THE UNIVERSITY OF WARWICK

Library Declaration and Deposit Agreement

1. STUDENT DETAILS

Please complete the following:

Full name: ……

University ID number: ……

2. THESIS DEPOSIT

2.1 I understand that under my registration at the University, I am required to deposit my thesis with the University in BOTH hard copy and in digital format. The digital version should normally be saved as a single pdf file.

2.2 The hard copy will be housed in the University Library. The digital version will be deposited in the University's Institutional Repository (WRAP). Unless otherwise indicated (see 2.3 below) this will be made openly accessible on the Internet and will be supplied to the British Library to be made available online via its Electronic Theses Online Service (EThOS) service.
[At present, theses submitted for a Master's degree by Research (MA, MSc, LLM, MS or MMedSci) are not being deposited in WRAP and not being made available via EthOS. This may change in future.]

2.3 In exceptional circumstances, the Chair of the Board of Graduate Studies may grant permission for an embargo to be placed on public access to the hard copy thesis for a limited period. It is also possible to apply separately for an embargo on the digital version. (Further information is available in the *Guide to Examinations for Higher Degrees by Research*.)

2.4 *If you are depositing a thesis for a Master's degree by Research, please complete section (a) below. For all other research degrees, please complete both sections (a) and (b) below:*

(a) Hard Copy

I hereby deposit a hard copy of my thesis in the University Library to be made publicly available to readers (please delete as appropriate) EITHER immediately OR after an embargo period of ……..... months/years as agreed by the Chair of the Board of Graduate Studies.

I agree that my thesis may be photocopied. YES / NO (*Please delete as appropriate*)

(b) Digital Copy

I hereby deposit a digital copy of my thesis to be held in WRAP and made available via EThOS.

Please choose one of the following options:

EITHER My thesis can be made publicly available online. YES / NO (*Please delete as appropriate*)

OR My thesis can be made publicly available only after…..[date] (Please give date)
YES / NO (*Please delete as appropriate*)

OR My full thesis cannot be made publicly available online but I am submitting a separately identified additional, abridged version that can be made available online.
YES / NO (*Please delete as appropriate*)

OR My thesis cannot be made publicly available online. YES / NO (*Please delete as appropriate*)

3. GRANTING OF NON-EXCLUSIVE RIGHTS

Whether I deposit my Work personally or through an assistant or other agent, I agree to the following:

Rights granted to the University of Warwick and the British Library and the user of the thesis through this agreement are non-exclusive. I retain all rights in the thesis in its present version or future versions. I agree that the institutional repository administrators and the British Library or their agents may, without changing content, digitise and migrate the thesis to any medium or format for the purpose of future preservation and accessibility.

4. DECLARATIONS

(a) I DECLARE THAT:

- I am the author and owner of the copyright in the thesis and/or I have the authority of the authors and owners of the copyright in the thesis to make this agreement. Reproduction of any part of this thesis for teaching or in academic or other forms of publication is subject to the normal limitations on the use of copyrighted materials and to the proper and full acknowledgement of its source.

- The digital version of the thesis I am supplying is the same version as the final, hard-bound copy submitted in completion of my degree, once any minor corrections have been completed.

- I have exercised reasonable care to ensure that the thesis is original, and does not to the best of my knowledge break any UK law or other Intellectual Property Right, or contain any confidential material.

- I understand that, through the medium of the Internet, files will be available to automated agents, and may be searched and copied by, for example, text mining and plagiarism detection software.

(b) IF I HAVE AGREED (in Section 2 above) TO MAKE MY THESIS PUBLICLY AVAILABLE DIGITALLY, I ALSO DECLARE THAT:

- I grant the University of Warwick and the British Library a licence to make available on the Internet the thesis in digitised format through the Institutional Repository and through the British Library via the EThOS service.

- If my thesis does include any substantial subsidiary material owned by third-party copyright holders, I have sought and obtained permission to include it in any version of my thesis available in digital format and that this permission encompasses the rights that I have granted to the University of Warwick and to the British Library.

5. LEGAL INFRINGEMENTS

I understand that neither the University of Warwick nor the British Library have any obligation to take legal action on behalf of myself, or other rights holders, in the event of infringement of intellectual property rights, breach of contract or of any other right, in the thesis.

Please sign this agreement and return it to the Graduate School Office when you submit your thesis.

Student's signature: .....…… Date: .....

Designing a Series of Clinical Trials

by

Siew Wan Hee

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Health Sciences

The University of Warwick, Warwick Medical School
July 2012

Contents

List of Tables

List of Figures

Acknowledgements

I am grateful to Roche Products Limited and Warwick Medical School for the financial support for this PhD project.

Words fail to describe my gratitude to my family and friends. Their humours kept me rooted and their beliefs kept me going.

Many thanks to Professor Nigel Stallard. It is a pleasure to be under your tutelage and I cannot thank you enough for always being patient with me.

My sincere gratitude to Dr Simon Day, Mr Nelson Kinnersley and Mr Paul Mahoney for their invaluable feedbacks, tips and advices throughout the PhD project.

A massive thank you to the staff of Warwick Medical School as well, for their little ways have always mysteriously made the day-to-day work works.

There is one constant thing that keeps me awake at night and it is that I may have forgotten to thank the people who have helped me. If I have forgotten you, please, forgive me.

Finally, I like to acknowledge the permission from Hal Leonard Corporation to reprint excerpts from the copyrighted material 'Putting It Together', words and music by Stephen Sondheim. Copyright ©1984 RILTING MUSIC, INC. All rights administered by WB MUSIC CORP. All rights reserved. Used by permission.

Declaration

I declare that this thesis is all my own work except where I have stated otherwise. I also confirm that the thesis has not been submitted for a degree at another university. The materials in Chapter 6 have recently been submitted for consideration for publication.

Abstract

This thesis presents designs for a series of clinical trials where instead of designing clinical trials individually, each of the trials is designed as part of a series of trials. The framework of the design is based on a combination of classical frequentist and Bayesian approaches which is sometimes known as the hybrid approach. The unknown parameter of the treatment efficacy is assumed to be random and follows a prior distribution in the design stage but at the end of the trial a frequentist test statistic is used on the observed data to infer the parameter. The design introduced in Chapter 5 aims to determine an optimum sample size for each trial by optimizing the average power of each trial and the overall resources while fixing the conventional type I error. The design has the flexibility to either run sequentially or concurrently. The design is then extended to allow interim analyses in each trial (Chapter 6). The focus of the extended design is on a series of Bayesian decision-theoretic phase II trials and one frequentist phase III trial. At each interim stage, a decision is made based on the expected utilities of subsequent actions. There are four possible actions to choose from, namely, to continue the current trial by recruiting more patients, to initiate a new phase II trial, to abandon the development plan or to proceed to a phase III trial with this treatment against a control arm. For the last action, the phase III trial is designed with the hybrid methodology as described above. Finally, the prior distributions for each treatments are assumed to be correlated and as information is gathered from the previous and current trials, the current and following prior distributions are updated (Chapter 7).

Thesis Overview

This thesis germinated from a scholarly idea of designing a series of clinical trials using a hybrid methodology, that is, frequentist and Bayesian. The work is motivated by practical problems from asthma clinical research so it is important that the design is easily adapted and adopted in clinical trials. Although examples from asthma clinical research are used as illustrations for the design, the application of the proposed design can easily be used for other diseases by simply adjusting the hypotheses and prior densities. Also, although the design is presented in a simplistic form, it can be extended to include more complex parameters to model real-life expectations, conduct and end of trial strategies.

Current treatments for asthma are highly effective but they are not preventive, curative, nor disease modifying. In addition, for a minority group of patients presenting with persistent and severe asthma, current therapies are not effective in controlling the disease. New therapies are constantly being developed either as a single agent or a combined regimen. Results from a search in the ClinicalTrials.gov and published articles showed that there are many ongoing and completed clinical trials which aim to identify a treatment regimen that provides the most complete asthma control.

Thesis Overview

Pharmaceutical companies may be simultaneously developing drugs from classes that target different cells and mediators. As these therapies are targeting the same population and development is constrained by limitation of resources such as time and budget, drugs are ranked and prioritized for selection for development. Project evaluation and prioritization is often a complex challenge. There are four main factors to be considered when assessing potential treatments, namely, costs, probability of success, the rewards if the treatment is successful and time to develop (Senn, 2007, Ch. 24). The last criterion, time to develop, may be accounted for in the cost and reward with appropriate adjustment.

The work in this thesis is thus built upon this scenario. Suppose all the potential treatments targeting the same population can be tried concurrently but with the constraints of resources; what is the optimal sample size for each trial if instead by fixing the power of the trials its expectation is maximized? Also suppose that the population is very small such that trials have to be run sequentially and its viability in a larger phase III trial has to be taken into account if it is successful in the phase II trial; what is the optimal sample size?

The thesis is divided into four main parts where chapters within each part are more similar. Chapters in Part I are introductory where Chapter 1 briefly introduces clinical trials and the various phases of drug trials. Chapter 2 discusses the asthma disease and some of the classes of standard treatments and common endpoints employed in asthma clinical trials. Chapter 3 introduces some of the common notation and statistical nomenclatures used in this thesis. Finally, Chapter 4 briefly describes some of the common designs

of clinical trials.

The core of the thesis is in Parts II and III, covering Chapters 5 to 7. In Chapter 5 the design considers a series of trials as a whole instead of designing each trial individually in order to optimize the resources. The design is based on a hybrid of frequentist and Bayesian approaches where the type I error of each trial is maintained and the assurance, that is, the Bayesian expected power, is optimized. The applicability of the design is tested with examples where the primary endpoint is a continuous variable, for example, asthma trials testing bronchodilators where the FEV_1 is usually the primary endpoint.

The design is subsequently modified to consider the scenario where the population is much smaller, presented in Chapter 6. The formulation of the design also adopts the hybrid approach but because of the small population, the proposed design uses a Bayesian decision theoretic approach where patients are entered into the trial sequentially and their results are used to update the prior beliefs. The applicability of the design is tested with examples from trials for severe asthmatic patients. The primary endpoint is a discrete variable, that is, number of patients with no exacerbation during the treatment duration (analogous to $1 - p$ where p is the proportion of patients with at least one exacerbation).

As described in Chapter 2, different classes of drugs treat the disease differently. Therefore, drugs from the same class are related to each other in their effectiveness than to other drugs from different classes. As more information are obtained in earlier trials, they could be used to update the prior beliefs of the subsequent drugs within the same family. Chapter 7

extends the decision theoretic design by considering the dependency of each drug within a family. The prior distribution of each of the treatments are assumed to be related to each other and observed data from the preceding trials in the series are used to update the prior beliefs of the subsequent trials in each interim analysis.

Finally, Part IV rounds up the thesis with a summary and discussion in Chapter 8. Proposals of future work are also presented.

Part I

Background

Ounce by ounce, putting it together

Small amounts, adding up to make a work of art

First of all you need a good foundation

Otherwise it's risky from the start

Takes a lot of earnest conversation

But without the proper preparation

Having just a vision's no solution

Everything depends on execution

The art of making art, is putting it together

Stephen Sondheim

Putting It Together

Chapter 1

Clinical Trials

Clinical trials are experiments done on human beings to study and assess the effect of an intervention. The intervention could be a new drug, a combination of drugs, a medical procedure, or a medical device for human use. The following review of the nomenclature system and designs of clinical trials are mainly based on pharmaceutical trials evaluating drug therapy. However, the designs are easily generalized to non-drug trials. Throughout this thesis, terms such as "therapy", "treatment", and "drug" will be used interchangeably.

The primary objective of a clinical trial is to compare the effect of an experimental treatment with a control treatment. A control treatment could be a non-active intervention which means that the "treatment" is a placebo or due to ethical reasons the control treatment could be an "existing established effective treatment" (Fitzpatrick, 2005). However, in single-arm trials, the efficacy of the experimental treatment is compared with a known historical value which is usually based on the known efficacy of the standard treatment.

Clinical Trials

The clinical development of a new drug can be divided into four phases, i) human pharmacology study, ii) therapeutic exploratory study, iii) therapeutic confirmatory study, or iv) therapeutic use study (ICH, 1997). A common nonmenclature that is used for each type of the study is phase I, II, III and IV, respectively. The description of each phase is by no means restrictive. For example, although a phase II trial is usually meant to assess the therapeutic effect of a new treatment, it is not restricted to only such study. It may also look into the human pharmacology and/or confirmation of the efficacy of the new treatment.

Typically, a development plan of a new drug begins with a phase I trial where the new treatment is first tested on humans. In the phase I trials of most diseases, healthy volunteers are recruited to determine the level of tolerability for later trials. An exception is if the new therapy is highly toxic such as cytotoxic chemotherapy. Patients are recruited instead to these trials and usually these patients have already tried and failed on existing standard therapies. The main objective of a phase I trial is to estimate the maximum dose level that is acceptable for a participant or patient without causing unacceptable toxicity. This dose is conventionally known as the maximally tolerated dose (MTD).

Once the range of safe dose levels has been established, the new treatment is tried in a phase II trial. The primary objective of a phase II trial is to explore therapeutic effect in patients and, as such, its aim is to compare the efficacy of the drug with that of the control treatment formally. However, it is not necessary to have a control treatment in a phase II trial. Thus, in the uncontrolled-trial the efficacy of the new drug is compared against a known

value of the current standard or historical control. Usually a single group of patients is selected to the phase II trial and they are usually a homogeneous group in terms of disease and stage of disease (Gehan, 1961, Schoenfeld, 1980).

If the new treatment has shown some minimally acceptable clinical effect, it would be recommended for further testing in larger phase III trials. A phase III trial is a definitive clinical trial and is comparative in nature. It is a large confirmatory trial where the results are submitted to regulatory authorities for drug approval. Due to the large sample size required in a phase III trial, it is often conducted concurrently by many centres, ranging from tens to hundreds, and as such is sometimes known as a multicentre trial. The advantage of the multicentre trials is the possibility of wider patient population recruitment and a broad range of clinical settings that is more typical of future use.

Phase IV trials are usually undertaken after or during the registration of a drug to monitor and discover more about the safety of the drug for the approved indication. Sometimes, the trials also assess efficacy in different populations. The sample size is usually very large and the trial may not have a control arm.

In an ideal situation, the development of a new drug would go through a series of clinical trials sequentially through phase I to IV. This is because the results from the preceding phase are used to motivate the design of the next phase. In practice, the development plan may not go through the same sequence. It is rather common for the results from a phase II exploratory study to prompt additional human pharmacology studies or to modify the

strategy of drug administration or to lead to more studies to investigate the dose-response relationship. Or the results from a phase III trial may prompt another phase III trial by narrowing the disease population.

The majority of clinical trials aim to demonstrate the superiority of the efficacy of the new treatment against that of the placebo or standard treatment. There are other types of comparison, namely, equivalence and non-inferiority. A common example of clinical equivalence trials is the demonstration of the clinical equivalence of a generic product to the marketed product. A non-inferiority trial aims to show that the efficacy of the new drug is not clinically inferior to the standard treatment.

This thesis concentrates on the designs of superiority clinical trials where the efficacy of the new treatment is compared to that of a control treatment or a known value from historical controls. As such, it is assumed that the dosage and safety issues of the new drug or treatment regimen has been addressed in phase I trials. As an illustration of the applicability of the proposed designs, examples from asthma clinical researches are used. The following chapter briefly describes asthma, some of the class of standard treatments and common endpoints employed in asthma clinical trials.

Chapter 2

Asthma

The definition of asthma has evolved over time as understanding of the disease has become clearer. However, the pathogenesis of the disease is still unclear and therefore, the definition by the Global Initiative for Asthma (GINA) is based on the functional consequences of airway inflammation (GINA, 2010). In 2010 GINA issued an updated revision of an operational description of asthma:

> Asthma is a chronic inflammatory disorder of the airways in which many cells and cellular elements play a role. The chronic inflammation is associated with airway hyperresponsiveness that leads to recurrent episodes of wheezing, breathlessness, chest tightness, and coughing, particularly at night or in the early morning. These episodes are usually associated with widespread, but variable, airflow obstruction within the lung that is often reversible either spontaneously or with treatment.

There is still a lack of consensus on the definition of asthma and as a result it is difficult to compare the prevalence rate from different parts of the world. However, it is estimated that 300 million individuals are affected worldwide and based on the application of standardized methods in children and adults, the prevalence "ranges from 1% to 18% of the population in different countries" (GINA, 2010, p. 3). The prevalence rate in the United Kingdom is 6% (3 million) and of this, 10% are children (Corrigan, 2009).

Although there is no clear definition of the asthma phenotype, there is much clearer understanding of asthma clinical manifestations. Appropriate treatments can then be prescribed to control the condition effectively. Most of the existing treatments may be classified into a few classes; bronchodilators, corticosteroid, mediator antagonist, or immunomodulatory and they aim "to minimize symptoms, optimize lung function, and prevent exacerbations" (Reddel *et al.*, 2009).

2.1 Therapies for asthma

The mainstay of bronchodilator agents is the β_2-adrenoceptor agonists. The β_2-adrenoceptor agonists are further classified into short-acting β_2-adrenoceptor agonists (SABA) and long-acting β_2-adrenoceptor agonists (LABA). Some of the frequently used SABA are salbutamol (or albuterol in the United States), terbutaline and fenoterol, while LABA are formoterol and salmeterol. The β_2-adrenoceptor agonists are effective in reversing airflow obstruction as well as protecting asthmatic patients against bronchoconstrictor challenge (Hall, 2009).

Another commmonly prescribed family of treatments to control asthma symptoms is the corticosteroids. These work by switching off the multiple inflammatory genes that are turned on in the airways by proinflammatory transcription factors (Barnes, 2009). Inhaled corticosteroids are the most common asthma management and they suppress the mucosal inflammation relatively rapid in the asthmatic airways. Some of the common prescribed inhaled corticosteroids are budesonide, fluticasone propionate and beclomethasone dipropionate.

A large number of inflammatory mediator receptors are involved in the pathophysiology of asthma but so far only one class of mediator antagonists has become the established treatment, that is, anti-leukotrienes. Leukotriene receptor antagonists cause bronchodilation and have an additive effect to the SABA. Their effectiveness against placebo has been shown in short-term clinical studies of 4-6 week duration (Chung and Barnes, 2009). Some of the drugs under this class have been approved as a first-line treatment in the United States and they are zileuton, zafirlukast and montelukast whereas in Europe, only montelukast and zafirlukast have been approved as a second-line add-on therapy.

Studies have shown that T-lymphocytes may be involved in the initiation and maintenance of the inflammatory process of asthma (Corrigan, 2009). There have been subsequently clinical trials investigating T-cell immunomodulatory agents for possible therapeutic effects especially in patients with severe asthma. One particular drug from this class that has been approved in Europe as an add-on therapy for severe persistent asthma that is caused by an allergy is omalizumab. Omalizumab is a chimeric antibody

that blocks IgE and it is delivered subcutaneously.

Asthma is a heterogenous condition. Most patients are effectively treated with either corticosteroids or bronchodilators but some patients with severe asthma however, are poorly controlled even with maximal doses. There is also a minority of patients resistant to the anti-inflammatory drugs. Therefore, combined treatments are sometimes recommended for these asthma patients. Inhaled β_2-adrenoceptor agonists and corticosteroids are frequently recommended to be used together and studies have shown that these two classes of drugs have important molecular interactions (Barnes, 2009). Increasingly, drugs from other classes have gone on trials to be used as concomitant medications to the inhaled corticosteroids if the asthma is not controlled.

2.2 Outcome variables

Spirometry has been one of the most fundamental measurements of asthma control. It is objective and highly reproducible in measuring lung function. The most common spirometric measurement is the forced expiratory volume in one second (FEV_1). In 2009, the American Thoracic Society (ATS) and European Respiratory Society (ERS) issued an offical guideline that FEV_1 should be included as a primary endpoint for bronchodilator clinical trials (Reddel *et al.*, 2009). An improvement of a minimum of 10% from the baseline measurement is considered to be clinically meaningful.

Even prior to the 2009 guideline by ATS/ERS, FEV_1 has been the primary endpoint in most clinical trials because it measures the airflow limitation which is the primary manifestation of asthma. However, the prevention

of asthma exacerbation is arguably the most important clinical outcome. In the past 10 years some clinical trials have used exacerbation as the primary endpoint but the criteria and definitions used by different studies are quite varied.

The ATS/ERS task force has thus recommended that severe exacerbations be events that include at least one of the following: use of systemic corticosteroids or increase dosage from the maintenance dose, and a hospitalization or emergency department visit because of asthma and require systemic corticosteroids. The definition of moderate asthma exacerbations include at least one of the following: deterioration in symptoms, deterioration in lung function, increased rescue bronchodilator use, and emergency department visits because of asthma but does not require systemic corticosteroids.

The ATS/ERS task force also proposed to include moderate and severe asthma exacerbations as the important outcome for clinical trials in primary care. Prior to the issuance of the guideline, studies have reported the percentage of patients with at least one exacerbation, the time to first severe exacerbation, or the rate of exacerbations. The time to first exacerbation is favoured as the effect of the experimental therapy may be examined before other rescue or add-on treatments are introduced. The rate of exacerbations is advantageous especially for comparing between patient populations.

There are other measurable outcome variables used to assess asthma control. Asthma symptoms are highly variable between patients and the use of diaries in clinical trials is thus very useful. Some of the questions asked in diaries are symptoms, adverse events, reliever use, interference to nor-

mal activities and health care utilization. Peak expiratory flow is another spirometric measurement but is considered to be inferior to FEV_1 as a measurement for airways obstruction.

Two main outcomes are considered for the illustrations of the applicability of the proposed designs in this thesis and they are FEV_1 and asthma exacerbation, the former as an example of a continuous variable and the latter as a binary variable where a patient with at least one episode of exacerbation is considered to be a "failure" whereas a "success", otherwise.

For clinical trials, it is important to measure asthma control in a predefined time point. Reddel *et al.* (2009, p. 66) pointed out that "by longstanding consensus, clinical asthma control is usually assessed over periods of 1 to 4 weeks". Following this, the proposed clinical trial designs in this thesis thus assume that the primary outcome is obtained within 4 weeks.

Chapter 3

Statistical Background

The statistical terms, notation and general frameworks used in this thesis are presented in this chapter. The topics are usually found under different chapters in standard statistical textbook but now are placed in one chapter under different sections. Some of the discrete and continuous distributions are discussed in Section 3.1. A general review of Bayes' theorem and frequentist hypothesis testing are given in Sections 3.2 and 3.3, respectively. An alternative to the computation of the power of a study is discussed in Section 3.4. A general decision theory methodology is given in Section 3.5 and finally, Section 3.6 concludes the chapter with a summary.

3.1 Distribution functions

3.1.1 Discrete random variables

Bernoulli random variables

Let X be a Bernoulli random variable. It can only take two values: 1 and 0. If the probability that $\Pr(X = 1) = p$ and $\Pr(X = 0) = 1-p$, the probability mass function may be represented as,

$$f(x) = p^x(1-p)^{1-x}, x = 0, 1$$

Drawing from the motivation of the asthma research, if a patient is able to achieve the targeted asthma control within the first four weeks, then the observed response is considered as a success. This may be represented numerically as $X = 1$. However, if a patient fails to achieve the targeted asthma control within the first four weeks, then the response is considered as a failure and numerically represented as $X = 0$.

The binomial distribution

Suppose that there are n independent patients in the trial and each patient's response is recorded either as a success or a failure with p probability of success, that is, a Bernoulli variable with parameter p. Assuming that all the binary responses are identical Bernoulli random variable with the same parameter p, let the accumulated number of successes be denoted by X. The variable X is said to be a binomial random variable with index n and

parameter p. As the sequence of the occurrence of successes is not important there are $\binom{n}{x}$ ways in which a total number of x successes may occur from the n patients. Thus, the probability mass function is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \tag{3.1}$$

for $x = 0, 1, \ldots, n$, and the parameter $0 < p < 1$. Its cumulative distribution function is

$$F(x) = \sum_{i=0}^{x} \binom{n}{i} p^i (1-p)^{n-i}.$$

The expected value of X, denoted by $E(X)$, is

$$E(X) = \sum_{x=0}^{n} x f(x) = np,$$

and the variance of X, denoted by $\text{var}(X)$, is

$$\text{var}(X) = E[(x - E(X))^2] = \sum_{x=0}^{n} (x - E(X))^2 f(x) = np(1-p).$$

The statement of X following a binomial distribution with (n, p) can be "rewritten" as $X \sim Bin(n, p)$.

The geometric distribution

The geometric distribution is another discrete distribution that is constructed from independent Bernoulli variables. The difference between the binomial and geometric distribution is that in the geometric distribution there is an unlimited number of patients in a trial. Suppose that in a trial, a sequence of

patients is recruited and the trial will stop when the first success is observed.

Let X be the total number of patients including the first successful outcome and let the probability of a success be p. Following from the independence of each patient, the probability mass function is

$$f(x) = (1-p)^{x-1}p, \quad x = 1, 2, \ldots. \tag{3.2}$$

The expected value of a geometric random variable is $E(X) = 1/p$ and the variance is $\text{var}(X) = (1-p)/p^2$. The statement of X following a geometric distribution with parameter p can be written as $X \sim Ge(p)$.

3.1.2 Continuous random variables

The normal distribution

In another example from the asthma research, spirometric measurements such as forced expiratory volume in one second (FEV_1) are used to assess the lung function and airway hyperresponsiveness. The outcome of interest is usually the difference between the measurement at baseline and after the administration of treatment. It may be measured either in the absolute unit (ml) or percentage. Let X be the difference of the FEV_1 in percentage and it is a continuous random variable that is often assumed to follow a normal distribution. The normal distribution which is also known as the Gaussian distribution is the most important continuous distribution and "plays a central role in probability and statistics" (Rice, 1995, p. 53). The probability

density function of a normal distribution is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\theta)^2/2\sigma^2}, \tag{3.3}$$

for $-\infty < x < \infty$. The probability density function depends on two parameters θ and σ where $-\infty < \theta < \infty$ and $\sigma > 0$. The cumulative distribution function is

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(u-\theta)^2/2\sigma^2}\, du.$$

The density of the normal distribution integrates to 1 in the whole space of $(-\infty, \infty)$. However, the cumulative distribution function cannot be evaluated in a closed form but has to be computed numerically. The expected value is

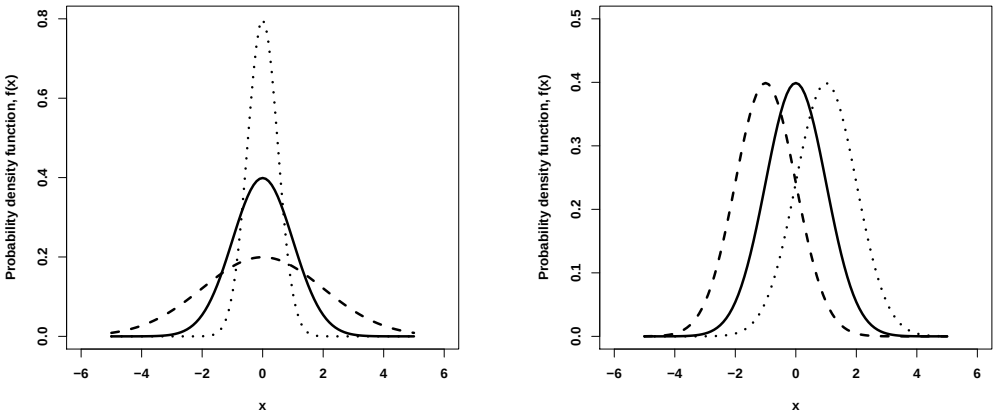
$$E(X) = \int_{-\infty}^{-\infty} x f(x)\, dx = \theta,$$

and the variance is

$$\mathrm{var}(X) = \int_{-\infty}^{-\infty} (x-\theta)^2 f(x)\, dx = \sigma^2.$$

For convenience, the statement that the random variable X follows a normal distribution with mean θ and variance σ^2 is written as $X \sim N(\theta, \sigma^2)$.

A special case of the normal distribution is the standard normal distribution where $\theta = 0$ and $\sigma^2 = 1$. Its density function is usually denoted by $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ and its cumulative distribution function is denoted by $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du$. The relationship between a normal and standard normal distribution can be stated by: $f(x) = \frac{1}{\sigma}\phi\left(\frac{x-\theta}{\sigma}\right)$ and $F(x) = \Phi(\frac{x-\theta}{\sigma})$.



(a) $\theta = 0$ (b) $\sigma = 1$

Figure 3.1: Normal densities (a) σ of 0.5 (dotted), 1 (solid), and 2 (dashed), and (b) θ of -1 (dashed), 0 (solid), and 1 (dotted).

The normal distribution when plotted in a plane of $f(x)$ against x has a bell-shaped curve (Fig. 3.1). It is symmetric about its mean, θ, and the shape of the curve, either narrow or wide, depends on the standard deviation, σ.

The beta distribution

The beta distribution is a distribution that has very flexible shapes with two parameters a and b, shown in Figure 3.2, from flat to narrow curves. Let X be the random variable that follows a beta distribution with non-negative parameters a and b, $X \sim Beta(a, b)$. Its probability density function is given by

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, \tag{3.4}$$

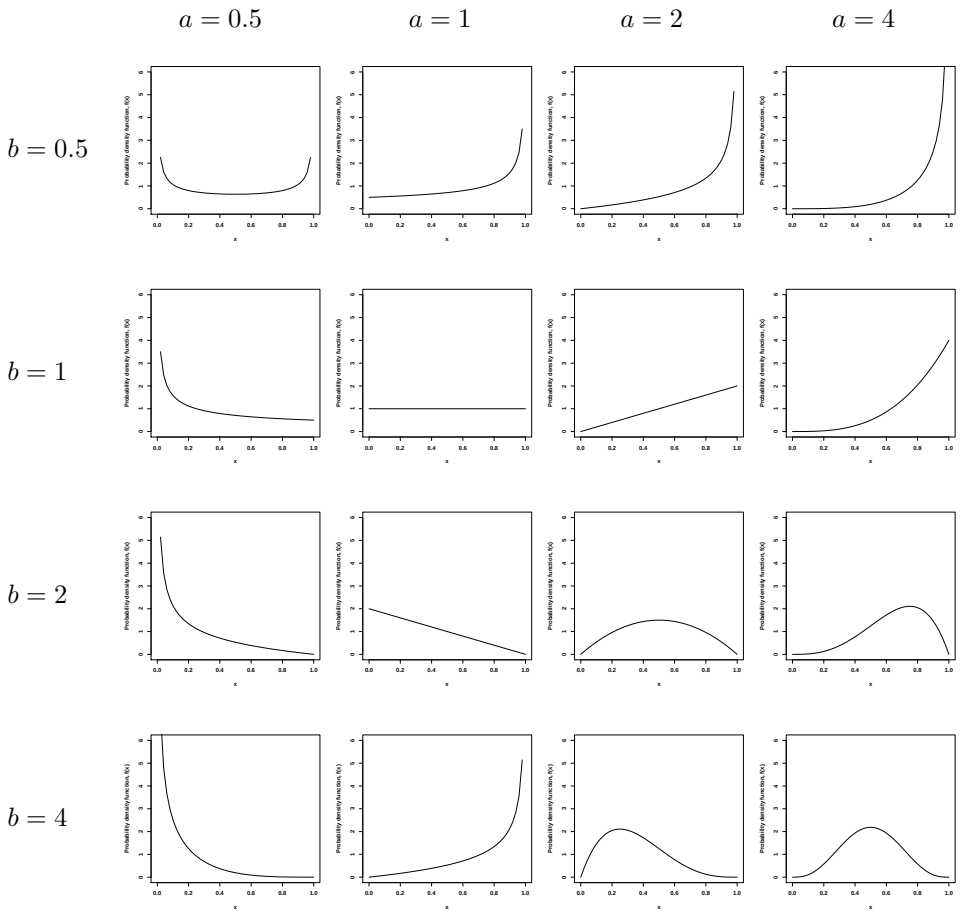


Figure 3.2: Beta densities with various values of a and b.

for $0 < x < 1$ where the beta function is defined as

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}.$$

The gamma function is defined as $\Gamma(a) = \int_0^\infty u^{a-1} e^{-u}\, du$ if a is a non-integer. If a is an integer the gamma function is a simple factorial function, $\Gamma(a) = (a - 1)!$.

The beta distribution belongs to the natural exponential family of distri-

butions. A k-parameter exponential family density can be written as

$$f(x,\theta) = r(x)\eta(\theta)\exp\left\{\sum_{i=1}^{k}\theta_i p_i(x)\right\}.$$

The beta distribution thus can be shown is a two-parameter exponential family with $r(x) = 1$, $\eta(\theta) = \Gamma(a+b)/(\Gamma(a)\Gamma(b))$, $\theta = (a-1, b-1)'$ and $p(x) = (\log(x), \log(1-x))'$ where $\log(\cdot)$ is the natural logarithm,

$$\begin{aligned} f(x,\theta) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\exp\Big\{(a-1)\log(x) + (b-1)\log(1-x)\Big\} \\ &= \frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}, \end{aligned}$$

The expected value of a beta random variable is $E(X) = a/(a+b)$, and the variance is $\text{var}(X) = ab/[(a+b)^2(a+b+1)]$.

3.1.3 Joint distribution

The Sarmanov distribution

One lesser known family of bivariate distribution is the family introduced by Sarmanov which appeared in Doklady (Soviet Mathematics) in 1966 (Lee, 1996). Define X_1 and X_2 as the random variables and $f_1(x_1)$ and $f_2(x_2)$ as the marginal probability density functions of X_1 and X_2, respectively. Let μ_i be the mean of X_i and σ_i the standard deviation for $i = 1, 2$. The general function of the joint density function is

$$h(x_1, x_2) = f_1(x_1)f_2(x_2)\Big(1 + \omega\phi_1(x_1)\phi_2(x_2)\Big),$$

where ω is a real number that satisfies the condition $1 + \omega\phi_1(x_1)\phi_2(x_2) \geq 0$ and $\phi_i(x_i)$ is a nonconstant mixing function bounded by $\int_{-\infty}^{\infty} \phi_i(x_i) f_i(x_i)\, dx_i = 0$ $(i = 1, 2)$ for all values of x_1 and x_2.

The correlation coefficient of X_1 and X_2 is given by $\rho = \omega\sigma_1\sigma_2$ where ω satisfies the condition

$$\max\left\{\frac{-1}{\mu_1\mu_2}, \frac{-1}{(1-\mu_1)(1-\mu_2)}\right\} \leq \omega \leq \min\left\{\frac{1}{\mu_1(1-\mu_2)}, \frac{1}{\mu_2(1-\mu_1)}\right\}.$$

Therefore, X_1 and X_2 are positively correlated if $\omega > 0$; negatively correlated if $\omega < 0$; and independent if $\omega = 0$.

In her paper, Lee (1996) discussed the properties and applications of the Sarmanov's family of bivariate distribution. Of relevance to this thesis is the case where the marginals of the bivariate distribution follow the beta distributions. Let X_i be a random variable that follows a beta distribution with parameters a_i and b_i, that is, $X_i \sim Beta(a_i, b_i)$, for $i = 1, 2$. Since the sample space of the random variable is contained in $[0, 1]$, Lee proposed the mixing function to be

$$\phi_i(u_i) = x_i - \mu_i,$$

where $\mu_i = a_i/(a_i + b_i)$. Therefore, the bivariate density of X_1 and X_2 is

$$h(x_1, x_2) = f_1(x_1) f_2(x_2)\left(1 + \omega\Big(x_1 - \frac{a_1}{a_1 + b_1}\Big)\Big(x_2 - \frac{a_2}{a_2 + b_2}\Big)\right), \qquad (3.5)$$

where $f_i(x_i)$ is the beta density. Note that the bivariate density is a linear combination of products of independent beta densities and ω is within the

range

$$\begin{aligned} \max\left\{\frac{-(a_1+b_1)(a_2+b_2)}{a_1a_2}, \frac{-(a_1+b_1)(a_2+b_2)}{b_1b_2}\right\} &\leq \omega \\ &\leq \min\left\{\frac{(a_1+b_1)(a_2+b_2)}{a_1b_2}, \frac{(a_1+b_1)(a_2+b_2)}{a_2b_1}\right\} \\ \Leftrightarrow \qquad \frac{-(a_1+b_1)(a_2+b_2)}{\max\{a_1a_2, b_1b_2\}} &\leq \omega \leq \frac{(a_1+b_1)(a_2+b_2)}{\max\{a_1b_2, a_2b_1\}}. \end{aligned}$$

In the same paper, Lee extended the family of Sarmanov's bivariate distribution to the multivariate case. Let X_i be a random variable with marginal density function f_i for $i = 1, 2, \ldots, k$ then the k-variate joint density is

$$h(x_1, \ldots, x_k) = \Big(\prod_{i=1}^{k} f_i(x_i)\Big)\Big(1 + R_{\Omega_k}(x_1, \ldots, x_k)\Big), \tag{3.6}$$

where

$$\begin{aligned} R_{\Omega_k}(x_1, x_2, \ldots, x_k) = &\sum_{i_1=1}^{k-1}\sum_{i_2=i_1+1}^{k} \omega_{i_1,i_2}\phi(x_{i_1})\phi(x_{i_2}) \\ &+ \sum_{i_1=1}^{k-2}\sum_{i_2=i_1+1}^{k-1}\sum_{i_3=i_2+1}^{k} \omega_{i_1,i_2,i_3}\phi(x_{i_1})\phi(x_{i_2})\phi(x_{i_3}) \\ &+ \cdots + \omega_{1,2,\ldots,k}\prod_{i=1}^{k}\phi(x_i), \end{aligned}$$

and $\Omega_k = \{\omega_{i_1,i_2}, \omega_{i_1,i_2,i_3}, \ldots, \omega_{1,2,\ldots,k}\}$ is a set of real numbers satisfying the condition $1 + R_{\Omega_k}(x_1, x_2, \ldots, x_k) \geq 0$. If all of the values of ω's (each element in Ω_k) are zero, then all the k variables are independent.

3.2 Bayes' theorem

Suppose that a parameter θ does not have some fixed value but is random. Its probable value is quantified in a probablity density function known as prior density, $f_\Theta(\theta)$. The random variable X depends on the unknown parameter θ in a known way and having observed some data $X = x$ the dependency is expressed by a density function, $f_{X|\Theta}(x|\theta)$, which is known as the likelihood function. The new opinion of the parameter θ is updated and by the Bayes' theorem it is

$$f_{\Theta|X}(\theta|x) = \frac{h(x,\theta)}{f_X(x)} = \frac{f_{X|\Theta}(x|\theta)f_\Theta(\theta)}{f_X(x)}, \tag{3.7}$$

where $h(x,\theta)$ is the joint density of X and θ. The function $f_{\Theta|X}(\theta|x)$ is called the posterior density. The marginal density of X is obtained by integrating $h(x,\theta)$ over the sample space of θ,

$$f_X(x) = \int h(x,\theta)\,d\theta = \int f_{X|\Theta}(x|\theta)f_\Theta(\theta)\,d\theta. \tag{3.8}$$

The marginal distribution of X is also called the predictive distribution.

3.2.1 Normal mean

The Bayes' theorem is one of the fundamental tools in the Bayesian analysis. Following are two illustrations of using Bayes' theorem to infer the unknown parameter of a random variable that follows a known distribution. Let X be a random variable that follows a normal distribution with unknown mean θ and known variance σ^2. The parameter θ is assumed to be random and to

also follow a normal distribution with known mean μ and variance τ^2. The prior density of θ is

$$f_\Theta(\theta) = \frac{1}{\tau}\phi\Big(\frac{\theta-\mu}{\tau}\Big) = \frac{1}{\sqrt{2\pi\tau^2}}\exp\Big\{-\frac{(\theta-\mu)^2}{2\tau^2}\Big\},$$

and the likelihood function of X is

$$f_{X|\Theta}(x|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\Big\{-\frac{(x-\theta)^2}{2\sigma^2}\Big\}.$$

The marginal density of X is thus,

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{X|\Theta}(x|\theta) f_\Theta(\theta)\, d\theta \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\tau\sigma}\exp\Big\{-\frac{1}{2\tau^2\sigma^2}(\tau^2(x-\theta)^2+\sigma^2(\theta-\mu)^2)\Big\}\, d\theta \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\tau\sigma}\exp\Big\{-\frac{1}{2\tau^2\sigma^2}\Big[\Big(\theta-\frac{\tau^2x+\sigma^2\mu}{\tau^2+\sigma^2}\Big)^2(\tau^2+\sigma^2) \\
&\quad + \frac{\tau^2\sigma^2}{\tau^2+\sigma^2}(x-\mu)^2\Big]\Big\}\, d\theta \\
&= \frac{1}{\sqrt{2\pi\tau^2\sigma^2}}\exp\Big\{-\frac{1}{2(\tau^2+\sigma^2)}(x-\mu)^2\Big\} \\
&\quad \times \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\Big\{-\frac{\tau^2+\sigma^2}{2\tau^2\sigma^2}\Big(\theta-\frac{\tau^2x+\sigma^2\mu}{\tau^2+\sigma^2}\Big)^2\Big\}\, d\theta \\
&= \frac{1}{\sqrt{2\pi\tau^2\sigma^2}}\exp\Big\{-\frac{1}{2(\tau^2+\sigma^2)}(x-\mu)^2\Big\}\cdot\sqrt{\frac{\tau^2\sigma^2}{\tau^2+\sigma^2}} \\
&= \frac{1}{\sqrt{2\pi(\tau^2+\sigma^2)}}\exp\Big\{-\frac{1}{2(\tau^2+\sigma^2)}(x-\mu)^2\Big\}. \qquad (3.9)
\end{aligned}$$

The marginal distribution of the random variable X has the form of a normal distribution with mean μ and variance $(\tau^2+\sigma^2)$. Upon observing data x, the prior belief of the parameter θ can be updated. For ease of notation

let $\lambda = (\tau^2 x + \sigma^2\mu)/(\tau^2+\sigma^2)$ and $\nu = (\tau^2\sigma^2)/(\tau^2+\sigma^2)$, the posterior density of θ is

$$\begin{aligned} f_{\Theta|X}(\theta|x) &= \frac{h(x,\theta)}{f_X(x)} \\ &= \frac{\exp\Big\{ -\Big((\theta-\lambda)^2(\tau^2+\sigma^2) + \nu(x-\mu)^2\Big)/(2\tau^2\sigma^2)\Big\}/(2\pi\tau\sigma)}{\exp\Big\{ -(x-\mu)^2/(2(\tau^2+\sigma^2))\Big\}/\sqrt{2\pi(\tau^2+\sigma^2)}} \\ &= \frac{1}{\sqrt{2\pi\nu}}\exp\Big\{ -\frac{1}{2\nu}(\theta-\lambda)^2\Big\}. \end{aligned} \tag{3.10}$$

The posterior distribution of the random parameter θ is also a normal distribution but with mean λ and variance ν.

3.2.2 Binomial distribution

For the second illustration, let X be a discrete random variable and assume that it follows a binomial distribution such that $X|p \sim Bin(n,p)$. The parameter p is assumed to be random and as $0 < p < 1$ a convenient choice for the prior distribution is a beta distribution. Assume that $p \sim Beta(a,b)$ with known parameters a and b. From equations (3.1) and (3.4) the marginal density of X is therefore,

$$\begin{aligned} f_X(x|a,b) &= \int_0^1 f_{X|p}(x|p) f_p(p)\, dp \\ &= \binom{n}{x}\frac{1}{B(a,b)}\int_0^1 p^{a+x-1}(1-p)^{b+n-x-1}\, dp \\ &= \binom{n}{x}\frac{B(a+x,b+n-x)}{B(a,b)}. \end{aligned} \tag{3.11}$$

The marginal distribution of X is known as the beta-binomial distribution with index n, and parameters a, and b.

The posterior density of p given data x is

$$\begin{aligned} f_{p|X}(p|x) &= \frac{f_{X|p}(x|p)f_p(p)}{f_X(x)} \\ &= \frac{1}{B(a+x, b+n-x)} p^{a+x-1}(1-p)^{b+n-x-1}, \end{aligned} \tag{3.12}$$

which has the same form as a beta distribution. The posterior distribution of p given x is thus a beta distribution with parameters $(a+x, b+n-x)$.

In both examples, it is shown that if X is normal with parameter θ which prior distribution is also normal, its posterior distribution is likewise a normal distribution but with different parameters. Similarly, if X is binomial with parameter p and if it follows a beta distribution, its posterior distribution is also a beta distribution with different parameters. In general, if $L(\theta; \boldsymbol{x})$ is a likelihood function and f_Θ is a prior distribution belongs to a family of G where the posterior density

$$f_{\Theta|\boldsymbol{x}}(\theta|\boldsymbol{x}) \propto f_\Theta(\theta) L(\theta; \boldsymbol{x}),$$

also belongs to the family G, then G is said to be a family of conjugate priors for all $\boldsymbol{x}$. Due to the "nice" form of the conjugate priors, it is therefore, a mathematical convenience to choose a prior distribution from a conjugate family so that the posterior can be evaluated easily.

3.2.3 Sarmanov distribution

The next example is based on the k-variate joint distribution from the Sarmanov's family. Let X_i be a random variable that follows a binomial distribution with index n_i and an unknown parameter p_i $(i = 1, 2, \ldots, k)$. Suppose that the likelihood functions of $X_1, X_2, \ldots, X_k$ are independent from each other, then the joint conditional density, denoted by $h_{\boldsymbol{X}|\boldsymbol{p}}(x_1, \ldots, x_k|p_1, \ldots, p_k)$, is the product of all the likelihood functions,

$$h_{\boldsymbol{X}|\boldsymbol{p}}(x_1, \ldots, x_k|p_1, \ldots, p_k) = \prod_{i=1}^{k} f_{X|p}(x_i|p_i) = f_{X|p}(x_1|p_1) \ldots f_{X|p}(x_k|p_k). \tag{3.13}$$

Suppose that p_i is a random variable and has a beta distribution with known parameters a_i and b_i, and let the joint distribution of $p_1, p_2, \ldots, p_k$ follows the k-variate Sarmanov's family as seen in (3.6). Denote $h_{\boldsymbol{p}}(p_1, \ldots, p_k)$ as the joint density of p_i's and let the mixing function be $\phi_i(p_i) = x_i - \mu_i$ where $\mu_i = a_i/(a_i + b_i)$ is the expected value of the beta distribution. Therefore, the unconditional joint density of $X_1, X_2 \ldots, X_k$, denoted by $h_{\boldsymbol{X}}(x_1, \ldots, x_k)$, is

$$h_{\boldsymbol{X}}(x_1, \ldots, x_k) = \int \cdots \int \prod_{i=1}^{k} f_{X|p}(x_i|p_i) h(p_1, \ldots, p_k)\, dp_1 \cdots dp_k,$$

an iterated integral. The detailed working of the integration by parts is shown in Appendix A. From equation (A.8), the unconditional joint density of $X_1, \ldots, X_k$ is

$$h_{\boldsymbol{X}}(x_1, \ldots, x_k) = \Big(\prod_{i=1}^{k} f_X(x_i) \Big)(1 + D_{\Omega_k}(x_1, \ldots, x_k)), \tag{3.14}$$

where $f_X(x_i) = \binom{n_i}{x_i} Beta(a_i + x_i, b_i + n_i - x_i)/Beta(a_i, b_i)$ is the marginal density of X_i and

$$\begin{aligned} D_{\Omega_k}(x_1, \ldots, x_k) = &\sum_{i_1=1}^{k-1} \sum_{i_2=i_1+1}^{k} \omega_{i_1,i_2} \psi(x_{i_1}) \psi(x_{i_2}) \\ &+ \sum_{i_1=1}^{k-2} \sum_{i_2=i_1+1}^{k-1} \sum_{i_3=i_2+1}^{k} \omega_{i_1,i_2,i_3} \psi(x_{i_1}) \psi(x_{i_2}) \psi(x_{i_3}) \\ &+ \cdots + \omega_{1,2,\ldots,k} \prod_{i=1}^{k} \psi(x_i), \end{aligned}$$

where the function ψ is defined as $\psi(x_i) = (x_i - \mu_i n_i)/(a_i + b_i + n_i)$.

From equations (3.6), (3.13), and (3.14) the joint posterior density is

$$\begin{aligned} &h_{\boldsymbol{p}|\boldsymbol{X}}(p_1, \ldots, p_k | x_1, \ldots, x_k) \\ &\quad = \frac{h_{\boldsymbol{X}|\boldsymbol{p}}(x_1, \ldots, x_k | p_1, \ldots, p_k) h_{\boldsymbol{p}}(p_1, \ldots, p_k)}{h_{\boldsymbol{X}}(x_1, \ldots, x_k)} \\ &\quad = \Big(\prod_{i=1}^{k} \frac{f_{X|p}(x_i|p_i) f_p(p_i)}{f_X(x_i)} \Big) \left(\frac{1 + R_{\Omega_k}(p_1, \ldots, p_k)}{1 + D_{\Omega_k}(x_1, \ldots, x_k)} \right) \\ &\quad = \frac{1 + R_{\Omega_k}(p_1, \ldots, p_k)}{1 + D_{\Omega_k}(x_1, \ldots, x_k)} \prod_{i=1}^{k} f_{p|X}(p_i|x_i) \end{aligned} \tag{3.15}$$

The posterior is a linear combination of products of the posterior beta densities and it is known as pseudo-conjugate. Although it does not have as convenient a form as its prior, it is still relatively easy to compute the posterior density numerically.

3.3 Hypothesis testing and power function

The theory of statistical inference is broadly divided into two branches, namely, estimation and hypothesis testing. It is the latter branch that is discussed in this section. Due to the inherent variability in observing an outcome in each situation, a probability distribution is used to describe the variability. However, the true probability distribution is also unknown to us. The inference problem is thus to infer something of the true distribution or the true parameter. Observations from a certain sample space are more likely to belong to some known distributions, for example, continuous variables may follow the normal distribution or the beta distribution and discrete random variable may tend to follow the binomial distribution. Therefore, for this thesis, it is assumed that the inherent variability of observations are adequately explained by a known probability distribution. The inference problem is then to make use of the observed outcomes to estimate the true parameter.

We generally wish to test a statement that the true parameter θ belongs to a subset of the parameter space Θ. This statement is known as a hypothesis. The testing of the hypothesis is to use statistical methods to check if the observations are consistent with the stated hypothesis or not. A statistical rule is used to assign "each possible observation to one of two exclusive categories: 'consistent with the hypothesis under consideration' and 'not consistent with this hypothesis"' (Silvey, 1975, pp. 95).

In the classical approach of hypothesis testing which is also known as the frequentist method, there are two hypotheses. The first is the null hypothesis

which states that the parameter θ belongs to ω which is a subset of Θ. The other hypothesis is simply known as the alternative hypothesis which states that the parameter θ does not belong to the subset ω but belongs to $\Theta - \omega$. If there is only one element in ω, the hypothesis is known as a simple null hypothesis because it is in its simplest form, and similarly, if there is only one element in $\Theta - \omega$ the alternative hypothesis is a simple alternative hypothesis. Suppose that the elements in ω and $\Theta - \omega$ are θ_0 and θ_A, respectively, the hypotheses can be formulated as

$$H_0 : \theta = \theta_0 \quad \text{against} \quad H_1 : \theta = \theta_A,$$

where the statement H_0 is the null hypothesis and H_1 is the alternative hypothesis. The null hypothesis is always assumed to be true until proven to be otherwise. The statistical rule to reject H_0 is called a statistical test.

Two possible decisions can be made based on the observed data at the end of the trial: (1) reject the null hypothesis, or (2) do not reject the null hypothesis. Inevitably, errors may occur when rejecting or not rejecting H_0. The type I error is an error incurred when the null hypothesis is rejected when it is true. Another type of error that can be incurred is the type II error. It is an error incurred when the null hypothesis is accepted when it is false. The probability of incurring the type I error is usually capped at a predetermined value α such that,

$$\Pr(\text{Reject } H_0 | H_0 \text{ is true}) \leq \alpha,$$

and similary, the probability of incurring the type II error is capped by a predetermined value β,

$$\Pr(\text{Non-rejection of } H_0 | H_1 \text{ is true}) \leq \beta.$$

The probability that the null hypothesis is rejected when it is false is called the power of the test and it is simply $1-\beta$. Although the choice of α could be arbitrary, it is customary to have α at small values such as 0.1, 0.05, or 0.01. Similarly, the customary values of β are 0.2, 0.1, or 0.05. Correspondingly, the power of the test is 0.8, 0.9, or 0.95, respectively.

If $X_1, X_2, \ldots, X_n$ are n independent continuous random variables and each is normally distributed with unknown mean θ and known variance σ^2, let $\overline{X} = \sum_{i=1}^{n} X_i/n$, then $\overline{X} \sim N(\theta, \sigma^2/n)$. Let $(x_1, x_2, \ldots, x_n)$ be the sample of $X_1, X_2, \ldots, X_n$. It is desired to test whether the true mean is equal to some constants θ_0 or θ_A. The simple hypotheses are

$$H_0 : \theta = \theta_0 \quad \text{against} \quad H_1 : \theta = \theta_A.$$

The decision to either reject H_0 or not is made on the basis of the test statistic upon observing the responses at the end of the trial. The test statistic is most powerful if $\overline{X} > c$ where c is some constant such that the size of the test is α,

$$\Pr(\overline{X} > c) = \alpha.$$

Under the null hypothesis, the distribution of $\overline{X}$ is normal with mean θ_0 and variance σ^2/n. Let $Z = \sqrt{n}(\overline{X} - \theta)/\sigma$, then Z has a standard normal

distribution, $Z \sim N(0,1)$. Solving for c under the null hypothesis,

$$\begin{aligned}
& \Pr(\text{Reject } H_0|\theta=\theta_0)=\alpha \\
\Leftrightarrow \quad & \Pr(\overline{X}>c|\theta=\theta_0)=\alpha \\
\Leftrightarrow \quad & \Pr\left(\frac{\overline{X}-\theta}{\sqrt{\sigma^2/n}}>\frac{c-\theta}{\sqrt{\sigma^2/n}}\middle|\theta=\theta_0\right)=\alpha \\
\Leftrightarrow \quad & \Pr\left(\frac{\overline{X}-\theta_0}{\sqrt{\sigma^2/n}}>\frac{c-\theta_0}{\sqrt{\sigma^2/n}}\right)=\alpha \\
\Leftrightarrow \quad & \Pr\left(Z>\frac{c-\theta_0}{\sqrt{\sigma^2/n}}\right)=\alpha \\
\Leftrightarrow \quad & 1-\Phi\left(\frac{c-\theta_0}{\sqrt{\sigma^2/n}}\right)=\alpha \\
\Leftrightarrow \quad & c=z_{1-\alpha}\sqrt{\sigma^2/n}+\theta_0 \qquad (3.16)
\end{aligned}$$

where z_γ is the lower 100γ percentile of the standard normal distribution.

The computation of the power on the other hand is important when designing a trial. The power calculation is one of the standard statistical methods in determining sample size. Under the alternative hypothesis, $\overline{X} \sim N(\theta_A, \sigma^2/n)$ and from equation (3.16),

$$\begin{aligned}
\text{Power} &= \Pr(\text{Reject } H_0|\theta=\theta_A) \\
1-\beta &= \Pr\left(\frac{\overline{X}-\theta}{\sqrt{\sigma^2/n}}>\frac{c-\theta}{\sqrt{\sigma^2/n}}\middle|\theta=\theta_A\right) \\
&= \Pr\left(Z>\frac{z_{1-\alpha}\sqrt{\sigma^2/n}+\theta_0-\theta_A}{\sqrt{\sigma^2/n}}\right) \\
&= 1-\Phi\left(z_{1-\alpha}-\left(\frac{\theta_A-\theta_0}{\sqrt{\sigma^2/n}}\right)\right). \qquad (3.17)
\end{aligned}$$

The equation (3.17) shows that if θ_A is fixed, then the expression is a function

of n and if n is fixed, it can be evaluated as a function of θ_A.

As an example, a new bronchodilator for asthma control is ready to be put on clinical trials. The primary endpoint is the mean percentage change in FEV_1 from baseline. Let θ be the difference of mean change between the new bronchodilator and a placebo. The new bronchodilator is considered to be effective in controlling asthma if the mean change difference is at least 10 percent, that is, let $\theta_0 = 0$ (there is no difference in mean change between the new bronchodilator and placebo) and $\theta_A = 10$. Assume that the population standard deviation is known and fixed at $\sigma = 14$, and the size of the hypothesis test is $\alpha = 0.05$. The power function which for a fixed θ_A is a function of n is shown in Figure 3.3. According to the figure, in order to achieve a power of at least 90%, the sample size has to be at least 17.

3.4 Assurance

In 2001, O'Hagan and Stevens presented an alternative to the power calculation in order to determine a sample size in the context of cost-effectiveness (O'Hagan and Stevens, 2001). The methodology has subsequently been adopted by other researchers and it is easily applicable in the context of demonstrating treatment efficacy. In the frequentist approach, as shown in the preceding section, the parameter is assumed to be a fixed value so that a specific power can be achieved when the true parameter is not equal to θ_0. However, under the Bayesian approach, the true parameter does not need to assume a fixed value under H_1. Let the true parameter θ be random and be represented by a distribution which consequently implies a prior distribution

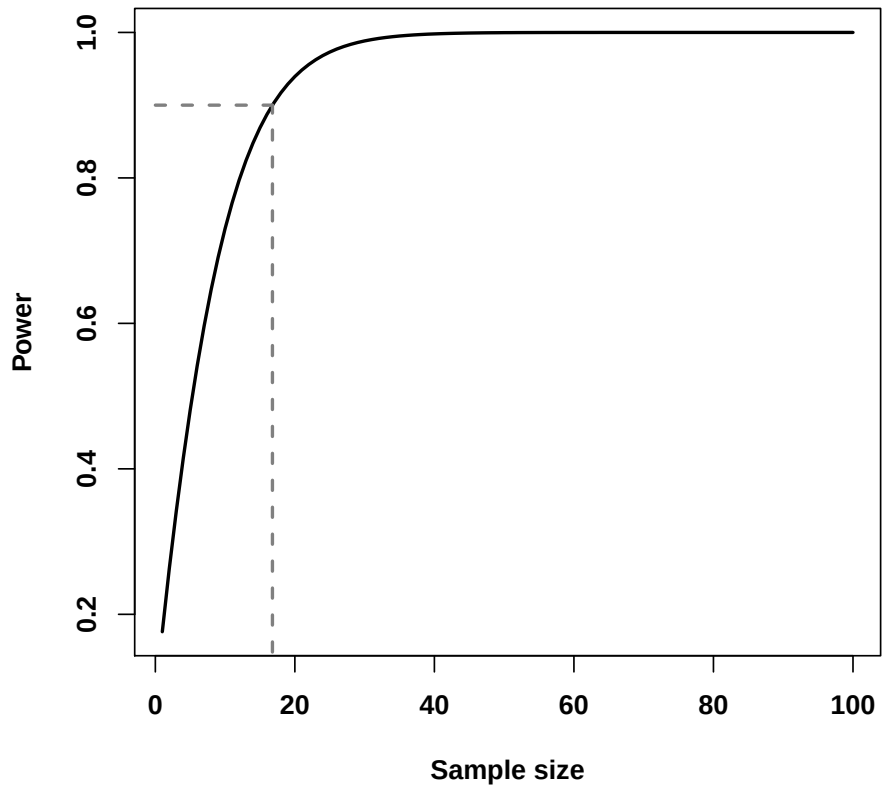


Figure 3.3: Power as a function of n.

for θ. Therefore, the power is now a kind of average power and is called assurance. Denote the assurance as A and from equation (3.17) the assurance is given by

$$A = E(\text{Power}) = \int_{\Theta} \left[1 - \Phi\left(z_{1-\alpha} - \left(\frac{\theta - \theta_0}{\sqrt{\sigma^2/n}}\right)\right)\right] f_{\Theta}(\theta)\, d\theta, \tag{3.18}$$

which is an integral over the whole parameter space.

3.4.1 Normal distribution

As an illustration, consider a random variable $\overline{X}$ whose likelihood is $\overline{X}|\theta \sim N(\theta, \sigma^2/n)$ where the variance σ^2/n is known and θ is a random parameter such that $\theta \sim N(\mu, \tau^2)$ where μ and τ are known. Similar to the workings in equation (3.9), the marginal distribution of $\overline{X}$ can be shown to be normal with mean μ and variance $(\tau^2 + \sigma^2/n)$. Thus, the assurance is

$$\begin{aligned} A &= \Pr(\overline{X} > z_{1-\alpha}\sqrt{\sigma^2/n} + \theta_0) \\ &= \Pr\Big(Z > \frac{z_{1-\alpha}\sqrt{\sigma^2/n} + \theta_0 - \mu}{\tau^2 + \sigma^2/n}\Big) \\ &= 1 - \Phi\Big(\frac{z_{1-\alpha} - \sqrt{n/\sigma^2}(\mu - \theta_0)}{\sqrt{1 + n\tau^2/\sigma^2}}\Big), \end{aligned} \tag{3.19}$$

which is a function of n.

3.4.2 Binomial distribution

For another illustration, let X_1 and X_2 be binary random variables and $X_1 \sim Bin(n_1, p_1)$ and $X_2 \sim Bin(n_2, p_2)$. Under the frequentist setting, it is desired to test the hypothesis,

$$H_0 : p_1 = p_2 \quad \text{against} \quad H_1 : p_1 \neq p_2.$$

A simple measurement to test the hypothesis is $\delta = p_2 - p_1$ which lies between -1 and 1. However, the restricted parameter space of δ may lead to anomalies (Whitehead, 1997, Ch. 3) and so the log odds ratio is used instead. It is

defined as

$$\theta = \log\Big(\frac{p_2(1-p_1)}{p_1(1-p_2)}\Big),$$

and because the log odds ratio lies between $-\infty$ and ∞, it is a more appealing measurement than the simple measurement of difference of proportions. The hypotheses are now rewritten as

$$H_0 : \theta = 0 \quad \text{against} \quad H_1 : \theta \neq 0,$$

The score statistic for θ is

$$B = \frac{n_1 S_2 - n_2 S_1}{n}, \tag{3.20}$$

and the Fisher's information is

$$V = \frac{n_1 n_2 S(n-S)}{n^3}, \tag{3.21}$$

where S_i is the number of successes out of n_i for $i = 1, 2$, $S = S_1 + S_2$ and $n = n_1 + n_2$. The score statistic B is approximately normally distributed with mean θV and variance V, $B \sim N(\theta V, V)$. Assume that $n_1 = n_2$, and that the probability of success of the whole trial is $\bar{p} = (p_1 + p_2)/2$. For large sample size, $S \approx n\bar{p}$, and so the equation (3.21) becomes

$$V \approx \frac{n\bar{p}(1-\bar{p})}{4}. \tag{3.22}$$

For a two-sided hypothesis, under the null hypothesis,

$$\begin{aligned} & \Pr(B > c|\theta = 0) = \alpha/2 \\ \Leftrightarrow \quad & \Pr\Big(Z > \frac{c - \theta V}{\sqrt{V}}\Big) = \alpha/2 \\ \Leftrightarrow \quad & 1 - \Phi\Big(\frac{c}{\sqrt{V}}\Big) = \alpha/2 \\ \Leftrightarrow \quad & c = z_{1-\alpha/2}\sqrt{V}. \end{aligned}$$

Let θ_A be the anticipated log odds ratio to be detected from the trial and it is one of the parameters in the alternative space $\Omega - \omega$. The power of the trial is,

$$\begin{aligned} 1 - \beta &= \Pr(B > c|\theta = \theta_A) \\ &= \Pr\Big(Z > \frac{z_{1-\alpha/2}\sqrt{V} - \theta_A V}{\sqrt{V}}\Big) \\ &= \Pr\Big(Z > z_{1-\alpha/2} - \theta_A\sqrt{V}\Big) \\ &= 1 - \Phi\Big(z_{1-\alpha/2} - \theta_A\sqrt{V}\Big). \end{aligned}$$

Assume that p_1 is a fixed constant while p_2 is a random parameter that follows a beta distribution with fixed parameters a and b. The power now has to be averaged over all possible values of p_2,

$$\begin{aligned} A &= \int_0^1 \Big(1 - \Phi\Big(z_{1-\alpha/2} - \theta\sqrt{V}\Big)\Big) f_2(p_2)\, dp_2 \\ &= \int_0^1 \Big(1 - \Phi\Big(z_{1-\alpha/2} - \log\Big(\frac{p_2(1-p_1)}{p_1(1-p_2)}\Big)\sqrt{\frac{n\bar{p}(1-\bar{p})}{4}}\Big)\Big) f_2(p_2)\, dp_2, \end{aligned} \tag{3.23}$$

which is also a function of n and can only be evaluated numerically.

3.5 Decision theory

The simple hypothesis testing is one of inference problems where the results from the trial is used to infer the value of the unknown true parameter θ of a random variable X. In a broad sense, it is also a decision problem. After the collection of observations, a decision is made from a choice of two. These decisions are:

d_0**:** The hypothesis that the unknown θ belongs to ω is true.

d_1**:** The hypothesis is false.

Another technique of formally making informed decision is the statistical decision theory. Decision theory is concerned about making decision under uncertainty and each decision has its consequence and "value". Under the uncertain circumstances, a decision has to be made such that it is the best possible one with the knowledge that the worst scenario could happen (Pratt *et al.*, 1995, Ch. 1).

The "value" of a consequence could be a monetary reward which is measurable in existing scale or it could be a value that has no obvious scale of measurement, such as happier feeling. However, to work on these "values", numbers are assigned and they are called utilities. The function of decision and parameter is called the utility function and is denoted by $G(d, \theta)$. A decision problem is solved by maximising the expected utility.

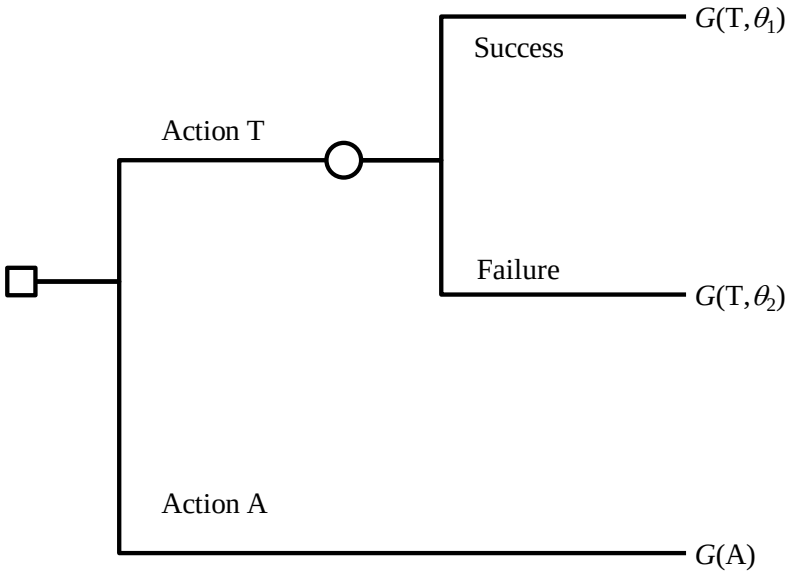


Figure 3.4: A simple decision tree for a phase II trial.

For the review of the decision theory, a simple illustration is used. Suppose that a new treatment is available for a phase II clinical trial then there are two possible actions to choose:

Action T: Try the new treatment in the phase II trial, or

Action A: Do not try the new treatment in the phase II trial.

There are two possible states of nature from the new treatment:

θ_1**:** The new treatment is effective,

θ_2**:** The new treatment is not effective.

The decision problem for the above scenario can be represented by a decision tree (Fig. 3.4). If the new treatment is effective, the trial is declared as a success and it has 1 unit of gain. On the other hand, if it is not effective,

that is, a failure, it has 0 unit of gain. Let the cost of starting a trial be m which is relative to the one unit of gain. The gain of taking action T and if the treatment is effective is, $G(T, \theta_1) = 1 - m$, and if the treatment is not effective it is, $G(T, \theta_2) = -m$. If action A is taken, then there is no cost incurred and so $G(A, \theta_i) = 0$ for $i = 1, 2$ (Hilden, 1990). The utility table is as shown in Table 3.1.

Suppose that the probability of the trial being a success is p and the probability of it failing is $1 - p$. The expected utility function of action $a \in \{T, A\}$ is

$$\mathcal{G}(a) = \sum_{\theta \in \Theta} G(a, \theta) p(\theta).$$

Thus, the expected utility for action T is

$$\mathcal{G}(T) = p(1 - m) - (1 - p)m = p - m,$$

and the expected utility for action A is

$$\mathcal{G}(A) = 0.$$

Therefore, if the probability of success is greater than the relative start-up cost, the optimal action is action T, otherwise, action A. For example, if the relative start-up cost is 0.02, then if $p > 0.02$, the treatment should be put on trial but if $p < 0.02$ then the trial should be abandoned.

In some clinical trials, it may be possible to recruit a group of patients or volunteers to start on treatments at the same time. For most clinical trials however, patients are recruited to the trial and treated serially. Although the

Table 3.1: Utility table for a phase II trial.

Action	State of nature	
	θ_1	θ_2
T	$1-m$	$-m$
A	0	0

responses from these patients are available in a sequential order, the analysis to test the hypothesis is still performed at the end of the trial—after the data from the last patient has been obtained. Nevertheless, due to the sequential nature there is a feasibility to analyse the data as they are made available especially if it deems more advantageous to do so. One of the advantages of doing sequential analysis is the flexibility to stop a trial early.

After each sequential analysis, there is a possible set of decisions to be made: 1) to stop the trial because the new treatment is not efficacious and thus subject fewer patients to the inferior treatment, 2) to stop the trial and recommend the drug for larger confirmatory trials or for marketing, thus making it available for more patients quicker, or 3) to recruit more patients as the results are inconclusive to decide if the treatment is effective or not.

One of the key issues in sequential analyses is that the stopping rules have to be laid down in the design stage of the trial. The conditions and rules state when and how the trial should stop or continue with patient recruitment. There are a few methodologies to construct the rules to decide if the trial should stop either for futility or efficacy, or to recruit more patients. In the frequentist setting, the size of the sequential test, α', is set to be smaller than the conventional one-stage analysis α (Jennison and Turnbull, 2000). The reason is that the null hypothesis is tested continuously and thus increase

the chance of claiming efficacy when in fact it is not. Another method is to construct stopping boundaries where a set of critical values are determined before the trial begins. The observed data are used to calculate the test statistics and then compared with the critical values to determine if the trial should stop or continue.

3.5.1 Backward induction

The formulation of the design of sequential clinical trials to be discussed in Chapters 6 and 7 is based on the Bayesian decision theoretic approach and thus, will be discussed in slightly greater length here. For an illustration, in a clinical trial, n_1 patients are recruited in the first stage and upon the collection of observations a decision is made from a choice of:

Action R: Recruit another patient, or

Action A: Abandon the trial.

If action R is taken, n_2 patients are recruited in the second stage and a decision to take action R or A is made from the accumulated responses. At stage k the decision making is based on the accumulated data from $n_1, n_2, \ldots, n_k$ patients and either the trial terminates or continue by recruiting n_{k+1} patients. Suppose that there are only N patients eligible for trial and $N = \sum_{i=1}^{k} n_i$. If all N patients were observed, then the only action available is to terminate the trial, action A. The sequential decision tree is shown in Figure 3.5.

The decision problem is solved by backward induction which begins by considering the last stage of the decision tree and then moves backward to

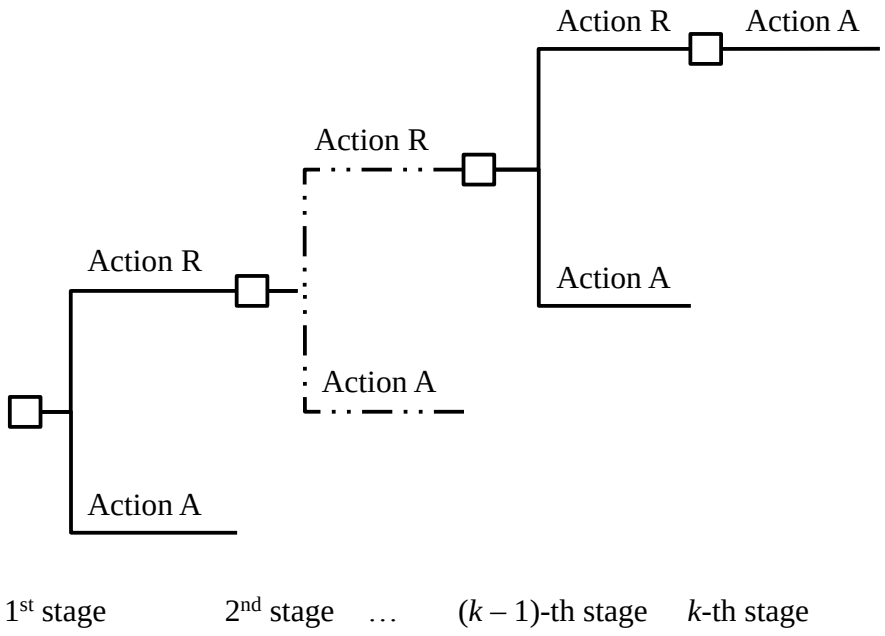


Figure 3.5: A simple sequential decision tree for a phase II trial.

the first stage of the observation (DeGroot, 1970, Ch. 12). Analogous to the simple decision tree, the gain function of recruiting n_1 patients must be greater than the gain function of not recruiting in order for the trial to commence. Based upon the observation from n_1 patients, if there is benefit in recruiting n_2 patients for more information than not recruiting, then action R should be taken. Otherwise, action A. In the former scenario n_2 patients are recruited in the second stage and based on the accumulated information from n_1 and n_2 if there is benefit in having more information from n_3 patients, then the optimal decision is to take action R, that is, recruit n_3 patients in the third stage. Continuing in this manner of evaluation, at the penultimate stage where given information from $n_1, n_2, \ldots, n_{k-1}$ patients, if the benefit of recruiting n_k patients is greater than not recruiting, then the last group of

patients should be recruited to the k-th stage before the trial terminates.

Let x_i be the observed events from the i-th stage, $i = 1, 2, \ldots, k$. The gain function at the k-th stage can be denoted by $G(A|x_1, \ldots, x_k)$ because only action A is available after obtaining information from all N patients. At the $(k-1)$-th stage, the gain function of action R is

$$G(R|x_1, \ldots, x_{k-1}) = \sum_{x_k} G(A|x_1, \ldots, x_k) f(x_k|x_1, \ldots, x_{k-1}),$$

which depends on the benefit of recruiting n_k more patients and the possible values that may be observed. The expression $f(x_k|x_1, \ldots, x_{k-1})$ is the function of the possible observed events in the k-th stage given the observed events $x_1, x_2, \ldots, x_{k-1}$.

Similarly, the gain function of action A depends on all the $x_1, x_2, \ldots, x_{k-1}$ observations, denoted by $G(A|x_1, \ldots, x_{k-1})$. If

$$G(R|x_1, \ldots, x_{k-1}) > G(A|x_1, \ldots, x_{k-1}),$$

then the optimal decision is to take action R and if otherwise, action A. By working recursively back to the first stage, an optimal decision can be made for all possible observations from n_1 patients.

As described by Lindley (1961), the optimality problem at each present stage is solved by considering the optimum future. The sequential recruitment of patients by blocks of n_i's is known as group sequential. When $n_i = 1$, for $i = 1, 2, \ldots, k$ the sequential recruitment is called fully sequential. In Chapters 6 and 7, the illustrations of the patient recruitments for the design

for a series of trials are based on fully sequential.

3.6 Concluding remarks

In the design of clinical trials, one of the key components is to have an appropriate sample size such that the minimally clinical accepted efficacy is detected with small errors. Some of the methodology to determine sample size is discussed in the following chapter. Most of the designs regard each clinical trial individually even though the success or the failure of one may have an impact on subsequent trials. A new design for a series of trials is proposed in Chapter 5 so that the optimality of the whole of the project development is considered. Following on that, a series of sequential trials with sequential sampling is proposed in Chapter 6. Patients are recruited sequentially and observations from the patients are then used to support if the current trial should continue recruitment, stop and initiate another clinical trial or abandon the development programme. Treatments targeting the same population may be more similar and thus may be correlated. The design for a series of sequential trials is therefore extended by considering the correlation between treatments (Chapter 7).

All the proposed designs make use of the Bayesian approach by maximising the assurance of each trial which optimize the expected gain of the whole series of trials. On the basis of that, the random variable X is assumed to take on values from the real line and it follows a known form of distribution depending on a random parameter θ. The parameter θ is assumed to follow a known distribution with fixed parameters, implying a prior distribution for

θ. The elicitation of the prior distribution could be from the opinion and judgments of an expert, a group of experts, data from previous studies or published results (O'Hagan *et al.*, 2006). The prior densities in this thesis are estimated from data of published trials using maximum likelihood estimate methodology.

Chapter 4

Sample Size Determination

Due to the inherent biologic variations in patients presented with the same condition it is necessary to have a group of patients in clinical trials. The results from the sampled patients are consequently used to infer how the treatments may behave in the population. Thus, one of the fundamental issues in the design of a clinical trial is the number of patients that should be sampled. On the one hand, a sample size that is too large may delay the treatment from being made available to the population when it has shown some minimum clinical efficacy. On the other hand, an inadequate sample size may not be able to draw a valid conclusion thus, subjecting patients unnecessarily to "questionable" treatments.

This chapter discusses some of the common methods used to determine sample size for a superiority trial which aims to establish the superiority of the experimental treatment to a control treatment that could either be a placebo or the current standard treatment. In a phase II setting however, it may not be necessary to have a controlled arm. Instead the efficacy of

the experimental treatment is compared against a known value of an historical control. The designs discussed in the later sections thus include both controlled (two-arm) and uncontrolled (one-arm) trials, and the primary endpoint is assumed to be either a continuous variable or a binary variable.

This thesis assumes that the trial's primary outcome, X, is a random variable that has a known form of a probability distribution function with an unknown parameter θ. The probability density function (or equivalently the probability mass function for a discrete variable) is represented by $f(x|\theta)$. Patients' outcomes are independent of each other and so each random variable is assumed to be independently and identically distributed with the same probability density function, $f(x|\theta)$.

The sample size is determined based on the analysis of the primary endpoint that is to be done at the end of the trial (ICH, 1998), that is, an inference on parameter θ is made. Issues such as patient withdrawal and protocol violation may affect the actual number of patients in a trial and thus, methods that deal with these issues are employed alongside the common sample size calculation formulations to determine the minimum sample size that is necessary. However, these methods will not be discussed in this thesis.

Sample size determination methods are broadly classified into two main groups, namely, frequentist and Bayesian. The Bayesian methods are further divided into two general categories; an inferential technique and a decision theoretic technique which treats the inference problem as a decision problem based on utility or loss function. Some designs make use of both frequentist and Bayesian methods and they are commonly known as the hybrid approach.

As discussed briefly in Spiegelhalter *et al.* (2004, Ch. 6), some authors advocated the decision-theoretic framework for clinical trial designs and some advocated the inferential framework. For the former framework the argument is that a decision is ultimately made whether to stop the trial or not and therefore, assessing utilities. Under the inferential framework the argument is that the estimated efficacy of interest should be reported with sufficient confidence because the population who may be receiving the treatment after the trial is usually more heterogeneous than the population in the trial. Both frameworks have their merits and Whitehead (1993) argues that Bayesian decision-theoretic framework is appropriate in early phase trials whereas a frequentist approach is suitable for phase III trial.

This thesis concentrates on the designs of a series of trials based on the hybrid approach. As such the general frameworks of both frequentist and Bayesian methodology will be discussed. The first section is on the frequentist method for continuous and binary random variables. There is a subsection on sample size determination for phase II trials based on the frequentist approach (Section 4.1.3). Following on, Section 4.2 discusses some general clinical trials design based on the Bayesian methodology. The hybrid method is discussed in Section 4.3 and Section 4.4 presents some published works on designing a series of clinical trials. Finally, Section 4.5 concludes the chapter. Details of the methods discussed in this chapter can be found in these textbooks: Friedman *et al.* (2010), Joseph and Belisle (1997), Julious (2010), Lachin (1981) and Machin *et al.* (2009).

4.1 Frequentist method

The principle of sample size determination is to choose a sample size so that an analysis is performed on the observed primary outcomes and consequently infer the unknown parameter θ. Therefore, it is necessary to ascertain the primary patient outcome which is chosen based on the main objective of the trial.

In the classical frequentist approach, the objective of the trial is delineated in the hypothesis where the null hypothesis states that the true difference between the experimental and control treatments belongs to a subset of the parameter space. Let Ω be the parameter space and let ω be the subset of Ω where the true difference belongs to. For the purpose of designing a trial, the alternative hypothesis needs to be specified, too, where it simply states that the true treatment difference does not belong to ω but to $\Omega - \omega$.

4.1.1 Continuous variable

Controlled trial

Consider a trial examining the efficacy between the experimental treatment and a control treatment, and that the primary outcome is a continuous variable, that is, it takes a range of values from the real line. Suppose that the trial's objective is to estimate the true difference between the mean of the true outcome of the experimental treatment and the mean of the true outcome of the control treatment. Let $X_1, X_2, \ldots, X_{n_1}$ be n_1 independent continuous outcomes from the control group where $X_i \sim N(\mu_1, \sigma_1^2)$, $i = 1, \ldots, n_1$. Let

$\overline{X} = \sum_{i=1}^{n_1} X_i/n_1$ be the mean of $X_1, \ldots, X_{n_1}$ and as the observations from the patients are assumed to be independent from each other, $\overline{X} \sim N(\mu_1, \sigma_1^2/n_1)$. Similarly, let $Y_1, Y_2, \ldots, Y_{n_2}$ be the n_2 independent continuous outcomes from the experimental treatment group where $Y_j \sim N(\mu_2, \sigma_2^2)$, $j = 1, \ldots, n_2$, and let $\overline{Y} = \sum_{j=1}^{n_2} Y_j/n_2$ be the mean from the experimental group where $\overline{Y} \sim N(\mu_2, \sigma_2^2/n_2)$.

Let the difference of the mean of the two treatment arms denoted by $\delta = \overline{Y} - \overline{X}$. The observations from both arms are assumed to be independent of each other and so $\delta \sim N(\mu_2 - \mu_1, \sigma_1^2/n_1 + \sigma_2^2/n_2)$, that is, δ also follows the normal distribution with mean $\mu_2 - \mu_1$ and variance $\sigma_1^2/n_1 + \sigma_2^2/n_2$. For ease of notation, let $\theta = \mu_2 - \mu_1$ and $n_1 = rn_2$ where r is the allocation ratio. The total number of patients to be recruited to the trial is $N = n_1 + n_2 = (1 + r)n_2$. Therefore, rewriting the notation, $\delta \sim N(\theta, (\sigma_1^2 + r\sigma_2^2)/(rn_2))$, which is equivalent to

$$\frac{\delta - \theta}{\sqrt{(\sigma_1^2 + r\sigma_2^2)/(rn_2)}} \sim N(0, 1). \tag{4.1}$$

The objective of the trial is to estimate the true difference between the two means. Following on the review in Chapter 3.3, assuming that there is only one element in ω and that element is denoted by θ_0 then the null hypothesis is written as $H_0 : \theta = \theta_0$. Under the alternative hypothesis, also, assumed that there is also only one element in $\Omega - \omega$ and this is denoted as θ_A, so the alternative hypothesis is $H_1 : \theta = \theta_A$. In determining the sample size, both θ_0 and θ_A are specified *a priori*. These two values may be obtained by assuming the true mean of the control arm and the smallest difference

that is of clinical significance between the two arms.

In the process of inferring the population mean, two types of errors may be incurred and they are type I and II errors. Type I error is the error of rejecting the null hypothesis when it is true and type II error is the error of not rejecting the null hypothesis when it is false. The maximum allowable levels of type I and II error rates need to be specified *a priori*, too, in the design stage when determining the sample size. Let α and β be the maximal allowable levels for the type I and II errors, respectively,

$$\begin{aligned} \Pr(\text{Reject } H_0|\theta = \theta_0) &= \Pr(\text{type I error}) = \alpha \\ \Pr(\text{Non-rejection of } H_0|\theta = \theta_A) &= \Pr(\text{type II error}) = \beta. \end{aligned} \tag{4.2}$$

Assuming that the variance of δ is the same in both null and alternative hypotheses, a general equation for sample size determination can be developed from (4.1) and (4.2). Similar to the steps seen in (3.16), under the null hypothesis and a two-sided α-level of significance, the critical region to reject H_0 is given by

$$\begin{aligned} & \Pr(\text{Reject } H_0|\theta = \theta_0) = \alpha/2 \\ \Leftrightarrow \quad & \Pr(\delta > c|\theta = \theta_0) = \alpha/2 \\ \Leftrightarrow \quad & \Pr\left(Z > \frac{c - \theta_0}{\sqrt{(\sigma_1^2 + r\sigma_2^2)/(rn_2)}}\right) = \alpha/2 \\ \Leftrightarrow \quad & 1 - \Phi\left(\frac{c - \theta_0}{\sqrt{(\sigma_1^2 + r\sigma_2^2)/(rn_2)}}\right) = \alpha/2 \\ \Leftrightarrow \quad & c = z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2 + r\sigma_2^2}{rn_2}} + \theta_0 \end{aligned} \tag{4.3}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution and z_γ is the lower 100γ percentile of the standard normal distribution.

Under the alternative hypothesis, the random variable δ follows the normal distribution with mean θ_A and variance $(\sigma_1^2 + r\sigma_2^2)/(rn_2)$, equivalently, this is rewritten as $(\delta - \theta_A)/(\sqrt{(\sigma_1^2 + r\sigma_2^2)/(rn_2)}) \sim N(0,1)$. As shown earlier, in (3.17) the power function is

$$\begin{aligned} & \Pr(\text{Reject } H_0 | \theta = \theta_A) = \text{ Power} \\ \Leftrightarrow \quad & \Pr(\delta > c | \theta = \theta_A) = 1 - \beta \end{aligned} \tag{4.4}$$

Substituting (4.3) into (4.4) to solve for n_2,

$$\begin{aligned} & \Pr\left(Z > z_{1-\alpha/2} + \frac{\theta_0 - \theta_A}{\sqrt{(\sigma_1^2 + r\sigma_2^2)/(rn_2)}} \right) = 1 - \beta \\ \Leftrightarrow \quad & z_{1-\alpha/2} - \frac{\theta_A - \theta_0}{\sqrt{(\sigma_1^2 + r\sigma_2^2)/(rn_2)}} = z_\beta \\ \Leftrightarrow \quad & \sqrt{rn_2} = \frac{(z_{1-\alpha/2} - z_\beta)\sqrt{\sigma_1^2 + r\sigma_2^2}}{\theta_A - \theta_0} \\ \Leftrightarrow \quad & n_2 = \frac{(\sigma_1^2 + r\sigma_2^2)(z_{1-\alpha/2} - z_\beta)^2}{r(\theta_A - \theta_0)^2}. \end{aligned} \tag{4.5}$$

Therefore, the total number of patients needed is $N = (1 + r)n_2$ where $n_1 = rn_2$ patients are randomized to the control arm and n_2 patients are randomized to the experimental treatment arm.

Assuming that the variances of both arms are equal, $\sigma^2 = \sigma_1^2 = \sigma_2^2$, and

the allocation ratio is 1:1 then (4.5) is simplified into

$$n_2 = \frac{2\sigma^2(z_{1-\alpha/2} - z_\beta)^2}{(\theta_A - \theta_0)^2}. \tag{4.6}$$

The total number of patients to be recruited is thus, $N = 2n_2$, a positive number. In practice, N is rounded up to the nearest even integer. For example, consider an asthma clinical trial whose objective is to detect a difference of 10% between treatments in mean percentage change from baseline in the forced expiratory volume in one second (FEV_1) observed 12 hours after the treatment dose against a null hypothesis of no difference between treatments, that is, $\theta_0 = 0$ and $\theta_A = 10$. The standard deviations of both treatment arms are assumed to be equal and fixed at $\sigma = 14$. Let $\alpha = 0.05$ and $\beta = 0.10$, and the randomization ratio is 1:1 then from equation (4.6), the number of patients needed per arm is $n_1 = n_2 = 41.189$. Rounding up to the nearest even integer the total number of patients needed is $N = 84$.

Uncontrolled trial

Some phase II trials are conducted with only one treatment arm, that is, the experimental treatment is tested against a known historical control value. In this setting, the sample size determination is done by testing the parameter μ_2 of the random variable $\overline{Y}$ under these hypotheses: $H_0 : \mu_2 = \theta_0$ against $H_1 : \mu_2 = \theta_A$ where θ_0 is the known historical value and θ_A is the minimum value that is of clinical significance that may be attained by the experimental treatment. Given the type I and II error rates and assuming that the variance

is known, the desired sample size is

$$n = \frac{(z_{1-\alpha/2} - z_\beta)^2 \sigma_2^2}{(\theta_A - \theta_0)^2}. \tag{4.7}$$

Consider the example given in the controlled trial setting but instead of randomizing patients to two treatment arms, all n patients are given the experimental treatment. Assume that the mean percentage change from baseline in FEV_1 of the historical control is $\theta_0 = 0$ and that the mean percentage change of the experimental treatment is $\theta_A = 10$. The standard deviation is as given in the earlier example, $\sigma_2 = 14$ and so, for a two-sided test at 0.05 level of significance and power of 0.90 the desired sample size is $n = 21$ (rounding up from $n = 20.59$).

4.1.2 Binary variable

This section considers clinical trials with binary data as the primary endpoint, that is, the response from each patient is dichotomized into, for example, success or failure, yes or no. Usually, a positive response such as success is represented numerically by 1 and a negative response, failure, by 0. Consider a two-arm randomized trial where patients are randomly allocated to either treatment. Let $X_1, X_2, \ldots, X_{n_1}$ be n_1 independent binary outcomes from the control group where $X_i = 1$ if the response is a success and $X_i = 0$, otherwise, for $i = 1, 2, \ldots, n_1$. Let $S_1 = X_1 + X_2 + \ldots + X_{n_1}$ be the sum of successes from n_1 patients. A common and convenient probability distribution for X_i is the Bernoulli distribution and therefore, as seen earlier in Section 3.1.1, S_1 is assumed to follow the binomial distribution with in-

Table 4.1: Statistics for a two-arm randomized controlled trial with binary response.

Treatment arm	Number of successes	Number of observations	Observed success rate	Anticipated success rate
Control	s_1	n_1	s_1/n_1	p_1
Experimental	s_2	n_2	s_2/n_2	p_2
Overall	s	N		

dex n_1 and an unknown parameter p_1, the probability of success. Similarly, let $Y_1, Y_2, \ldots, Y_{n_2}$ be n_2 independent binary outcomes from the experimental treatment arm and let $S_2 = Y_1 + Y_2 + \ldots + Y_{n_2}$ be the sum of successes from n_2 patients. Likewise, S_2 is assumed to follow the binomial distribution with index n_2 and an unknown probability of success, p_2. Let $S_1 = s_1$ and $S_2 = s_2$ be the observed successes at the end of the trial, then the statistics may be recorded as shown in Table 4.1.

There are a few methods to summarize the binary effects between two treatments at the end of the trial. Some of the common summary measures are absolute risk difference, odds ratio, log odds ratio and relative risk (Table 4.2). The sample size determination depends on the hypothesis which depends on the summary measure that would be used for the final analysis at the end of the trial. This thesis will concentrate on the log odds ratio but examples will be given for both absolute risk difference and log odds ratio summary measures.

Absolute risk difference: controlled trial

Let r be the allocation ratio such that $n_1 = rn_2$, then the total number of patients needed for the trial is $N = (r + 1)n_2$. First, consider the setting of

Table 4.2: Summary measures for binary response.

Measure	Definition
Absolute risk difference	$p_2 - p_1$
Odds ratio	$\frac{p_2(1-p_1)}{p_1(1-p_2)}$
Log odds ratio	$\log\left(\frac{p_2(1-p_1)}{p_1(1-p_2)}\right)$
Relative risk	p_2/p_1

testing the absolute risk difference between the two treatment groups. The null and alternative hypotheses are thus written as

$$H_0 : p_2 = p_1 \quad \text{against} \quad H_1 : p_2 \neq p_1.$$

For some values of n_1 and n_2 that satisfy these conditions $n_1 p_1(1-p_1) \geq 10$ and $n_2 p_2(1-p_2) \geq 10$, respectively, the normal distribution is a good approximation to the binomial distribution. As such, the test statistic to test the null hypothesis is defined as

$$Z = \frac{\hat{p}_2 - \hat{p}_1}{\hat{\sigma}},$$

where $\hat{p}_1 = s_1/n_1$, $\hat{p}_2 = s_2/n_2$ and $\hat{\sigma}^2 = \hat{\bar{p}}(1-\hat{\bar{p}})(1/n_1 + 1/n_2)$. The average proportions of events is estimated as $\hat{\bar{p}} = (r\hat{p}_1 + \hat{p}_2)/(r+1)$.

The null hypothesis is defined as, $H_0 : p_2 = p_1 = \bar{p}$, where $\bar{p} = (rp_1 + p_2)/(r+1)$ and the variance is $\sigma_0^2 = (1+r)\bar{p}(1-\bar{p})/(rn_2)$. Under the null hypothesis and for large sample sizes, the test statistic, Z, is assumed to follow the standard normal distribution, that is, $Z \sim N(0,1)$. Following the steps seen in Section 4.1.1, for a two-sided α-level of significance the critical

region to reject the null hypothesis is

$$\begin{aligned} \Pr(\text{Reject } H_0 | p_2 = p_1) &= \alpha/2 \\ \Pr\left(Z > \frac{c - (p_2 - p_1)}{\sigma_0}\right) &= \alpha/2 \\ 1 - \Phi\left(\frac{c}{\sigma_0}\right) &= \alpha/2 \\ c &= z_{1-\alpha/2}\sigma_0. \end{aligned} \tag{4.8}$$

As the variance of the test statistic depends on the anticipated probabilities of success of each treatment arm, it is necessary to specify both p_1 and p_2 *a priori* at the design stage for sample size calculation. Under the alternative hypothesis, $H_1 : p_2 - p_1 = p_A$, the variance is $\sigma_1^2 = p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2$, and the power function is

$$\Pr\left(Z > \frac{c - (p_2 - p_1)}{\sigma_1}\right) = 1 - \beta. \tag{4.9}$$

Substituting (4.8) into (4.9) to solve for n_2,

$$\begin{aligned} &\Pr\left(Z > \frac{z_{1-\alpha/2}\sigma_0 - (p_2 - p_1)}{\sigma_1}\right) = 1 - \beta \\ \Leftrightarrow \quad & n_2 = \frac{\left(z_{1-\alpha/2}\sqrt{(1+r)\bar{p}(1-\bar{p})} - z_\beta\sqrt{p_1(1-p_1) + rp_2(1-p_2)}\right)^2}{r(p_2 - p_1)^2}. \end{aligned} \tag{4.10}$$

Therefore, the minimum total sample size needed is $N = (1+r)n_2$ and of this, $n_1 = rn_2$ is randomized to the control arm and n_2 patients to the experimental treatment arm.

As an illustration of the sample size calculation, consider an asthma clinical trial whose objective is to test if the experimental treatment is better in controlling asthma exacerbation than the placebo. The primary endpoint is binary where at least an episode of moderate or severe asthma exacerbation during a 4-week treatment period is considered as a failure and the absence of any asthma exacerbation event is considered as a success. Assume that the probability of success of the placebo is $p_1 = 0.80$ and that the probability of success of the experimental treatment is $p_2 = 0.90$. For a two-sided significance level of $\alpha = 0.05$ and power, $1 - \beta = 0.90$, and an equal allocation, $r = 1$, the minimum sample size that is required per arm is $n_1 = n_2 = 265.86$. So, rounding up to an even integer, the minimum total number of patients needed is $N = 532$.

Absolute risk difference: uncontrolled trial

In a single-arm trial, the true probability of success of the experimental treatment is tested against a fixed and known historical control, p_0. For sample size calculation, the probability of success of the experimental treatment, p_A, is specified *a priori*. Thus, the null hypothesis is written as, $H_0 : p_2 = p_0$, which yields $\sigma_0^2 = p_0(1 - p_0)/n_2$ where n_2 is the sample size required for the trial. The alternative hypothesis, on the other hand, is $H_1 : p_2 = p_A$ and the variance is $\sigma_1^2 = p_A(1 - p_A)/n_2$. The sample size is

$$n_2 = \left(\frac{z_{1-\alpha/2}\sqrt{p_0(1-p_0)} - z_{\beta}\sqrt{p_A(1-p_A)}}{p_A - p_0} \right)^2. \tag{4.11}$$

For illustration, let $p_0 = 0.80$, $p_A = 0.90$, $\alpha = 0.05$ and $\beta = 0.10$, then from (4.11), $n_2 = 136.52$ and so the minimum sample size required for the trial is 137.

Log odds ratio: controlled trial

The absolute risk difference discussed in the preceding section is a simple measurement bounded by $(-1, 1)$. This may however, lead to anomalies especially if the response is very near to the boundary (Julious, 2010, Whitehead, 1997). As such, the log odds ratio may be a better choice as a summary measure. The odds of success of the control arm is $p_1/(1-p_1)$ and the odds of success of the experimental treatment arm is $p_2/(1-p_2)$. The ratio of these odds, $OR = p_2(1-p_1)/(p_1(1-p_2))$, is the odds ratio and it can take any positive values. However, by taking the log of the odds ratio with respect to the natural base, the log odds ratio lies between $-\infty$ and ∞, an unbounded random variable. As such, it has a more appealing property than the absolute risk difference.

Let $\theta = \log\{p_2(1-p_1)/(p_1(1-p_2))\}$ be the log odds ratio, then the hypotheses of the absolute risk difference, $H_0 : p_2 = p_1$ against $H_1 : p_2 - p_1 = p_A$ is rewritten as

$$H_0 : \theta = 0 \quad \text{against} \quad H_1 : \theta = \theta_A.$$

Following the same notation in Table 4.1, let $S = S_1 + S_2$ be the sum of successes from both treatment arms, the score statistic of the log odds ratio

is

$$B = \frac{n_1 S_2 - n_2 S_1}{N}. \tag{4.12}$$

Following Whitehead's definition, the Fisher's information of the score statistic is

$$V = \frac{n_1 n_2 S(N-S)}{N^3}. \tag{4.13}$$

The score statistic, B, is approximately normally distributed with mean θV and variance V, thus, providing an attractive scale for design, analysis and inference.

Under the null hypothesis, $H_0 : \theta = 0$, the test statistic is $B/\sqrt{V}$. Therefore, for a two-sided α-level of significance, the critical region to reject H_0 is

$$\begin{aligned} & \Pr(B > c|\theta = 0) = \alpha/2 \\ \Leftrightarrow \quad & \Pr(Z > c/\sqrt{V}) = \alpha/2 \\ \Leftrightarrow \quad & c = z_{1-\alpha/2}\sqrt{V}. \end{aligned} \tag{4.14}$$

Under the alternative hypothesis, the power function is

$$\begin{aligned} & \Pr(B > c|\theta = \theta_A) = 1 - \beta \\ \Leftrightarrow \quad & \Pr\left(Z > \frac{c - \theta_A V}{\sqrt{V}}\right) = 1 - \beta \\ \Leftrightarrow \quad & 1 - \Phi\Big(z_{1-\alpha/2} - \theta_A\sqrt{V}\Big) = 1 - \beta \\ \Leftrightarrow \quad & \sqrt{V} = (z_{1-\alpha/2} - z_\beta)/\theta_A. \end{aligned} \tag{4.15}$$

Let r be the allocation ratio, then for large sample trial, $S \approx N\bar{p}$, where $\bar{p} = (rp_1 + p_2)/(r+1)$, the Fisher's information in equation (4.13) is now approximately

$$V \approx \frac{r}{1+r}\bar{p}(1-\bar{p})n_2, \tag{4.16}$$

which is a function of sample size. Substituting (4.16) into (4.15) the sample size can be approximately determined by

$$\begin{aligned}\sqrt{\frac{r}{1+r}\bar{p}(1-\bar{p})n_2} &\approx \frac{z_{1-\alpha/2} - z_\beta}{\theta_A} \\ n_2 &\approx \frac{(1+r)(z_{1-\alpha/2} - z_\beta)^2}{r\bar{p}(1-\bar{p})\theta_A^2}.\end{aligned} \tag{4.17}$$

The desired approximate total sample size is thus, $N = (1+r)n_2$. Based on the same example for absolute risk difference measurement, given $p_1 = 0.80$, $p_2 = 0.90$, $\alpha = 0.05$, $\beta = 0.90$ and a 1:1 allocation ratio, the log odds ratio under the alternative hypothesis is $\theta_A = 0.8109$. Therefore, the approximate sample size per arm is 250.64 and the total sample size required is 502.

The formulation using the absolute risk difference (4.10) is not very dissimilar to the formulation using the log odds ratio (4.17). Referring to Julious and Campbell (1996), the log odds ratio can be approximated by $\theta \approx 2(OR-1)/(OR+1)$ and so

$$\frac{1}{\theta} \approx \frac{\bar{p}(1-\bar{p})}{p_2 - p_1}.$$

Without loss of generality, let $r = 1$, then equation (4.17) becomes

$$n \approx \frac{2(z_{1-\alpha/2} - z_\beta)^2 \bar{p}(1-\bar{p})}{(p_2 - p_1)^2}.$$

In addition, $p_1(1-p_1)+p_2(1-p_2) \approx 2\bar{p}(1-\bar{p})$ and so equation (4.10) becomes

$$n \approx \frac{(z_{1-\alpha/2} - z_\beta)^2 (2\bar{p}(1-\bar{p}))}{(p_2 - p_1)^2}.$$

Therefore, the two methods are approximately equivalent to each other.

4.1.3 Sample size for phase II trials

One of the aims of phase II trials is to screen out nonpromising treatments. Some phase II trials are designed as single-arm and the sample size is relatively smaller than what is required in a phase III setting. Many designs of phase II clinical trials are modelled from oncology trials. As such, the primary objective is usually to look for anti-tumour activity where if the tumour shrinks by at least 30% according to the RECIST criteria (Eisenhauer *et al.*, 2009) it is considered as a success and a failure otherwise. The primary endpoint is thus, a binary random variable. This section presents some of the common sample size determination methodologies for single-arm phase II trials with binary primary endpoint.

In screening out nonpromising treatments, the phase II trial is to test if the experimental treatment has met the minimum level of efficacy, p_A, in order to be recommended for further testing in a randomized phase III trial. If however, the experimental treatment is worse than a prespecified level of

efficacy, p_0, then it is not worthy of any further trial. The hypotheses are thus, written as

$$H_0 : p \leq p_0 \quad \text{against} \quad H_1 : p \geq p_A,$$

where p is the true probability of success of the experimental arm.

The origin of designs of trials has its roots in the agricultural field experiments where the field is plotted and sown, the crops grown and finally, harvested simultaneously for analysis (Whitehead, 1997). The data accumulation from clinical trials, on the other hand, happens steadily over a period of time. The results from patients who have been recruited earlier could be used for analysis, interpretation and decision-making. There are many reasons for the use of interim analyses but they can be loosely categorized into three: ethical, economic and administrative (Jennison and Turnbull, 2000, Ch.1).

Ethically, if the results from the interim analyses showed that the treatments are greatly ineffective, unsafe or inferior, patients should not be exposed to them anymore. Additionally, resources can then be channelled to other trials to study the next promising treatment. On the other hand, if the interim analyses showed that the result is positive, an early stopping of the trial will allow the treatments to be marketed sooner and consequently available for more patients. Economically, the expected sample size for a multi-stage trial is usually smaller compare to a single-stage trial's. Administratively, for example, if the non-compliance rate from patients are very high, this could be identified during the interim analyses. Thus, the trial can be protected from future non-compliance without losing any power to detect

a significant result.

Recruitment of patients to multi-stage trial can be done either in groups or individually. The latter sequential sampling is sometimes known as fully sequential trial where patients are recruited one at a time and analysis is performed after every response is obtained. Whereas, in group recruitment, analysis is performed after all the responses from the current and preceding groups are observed.

A practical issue regarding multi-stage trials is that patients accrual may have to be suspended for a few weeks or months while observing recruited patients' responses. The disruption may cause awkwardness if patient has been recruited but could not be treated until a decision is made. Also, the interest in the trial may wane if the duration of suspension is too long. As such, trials that are able to recruit patients quickly may benefit from single-stage designs.

Single-stage

One methodology of sample size determination for a single-stage single-arm phase II trial with binary primary endpoint is as shown in equation (4.11) but instead of testing at a two-sided α it is only tested at a one-sided level of significance (Fleming, 1982, Schoenfeld, 1980). Thus, the sample size is simplified to

$$n = \left(\frac{z_{1-\alpha}\sqrt{p_0(1-p_0)} - z_{\beta}\sqrt{p_A(1-p_A)}}{p_A - p_0} \right)^2. \tag{4.18}$$

Back to the same example where $p_0 = 0.80$, $p_A = 0.90$, $\alpha = 0.05$ and $\beta = 0.10$, at a one-sided level of significance, the sample size that is required is 109. As shown, the sample size obtained from this formulation may be quite large and so, to ensure that the required sample size is relatively smaller, the difference between experimental treatment and known historical control is set large or the type I and II error rates are set higher.

Schoenfeld (1980) argued that in the phase II setting, the type II error is more serious than the type I error because by incurring a type II error a better drug would be denied the chance of being studied further and patients are not able to benefit from a more superior drug. Thus, the type I error could be set as high as 0.25 while the type II error is still restricted to a low value, between 0.10 and 0.20. As an illustration, consider the same example where $p_0 = 0.80$, $p_A = 0.90$ and $\beta = 0.10$ but the type I error rate is set higher to $\alpha = 0.25$, then the required sample size is $n = 42.81$ and rounding up to the nearest integer the number of patients needed is 43, less than half of what would be required if $\alpha = 0.05$.

The sample size determination based on (4.18) is an approximation based on the normal distribution although the proportion of success has a binomial distribution. Referring back to the preceding example where $p_0 = 0.80$, $p_A = 0.90$, $\alpha = 0.05$ and $\beta = 0.10$, the minimum number of successes needed in order to conclude that the experimental treatment is worthy of further investigation is

$$S = cn = z_{1-\alpha}\sigma_0 + p_0 = 94.08.$$

If the true response rate was 0.80, the probability of observing at least 95

successes out of 109 is 3.52%. If the true response rate was 0.90, then the probability of observing at least 95 successes out of 109 is 87.29%. Therefore, the actual type I and II error rates are 3.52% and 12.71%, respectively. Note that the type II error is slightly higher than the prespecified 10%. Thus, for trials with small sample sizes, the normal approximation may not be ideal.

A'Hern (2001) proposed an exact binomial computation to determine sample sizes. The search for the minimum sample size and cut-off is done by using the exact binomial distribution and the cumulative binomial distribution function. For the same example of $p_0 = 0.80$, $p_A = 0.90$, $\alpha = 0.05$ and $\beta = 0.10$, the required sample size is 112 and the minimum number of successes needed to recommend the experimental treatment for further trial is 97. The actual type I and II errors are 4.67% and 9.22%, respectively.

Multi-stage

One of the earliest works on sequential methodology was by Wald (1945) where trials are conducted in stages. At each stage, a decision is made whether to stop the current trial or to continue recruiting patients to the next stage. If the latter decision is made, at the second interim stage a decision is made whether to stop the current trial or to continue recruitment based on the cumulative observed responses. The trial may go on until a definite decision is made at the last interim stage, that is, whether to conclude that the treatment is worthy of further trial or to abandon the trial. The decision is made by testing the null hypothesis and there is a theoretical optimal property where the expected sample size is smallest when the type I and II error rates are bounded by α and β, respectively, under H_0 and H_1 (Wald

and Wolfowitz, 1948). Other works that are commonly referred to and cited are, for example, from these authors Lan and DeMets (1983), Pocock (1977) and O'Brien and Fleming (1979).

There is more literature available on multi-stage trials with binary endpoint than for continuous endpoint due to the nature of discreteness. The calculations are more straightforward and the boundaries are easily obtained in the design stage. Gehan (1961) proposed a two-stage phase II trial although when it was first proposed, Gehan referred to it as preliminary and follow-up trials where preliminary is now commonly known as first stage and the follow-up is now known as second stage. In the first stage, there is no hypothesis testing, instead a decision is made based on the observed responses from patients. Suppose that the true probability of success of the experimental treatment is 90%, then there are two choices available for decision-making. The first choice is that the experimental treatment is unlikely to be effective in 90% or more of the patients. The second choice is that the experimental treatment could be effective in 90% or more of the patients. Assuming that the type II error rate is 5% then the probability that the first two consecutive responses are failures is 0.01, that is, the probability of observing at least one success in two consecutive patients enrolled into the trial is greater than 95%. The sample size for the first stage is thus, $n_1 = 2$. If there was no success observed among the two patients, the decision is to reject the experimental treatment. If however, at least one success was observed, then the second decision is made, that is, recruit more patients to pinpoint the treatment's effectiveness. The additional number of patients to be recruited following the initial observed success(es) is chosen so that the true probability of success

"is estimated with given precision, i.e. standard error".

Gehan's method requires the number of successes in the first stage to determine the sample size needed for second stage. Thus, the total number of patients cannot be determined at the design stage. Simon (1989) proposed a two-stage phase II design that determined the number of patients needed for stage one (n_1) and two (n_2) at the design stage. The decision whether to proceed to stage two of the trial is based upon the minimum number of successes observed at the end of stage one and the cut-offs are also determined in the design stage. If the true probability of response is p and the number of success observed is c_1 or less at the end of stage one, then the trial will end early. The probability of terminating the trial at stage one is $\Pr(X \leq c_1) = \sum_{x=0}^{c_1} \binom{n_1}{x} p^x (1-p)^{n_1-x}$ and the expected total sample size is $E(N) = n_1 + n_2(1 - \Pr(X \leq c_1))$. The sample sizes ($n_1$ and n_2) are determined under the constraints of α and β error rates by minimizing $E(N)$ and assuming that the true response rate is p_0. There are two optimization methods in Simon's design. One of the methods is known as the optimum design where the number of patients needed for stage one (n_1) is kept to a minimum to ensure that not many patients are subjected to an inferior treatment. The other method is known as a minimax design which is to choose the smallest maximum total sample size N that satisfies the design error probability constraints. Other multi-stage phase II trials have been proposed by, for example, Fleming (1982) and Chen (1997).

4.2 Bayesian method

In the frequentist approach, the measure of probabilities is in the data space whilst the Bayesian measures probabilities in both data space and parameter space. The principle of the frequentist approach is to use the observations from the sampled population to make inferences about the population. The interpretation of the probability assumes that in the long run, on average, the true parameter is θ. The principle of the Bayesian approach, on the other hand, considers the current trial in the context of other similar trials. Under the Bayesian framework the unknown parameter θ is assumed to follow a probability distribution function with known parameters. In this thesis, the probability distribution function of θ is assumed to be of some known parametric form, denoted by $f_{\Theta}(\theta)$.

The interpretation of the Bayesian probability is based on the degree of belief where the prior probability is the belief of the parameter space before any observation is obtained. Upon observing the responses from patients, the prior is updated and the posterior probability is the combined belief of the prior probability given the observed data (Stangl and Berry, 1998).

There are generally two main branches under the Bayesian framework for sample size determination, namely, a statistical decision theory technique and an inferential technique. Under the decision theory framework, consequences of all the available actions are expressed as either loss or utility functions. An optimum action, one that optimizes the loss or utility function, is then chosen. Under the inferential technique there is no explicit loss or utility function but the conclusion at the end of the trial is based on the posterior

distribution of the parameter of interest.

Most of the published works under the Bayesian framework are based on the decision theory framework and address sequential sampling. Hence, there is a considerably lengthy discussion on the decision theory framework than on the inferential technique in this section. In addition, most of the designs are proposed for a phase II setting. In general, phase II trials can be thought of as a screening process where promising treatments can be identified early and be recommended for further trial in a phase III setting. If on the other hand, they are nonpromising the trial can stop early thus, not subjecting patients unnecessarily to a nonpromising treatment. In between the two terminal decisions, the accumulated data may suggest that patient recruitment should continue.

Single-stage

Under the decision theory framework, the utility function, $G(n, x, d, \theta)$, is made up of sampling n patients, observing X responses and taking a decision d given that the true parameter is θ. Let $f_\Theta(\theta)$ be the probability density function of θ. The prior opinion of θ is not influenced by the decision d that is to be made at the end of the trial nor the sample size n that is to be sampled. From (3.7), the posterior density of θ upon observing $X = x$ responses from n patients is

$$f_{\Theta|X}(\theta|x, n) = \frac{f_{X|\Theta}(x|\theta, n) f_\Theta(\theta)}{f_X(x|n)}.$$

The marginal density of X, as shown earlier in (3.8), is

$$f_X(x|n) = \int f_{X|\Theta}(x|\theta, n) f_\Theta(\theta)\, d\theta.$$

The optimum sample size can be determined by backward induction which was discussed earlier in Chapter 3.5.1. For designing a single-stage trial the backward induction is simple. First, taking the expectations of the utility function over the random variable θ given (x, n). Secondly, maximize the expectation function over the deterministic variable d. Thirdly, take the expectation of the maximized expectation function over X, and finally, maximize over the deterministic value n (Lindley, 1997). The optimum sample size is thus,

$$\max_n \Big\{ \int_X \max_d \Big\{ \int_\Theta G(n, x, d, \theta) f_{\Theta|X}(\theta|x, n)\, d\theta \Big\} f_X(x|n)\, dx \Big\}.$$

If X is a discrete random variable, the integral is replaced by summation over all possible values of X.

The utility function can be made up of the costs of treating patients and conducting the trial, and the profits gained from a successful treatment (for example, Hilden *et al.* (1987) and Sylvester (1988)). It may also be made up of significance testing based on the classical frequentist power which then "classifies" the methodology as a hybrid design as it uses the prior distribution of the parameter in the design stage and assuming a classical frequentist hypothesis testing at the end of the trial (for example, Brunier and Whitehead (1994) and Stallard (1998)). Further discussion on multi-

stage trials based on decision-theoretic approach is presented in Section 4.3.

The optimum sample size determination under the Bayesian inferential framework is analogous to the frequentist approach except that the uncertainty about the parameter θ holds and that there is no hypothesis testing at the end of the trial. The Bayesian approach utilizes the posterior distribution which if the estimate based on the posterior distribution is within the credibility intervals, (L, U), the result is significant. As such, the sample size is obtained through the posterior distribution. For a fixed posterior interval (L, U), the optimum n^* is the smallest n that satisfies the equation

$$\int_X \Big(\int_L^U f_{\Theta|X}(\theta|x,n)\, d\theta \Big) f_X(x|n)\, dx \geq 1 - \alpha/2.$$

Joseph and Belisle (1997) presented other fully Bayesian methods for single-stage trials.

Multi-stage

Thall and Simon (1994) presented a design of a single-arm phase II trial where patients are recruited sequentially. The maximum number of patients to be recruited is fixed and will be denoted by $n_{\max}$. If the cumulative number of successes up to the i-th patient, S_i, is greater than or equal to the upper bound U_i then the trial will terminate and the new treatment is declared promising. If S_i is less than or equal to the lower bound L_i then the trial will terminate and the new treatment is declared nonpromising. If $L_i < S_i < U_i$ then the trial will continue by recruiting another patient. If however, at $i = n_{\max}$ and $L_i < S_i < U_i$ then the trial is concluded as inconclusive.

Other designs similar to Thall and Simon's have been proposed, for example, by Tan and Machin (2002) and Sambucini (2008). However, instead of monitoring patients continuously, these designs recruit patients in groups. At the interim stage, the posterior probability of the true response rate from the first group of patients is estimated. A decision is then made to either abandon the trial early or to continue to recruit more patients to the second stage. For the latter decision, once again the posterior probability of the true response is estimated based on all the patients and a decision is made from two choices, namely, to recommend the experimental treatment for further trial or to conclude that the experimental treatment is nonpromising.

It seems more intuitive and appealing to use Bayesian decision theory in multi-stage trials as at each interim analysis we are more informed on the efficacy of the treatment as data are amassed and consequently assisted the decision-maker in making informed decision. Some of the multi-stage trials designs based on Bayesian decision theoretic approach are by, for example, Brunier and Whitehead (1994), Hilden *et al.* (1987), Lewis and Berry (1994), Stallard (1998), Sylvester (1988), and Sylvester and Staquet (1980). These designs are briefly discussed in the next section as most of them are based on hybrid decision theory.

4.3 Hybrid method

Hybrid designs are designs that utilize both frequentist and Bayesian approaches. Most of the common designs assume a prior distribution for the parameter but the conclusion at the end of the trial is based on the classical

frequentist testing, thus, not making use of the prior belief. As such it is assumed that the point estimate of the treatment difference is reported with its confidence interval and p-value which is the probability of observing a treatment difference when in fact the two treatments are equally effective.

One example from among the hybrid decision-theoretic designs cited above is one proposed by Brunier and Whitehead (1994). The proposed phase II design assumes that the primary endpoint is binary and that the sum of successes, S, follows the binomial distribution with parameter p. Likewise, the parameter p is assumed to follow a prior distribution and its density is denoted by $f_\Theta(\theta)$. At the end of the phase II trial, the treatment is either concluded as nonpromising or recommended to proceed to a larger phase III trial. In the latter scenario it is assumed that the design of the phase III trial is based on the conventional frequentist approach (discussed in Section 4.1.2).

The utility function is based on the formulation of the cost of conducting an ineffective treatment (considered as a loss) and the expected gain if the treatment is found to be effective, and also the loss of rejecting an effective treatment. The proposed design considers the number of patients who will be treated with the new treatment in the phase III, denoted by n_{III}, if it has shown to be promising in the phase II setting. Let $B(p)$ denote the power of the phase III trial which is the probability that the experimental treatment is concluded as more effective than the standard treatment when the true probability of success is p. Consequently, it is assumed that n_F number of future patients will be treated with the new recommended treatment till a successor is found. Let n_{II} denote the sample size for the phase II trial and

let $A_n(p, c)$ denote the probability that the experimental treatment proceeds from phase II to phase III trial when the true probability of success is p, that is,

$$A_n(p, c) = \sum_{s=c+1}^{n_{\text{II}}} \binom{n_{\text{II}}}{s} p^s (1-p)^{n_{\text{II}}-s}.$$

The utility function is, thus,

$$G(n, c, \theta) = (p - p_0)\Big(n_{\text{II}} + A_n(p, c)(n_{\text{III}} + B(p)n_F)\Big)$$

The optimal design is one with the combination (n, c) that maximizes the expected utility function

$$\mathcal{G}(n, c) = \int_0^1 G(n, c, \theta) f_\Theta(\theta)\, dp, \tag{4.19}$$

where p_0 is the probability of success of the standard treatment. The utility function given in (4.19) may be extended to include the cost of conducting phase II and III trials. In addition, the number of future patients n_F may be assumed to be dependent on the time it needs to conduct both phase II and III trials (Stallard, 1998).

An example of a multi-stage hybrid decision-theoretic trial is one by Stallard (1998). The principle involved in obtaining an optimal multi-stage hybrid decision-theoretic trial is very similar to the single-stage design but the utility of each possible decision, d, is denoted by $G(n, c, d, \theta)$. If the optimum decision at an interim stage is to recruit more patients, the expected utility depends on the subsequent responses and the resulting decisions. The expected utility function is an averaged over all the posterior density of the

unknown parameter of the treatment efficacy given the observed responses.

4.4 Series of trials

According to Mariani and Marubini (1996) there may be two general scenarios that determine the designs of trials. One scenario is when the rate of eligibility patients is relatively higher than the availability of potential treatments. The second scenario is when there are a few new treatments available for trials simultaneously and the rate of eligible patients is relatively low such that the common designs (as introduced above) may not be suitable. In the latter scenario there is a need to identify promising treatments effectively and efficiently.

Whitehead (1985) proposed a design for a series of phase II trials where each trial is designed in the context of others and as such, they are considered as part of a single development plan. The design assumes that the total number of patients available for study is known and fixed. Let N denote the total number of patients and assume that n patients are assigned to each of the distinct trial. There are then a total of $K = N/n$ trials.

The design assumes that the primary endpoint is binary. Let p_i be the probability of success for the i-th trial, $i = 1, 2, \ldots, K$, and assume they are independent and identical random variables with prior density $f_p(p)$. The cumulated successes from each trial are ranked and the trial with the highest number of successes is subsequently recommended for phase III trial. Let $p_{[1]}$ denote the p_i from the most promising treatment, then the optimum number of treatments to be on trial, K^*, is obtained by maximizing the expected

probability of success, $E(p_{[1]})$, subject to the constraint of $N = nK$ where n and K are integers.

The proposed design is subsequently extended by including the design of a phase III trial in the series of trials (Whitehead, 1986). As mentioned above, from among the K^* trials, the treatment with the highest number of successes is eventually recommended for a phase III trial. Therefore, the design of the phase III trial is considered simultaneously in the extended development plan. This integrated approach is appealing if the population of interest is small as it is of importance to optimize the number of patients for each phase II trials and the eventual phase III trial.

More recent work by Yao *et al.* (1996) assumes that the total number of patients is not fixed. Instead, the main objective of the design is to minimize the total number of patients, N. In Whitehead's design, the trials can either run sequentially or concurrently whereas Yao's *et al.* design assumes that the trials run sequentially until a treatment is declared to be promising. The development plan is then considered as completed and a new one can commence.

Similar to Whitehead's designs, Yao *et al.* also assumes that the primary endpoint is binary. Let S_i be the sum of successes from trial i for $i = 1, 2, \ldots,$ and let p_i be the probability of success of treatment i. At the end of the trial, the observed data are used to test the null hypothesis $H_0 : p_i \leq p_0$ against the alternative $H_1 : p_i \geq p_A$. If the observed successes, $S_i = s_i$, is greater than the critical value, c, the treatment is declared promising. Suppose that K-th treatment is the first treatment to be declared promising. There are two types of errors that could be made in the conclusion in the series of trials.

One of the errors is accepting a nonpromising treatment and the other error is rejecting one or more promising treatments. The probabilities of these errors are, respectively,

$$\alpha_1 = \sum_{i=1}^{\infty} \Pr(S_1 \leq c, S_2 \leq c, \ldots, S_{i-1} \leq c, S_i > c, p_i < p_A),$$

and

$$\alpha_2 = \sum_{i=1}^{\infty} \Pr(S_1 \leq c, S_2 \leq c, \ldots, S_{i-1} \leq c, S_i > c, \bar{g}_i),$$

where

$$g_i = \{p_1 < p_A, p_2 < p_A, \ldots, p_{i-1} < p_A\},$$

and $\bar{g}_i$ is the complementary set to g_i. The optimum sample size, n^*, and cut-off, c^*, are obtained from a search algorithm by constraining the two posterior error probabilities, $\alpha_1 < e_1$ and $\alpha_2 < e_2$ where e_1 and e_2 are predefined maximum tolerable error rates.

Yao *et al.* design can be extended by including a phase III trial (akin to Whitehead's extension) where a treatment that is declared promising is then recommended to a phase III trial. Consequently, the development plan is considered as complete at the end of the phase III trial. The errors that may be made in the conclusion of the development programme now include the possible errors that may be made in the phase III setting. These errors are:

Error 1: Accepting a promising treatment in the phase II setting but subsequently rejecting it in the phase III setting.

Error 2: Accepting a nonpromising treatment in both phase II and III set-

tings.

Error 3: Rejecting one or more promising treatments in the phase II setting.

Wang and Leung (1998) extended Yao *et al.* design by considering group sequential sampling in each of the phase II trials, that is, patients are recruited sequentially and a decision is made based on the cumulative observed data. The available actions to choose from are to stop the current trial for efficacy and recommending the current treatment for larger trial, to recruit more patients to the current trial, and to cease current trial because of futility.

The optimum decision is obtained under the constraints of the prespecified posterior error rates introduced by Yao *et al.* This is computed using backward induction which is based on the principle of dynamic programming (French and Ros Insua, 2000, Parmigiani and Inoue, 2009). At each interim stage a decision is made from the available choices. If the optimum action is to continue patient recruitment, the utility of the action depends on the subsequent data and possible resulting actions. The optimum strategy is obtained by considering the whole programme using Bellman's principle of optimality which states that regardless of the initial state of nature and the initial decision, if observations have been taken so far leading to the current stage then the continuation of the strategy resulting from the first decision is the optimum strategy.

A recent work by Stallard (2012) is similar to Whitehead (1986) and Yao *et al.* (1996) where a series of phase II and III trials are considered simultaneously with the phase II trials running sequentially.

4.5 Concluding remarks

This thesis is about designing a series of clinical trials. The motivation for the design is that a pharmaceutical company may have a number of new treatments in development and ready for trial or a large public sector body has funding to identify a promising treatment from a pool of treatments for a particular disease population. The main assumption in the designs is that the total number of patients is assumed to be fixed and known. Equivalently, it is sufficient to fix the total funding available as this can be adjusted accordingly per the total number of patients.

The proposed designs utilize both classical frequentist and Bayesian formulations. The designs introduced in Chapter 5 consider each trial as part of a development plan. At the end of each trial the classical frequentist hypothesis testing is performed. However, the parameter of the random variable is assumed to follow a prior distribution. Therefore, the classical sample size determination for a continuous variable as discussed in Section 4.1.1 is used in the formulation but the parameter θ is assumed to be random.

The designs in Chapter 5 are assumed to be single-stage. Following on, the designs are extended to be multi-stage and these are presented in Chapters 6 and 7. As such, a convenient framework is to adopt the Bayesian decision-theoretic framework.

Part II

Designs for a Series of Hybrid Trials

Note by note, working on projection
Lips, teeth, throat, looking for a moment to inhale
Keeping the emotional connection
Even when your fellow actors fail
Pointing at the subtext by inflection
Helping your director reach perfection
Even though you have a strong objection
To the way he's handling the direction

Stephen Sondheim

Putting It Together

Chapter 5

A Series of Hybrid Trials

It is not uncommon for a pharmaceutical company to develop a number of potential new drugs simultaneously for the same population. Traditionally, clinical trial designs are chosen and planned individually without taking account of other trials. Resources are essentially finite and as these new treatments are made available for clinical trials it seems imperative to design the trial by considering it as part of a series of trials such that the long-term yield is maximized.

Some authors (for example, Whitehead (1985) and Yao *et al.* (1996)) have proposed to consider designs for a series of trials as a whole and these are in the setting of phase II trials. The works in this chapter extend some of the methodology developed by them. The aim is to view a clinical trial as part of a series of trials and to identify the optimal sample size for each trial by taking the whole series into account. The methodology used is a hybrid of frequentist and Bayesian approaches. It is assumed that at the end of each trial, the conventional frequentist method will be used to analyse the

observed responses and to test the hypotheses. The type I error for each trial is fixed but the type II error is not so that this is determined by the trial sample size. The type II error will be minimized which is equivalent to optimizing the power of each trial.

The next section introduces the statistical framework that is to be used in the designs. Section 5.2 presents the first proposed design which aims to maximize the number of successful trials. Following on, the design is extended by incorporating a start-up cost for each trial and its aim is to maximize the expected utility (Section 5.3). Finally, the design is generalized to a limiting setting where the size of the total population is not fixed. This is presented in Section 5.4. The chapter ends with some discussion and concluding remarks.

5.1 Assurance

The development in this section employs most of the notation and concepts introduced in Chapter 3. Throughout this chapter, the trials are assumed to be single-arm clinical trials. Let N be the known total number of patients eligible for trials. This value could be the estimated size of the patient population or in the case of small population diseases, it is not unreasonable to assume that it is known at least approximately. If n ($n \leq N$) patients are entered into each trial, then the number of trials that can be tried is, $K = N/n$.

Suppose that the primary endpoint is a continuous variable and denote X_k as the sample mean from n observations from trial k, $k = 1, 2, \ldots, K$.

Let the sample mean be normally distributed with an unknown mean θ_k and known variance σ^2/n which can also be stated as $X_k \sim N(\theta_k, \sigma^2/n)$. The unknown parameter θ_k is assumed to be random and a common and convenient choice of its prior distribution is the normal distribution. Let $\theta_k \sim N(\mu, \tau^2)$ where μ and τ are some fixed known parameters.

Upon observing the n responses at the end of each trial the frequentist method is used to test the null hypothesis. For a single-arm trial, the true mean θ_k is compared with the historical value θ_0 and the size of the hypothesis is set at $\alpha/2$. Thus, the null hypothesis is

$$H_0 : \theta_k = \theta_0.$$

From equation (3.19), the assurance (the average power) of a trial is

$$A(n) = 1 - \Phi\Big(\frac{z_{1-\alpha/2} - \sqrt{n/\sigma^2}(\mu - \theta_0)}{\sqrt{1 + n\tau^2/\sigma^2}}\Big),$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution function and z_γ is the lower 100γ percentile of the standard normal distribution. The assurance is a function of n and although in practice n is an integer, in this chapter, as an idealization, n is assumed to be continuous. If we let $n \to 0$,

$$\lim_{n\to 0} A(n) = 1 - \Phi(z_{1-\alpha/2}) = \alpha/2,$$

half of the specified two-sided type I error rate. This suggests that there is a minimum average power, albeit very small, that can be attained when the sample size goes to 0, which means that we may still benefit from a successful

trial without even starting a trial!

As $n \to \infty$,

$$\lim_{n\to\infty} A(n) \approx \lim_{n\to\infty} 1 - \Phi\left(\frac{z_{1-\alpha/2} - \sqrt{n/\sigma^2}(\mu - \theta_0)}{\sqrt{n\tau^2/\sigma^2}}\right) = \Phi\left(\frac{\mu - \theta_0}{\tau}\right),$$

which is the prior probability that $\theta_k > \theta_0$. Therefore, a trial with infinite sample size will lead to rejection of H_0 whenever $\theta_k > \theta_0$. This is in contrast to the power of a trial which approaches 1 when $n \to \infty$. Therefore, if the prior belief is positive, the assurance will be high and if the prior is negative, the assurance will be low.

5.2 Maximization of the number of successful trials

The assurance can be interpreted as the average probability of rejecting H_0 over all possible values of the parameter of interest based on the prior density. Let $\widetilde{K}(n)$ denote the number of trials that reject H_0 where each trial is of size n and the total sample size is constraint to be at most N. The first design that we are proposing is to find the optimal number of patients per trial, n^*, that maximizes the expected number of trials that reject H_0 when H_0 is not true, $E(\widetilde{K}(n))$. Under the model that we assume for this chapter,

the expected number of trials that reject H_0 is simply,

$$\begin{aligned} E(\widetilde{K}(n)) &= KA(n) \\ &= \frac{N}{n}\left(1 - \Phi\left(\frac{z_{1-\alpha/2} - \sqrt{n/\sigma^2}(\mu - \theta_0)}{\sqrt{1 + n\tau^2/\sigma^2}}\right)\right). \end{aligned} \tag{5.1}$$

Considering n as a continuous variable, the assurance can be maximized by differentiation. In theory, the maximum value of $E(\widetilde{K}(n))$ is obtained by differentiating $E(\widetilde{K}(n))$ with respect to n and subsequently solve for n by equating the derivative to 0. The derivative of $E(\widetilde{K}(n))$ is presented in Appendix B. However, the solution to $dE(\widetilde{K}(n))/\,dn = 0$ is not tractable and thus can only be solved numerically.

As seen in the preceding section, the sample size n is finite and also the assurance is bounded between 0 and 1 for $n > 0$. This implies that $E(\widetilde{K}(n))$ is also finite for all $n > 0$. Suppose that, $n \to N \to \infty$ then,

$$\lim_{n\to\infty} E(\widetilde{K}(n)) = \Phi\left(\frac{\mu - \theta_0}{\tau}\right).$$

Whereas as $n \to 0$, $N/n \to \infty$ and as we have seen earlier, the assurance, $A(n) \to \alpha/2$, a finite value. Consequently,

$$\lim_{n\to 0} E(\widetilde{K}(n)) \to \infty.$$

As there is no other value of n that could give greater value to $E(\widetilde{K}(n))$, the optimal sample size is thus $n^* = 0$, that is, the maximum of the expected number of successful trials is attainable by not recruiting any patient to the

phase II trials.

Corresponding to the examination of its properties, consider an example where the variance of the sample mean is $\sigma = 5$, and the parameters of the priof belief are $\mu = 1$ and $\tau = 1$. Under the null hypothesis, $H_0 : \theta_k = \theta_0 = 0$, Figure 5.1 shows that to maximize the number of trials that conclude that the treatments are promising when they are showing efficacy, we should have many small individual trials. In fact, if n is fixed to have only non-zero integer value instead of any real positive value, each individual trial should have only one patient. By having only one patient per trial, naturally more drugs can be put on trial and consequently, increase the expected total number of trials that reject H_0 since there is positive probability of rejection of H_0 even with a very small sample size.

It appears from Figure 5.1 that $E(\widetilde{K}(n))$ is a monotonic decreasing function of n. However, upon closer examination, this is not so. For some very small type I error rates, there is a local maximum at $n > 0$, illustrated in Figure 5.2. However, as it is not practical to design trials with very small type I error, this interesting property is not to be examined further in this thesis.

5.3 Maximization of the expected utility

So far we have ignored the start-up cost per trial and hence from the preceding result, we "get something for nothing" and as such it leads to the very small optimal sample size. In practice, there is a start-up cost associated with each clinical trial. By having as many trials as the total number of

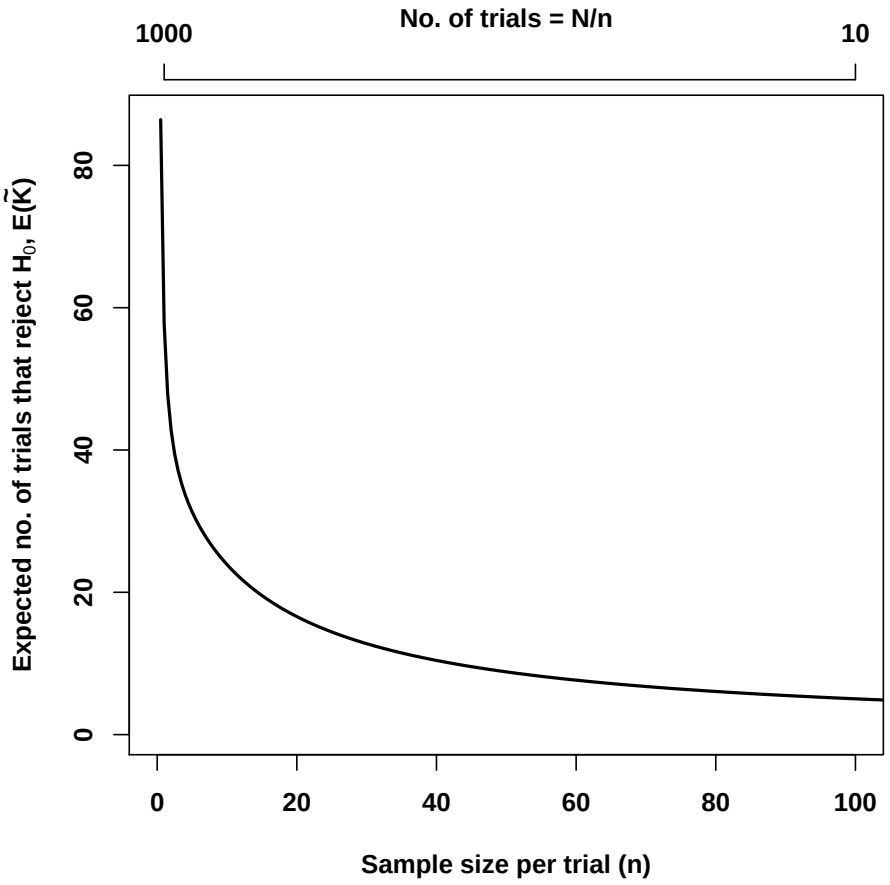
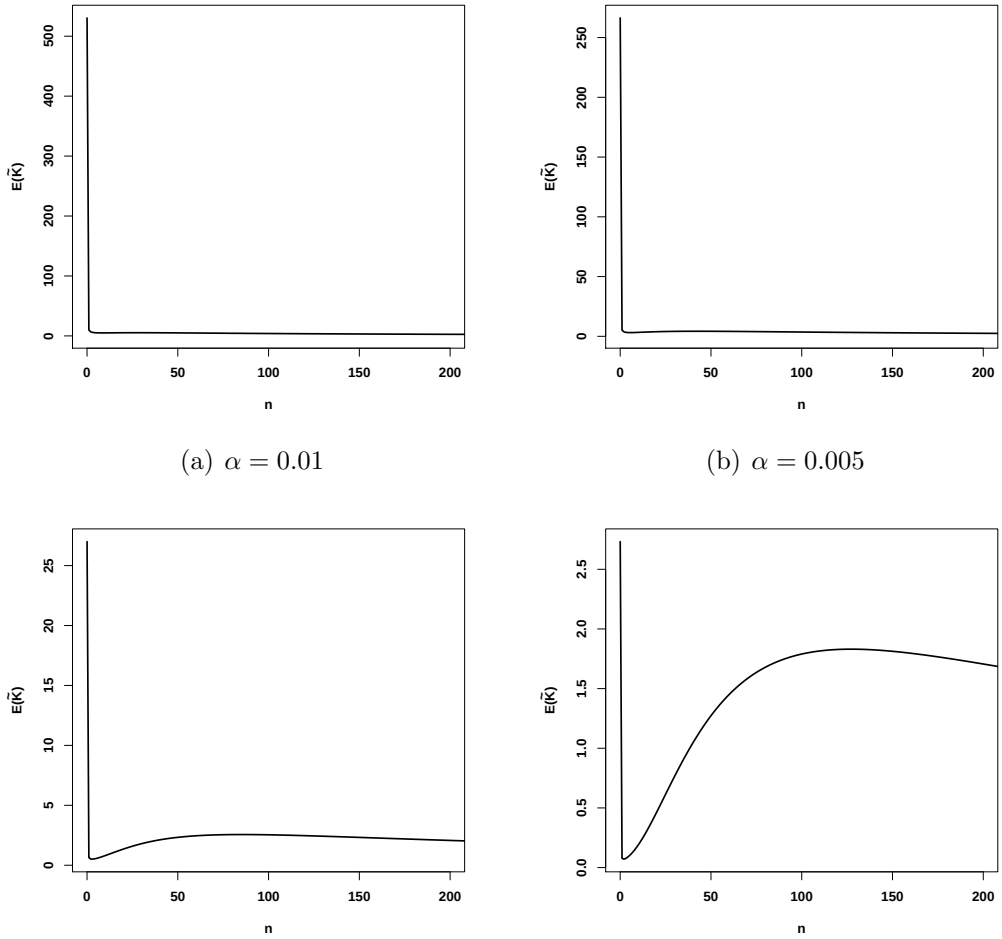


Figure 5.1: The expected number of trials that reject H_0 as a function of n.

patients available, the total start-up cost will be greatly inflated. The second design that is to be discussed in this section is to extend the first design that maximizes the number of successful trials (Section 5.2) by considering the start-up cost. The start-up cost could be the money or time spent on planning, designing, submitting for ethics approval, and so on.

Suppose that one unit of gain is assigned to each successful trial (rejecting H_0 correctly), then the total expected gain is $E(\widetilde{K}(n))$ as before. Let l_{II} be the fixed start-up cost which is relative to one unit of gain. The total start-up



(a) $\alpha = 0.01$ (b) $\alpha = 0.005$

(c) $\alpha = 0.0005$ (d) $\alpha = 0.00005$

Figure 5.2: The expected number of trials that reject H_0 under the fixed error rates, (a) $\alpha = 0.01$, (b) $\alpha = 0.005$, (c) $\alpha = 0.0005$, and (d) $\alpha = 0.00005$.

cost for all trials is then $l_{\mathrm{II}}K$. The expected utility is thus,

$$\begin{aligned} \mathcal{G}(n) &= E(\widetilde{K}(n)) - l_{\mathrm{II}}K \\ &= \Big(1 - \Phi(f(n)) - l_{\mathrm{II}}\Big)\frac{N}{n}, \end{aligned} \tag{5.2}$$

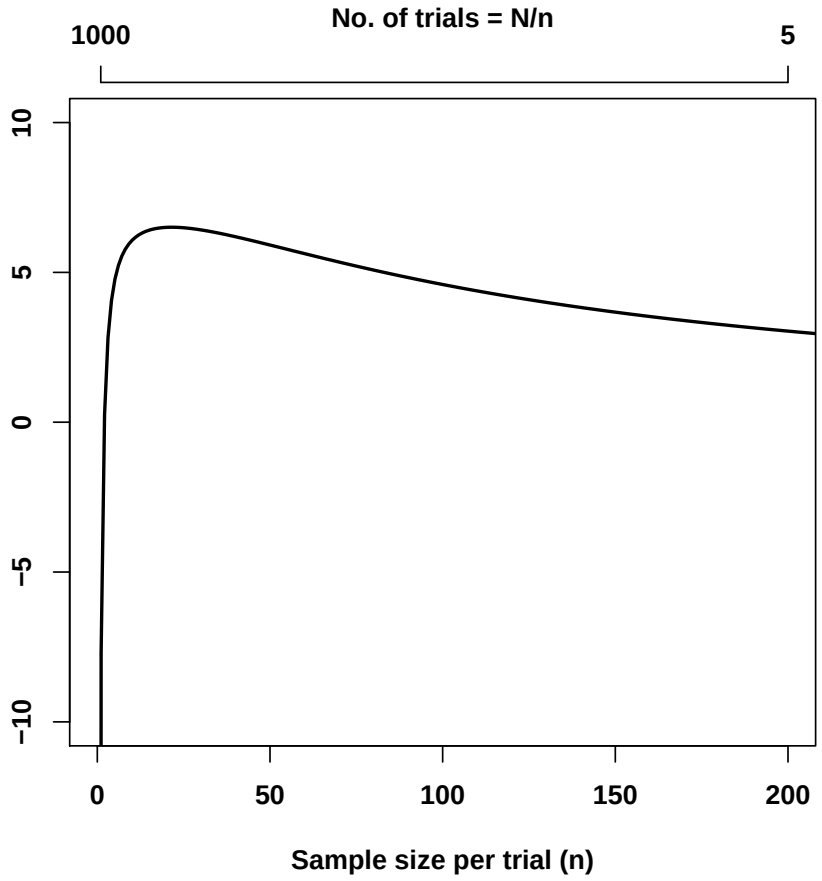


Figure 5.3: The expected utility, $\mathcal{G}(n)$ as a function of n.

where $f(n) = \sigma(z_{1-\alpha/2} - \sqrt{n/\sigma^2}(\mu - \theta_0))/(\sqrt{1+n\tau^2})$. Note that the cost for each patient is excluded from the start-up cost because the total number of patients is known and fixed. Hence, the total patients cost is a constant and will not affect the optimization of the design.

The optimization methodology is to find an n that maximizes the expected utility, $\mathcal{G}(n)$. Similarly, to find the maximum of $\mathcal{G}(n)$ is to differentiate

it with respect to n,

$$\frac{d}{dn}\mathcal{G}(n) = -f'(n)\phi(f(n))\frac{N}{n} - \Big(1 - \Phi(f(n)) - l_{\text{II}}\Big)\frac{N}{n^2}. \tag{5.3}$$

Subsequently, solving for n by equating the derivative to 0. However, due to the cumulative distribution function term in the derivative, it can only be solved numerically. The search for the optimal n^* that maximizes the expected utility can alternatively be made by a direct computation of $\mathcal{G}(n)$ and the n that corresponds to the smallest value of $\mathcal{G}(n)$ is then taken as the optimal n^*. We do this for a range of values for n, for example, from 0.01 to N by an increment of 0.01.

There is a range of optimal solutions for different combinations of cost, parameter of likelihood function and *a priori* knowledge of the unknown parameters. Table 5.1 presents some of the optimal sample sizes for different parameters such that the standardized effect sizes (μ/σ) are 0.2 ("small" effect size), 0.5 ("moderate" effect size) and 0.8 ("large" effect size). The null hypothesis is, $H_0 : \theta_k = 0$ and the two-sided type I error is fixed at 0.05. Contrary to the $E(\widetilde{K}(n))$ function whose optimal solution is at $n^* = 0$, the $\mathcal{G}(n)$ function is concave for most combinations (for example, see Figure 5.3 where $\theta_0 = 0$, $\sigma = 7.5$, $\mu = 1.5$, $\tau = 1$). An example, if $\theta_0 = 0$, $\sigma = 5$, $\mu = 1$, $\tau = 1$, and the start-up cost of a trial is $l_{\text{II}} = 0.05$, the optimal solution is $n^* = 14.83$ (see Table 5.1). Rounding up to the nearest integer, the sample size needed for each trial is 15.

Note that when the parameters $(\mu = 1, \tau = 1, \sigma = 5)$ are all multiplied by a constant, for example, by 2 to give $(\mu = 2, \tau = 2, \sigma = 10)$, the n^* remains

Table 5.1: The optimal sample sizes per trial, n^*, and expected utility, $\mathcal{G}(n)$, of the design that maximizes $\mathcal{G}(n)$.

Effect size	σ	μ	τ	n^*	$\mathcal{G}(n)$
0.2	5	1	1	14.83	8.290
	5	1	2	6.38	16.154
	5	1	5	1.51	59.356
	7.5	1.5	1	21.52	6.507
	7.5	1.5	2	10.78	10.520
	7.5	1.5	5	2.96	31.805
	10	2	1	26.64	5.813
	10	2	2	14.83	8.290
	10	2	5	4.62	21.350
0.5	2	1	1	2.37	51.813
	2	1	2	1.02	100.963
	2	1	5	0.24	370.974
	3	1.5	1	3.44	40.666
	3	1.5	2	1.73	65.748
	3	1.5	5	0.47	198.779
	4	2	1	4.26	36.334
	4	2	2	2.37	51.813
	4	2	5	0.74	133.438
0.8	1.25	1	1	0.93	132.642
	1.25	1	2	0.4	258.465
	1.25	1	5	0.09	949.198
	1.875	1.5	1	1.34	104.104
	1.875	1.5	2	0.67	168.313
	1.875	1.5	5	0.19	508.797
	2.5	2	1	1.67	93.014
	2.5	2	2	0.93	132.642
	2.5	2	5	0.29	341.601

the same, that is, $n^* = 14.83$. This implies that the optimal sample size n^* is robust towards scaling of the unit of measurement. For example, if the original measurement is in inches and it has now changed to centimetres, the n^* will remain the same.

Table 5.2: Optimal sample sizes per trial, n^*, expected gain, $E(\widetilde{K}(n))$, and expected utility, $\mathcal{G}(n)$, for different start-up costs, l_{II}.

l_{II}	n^*	$E(\widetilde{K})$	$\mathcal{G}(n^*)$
0.001	0.01†	2621.561	2521.561
0.01	0.01†	2621.561	1621.561
0.02	0.01†	2621.561	621.561
0.03	0.84	47.942	12.228
0.04	9.99	13.098	9.094
0.05	14.84	11.659	8.290
0.1	29.45	9.396	6.000
0.2	55	7.181	3.544
0.3	87.11	5.538	2.094
0.4	135.07	4.126	1.164
0.5	218.61	2.860	0.573
0.6	398.42	1.730	0.224
0.7	985.06	0.763	0.052
0.8	0	0.752	−0.048
0.9	0	0.752	−0.148
0.99	0	0.752	−0.238

† In the direct search algorithm 0.01 was used as the minimum for n to start off the search. Hence, the minimum value that n^* can reach is 0.01.

From Table 5.1, the expected utility increases as τ increases while σ is held constant. When τ gets larger, the variance of the prior belief is wider. Thus, smaller sample sizes are required to give information regarding the θ_k of each trial. In contrast to what has been observed from the increment of τ, $\mathcal{G}(n)$ decreases as σ increases while τ is held constant. This property conforms to the idea that as the standard deviation of the likelihood function gets wider, a larger sample size is required to provide more information on the parameter of interest, θ_k.

So far, we have shown that the expected utility function has a local maximum. However, there exists some situations where no optimal solution is

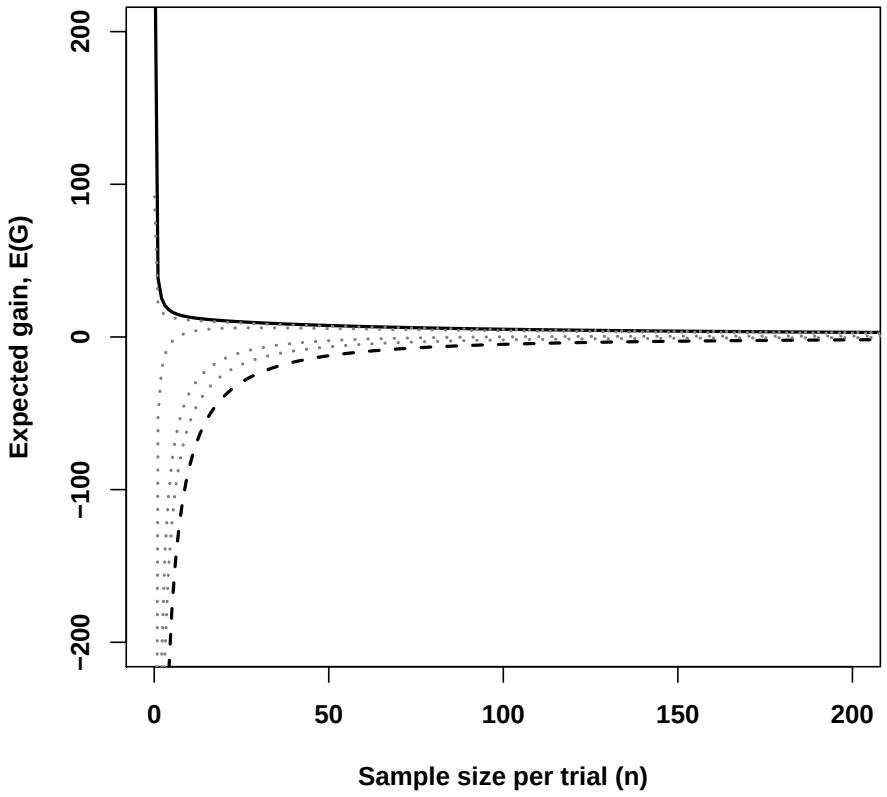


Figure 5.4: The expected utility as a function of n for $l_{\text{II}} = 0.001$ (dark solid line), and $l_{\text{II}} = 0.99$ (dark dashed line). The other grey dotted lines in between are various l_{II} values at 0.02, 0.1, 0.5, and 0.7.

found. If the start-up cost, $l_{\text{II}} \to 0$, the total cost for starting up trials, $l_{\text{II}}K$, is negligible and thus,

$$\lim_{l_{\text{II}} \to 0} \mathcal{G}(n) = E(\widetilde{K}(n).$$

The expected utility now is the same as the maximization of the number of successful trials (Section 5.2). Therefore, the optimal solution is $n^* = 0$. Table 5.2 presents some of the optimal solutions when l_{II} takes on various

values. When $l_{\text{II}} = 0.01$, $n^* = 0.01$ which is the minimum value used in the direct search algorithm. By modifying the minimum value in the direct search algorithm n^* changes and takes on that minimum value (result not shown). The illustration of the change of the expected utility as a function of n as l_{II} takes on various values is presented in Figure 5.4.

On the other hand, as l_{II} approaches 1, the total start-up costs, $l_{\text{II}}K$, will be greater than the expected gain. From equation (5.2),

$$\begin{aligned}\lim_{l_{\text{II}}\to 1} \mathcal{G}(n) &= A(n)K - K \\ &= \Big(1 - \Phi(f(n)) - 1\Big)K \\ &= -\Phi(f(n))K.\end{aligned}$$

As $\Phi(f(n))$ and K are always greater than 0 for all $n > 0$, the expected utility, $\lim_{l_{\text{II}}\to 1} \mathcal{G}(n) < 0$, so that it is not worth starting any trial at all.

5.4 Minimization of the expected loss

So far, our designs have had a constraint of a fixed total sample size N. Usually the availability of eligible patients is a continuous progress and as argued by Yao *et al.* (1996) it is more practical to take away the constraint on the total sample size. Suppose that the clinical trials are conducted sequentially, the objective is to recommend the first trial that is successful for further testing. When the first successful trial is obtained, the series can be considered to end and another series of trials to start.

Define T as the total number of trials up to and including the first suc-

cessful trial so that T is a random variable. Each trial is independent of each other and the prior probability that a trial is declared successful is equal to $A(n)$, the assurance as introduced earlier. The random variable T thus follows a geometric distribution,

$$T \sim Ge(A(n)),$$

with $\Pr(T = t)$ given by $(1 - A(n))^{t-1}A(n)$. The expected number of trials required to give one successful trial is $1/A(n)$. Denoting the start-up cost per trial as l_{II}, then the average total start-up cost for all trials up to and including the first successful trial is

$$\frac{l_{\text{II}}}{A(n)}.$$

Suppose that there are n patients in each trial, the expected total number of patients till a successful trial is found is,

$$E(N) = \frac{n}{A(n)}.$$

As the total number of patients to be required is not fixed, the total cost needed to be spent on patients is not fixed either. Let l be the cost per patient which is also relative to the one unit of gain, then the expected total cost of patients will be,

$$lE(N) = \frac{nl}{A(n)}.$$

Therefore, the expected total cost that would be spent till a successful trial

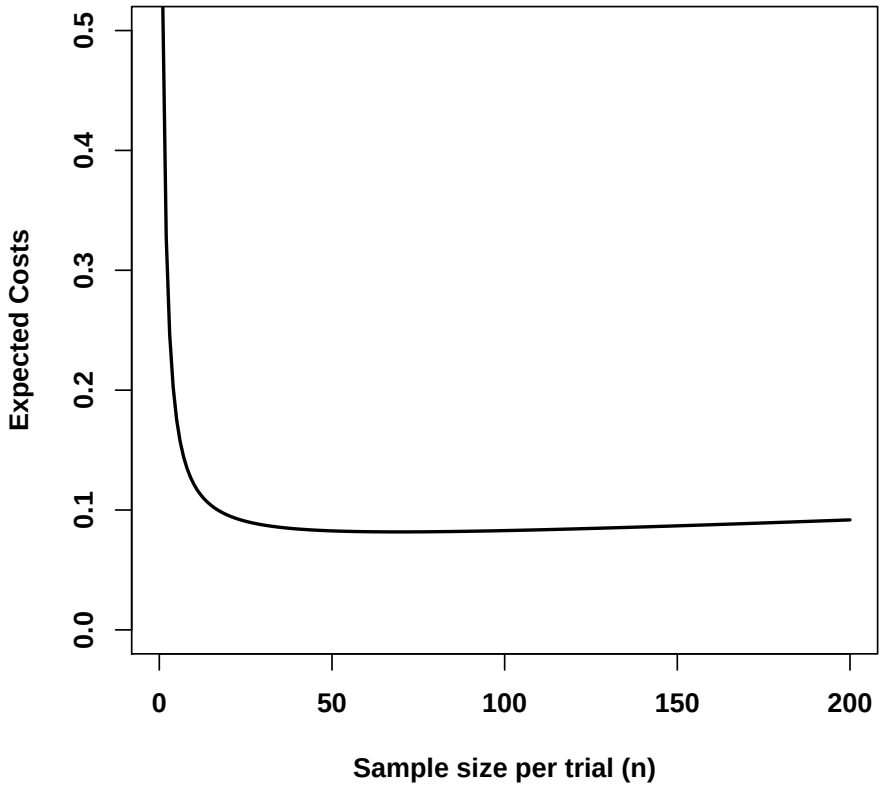


Figure 5.5: The expected loss, $\mathcal{L}(n)$, as a function of n, for $l_{\text{II}} = 0.05$ and $l = 0.0001$.

is declared is

$$\mathcal{L}(n) = \Big(l_{\text{II}} + nl\Big)\frac{1}{A(n)}. \tag{5.4}$$

The optimization problem is to find an n that minimizes the expected loss, $\mathcal{L}(n)$. The evaluation is done numerically, similar to the approach presented in the preceding section. Figure 5.5 shows the expected loss function of conducting a series of trials as a function of n for $l_{\text{II}} = 0.05$ and $l = 0.0001$.

There is a range of optimal solutions for various combinations of costs,

parameter of the likelihood function and *a priori* knowledge of the unknown parameter. Table 5.3 presents some of the optimal sample sizes for a few combinations of parameters such that the effect sizes (μ/σ) are 0.2, 0.5, and 0.8. The null hypothesis is, $H_0 : \theta_k = 0$ and the two-sided type I error rate is fixed at 0.05. The start-up cost for a clinical trial is fixed at $l_{\text{II}} = 0.05$ and the cost per patient is $l = 0.0001$. For an example, the optimal sample size for a likelihood, $X_k \sim N(\theta_k, 25)$, and *a priori* knowledge of $\theta_k \sim N(1, 1)$, is $n^* = 163.38$. Rounding up to the nearest integer, the number of patients required for a trial is $n = 164$.

5.5 Discussion and concluding remarks

The objectives of the optimization of the series of designs incorporating costs are different from each other. The first design that incorporates cost (presented in Section 5.3) is to maximize the expected utility of *a number of successful trials* in a series of trials whereas the second design (presented in Section 5.4) is to minimize the expected loss till *a succesful trial* is found in a series of trials. Thus, for the same set of likelihood and prior parameters, for example, $\sigma = 5$, $\mu = 1$, and $\tau = 1$, the optimal solution under the formulation of the design that maximizes $\mathcal{G}(n)$ is $n^*_{\mathcal{G}} = 14.83$. Given that the total sample size is $N = 1000$, 67.43 trials can be conducted. On average, the number of successful trials is 11.66, which is about one trial out of every six trials. On the other hand, the optimal solution under the formulation of the design that minimizes $\mathcal{L}(n)$ is $n^*_{\mathcal{L}} = 163.38$. The expected total number of patients to be recruited under this design is $E(N) = 278.8$ and thus, the

Table 5.3: Optimal sample sizes per trial, n^*, and expected loss, $\mathcal{L}(n)$, of the design that minimizes $\mathcal{L}(n)$.

Effect size	σ	μ	τ	n^*	$\mathcal{L}(n)$
0.2	5	1	1	163.38	0.113
	5	1	2	108.67	0.119
	5	1	5	58.71	0.117
	7.5	1.5	1	197.73	0.106
	7.5	1.5	2	139.28	0.117
	7.5	1.5	5	77.55	0.119
	10	2	1	218.72	0.101
	10	2	2	163.38	0.113
	10	2	5	93.96	0.119
0.5	2	1	1	69.31	0.082
	2	1	2	50.53	0.094
	2	1	5	29.13	0.101
	3	1.5	1	77.07	0.074
	3	1.5	2	61.75	0.087
	3	1.5	5	37.63	0.099
	4	2	1	79.08	0.068
	4	2	2	69.31	0.082
	4	2	5	44.59	0.096
0.8	1.25	1	1	45.81	0.074
	1.25	1	2	34.84	0.087
	1.25	1	5	20.72	0.097
	1.875	1.5	1	48.68	0.066
	1.875	1.5	2	41.68	0.080
	1.875	1.5	5	26.44	0.094
	2.5	2	1	47.62	0.062
	2.5	2	2	45.81	0.074
	2.5	2	5	31.03	0.090

expected number of trials to be tried till a successful one is found is 1.71.

In this chapter we have proposed to design each trial as part of a series of trials so that the long-term gain will be the greatest. The first two proposed designs (presented in Sections 5.2 and 5.3) assume that the total population

is known, similar to the design proposed by Whitehead (1985). Although the clinical trials are planned as a group, there is a flexibility to either run the trials sequentially or concurrently. Each of the trials is independent from each other and if at any point at least one potential treatment has shown sufficient efficacy, it may be recommended for further testing regardless of the results of the others.

Patient recruitment is an ongoing and fluid process and it seems practical not to impose any constraint on the total sample size. As such, the design proposed in Section 5.4 assumes that N is not fixed. The objective is to run the trials sequentially and when the first trial is declared promising the running of the series of trials ceases and a new series of trials is initiated.

The traditional design of a clinical trial is based on data estimation and inference method. Subsequently, based on evidence of efficacy a decision is made if the treatment should be developed further or not, or be approved by the regulatory agency or not. However, this approach does not inform decision-makers (for example, regulatory and reimbursement agencies) on whether the experimental treatment is worthy to be approved based on current evidence of cost-effectiveness or if further evidence is required to eliminate uncertainty surrounding the adoption of a decision (Briggs *et al.*, 2006). Models based on decision theory and expected value of perfect information (EVPI) are common approaches proposed to aid such decision-making.

Both the traditional and EVPI approaches aim to accomplish different objectives and so do not seem to be connected. The models proposed in Sections 5.3 and 5.4 represent a hybrid of the traditional approach and a simplified EVPI method. Only a minimal set of economic parameters normally covered

by EVPI designs were incorporated into these models. However, they can be expanded to incorporate further economic parameters and probabilistic distributions to assess whether any additional evidence in the future is required to support the current decision.

Resources are necessarily finite and it may be worth considering a fixed population which is equivalent to fixing the budget when developing a portfolio of drugs. Thus far, the planning of the clinical trials presented in this chapter are in the context of a series of trials with the same setting. In the next chapter, the series of trials is modified and expanded for the special case of small population diseases where quite often the total population is known. The framework of the design is in the context of a series of phase II trial and it also incorporates the viability of the treatment in the phase III setting in the event that it is recommended for phase III based on its responses in the phase II. This is necessary to address the inherent issue of the limited population size.

Part III

Designs for a Series of Decision-Theoretic Trials

Beat by beat, losing inhibition
Head, hands, feet, trying to relax, but not too much
Trying to lay out the exposition
But without exposing it as such
Trying to perform but not audition
Trying to establish recognition
Trying to persuade the electrician
That he should destroy the competition

Stephen Sondheim

Putting It Together

Chapter 6

A Series of Decision-Theoretic Phase II Trials

The design of a series of trials in the preceding chapter is applicable for both phase II and III settings. This chapter extends and modifies the design by encompassing a series of decision-theoretic single-arm phase II trials and a randomized controlled two-arm phase III trial. The motivation is based on the scenario where a pharmaceutical company has developed a few new drugs for the same disease or a large public sector or charity organization has funding for a few clinical trials. It is assumed that the intended population is considerably small and there is an interest to find an effective treatment efficiently. Due to the smaller population it may not be feasible to try all the new drugs concurrently. A practical approach is to consider one trial at a time and if the treatment is recommended for a phase III trial then the other treatments are temporarily suspended from trials so that the remaining patients can be admitted to the phase III trial. Depending on the availability

of resources, drugs that have not been tried can be considered for future trials. These "leftover" drugs may be considered in a "new" series of trials. Alternatively, individual trials may be designed for these "leftover" drugs.

This chapter introduces a design that considers each of the phase II trial as part of a series of trials where the trials are conducted sequentially with interim decision-making. Section 6.1 introduces the statistical framework of the design. The design is illustrated with examples from treatments for persistent asthmatic patients. The new treatments belong to the same type of class, mediator antagonists, and this is shown in Section 6.3. The chapter concludes with concluding remarks.

6.1 Design

The formulation of the design is an extension of the designs proposed by Whitehead (1986) and Stallard (2012), reviewed in Chapter 4, to allow sequential decision-making in the phase II trials. Phase II trials are sometimes used as exploratory trials to rule out nonpromising treatments. Hence, it seems intuitive to use Bayesian decision-theoretic framework in phase II trials where decision can be made in order to prevent loss of resources due to inferential conclusion.

The methodology of the proposed design is based on the hybrid approach, namely, a combination of classical frequentist testing, Bayesian decision-theory and a prior distribution for the parameter that corresponds to the treatment effects for the experimental treatment. The design of the phase II trials is based on Bayesian decision-theoretic methodology. At each interim

stage, a decision is made from four possible actions:

Action R Resume the current phase II trial by recruiting more patients,

Action P Stop the current phase II trial and proceed to a two-arm phase III trial where the experimental treatment will be tried against a control treatment with the remaining patients,

Action T Stop the current phase II trial and try a new treatment in a new single-arm phase II trial using the remaining patients; or

Action A Stop the current phase II trial and abandon the development plan.

If the decision is to take action P, then the recommended experimental treatment is tested in a phase III trial with a control treatment that is either a standard treatment or a placebo. At the end of the trial, the data from the phase III trials are tested using the classical frequentist analysis, that is, the point estimate is reported with its confidence interval and p-value to infer the true parameter.

The proposed design assumes that the size of the population and the number of treatments available for trials are fixed and known. In addition, both phase II and III trial are assumed to have the same binary primary endpoint.

6.1.1 Notation

Let N denote the fixed and known total population size eligible for trials. Also, let K denote the fixed and known number of potential new treatments

available for trials for the population. Patients are recruited sequentially to each phase II trial and a decision is made based on the accumulated responses at every interim stage. At the i-th stage of k-th trial, $k = 1, 2, \ldots, K$, m_{ki} patients are recruited. As the sampling is done fully sequentially, $m_{ki} = 1$. The variable m_{ki} can however, take on any integer value but the proposed design is illustrated as a fully sequential trial.

Let a success response from a patient be represented numerically by 1 and a failure by 0. Let X_{ki} denote the sum of successes from the i-th stage of the k-th trial then the cumulative successes from the first stage up to and including the current i-th stage of k-th trial is $S_{ki} = X_{k1} + X_{k2} + \ldots + X_{ki}$. Patients' responses are assumed to be independent from each other and as such S_{ki} is assumed to follow a binomial distribution with index $n_{ki} = m_{k1} + m_{k2} + \ldots + m_{ki}$ and unknown probability of success, p_k, where n_{ki} is the total number of patients recruited from the first stage up to and including the i-th stage of k-th trial. The likelihood function of S_{ki} is

$$f_{S|p}(s_{ki}|p_k) = \binom{n_{ki}}{s_{ki}} p_k^{s_{ki}} (1 - p_k)^{n_{ki} - s_{ki}}.$$

The unknown probabilities of success, $p_1, p_2, \ldots, p_K$ are assumed to be independent identically distributed random variables and follow a known parametric distribution. Its density is denoted by $f_p(p_k)$. The prior density summarizes the degree of belief about the efficacy of the treatments before any responses have been observed. Upon observing patients' responses from the first i stages of the k-th trial, the prior is updated and the posterior density is now a combined belief of the prior probability given the observed

responses. Having observed $S_{ki} = s_{ki}$, according to the Bayes' theorem, from (3.7) the posterior density is

$$f_{p|S}(p|s_{ki}) = \frac{f_{S|p}(s_{ki}|p)f_p(p)}{f_S(s_{ki})},$$

where

$$f_S(s_{ki}) = \int_0^1 f_{S|p}(s_{ki}|p_k)f_p(p_k)\,dp_k,$$

is the marginal density of S_{ki}.

A convenient and common choice for p_k prior distribution is the beta distribution as it is a conjugate family for binomial data. Assuming that the parameter p_k follows the beta distribution with known parameters a_k and b_k, for $k = 1, 2, \ldots, K$. The probability density function is $f_p(p_k) = p_k^{a_k-1}(1-p_k)^{b_k-1}/B(a_k, b_k)$ where $B(a_k, b_k) = \Gamma(a_k)\Gamma(b_k)/\Gamma(a_k + b_k)$ is the beta function. From (3.11), the marginal density of S_{ki} is

$$f_S(s_{ki}) = \binom{n_{ki}}{s_{ki}} \frac{B(a_k + s_{ki}, b_k + n_{ki} - s_{ki})}{B(a_k, b_k)}$$

which is a beta-binomial distribution function with index n_{ki}, and parameters a_k, and b_k. As shown earlier in equation (3.12), the posterior density function of p_k given s_{ki} is

$$f_{p|S}(p|s_{ki}) = \frac{1}{B(a_k + s_{ki}, b_k + n_{ki} - s_{ki})} p_k^{a_k+s_{ki}-1}(1-p_k)^{b_k+n_{ki}-s_{ki}-1},$$

which is a beta distribution with parameters $a_k + s_{ki}$ and $b_k + n_{ki} - s_{ki}$.

6.1.2 Expected utility

The objective of the series of decision-theoretic trials is to derive a decision-making tree for each phase II trials given the constraints of a fixed and known population and number of potential treatments by optimizing the long term expected gain of the development plan. Let $\mathcal{G}_{\text{Total}}(N, K)$ be the expected utility of the optimum development plan with total population N and K potential treatments. The expected gain of the development plan is relative to a fixed reference value that is constant throughout the decision making process (Hilden, 1990, Pratt *et al.*, 1995, Ch. 4). It is easier to fix the reference point to be the current asset position and any utility function is measured from it. There are some costs involved in conducting clinical trials and recruiting patients to the trials. At the initial stage of the development plan, the gain function of not starting any trial at all is taken to be zero because no cost is spent. If however, the information from patients' responses are worthier than the cost of sampling then it is warranted to conduct the trial. Therefore, the optimal development plan may start to recruit its first batch of m_{11} patients if the expected value of the information less the cost of conducting and sampling is greater than zero, that is, $\mathcal{G}_{\text{Total}}(N, K) > 0$.

There are two types of economic factors that are involved in the utility function, namely, cash outflows and profits. For simplicity, only the cash outflows is considered in the proposed design. The profits, for example, from the sales of the new treatment if the trial is a success, is not considered. There are two types of costs, fixed and variable costs (Patel and Ankolekar, 2007). The fixed costs are the costs of setting up and running the trials.

These costs are incurred once the trial is committed to go on and are not dependent on the size of the trial. Let l_{II} and l_{III} be the costs of setting up and conducting a phase II and III trials, respectively. These costs may reflect the advertising and recruitment of potential patients, the setting up of the necessary equipments, periodic audits and site visits by clinical research assistants, the management of the data, *etc.*

The variable cash outflow is the cost that depends on the size of the trial. This cost may reflect the costs of the medical supplies, screening tests, clinical and biochemical tests during the intervention and follow-up periods, *etc.* This cost is known as the cost per patient and it is denoted by l. For simplicity it is assumed that the cost per patient is equal in phase II and III trials. The non-negative costs l, l_{II} and l_{III} are taken to be relative to the one unit of gain which is given by a successful phase III trial. The details of the gain of a successful trial is given in the section where the expected utility of action P is derived.

The decision whether to choose actions R, P, T or A at each interim analysis depends on the desirability of each action which in turn depends on the true parameter of the experimental treatment, the total number of patients and the number of potential treatments. During the k-th trial, let $G_a(p_k, N_k, K_k)$ denote the utility function of action a, $(a \in \{R, P, T, A\})$ where p_k is the true parameter of the experimental treatment, N_k is the total population and K_k is the number of potential treatments at the start of this trial. As the unknown parameter follows some known distribution, the utility function of action a is an expectation over all possible values of p_k. Having observed $S_{ki} = s_{ki}$ successes from n_{ki} patients from the first i stages of k-th

trial, the expected utility of action a is the expectation over the posterior density of p_k given s_{ki} and n_{ki} is

$$\begin{aligned}\mathcal{G}_a(s_{ki}, n_{ki}, N_k, K_k) &= E(G_a(p_k, N_k, K_k)|s_{ki}, n_{ki}) \\ &= \int_0^1 G_a(p_k, N_k, K_k) f_{p|S}(p_k|s_{ki}, n_{ki})\, dp_k. \end{aligned} \tag{6.1}$$

Note that the expected utilities in the k-th trial depend on the observations from that trial only as $p_1, \ldots, p_K$ are assumed to be independent.

Expected utility of the whole development plan

The development plan is consisted of a series of sequential phase II trials and it commences with the first phase II trial by recruiting $n_{11} = m_{11}$ patients to it. Of this, $S_{11} = s_{11}$, successes are observed. The expected utility of the whole development plan is obtained by first, calculating the expected utilities of each action. Secondly, the expected utilities are compared. Then the maximized expected utility is averaged over all possible values of s_{11} less the cost of conducting a phase II trial.

Let N_1 and K_1 be the population size and number of potential treatments, respectively, at the start of the first phase II trial of the development plan, that is, $N_1 = N$ and $K_1 = K$. Thus, the expected utility of the whole development programme is

$$\mathcal{G}_{\text{Total}}(N_1, K_1) = \sum_{s_{11}=0}^{n_{11}} \max_{a \in \{R,P,T,A\}} \{\mathcal{G}_a(s_{11}, n_{11}, N_1, K_1)\} f_S(s_{11}) - l_{\text{II}}, \tag{6.2}$$

where $\mathcal{G}_a(s_{11}, n_{11}, N_1, K_1)$ is the expected utility of action a (as given in

equation (6.1)), $f_S(s_{11})$ is the marginal density of S_{11}, a beta-binomial density function with index n_{11} and parameters a_1 and b_1 (as shown above).

Expected utility of action A (abandon the programme)

At the i-th stage of the first trial, a total of $n_{1i} = \sum_{j=1}^{i} m_{1j}$ patients have been recruited and of this, $S_{1i} = \sum_{j=1}^{i} X_{1j}$ successes are observed. The utility function of doing nothing is fixed as zero, the reference point, so the utility function of action A is less the cost of patients recruited in the first i stages of the first trial. Let s_{1i} be the observed value then the utility function is $G_A(p_1, N_1, K_1) = -ln_{1i}$. Taking expectation by integrating it with respect to p_1 over all its possible values gives

$$\mathcal{G}_A(s_{1i}, n_{1i}, N_1, K_1) = -ln_{1i}. \tag{6.3}$$

Expected utility of action P (proceed to phase III trial)

If the optimal action is to take action P upon observing patients' responses from the first i stages of the first trial, then the remaining number of patients, $n_{\text{III}} = N_1 - n_{1i}$ are recruited to the randomized controlled two-arm phase III trial. It is assumed that treatment allocation is in a 1:1 ratio, therefore, $n_{\text{III}}/2$ patients per arm. The randomization list can be prepared either by a simple or a block randomization method (Pocock, 1983, Ch. 5). The randomization technique, however, will not be discussed in this thesis.

The utility of action P depends on the reward of the phase III trial and the cost of conducting it. The primary endpoint of the phase III trial is assumed to be binary. At the end of the phase III trial, a frequentist analysis

is done on the observed responses from this trial to test whether the efficacy of the experimental treatment is different than the control treatment. The null and alternative hypotheses are written as

$$H_0 : p_1 = p_C \quad \text{against} \quad H_1 : p_1 \neq p_C,$$

where p_C is the true probability of success of the control treatment. This value is assumed to be fixed and known. The size of the two-sided hypothesis test will be denoted by α.

As discussed in 4.1.2, there are a few methods used to summarize the binary effects between two treatments. Following the design of Whitehead (1986), the log odds ratio measurement is used. Let $\theta_1 = \log\{p_1(1-p_C)/(p_C(1-p_1))\}$ denote the log odds ratio, then the null hypothesis is rewritten as $H_0 : \theta_1 = 0$. From equation (4.12) the score statistic is

$$B = \frac{n_{\text{III}} S_E/2 - n_{\text{III}} S_C/2}{n_{\text{III}}} = \frac{S_E - S_C}{4},$$

where S_E and S_C are the sums of successes from the experimental and control treatments, respectively. The score statistic B is approximately normally distributed with mean $\theta_1 V_1$ and variance V_1 where V_1 is the Fisher's information,

$$V_1 \approx n_{\text{III}} \bar{p}(1-\bar{p})/4,$$

where $\bar{p} = (p_1 + p_C)/2$ assuming that the sample size is sufficiently large. The test statistic under H_0 is $B/\sqrt{V}$.

The phase III trial is considered a success if the null hypothesis is rejected

correctly, that is, concluding that the experimental treatment is more effective than the control arm when it is true. A success has thus, one unit of gain. The probability that the null hypothesis is rejected at the upper end of the level of significance under the alternative hypothesis is

$$1 - \Phi\Big(z_{1-\alpha/2} - \theta_1\sqrt{V_1}\Big), \tag{6.4}$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution function and z_γ is the lower 100γ percentile of the standard normal density.

The reward of identifying a successful treatment correctly has one unit of gain and so the gain of the phase III trial is the probability of success which is given by the power function. It is shown from equation (6.4) the power function is a function of p_1, the true probability of success of the first treatment. Therefore, the gain of action P is the gain of the phase III trial less the cost of all N_1 patients and the cost of conducting a phase III trial,

$$G_P(p_1, N_1, K_1) = 1 - \Phi\Big(z_{1-\alpha/2} - \theta_1\sqrt{V_1}\Big) - lN_1 - l_{\text{III}}.$$

Taking expectation by integrating it with respect to p_1 over its posterior density given $S_{1i} = s_{1i}$ successes from n_{1i} patients gives

$$\begin{aligned}&\mathcal{G}_P(s_{1i}, n_{1i}, N_1, K_1)\\ &\quad= \int_0^1 \left(1 - \Phi\Big(z_{1-\alpha/2} - \theta_1\sqrt{V_1}\Big) - lN_1 - l_{\text{III}}\right) f_{p|S}(p_1|s_{1i}, n_{1i})\, dp_1. \end{aligned}\tag{6.5}$$

From (3.23), the expected power which is the assurance (O'Hagan and Stevens,

2001) is shown to be

$$A(s_{1i}, n_{1i}, N_1, K_1) = \int_0^1 \Big(1 - \Phi\big(z_{1-\alpha/2} - \theta_1\sqrt{V_1}\big)\Big) f_{p|S}(p_1|s_{1i}, n_{1i})\, dp_1.$$

Therefore, (6.5) is simplified to

$$\mathcal{G}_P(s_{1i}, n_{1i}, N_1, K_1) = A(s_{1i}, n_{1i}, N_1, K_1) - lN_1 - l_{\text{III}}. \tag{6.6}$$

Expected utility of action R (resume the current phase II trial)

The expected utility of action R is not as straightforward to find as for actions A and P. Its expectation is made up of the assumption of resuming the current trial by recruiting more patients and consequently take the best possible decisions from then onwards (Lindley, 1961). At the i-th interim analysis of the first trial, $n_{1i} = \sum_{j=1}^{i} m_{1j}$ patients would have been recruited and of this, $S_{1i} = s_{1i}$ successes are observed. The gain function of action R depends on the action taken after observing the responses from $m_{1,i+1}$ patients. Upon observing $X_{1,i+1} = x_{1,i+1}$ successes, the optimal decision at the $(i+1)$-th stage is the action with the highest expected utility given $s_{1i} + x_{1,i+1}$ which is

$$\max_{a\in\{R,P,T,A\}} \{\mathcal{G}_a(s_{1i} + x_{1,i+1}, n_{1i} + m_{1,i+1}, N_1, K_1)\}.$$

Therefore, the expected utility of action R at the i-th stage depends on the reward after recruiting the additional $m_{1,i+1}$ patients averaged over all

the possible responses that may be observed given s_{1i}, that is

$$\begin{aligned}&\mathcal{G}_R(s_{1i}, n_{1i}, N_1, K_1)\\ &= \sum_{x_{1,i+1}=0}^{m_{1,i+1}} \max_{a\in\{R,P,T,A\}} \{\mathcal{G}_a(s_{1i}+x_{1,i+1}, n_{1i}+m_{1,i+1}, N_1, K_1)\} f_S(x_{1,i+1}|s_{1i}, n_{1i}),\end{aligned} \tag{6.7}$$

where $f_S(x_{1,i+1}|s_{1i}, n_{1i})$ is the marginal density of $X_{1,i+1}$ given $S_{1i} = s_{1i}$.

Patients' responses are assumed to be independent from each other and so the cumulative successes, $X_{1,i+1}$, follows a binomial distribution with index $m_{1,i+1}$ and the unknown parameter p_1 which now follows the posterior beta distribution, that is, a beta distribution with parameters $(a_1 + s_{1i})$ and $(b_1 + n_{1i} - s_{1i})$. The marginal density of $X_{1,i+1}$ given $S_{1i} = s_{1i}$ is thus,

$$\begin{aligned}&f_S(x_{1,i+1}|s_{1i}, n_{1i})\\ &\quad= \int_0^1 f_{S|p}(x_{1,i+1}|p_1) f_{p|S}(p_1|s_{1i}, n_{1i})\, dp_1\\ &\quad= \binom{m_{1,i+1}}{x_{1,i+1}} \frac{B(a_1+s_{1i}+x_{1,i+1}, b_1+n_{1i}+m_{1,i+1}-(s_{1i}+x_{1,i+1}))}{B(a_1+s_{1i}, b_1+n_{1i}-s_{1i})}\\ &\quad= f_S(x_{1,i+1}|a_1+s_{1i}, b_1+n_{1i}-s_{1i}),\end{aligned}$$

which is a beta-binomial density with index $m_{1,i+1}$ and parameters $(a_1 + s_{1i})$ and $(b_1 + n_{1i} - s_{1i})$, as shown above.

It is essential to know the future optimal actions in order to decide whether to recruit more patients or not to the next stage. As the sequential scheme is consisted of finite sequence of actions, the expected utility can be computed by backward induction which is moving backward from the ter-

minal branch to the present stage (DeGroot, 1970, Ch. 12). Section 3.5.1 gives a brief review of the backward induction methodology. A more detailed description of the methodology for the proposed design is given in Section 6.2.

Expected utility of action T (start a new phase II trial)

The treatment effects are assumed to be independent and as such the utility of action T depends on the number of patients remaining, $N_2 = N_1 - n_{1i}$, and the remaining number of potential new drugs, $K_2 = K_1 - 1$. The gain function is obtained by considering a new portfolio of trials starting with this smaller population size and fewer treatments less the total cost of patients from the current trial. Let $\mathcal{G}_{\text{Total}}(N_2, K_2)$ denote the expected utility of a whole programme with population N_2 and K_2 treatments, then the expected utility of action T is

$$\mathcal{G}_T(s_{1i}, n_{1i}, N_1, K_1) = \mathcal{G}_{\text{Total}}(N_2, K_2) - l n_{1i}. \tag{6.8}$$

The gain $\mathcal{G}_{\text{Total}}(N_2, K_2)$ is given by

$$\sum_{s_{21}=0}^{n_{21}} \max_{a \in \{R,P,T,A\}} \{\mathcal{G}_a(s_{21}, n_{21}, N_2, K_2)\} f_S(s_{21}) - l_{\text{II}}$$

and the individual expected utility of action a are obtained as given in equations (6.3), (6.6), (6.7), and (6.8).

In a more general notation, the expected utility of the whole development

at the k-th trial is denoted by

$$\mathcal{G}_{\text{Total}}(N_k, K_k) = \sum_{s_{k1}=0}^{n_{k1}} \max_{a\in\{R,P,T,A\}} \{\mathcal{G}_a(s_{k1}, n_{k1}, N_k, K_k)\} f_S(s_k) - l_{\text{II}}$$

and the expected utility of action a at i-th stage of k-th trial is

$$\begin{aligned}
&\mathcal{G}_a(s_{ki}, n_{ki}, N_k, K_k) \\
&= \begin{cases}
\displaystyle\sum_{x_{k,i+1}=0}^{m_{k,i+1}} \max_{a\in\{R,P,T,A\}} \{\mathcal{G}_a(s_{ki} + x_{k,i+1}, n_{ki} + m_{k,i+1}, N_k, K_k)\} & \\
\qquad \times f_S(x_{k,i+1}|s_{ki}, n_{ki}), & a = R \\
A(s_{ki}, n_{ki}, N_k, K_k) - lN_k - l_{\text{III}}, & a = P \\
\mathcal{G}_{\text{Total}}(N_{k+1}, K_{k+1}) - ln_{ki}, & a = T \\
-ln_{ki}, & a = A
\end{cases}
\end{aligned} \tag{6.9}$$

where $A(s_{ki}, n_{ki}, N_k, K_k) = \int_0^1 \left(1 - \Phi\left(z_{1-\alpha/2} - \theta_k\sqrt{V_k}\right)\right) f_{p|s}(p_k|s_{ki}, n_{ki})\, dp_k$, and the remaining patients and treatments for action T are $N_{k+1} = N_k - n_{ki}$ and $K_{k+1} = K_k - 1$, respectively.

The expected utility of action T depends on the subsequent actions and thus its computation is also solved by backward induction. The computation of the expected utility of action T is just like the computation of the expected utility of the whole programme and this is discussed in the following section.

6.2 Backward induction

The expected utility of the whole development plan as given in (6.2) depends on the expected utilities of actions R, P, T and A. Their expressions in turn are given by (6.7), (6.6), (6.8), and (6.3), respectively, showing that some expected utilities may depend on the expected utilities of subsequent actions. The development plan may be represented by a decision tree where the plan begins with an initial state on the left and the future lies on the right (Lindley, 1985). Just like a real tree, the decision tree has branches where each branch is thought of as an interim stage (Figure 6.1). Each branch "contains several parts that act together" and so, different parts that are easier to solve should be done first. Then, using the rules of probability, the computations are brought together to solve for the expected utility of the whole series of trials.

The computational strategy is to compute the expected utilities for branches further to the right. At each point, the expected utilities from different actions are compared and consequently, the optimal decision is obtained. In practice, the computation of the expected utility of a programme with an initial population of size N and K potential treatments is to, first, solved for action T at the further right. As shown from equation (6.9), at the k-th trial, the expected utility of action T is made up of the expected utility of a whole programme with n_{ki} patients and 1 treatment less than the current k-th trial. Therefore, it is easier to calculate the expected utility of the last treatment (K-th trial) first and this is done using backward induction. At the ultimate stage within this "branch", there can only be two

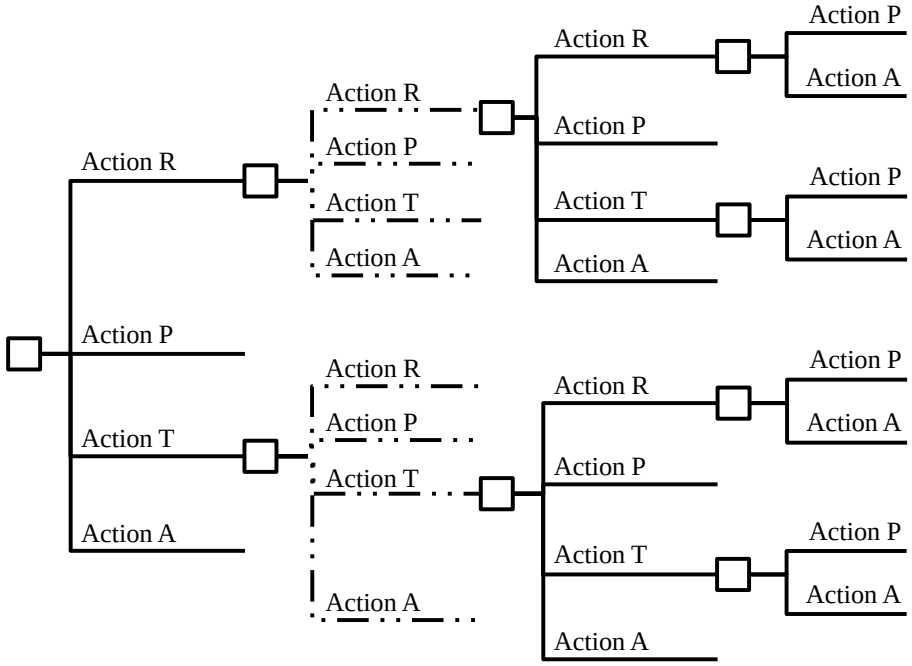


Figure 6.1: The decision tree for a series of decision-theoretic phase II trials

actions to choose from, namely, actions P (proceed to phase III trial) and A (abandon the development plan). Action R (resume current trial) is not available because there would not be enough patients to go on to phase III trial, that is, action P will not be available at some point in the future. For ease of computation, at the ultimate interim analysis of K-th trial, the expected utility of action R is $\mathcal{G}_R(s_{Ki}, n_{Ki}, N_K, K_K) \to -\infty$ so that this action will never be chosen as the optimum action. The expected utilities of actions P and A are easily computed, as given in equations (6.6) and (6.3). If $\mathcal{G}_P(s_{Ki}, n_{Ki}, N_K, K_K) > \mathcal{G}_A(s_{Ki}, n_{Ki}, N_K, K_K)$ then the optimum action is action P, and if otherwise, action A. Consequently, the expected utility of action R at the penultimate interim stage can be evaluated. At the penultimate stage, the expected utilities of actions R, P and A are compared and the

highest is the optimum action. Following on this manner of evaluation, the expected utility of the whole programme with N_K population and $K_K = 0$ treatment can be obtained. Thus, the expected utility of all actions T of the $(K-1)$-th trial, $\mathcal{G}_T(s_{K-1,i}, n_{K-1,i}, N_{K-1}, K_{K-1}) = \mathcal{G}_{\text{Total}}(N_K, K_K) - ln_{K-1,i}$, is obtained for all $n_{K-1,i}$ values.

Once this "branch" of the decision tree is solved, the next "branch" to be considered is the action T with N_{K-1} population and K_{K-1} treatments. Similarly, at the ultimate stage of the $(K-1)$-th trial, there are only actions P and A to choose from. By knowing the optimum action at the ultimate stage, the expected utility of action R at the penultimate stage is solved and compared with those of actions P, T and A. The expected utility of action T is already computed as described in the preceding paragraph.

Finally, the expected utility of the whole programme is solved by evaluating in this iterative manner; firstly, by moving up to a "branch", that is, increasing the number of treatments by one. Secondly, within the "branch", evaluate the expected utility of action T from the terminal branch back to the initial stage. Therefore, all expected utilities and optimal actions can be determined and the expected utility of the whole programme obtained.

6.3 Application

In the earlier introductory chapter on asthma (Chapter 2), the prevalence rate of asthma in the United Kingdom is 6% (Corrigan, 2009). Although only about 5–10% of them have the most severe form of asthma, they account for at least 50% of the total asthma-related health care cost (Humbert *et al.*,

2005, p. 309). These patients are usually at higher risk of asthma exacerbations and the usual prescribed treatments for them are combined regimens; bronchodilators and corticosteroid. In the recent years, with the advent of targeted therapies, mediator antagonists and immunomodulatory have gone on trials and some have been approved as first line therapies. Drugs from the class of mediator antagonists are used as illustrations in this section as there are more published randomized clinical trials for this class of drugs.

These new treatments aim to control asthma exacerbations, where severe exacerbations are defined as events of at least one of the following: use of systemic corticosteroids or increase dosage from the maintenance dose, and a hospitalization or emergency department visit because of asthma and requires systemic corticosteroids. The definition of moderate asthma exacerbations includes at least one of the following: deterioration in symptoms, deterioration in lung function, increased rescue bronchodilator use, and emergency department visits because of asthma but does not require systemic corticosteroids (Reddel *et al.*, 2009).

The first illustration will assume that there is an unlimited number of potential treatments ($K \to \infty$), that is, new treatments are continuously being rolled out for phase II clinical trials and the second illustration will assume that there are only a few limited number of treatments available. In the latter scenario, when the K-th treatment (the last treatment) is on trial, action T is not available at the interim stages, leaving only three actions to choose from, namely, actions R, P and A.

The primary endpoint is the efficacy of the treatment in controlling asthma exacerbations. At least an episode of asthma exacerbation during

the period of the treatment (assumed to be four weeks) is considered as a failure and a non-event as a success. If action P is taken, then the experimental treatment is tested against a placebo control arm at a two-sided significance level of $\alpha = 0.05$. Under the null hypothesis of the phase III trial, the probability of success of the placebo is assumed to be $p_C = 0.80$ whereas the probability of success of the experimental treatments, $p_1, p_2, \ldots, p_K$, are independent unknown variables. The unknown parameter of the efficacy, p_k, is assumed to be a random variable following a beta distribution with known parameters a_k and b_k. All K treatments are assumed to have the same prior densities and so, for convenience, the subscript k is suppressed.

It is not easy to elicit from experts for what these parameters may be. As there are some published randomized clinical trials comparing mediator antagonist therapies with placebo for severe and persistent asthmatic patients, it is possible to estimate the parameters empirically. The criteria used to search for published trials on randomized clinical trials were persistent severe asthmatic population and mediator antagonists interventions. Publications from these authors were obtained: Busse *et al.* (1999), Humbert *et al.* (2005, 2009) and Malmstrom *et al.* (1999).

The primary endpoint from some of these trials was asthma exacerbation whilst some trials reported it as the secondary endpoint. All these trials were conducted prior to the published guideline by the American Thoracic Society (ATS) and European Respiratory Society (ERS) in 2009 (Reddel *et al.*, 2009). As such, the definitions of asthma exacerbations may be slightly different than the one recommended by the ATS/ERS Task Force. However, they reported the number of patients who needed rescue medication, un-

scheduled physicians visits, emergency department visits or hospitalization. Henceforth, it is still possible to obtain the number of patients with at least an episode of asthma exacerbations at the predefined period of observation.

Another difficulty in summarizing the proportion of patients with asthma exacerbations is the treatment period. Some of the trials were measuring short-term effect and some long-term. It is thus, not feasible to compare the number of patients with at least one exacerbation in a four-week treatment period with a trial that investigated a 12-week treatment regimen. In order to standardize treatment comparison, the treatment period is scaled to be the same. A four-week treatment period is chosen as the treatment period as it is possibly the shortest treatment period in investigating targeted therapy such as the mediator antagonist. Let T be the time when the first episode of asthma exacerbation (either a moderate or a severe exacerbation) is reported. Therefore, T is a continuous variable and assuming that the event of exacerbation occurs constantly, a reasonable assumption for its distribution is the exponential distribution. Let $S(t)$ denote the cumulative density function of T,

$$S(t) = \Pr(T > t) = e^{-\lambda t}. \tag{6.10}$$

Humbert *et al.* (2009) conducted a trial comparing masitinib to a placebo and the treatment duration was $t = 16$ weeks. It was reported that a total of $f = 14$ patients out of $n = 33$ in the masitinib arm experienced at least

Table 6.1: Summary of estimated proportions of no asthma exacerbation from published randomized clinical trials comparing mediator antagonist therapies with placebo for severe and persistent asthmatic patients

Published trials	n	f	Treatment period, t	$S(t=4)$
Humbert *et al.* (2009)	33	14	16 weeks	0.87108
Malmstrom *et al.* (1999)	387	60	12 weeks	0.94539
Busse *et al.* (1999)	145	9	4 weeks	0.93793

one exacerbation during the treatment period. Therefore, from (6.10),

$$\begin{aligned} & S(t=16) = 1 - f/n \\ \Leftrightarrow \quad & \exp\{-16\hat{\lambda}\} = 1 - 14/33 \\ \Leftrightarrow \quad & \hat{\lambda} = 0.03450. \end{aligned}$$

Therefore, the estimated proportion of non-event is

$$S(t=4) = \exp\{-4\hat{\lambda}\} = 0.87108.$$

A summary of the estimated proportions of non-event from the other published trials is given in Table 6.1. Using these estimated proportions the parameters for the beta distribution can be estimated using the maximum likelihood estimation method. The estimated parameters are $\hat{a} = 68.9870$ and $\hat{b} = 6.1569$. Rounding them off to the nearest integer the prior density of p is $Beta(69, 6)$. The degree of the prior belief is equivalent to obtaining information from a sample size of 75.

Assuming that the projected total size of the population is $N = 300$. The variable and fixed costs are taken to be relative to one unit of gain and

assuming that the cost per patient is $l = 0.0001$ and the costs of conducting a phase II and III trials are $l_{\text{II}} = 0.002$ and $l_{\text{III}} = 0.02$, respectively. At each stage of the k-th decision-theoretic phase II trials, $m_{ki} = 1$ patient is recruited. For a simplistic illustration, the minimum sample size for the phase III trial is set to 1. Although it is not practical to have only one patient in a phase III trial, it is deemed necessary to have a restriction of at least one patient. The design can easily be extended to restrict the minimum sample size to be more than one patient.

6.3.1 Unlimited number of treatments

Figure 6.2 shows that the expected utility of the whole development plan is $0.83 > 0$ and so the development plan is worthy to start. The decision-making process for the first trial is shown in Figure 6.3. If the first patient was a failure then the optimal action is to stop the current trial and initiate a new decision-theoretic phase II trial, that is, an action T. If however, the first patient did not have asthma exacerbation in the first four weeks of treatment, then the optimal action is to continue recruiting another patient to the current trial. If the first 24 patients were all successes, then the optimum action is action P, that is, to stop the current phase II trial and proceed to a phase III trial with the remaining 276 patients where the experimental mediator antagonist drug is compared to a placebo. Therefore, the minimum number of patients needed for the phase II trial before a phase III trial could be initiated is 24. The maximum number of patients needed is 79 which is very close to the prior size of 75. If the cumulative number of successes is

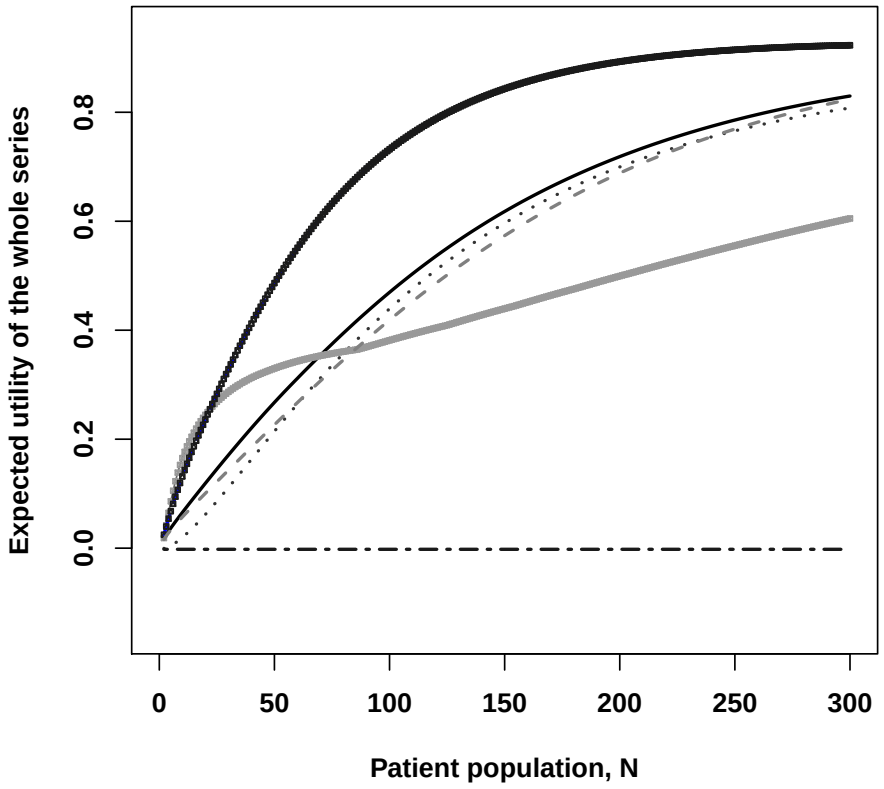


Figure 6.2: Expected gain of the whole development plan for $Beta(a = 69, b = 6)$ (solid line), $Beta(a = 1, b = 1)$ (dotted line), $Beta(a = 1.15 \times 10^{-5}, b = 10^{-6})$ (dashed-dotted line), $Beta(a = 1.01, b = 0.088)$ (thick grey line), $Beta(a = 11.5, b = 1)$ (thick black line) and $Beta(a = 676.2, b = 58.8)$ (dashed line).

74, then the optimal action is action P but if the cumulative successes is less than 74, the optimal action is to stop the current trial and a new phase II is initiated. In the latter scenario, the new phase II trial will begin with a population of $N_2 = 221$.

Note that the boundaries for actions P and T are very close together. The expected value of the informative prior is $E(p) = a/(a + b) = 0.92$ and

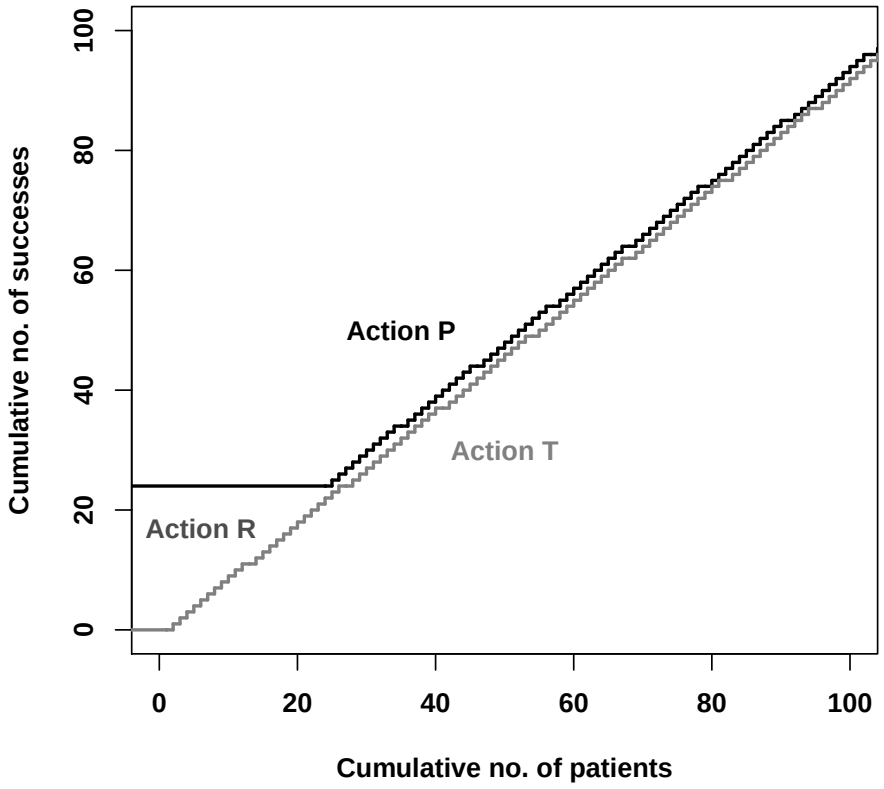


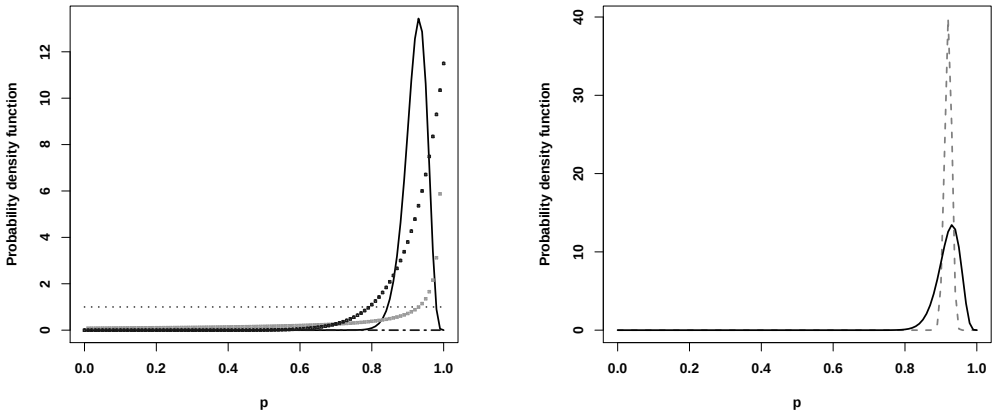
Figure 6.3: Decision rules for example in Section 6.3.1 based on prior $Beta(69, 6)$ and assuming $p_C = 0.80$.

the variance is 0.000968. This suggests that the treatment is highly effective and thus, the threshold to proceed to the phase III trial must be high, too. The threshold for action T is also severely high. As given by the very small variance, it is expected that the probability of success is very close to 0.92 therefore, if a failure is observed early the trial should stop and a new trial with a new mediator antagonist should be initiated instead.

Decision-theoretic development plans with different underlying prior den-

sities are compared. A reference prior is sometimes preferred instead of eliciting information from experts as their judgments may be subjective (Stangl and Berry, 1998). The standard uniform prior is set as the reference prior, which is a special case of a beta distribution with parameters $a = 1$, $b = 1$. The expected value of a uniform prior is 0.5 and the variance is 0.0833. In addition, other parameters such as $(a = 1.15 \times 10^{-5}, b = 10^{-6})$, $(a = 1.01, b = 0.088)$, $(a = 11.5, b = 1)$, and $(a = 676.2, b = 58.8)$ are used to compare the characteristics of the optimal development plan. All these prior densities; $Beta(1.15 \times 10^{-5}, 10^{-6})$, $Beta(1.01, 0.088)$, $Beta(11.5, 1)$, and $Beta(676.2, 58.8)$ give the same expected value, 0.92, but different variances. The prior $Beta(1.15 \times 10^{-5}, 10^{-6})$ gives the largest possible variance, 0.0736, subject to the constraints of $E(p) = 0.92$ and $a > 0, b > 0$. On the one hand, prior $Beta(1.01, 0.088)$ gives the largest possible variance, 0.03514, subject to the constraints of $E(p) = 0.92$, $a \geq 1$ and $b > 0$. On the other hand, prior $Beta(a = 11.5, b = 1)$ gives the largest variance, 0.00545, subject to the constraints of $E(p) = 0.92$, $a > 0$ and $b \geq 1$. Finally, the prior $Beta(676.2, 58.8)$ gives the smallest computationally possible variance, 0.0001, subject to the constraints of $E(p) = 0.92$ and $a > 0, b > 0$. The shapes of these densities are shown in Figure 6.4.

As shown in Figure 6.4(a) the density $Beta(11.5, 1)$ put most of the weight on the extreme right, that is, it is more likely to have high probability of success. Unsurprisingly, the expected gain of the whole development plan is also higher than other prior densities (Figure 6.2). Although the expected utility of the whole plan is higher than the informative prior $Beta(69, 6)$, its larger variance means that more patients are needed before an informed



(a) (b)

Figure 6.4: Beta densities for various parameters. Panel (a) $Beta(a = 69, b = 6)$ (solid line), $Beta(a = 1, b = 1)$ (dotted line), $Beta(a = 1.15 \times 10^{-5}, b = 10^{-6})$ (dashed-dotted line), $Beta(a = 1.01, b = 0.088)$ (thick grey dots) and $Beta(a = 11.5, b = 1)$ (thick black dots). Panel (b) $Beta(a = 69, b = 6)$ (solid line) and $Beta(a = 676.2, b = 58.8)$ (grey dashed line)

choice can be made to progress to the phase III trial (Table 6.2).

If the prior density is $Beta(1.01, 0.088)$, it has a slightly higher overall expected gain than others if the population is small but as its shape is slightly "wider" it is not as informative as the other priors as the population size gets larger. The variance given by the parameters $(a = 1.15 \times 10^{-5}, b = 10^{-6})$ is the largest variance that can be attained under the constraints stated earlier. The expected utility of the whole development plan is less than zero (the dashed line just below zero shown in Figure 6.2), that is, given a population of 300 with a true probability of success p following a prior distribution $Beta(1.15 \times 10^{-5}, 10^{-6})$ and the probability of success of a placebo is assumed to be 0.80, it is not worthy to start any trial at all.

For the reference prior which is the standard uniform distribution, the expected gain of the whole series tends to the expected gain of an informative prior ($Beta(69, 6)$) for large population, $N > 300$. This conforms to the belief that regardless if the prior is flat or informative, given sufficiently large population, similar results will be attained in the long run. The minimum and maximum sample size that are needed in order to go on to the phase III trial is, as suspected, larger than what is necessary for the $Beta(69, 6)$ prior.

The parameters ($a = 676.2, b = 58.8$) is about 10 times higher than the informative $(69, 6)$ and its variance is the smallest possible variance attainable for illustration purposes. Due to its extremity, a comparison between its density function and the one given by $a = 69, b = 6$ is shown in Figure 6.4(b). Interestingly, the expected utility of the whole series with prior density $Beta(676.2, 58.8)$ is slightly lower than the $Beta(69, 6)$ prior. However, due to its small variance, the minimum sample size needed for the decision to go on to a phase III trial is only one (Table 6.2).

The expected utility of the whole series is a non-decreasing function of N (patient population). It increases rapidly for small values of N but the increase levels off and becomes rather more linear for large values of N. The most obvious examples are priors $Beta(1.01, 0.088)$ and $Beta(a = 11.5, b = 1)$. For the former prior, the expected utility function (thick grey line, Figure 6.2) seemed to dip slightly between a population of size 75 and 100. However, upon closer inspection, the expected utility of the whole series was increasing at a slower rate. The inflection for the prior $Beta(a = 11.5, b = 1)$ (thick black line, Figure 6.2) occurred at about the same interval as $Beta(1.01, 0.088)$, on the other hand, was smoother.

Table 6.2: Optimal designs for a series of decision-theoretic phase II trials based on beta prior distributions with parameters a and b, and assuming $p_C = 0.80$.

			Minimum	Maximum	
			sample size to proceed		
a	b	N	to phase III trial		$\mathcal{G}_{\text{Total}}(N, K)$
69	6	100	7	7	0.4695
1	1		17	35	0.4393
1.15×10^{-5}	10^{-6}		A	A	-0.0021
1.01	0.088		19	73	0.3814
11.5	1		20	20	0.7322
676.2	58.8		1	1	0.4187
69	6	200	16	27	0.7186
1	1		29	69	0.6996
1.15×10^{-5}	10^{-6}		A	A	-0.0021
1.01	0.088		25	185	0.4994
11.5	1		37	66	0.8936
676.2	58.8		1	1	0.6884
69	6	300	24	79	0.8298
1	1		40	110	0.8079
1.15×10^{-5}	10^{-6}		A	A	-0.0021
1.01	0.088		31	290	0.6051
11.5	1		49	107	0.9231
676.2	58.8		1	1	0.8238

A, Abandon the development programme without starting any phase II trial

6.3.2 Limited number of treatments

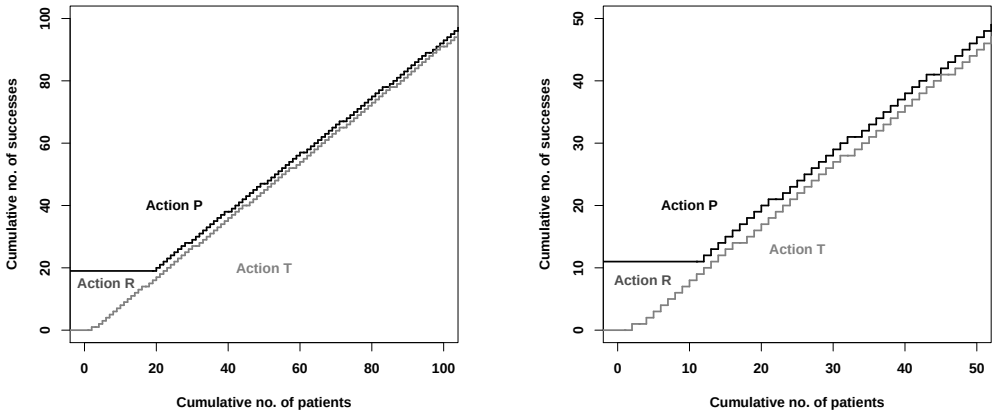
inal likelihood of $S_{1i_1}, \ldots, S_{Ki_K}$. In practice, the number of potential new treatments available for trial is usually between three and five. Suppose that the number of promising treatments is $K = 3$, and the other variables are as given above, $p_C = 0.80$, $N = 300$, $l = 0.0001$, $l_{\text{II}} = 0.002$, $l_{\text{III}} = 0.02$, $m_{ki} = 1$ and there must be at least one patient for a phase III trial. Only the

result from the informative prior $Beta(69, 6)$ is given here. The boundaries for optimal actions are shown in Figure 6.5.

At the start of the development plan, there are 300 patients in the population and 3 potential treatments. The minimum number of patients that will be recruited to the decision-theoretic phase II trial in order to proceed to the phase III trial is 19, fewer than when $K \to \infty$. On the other hand, the maximum number of patients is 84, slightly higher than the prior sample size 75 and the case of $K \to \infty$. If there are at least 78 successes out of 84 patients, the optimal action is to go on to the phase III trial (Figure 6.5(a)). If however, less than 78 successes were observed then the optimal action is action T, to stop the current trial and initiate a new phase II trial.

In the latter case, starting a new phase II trial, the population for the "new" development plan is $N_2 = 216$ and the number of potential treatments is $K_2 = 2$. The minimum number of patients needed to move on to a phase III trial is 11. The maximum number of patients that will be recruited to the current phase II trial is 44 and if the cumulative number of successes is 41 or higher, then action P should be taken. On the other hand, if the cumulative number of successes is 40 or less then the optimal action is to stop the current trial and try a new treatment in a new phase II trial (Figure 6.5(b)).

Suppose that the third phase II trial has to be initiated, then the development plan now has only a population of size $N_3 = 172$ and $K_3 = 1$ treatment. As there is only one treatment left, action T is not available to choose from in each interim stage. Interestingly, regardless if the first patient recruited to the third phase II trial was either a failure or a success, the optimal action



(a) (b)

Figure 6.5: Decision rules for optimal actions for example in Section 6.3.2.

is to proceed to a phase III trial.

The boundaries between Action T (stopping the trial and try a new treatment) and action P (stop the trial and proceed to the phase III with the current treatment) are very close together. The efficacy of the treatment is very high and so the boundary to proceed to phase III trial is also set higher to ensure that the treatment is genuinely effective. The narrow gap, in order to resume the current trial, suggests that if a success has been observed, it is always worthwhile to recruit another patient. The aim of the design is to reach the "tipping point" where it is sufficient to make an informed decision.

6.4 Discussion and concluding remarks

The proposed design is for a series of Bayesian decision-theoretic phase II trials and one phase III trial. At each interim stage of each phase II trial,

the viability of the experimental treatment in a phase III trial is evaluated and incorporated into the utility function to aid the decision-making. This expected gain is compared with other expected gains, such as, the gain of initiating a new phase II trial with the remaining potential treatments, the gain of recruiting more patients to the current trial to obtain more information, and the gain of ceasing the development portfolio.

At the design stage, the unknown parameter of the efficacy of the experimental treatment is assumed to be random. In turn, due to the uncertainty of belief, the sample size cannot be determined at the design stage. Only the minimum and maximum sample sizes can be computed. On average though, the sample size required in a hybrid design is smaller than one obtained using the frequentist methodology.

For simplification, the economic factors that makes up the utility function only considers the cash outflows. The profits from the future sales of the successful treatment could be included for a realistic modelling. In reality, profits may increase steadily in the first few years upon being marketed and then hit a plateau before finally, decrease after the patent of the new drug has expired or a successor treatment is available. Therefore, the projected duration of the treatment period and sales of the drug may be included in the utility function.

Another simplification assumed for the utility function is that the cost per patient for a phase II trial is equal to that of the cost per patient recruited to a phase III trial. In practice, there may be more tests and longer follow-up periods for patients in a phase III trial and so the total cost per patient in a phase III trial may be higher than the total cost per patient in a phase II

trial. Consequently, this will decrease the expected utility function of action P slightly and affects the sequential scheme.

The proposed design may be extended by incorporating multi-arm designs in the phase II setting. There are a few possible frameworks for the design of such studies. One approach is to assign one patient at a time to one of the available treatments then make a decision on the basis of all the observed responses as to how the trial should proceed. The available choices are similar to the actions introduced in Section 6.1, namely, to recruit another patient to the same treatment (action R), to proceed to a phase III trial with the same treatment (action P), to abandon the development plan (action A) or to randomize another patient to one of the other treatments (similar to action T). In the context of the examples given above (Section 6.3), the expected utility of the whole series may not be too different to that from the single-arm design as the expected gain from action P has most influence in the expected utility of the whole series.

Another framework for a multi-arm design is to randomize a cohort of patients at the initial stage of the development plan to all the available treatments. This is similar to the design by Stallard and Thall (2001) except that a decision is made at each interim stage from all the observed data. The actions at each interim stage are to drop a treatment, to proceed to a phase III setting with one treatment or to abandon the development plan. If a treatment is dropped then the next cohort of patients is randomized to the remaining treatments. If the optimal action is to proceed to a phase III trial then the remaining patients would be randomized to either the chosen treatment or the control treatment. The framework can also be modified to

allow a new treatment to be added at the interim stage.

Although we have assumed that all the potential treatments are from the same class of drugs, it is possible that both fixed and variable costs may be different for different treatments. These information can be easily incorporated into the utility function.

The proposed design in this chapter assumes that the potential treatments are related. As these treatments are from the same class of drug and they are intended for the same population it is not unrealistic to consider these treatments to be correlated with each other. Therefore, the next chapter considers the correlation between treatments. The prior distributions of each treatments are assumed to be correlated. Therefore, given the observed data from the preceding trials and the prior beliefs of the preceding treatments the prior densities of subsequent treatments are updated.

Chapter 7

A Series of Decision-Theoretic Phase II Trials of Related Treatments

The introductory chapter on asthma (Chapter 2) briefly describes that most of the current drugs for asthma can generally be classified into a few major families, namely, bronchodilators, corticosteroid, mediator antagonist, and immunomodulatory. Each of them has different mechanism in controlling asthma and it is expected that drugs from the same class are more similar in controlling the disease than drugs from other classes. Therefore, drugs from the same class are more related than those that are not.

In the preceding chapter, the design for a series of phase II trials assumes that each of the drugs available for trials has the same prior distribution and the probabilities of success are independent. However, as drugs from the same class are expected to be related, it is worth considering the correlation

between them in the design. As such, the prior distributions of the subsequent drugs may be affected by the data from the preceding drugs and consequently informing us the performance of the class of drugs under evaluation. In this chapter, the dependency between drugs of the same family is quantified such that the prior density functions of the following drugs are updated as responses from the preceding drugs are obtained.

The design of the series of trials is introduced in Section 7.1. The notation used in this chapter is those introduced earlier in this thesis but is presented in Section 7.2 for completeness. Section 7.3 presents an illustration of the design with examples from drugs from the class of mediator antagonist for patients with severe and persistent asthma. Finally, the chapter concludes with a discussion in Section 7.4.

7.1 Design

The statistical framework for the design for a series of decision-theoretic phase II trials of related treatments follows the design introduced in the preceding chapter very closely. The design encompasses a series of decision-theoretic single-arm phase II trials and a randomized controlled two-arm phase III trial. The phase II trials are conducted sequentially with interim decision making. At each interim stage, a decision is made from the following actions:

Action R Resume the current phase II trial by recruiting more patients,

Action P Stop the current phase II trial and proceed to a randomized two-

arm phase III trial where the experimental treatment is compared to the standard treatment using the remaining patients,

Action T Stop the current phase II trial and initiate a new single-arm phase II trial of a different treatment using the remaining patients; or

Action A Stop the current phase II trial and abandon the development plan.

The design of the phase III trial following action P is based on the standard frequentist method (discussed in Section 4.1). Therefore, at the end of the phase III trial, the data from that trial are analysed using the standard frequentist analysis where the point estimate of the treatment efficacy is reported with its confidence interval and the corresponding p-value to infer the true treatment efficacy. At the design stage of the whole development plan though, the unknown parameter of the treatment efficacy is assumed to be random and follows a prior distribution. Each treatment has a prior distribution and they are all correlated. As the trials are run sequentially, the accumulated observed data from patients are used to update the prior densities of the current and subsequent treatments. The posterior densities are made up of combined beliefs of the preceding priors given the observed data.

The process of the decision making at each interim analysis is based on the utility of each action which depends on the true efficacy of the experimental treatment, and the number of patients and treatments available at the start of the series. As the true parameter of the efficacy is a random variable following a known distribution, the utility function of each action is an expectation

function over all the possible values of the true parameter. The expected utility in turn depends on the accumulated information.

Similar to the preceding chapter, the utility function is defined to be relative to a constant reference value. This baseline is fixed to the initial stage of the development plan where no cost is spent, as such, the fixed reference value is zero. Therefore, the development plan will only commence if the value of information from patients' responses is worthier than the costs of recruiting patients, setting up and conducting the trials.

The proposed design is made up of a finite sequence of decisions. Some of the expected utilities functions depend on the expected utilities of subsequent actions. Therefore, the optimum sequential scheme is obtained by backward induction, similar to the design of a series of decision-theoretic phase II trials (Section 6.2). The computational strategy is to first consider the expected utility of the whole decision tree of the ultimate treatment. As it is the last treatment, action T will not be available at any interim stage. Then compute the expected utility of the whole decision tree of the penultimate trial, noting that all the actions are available to choose from. Following in this manner of iterative evaluation, the optimal expected utility of the whole development plan is computed. The expected gain functions are described in detail in the following sections.

7.2 Notation

Following on the notation used in the previous chapter, let N denote the total size of the population which is assumed to be fixed and known. Let K

denote the number of potential new treatments available for trials for this population and it is also assumed to be fixed and known. At the i_k-th stage of k-th phase II trial which is equivalent to k-th drug ($k = 1, 2, \ldots, K$), m_{ki_k} patients are recruited to the trial. As in the preceding chapter, the design proposed in this chapter is illustrated with $m_{ki_k} = 1$, that is, a fully sequential phase II trial where a decision is made after every patient. As the design can be extended to recruit patients in groups of size greater than one, the more general notation m_{ki_k} is used. The total number of patients recruited up to and including the i_k-th stage of k-th trial is $n_{ki_k} = m_{k1} + m_{k2} + \ldots + m_{ki_k}$.

The primary endpoints for both phase II and III trials are assumed to be binary. Let a success response be represented numerically by 1 and a failure by 0. Let X_{ki_k} denote the sum of successes out of m_{ki_k} patients in the i_k-th stage of k-th trial. Let the accumulated total number of successes from the first stage up to and including the i_k-th stage of k-th trial be denoted by $S_{ki_k} = X_{k1} + X_{k2} + \ldots + X_{ki_k}$. Patients' responses are assumed to be independent from each other and so the sum of successes of the current trial, S_{ki_K}, is assumed to follow the binomial distribution with index n_{ki_k} and an unknown probability of success, p_k. The likelihood function of S_{ki_k} is

$$f_{S|p}(s_{ki_k}|p_k) = \binom{n_{ki_k}}{s_{ki_k}} p_k^{s_{ki_k}} (1 - p_k)^{n_{ki_k} - s_{ki_k}}.$$

In addition, the responses from each trial are also assumed to be independent from each other. Therefore, the joint conditional K-variate distribution

of $(S_{1i_1}, \ldots, S_{Ki_K})$ is the product of each of the likelihood function,

$$h_{\boldsymbol{S}|\boldsymbol{p}}(S_{1i_1} = s_{1i_1}, \ldots, S_{Ki_K} = s_{Ki_K}|p_1, \ldots, p_K) = \prod_{k=1}^{K} f_{S|p}(s_{ki_k}|p_k). \qquad (7.1)$$

As the trials run sequentially, at k-th trial there is no observation from the later trials, that is, $S_{k+1,i_{k+1}} = \ldots = S_{Ki_K} = 0$.

Similar to the previous chapter, the unknown parameter p_k is assumed to be random and let $f_p(p_k)$ be the marginal prior density of p_k. The prior density summarizes the belief of the efficacy of the k-th drug prior to observing any response. The probabilities of success are, however, not independent. Let $h_{\boldsymbol{p}}(p_1, \ldots, p_K)$ denote the joint prior density of $p_1, p_2, \ldots, p_K$.

Upon observing $S_{ki_k} = s_{ki_k}$ successes out of n_{ki_k} patients in the first i_k stages of the k-th trial and given the observed successes from the preceding trials, $s_{1i_1}, \ldots, s_{k-1,i_{k-1}}$, the information of $(p_1, \ldots, p_K)$ is updated. According to Bayes' theorem, the joint posterior density is

$$\begin{aligned} &h_{\boldsymbol{p}|\boldsymbol{S}}(p_1, \ldots, p_K|s_{1i_1}, \ldots, s_{k,i_k}, n_{1i_1}, \ldots, n_{ki_k}) \\ &= \frac{h_{\boldsymbol{S}|\boldsymbol{p}}(s_{1i_1}, \ldots, s_{ki_k}, 0, \ldots, 0, n_{1i_1}, \ldots, n_{ki_k}, 0, \ldots, 0|p_1, \ldots, p_K)}{h_{\boldsymbol{S}}(S_{1i_1} = s_{1i_1}, \ldots, S_{ki_k} = s_{ki_k}, S_{k+1,i_{k+1}} = 0, \ldots, S_{Ki_K} = 0)} \\ &\quad \times h_{\boldsymbol{p}}(p_1, \ldots, p_K) \end{aligned} \qquad (7.2)$$

where

$$\begin{aligned} & h_{\boldsymbol{S}}(S_{1i_1} = s_{1i_1}, \ldots, S_{ki_k} = s_{ki_k}, S_{k+1,i_{k+1}} = 0, \ldots, S_{Ki_K} = 0) \\ & = \int \cdots \int h_{\boldsymbol{S}|\boldsymbol{p}}(s_{1i_1}, \ldots, s_{ki_k}, 0, \ldots, 0 | p_1, \ldots, p_K) \\ & \quad \times h_{\boldsymbol{p}}(p_1, \ldots, p_K)\, dp_1 \cdots dp_K, \end{aligned} \tag{7.3}$$

is the joint marginal likelihood of $S_{1i_1}, \ldots, S_{Ki_K}$.

Similar to the preceding chapter, each of the unknown parameters p_k is assumed to follow the beta distribution with fixed parameters a_k and b_k, and the marginal probability density function is thus,

$$f_p(p_k) = \frac{1}{B(a_k, b_k)} p_k^{a_k - 1} (1 - p_k)^{b_k - 1},$$

where $B(a_k, b_k) = \Gamma(a_k)\Gamma(b_k)/\Gamma(a_k + b_k)$ is the beta function. As introduced earlier in Section 3.1.3, the joint distribution of $p_1, \ldots, p_K$ is assumed to follow the K-variate Sarmanov's family with parameters $\Omega_K = \{\omega_{i_1,i_2}, \omega_{i_1,i_2,i_3}, \ldots, \omega_{1,2,\ldots,K}\}$ and mixing function $\phi(p_k)$ (Lee, 1996). Its expression is

$$h_{\boldsymbol{p}}(p_1, \ldots, p_K) = \Big(\prod_{k=1}^{K} f_p(p_k) \Big) \Big(1 + R_{\Omega_K}(p_1, \ldots, p_K) \Big), \tag{7.4}$$

where

$$R_{\Omega_K}(p_1,\ldots,p_K) = \sum_{j_1=1}^{K-1}\sum_{j_2=j_1+1}^{K}\omega_{j_1,j_2}\phi(p_{j_1})\phi(p_{j_2}) + \sum_{j_1=1}^{K-2}\sum_{j_2=j_1+1}^{K-1}\sum_{j_3=j_2+1}^{K}\omega_{j_1,j_2,j_3}\phi(p_{j_1})\phi(p_{j_2})\phi(p_{j_3}) + \cdots + \omega_{1,2,\ldots,K}\prod_{j=1}^{K}\phi(p_j).$$

The mixing function is assumed to be $\phi(p_k) = p_k - \mu_k$ where $\mu_k = a_k/(a_k+b_k)$ is the expected value of a beta distribution with parameters a_k and b_k.

The unconditional joint density of $S_{1i_1},\ldots,S_{Ki_K}$ is obtained as in (7.3). The expression, derived earlier in Section 3.2.3 (equation (3.14)), is

$$h_{\boldsymbol{S}}(S_{1i_1},\ldots,S_{Ki_K}) = \Big(\prod_{k=1}^{K} f_S(s_{ki_k},n_{ki_k}|a_k,b_k)\Big)\Big(1+D_{\Omega_K}(s_{1i_1},\ldots,s_{Ki_K})\Big), \tag{7.5}$$

where $f_S(s_{ki_k},n_{ki_k}|a_k,b_k) = \binom{n_{ki_k}}{s_{ki_k}}B(a_k+s_{ki_k},b_k+n_{ki_k}-s_{ki_k})/B(a_k,b_k)$ is a marginal beta-binomial density with index n_{ki_k} and parameters (a_k,b_k), and the expression in the second term is

$$D_{\Omega_K}(s_{1i_1},\ldots,s_{Ki_K}) = \sum_{j_1=1}^{K-1}\sum_{j_2=j_1+1}^{K}\omega_{j_1,j_2}\psi(s_{j_1})\psi(s_{j_2}) + \sum_{j_1=1}^{K-2}\sum_{j_2=j_1+1}^{K-1}\sum_{j_3=j_2+1}^{K}\omega_{j_1,j_2,j_3}\psi(s_{j_1})\psi(s_{j_2})\psi(s_{j_3}) + \cdots + \omega_{1,2,\ldots,K}\prod_{j=1}^{K}\psi(s_{ji_j}),$$

where $\psi(s_{ki_k}) = (s_{ki_k} - \mu_k n_{ki_k})/(a_k+b_k+n_{ki_k})$.

The joint posterior density is subsequently obtained as in (7.2) which was also derived earlier and is given by (3.15),

$$h_{\boldsymbol{p}|\boldsymbol{S}}(p_1,\ldots,p_K|s_{1i_1},\ldots,s_{Ki_K},n_{1i_1},\ldots,n_{Ki_K}) = \Big(\prod_{k=1}^{K} f_{p|S}(p_k|s_{ki_k},n_{ki_k})\Big)\left(\frac{1+R_{\Omega_K}(p_1,\ldots,p_K)}{1+D_{\Omega_K}(s_{1i_1},\ldots,s_{Ki_K})}\right), \tag{7.6}$$

where $f_{p|S}(p_k|s_{ki_k},n_{ki_k}) = p_k^{a_k+s_{ki_k}-1}(1-p_k)^{b_k+n_{ki_k}-s_{ki_k}-1}/B(a_k+s_{ki_k},b_k+n_{ki_k}-s_{ki_k})$ is the posterior beta density.

If action P is taken, the design of the randomized controlled trial is based on the log odds ratio summary measure. This design is discussed in Section 4.1.2. At the end of the trial, a frequentist analysis is used to test the null hypothesis, $H_0:\theta_k=0$ where $\theta_k=\log\{p_k(1-p_C)/(p_C(1-p_k))\}$ is the log odds ratio and p_C is the true probability of success of the control arm. The minimum number of patients needed for phase III trial is set to $n_{\text{III}}=1$ as in the preceding chapter.

7.2.1 Expected utility

Let $\mathcal{G}_{\text{Total}}(N,K)$ denote the expected utility of the optimum development plan with a total population of size N and K potential related treatments. If $\mathcal{G}_{\text{Total}}(N,K)>0$, then the development plan commences by recruiting its first batch of $m_{11}=1$ patient to the first phase II trial. Due to the complexity of the correlation among the prior distributions and the sequential nature of the design, it is easier to define the expected utility of the whole programme with an illustration. Consider a development plan with three

potential treatments, that is, $K = 3$. By backward induction methodology, the development plan of the third trial is computed first followed by the computation for the second trial and finally, the computation for the first trial which then gives the optimal strategy for the whole development plan.

7.2.2 Illustration: Expected utilities of a development plan with N_3 and K_3

At the start of the third trial, the size of the population is $N_3 = N - n_{1i_i} - n_{2i_2}$, a smaller size than the original population N where n_{1i_1} is the total number of patients recruited from the first stage up to the i_1-th stage of the first trial before it is aborted in favour of starting the second trial and n_{2i_2} is the total number of patients recruited from the first stage up to the i_2-th stage of the second trial before it is also aborted in favour of starting the third and final trial. Denote the total number of successes observed from the first and second trials by $S_{1i_1} = s_{1i_1}$ and $S_{2i_2} = s_{2i_2}$, respectively. At each interim analysis of the third trial, there are only three choices of actions, namely, actions R, P and A, except when $N_3 - 1$ patients have been recruited, then there would only be actions P and A to choose from.

Expected utility of action A (abandon the programme)

At the i_3-th interim analysis, let $n_{3i_3} = \sum_{j=1}^{i_3} m_{3j}$ be the total number of patients recruited thus far to the third trial and let $S_{3i_3} = \sum_{j=1}^{i_3} X_{3j}$ be the cumulative successes from this trial. Given $S_{3i_3} = s_{3i_3}$ successes, the utility function of action A is less the cost of patients recruited from the first stage up

to and including the i_3-th stage of this trial, that is, $G_A(p_3, N_3, K_3) = -ln_{3i_3}$. Taking expectation by integrating it with respect to p_3 over all its possible value,

$$\mathcal{G}_A(s_{3i_3}, n_{3i_3}, N_3, K_3) = -ln_{3i_3}. \tag{7.7}$$

Expected utility of action P (proceed to phase III trial)

The gain function of action P depends on the success of the phase III trial and the cost of conducting it. The probability that the phase III trial is a success is given by the power function. From equation (6.4) the probability that the null hypothesis, $H_0 : \theta_3 = 0$, is rejected at the upper end of the α-level of significance under the alternative hypothesis is $1 - \Phi\Big(z_{1-\alpha/2} - \theta_3\sqrt{V_3}\Big)$ where $\theta_3 = \log\{p_3(1-p_C)/(p_C(1-p_3))\}$ is the log odds ratio, p_C is the true probability of success of the control arm, $\Phi(\cdot)$ is the cumulative standard normal distribution function and z_γ is the upper 100γ percentile of the standard normal density. Therefore, the gain function of action P is

$$G_P(p_3, N_3, K_3) = 1 - \Phi\Big(z_{1-\alpha/2} - \theta_3\sqrt{V_3}\Big) - lN_3 - l_{\text{III}}.$$

Given the observed data from the preceding two trials and cumulative responses from the first i_3 stages of the third trial, the expected utility of

action P is (the detailed integration by parts is shown in Appendix C)

$$
\begin{aligned}
&\mathcal{G}_P(s_{3i_3}, n_{3i_3}, N_3, K_3) \\
&\quad = \iiint \left(1 - \Phi\Big(z_{1-\alpha/2} - \theta_3\sqrt{V_3}\Big)\right) \\
&\qquad \times h_{\boldsymbol{p}|\boldsymbol{S}}(p_1, p_2, p_3 | s_{1i_1}, s_{2i_2}, s_{3i_3}, n_{1i_1}, n_{2i_2}, n_{3i_3})\, dp_1\, dp_2\, dp_3 - lN_3 - l_{\text{III}} \\
&\quad = \frac{1}{1 + D_{\Omega_3}(s_{1i_1}, s_{2i_2}, s_{3i_3})} \Bigg[\Big(1 + \omega_{12}\psi(s_{1i_1})\psi(s_{2i_2})\Big) A(s_{3i_3}, n_{3i_3}) \\
&\qquad + \Big(\omega_{13}\psi(s_{1i_1}) + \omega_{23}\psi(s_{2i_2}) + \omega_{123}\psi(s_{1i_1})\psi(s_{2i_2})\Big) \\
&\qquad\quad \times \int_0^1 \phi(p_3) \left(1 - \Phi\Big(z_{1-\alpha/2} - \theta_3\sqrt{V_3}\Big)\right) f_{p|S}(p_3 | s_{3i_3}, n_{3i_3})\, dp_3 \Bigg] \\
&\qquad - lN_3 - l_{\text{III}}, \qquad (7.8)
\end{aligned}
$$

where $A(s_{3i_3}, n_{3i_3}) = \int_0^1 \Big(1 - \Phi\big(z_{1-\alpha/2} - \theta_3\sqrt{V_3}\big)\Big) f_{p|s}(p_3 | s_{3i_3}, n_{3i_3})\, dp_3$ is the assurance.

Expected utility of action R (resume the current phase II trial)

The expected utility function of action R at the i_3-th stage of the third trial depends on the responses from the patients recruited to the immediate following stage and the decisions taken thereafter. The computation is similar to that described earlier in Section 6.1.2 except that its expectation is averaged over all the possible responses that may be observed in this stage having observed all the successes from the first trial up to and including the third trial.

Before deriving the expected utility of action R at the i_3-th interim analysis of the third trial, first consider the marginal density of $X_{3,i_3+1} = x_{3,i_3+1}$

given $S_{1i_1} = s_{1i_1}, S_{2i_2} = s_{2i_2}, S_{3i_3} = \sum_{j=1}^{i_3} X_{3j}$. Patients' responses are independent from each other. Therefore, the marginal likelihood of S_{3i_3} has a binomial distribution with index n_{3i_3} and parameter p_3. From the law of succession, the likelihood function of X_{3,i_3+1} is similarly, a binomial distribution but with index m_{3,i_3+1} given $S_{1i_1} = s_{1i_1}, S_{2i_2} = s_{2i_2}$ and $S_{3i_3} = s_{3i_3}$. Thus, as shown in detail in Appendix D, the "posterior" marginal density is

$$
\begin{aligned}
&g_{X|\boldsymbol{S}}(x_{3,i_3+1}, m_{3,i_3+1}|s_{1i_1}, s_{2i_2}, s_{3i_3}, n_{1i_1}, n_{2i_2}, n_{3i_3}) \\
&= \iiint f_{S|p}(x_{3,i_3+1}|p_3) h_{\boldsymbol{p}|\boldsymbol{S}}(p_1, p_2, p_3|s_{1i_1}, s_{2i_2}, s_{3i_3}, n_{1i_1}, n_{2i_2}, n_{3i_3})\, dp_1\, dp_2\, dp_3 \\
&= \frac{1}{1 + D_{\Omega_3}(s_{1i_1}, s_{2i_2}, s_{3i_3})} \Bigg[1 + \omega_{12}\psi(s_{1i_1})\psi(s_{2i_1}) + \Big(\omega_{13}\psi(s_{1i_1}) + \omega_{23}\psi(s_{2i_1}) \\
&\quad + \omega_{123}\psi(s_{1i_1})\psi(s_{2i_1}) \Big) \Big(\frac{s_{3i_3} + x_{3,i_3+1} - \mu_3(n_{3i_3} + m_{3,i_3+1})}{a_3 + b_3 + n_{3i_3} + m_{3,i_3+1}} \Big) \Bigg] \\
&\quad \times f_S(x_{3,i_3+1}|a_3 + s_{3i_3}, b_3 + n_{3i_3} - s_{3i_3}),
\end{aligned} \tag{7.9}
$$

where μ_3 is the expected value of the prior beta distribution with parameters a_3 and b_3, and $f_S(x_{3,i_3+1}|a_3 + s_{3i_3}, b_3 + n_{3i_3} - s_{3i_3})$ is the marginal density of X_{3,i_3+1} with parameters $a_3 + s_{3i_3}$ and $b_3 + n_{3i_3} - s_{3i_3}$.

The expected utility of action R at the i_3-th stage of the third trial is

$$
\begin{aligned}
&\mathcal{G}_R(s_{3i_3}, n_{3i_3}, N_3, K_3) \\
&\quad = \sum_{x_{3,i_3+1}=0}^{m_{3,i_3+1}} \max_{a \in \{R,P,A\}} \Big\{ \mathcal{G}_a(s_{3i_3} + x_{3,i_3+1}, n_{3i_3} + m_{3,i_3+1}, N_3, K_3) \Big\} \\
&\qquad \times g_{X|\boldsymbol{S}}(x_{3,i_3+1}, m_{3,i_3+1}|s_{1i_1}, s_{2i_2}, s_{3i_3}, n_{1i_1}, n_{2i_2}, n_{3i_3}).
\end{aligned} \tag{7.10}
$$

It is solved by backward induction, starting from the terminal branch of the

decision tree and moving back to the present stage.

Expected utility of the whole development plan

After conducting the first two trials, the third trial commences with a population of size N_3 and $K_3 = 1$ treatment. The third trial starts by recruiting $n_{31} = m_{31}$ patients to the first stage of the third treatment. Based on the observed, $S_{31} = s_{31}$, successes from the n_{31} patients and having already observed $S_{1i_1} = s_{1i_1}$ and $S_{2i_2} = s_{2i_2}$ successes from the preceding trials, the expected utility of the whole development plan of the third trial can be obtained. Its expectation is solved by maximizing the expected utility of each action and averaging it over all possible values of S_{31} given S_{1i_1} and S_{2i_2}, and less the cost of conducting a phase II trial,

$$\begin{aligned}
&\mathcal{G}_{\text{Total}}(N_3, K_3 = 1) \\
&\quad = \sum_{s_{31}=0}^{n_{31}} \max_{a\in\{R,P,A\}} \Big\{\mathcal{G}_a(s_{31}, n_{31}, N_3, K_3)\Big\} \\
&\qquad \times g_{X|\boldsymbol{S}}(s_{31}, n_{31}|s_{1i_1}, s_{2i_2}, 0, n_{1i_1}, n_{2i_2}, 0) - l_{\text{II}}, \qquad (7.11)
\end{aligned}$$

where the individual expected utilities are obtained as given in equations (7.7), (7.8) and (7.10), and $g_{X|\boldsymbol{S}}(s_{31}, n_{31}|s_{1i_1}, s_{2i_2}, 0, n_{1i_1}, n_{2i_2}, 0)$ is as given by equation (7.9).

The expression in (7.11) shows that the expected utility of the whole development plan of the third trial with a population of size N_3 and K_3 treatment depends on the subsequent actions. Thus, the optimal strategy of the third trial is solved by backward induction methodology.

7.2.3 Illustration: Expected utilities of a development plan with N_2 and K_2

At the start of the second trial, there are $N_2 = N - n_{1i_1}$ population where n_{1i_1} is the total number of patients recruited from the first stage up to and including the i_1-th stage of the first trial before the optimal action to stop the first trial in favour of the second trial is made. Let $S_{1i_1} = s_{1i_1}$ be the successes observed out of n_{1i_1} patients from the first trial. As the trials are designed to run sequentially, there is no information from the third trial. Also, at the start of the second trial there are $K_2 = K - 1 = 2$ potential treatments.

Expected utility of action A (abandon the programme)

Similar to the utility of action A of the third trial, the utility of action A of the second trial at the i_2-th interim analysis is less the cost of recruiting n_{2i_2} from the first i_2 stages of the second trial, $G_A(p_2, N_2, K_2) = -ln_{2i_2}$. Therefore, having observed $S_{2i_2} = s_{2i_2}$ successes out of n_{2i_2} patients, the expected utility of action A of the second trial is

$$\mathcal{G}_A(s_{2i_2}, n_{2i_2}, N_2, K_2) = -ln_{2i_2}. \tag{7.12}$$

Expected utility of action P (proceed to phase III trial)

The formulation of the expected utility of action P of the second trial is similar to the one in the third trial described above. If the second treatment is recommended for further testing in a phase III setting against a control

arm, the data from this phase III trial are used to test the null hypothesis $H_0 : \theta_2 = 0$ where $\theta_2 = \log\{p_2(1-p_C)/(p_C(1-p_2))\}$ is the log odds ratio and p_C is the true probability of success of the control arm. Let $S_{2i_2} = s_{2i_2}$ be the successes from the first i_2 stages of the second trial, the gain function of action P is

$$G_P(p_2, N_2, K_2) = 1 - \Phi\Big(z_{1-\alpha/2} - \theta_2\sqrt{V_2}\Big) - lN_2 - l_{\text{III}}.$$

Similar to the workings in (7.8), having observed $S_{1i_1} = s_{1i_1}, S_{2i_2} = s_{2i_2}$ and $S_{3i_3} = 0$ because the third trial has yet to commence, as shown in Appendix E, the expected gain function of action P is

$$\begin{aligned}
&\mathcal{G}_P(s_{2i_2}, n_{2i_2}, N_2, K_2) \\
&= \iiint G_P(p_2, N_2, K_2) h_{\boldsymbol{p}|\boldsymbol{S}}(p_1, p_2, p_3 | s_{1i_1}, s_{2i_2}, s_{3i_3}, n_{1i_1}, n_{2i_2}, n_{3i_3})\, dp_1\, dp_2\, dp_3 \\
&= \frac{1}{1 + \omega_{12}\psi(s_{1i_1})\psi(s_{2i_2})} \Bigg[A(s_{2i_2}, n_{2i_2})) \\
&\qquad + \omega_{12}\psi(s_{1i_1}) \int_0^1 \phi(p_2) \Big(1 - \Phi\Big(z_{1-\alpha/2} - \theta_2\sqrt{V_2}\Big)\Big) f_{p|S}(p_2 | s_{2i_2}, n_{2i_2})\, dp_2 \Bigg] \\
&\quad - lN_2 - l_{\text{III}}
\end{aligned} \tag{7.13}$$

where $A(s_{2i_2}, n_{2i_2}) = \int_0^1 \Big(1 - \Phi\big(z_{1-\alpha/2} - \theta_2\sqrt{V_2}\big)\Big) f_{p|s}(p_2 | s_{2i_2}, n_{2i_2})\, dp_2$ is the assurance.

Expected utility of action T (start a new phase II trial)

If the optimal action at i_2-th stage of the second trial is the action T, then the second trial is abandoned and a new phase II trial is initiated with the

remaining $N_3 = N_2 - n_{2i_2}$ patients and $K_3 = K_2 - 1 = 1$ treatment. The expected utility of action T is the expected utility of the whole programme of the third trial less the cost of the patients recruited thus far to the second trial, that is,

$$\mathcal{G}_T(s_{2i_2}, n_{2i_2}, N_2, K_2) = \mathcal{G}_{\text{Total}}(N_3, K_3 = 1) - l n_{2i_2}, \tag{7.14}$$

where $\mathcal{G}_{\text{Total}}(N_3, K_3 = 1)$ is as given in (7.11).

Expected utility of action R (resume the current phase II trial)

The expected utility of action R at i_2-th stage of the second trial depends on the future responses and the resulting actions. Thus, its expectation function is also solved by backward induction. Similar to the expression derived earlier for the third trial, the expectation utility function is

$$\begin{aligned}
&\mathcal{G}_R(s_{2i_2}, n_{2i_2}, N_2, K_2) \\
&\quad = \sum_{x_{2,i_2+1}=0}^{m_{2,i_2+1}} \max_{a\in\{R,P,T,A\}} \Big\{\mathcal{G}_a(s_{2i_2} + x_{2,i_2+1}, n_{2i_2} + m_{2,i_2+1}, N_2, K_2)\Big\} \\
&\qquad \times g_{X|\boldsymbol{S}}(x_{2,i_2+1}, m_{2,i_2+1}|s_{1i_1}, s_{2i_2}, 0, n_{1i_1}, n_{2i_2}, 0),
\end{aligned} \tag{7.15}$$

where

$$\begin{aligned}
&g_{X|\boldsymbol{S}}(x_{2,i_2+1}, m_{2,i_2+1}|s_{1i_1}, s_{2i_2}, 0, n_{1i_1}, n_{2i_2}, 0) \\
&= \frac{1}{1 + \omega_{12}\psi(s_{1i_1})\psi(s_{2i_2})}\big(1 + \omega_{12}\psi(s_{1i_1})\big)\left(\frac{s_{2i_2} + x_{2,i_2+1} - \mu_2(n_{2i_3} + m_{2,i_2+1})}{a_2 + b_2 + n_{2i_2} + m_{2,i_2+1}}\right) \\
&\quad \times f_S(x_{2,i_2+1}|a_2 + s_{2i_2}, b_2 + n_{2i_2} - s_{2i_2}),
\end{aligned} \tag{7.16}$$

is the marginal density of X_{2,i_2+1} given $S_{1i_1} = s_{1i_1}, S_{2i_2} = s_{2i_2}$ and $S_{3i_3} = 0$. The details of the derivation of the marginal density is in Appendix F.

Expected utility of the whole development plan

The second trial commences by recruiting $n_{21} = m_{21}$ patients into the first stage. Having observed $S_{1i_1} = s_{1i_1}$ successes from the first i_1 stages of the first trial and let $S_{21} = s_{21}$ be the successes from the first n_{21} patients, then the expected utility of the whole development plan of the second trial with an initial population of N_2 and K_2 treatments is solved by averaging the maximum expected utility of the available actions over all possible values of S_{21} given S_{1i_1} less the cost of conducting a phase II trial,

$$\begin{aligned}\mathcal{G}_{\text{Total}}(N_2, K_2) = \sum_{s_{21}=0}^{n_{21}} \max_{a\in\{R,P,T,A\}} \Big\{\mathcal{G}_a(s_{21}, n_{21}, N_2, K_2)\Big\} \\ \times g_{X|\boldsymbol{S}}(s_{21}, n_{21}|s_{1i_1}, 0, 0, n_{1i_1}, 0, 0) - l_{\text{II}},\end{aligned} \tag{7.17}$$

where the individual expected utilities are obtained as given in equations (7.12), (7.13), (7.14) and (7.15), and $g_{X|\boldsymbol{S}}(s_{21}, n_{21}|s_{1i_1}, 0, 0, n_{1i_1}, 0, 0)$ is as given by equation (7.16).

7.2.4 Illustration: Expected utilities of a development plan with N_1 and K_1

According to the backward induction methodology, having already first solved the expected utility of the whole programme of the third and final trial followed by the computation of the expected utility of the whole programme of

the second trial, the last step is to solve for the expected utility of the whole programme of the first trial which is also the expected utility function of a series of trials with an initial population of N and K treatments.

At the start of the first trial, the size of the population is $N_1 = N$ and there are $K_1 = K$ potential treatments available for trial. Throughout the conduct of the first trial there is no information from the second and third trials. Therefore, at the i_1-th interim stage of the first trial, there are only S_{1i_1} successes out of n_{1i_1} patients while $S_{2i_2} = S_{3i_3} = 0$.

Expected utility of action A (abandon the programme)

The gain function of action A at i_1-th interim stage of the first trial is less the cost of recruited patients, $G_A(p_1, N_1, K_1) = -ln_{1i_1}$. Therefore, having observed $S_{1i_1} = s_{1i_1}$, the expected utility of action A of the first trial is

$$\mathcal{G}_A(s_{1i_1}, n_{1i_1}, N_1, K_1) = -ln_{1i_1}. \tag{7.18}$$

Expected utility of action P (proceed to phase III trial)

If action P is taken, then the recommended treatment is compared to the control arm in a randomized controlled trial and the data from this phase III trial are used to test the hypothesis $H_0 : \theta_1 = 0$ where $\theta_1 = \log\{p_1(1 - p_C)/(p_C(1 - p_1))\}$ is the log odds ratio. Let $S_{1i_1} = s_{1i_1}$ be the successes from the first stage up to and including the i_1-th stage of the first trial. The gain function of action P is

$$G_P(p_1, N_1, K_1) = 1 - \Phi\Big(z_{1-\alpha/2} - \theta_1\sqrt{V_1}\Big) - lN_1 - l_{\text{III}}.$$

As $n_{2i_2} = 0$ and $n_{3i_3} = 0$, the cumulative successes are also $S_{2i_2} = 0$ and $S_{3i_3} = 0$. Thus, $D_{\Omega_3}(s_{1i_1}, s_{2i_2}, s_{3i_3}) = 0$. From (7.6) and following the similar workings for the second and third trials, the expected utility of action P is

$$
\begin{aligned}
&\mathcal{G}_P(s_{1i_1}, n_{1i_1}, N_1, K_1) \\
&= \iiint G_P(p_1, N_1, K_1) h_{\boldsymbol{p}|\boldsymbol{S}}(p_1, p_2, p_3 | s_{1i_1}, s_{2i_2}, s_{3i_3}, n_{1i_1}, n_{2i_2}, n_{3i_3})\, dp_1\, dp_2\, dp_3 \\
&= \iiint \left(1 - \Phi\Big(z_{1-\alpha/2} - \theta_1 \sqrt{V_1}\Big)\right) f_{p|S}(p_1 | s_{1i_1}, n_{1i_1}) f_p(p_2) f_p(p_3) \\
&\quad \times \Big(1 + \omega_{12}\phi(p_1)\phi(p_2) + \omega_{13}\phi(p_1)\phi(p_3) + \omega_{23}\phi(p_2)\phi(p_3) \\
&\qquad + \omega_{123}\phi(p_1)\phi(p_2)\phi(p_3)\Big)\, dp_1\, dp_2\, dp_3 - lN_1 - l_{\text{III}} \\
&= \int_0^1 \left(1 - \Phi\Big(z_{1-\alpha/2} - \theta_1 \sqrt{V_1}\Big)\right) f_{p|S}(p_1 | s_{1i_1}, n_{1i_1})\, dp_1 - lN_1 - l_{\text{III}} \\
&= A(s_{1i_1}, n_{1i_1}) - lN_1 - l_{\text{III}} \qquad (7.19)
\end{aligned}
$$

where $A(s_{1i_1}, n_{1i_1}) = \int_0^1 \left(1 - \Phi\big(z_{1-\alpha/2} - \theta_1 \sqrt{V_1}\big)\right) f_{p|s}(p_1 | s_{1i_1}, n_{1i_1})\, dp_1$ is the assurance. Note that as there is no information from the subsequent trials, so, the correlation coefficients are not utilized in the expectation of action P in the first phase II trial, that is, the expected utility function of action P does not depend on the prior beliefs of the following treatments. Thus, its expected function is the same as the expected utility of action P when the treatments are independent of each other (Section 6.1.2).

Expected utility of action T (start a new phase II trial)

Action T is to stop the current first trial and initiate a second trial with a different treatment with the remaining patients, $N_2 = N_1 - n_{1i_1}$ and $K_2 =$

$K_1 - 1$ treatments where n_{1i_1} is the number of patients recruited from the first stage up to and including the i_1-th stage of the first trial. The expected utility of action T at the i_1-th interim stage is thus, the expected utility of the whole programme of the second trial less the cost of the patients recruited thus far to the first trial, that is,

$$\mathcal{G}_T(s_{1i_1}, n_{1i_1}, N_1, K_1) = \mathcal{G}_{\text{Total}}(N_2, K_2) - ln_{1i_1}, \tag{7.20}$$

where $\mathcal{G}_{\text{Total}}(N_2, K_2)$ is as given in (7.17).

Expected utility of action R (resume the current phase II trial)

The expected utility of action R at the i_1-th stage of the first trial depends on the responses that are to be observed in the $(i_1 + 1)$-th stage and the resulting actions. The expected utility function is

$$\begin{aligned}
&\mathcal{G}_R(s_{1i_1}, n_{1i_1}, N_1, K_1) \\
&\quad = \sum_{x_{1,i_1+1}=0}^{m_{1,i_1+1}} \max_{a\in\{R,P,T,A\}} \Big\{\mathcal{G}_a(s_{1i_1} + x_{1,i_1+1}, n_{1i_1} + m_{1,i_1+1}, N_1, K_1)\Big\} \\
&\qquad \times g_{X|\boldsymbol{S}}(x_{1,i_1+1}, m_{1,i_1+1}|s_{1i_1}, 0, 0, n_{1i_1}, 0, 0),
\end{aligned} \tag{7.21}$$

where

$$\begin{aligned}
&g_{X|\boldsymbol{S}}(x_{1,i_1+1}, m_{1,i_1+1}|s_{1i_1}, 0, 0, n_{1i_1}, 0, 0) \\
&\qquad = f_S(x_{1,i_1+1}|a_1 + s_{1i_1}, b_1 + n_{1i_1} - s_{1i_1}),
\end{aligned}$$

is the marginal beta-binomial density of X_{1,i_1+1} with index $n_{1i_1}+m_{1,i_1+1}$ and parameters $a_1+s_{1i_1}$ and $b_1+n_{1i_1}-s_{1i_1}$. The form of the expected utility function of action R at the i_1-th stage of the first trial looks similar to the expected utility function of action R when the treatments are independent (Section 6.1.2). However, as the expectation may depend on the subsequent action T which in turns utilizes the information from the prespecified correlation coefficents and the observed data, it is not the same as the one given by equation (6.7). As action R depends on the future optimal actions, its expectation function is solved by backward induction.

Expected utility of the whole development plan

The series of trials with an initial population of size N and K potential treatments begin by recruiting $n_{11}=m_{11}$ patients to the first trial. Let $S_{11}=s_{11}$ be the observed successes from these n_{11} patients. The expected utility of the whole development plan is solved by taking the expectation of the maximum expected utility of the actions over all the possible values of S_{11} less the cost of conducting a phase II trial,

$$\begin{aligned}\mathcal{G}_{\text{Total}}(N,K) &= \mathcal{G}_{\text{Total}}(N_1,K_1)\\ &= \sum_{s_{11}=0}^{n_{11}} \max_{a\in\{R,P,T,A\}}\{\mathcal{G}_a(s_{11},n_{11},N_1,K_1)\}f_S(s_{11}) - l_{\text{II}}, \qquad (7.22)\end{aligned}$$

where the individual expected utilities are as given in equations (7.18), (7.19), (7.20) and (7.21), and $f_S(s_{11})$ is the marginal beta-binomial density with index n_{11} and parameters a_1 and b_1.

7.3 Application

The proposed design is illustrated with the same asthma clinical examples seen earlier in Section 6.3. Suppose that there are $K = 3$ potential drugs from the family of mediator antagonists. The primary endpoint is binary where at least an episode of moderate or severe exacerbation in a four-week treatment period is considered as a failure and a non-event as a success. The definitions of moderate and severe asthma exacerbations are as given earlier in the preceding chapter. The unknown probability of success of each treatment, p_k, $(k = 1, 2, 3)$, is assumed to be random and follows the beta distribution with known parameters a_k and b_k.

In a multivariate setting, assuming that $p_1, p_2, \ldots, p_K$ has the joint probability density function as defined in (7.4), Lee (1996) shows that any subvector $p_{i_1}, p_{i_2}, \ldots, p_{i_m}$, $1 \leq i_1 < i_2 < \ldots < i_m \leq K$, has the following joint density

$$h_{\boldsymbol{p}}(p_{i_1}, \ldots, p_{i_m}) = \Big(\prod_{j=1}^{m} f_p(p_{i_j})\Big)\Big(1 + R_{\Omega_m}(p_{i_1}, \ldots, p_{i_m})\Big), \tag{7.23}$$

where $R_{\Omega_1} = 0$ and Ω_m is a subset of Ω_K such that Ω_m is only for the combinations of m variables, $p_{i_1}, p_{i_2}, \ldots, p_{i_m}$.

For this illustration, consider a subvector with only two variables out of the given $K = 3$ variables. As seen earlier in Section 3.1.3, the correlation coefficient between p_i and p_j is given by $\rho_{ij} = \omega_{ij}\sigma_i\sigma_j$ where σ_i (σ_j) is the

standard deviation of p_i (p_j), and ω_{ij} satisfies the condition

$$\max\left\{\frac{-1}{\mu_i\mu_j}, \frac{-1}{(1-\mu_i)(1-\mu_j)}\right\} \leq \omega_{ij} \leq \min\left\{\frac{1}{\mu_i(1-\mu_j)}, \frac{1}{\mu_j(1-\mu_i)}\right\}.$$

For a bivariate density of independent beta densities, the range is,

$$\frac{-(a_i+b_i)(a_j+b_j)}{\max\{a_ia_j, b_ib_j\}} \leq \omega_{ij} \leq \frac{(a_i+b_i)(a_j+b_j)}{\max\{a_ib_j, a_jb_i\}}.$$

Based on this subvector relationship, the range of values for ω_{12}, ω_{13}, and ω_{23} and correspondingly, the correlation coefficients ρ_{12}, ρ_{13} and ρ_{23} can be obtained. The correlation between (p_1, p_2) and p_3 are defined by ω_{123} and as it is difficult to elicit the possible values of this correlation, it is assumed to be $\omega_{123} = 0$.

All the three treatments are assumed to have the same prior densities and so the subscript k will be suppressed for convenience, that is, $p \sim Beta(a, b)$. The range of values for the correlation coefficient between any two treatments is then

$$\frac{-ab}{(a+b+1)\max\{a^2, b^2\}} \leq \rho \leq \frac{1}{a+b+1}, \tag{7.24}$$

and

$$\frac{-(a+b)^2}{\max\{a^2, b^2\}} \leq \omega \leq \frac{(a+b)^2}{ab}. \tag{7.25}$$

The computation of the expected utility function of the whole series is very tedious and as such, only $N = 100$ is considered for illustration. Following the same assumed values as in the preceding chapter, the probability of success of the control arm is assumed to be, $p_C = 0.80$, and the two-

sided significance level is $\alpha = 0.05$. The known parameters of the prior beta density function are $a = 69$ and $b = 6$, which is equivalent to obtaining information from a sample size of 75. From equations (7.24), the possible range of the correlation coefficients is $[-0.00114, 0.013]$ and the corresponding ω range is $[-1.1813, 13.4239]$. For illustration, $\omega = -1$ (correspondingly, $\rho = -0.0009684$) and $\omega = 10$ (correspondingly, $\rho = 0.009684$) are used. The variable and fixed costs are taken to be relative to one unit of gain. The cost per patient is $l = 0.0001$, the cost of conducting a phase II trial is $l_{\text{II}} = 0.002$ and the cost of conducting a phase III trial is $l_{\text{III}} = 0.02$. At each stage of the k-th decision-theoretic phase II trials, $m_{ki_k} = 1$ patient is recruited and the minimum sample size for the phase III trial is set to 1.

The expected utility of the whole development plan when the treatments are negatively correlated, $\rho = -0.0009684$, is 0.465356 (Table 7.1). As the expected utility is greater than zero, it is worth starting the development programme. Figure 7.1(a) shows the decision rules for the optimal actions when given an initial population of $N_1 = N = 100$ and $K_1 = K = 3$ potential treatments. The optimal sample size needed to proceed to a phase III trial is 4. If all of the patients responded positively to the treatments then the optimal action is action P. Otherwise, the optimal action is action T, that is, a new phase II trial is to commence with a population of $N_2 = 96$ and $K_2 = 2$ treatments.

Suppose that action T was taken after 4 patients had been recruited to the first trial. The maximum number of patients needed in order to make a definite decision, that is, action P, T or A in the second phase II trial is 5. If the accumulated successes is 5 out of 5 then the optimal action is

Table 7.1: Expected utility of a whole programme of population N and K treatments for various correlation coefficients, ρ.

Programme with $N = N_i$ population, $K = K_i$ treatments	$\mathcal{G}_{\text{Total}}(N_i, K_i)$		
	$\rho = -0.0009684$	$\rho = 0.009684$	$\rho = 0$
$N_1 = 100, K_1 = 3$	0.465356	0.465034	0.467494
$N_2 = 96, K_2 = 2$	0.452369	0.451522	0.463745
$N_3 = 91, K_3 = 1$	0.425837	0.424690	0.452788

action P and if less than 5, the optimal action is action T (Figure 7.1(b)). If the latter action is taken, the third trial with the third treatment is to be initiated with a population of $N_3 = 91$. As this is the last treatment, action T is not available at each interim stage. Regardless of the response from the first patient, the optimal action is to go on to the phase III trial (action P) and so 90 patients will be randomized to either the third treatment or to the placebo in a 1:1 ratio.

In the next example, consider the case of positive correlation between any two treatments, $\rho = 0.009684$. The expected utility of the whole programme is 0.465034, slightly lower than the case when the treatments are weakly and negatively correlated (Table 7.1). However, similar to the case of negative correlation, the maximum sample size that is needed in order to make a definitive decision in the first phase II trial is 4 and of this, if all are successes then the optimal action is action P, otherwise it is action T (Figure 7.1(c)). For the latter action, the second trial will commence with a population of $N_2 = 96$ and $K_2 = 2$ potential treatments. The maximum sample size for the second phase II trial is also 5 where a definitive decision is made to either proceed to a phase III trial or to stop the current trial and initiate the third

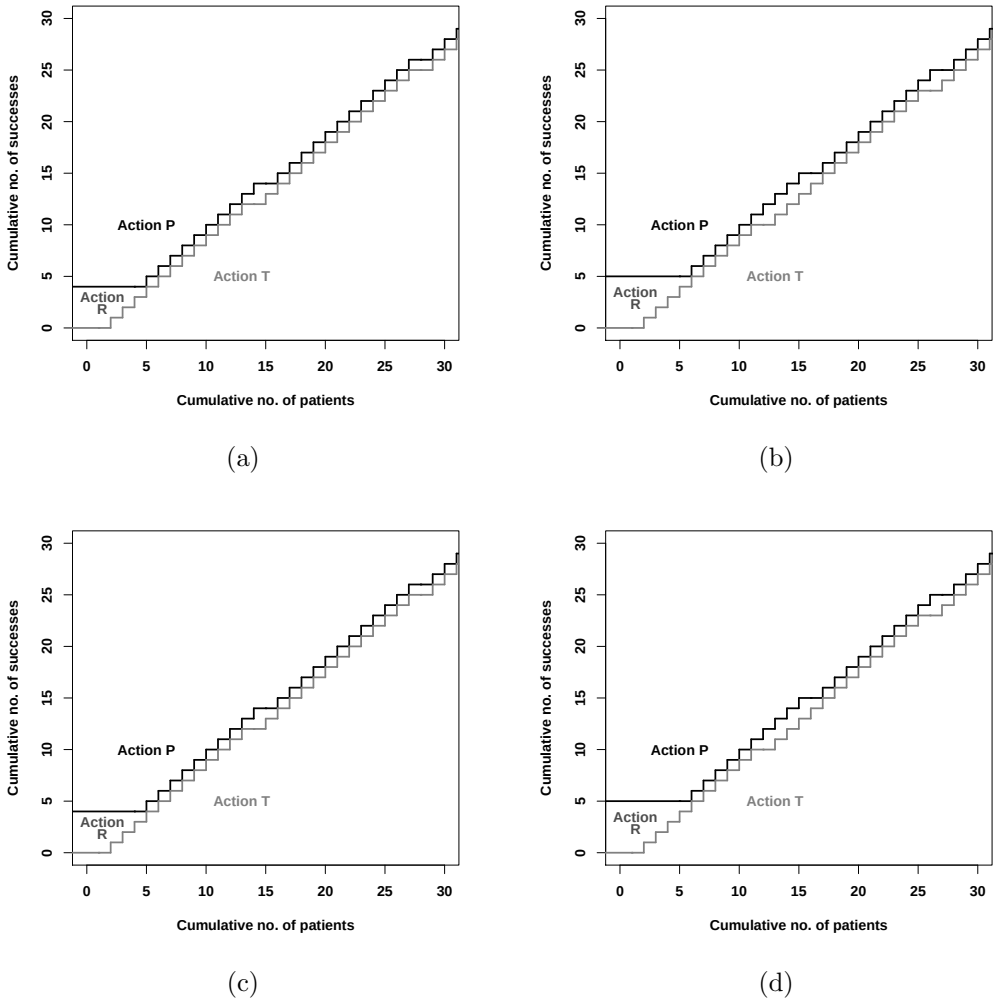


Figure 7.1: Decision rules for optimal actions for a series of related treatments for $\rho = -0.0009684$ (upper row) and $\rho = 0.009684$ (lower row).

and final trial with the last available treatment (Figure 7.1(d)). Also, if the third trial is to commence with $N_3 = 91$ patients and $K_3 = 1$ treatment, the optimal action after observing the response from the first patient is action P, regardless if it is a failure or success.

The illustrations given here are limited due to the narrow range of a

correlation coefficient between two equal and highly informative prior densities. In the example of negative correlation ($\rho = -0.0009684$), it is also the weaker correlation as it is nearer to 0 compare to the positive correlation ($\rho = 0.009684$). Interestingly, the expected utility of the whole programme with an initial population of $N_1 = 100$ and $K_1 = 3$ treatments when the treatments are independent $\rho = 0$ is the highest (Table 7.1), followed by the series of trials that is negatively and weaker correlated. The series of trials that is positively and higher correlated has the lowest expected utility.

The maximum number of patients that is needed in the first phase II trial to proceed to the phase III trial for both series of trials with related treatments is 4 whereas if the treatments are independent, the maximum number of patients needed is 6. In the case of independent treatments, nothing more is known about subsequent treatments except their prior densities and as such, as the prior density of subsequent treatments is highly informative, it is worthier to try a new treatment than to proceed to the larger definitive phase III trial. Contrary to the series of related treatments, as information of the current trial is gathered, more is known about the posterior densities of subsequent treatments and it seems to be worthier to go on to the larger phase III trial earlier.

For the second trial to commence with $N_2 = 96$ patients, the first trial of the correlated series would have observed 3 successes out of the 4 patients. The expected utility of the whole programme depends on the subsequent actions and as the information from the first trial suggest that the treatment is slightly worse than expected (as having observed one failure so soon in the trial), it affects the posterior densities of subsequent treatments much more

than when they are independent. Thus, the expected utilities of the series of related trials are much lower; 0.452369 for the negative correlated trials, 0.451522 for the positive correlated trials, and 0.463745 for the independent trials.

For the third trial to commence with $N_3 = 91$ patients, the second trial would have observed 4 successes only. Clearly, as both first and second trials seem to have "failed", the belief of the third treatment to be more effective than the placebo is affected which in turn affects the expected utility of the whole series.

Simulation

A simple simulation was conducted to estimate the distribution of the number of phase II trials before proceeding to a phase III trial. The algorithm and codes of the simulation are given in Appendix G. The optimal action in the third phase II trial is always action P, proceed to the phase III trial, regardless of the observed response. Therefore, the simulation study was only on the scenario of the first two phase II trials. The results for both $\rho = -0.0009684$ and $\rho = 0.009684$ are about the same (Table 7.2).

As expected, the probability of the first treatment to proceed to a phase III trial is higher than the second treatment (0.72 against 0.67). The average sample size required for the first phase II trial is 3.5 which is close to the required maximum size 4. This is unsurprising as the probability to proceed to phase III trial is rather high. The average sample size for the second phase II trial is 4.3 which is comparatively slightly lower than the required maximum size since the probability to proceed to phase III trial is not as

Table 7.2: Simulation results for both negative and positive correlation coefficients, ρ.

	$\rho = -0.0009684$	$\rho = 0.009684$
Average sample size for first phase II trial	3.5462	3.5462
Probability of going to phase III after one phase II trial	0.7187	0.7187
Average sample size for second phase II trial	4.2769	4.2732
Probability of going to phase III after two phase II trials	0.6680	0.6673

high as for the first phase II trial.

7.4 Discussion and concluding remarks

The prior density assumed for illustration is highly informative. As a result, the maximum number of patients needed to make a definitive decision is very small. Based on the assumed joint probability density function, $h_{\boldsymbol{p}}$, for parameters $p_1, p_2, \ldots, p_K$, the correlation coefficient is within a very narrow range and they are also very close to 0 which is the independent case.

The computational time needed to solve for the optimal decision strategy is very long and so it is not feasible to explore further in this thesis the characteristics of the sequential scheme under various scenarios; a wider range of correlation coefficients with different prior densities.

The prior densities for all the treatments are assumed to be the same. This is not unreasonable considering that all treatments may be equally effective before any trial is undertaken. However, in the scenario where a large funding body such as a public sector or a charity organization is funding a

series of clinical trials with treatments from different pharmaceutical companies for the same population, the prior densities may be different. Also, treatments from the same company may be highly correlated than treatments from different companies. The formulation given in Section 7.2.1 can be used to allow for different correlation coefficients but the computation can be quite taxing and challenging.

The result from the simulation study showed that the ordering of the treatments to be evaluated in the phase II setting affects the probability of any given treatment being evaluated further in the phase III setting. The treatments in this example are assumed to be equally promising and thus, they are interchangeable. Accordingly, the treatment order can be random. However, in practice, even if identical prior distributions are used it is likely that some treatments may be considered to be more promising. Therefore, these treatments should be chosen to be evaluated first so that they may have higher chance to be evaluated further in the phase III setting.

Part IV

Summary

Filling up the holes with animation
Covering the flaws in the construction
Wiping all the scenic ostentation
Knowing it's a "Macintosh" production
Working for a tiny compensation
Hoping for a thunderous ovation
The art of making art is putting it together

Stephen Sondheim

Putting It Together

Chapter 8

Summary

8.1 Conclusion

This thesis proposed designs for a series of trials. The motivation is based on a pharmaceutical company that has developed a few treatments concurrently for the same population. In another scenario, a charity organization or a public sector with large fundings may be interested to identify a promising treatment from a pool of treatments for the same population.

Resources such as patients and money are essentially finite and limited. As such, the decision from each trial will affect subsequent trials either in the design of the next trial or allocation of resources to other trials. Therefore, it seems intuitive and appealing to design clinical trials in the context of other trials so that the clinical development plan is more cohesive (Senn, 1996). As it is essentially a decision problem whether to recommend a treatment for trial or further development, the statistical decision theory seems to be an obvious choice to model the clinical development plan (Julious and Swank,

2005). The method allows the quantification of the risk of futile treatments, the value of information, the gains from success trials and the losses from futile trials which in turn aid decision making.

The methodology used in the designs is based on a hybrid approach which is a combination of both classical frequentist and Bayesian frameworks. In the design for a series of trials (Chapter 5), it is assumed that at the end of each trial, the responses observed from within each trial are analysed using the classical frequentist analysis to test the null hypothesis. A point estimate of the treatment efficacy is reported with its confidence interval and the corresponding p-value. Therefore, the design of the trial is based on the final analysis that is to be done. However, the unknown parameter of the treatment efficacy is assumed to be random and follows a prior distribution. Thus, the methodology used at the design stage is a Bayesian one.

Following on, in the design for a series of decision-theoretic trials (Chapters 6 and 7), a series of phase II trials are designed based on the Bayesian decision-theoretic approach where an optimal decision is chosen based on the expected utilities of each actions. As above, the unknown parameter of the treatment efficacy is also assumed to be random and follows a prior distribution. If the treatment from the current phase II trial is recommended for further testing in a phase III trial setting, it is assumed that at the end of the phase III trial, the observed data from this trial are analysed using the classical frequentist analysis to test the null hypothesis. Therefore, the conclusion is based on the inference from the trial unlike in a fully Bayesian inferential technique where the conclusion is based on the posterior distribution.

Beliefs of how effective a treatment may be are required in sample size

calculation and quite often these beliefs are elicited from experts either as the best guess or the minimum clinically meaningful effect. In a frequentist setting, these beliefs form the null and alternative hypotheses. There is inevitably some subjectivity and variability on these beliefs and the Bayesian approach quantifies these variability. Bayesian approach takes into account the different results and other available evidence and information to develop the prior distribution. In addition, it puts the current trial in the context of previous trials regardless if the trials have very similar or very different settings. There are many published works on how to elicit prior beliefs from experts, for example, Chaloner *et al.* (1993), Chaloner and Rhame (2001), O'Hagan *et al.* (2006) and Stangl and Berry (1998). Another technique that may be used and which is used in this thesis is to obtain prior beliefs empirically from published literatures. It is not as proficient as eliciting from experts and there may be publication bias as there are more effective trials than futile trials being published. A reference prior, the standard uniform prior, together with the prior elicited from published literatures are used to illustrate the application of the designs.

The optimal sample size obtained with the proposed designs is, on average, smaller than the usual sample size obtained with the frequentist approach. Although it is difficult to compare the sample sizes directly between a frequentist method and a hybrid method, consider the following example. Let the probability of success of a historical control be 0.80 and it is desired to detect a minimum clinical meaningful success rate of 0.92 from the experimental treatment arm. Fixing the type I and II error rates to $\alpha = 0.05$ and $\beta = 0.10$, at a two-sided significance test, the minimum sample size is

89. From Chapter 6, the unknown parameter, p, follows the beta density with parameters $a = 69$ and $b = 6$ with an expected value of 0.92, that is, the average success rate of the experimental treatment is believed to be 0.92. The minimum number of patients required for the phase II trial in order to proceed to a phase III trial is 24 whereas the maximum number of patients is 79, ten patients less than the frequentist approach.

Thus, an advantage over the frequentist approach as trials can be conducted more efficiently. However, in the series of decision-theoretic trials, the sampling is a fully sequential one and as such, it is not practical if the duration that is needed to observe the primary endpoint is more than a few weeks as that will increase the time to conduct a trial and is contrary to an efficient design.

8.2 Further works

The final analysis at the end of a trial is based on the primary endpoint using all the data from the patients. Traditionally, subgroup analysis is also routinely performed regardless if it has been prespecified in the analysis plan or on a post-hoc basis. In order for the result of subgroup analysis to be admissible for approval by the regulatory agency it is necessary to prespecify the subgroup and the type of analysis to be performed. Also of importance, the sample size has to be adequate so that both principal and subgroup analyses can be performed (EMEA, 2011). Therefore, it would be of interest to consider subgroup analysis in the proposed design. In the sequential decision-theoretic design, the subgroup analysis may be built in

each interim stage or only at the end of the trial or both interim and final analyses.

Julious and Swank (2005) suggested using statistical decision theory in the design of a series of trials but the probability of success of each trial is different from each other. Thus, it is of interest to expand the proposed designs by considering different priors for all the treatments. Consequently, this may affect the optimality of the development plan. The expected utility of the whole series can be evaluated for each permutation of sequence of treatments and so the ordering of the treatments to be evaluated in the phase II setting is the one that gives the highest expected utility for the whole series. In addition, the specification of prior distribution should not be restricted to the experimental treatments only but should also be applied to the control arm. Some of the more straightforward and obvious extensions that can be applied to the proposed designs are given below.

A Series of Hybrid Trials

The framework of the design of a series of hybrid trials is based on the assumption that each of the trial is a single-arm trial (Chapter 5). The formulation can be extended by considering each trial as a two-arm trial. Following the same framework introduced in Chapter 5, let us assume that the primary endpoint is a continuous variable. Let δ denote the mean difference between the two treatments. From Section 4.1, $\delta \sim N(\theta, (\sigma_1^2 + r\sigma_2^2)/(rn_2))$ where θ is an unknown parameter, σ_1^2 and σ_2^2 are the known variances of the control and experimental arms, respectively, r is the allocation ratio and n_2 is the sample size for the experimental treatment. Assuming that at the end of the trial the

classical frequentist testing is done at α-level of significance. The unknown mean of the random variable, θ, is assumed to be random and follows a prior distribution, $f_\Theta(\theta)$. From equations (4.3) and (4.4) the assurance is

$$A(n) = \int_\Theta \left[1 - \Phi\left(z_{1-\alpha/2} - \frac{\theta_A - \theta_0}{\sqrt{(\sigma_1^2 + r\sigma_2^2)/(rn_2)}}\right)\right] f_\Theta(\theta)\, d\theta,$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution and z_γ is the lower 100γ percentile of the standard normal distribution.

The design can also be modified by considering equal and unequal variances. It can also be extended to assume that the variances are unknown and follow some prior distributions.

The primary endpoint thus far, is assumed to be continuous and its likelihood function is a normal distribution. It is also of interest to modify the design by considering other endpoints with different likelihood functions.

A Series of Decision-Theoretic Trials

The proposed design has considered a binary outcome where the data is assumed to follow the binomial distribution while its parameter is assumed to follow the beta distribution. It is of interest to modify the design by considering other endpoints with different likelihood functions and prior distributions, for example, normal distribution (for example, the forced expiratory volume in one second, FEV_1), survival function (for example, the time to first exacerbation), and Poisson regression (for example, the rate of exacerbation in a predefined period).

This design assumes that the population, primary endpoint and treatment duration for both phase II and III trials are the same. This is usually not the case in practice as phase II trials have shorter treatment period whilst phase III trials aim to examine longer treatment effect. It may be more realistic to consider different endpoints for the phase II and III trials. In such scenario, the phase II primary endpoint and the phase III primary endpoint are correlated and this can be incorporated into the design albeit more challenging.

The design is illustrated with fully sequential sampling. The advantage is that it is more efficient; the minimum number of patients needed for phase II is fewer than what would be needed in a frequentist setting. In practice, however, group sequential may be more preferred because it is not pragmatic to monitor patients continuously and stop the trial frequently for a long period if a few weeks or months are needed to observe a response. It would be straightforward to implement group recruitment as described in Section 6.1 but it may lead to a decrease in the expected gain.

8.3 Discussion

In the proposed design, all the treatments are assumed to be equally effective and have the same prior distribution. Consequently, they are interchangeable and so the ordering of the treatments to be admitted into trials may be random. However, as seen from the simulation results (see Section 7.3), the ordering does affect the chance of any given treatment being recommended for phase III trial evaluation.

The first treatment is more likely to go on to the phase III trial, mainly because if it is successful in the phase II setting, the remaining treatments will not be tested. However, as the development plan progresses, the size of the population gets smaller and there is a greater urgency to promote a treatment to a phase III trial. Thus, less positive results are required to justify further testing in a phase III trial.

The population in a trial is never truly homogeneous. If a treatment is not statistical significantly different than a control treatment in the final confirmatory analysis, it may be statistically different for a subgroup of patients. Patients in this subgroup are usually thought to be more homogeneous as they have been clearly defined by a narrower clinical characteristics such as sex, age and ethnicity.

Some of the newer treatments developed in the past few years are targeted therapies; designed to target the affected receptors or pathways in order to control, modify, suppress or kill the cause of disease. Such targeted therapies although may be more effective and beneficial to the patients, combined with clinical characteristics have further delineated patients into a much smaller subgroup.

These issues may pose a dilemma on the design and conduct of clinical trials because patients would need to fulfill much more stringent eligibility criteria which consequently affects the size of the total population. In an expert workshop on subgroup analysis conducted by the European Medicines Agency (EMEA, 2011) in November 2011, there is a call for closer and early dialogue between industry and regulatory agency or health technology assessment bodies in order to incorporate relevant subgroups analyses in the

designs of clinical trials as this will help in guiding the assessment and interpretation of treatments efficacies.

Some novel clinical trials designs have been proposed for the targeted therapies, and of particular interest are the Bayesian adaptive designs. In the recent years, Bayesian adaptive designs have generated much interest and excitement in the research community especially when a couple of clinical trials adopted this methodology to deliver targeted therapies to patients with the same condition. A recently concluded and published clinical trial is the BATTLE trial, a phase II trial that adaptively randomized patients with refractory non-small cell lung cancer to one of the four targeted treatments (Kim *et al.*, 2011, Zhou *et al.*, 2008).

Another trial that was launched in 2010 is the I-SPY 2 trial where patients with locally advanced breast cancer are randomized to one of the 12 different cancer drugs in a neoadjuvant setting. The design allows treatments that are not promising to be dropped from the trial and those that have shown sufficient efficacy are pass on to be tried in larger phase III trial (Barker *et al.*, 2009, Printz, 2010). The underlying scenario of this programme is very similar to the motivation of this PhD project where a large funder is interested to identify promising treatments from a pool of available treatments.

In the I-SPY 2 trial, patients are randomized to treatment regimens based on their biomarker signatures which is analogous to stratifying patients in the design stage. At the interim stage, Bayesian predictive probability methodology is used to recommend the drug to be dropped from the trial or not. For the latter action, the biomarker signature(s) that correspond(s) to the "graduated" drug will also graduate from the trial. As patients are already

stratified according to their molecular characteristics, this is comparable to performing "subgroup analysis".

The advantage of the adaptive designs cited above is that a few treatments can be tried concurrently. However, this may not be feasible if the population is small and as such, alternative designs that are "customized to address the clinical research question and study population" should be considered (Institute of Medicine, 2001, p. 11). The proposed series of Bayesian decision-theoretic phase II trials and one phase III trial aims to address this issue. In addition, treatments targeting the same population are inevitably correlated, regardless of whether they are highly or lowly correlated. Although it may be computationally challenging, it is worthy to investigate further its effect on the design. It is also of interest to incorporate subgroup analysis formally in the proposed design based on a hybrid methodology.

Part V

Appendices

Appendix A

Derivation of the Unconditional Joint Density of $X_1, \ldots, X_k$ from the Sarmanov's Family

Let $X_i \sim Bin(n_i, p_i)$ and $p_i \sim Beta(a_i, b_i)$, for $i = 1, 2, \ldots, k$. Let $f_{X|p}(x_i|p_i)$ denote the marginal likelihood function of X_i,

$$f_{X|p}(x_i|p_i) = \binom{n_i}{x_i} p_i^{x_i}(1-p_i)^{n_i - x_i}.$$

Let $f_p(p_i)$ denote the marginal prior density of p_i,

$$f_p(p_i) = \frac{1}{B(a_i, b_i)} p_i^{a_i - 1}(1-p_i)^{b_i - 1}.$$

Let $h_{\boldsymbol{X}|\boldsymbol{p}}(X_1 = x_1, \ldots, X_k = x_k|p_1, \ldots, p_k)$ denote the joint likelihood

function of $X_1, \ldots, X_k$,

$$h_{\boldsymbol{X}|\boldsymbol{p}}(X_1 = x_1, \ldots, X_k = x_k|p_1, \ldots, p_k) = \prod_{i=1}^{k} f_{X|p}(x_i|p_i).$$

Finally, let $h_{\boldsymbol{p}}(p_1, \ldots, p_k)$ denote the joint prior density of $p_1, \ldots, p_k$,

$$h_{\boldsymbol{p}}(p_1, \ldots, p_k) = \Big(\prod_{i=1}^{k} f_p(p_i)\Big)\Big(1 + R_{\Omega_k}(p_1, \ldots, p_k)\Big),$$

where

$$\begin{aligned} R_{\Omega_k}(p_1, \ldots, p_k) = &\sum_{i_1=1}^{k-1} \sum_{i_2=i_1+1}^{k} \omega_{i_1,i_2} \phi(p_{i_1}) \phi(p_{i_2}) \\ &+ \sum_{i_1=1}^{k-2} \sum_{i_2=i_1+1}^{k-1} \sum_{i_3=i_2+1}^{k} \omega_{i_1,i_2,i_3} \phi(p_{i_1}) \phi(p_{i_2}) \phi(p_{i_3}) \\ &+ \cdots + \omega_{1,2,\ldots,k} \prod_{i=1}^{k} \phi(p_i). \end{aligned}$$

The mixing function is $\phi(p_i) = p_i - \mu_i$ where $\mu_i = a_i/(a_i + b_i)$ is the expected value of p_i.

Let $h_{\boldsymbol{X}}(x_1, \ldots, x_k)$ denote the unconditional joint density of $X_1, \ldots, X_k$ which is obtained by integrating the product of $h_{\boldsymbol{X}|\boldsymbol{p}}(X_1 = x_1, \ldots, X_k =$

$x_k|p_1, \ldots, p_k)$ and $h_{\boldsymbol{p}}(p_1, \ldots, p_k)$ with respect to $p_1, \ldots, p_k$,

$$
\begin{aligned}
&h_{\boldsymbol{X}}(x_1, \ldots, x_k) \\
&= \int \cdots \int h_{\boldsymbol{X}|\boldsymbol{p}}(X_1 = x_1, \ldots, X_k = x_k|p_1, \ldots, p_k) h_{\boldsymbol{p}}(p_1, \ldots, p_k)\, dp_1 \ldots\, dp_k \\
&= \int \cdots \int \Big(\prod_{i=1}^{k} f_{X|p}(x_i|p_i) f_p(p_i) \Big) \Big(1 + R_{\Omega_k}(p_1, \ldots, p_k) \Big)\, dp_1 \ldots\, dp_k \\
&= \int \cdots \int \Big(\prod_{i=1}^{k} f_{X|p}(x_i|p_i) f_p(p_i) \Big)\, dp_1 \ldots\, dp_k \\
&\quad + \int \cdots \int \Big(\prod_{i=1}^{k} f_{X|p}(x_i|p_i) f_p(p_i) \Big) R_{\Omega_k}(p_1, \ldots, p_k)\, dp_1 \ldots\, dp_k
\end{aligned} \tag{A.1}
$$

Let $f_X(x_i)$ denote the marginal density of X_i which is a beta-binomial density, $f_X(x_i) = \binom{n_i}{x_i} Beta(a_i + x_i, b_i + n_i - x_i)/Beta(a_i, b_i)$. Then the first term of (A.1) integrates to the product of all the marginal densities of X_i,

$$
\prod_{i=1}^{k} \int \cdots \int \Big(f_{X|p}(x_i|p_i) f_p(p_i) \Big)\, dp_1 \ldots\, dp_k = \prod_{i=1}^{k} f_X(x_i). \tag{A.2}
$$

For the integration of the second term of (A.1), first consider a general

integrand that will appear iteratively,

$$\begin{aligned}
&\int_0^1 \phi(p_i) f_{X|p}(x_i|p_i) f_p(p_i)\, dp_i \\
&= \int_0^1 (p_i - \mu_i) \binom{n_i}{x_i} \frac{1}{B(a_i, b_i)} p_i^{a_i + x_i - 1} (1 - p_i)^{b_i + n_i - x_i - 1}\, dp_i \\
&= \binom{n_i}{x_i} \frac{1}{B(a_i, b_i)} \int_0^1 p_i^{a_i + x_i} (1 - p_i)^{b_i + n_i - x_i - 1}\, dp_i - \mu_i f_X(x_i) \\
&= \binom{n_i}{x_i} \frac{B(a_i + x_i + 1, b_i + n_i - x_i)}{B(a_i, b_i)} - \mu_i f_X(x_i) \\
&= \Big(\frac{a_i + x_i}{a_i + b_i + n_i} - \mu_i \Big) f_X(x_i) \\
&= \Big(\frac{(a_i + x_i)(a_i + b_i) - a_i(a_i + b_i + n_i)}{(a_i + b_i)(a_i + b_i + n_i)} \Big) f_X(x_i) \\
&= \Big(\frac{x_i - \mu_i n_i}{a_i + b_i + n_i} \Big) f_X(x_i)
\end{aligned}$$

For ease of notation, let $\psi(x_i) = (x_i - \mu_i n_i)/(a_i + b_i + n_i)$ then

$$\int_0^1 \phi(p_i) f_{X|p}(x_i|p_i) f_p(p_i)\, dp_i = \psi(x_i) f_X(x_i). \tag{A.3}$$

The integration of the second term of (A.1) is done iteratively, first inte-

grate it with respect to p_1,

$$
\begin{aligned}
&\int_0^1 \Big(\prod_{i=1}^{k} f_{X|p}(x_i|p_i) f_p(p_i) \Big) R_{\Omega_k}(p_1, \ldots, p_k)\, dp_1 \\
&= \Big(\prod_{i=2}^{k} f_{X|p}(x_i|p_i) f_p(p_i) \Big) \Big\{ \Big(\sum_{i=2}^{k} \omega_{1i} \phi(p_i) + \sum_{i_1=2}^{k-1} \sum_{i_2=3}^{k} \omega_{1,i_1,i_2} \phi(p_{i_1}) \phi(p_{i_2}) \\
&\qquad + \cdots + \omega_{1,\ldots,k} \prod_{i=2}^{k} \phi(p_i) \Big) \int_0^1 \phi(p_1) f_{X|p}(x_1|p_1) f_p(p_1)\, dp_1 \\
&\quad + \Big(\sum_{i_1=2}^{k-1} \sum_{i_2=i_1+1}^{k} \omega_{i_1,i_2} \phi(p_{i_1}) \phi(p_{i_2}) + \sum_{i_1=2}^{k-2} \sum_{i_2=i_1+1}^{k-1} \sum_{i_3=i_2+1}^{k} \omega_{i_1,i_2,i_3} \phi(p_{i_1}) \phi(p_{i_2}) \phi(p_{i_3}) \\
&\qquad + \cdots + \omega_{2,\ldots,k} \prod_{i=2}^{k} \phi(p_i) \Big) \int_0^1 f_{X|p}(x_1|p_1) f_p(p_1)\, dp_1 \Big\} \\
&= f_X(x_1) \Big(\prod_{i=2}^{k} f_{X|p}(x_i|p_i) f_p(p_i) \Big) \Big\{ \psi(x_1) \Big(\sum_{i=2}^{k} \omega_{1i} \phi(p_i) \\
&\qquad + \sum_{i_1=2}^{k-1} \sum_{i_2=i_1+1}^{k} \omega_{1,i_1,i_2} \phi(p_{i_1}) \phi(p_{i_2}) + \cdots + \omega_{1,\ldots,k} \prod_{i=2}^{k} \phi(p_i) \Big) \\
&\quad + \Big(\sum_{i_1=2}^{k-1} \sum_{i_2=i_1+1}^{k} \omega_{i_1,i_2} \phi(p_{i_1}) \phi(p_{i_2}) + \sum_{i_1=2}^{k-2} \sum_{i_2=i_1+1}^{k-1} \sum_{i_3=i_2+1}^{k} \omega_{i_1,i_2,i_3} \phi(p_{i_1}) \phi(p_{i_2}) \phi(p_{i_3}) \\
&\qquad + \cdots + \omega_{2,\ldots,k} \prod_{i=2}^{k} \phi(p_i) \Big) \Big\}. \qquad \text{(A.4)}
\end{aligned}
$$

Subsequently, integrate (A.4) with respect to p_2,

$$
\begin{aligned}
f_X(x_1)\Big(\prod_{i=3}^{k} f_{X|p}(x_i|p_i)f_p(p_i)\Big)\int_0^1 \Big\{\psi(x_1)\Big(\omega_{12}\phi(p_2)+\sum_{i=3}^{k}\omega_{1i}\phi(p_i) \\
+\phi(p_2)\sum_{i=3}^{k}\omega_{12i}\phi(p_i)+\sum_{i_1=3}^{k-1}\sum_{i_2=i_1+1}^{k}\omega_{1,i_1,i_2}\phi(p_{i_1})\phi(p_{i_2}) \\
+\cdots+\phi(p_2)\omega_{1,\ldots,k}\prod_{i=3}^{k}\phi(p_i)\Big) \\
+\Big(\phi(p_2)\sum_{i=3}^{k}\omega_{2i}\phi(p_i)+\sum_{i_1=3}^{k-1}\sum_{i_2=i_1+1}^{k}\omega_{i_1,i_2}\phi(p_{i_1})\phi(p_{i_2}) \\
+\phi(p_2)\sum_{i_1=3}^{k-1}\sum_{i_2=i_1+1}^{k}\omega_{2,i_1,i_2}\phi(p_{i_1})\phi(p_{i_2}) \\
+\sum_{i_1=3}^{k-2}\sum_{i_2=i_1+1}^{k-1}\sum_{i_3=i_2+1}^{k}\omega_{i_1,i_2,i_3}\phi(p_{i_1})\phi(p_{i_2})\phi(p_{i_3}) \\
+\cdots+\phi(p_2)\omega_{2,\ldots,k}\prod_{i=3}^{k}\phi(p_i)\Big)\Big\}f_{X|p}(x_2|p_2)f_p(p_2)\,dp_2 \\
= f_X(x_1)f_X(x_2)\Big(\prod_{i=3}^{k} f_{X|p}(x_i|p_i)f_p(p_i)\Big)\Big\{\omega_{12}\psi(x_1)\psi(x_2)+\psi(x_1)\sum_{i=3}^{k}\omega_{1i}\phi(p_i) \\
+\psi(x_1)\psi(x_2)\sum_{i=3}^{k}\omega_{12i}\phi(p_i)+\psi(x_1)\sum_{i_1=3}^{k-1}\sum_{i_2=i_1+1}^{k}\omega_{1,i_1,i_2}\phi(p_{i_1})\phi(p_{i_2}) \\
+\cdots+\psi(x_1)\psi(x_2)\omega_{1,\ldots,k}\prod_{i=3}^{k}\phi(p_i)+\psi(x_2)\sum_{i=3}^{k}\omega_{2i}\phi(p_i) \\
+\sum_{i_1=3}^{k-1}\sum_{i_2=i_1+1}^{k}\omega_{i_1,i_2}\phi(p_{i_1})\phi(p_{i_2})+\psi(x_2)\sum_{i_1=3}^{k-1}\sum_{i_2=i_1+1}^{k}\omega_{2,i_1,i_2}\phi(p_{i_1})\phi(p_{i_2}) \\
+\sum_{i_1=3}^{k-2}\sum_{i_2=i_1+1}^{k-1}\sum_{i_3=i_2+1}^{k}\omega_{i_1,i_2,i_3}\phi(p_{i_1})\phi(p_{i_2})\phi(p_{i_3})+\cdots+\psi(x_2)\omega_{2,\ldots,k}\prod_{i=3}^{k}\phi(p_i)\Big\}.
\end{aligned}
$$

Rearranging the variables the expression becomes,

$$
\begin{aligned}
&f_X(x_1)f_X(x_2)\Big(\prod_{i=3}^{k} f_{X|p}(x_i|p_i)f_p(p_i)\Big)\Big\{\omega_{12}\psi(x_1)\psi(x_2)+\psi(x_1)\sum_{i=3}^{k}\omega_{1i}\phi(p_i)\\
&+\psi(x_2)\sum_{i=3}^{k}\omega_{2i}\phi(p_i)+\sum_{i_1=3}^{k-1}\sum_{i_2=i_1+1}^{k}\omega_{i_1,i_2}\phi(p_{i_1})\phi(p_{i_2})+\psi(x_1)\psi(x_2)\sum_{i=3}^{k}\omega_{12i}\phi(p_i)\\
&+\psi(x_1)\sum_{i_1=3}^{k-1}\sum_{i_2=i_1+1}^{k}\omega_{1,i_1,i_2}\phi(p_{i_1})\phi(p_{i_2})+\psi(x_2)\sum_{i_1=3}^{k-1}\sum_{i_2=i_1+1}^{k}\omega_{2,i_1,i_2}\phi(p_{i_1})\phi(p_{i_2})\\
&+\sum_{i_1=3}^{k-2}\sum_{i_2=i_1+1}^{k-1}\sum_{i_3=i_2+1}^{k}\omega_{i_1,i_2,i_3}\phi(p_{i_1})\phi(p_{i_2})\phi(p_{i_3})\\
&+\cdots+\psi(x_2)\omega_{2,\ldots,k}\prod_{i=3}^{k}\phi(p_i)+\psi(x_1)\psi(x_2)\omega_{1,\ldots,k}\prod_{i=3}^{k}\phi(p_i)\Big\}
\end{aligned}
\tag{A.5}
$$

Following on, integrate (A.5) with respect to p_3,

$$
\begin{aligned}
&f_X(x_1)f_X(x_2)f_X(x_3)\Big(\prod_{i=4}^{k} f_{X|p}(x_i|p_i)f_p(p_i)\Big)\\
&\times\Big\{\omega_{12}\psi(x_1)\psi(x_2)+\omega_{13}\psi(x_1)\psi(x_3)+\psi(x_1)\sum_{i=4}^{k}\omega_{1i}\phi(p_i)+\omega_{23}\psi(x_2)\psi(x_3)\\
&+\psi(x_2)\sum_{i=4}^{k}\omega_{2i}\phi(p_i)+\psi(x_3)\sum_{i=4}^{k}\omega_{3i}\phi(p_i)+\sum_{i_1=4}^{k-1}\sum_{i_2=i_1+1}^{k}\omega_{i_1,i_2}\phi(p_{i_1})\phi(p_{i_2})\\
&+\omega_{123}\psi(x_1)\psi(x_2)\psi(x_3)+\psi(x_1)\psi(x_2)\sum_{i=4}^{k}\omega_{12i}\phi(p_i)\\
&+\psi(x_1)\psi(x_3)\sum_{i=4}^{k}\omega_{13i}\phi(p_i)+\psi(x_2)\psi(x_3)\sum_{i=4}^{k}\omega_{23i}\phi(p_i)\\
&+\sum_{i_1=4}^{k-2}\sum_{i_2=i_1+1}^{k-1}\sum_{i_3=i_2+1}^{k}\omega_{i_1,i_2,i_3}\phi(p_{i_1})\phi(p_{i_2})\phi(p_{i_3})\\
&+\cdots+\psi(x_1)\psi(x_2)\psi(x_3)\omega_{1,\ldots,k}\prod_{i=4}^{k}\phi(p_i)\Big\}
\end{aligned}
\tag{A.6}
$$

Following on the same manner of integration, the second term of (A.1) is integrated to

$$\begin{aligned}
&\int \cdots \int \Big(\prod_{i=1}^{k} f_{X|p}(x_i|p_i) f_p(p_i) \Big) R_{\Omega_k}(p_1, \ldots, p_k)\, dp_1 \ldots\, dp_k \\
&= \Big(\prod_{i=1}^{k} f_X(x_i) \Big) \Big\{ \sum_{i_1=1}^{k-1} \sum_{i_2=i_1+1}^{k} \omega_{i_1,i_2} \psi(x_{i_1}) \psi(x_{i_2}) \\
&\quad + \sum_{i_1=1}^{k-2} \sum_{i_2=i_1+1}^{k-1} \sum_{i_3=i_2+1}^{k} \omega_{i_1,i_2,i_3} \psi(x_{i_1}) \psi(x_{i_2}) \psi(x_{i_3}) \\
&\quad + \cdots + \omega_{1,\ldots,k} \prod_{i=1}^{k} \psi(x_i) \Big\}
\end{aligned} \tag{A.7}$$

For ease of notation, let $D_{\Omega_k}(x_1, \ldots, x_k)$ represents the second term of (A.7). Finally, with (A.2) and (A.7), the joint unconditional density of $X_1, \ldots, X_k$, is

$$\begin{aligned}
&h_{\boldsymbol{X}}(x_1, \ldots, x_k) \\
&= \int \cdots \int h_{\boldsymbol{X}|\boldsymbol{p}}(X_1 = x_1, \ldots, X_k = x_k | p_1, \ldots, p_k) h_{\boldsymbol{p}}(p_1, \ldots, p_k)\, dp_1 \ldots\, dp_k \\
&= \prod_{i=1}^{k} f_X(x_i) + \Big(\prod_{i=1}^{k} f_X(x_i) \Big) D_{\Omega_k}(x_1, \ldots, x_k) \\
&= \Big(\prod_{i=1}^{k} f_X(x_i) \Big) \Big(1 + D_{\Omega_k}(x_1, \ldots, x_k) \Big)
\end{aligned} \tag{A.8}$$

Appendix B

Derivation of the Expected Number of Hybrid Trials that Reject H_0

The expected number of trials that reject H_0 is,

$$E(\widetilde{K}) = \frac{N}{n}\Bigg(1 - \Phi\bigg(\frac{z_{1-\alpha/2} - \sqrt{n/\sigma^2}(\mu - \theta_0)}{\sqrt{1 + n\tau^2/\sigma^2}}\bigg)\Bigg).$$

Let $f(n) = \frac{z_{1-\alpha/2} - \sqrt{n/\sigma^2}(\mu - \theta_0)}{\sqrt{1+n\tau^2/\sigma^2}}$, and since the derivative of the cumulative distribution function (cdf) is the density function, $\Phi'(x) = \phi(x)$, the derivative of $E(\widetilde{K})$ is,

$$\frac{d}{dn}E(\widetilde{K}) = -\frac{N}{n^2}\Big(1 - \Phi\Big(f(n)\Big)\Big) - \frac{N}{n}f'(n)\phi\Big(f(n)\Big). \tag{B.1}$$

The cdf in the first term of (B.1) is bounded, $\Phi(f(n)) \in (0,1)$. Therefore,

$$\frac{N}{n^2}\Big(1-\Phi(f(n))\Big) > 0, \quad \text{for all } n > 0. \tag{B.2}$$

The derivative of $f'(n)$ in the second term of (B.1) is,

$$\begin{aligned} f'(n) &= \frac{-\dfrac{\sqrt{1+n\tau^2/\sigma^2}(\mu-\theta_0)}{2\sqrt{n\sigma^2}} - \dfrac{\tau^2(z_{1-\alpha/2}-\sqrt{n/\sigma^2}(\mu-\theta_0))}{2\sigma^2\sqrt{1+n\tau^2/\sigma^2}}}{1+n\tau^2/\sigma^2} \\ &= \frac{-\dfrac{(1+n\tau^2/\sigma^2)(\mu-\theta_0)}{2\sqrt{n\sigma^2}} - \dfrac{\tau^2(z_{1-\alpha/2}-\sqrt{n/\sigma^2}(\mu-\theta_0))}{2\sigma^2}}{(1+n\tau^2/\sigma^2)^{3/2}} \\ &= -\frac{z_{1-\alpha/2}(\tau^2/\sigma^2)+(\mu-\theta_0)/\sqrt{n\sigma^2}}{2(1+n\tau^2/\sigma^2)^{3/2}}. \end{aligned}$$

Appendix C

Derivation of the Expected Utility of Action P of the Third Trial for the Design of a Series of Related Treatments

The log odds ratio is defined as $\theta_3 = \log\{p_3(1-p_C)/(p_C(1-p_3))\}$. Let $g(s_3, n_3, N_3, K_3)$ be the integral given in (7.8), then

$$g(s_3, n_3, N_3, K_3) = \iiint \left(1 - \Phi\Big(z_{1-\alpha/2} - \theta_3\sqrt{V_3}\Big)\right) \times h_{\boldsymbol{p}|\boldsymbol{S}}(p_1, p_2, p_3|s_1, s_2, s_3, n_1, n_2, n_3)\, dp_1\, dp_2\, dp_3. \quad \text{(C.1)}$$

From equation (3.15), the integral is now

$$
\begin{aligned}
&g(s_3, n_3, N_3, K_3) \\
&= \frac{1}{1 + D_{\Omega_3}(s_1, s_2, s_3)} \iiint \left(1 - \Phi\Big(z_{1-\alpha/2} - \theta_3\sqrt{V_3}\Big)\right)\Big(\prod_{i=1}^{3} f_{p|S}(p_i|s_i, n_i)\Big) \\
&\quad \times \Big(1 + \omega_{12}\phi(p_1)\phi(p_2) + \omega_{13}\phi(p_1)\phi(p_3) + \omega_{23}\phi(p_2)\phi(p_3) \\
&\qquad + \omega_{123}\phi(p_1)\phi(p_2)\phi(p_3)\Big)\, dp_1\, dp_2\, dp_3. \qquad \text{(C.2)}
\end{aligned}
$$

First consider a general integrand that will appear iteratively,

$$
\begin{aligned}
&\int_0^1 \phi(p_i) f_{p|S}(p_i|s_i, n_i)\, dp_i \\
&\quad = \frac{1}{B(a_i + s_i, b_i + n_i - s_i)} \int_0^1 (p_i - \mu_i) p_i^{a_i + s_i - 1}(1 - p_i)^{b_i + n_i - s_i - 1}\, dp_i \\
&\quad = \frac{B(a_i + s_i + 1, b_i + n_i - s_i)}{B(a_i + s_i, b_i + n_i - s_i)} - \mu_i \\
&\quad = \frac{a_i + s_i}{a_i + b_i + n_i} - \mu_i \\
&\quad = \frac{s_i - \mu_i n_i}{a_i + b_i + n_i}. \qquad \text{(C.3)}
\end{aligned}
$$

First, integrate the inner expression in (C.2) with respect to p_1 over all its possible values,

$$
\begin{aligned}
&\int_0^1 f_{p|S}(p_1|s_1, n_1)\Big(1 + \omega_{12}\phi(p_1)\phi(p_2) + \omega_{13}\phi(p_1)\phi(p_3) \\
&\quad + \omega_{23}\phi(p_2)\phi(p_3) + \omega_{123}\phi(p_1)\phi(p_2)\phi(p_3)\Big)\, dp_1 \\
&= 1 + \omega_{23}\phi(p_2)\phi(p_3) + \Big(\omega_{12}\phi(p_2) + \omega_{13}\phi(p_3) + \omega_{123}\phi(p_2)\phi(p_3)\Big)\psi(s_1), \\
&\qquad \text{(C.4)}
\end{aligned}
$$

where as derived earlier in equation (C.3) $\psi(s_1) = (s_1 - \mu_1 n_1)/(a_1 + b_1 + n_1)$.

Substitute (C.4) back into (C.2) and integrate it with respect to p_2 over all its possible values,

$$\begin{aligned}
&\int_0^1 f_{p|S}(p_2|s_2, n_2)\Big(1 + \omega_{23}\phi(p_2)\phi(p_3) \\
&\quad + \Big(\omega_{12}\phi(p_2) + \omega_{13}\phi(p_3) + \omega_{123}\phi(p_2)\phi(p_3)\Big)\psi(s_1)\Big)\, dp_2 \\
&= 1 + \omega_{13}\phi(p_3)\psi(s_1) + \Big(\omega_{23}\phi(p_3) + \omega_{12}\psi(s_1) + \omega_{123}\phi(p_3)\psi(s_1)\Big)\psi(s_2).
\end{aligned} \tag{C.5}$$

Finally, substitute (C.5) back into (C.2) and integrate the whole expression with respect to p_3 over all its possible values,

$$\begin{aligned}
&g(s_3, n_3, N_3, K_3) \\
&= \frac{1}{1 + D_{\Omega_3}(s_1, s_2, s_3)} \int_0^1 \Big(1 - \Phi\Big(z_{1-\alpha/2} - \theta_3\sqrt{V_3}\Big)\Big) f_{p|S}(p_3|s_3, n_3) \\
&\times \Big(1 + \omega_{13}\phi(p_3)\psi(s_1) + \Big(\omega_{23}\phi(p_3) + \omega_{12}\psi(s_1) + \omega_{123}\phi(p_3)\psi(s_1)\Big)\psi(s_2)\Big)\, dp_3 \\
&= \frac{1}{1 + D_{\Omega_3}(s_1, s_2, s_3)}\Bigg[\Big(1 + \omega_{12}\psi(s_1)\psi(s_2)\Big) \\
&\quad \times \int_0^1 \Big(1 - \Phi\Big(z_{1-\alpha/2} - \theta_3\sqrt{V_3}\Big)\Big) f_{p|S}(p_3|s_3, n_3)\, dp_3 \\
&+ \Big(\omega_{13}\psi(s_1) + \omega_{23}\psi(s_2) + \omega_{123}\psi(s_1)\psi(s_2)\Big) \\
&\quad \times \int_0^1 \phi(p_3)\Big(1 - \Phi\Big(z_{1-\alpha/2} - \theta_3\sqrt{V_3}\Big)\Big) f_{p|S}(p_3|s_3, n_3)\, dp_3\Bigg].
\end{aligned} \tag{C.6}$$

Note that the expression $\int_0^1 \Big(1 - \Phi\big(z_{1-\alpha/2} - \theta_3\sqrt{V_3}\big)\Big) f_{p|S}(p_3|s_3, n_3)\, dp_3$ is the assurance (as defined in (3.23)) given observed responses, s_3, out of n_3

patients. Let the assurance be denoted by $A(s_3, n_3)$ thus,

$$
\begin{aligned}
&g(s_3, n_3, N_3, K_3) \\
&\quad = \frac{1}{1 + D_{\Omega_3}(s_1, s_2, s_3)} \Bigg[\Big(1 + \omega_{12}\psi(s_1)\psi(s_2)\Big) A(s_3, n_3) \\
&\qquad + \Big(\omega_{13}\psi(s_1) + \omega_{23}\psi(s_2) + \omega_{123}\psi(s_1)\psi(s_2)\Big) \\
&\qquad\quad \times \int_0^1 \phi(p_3) \Big(1 - \Phi\Big(z_{1-\alpha/2} - \theta_3\sqrt{V_3}\Big)\Big) f_{p|S}(p_3|s_3, n_3)\, dp_3 \Bigg].
\end{aligned}
$$

Appendix D

Derivation of the Expected Utility of Action R of the Third Trial for the Design of a Series of Related Treatments

The marginal density of X_3 given $S_1 = s_1, S_2 = s_2$ and $S_3 = s_3$ is defined as

$$g_{X|\boldsymbol{S}}(x_3, m_{3,i+1}|s_1, s_2, s_3, n_1, n_2, n_3)$$
$$= \iiint f_{S|p}(x_3|p_3) h_{\boldsymbol{p}|\boldsymbol{S}}(p_1, p_2, p_3|s_1, s_2, s_3, n_1, n_2, n_3)\, dp_1\, dp_2\, dp_3. \quad \text{(D.1)}$$

From equation (3.15), the integral in (D.1) becomes

$$
\begin{aligned}
&g_{X|\boldsymbol{S}}(x_3, m_{3,i+1}|s_1, s_2, s_3, n_1, n_2, n_3) \\
&\quad = \frac{1}{1 + D_{\Omega_3}(s_1, s_2, s_3)} \iiint f_{S|p}(x_3|p_3)\Big(\prod_{i=1}^{3} f_{p|S}(p_i|s_i, n_i)\Big) \\
&\qquad \times \Big(1 + \omega_{12}\phi(p_1)\phi(p_2) + \omega_{13}\phi(p_1)\phi(p_3) + \omega_{23}\phi(p_2)\phi(p_3) \\
&\qquad\quad + \omega_{123}\phi(p_1)\phi(p_2)\phi(p_3)\Big)\, dp_1\, dp_2\, dp_3. \qquad \text{(D.2)}
\end{aligned}
$$

First, integrate the inner expression in (D.2) with respect to p_1,

$$
\begin{aligned}
&\Big(1 + \omega_{23}\phi(p_2)\phi(p_3)\Big) \int_0^1 f_{p|S}(p_1|s_1, n_1)\, dp_1 \\
&\qquad + \Big(\omega_{12}\phi(p_2) + \omega_{13}\phi(p_3) + \omega_{123}\phi(p_2)\phi(p_3)\Big) \int_0^1 \phi(p_1) f_{p|S}(p_1|s_1, n_1)\, dp_1 \\
&\quad = 1 + \omega_{23}\phi(p_2)\phi(p_3) + \Big(\omega_{12}\phi(p_2) + \omega_{13}\phi(p_3) + \omega_{123}\phi(p_2)\phi(p_3)\Big)\psi(s_1), \qquad \text{(D.3)}
\end{aligned}
$$

where from equation (C.3), $\psi(s_1) = (s_1 - \mu_1 n_1)/(a_1 + b_1 + n_1)$.

Substitute (D.3) back into (D.2) and integrate it with respect to p_2,

$$\begin{aligned}
&\int_0^1 \Big(1 + \omega_{23}\phi(p_2)\phi(p_3) + \Big(\omega_{12}\phi(p_2) + \omega_{13}\phi(p_3) + \omega_{123}\phi(p_2)\phi(p_3)\Big)\psi(s_1)\Big) \\
&\quad \times f_{p|S}(p_2|s_2, n_2)\, dp_2 \\
&= \Big(1 + \omega_{13}\psi(s_1)\phi(p_3)\Big) \int_0^1 f_{p|S}(p_2|s_2, n_2)\, dp_2 \\
&\quad + \Big(\omega_{23}\phi(p_3) + \omega_{12}\psi(s_1) + \omega_{123}\psi(s_1)\phi(p_3)\Big) \int_0^1 \phi(p_2) f_{p|S}(p_2|s_2, n_2)\, dp_2 \\
&= 1 + \omega_{13}\psi(s_1)\phi(p_3) + \Big(\omega_{23}\phi(p_3) + \omega_{12}\psi(s_1) + \omega_{123}\psi(s_1)\phi(p_3)\Big)\psi(s_2).
\end{aligned} \tag{D.4}$$

Then, substitute (D.4) back into (D.2) and integrate it with respect to

p_3,

$$
\begin{aligned}
&g_{X|\boldsymbol{S}}(x_3, m_{3,i+1}|s_1, s_2, s_3, n_1, n_2, n_3) \\
&= \int_0^1 \Big[1 + \omega_{13}\psi(s_1)\phi(p_3) + \Big(\omega_{23}\phi(p_3) + \omega_{12}\psi(s_1) + \omega_{123}\psi(s_1)\phi(p_3)\Big)\psi(s_2)\Big] \\
&\quad \times f_{S|p}(x_3|p_3) f_{p|S}(p_3|s_3, n_3)\, dp_3 \\
&= \binom{m_{3,i+1}}{x_3} \frac{1}{B(a_3 + s_3, b_3 + n_3 - s_3)} \\
&\quad \times \Bigg[\Big(1 + \omega_{12}\psi(s_1)\psi(s_2)\Big) \int_0^1 p_3^{a_3+s_3+x_3-1}(1-p_3)^{b_3+n_3+m_{3,i+1}-(s_3+x_3)-1}\, dp_3 \\
&\qquad + \Big(\omega_{13}\psi(s_1) + \omega_{23}\psi(s_2) + \omega_{123}\psi(s_1)\psi(s_2)\Big) \\
&\qquad\quad \times \int_0^1 \phi(p_3) p_3^{a_3+s_3+x_3-1}(1-p_3)^{b_3+n_3+m_{3,i+1}-(s_3+x_3)-1}\, dp_3\Bigg] \\
&= \binom{m_{3,i+1}}{x_3} \frac{1}{B(a_3 + s_3, b_3 + n_3 - s_3)} \\
&\quad \times \Bigg[\Big(1 + \omega_{12}\psi(s_1)\psi(s_2)\Big) B\big(a_3 + s_3 + x_3, b_3 + n_3 + m_{3,i+1} - (s_3 + x_3)\big) \\
&\qquad + \Big(\omega_{13}\psi(s_1) + \omega_{23}\psi(s_2) + \omega_{123}\psi(s_1)\psi(s_2)\Big)\Big(\frac{s_3 + x_3 - \mu_3(n_3 + m_{3,i+1})}{a_3 + b_3 + n_3 + m_{3,i+1}}\Big) \\
&\qquad\quad \times B\big(a_3 + s_3 + x_3, b_3 + n_3 + m_{3,i+1} - (s_3 + x_3)\big)\Bigg]. \\
&= \Bigg[1 + \omega_{12}\psi(s_1)\psi(s_2) + \Big(\omega_{13}\psi(s_1) + \omega_{23}\psi(s_2) + \omega_{123}\psi(s_1)\psi(s_2)\Big) \\
&\quad \times \Big(\frac{s_3 + x_3 - \mu_3(n_3 + m_{3,i+1})}{a_3 + b_3 + n_3 + m_{3,i+1}}\Big)\Bigg] f_S(x_3|a_3 + s_3, b_3 + n_3 - s_3), \qquad \text{(D.5)}
\end{aligned}
$$

where $B(\alpha, \beta) = \Gamma(\alpha + \beta)/(\Gamma(\alpha)\Gamma(\beta))$ is the beta function and $f_S(x_3|a_3 + s_3, b_3 + n_3 - s_3) = \binom{m_{3,i+1}}{x_3} B(a_3 + s_3 + x_3, b_3 + n_3 + m_{3,i+1} - (s_3 + x_3))/B(a_3 + s_3, b_3 + n_3 - s_3)$ is the beta-binomial distribution of X_3 with index $m_{3,i+1}$ and parameters $a_3 + s_3$ and $b_3 + n_3 - s_3$.

Substituting (D.5) back into (D.2), the "posterior" marginal density of

X_3 given $S_1 = s_1, S_2 = s_2$ and $S_3 = s_3$ is

$$
\begin{aligned}
&g_{X|\boldsymbol{S}}(x_3, m_{3,i+1}|s_1, s_2, s_3, n_1, n_2, n_3) \\
&= \frac{1}{1 + D_{\Omega_3}(s_1, s_2, s_3)} \Bigg[1 + \omega_{12}\psi(s_1)\psi(s_2) \\
&\quad + \Big(\omega_{13}\psi(s_1) + \omega_{23}\psi(s_2) + \omega_{123}\psi(s_1)\psi(s_2) \Big) \Big(\frac{s_3 + x_3 - \mu_3(n_3 + m_{3,i+1})}{a_3 + b_3 + n_3 + m_{3,i+1}} \Big) \Bigg] \\
&\qquad \times f_S(x_3|a_3 + s_3, b_3 + n_3 - s_3).
\end{aligned}
$$

Appendix E

Derivation of the Expected Utility of Action P of the Second Trial for the Design of a Series of Related Treatments

The log odds ratio is defined as $\theta_2 = \log\{p_2(1-p_C)/(p_C(1-p_2))\}$. Let

$$g(s_2, n_2, N_2, K_2) = \iiint \left(1 - \Phi\Big(z_{1-\alpha/2} - \theta_2\sqrt{V_2}\Big)\right) \times h_{\boldsymbol{p}|\boldsymbol{S}}(p_1, p_2, p_3|s_1, s_2, s_3, n_1, n_2, n_3)\, dp_1\, dp_2\, dp_3, \quad \text{(E.1)}$$

be the integral given in (7.13). As $S_3 = 0$ and $n_3 = 0$, then $D_{\Omega_3}(s_1, s_2, s_3) = \omega_{12}\psi(s_1)\psi(s_2)$. From (3.15), the above equation (E.1) is simplified to

$$\begin{aligned} & g(s_2, n_2, N_2, K_2) \\ & \quad = \frac{1}{1+\omega_{12}\psi(s_1)\psi(s_2)} \iiint \left(1 - \Phi\Big(z_{1-\alpha/2} - \theta_2\sqrt{V_2}\Big)\right) \\ & \qquad \times f_{p|S}(p_1|s_1, n_1) f_{p|S}(p_2|s_2, n_2) f_p(p_3) \\ & \qquad \times \Big(1 + \omega_{12}\phi(p_1)\phi(p_2) + \omega_{13}\phi(p_1)\phi(p_3) + \omega_{23}\phi(p_2)\phi(p_3) \\ & \qquad\quad + \omega_{123}\phi(p_1)\phi(p_2)\phi(p_3)\Big)\, dp_1\, dp_2\, dp_3. \end{aligned} \tag{E.2}$$

First, integrate the inner expression (E.2) with respect to p_3,

$$\begin{aligned} & \Big(1 + \omega_{12}\phi(p_1)\phi(p_2)\Big) \int_0^1 f_p(p_3)\, dp_3 + \\ & \qquad \Big(\omega_{13}\phi(p_1) + \omega_{23}\phi(p_2) + \omega_{123}\phi(p_1)\phi(p_2)\Big) \int_0^1 \phi(p_3) f_p(p_3)\, dp_3, \end{aligned} \tag{E.3}$$

where the mixing function is defined as $\phi(p_3) = p_3 - \mu_3$. Therefore, the second integrand of (E.3) integrates to 0 whilst the first integrand integrates to 1 leaving the expression in (E.3) simply as $\big(1+\omega_{12}\phi(p_1)\phi(p_2)\big)$. Substitute this expression back into (E.2) and integrates it with respect to p_1,

$$\begin{aligned} & \int_0^1 \big(1 + \omega_{12}\phi(p_1)\phi(p_2)\big) f_{p|S}(p_1|s_1, n_1)\, dp_1 \\ & \qquad = 1 + \omega_{12}\phi(p_2) \int_0^1 \phi(p_1) f_{p|S}(p_1|s_1, n_1)\, dp_1 \\ & \qquad = 1 + \omega_{12}\phi(p_2)\psi(s_1) \end{aligned} \tag{E.4}$$

where $\psi(s_1) = (s_1 - \mu_1 n_1)/(a_1 + b_1 + n_1)$ as given in equation (C.3).

Finally, substitute (E.4) back into (E.2) and integrate it with respect to p_2,

$$
\begin{aligned}
&g(s_2, n_2, N_2, K_2) \\
&\quad = \frac{1}{1+\omega_{12}\psi(s_1)\psi(s_2)} \int_0^1 \left(1 - \Phi\Big(z_{1-\alpha/2} - \theta_2\sqrt{V_2}\Big)\right) f_{p|S}(p_2|s_2, n_2) \\
&\qquad \times \big(1 + \omega_{12}\psi(s_1)\phi(p_2)\big)\, dp_2 \\
&\quad = \frac{1}{1+\omega_{12}\psi(s_1)\psi(s_2)} \Bigg[A(s_2, n_2) \\
&\qquad + \omega_{12}\psi(s_1) \int_0^1 \phi(p_2) \left(1 - \Phi\Big(z_{1-\alpha/2} - \theta_2\sqrt{V_2}\Big)\right) f_{p|S}(p_2|s_2, n_2)\, dp_2, \Bigg]
\end{aligned}
$$

where $A(s_2, n_2) = \int_0^1 \left(1 - \Phi\big(z_{1-\alpha/2} - \theta_2\sqrt{V_2}\big)\right) f_{p|S}(p_2|s_2, n_2)\, dp_2$ is the assurance (as defined in (3.23)) given observed responses, s_2, out of n_2 patients.

Appendix F

Derivation of the Expected Utility of Action R of the Second Trial for the Design of a Series of Related Treatments

The marginal density of X_2 given $S_1 = s_1, S_2 = s_2$ and $S_3 = 0$ is defined as

$$g_{X|\boldsymbol{S}}(x_2, m_{2,i+1}|s_1, s_2, 0, n_1, n_2, 0) = \iiint f_{S|p}(x_2|p_2) h_{\boldsymbol{p}|\boldsymbol{S}}(p_1, p_2, p_3|s_1, s_2, 0, n_1, n_2, 0)\, dp_1\, dp_2\, dp_3. \quad \text{(F.1)}$$

Given that $S_3 = 0$ and $n_3 = 0$ then $D_{\Omega_3}(s_1, s_2, s_3) = \omega_{12}\psi(s_1)\psi(s_2)$. From

equation (3.15), the integral in (F.1) becomes

$$
\begin{aligned}
&g_{X|\boldsymbol{S}}(x_2, m_{2,i+1}|s_1, s_2, 0, n_1, n_2, 0) \\
&= \frac{1}{1+\omega_{12}\psi(s_1)\psi(s_2)} \iiint f_{S|p}(x_2|p_2) f_{p|S}(p_1|s_1, n_1) f_{p|S}(p_2|s_2, n_2) f_p(p_3) \\
&\quad \times \Big(1 + \omega_{12}\phi(p_1)\phi(p_2) + \omega_{13}\phi(p_1)\phi(p_3) + \omega_{23}\phi(p_2)\phi(p_3) \\
&\qquad + \omega_{123}\phi(p_1)\phi(p_2)\phi(p_3)\Big)\, dp_1\, dp_2\, dp_3. \qquad \text{(F.2)}
\end{aligned}
$$

First, integrate (F.2) with respect to p_3,

$$
\begin{aligned}
&\int_0^1 \Big(1 + \omega_{12}\phi(p_1)\phi(p_2) + \omega_{13}\phi(p_1)\phi(p_3) + \omega_{23}\phi(p_2)\phi(p_3) \\
&\qquad + \omega_{123}\phi(p_1)\phi(p_2)\phi(p_3)\Big) f_p(p_3)\, dp_3 \\
&= \big(1 + \omega_{12}\phi(p_1)\phi(p_2)\big) \int_0^1 f_p(p_3)\, dp_3 \\
&\qquad + \big(\omega_{13}\phi(p_1) + \omega_{23}\phi(p_2) + \omega_{123}\phi(p_1)\phi(p_2)\big) \int_0^1 \phi(p_3) f_p(p_3)\, dp_3. \qquad \text{(F.3)}
\end{aligned}
$$

Note that the mixing function is defined as $\phi(p_3) = p_3 - \mu_3$ and so the second integrand of (F.3) integrates to 0 whilst the first integrand integrates to 1, leaving (F.3) simply as $1 + \omega_{12}\phi(p_1)\phi(p_2)$. Substitute this expression into (F.2) and integrate it with respect to p_1,

$$
\begin{aligned}
&\int_0^1 \big(1 + \omega_{12}\phi(p_1)\phi(p_2)\big) f_{p|S}(p_1|s_1, n_1) \\
&\qquad = 1 + \omega_{12}\psi(s_1)\phi(p_2). \qquad \text{(F.4)}
\end{aligned}
$$

Finally, substitute (F.4) back into (F.2) and integrate it with respect to

p_2,

$$
\begin{aligned}
&g_{X|\boldsymbol{S}}(x_2, m_{2,i+1}|s_1, s_2, 0, n_1, n_2, 0) \\
&= \frac{1}{\big(1+\omega_{12}\psi(s_1)\psi(s_2)\big)B(a_2+s_2, b_2+n_2-s_2)}\binom{m_{2,i+1}}{x_2} \\
&\quad \times \int_0^1 \big(1+\omega_{12}\psi(s_1)\phi(p_2)\big)p_2^{a_2+s_2+x_2-1}(1-p_2)^{b_2+n_2+m_{2,i+1}-(s_2+x_2)-1}\,dp_2 \\
&= \frac{1}{1+\omega_{12}\psi(s_1)\psi(s_2)}\big(1+\omega_{12}\psi(s_1)\big)\Big(\frac{s_2+x_2-\mu_2(n_2+m_{2,i+1})}{a_2+b_2+n_2+m_{2,i+1}}\Big) \\
&\qquad \times f_S(x_2|a_2+s_2, b_2+n_2-s_2). \qquad \text{(F.5)}
\end{aligned}
$$

Appendix G

Simulation for a Series of Related Treatments

The algorithm of the simulation study is

1. Generate p_1 from $Beta(a_1, b_1)$.
2. Generate s_1 from $Bin(4, p_1)$.
3. Generate p_2 from the probability density function, $f(p_2|p_1)$.
4. Generate s_2 from $Bin(5, p_2)$.

As the probability density function $f(p_2|p_1)$ is not of a standard parametric form, the sampling of p_2 from the function $f(p_2|p_1)$ is based on rejection sampling (Ripley, 1987). The joint distribution of p_1, p_2, p_3 is denoted by $h(p_1, p_2, p_3) = \prod_{i=1}^{3} f_p(p_i)(1 + R_{\Omega_3}(p_1, p_2, p_3))$ where $f_p(p_i) = p_i^{a_i - 1}(1 -$

$p_i)^{b_i-1}/B(a_i, b_i)$ is the marginal probability density function,

$$R_{\Omega_3}(p_1, p_2, p_3) = \omega_{1,2}\phi(p_1)\phi(p_2) + \omega_{1,3}\phi(p_1)\phi(p_3) + \omega_{2,3}\phi(p_2)\phi(p_3) + \omega_{1,2,3}\phi(p_1)\phi(p_2)\phi(p_3),$$

and $\phi(p_i) = p_i - \mu_i$ and $\mu_i = a_i/(a_i + b_i)$ for $i = 1, 2, 3$.

Let $f(p_2|p_1)$ denote the conditional distribution of p_2 given p_1 and from Bayes' theorem it is

$$f(p_2|p_1) = \frac{\int_0^1 h(p_1, p_2, p_3)\, dp_3}{f_p(p_1)}. \tag{G.1}$$

Consider only the numerator of equation (G.1),

$$\begin{aligned}
\int_0^1 h(p_1, p_2, p_3)\, dp_3 &= f_p(p_1)f_p(p_2)\int_0^1 f_p(p_3)\Big(1 + \omega_{1,2}\phi(p_1)\phi(p_2) \\
&\quad + \omega_{1,3}\phi(p_1)\phi(p_3) + \omega_{2,3}\phi(p_2)\phi(p_3) \\
&\quad + \omega_{1,2,3}\phi(p_1)\phi(p_2)\phi(p_3)\Big)\, dp_3 \\
&= f_p(p_1)f_p(p_2)\Bigg[\Big(1 + \omega_{1,2}\phi(p_1)\phi(p_2)\Big) \\
&\quad + \Big(\omega_{1,3}\phi(p_1) + \omega_{2,3}\phi(p_2) + \omega_{1,2,3}\phi(p_1)\phi(p_2)\Big) \\
&\quad \times \int_0^1 \phi(p_3) f_p(p_3)\, dp_3\Bigg] \\
&= f_p(p_1)f_p(p_2)\Big(1 + \omega_{1,2}\phi(p_1)\phi(p_2)\Big).
\end{aligned} \tag{G.2}$$

Substitute (G.2) back into (G.1),

$$f(p_2|p_1) = \Big(1 + \omega_{1,2}\phi(p_1)\phi(p_2)\Big) f_p(p_2).$$

Let $g(p_2) = \Big(1 + \omega_{1,2}(p_1 - \mu_1)(p_2 - \mu_2)\Big) p_2^{a_2-1}(1-p_2)^{b_2-1}$. As shown in equation (7.25), for $a_1 = a_2$ and $b_1 = b_2$, ω is bounded by $(a+b)^2/ab$. Thus, $g(p_2)$ is bounded by

$$c = \Big(1 + \frac{(a+b)^2}{ab}\Big) \frac{(a-1)^{a-1}(b-1)^{b-1}}{(a+b-2)^{a+b-2}}.$$

Therefore, the algorithm for the rejection sampling is:

1. Generate Y from $U(0,1)$.
2. Generate U from $U(0,1)$ until $U \leq g(p_2)/c$.
3. Return $p_2 = Y$.

The program in R[1] for the simulation is given below.

```
# Define the parameters
a = 69
b = 6

user.seed = 20120607
n.sim = 100000

set.seed(user.seed)

#------------------------------------------------------------
# First phase II trial
#------------------------------------------------------------

p.sim <- rbeta(n.sim,a,b)

# Generate a random sequence of "0" and "1" out of n=4
# observations given p.sim
s.sim <- rbinom(n=4*n.sim,size=1,rep(p.sim,rep(4,length(p.sim))))
```

[1] http://www.r-project.org/

```
# The first column is the list of randomly generated p.sim
# and the next 4 columns are the randomly generated observations
m.sim <- cbind(p.sim,matrix(s.sim,byrow=TRUE,ncol=4))

# If the sum of successes is NOT EQUAL to 4 then
# the optimal action is Action T
samp.size <- function(obs.sim){
  if(sum(obs.sim[2:5])==4){
    n.II = 4
    p.III = 1
  }
  else{
    n.II = which.min(obs.sim[2:5])
    p.III = 0
  }
  c(n.II,p.III)
}

sum.sim <- apply(m.sim,MARGIN=1, samp.size)

#-------------------------------------------------------------
# Second phase II trial
#-------------------------------------------------------------

#-------------------------------------------------------------
# 1. For every p in p.sim
# 2. Generate a variate Y from Unif(0,1)
# 3. Generate a variate U from Unif(0,1)
# 4. (a) If U <= f.p2(Y)/c then p2 = Y
# 4. (b) Else, go to Step 1

# Calculate the expected value
mu = a/(a+b)

# Negative correlation
# (change it to omega.12=10 for positive correlation)
omega.12 = -1

# Compute the constant
c = (1+((a+b)^2)/(a*b))*((a-1)^(a-1))*((b-1)^(b-1))/(a+b-2)^(a+b-2)

# In order to generate values from
# f(p2|p1) = (1 + omega.12*phi(p1)*phi(p2))*f(p2)
```

```
# where p1 ~ beta(a1,b1) and p2 ~ beta(a2,b2),
# and phi(p.i)= p.i - mu.i and mu.i = a.i/(a.i+b.i), for i=1,2,
# let f(x) = (1 + omega.12*phi(p1)*phi(p2))*p2^(a2-1)*(1-p2)^(b2-1)

phi <- function(x){
  x - mu
}

rand.p2condp1 <- function(X){
  p = X[1]

  d=1
  while(d>0){
    Y = runif(1,0,1)
    U = runif(1,0,1)

    # Compute f(Y) where Y is the random variate from Unif(0,1)
    # and f(x) is a function given above
    f.p2 = (1 + omega.12*phi(p)*phi(Y))*(Y^(a-1))*((1-Y)^(b-1))

    # If the random variate U is within the density of
    # f(p2|p1) then accept
    if(U <= f.p2/c){
      p2.rand = Y
      d = -1  #stop the while loop
    }
    else d = 1

  } #close "while" loop

  return(p2.rand)
}

# Generate a variate p.2 from f(p2|p1)
p2givenp1 = apply(m.sim,MARGIN=1,rand.p2condp1)

# Combine the whole matrix:
# 1) the randomly generated p1 from beta(a,b)
# 2) the randomly generated observations from bin(4,p1)
# 3) the randomly generated p2 from f(p2|p1)

combi.p1p2 = cbind(m.sim,p2givenp1)
```

```
# Generate a random sequence of "0" and "1" out of n=5
# observations given p2
s2.sim = rbinom(n=5*n.sim,size=1,rep(combi.p1p2[,6],
                 rep(5,length(combi.p1p2[,6]))))

fin.mat = cbind(combi.p1p2,matrix(s2.sim,byrow=TRUE,ncol=5))

# If the sum of successes is NOT EQUAL to 5 then
# the optimal action is Action T
samp.size.02 <- function(obs.sim){
  if(sum(obs.sim[7:11])==5){
    n.II = 5
    p.III = 1
  }
  else{
    n.II = which.min(obs.sim[7:11])
    p.III = 0
  }
  c(n.II,p.III)
}

sum.sim02 <- apply(fin.mat,MARGIN=1, samp.size.02)

#-------------------------------------------------------------
# Print the results
#-------------------------------------------------------------

# Probability of taking Action P in the first phase II trial
prob.III <- sum(sum.sim[2,])/dim(sum.sim)[2]

# Average sample size for the first phase II trial
mean.n <- mean(sum.sim[1,])

# Probability of taking Action P in the second phase II trial
prob.III.02 <- sum(sum.sim02[2,])/dim(sum.sim02)[2]

# Average sample size for the second phase II trial
mean.n02 <- mean(sum.sim02[1,])
```

Bibliography

A'Hern, R. P. (2001). Sample size tables for exact single-stage phase II designs. *Statistics in Medicine*, 20(6):859–866.

Barker, A. D., Sigman, C. C., Kelloff, G. J., Hylton, N. M., Berry, D. A., and Esserman, L. J. (2009). I-SPY 2: An adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clinical Pharmacology & Therapeutics*, 86(1):97–100.

Barnes, P. J. (2009). Corticosteroids. In Barnes, P. J., Drazen, J. M., Rennard, S. I., and Thomson, N. C., editors, *Asthma and COPD: Basic Mechanisms and Clinical Management*, chapter 51, pages 639–653. Academic Press, second edition.

Briggs, A., Claxton, K., and Sculpher, M. J. (2006). *Decision Modelling for Health Economic Evaluation*. Oxford University Press, Oxford.

Brunier, H. C. and Whitehead, J. (1994). Sample sizes for phase II clinical trials derived from bayesian decision theory. *Statistics in Medicine*, 13(23-24):2493–2502.

Busse, W., Nelson, H., Wolfe, J., Kalberg, C., Yancey, S. W., and Rickard, K. A. (1999). Comparison of inhaled salmeterol and oral zafirlukast in patients with asthma. *Journal of Allergy and Clinical Immunology*, 103(6):1075–1080.

Chaloner, K., Church, T., Louis, T. A., and Matts, J. P. (1993). Graphical elicitation of a prior distribution for a clinical trial. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 42(4):341–353.

Chaloner, K. and Rhame, F. S. (2001). Quantifying and documenting prior beliefs in clinical trials. *Statistics in Medicine*, 20(4):581–600.

Chen, T. T. (1997). Optimal three-stage designs for phase II cancer clinical trials. *Statistics in Medicine*, 16(23):2701–2711.

Chung, K. F. and Barnes, P. J. (2009). Mediator antagonists. In Barnes, P. J., Drazen, J. M., Rennard, S. I., and Thomson, N. C., editors, *Asthma and COPD: Basic Mechanisms and Clinical Management*, chapter 52, pages 655–662. Academic Press, second edition.

Corrigan, C. J. (2009). Immunomodulators. In Barnes, P. J., Drazen, J. M., Rennard, S. I., and Thomson, N. C., editors, *Asthma and COPD: Basic Mechanisms and Clinical Management*, chapter 55, pages 689–698. Academic Press, second edition.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.

Eisenhauer, E., Therasse, P., Bogaerts, J., Schwartz, L., Sargent, D., Ford, R., Dancey, J., Arbuck, S., Gwyther, S., Mooney, M., Rubinstein, L., Shankar, L., Dodd, L., Kaplan, R., Lacombe, D., and Verweij, J. (2009). New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *European Journal of Cancer*, 45(2):228–247.

EMEA (2011). Expert workshop on subgroup analysis: Workshop report. European Medicines Agency. EMA/916257/2011. `http://www.ema.europa.eu/docs/en_GB/document_library/Report/2011/11/WC500118092.pdf`.

Fitzpatrick, S. (2005). *Clinical Trial Design*. Institute of Clinical Research.

Fleming, T. R. (1982). One-sample multiple testing procedure for phase II clinical trials. *Biometrics*, 38(1):143–151.

French, S. and Ros Insua, D. (2000). *Statistical Decision Theory*. Arnold, London.

Friedman, L. M., Furberg, C., and DeMets, D. L. (2010). *Fundamentals of Clinical Trials*. Springer, New York.

Gehan, E. A. (1961). The determination of the number of patients required in a preliminary and a follow-up trial of a new chemotherapeutic agent. *Journal of Chronic Diseases*, 13(4):346–353.

GINA (2010). Global Initiative for Asthma. Global strategy for asthma management and prevention: Updated 2010. Technical report. `http://www.ginasthma.org/guidelines-gina-report-global-strategy-for-asthma.html`.

Hall, I. P. (2009). β_2-adrenoceptor agonists. In Barnes, P. J., Drazen, J. M., Rennard, S. I., and Thomson, N. C., editors, *Asthma and COPD: Basic Mechanisms and Clinical Management*, chapter 48, pages 609–614. Academic Press, second edition.

Hilden, J. (1990). Corrected loss calculation for phase II trials. *Biometrics*, 46(2):535–537.

Hilden, J., Bock, J. E., Andreasson, B., and Visfeldt, J. (1987). Ethics and decision theory in a clinical trial involving severe disfigurement. *Theoretical Surgery*, 1:183–189.

Humbert, M., Beasley, R., Ayres, J., Slavin, R., Hbert, J., Bousquet, J., Beeh, K.-M., Ramos, S., Canonica, G. W., Hedgecock, S., Fox, H., Blogg, M., and Surrey, K. (2005). Benefits of omalizumab as add-on therapy in patients with severe persistent asthma who are inadequately controlled despite best available therapy (GINA 2002 step 4 treatment): INNOVATE. *Allergy*, 60(3):309–316.

Humbert, M., De Blay, F., Garcia, G., Prudhomme, A., Leroyer, C., Magnan, A., Tunon-de Lara, J. M., Pison, C., Aubier, M., Charpin, D., Vachier, I., Purohit, A., Gineste, P., Bader, T., Moussy, A., Hermine, O., and Chanez, P. (2009). Masitinib, a c-kit/PDGF receptor tyrosine kinase inhibitor, improves disease control in severe corticosteroid-dependent asthmatics. *Allergy*, 64(8):1194–1201.

ICH (1997). ICH Harmonised Tripartite Guideline: General considerations for clinical trials E8. The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. `http://www.ich.org/LOB/media/MEDIA484.pdf`.

ICH (1998). ICH Harmonised Tripartite Guideline: Statistical principles for clinical trials E9. The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. `http://www.ich.org/LOB/media/MEDIA485.pdf`.

Institute of Medicine (2001). *Small Clinical Trials: Issues and Challenges*. National Academy Press. Committee on Strategies for Small-Number-Participant Clinical Research Trials `http://www.nap.edu/catalog/10078.html`.

Jennison, C. and Turnbull, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*. Chapman & Hall/CRC, Boca Raton.

Joseph, L. and Belisle, P. (1997). Bayesian sample size determination for normal means and differences between normal means. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 46(2):209–226.

Julious, S. A. (2010). *Sample Sizes for Clinical Trials*. CRC Press/Taylor & Francis, Boca Raton.

Julious, S. A. and Campbell, M. J. (1996). Letter to the editor: Sample sizes calculations for ordered categorical data by J. Whitehead, Statistics in Medicine, 12, 2257-2272 (1993). *Statistics in Medicine*, 15(10):1065–1066.

Julious, S. A. and Swank, D. J. (2005). Moving statistics beyond the individual clinical trial: applying decision science to optimize a clinical development plan. *Pharmaceutical Statistics*, 4(1):37–46.

Kim, E. S., Herbst, R. S., Wistuba, I. I., Lee, J. J., Blumenschein, G. R., Tsao, A., Stewart, D. J., Hicks, M. E., Erasmus, J., Gupta, S., Alden, C. M., Liu, S., Tang, X., Khuri, F. R., Tran, H. T., Johnson, B. E., Heymach, J. V., Mao, L., Fossella, F., Kies, M. S., Papadimitrakopoulou, V., Davis, S. E., Lippman, S. M., and Hong, W. K. (2011). The BATTLE trial: Personalizing therapy for lung cancer. *Cancer Discovery*, pages 44–53.

Lachin, J. M. (1981). Introduction to sample size determination and power analysis for clinical trials. *Controlled Clinical Trials*, 2(2):93–113.

Lan, K. K. G. and DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika*, 70(3):659–663.

Lee, M.-L. T. (1996). Properties and applications of the Sarmanov family of bivariate distributions. *Communications in Statistics - Theory and Methods*, 25(6):1207–1222.

Lewis, R. J. and Berry, D. A. (1994). Group sequential clinical trials: A classical evaluation of Bayesian decision-theoretic designs. *Journal of the American Statistical Association*, 89(428):1528–1534.

Lindley, D. V. (1961). Dynamic programming and decision theory. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 10(1):39–51.

Lindley, D. V. (1985). *Making Decisions*. Wiley, London, second edition.

Lindley, D. V. (1997). The choice of sample size. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 46(2):129–138.

Machin, D., Campbell, M. J., Tan, S., and Tan, S. (2009). *Sample Size Tables for Clinical Studies*. Wiley-Blackwell, Chichester, third edition.

Malmstrom, K., Rodriguez-Gomez, G., Guerra, J., Villaran, C., Pieiro, A., Wei, L. X., Seidenberg, B. C., Reiss, T. F., and for the Montelukast/Beclomethasone Study Group* (1999). Oral montelukast, inhaled beclomethasone, and placebo for chronic asthma. *Annals of Internal Medicine*, 130(6):487–495.

Mariani, L. and Marubini, E. (1996). Design and analysis of phase II cancer trials: A review of statistical methods and guidelines for medical researchers. *International Statistical Review / Revue Internationale de Statistique*, 64(1):61–88.

O'Brien, P. C. and Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics*, 35(3):549–556.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*. John Wiley & Sons Ltd., Chichester.

O'Hagan, A. and Stevens, J. W. (2001). Bayesian assessment of sample size for clinical trials of cost-effectiveness. *Medical Decision Making*, 21(3):219–230.

Parmigiani, G. and Inoue, L. Y. T. (2009). *Decision Theory: Principles and Approaches*. John Wiley & Sons, Chichester.

Patel, N. R. and Ankolekar, S. (2007). A Bayesian approach for incorporating economic factors in sample size design for clinical trials of individual drugs and portfolios of drugs. *Statistics in Medicine*, 26(27):4976–4988.

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika*, 64(2):191–199.

Pocock, S. J. (1983). *Clinical Trials: A Practical Approach*. Wiley, Chichester.

Pratt, J. W., Raiffa, H., and Schlaifer, R. (1995). *Introduction to Statistical Decision Theory*. MIT Press, Cambridge.

Printz, C. (2010). I-SPY 2 breast cancer clinical trial launches nationwide. *Cancer*, 116(14):3308–3308.

Reddel, H. K., Taylor, D. R., Bateman, E. D., Boulet, L.-P., Boushey, H. A., Busse, W. W., Casale, T. B., Chanez, P., Enright, P. L., Gibson, P. G., de Jongste, J. C., Kerstjens, H. A. M., Lazarus, S. C., Levy, M. L., O'Byrne, P. M., Partridge, M. R., Pavord, I. D., Sears, M. R., Sterk, P. J., Stoloff, S. W., Sullivan, S. D., Szefler, S. J., Thomas, M. D., and Wenzel, S. E. on behalf of the American Thoracic Society/European Respiratory Society Task Force on Asthma Control and Exacerbations (2009). An official American Thoracic Society/European Respiratory Society statement: Asthma control and exacerbations: Standardizing endpoints for clinical asthma trials and clinical practice. *American Journal of Respiratory and Critical Care Medicine*, 180(1):59–99.

Rice, J. A. (1995). *Mathematical Statistics and Data Analysis*. Duxbury Press, Belmont, second edition.

Ripley, B. D. (1987). *Stochastic Simulation*. Wiley Series in Probability and Statistics. John Wiley & Sons.

Sambucini, V. (2008). A Bayesian predictive two-stage design for phase II clinical trials. *Statistics in Medicine*, 27(8):1199–1224.

Schoenfeld, D. (1980). Statistical considerations for pilot-studies. *International Journal of Radiation Oncology Biology Physics*, 6(3):371–374.

Senn, S. (1996). Some statistical issues in project prioritization in the pharmaceutical industry. *Statistics in Medicine*, 15(24):2689–2702.

Senn, S. (2007). *Statistical Issues in Drug Development*. John Wiley & Sons Ltd., Chichester, second edition.

Silvey, S. D. (1975). *Statistical Inference*. Chapman and Hall, London.

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

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley, Chichester.

Stallard, N. (1998). Sample size determination for phase II clinical trials based on Bayesian decision theory. *Biometrics*, 54(1):279–294.

Stallard, N. (2012). Optimal sample sizes for phase II clinical trials and pilot studies. *Statistics in Medicine*, 31(11-12):1031–1042.

Stallard, N. and Thall, P. F. (2001). Decision-theoretic designs for pre-phase II screening trials in oncology. *Biometrics*, 57(4):1089–1095.

Stangl, D. K. and Berry, D. A. (1998). Bayesian statistics in medicine: Where are we and where should we be going? *Sankhy: The Indian Journal of Statistics, Series B*, 60(1):176–195.

Sylvester, R. J. (1988). A Bayesian approach to the design of phase II clinical trials. *Biometrics*, 44(3):823–836.

Sylvester, R. J. and Staquet, M. J. (1980). Design of phase II clinical trials in cancer using decision theory. *Cancer Treatment Reports*, 64(2-3):519–524.

Tan, S.-B. and Machin, D. (2002). Bayesian two-stage designs for phase II clinical trials. *Statistics in Medicine*, 21(14):1991–2012.

Thall, P. F. and Simon, R. (1994). Practical Bayesian guidelines for phase IIB clinical trials. *Biometrics*, 50(2):337–349.

Wald, A. (1945). Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16(2):117–186.

Wald, A. and Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, 19(3):326–339.

Wang, Y.-G. and Leung, D. H.-Y. (1998). An optimal design for screening trials. *Biometrics*, 54(1):243–250.

Whitehead, J. (1985). Designing phase II studies in the context of a programme of clinical research. *Biometrics*, 41(2):373–383.

Whitehead, J. (1986). Sample sizes for phase II and phase III clinical trials: An integrated approach. *Statistics in Medicine*, 5(5):459–464.

Whitehead, J. (1993). The case for frequentism in clinical trials. *Statistics in Medicine*, 12(15-16):1405–1413.

Whitehead, J. (1997). *The Design and Analysis of Sequential Clinical Trials*. Wiley, Chichester, revised second edition.

Yao, T.-J., Begg, C. B., and Livingston, P. O. (1996). Optimal sample size for a series of pilot trials of new agents. *Biometrics*, 52(3):992–1001.

Zhou, X., Liu, S., Kim, E. S., Herbst, R. S., and Lee, J. J. (2008). Bayesian adaptive design for targeted therapy development in lung cancer—a step toward personalized medicine. *Clinical Trials*, 5(3):181–193.