and cost of drug treatment. *Heart 82*, 325–332.

Pinto, E., A. Willan, and B. O'Brien (2005). Cost-effectiveness analysis for multinational clinical trials. *Statistics in Medicine 24*, 1965–1982.

Pocock, S. (1979). Allocation of patients to treatment in clinical trials. *Biometrics 35*, 183–97.

Pratt, J., H. Raiffa, and R. Schlaifer (1994). *Introduction to statistical decision theory. In: Pratt JW, Raiffa H, Schlaifer R, eds.* Cambridge, Massachusetts: The MIT Press.

Present, D., P. Rutgeerts, S. Targan, S. Hanauer, L. Mayer, R. Van Hogezand, D. Podolsky, B. Sand, T. Braakman, K. DeWoody, T. Schaible, and S. Van Deventer (1999). Infliximab for the treatment of fistulas in patients with Crohn's disease. *The New England Journal of Medicine 340*(18), 1398–1405.

Press Release (2007, February). HUMIRA (adalimumab) receives FDA approval for treatment of crohn's disease. http://www.abbott.com/global/url/pressRelease/enUS/ 60.5:5/PressRelease0428.htm.

Raiffa, H. and R. Schlaifer (1961). *Applied Statistical Decision Theory.* Boston, MA: Division of Research, Harvard Business.

Rutgeerts, P., G. D'haens, S. Targan, E. Vasiliauskas, S. Schreiber, D. Present, L. Mayer, R. Hogezand, T. Braakman, K. Dewoody, T. Schaible, and V. D. S.J.H. (1999). Efficacy and safety of retreatment with anti-tumor necrosis factor antibody (infliximab) to maintain remission in Crohn's disease. *Gastroenterology 117*, 761–769.

Saito, Y., Y. Goto, A. Dane, K. Strutt, and A. Roza (2003). Randomized

dose-response study of Rosuvastatin in Japanese patients with hypercholesterolemia. *Journal of Atherosclerosis and Thrombosis 10*(6), 329–336.

Sakai, F., M. Iwata, Y. Itoyama, S. Tsuji, Y. Fukuuchi, G. Sobue, K. Nakashima, and M. Morimatsu (2002). Zolmitriptan is effective and well tolerated in Japanese patients with migraine: a dose-response study. *Cephalalgia 22*, 376–383.

Sambucini, V. (2010). A bayesian predictive strategy for an adaptive two-stage design in phase ii clinical trials. *Statistics in Medicine 29*(13), 1430–1442.

Samuelson, W. and R. Zeckhauser (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty 1*, 7–59.

Schwartz, G., M. Bolognese, B. Tremblay, R. Caplan, H. Hutchinson, A. Raza, and M. Cressman (2004). Efficacy and safety of rosuvastatin and atorvastatin in patients with hypercholesterolemia and a high risk of coronary heart disease: A randomized, controlled trial. *American Heart Journal 148*(1), 105–113.

Shaper, A., S. Pocock, M. Walker, N. Choen, C. Wale, and A. Thompson (1981). British regional heart study: cardiovascular risk factors in middle-aged men in 24 towns. *British Medical Journal 283*, 179–186.

Shipley, M., P. Smith, and M. Dramix (1989). Calculation of power for matched pair studies when randomization is by group. *International Journal of Epidemiology 18*, 457–461.

Shoukri, M. and S. Martin (1992). Estimating the number of clusters for the analysis of correlated binary response variables from unbalanced data. *Statistics in Medicine 11*, 751–760.

Simon, R. (1989). Optimal two-stage designs for phase II clinical trials. *Controlled Clinical Trials 10*(1), 1–10.

Smith, P. and R. Morrow (1996). *Field Trials of Health Interventions in Developing Countries: A Toolbox*. London: Macmillan.

Spiegelhalter, D. (2001). Bayesian methods of analysis for cluster randomized trials with continuous responses. *Statistics in Medicine 20*, 453–452.

Spiegelhalter, D., K. Abrams, and J. Myles (2003). *Bayesian approaches to clinical trials and health-care evaluation*. John Wiley & Sons,Ltd.

Spiegelhalter, D. and L. Freedman (1986). A predictive approach to selecting

the size of a clinical trial, based on subjective clinical opinion. *Statistics in Medicine 5*, 1–13.

Spiegelhalter, D., L. Freeman, and M. Parmar (1994). Bayesian approaches to randomized trials. *Journal of the Royal Statistical Society A 157*, 357–416.

Tan, S. and D. Machin (2002). Bayesian two-stage designs for phase II clinical trials. *Statistics in Medicine 21*, 1991–2012.

Targan, S., S. Hanauer, S. van Deventer, L. Mayer, D. Present, T. Braakman, K. DeWoody, T. Schaible, P. Rutgeertes, and F. the Crohn's Disease cA2 Study Group (1997). A short-term study of chimeric monoclonal antibody cA2 to tumor necrosis factor α for Crohn's disease. *The New England Journal of Medicine 337*, 1029–35.

Teramoto, T., J. Sasaki, H. Ueshima, G. Egusa, M. Kinoshita, K. Shimamoto, H. Daida, S. Biro, K. Hirobe, T. Funahashi, K. Yokote, and M. Yokode (2007). Risk factors of atherosclerotic diseases. executive summary of Japan Atherosclerosis Society (jas) guideline for diagnosis and prevention of atherosclerosis cardiovascular diseases for Japanese. *Journal of Atherosclerosis and Thrombosis 14*(6), 247–275.

Thompson, S., . Pyke, S.D, and H. R.J. (1997). The design and analysis of paired cluster randomized trials: an application of meta-analysis techniques. *Statistics in Medicine 16*, 2063–2079.

Thompson, S., D. Warn, and R. Turner (2004). Bayesian methods of analysis for binary outcome data in cluster randomized trials on the absolute risk scale. *Statistics in Medicine 23*, 389–410.

Tunstall-Pedoe, H. (1991). The Dundee coronary risk disk for management of change in risk factors. *BMJ 303*, 744–747.

Turner, R., R. Omar, and S. Thompson (2001). Bayesian methods of analysis for cluster randomized trials with binary outcome data. *Statistics in Medicine 20*, 453–472.

van Gijn, J. (1999). Low doses of aspirin in stroke prevention. *The Lancet 353*(9171), 2172–2173.

Van Hout, B. A., M. J. Al, G. S. Gordon, and F. Rutten (1994). Costs, effects and C/E-ratios alongside a clinical trial. *Health Economics 3*, 309–319.

Wang, M. and A. Gelfand (2002). A simulation-based approach to Bayesian

sample size determination for performance under a given model and for separating models. *Statistical Science 17*(2), 193–208.

Willan, A. (2001). Analysis, sample size, and power for estimating incremental net health benefit from clinical trial data. *Controlled Clinical Trials 22*, 228–237.

Willan, A. (2008). Optimal sample size determinations from an industry perspective based on the expected value of information. *Clinical Trials 5*, 587–594.

Willan, A. and S. Eckermann (2010). Optimal clinical trial design using value of information methods with imperfect implementation. *Health Economics 19*(5), 549–561.

Willan, A. and M. Kowgier (2008). Determining optimal sample size for multi-stage randomised clinical trials using value of information methods. *Clinical Trials 5*, 289–300.

Willan, A. and B. O'Brien (1999). Sample size and power issues in estimating incremental cost-effectiveness ratios from clinical trials data. *Health Economics 8*, 203–11.

Willan, A. and E. Pinto (2005). The expected value of information and optimal clinical trial design. *Statistics in Medicine 24*, 1791–1806.

Willan, A. R. (2007). Clinical decision making and the expected value of information. *Clinical Trials 4*(3), 279–285.

Wonderling, D., C. McDermott, M. Buxton, A. Kinmonth, S. Pyke, S. Thompson, and D. Wood (1996). Costs and cost effectiveness of cardiovascular screening and intervention: the British family heart study. *BMJ 312*, 1269–1273.

Wood, D., A. Kinmonth, G. Davies, J. Yarwood, S. Thompson, S. Pyke, Y. Kok, R. Cramb, C. Guen, T. Marteau, and P. Durrington (1994a). British family heart study: its design and method, and prevalence of cardiovascular risk factors. *British Journal of General Practice 44*, 62–67.

Wood, D., A. Kinmonth, G. Davies, J. Yarwood, S. Thompson, S. Pyke, Y. Kok, R. Cramb, C. Guen, T. Marteau, and P. Durrington (1994b). Randomised controlled trial evaluating cardiovascular screening and intervention in general practice: principal results of British family heart study. *BMJ 308*, 313–320.

Yuh, L., S. Beal, M. Davidian, F. Harrison, A. Hester, K. Kowalski, E. Vonesh, and R. Wolfinger (1994). Population pharmacokinetic/pharmacodynamic methodology and applications: a bibliography. *Biometrics 50*, 566–575.