Thesis for the Degree of Licentiate of Philosophy

Optimisation of Clinical Trials using Bayesian Decision Theory

Sebastian Jobjörnsson

Division of Applied Mathematics and Statistics
Department of Mathematical Sciences
Chalmers University of Technology and University of Gothenburg
Göteborg, Sweden 2016

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under the IDEAL Grant Agreement no 602552.

Optimisation of Clinical Trials
using Bayesian Decision Theory
Sebastian Jobjörnsson

Department of Mathematical Sciences
Chalmers University of Technology and University of Gothenburg
SE-412 96 GÖTEBORG, Sweden
Phone: +46 (0)31 772 1000

Author e-mail: jobjorns@chalmers.se

Typeset with LaTeX.
Department of Mathematical Sciences
Printed in Göteborg, Sweden 2016

Optimisation of Clinical Trials using Bayesian Decision Theory

Sebastian Jobjörnsson

Department of Mathematical Sciences
Chalmers University of Technology and University of Gothenburg

Abstract

A decision maker confronted with the task of designing a clinical trial has to consider a multitude of aspects. Large trials lead to more evidence, which in turn makes it less likely that harmful decisions are taken when deciding on future treatments for patients. On the other hand, large trials typically require substantial financial resources and may take a long time to finish. This trade-off between the cost of pharmaceutical R&D and the value of the data generated motivates a study of how clinical trials can be optimised in practice.

The size of the trial is but one aspect of its design. In a setting where there is some prior evidence that a treatment works better for a subpopulation defined by a biomarker, it is natural to ask whether the resources available are best spent by restricting trial recruitment via biomarker screening. The optimal trial design also depends on the type of the decision maker. For a commercial sponsor, it is vital that invested resources may eventually be recouped via incomes from sales. On the other hand, a publicly funded trial might instead be optimised purely from a public health perspective. By viewing the trial as a stage in a sequential problem, post-trial decisions regarding pricing for a new treatment may also affect optimality.

Bayesian decision theory is a flexible framework that may be applied when searching for an optimal course of action in an uncertain environment. In particular, it allows for different beliefs prior to the trial and different preferences for the trial outcomes. This thesis presents two papers in which Bayesian decision theory is used to find the optimal trial design under two different models for the contemporary regulatory environment. In addition to providing a methodology for trial design in the specific situations considered, the analysis also leads to insights regarding the impact of typical regulatory rules on the behaviour of trial sponsors.

Keywords: Bayesian statistics, decision theory, clinical trials, pharmaceutical R&D, drug regulation, subgroup analysis, multiple testing, health economics.

Acknowledgements

First of all, I would like to express my gratitude to Carl-Fredrik Burman who has provided constant support and guidance during my work on the IDEAL project.

I thank Martin Forster and Paolo Pertile for the hospitality shown during my visit to Verona and their tireless efforts during our work with paper I.

I also thank Thomas Ondra, Martin Posch and Franz König for receiving me several times in Vienna and for their great work on paper II.

Sebastian Jobjörnsson
Gothenburg, August 17, 2016

List of appended papers

Paper I **Jobjörnsson, S.**, M. Forster, P. Pertile, and C.-F. Burman (2016). Late-Stage Pharmaceutical R&D and Pricing Policies under Two-Stage Regulation. *Forthcoming in Journal of Health Economics*.

Paper II Ondra, T., **S. Jobjörnsson**, R. A. Beckman, C.-F. Burman, F. König, N. Stallard, and M. Posch (2016). Optimizing Trial Designs for Targeted Therapies. *Submitted to PLOS ONE*.

My contribution to the appended papers:

Paper I: I jointly developed the model and wrote the manuscript together with the other authors. I did most of the proofs and implemented the code for the application of the model.

Paper II: The model was jointly formulated by all authors at the beginning of the project. I and Thomas Ondra (first author) contributed equally to the derivation of some simplifying formulas for computing expected utilities and to the implementation of the code for trial design optimisation.

List of abbreviations

BDT	Bayesian Decision Theory
EMA	European Medicines Agency
EVPI	Expected Value of Perfect Information
FDA	U.S. Food and Drug Administration
FDR	False Discovery Rate
FWER	Family-Wise Error Rate
HCI	Health Care Insurer
HTP	Health Technology Provider
ICER	Incremental Cost-Effectiveness Ratio
IDEAL	Integrated DEsign and AnaLysis
MTP	Multiple Testing Procedure
NICE	National Institute of Clinical Excellence
NPV	Net Present Value
PCER	Per Comparison-Error Rate
QALY	Quality Adjusted Life Year
RA	Regulatory Authority
WTP	Willingness To Pay

Contents

Chapter 1

Introduction

This licentiate thesis consists of two papers in which Bayesian Decision Theory (BDT) is used to optimise clinical trial designs. Both papers consider optimisation problems where a pharmaceutical company's expected profit is to be maximised. From such a perspective, the aim is to choose a trial design that balances the probability of getting regulatory approval for market introduction of a new medical treatment with the cost and time duration associated with the design. Paper II also considers trial optimisation from a public health perspective, replacing the profit maximisation goal with that of maximising the total health benefits for the target population.

The idea of using utility functions to associate a numerical value with every possible outcome of a decision process is basic to the BDT framework. As compared to, for example, the classical power-based method of determining a sample size, this brings additional flexibility since all relevant gains and costs associated with the trial can be aggregated and represented in terms of a utility function. Given a fixed set of regulatory rules for the market introduction of a new treatment, the specification of a probabilistic model and an appropriate utility function leads directly to a maximisation problem. The purpose of this thesis is to provide an analysis of how rational decision makers act in some different models for the regulatory environment for approval and reimbursement of new medical treatments.

The work presented in this thesis is part of an EU project called Integrated DEsign and AnaLysis of clinical trials in small population groups (IDEAL). Small population groups may arise because the treatment is for a rare disease or is only expected to work in a small subset of the total population. Trial design in such situations often requires a more careful analysis of how to make the most of the necessarily small sample sizes. Consisting of a multidisciplinary consortium of researchers from universities, research institutes and the industry,

the purpose of the IDEAL project is to develop statistical methods that are sufficiently general to handle also the special case of small population groups. From a regulatory perspective, small population groups also raise questions that are not purely statistical. If a pharmaceutical company expects the final market to be small, will it view basic research in this area as a viable investment? If not, what needs to be changed in the regulatory structure so as to incentivise more research targeting small population groups? The economic resources of any society are limited, so how should the cost of implementing the incentives be balanced against the potential health gains? To answer these questions, it is necessary to have a good understanding of how rational decision makers act in different regulatory situations.

The content of this thesis is structured as follows. Chapter 2 contains a brief introduction to the basic terminology and history of clinical trials. It also describes the typical regulatory structure in place for the approval of new medical treatments and introduces some terminology for biomarkers which is used in paper II. Chapter 3 contains an introduction to the BDT methodology, which is the basic tool used to frame the problems of both paper I and paper II. To illustrate the typical steps involved in modelling and optimisation, this section contains three example applications: the optimisation of sample size in a clinical trial from a public health perspective, the computation of the expected value of perfect information about a certain regulatory parameter, and the sequential optimisation of the sample size for a phase II trial and the subsequent GO or NO GO decision for two phase III trials. Chapter 4 presents the basic notions of multiple hypothesis testing. In particular, the Speissens-Debois adjustment procedure used in paper II is discussed in detail. Two basic results from mathematical analysis, the implicit function theorem and the envelope theorem, are briefly stated in Chapter 5. Both are extensively used in the theory presented in paper I. Chapter 6 contains short summaries of the contents of paper I and paper II. Chapter 7 contains a discussion and a conclusion.

Chapter 2

Clinical trials

A *clinical trial* is an experiment performed on human subjects for the purpose of generating data that may be used to evaluate the efficacy and safety of a medical treatment. The treatment need not be a drug, but could just as well be a certain diet, a specific usage of a medical device or any other well-defined intervention in the life of a patient. Often, the term 'drug' will be used in this thesis to refer to a chemical substance under clinical development. However, in many cases the discussion also carries over to the more general concept of a medical treatment.

The purpose of performing a late-stage *confirmatory* or *pivotal* clinical trial is to generate information about a treatment that may aid a governmental body tasked with determining if the benefit outweighs the associated risk in a certain target population. Often, the treatment is compared with a *placebo* alternative, which could for example be inert tablets (sugar pills). After the trial, when a decision is to be made on whether or not to introduce a new treatment, there may still remain some uncertainty as to which choice that is the right one. On the one hand, approval of a new treatment with unknown safety issues could lead to serious long-term consequences for some patients. On the other hand, an overly strict policy of rejection could mean that potentially beneficial treatments would never reach the patients. The information obtained from a well-designed clinical trial reduces the probability that the wrong approval decision is made. In general, more participants in the trial leads to a greater body of evidence for supporting the approval decision. At the same time, clinical trials are typically very costly and the resources spent on a particular trial might have been better spent elsewhere. This is the basic trade-off: how much of available resources is it worthwhile to invest in a trial in order to reduce the uncertainties regarding the efficacy and safety properties of a given treatment? In addition to the increased cost of a larger trial, it is also important to keep in mind that a larger trial will

typically take a longer time to perform. Since the new treatment is unavailable to the general population for the duration of the trial, this provides another reason to keep it at a moderate size as long as the amount of evidence obtained is not overly compromised.

The following sections in this chapter give a brief introduction to the history and current status of clinical trial methodology. There are a number of books and review articles available for a reader who wishes to learn more about the subject. Chow and Liu (2014) provides a broad overview of both the practical and theoretical issues of the design and analysis of clinical trials. They cover the regulatory structure for approval of new treatments as implemented by the FDA and the basic statistical methods available for different kinds of trials. Speigelhalter et al. (2004) focus on the use of the Bayesian approach but also cover issues connected to health-care evaluation, such as cost-effectiveness analysis. An introduction to adaptive Bayesian designs is given by Berry et al. (2011). Ondra et al. (2016) present a review of different trial designs involving biomarkers that have been proposed in the literature.

2.1 History of clinical trials

The regulations controlling the testing and market introduction of new treatments have been under continuous development during the last century, eventually leading up to the present day situation in which substantial financial investments and many years of development are needed in order to bring a new drug from discovery to patient usage. DiMasi et al. (2016) estimate the average out-of-pocket cost per approved new compound to be 1395 million USD (2013 dollars). Kaitin and DiMasi (2011) present data on the average times required for the total clinical phase (phase I, II and III trials) and subsequent approval phase (evaluation of trial evidence by regulators) for new drugs approved in the U.S. For the 5-year period 2005-2009, average times of 6.4 years for the total clinical phase and 1.2 years for the approval phase are reported. Before a drug enters the clinical development stage, several years have typically been spent on pre-clinical research.

The regulations that exist for the marketing of new treatments are not the same across different regions of the world. From the viewpoint of this thesis, the most important regulatory agencies are the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA). The FDA was formed in 1931. Its main objectives, as regards drug regulation, is to ensure that drugs are safe and effective and that such products are honestly and informatively labelled (Chow and Liu, 2014). As an example of the importance of testing new treatments on human subjects in clinical trials, before making a decision on whether or not to bring the treatment to the open market, we

may consider the well-known example of the Elixir Sulfanilamide disaster which occurred in the late 1930s. This drug had never been tested in human subjects before market introduction. Safety issues with this drug eventually led to more than 100 deaths and to the passing of The Federal Food, Drug and Cosmetic Act in 1938, requiring pharmaceutical companies to submit reports on the safety of new drugs. A later amendment (the Kefauver-Harris Amendment of 1962) strengthened the safety requirements and also introduced requirements on efficacy for the first time.

EMA was founded in 1995 with the purpose to harmonise the work of the national level regulatory bodies within the EU (EMA, 2016). It works to ensure the efficacy and safety of human and veterinary medicines in Europe and provides a centralised procedure for the evaluation of applications for marketing authorisation. EMA does not decide on the price or availability of authorised treatments. Such decisions are made on the national level by the EU member states' health technology assessment bodies.

2.2 Contemporary clinical development

Speigelhalter et al. (2004) use the following classification for the stakeholders involved in the clinical development process. *Sponsors* are agents (e.g., pharmaceutical companies) paying for the trial and are generally interested in eventually being able to make a profit from the investments in research. *Investigators* are responsible for the actual conduct of the study. *Reviewers* are responsible for analysing the safety and efficacy issues and decide on marketing approval, whereas *policy-makers* are more concerned about the total cost-benefit impact of a new treatment. The *consumers* are the individual patients that derive benefit from a treatment.

In this thesis, a slightly different terminology is used. For-profit pharmaceutical companies that aim to get regulatory approval for introducing their products on the market are referred to as *commercial sponsors*. However, a sponsor could also be a publicly funded research institute or health care organisation that aims to maximise the quality and quantity of the health care benefiting society at large. Hence, a sponsor in the most general sense is an agent paying for the trial. A reviewer is referred to as a *Regulatory Authority* (RA). Using the clinical trial evidence provided by a sponsor, it decides on marketing approval for a proposed treatment. Given that such approval has been granted, it is up to a *Health Care Insurer* (HCI) to decide on the level of payment for the new treatment, balancing economic considerations with the overall benefit-risk profile of the new treatment. The general terminology for the stakeholders involved in the process of bringing a new treatment to a market can be applied to the regional EU and U.S. markets. In the U.S., the FDA takes

on the role as the RA. Payment is typically provided by insurance companies, which thus constitute the HCI's. In the EU, EMA is the RA, while the HCI's are taken to be the country-level health care authorities or insurance companies.

We will now consider the different viewpoints and aims of a sponsor, an RA and a HCI. For a commercial sponsor, the goal of the entire development process is to obtain a return on the investment made in the clinical trials for the new treatment. When the sponsor deliberates on which decision to make at a certain point in the process, its objective is assumed to be to maximise its expected future profits. The main decision rule used by the RA is assumed to be the demonstration of a statistically significant effect for the new treatment vs. placebo relative to some specified value for the type I error. This demonstration should be done by means of one or two (independent) clinical trials. The conventional requirement may be found in section 3.5 of the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use's (ICH) guidance on efficacy E9 (ICH, 2016),

> Conventionally the probability of type I error is set at 5% or less or as dictated by any adjustments made necessary for multiplicity considerations; the precise choice may be influenced by the prior plausibility of the hypothesis under test and the desired impact of the results. The probability of type II error is conventionally set at 10% to 20%; it is in the sponsor's interest to keep this figure as low as feasible especially in the case of trials that are difficult or impossible to repeat. Alternative values to the conventional levels of type I and type II error may be acceptable or even preferable in some cases.

A significance level of 5% was suggested by Fisher (1946) as a convenient cutoff level to reject a null hypothesis. Although this suggestion has developed into a regulatory convention, Fisher did not intend that this cutoff level be fixed at the same level regardless of application. Instead, he recommended that a specific significance level be set according to circumstances (Fisher, 1956). Typically, a requirement on power is set at between 80 and 90%. The RA will certainly also consider safety aspects of a treatment, although the precise nature of safety evaluations are often not as explicit as the requirements on efficacy. Neither paper I nor paper II contains detailed models of the RA's safety considerations.

Given that the treatment has been deemed sufficiently safe and efficacious for market introduction by the RA, the HCI decides on reimbursement based on both the benefit-risk balance and the cost of the treatment. The extent to which monetary and purely health related concerns are (or should be) combined is a complex ethical issue. Hence, it is unsurprising that the HCI conduct differs greatly between individual nations and regions. One HCI of particular importance to this thesis is the National Institute of Clinical Excellence (NICE) in

the U.K. The role of NICE is to provide guidance to health providers and help support the decisions on which treatments that are to be covered by governmental reimbursement schemes. It is used, in paper I, as a specific example of a HCI that is willing to directly associate an incremental health benefit with a monetary cost. For this, it combines both positive and negative health effects of the treatment into a common measure referred to as Quality Adjusted Life Years (QALYs).

Since the late 1970s, the FDA divides clinical development into three sequential stages. These stages are referred to as phase I, II and III (Chow and Liu, 2014). Phase I trials include a small number of healthy volunteers. The main objective of these early trials is to obtain information about the drug's pharmacokinetic and pharmacologic properties and how the efficacy and side effects vary with the dose level. Barring any serious safety events, the testing proceeds to phase II. The participants in this stage typically number in the hundreds. The main objective of phase II is to find the proper dosing for the new drug, striking a balance between its beneficial effects and the risks associated with any potential side effects. The procedure is concluded in phase III, which consists of one or several (independent) relatively large clinical trials (participants may number in the thousands), the purpose of which is to provide the major part of the evidence that the overall benefit-risk balance of the new drug is acceptable. The two papers which constitute this thesis are concerned with phase III trials.

2.3 Biomarkers and subgroup analyses

Paper II is concerned with optimising pivotal clinical trials investigating treatments for which there exists some indication of a superior performance in a subgroup of the target population. One way to define subgroups is through the concept of a *biomarker*. A definition which fits the purpose of this thesis is the one provided by the Biomarkers Definitions Working Group (2001),

> Biological marker (biomarker): A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.

A biomarker can be either *prognostic* or *predictive* (or both). A prognostic biomarker is one which can be used to characterise the general outcome for a patient with a certain condition, independently of any specific treatment. In contrast, a biomarker is predictive only with respect to a specific treatment, the idea being that the biomarker status may be used to predict how well a patient responds to the treatment in question. In general, a patient response consists

of an efficacy response and a safety response, corresponding to any side-effects of the treatment. Whether or not the biomarker can be used to predict the efficacy response or the safety response (or both) depends on the type of the biomarker and the properties of the screening procedure used to determine its status for a specific patient.

In general, there are both discrete and continuous biomarkers. However, paper II only considers the special case of a predictive binary biomarker. Note that the biomarker status of a patient, as determined by the screening procedure, is not treated as a response variable in this thesis. It is only used as a covariate that influences the efficacy response. Although a biomarker need not be genetically based, an important area where such examples do occur is that of clinical trials investigating cancer treatments (see, for example, Ziegler et al. (2012) for a high-level review). Cancer involves changes to the DNA at a cellular level which in some cases can be directly measured in the tumour. These changes may sometimes be mapped to a binary biomarker status corresponding to, for example, the presence or absence of a specific gene mutation. The biomarker status can then be used to decide on the best course of treatment.

2.4 Bayesian vs. frequentist methods in clinical trials

Although Bayesian methods continue to see an increasing use, the dominating approach to the design and analysis of confirmatory clinical trials is still what is commonly referred to as the *classical* or *frequentist* approach. As noted by Senn (2007), this classical approach is really a hybrid one, employing the Neyman-Pearson framework during the design stage while being Fisherian during analysis. The purpose of this thesis is not to enter into a discussion about the respective virtues and drawbacks of the frequentist and Bayesian approaches. Rather, it is recognised that the Bayesian decision theoretic approach seems a viable one for handling the practical problems of specific concern to this thesis, namely, how to make the most of the limited sample sizes achievable for rare disease trials and how to combine different aspects of the possible outcomes of a trial by means of utility functions.

It is important to note that when a particular decision problem is solved in this thesis, it is not assumed that all stakeholders involved in the process are taking decisions which are optimal relative to some BDT model. Instead, optimisation with respect to a full BDT model is done only for the sponsor. The decision rules for the RA and the HCI are modelled after current conventions. For the RA, this typically means that the rule is based on a classical significance test.

Chapter 3

Bayesian decision theory

Bayesian decision theory is a prescriptive theory that can be used to determine the optimal course of action for a decision maker faced with a situation involving uncertainty. The theory is referred to as 'Bayesian' because any uncertainty, regardless of its source, is associated with a probability distribution. An argument which is often put forward to justify the use of BDT is that a decision maker which aims to avoid inconsistent behaviour must act as if it represents degrees of beliefs by probability distributions and preferences about possible consequences by a utility function. Moreover, the action to be chosen should be the one that maximises the expected utility. Formally, this result is shown by imposing certain axioms (often referred to as 'axioms of rationality') on a mathematical abstraction of a general decision problem. The purpose of the axioms is to ensure that certain behaviour that many would regard as obviously inconsistent, or irrational, is avoided by any decision maker that follows them. Many variations of this formal argument can be found in the literature. For one of the earlier expositions, see Savage (1972). Another comprehensive treatment of the foundations of BDT is given by Bernardo and Smith (1994).

The clinical trial design problems that are analysed in papers I and II can both be formulated as *statistical decision problems* (DeGroot, 1970, p. 136). In this class of problems, the decision maker is to make a *final decision*, the consequence of which depends on the value of some unknown parameter. Before making the final decision, the decision maker has the option to perform an experiment in order to increase the knowledge that it has about the parameter value. The data generated by the experiment is used to increase the probability of making a final decision leading to the most desirable consequence. Following Raiffa and Schlaifer (1961), a statistical decision problem may formally be described as an ordered list with four components, $(\mathcal{E}, \mathcal{X}, \mathcal{D}, \Theta)$, where

- $\mathcal{E}$ is a set of experiments,
- $\mathcal{X}$ is a sample space of possible observations that may result from the experiments in $\mathcal{E}$,
- $\mathcal{D}$ is a set of possible decisions, and,
- Θ is a parameter space.

The decision maker selects an experiment $e \in \mathcal{E}$, observes a result $X \in \mathcal{X}$[1], and selects a final decision $d \in \mathcal{D}$. The consequence that results from d depends on the value of $\theta \in \Theta$.

In addition to the basic structure formalised by $(\mathcal{E}, \mathcal{X}, \mathcal{D}, \Theta)$, in order to complete the specification of a statistical decision problem and thus enable the computation of the optimal course of action, the decision maker must specify its beliefs regarding X and θ in terms of a joint probability distribution. Further, it must also specify its preferences regarding any combination of decisions, observations and parameter values by means of a real-valued utility function $u(e, x, d, \theta)$. Typically, the joint distribution of (X, θ) is defined by specifying a prior distribution for θ and a conditional distribution for the observation X given that the experiment e is performed and that θ is the true parameter value.

Having fully formalised the decision problem, the goal of the decision maker is to find a *decision rule* δ which maximises its expected utility. Such a decision rule consists of an experiment e_δ and a function d_δ that maps every possible outcome of the experiment to a final decision $d_\delta(x)$. Hence, the problem may be written as

$$\max_{\delta} \mathbb{E}\left[u(e_\delta, X, d_\delta(X), \theta)\right]. \tag{3.1}$$

The decision problem analysed in paper I has two stages. In the first stage, a pharmaceutical company chooses a sample size for a phase III trial and then observes the outcome. In the second stage, the company chooses a price for the new treatment. A HCI then decides whether or not to reimburse the company for the new treatment. This problem is easily reinterpreted as a statistical decision problem. Specifically, one may take $\mathcal{E}$ to be the union of a NO GO decision and the set of possible sample sizes corresponding to a GO decision in the first stage. $\mathcal{D}$ can be chosen as the set of possible prices in the second stage. In paper II, there is only a single stage, corresponding to the selection of a clinical trial design (of which the sample size decision constitutes one part). However, that situation may also be mapped to a statistical decision problem by simply taking $\mathcal{E}$ to be the set of available trial designs and letting $\mathcal{D}$ be a

[1] We follow the convention of denoting random variables by capital letters and specific realisations by the corresponding small letters. For example, X is a random variable before the experiment, with some specific realisation x after the experiment is completed.

set containing a single element. Such a definition for $\mathcal{D}$ implies that there is no choice involved and that the statistical decision problem therefore is reduced to a single stage problem.

3.1 The method of backward induction

An optimal decision rule δ^* can be found using the method of backward induction, which proceeds backwards in the decision problem while computing a sequence of induced utilities. The first step is to compute the expected utility of choosing d given that x was observed in the experiment e,

$$\bar{u}(e, x, d) \equiv \mathbb{E}\left[u(e, x, d, \theta)|e, x\right].$$

This is done for all triples $(e, x, d) \in \mathcal{E} \times \mathcal{X} \times \mathcal{D}$. Since the decision maker is free to choose any $d \in \mathcal{D}$, the rational choice is to select the one maximising $\bar{u}(e, x, d)$. Therefore, the optimal decision rule given e is

$$d_{\delta^*}(x) = \arg\max_{d \in \mathcal{D}} \bar{u}(e, x, d),$$

with a corresponding induced expected utility

$$\bar{u}^*(e, x) \equiv \max_{d \in \mathcal{D}} \bar{u}(e, x, d).$$

The next step is to find the experiment that maximises the expected utility of observing the outcome X, given that the optimal rule $d_{\delta^*}(x)$ is subsequently followed. The induced expected utility corresponding to a specific e may be written as

$$\tilde{u}(e) \equiv \mathbb{E}\left[\bar{u}^*(e, X)|e\right],$$

and the optimal choice of e is therefore $e^* \equiv \arg\max_{e \in \mathcal{E}} \tilde{u}(e)$, with a corresponding optimal expected utility $\tilde{u}^* = \max_{e \in \mathcal{E}} \tilde{u}(e)$. The optimal decision rule $\delta^* = (e^*, d_{\delta^*})$ constructed in this way solves the problem in Eq. (3.1). The interleaving operations of expectation and maximisation used to obtain the rule implies that the corresponding optimal utility may be written as

$$\tilde{u}^* = \max_{e \in \mathcal{E}} \mathbb{E}\left[\max_{d \in \mathcal{D}} \mathbb{E}\left[u(e, X, d, \theta)|e, X\right] \middle| e\right].$$

The formalism for a single experiment can easily be generalised to situations involving a sequence of experiments with corresponding observations. Such a *sequential statistical decision problem*[2] of k stages prior to the final decision

[2] Clearly, the original problem already contains two stages, and may therefore be considered a sequential decision problem. However, since the notion of a statistical decision problem always formally contain at least one experiment and one final decision, 'sequential' is used when more than one experiment is performed before the final decision.

may be specified as $(\mathcal{E}_1, \mathcal{X}_1, \ldots, \mathcal{E}_k, \mathcal{X}_k, \mathcal{D}, \Theta)$. This generalisation leads to no conceptual difficulties since the method of backward induction may still be used to construct an optimal decision rule. However, the computational effort required to find the optimal rule obviously increases with the number of stages.

In order to illustrate the method of backward induction, the remaining sections of this chapter describes three different examples.

3.2 Optimisation of sample size

This section presents an example of sample size optimisation in the context of a parallel-group trial comparing a new treatment A with a placebo alternative B. An equal randomisation of n patients to the two groups is assumed, implying a per-group sample size of $n/2$. In practice, different group sizes may result as a consequence of the recruitment or randomisation process. Such issues are ignored in the simple model treated here.

The unknown, incremental clinical utility[3] of A vs. B is denoted by θ. The individual responses in the two arms are assumed to be i.i.d. random variables which are combined into two sample means, $\bar{X}_A$ and $\bar{X}_B$. The variance of each individual response is for simplicity assumed to be known and equal to σ^2. The difference $\bar{X} = \bar{X}_A - \bar{X}_B$ is used to estimate θ under the assumption that

$$\bar{X} \mid n, \theta \sim \mathcal{N}\left(\theta, 4\sigma^2/n\right).$$

A normal conjugate prior with zero mean is assumed for θ, so that

$$\theta \sim \mathcal{N}\left(0, 4\sigma^2/n_0\right).$$

This prior formalises the knowledge about θ that is available to the decision maker before the data from the trial has been observed. By the probabilistic assumption of a normal conjugate model, the posterior distribution of θ given the observed value of the sample mean $\bar{X}$ (and n) is also normal. Writing the prior variance as $4\sigma^2/n_0$ makes it easy to compare the amount of information about θ in the prior with the information about θ provided by the trial sample.

The decision procedure is as follows. First, a non-negative sample size $n \in [0, N]$ is selected, where $N > 0$ denotes the total size of the target population for the new treatment. To facilitate the use of calculus to solve the problem, n is assumed to be a continuous quantity, although in practice it must of course be an integer. After the trial, a decision $d \in \{0, 1\}$ is taken on which treatment to give to the patients in the target population, where $d = 1$ and $d = 0$ correspond

[3] *Clinical utility* is taken to be some measure that incorporates both efficacy and safety aspects (side-effects) of a treatment. It should be thought of as a real number corresponding to the net health benefit that a treatment brings to a patient.

to treatment A and B, respectively. Hence, in terms of the general list structure, for this particular problem we have $(\mathcal{E} = [0, N], \mathcal{X} = \mathbb{R}, \mathcal{D} = \{0, 1\}, \Theta = \mathbb{R})$.

The utility of choosing a sample size n, observing a sample mean $\bar{x}$, and taking the post-trial decision d, given that θ is the true incremental clinical utility in the population, is taken to be the aggregated, incremental clinical utility for both in-trial and post-trial patients. Note that, in order to simplify the problem for presentation purposes, any monetary trial costs are ignored. However, there are still 'trial costs' involved (in units of clinical utility) since half of the patients in the trial are given an inferior treatment. Formally,

$$u(n, \bar{x}, d, \theta) = \frac{n\bar{x}}{2} + d(N - n)\theta.$$

We now proceed to solve the problem of finding the optimal sample size n^* maximising the expected utility. By the standard result for normal conjugate updating (Raiffa and Schlaifer, 1961), the posterior distribution of θ given n and $\bar{X}$ is

$$\theta \mid n, \bar{X} \sim \mathcal{N}\left(\frac{n\bar{X}}{n_0 + n}, \frac{4\sigma^2}{n_0 + n}\right).$$

It follows that the induced utility of choosing n, observing $\bar{x}$ and subsequently selecting a treatment option d is

$$\begin{aligned}\bar{u}(n, \bar{x}, d) &= \mathbb{E}\left[\frac{n\bar{x}}{2} + d(N - n)\theta \middle| n, \bar{x}\right] = \frac{n\bar{x}}{2} + d(N - n)\mathbb{E}\left[\theta | n, \bar{x}\right] \\ &= \frac{n\bar{x}}{2} + d(N - n)\left(\frac{n\bar{x}}{n_0 + n}\right).\end{aligned}$$

Maximising over $d \in \{0, 1\}$ gives

$$\bar{u}^*(n, \bar{x}) = \frac{n\bar{x}}{2} + \frac{(N - n)n}{n_0 + n}\max\left(0, \bar{x}\right).$$

Now, it may easily be shown that if Y is a random variable distributed as $Y \sim \mathcal{N}\left(0, \sigma_Y^2\right)$, then $\mathbb{E}\left[\max(0, Y)\right] = \sigma_Y/\sqrt{2\pi}$. Therefore, since the prior predictive distribution for $\bar{X}$ is given by $\bar{X} \sim \mathcal{N}\left(0, 4\sigma^2/n_0 + 4\sigma^2/n\right)$,

$$\tilde{u}(n) = \frac{(N - n)n}{n_0 + n}\mathbb{E}\left[\max\left(0, \bar{X}\right)\middle| n\right] = \frac{N - n}{\sqrt{2\pi}}\sqrt{\frac{4\sigma^2}{n_0}}\sqrt{\frac{n}{n_0 + n}}. \tag{3.2}$$

It follows directly from the expression for $\tilde{u}(n)$ in Eq. (3.2) that it is a differentiable function of n and that there must exist a global maximiser in $(0, N)$. By computing $\tilde{u}'(n)$ and solving $\tilde{u}'(n) = 0$, it follows that the optimal sample size is given by

$$n^* = \frac{n_0}{4}\left(\sqrt{9 + \frac{8N}{n_0}} - 3\right). \tag{3.3}$$

Fixing n_0 and using Eq. (3.3), it is straightforward to show that

$$n^* \sim \sqrt{\frac{n_0}{2}}\sqrt{N}, \quad N \to \infty.$$

If N is held fixed, then $\lim_{n_0 \to 0} n^* = 0$ and $\lim_{n_0 \to \infty} n^* = N/3$.

An interesting recent contribution by Stallard et al. (2016) contains a detailed analysis of a generalised version of the problem considered in this section. The authors show that the optimal sample size is $O(\sqrt{N})$ as $N \to \infty$, under the assumptions that the distribution for the primary endpoint has a one-parameter exponential family form and that the utility for each patient is a continuous function of the parameter. In addition to the result for a fixed value of N, they also extend the asymptotic analysis to the case in which N is unknown and has a geometric distribution.

3.3 EVPI analysis for a sponsor

This section presents an example using the concept of Expected Value of Perfect Information (EVPI). For a formal, general definition, see Raiffa and Schlaifer (1961, p. 88). To simplify the presentation, the concept of EVPI will be defined here only within the context of the specific example discussed. As noted previously, although both efficacy and safety concerns are important, the precise way in which these are combined and jointly evaluated by a RA is often unclear. Such uncertainty about the RA's decision rule impacts the optimal decision-making behaviour for a commercial sponsor. The EVPI provides a formal way to quantify the sponsor's expected gain if the RA clarifies its decision rule.

Suppose that a commercial sponsor is planning a confirmatory, parallel-group clinical trial comparing a new treatment with placebo. Assume an equal randomisation resulting in $n/2$ patients per group. The incremental mean efficacy and safety responses in the target population are denoted by θ_e and θ_s, respectively. It will be assumed that the data from the trial is combined into estimates $\bar{X}$ and $\bar{Y}$, distributed according to

$$\begin{bmatrix} \bar{X} \\ \bar{Y} \end{bmatrix} \mid n, \begin{bmatrix} \theta_e \\ \theta_s \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \theta_e \\ \theta_s \end{bmatrix}, \frac{4}{n} \begin{bmatrix} \sigma_e^2 & \rho\sigma_e\sigma_s \\ \rho\sigma_e\sigma_s & \sigma_s^2 \end{bmatrix} \right). \tag{3.4}$$

Further, it will also be assumed that the RA combines the efficacy and safety responses into a clinical utility response $\bar{X} - \kappa\bar{Y}$, and that market approval is granted if the null hypothesis $H_0 : \theta_e - \kappa\theta_s = 0$ can be rejected for a one-sided significance level α. For simplicity, the parameters θ_e and θ_s are assumed to be known by the sponsor. The computations that follow can also be carried out

when a prior is placed on them, but the focus in this section is that the sponsor is uncertain about the value of κ used by the RA.

By the distributional assumption in Eq. (3.4), it follows that

$$\bar{X} - \kappa\bar{Y} \mid n, \begin{bmatrix}\theta_e \\ \theta_s\end{bmatrix} \sim \mathcal{N}\left(\theta_e - \kappa\theta_s, \frac{4}{n}\left(\sigma_e^2 - 2\kappa\rho\sigma_e\sigma_s + \kappa^2\sigma_s^2\right)\right),$$

and that the RA rejects H_0 and approves the new treatment for marketing if

$$\frac{\bar{X} - \kappa\bar{Y}}{\sqrt{\frac{4}{n}\left(\sigma_e^2 - 2\kappa\rho\sigma_e\sigma_s + \kappa^2\sigma_s^2\right)}} > z_\alpha \iff \bar{X} - \kappa\bar{Y} > z_\alpha\sqrt{\frac{4}{n}}\sigma_c,$$

where $z_\alpha = \Phi^{-1}(1-\alpha)$ and $\sigma_c = \sqrt{\sigma_e^2 - 2\kappa\rho\sigma_e\sigma_s + \kappa^2\sigma_s^2}$.

Given a total target population of N patients, define the utility for the sponsor to be $b(N-n) - (c_0 + cn)$ if there is RA approval and $-(c_0 + cn)$ otherwise, where b is the monetary benefit per patient treated after the trial, c_0 is a fixed setup cost for the trial and c is the marginal cost per patient included in the trial. The utility of choosing a sample size n and obtaining estimates $\bar{X} = \bar{x}$ and $\bar{Y} = \bar{y}$, given that κ is the weight used by the RA, is then

$$u(n, \bar{x}, \bar{y}, \kappa) = b(N-n)\mathbb{I}\left\{\bar{x} - \kappa\bar{y} > z_\alpha\sqrt{\frac{4}{n}}\sigma_c\right\} - (c_0 + cn),$$

implying that the expected utility for the sponsor of choosing a sample of size n given κ is

$$\begin{aligned}\bar{u}(n,\kappa) &= b(N-n)\mathbb{P}\left(\bar{X} - \kappa\bar{Y} > z_\alpha\sqrt{\frac{4}{n}}\sigma_c \middle| \kappa\right) - (c_0 + cn) \\ &= b(N-n)\Phi\left(\sqrt{\frac{n}{4}}\left(\frac{\theta_e - \kappa\theta_s}{\sigma_c}\right) - z_\alpha\right) - (c_0 + cn). \qquad (3.5)\end{aligned}$$

The expected utility of choosing a sample of size n is obtained by taking the expectation of the expression above with respect to the distribution for κ,

$$\tilde{u}(n) = \mathbb{E}\left[\bar{u}(n,\kappa)\right] = b(N-n)\mathbb{E}\left[\Phi\left(\sqrt{\frac{n}{4}}\left(\frac{\theta_e - \kappa\theta_s}{\sigma_c}\right) - z_\alpha\right)\right] - (c_0 + cn). \quad (3.6)$$

For the present situation, the EVPI about κ for the sponsor may be interpreted as the difference in the sponsor's expected utility when acting optimally in the following two situations

1. The sponsor is provided with the true value of κ, thus obtaining full knowledge about the RA's decision rule for approval, and then chooses a sample size.

2. The sponsor chooses a sample size based on its prior beliefs about κ.

Note that the expected value over the prior for κ is taken for both situations. Although the sponsor knows that κ will be revealed before the sample size decision in the first situation, an integration over the prior is necessary also for this case. The reason is that the precise value communicated is still uncertain when the expected utility is evaluated.

Let $n^*(\kappa)$ be the sample size maximising $\bar{u}(n, \kappa)$ as defined in Eq. (3.5) and let n^* be the sample size maximising $\tilde{u}(n)$ as defined in Eq. (3.6). The EVPI about κ may then be written as

$$\text{EVPI} = \mathbb{E}\left[\bar{u}(n^*(\kappa), \kappa) - \bar{u}(n^*, \kappa)\right] = \mathbb{E}\left[\bar{u}(n^*(\kappa), \kappa)\right] - \tilde{u}(n^*).$$

Since $n^*(\kappa)$ is by definition the optimal sample size given κ, $\bar{u}(n^*(\kappa), \kappa) - \bar{u}(n^*, \kappa)$ is always non-negative. Hence, EVPI ≥ 0 always holds. For a numerical example, see Figure 3.1, in which $\log \kappa \sim \mathcal{N}(\mu_\kappa, \sigma_\kappa^2)$. The sample sizes were optimised under the restrictions $n^*(\kappa) \geq 1$ and $n^* \geq 1$. As the uncertainty about κ is reduced, i.e., as $\sigma_\kappa \to 0$, the EVPI tends to 0. Note that for $\sigma_\kappa = 0.2$, the expected utility for the optimal sample size is negative for situation 2 while positive for situation 1. Hence, a sponsor with such a prior for κ could be convinced to switch from a NO GO to a GO decision under the assurance that the true value of κ will be revealed before the sample size is chosen.

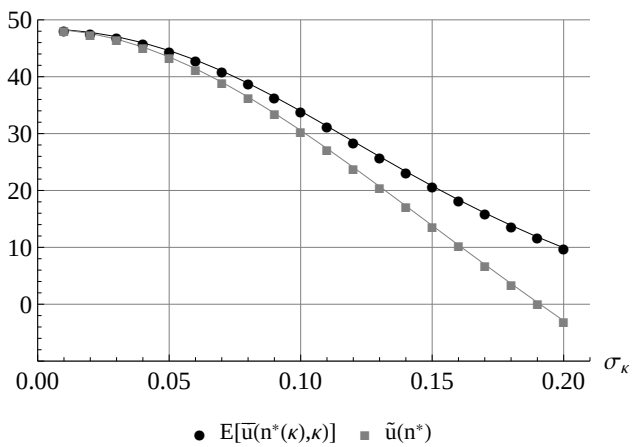

Figure 3.1: $\mathbb{E}\left[\bar{u}(n^*(\kappa), \kappa)\right]$ (black circles) and $\tilde{u}(n^*)$ (grey squares) as functions of σ_κ. Parameter values: $\theta_e = 1$, $\theta_s = 0.5$, $\sigma_e = \sigma_s = 1$, $\rho = 0.5$, $b = 1$, $c = 2$, $c_0 = 400$, $\alpha = 0.025$, $N = 1000$, $\mu_\kappa = 0$.

3.4 Optimisation of phase II sample size and phase III GO decision

This section solves a multi-stage decision problem in which a sponsor first optimises the sample size for a phase II trial and then the decision of whether or not to invest in a pair of phase III trials for market authorisation.

When planning the phase II trial, the sponsor is to select a sample size n_2 for the total number of patients to be included in a parallel group trial with equal randomisation (giving $n_2/2$ patients per group). The trial results in an estimate

$$\bar{X} \mid \theta_2 \sim \mathcal{N}\left(\theta_2, \frac{4\sigma_2^2}{n_2}\right),$$

where θ_2 is the unknown population mean of incremental efficacy for patients recruited to phase II and σ_2^2 is the variance for each individual patient response. For simplicity, σ_2^2 is assumed to be known. Before the trial is started, the sponsor's prior distribution for θ_2 is assumed to be

$$\theta_2 \sim \mathcal{N}\left(\mu_1, \frac{4\sigma_2^2}{n_1}\right),$$

implying that the posterior distribution for θ_2 given n_2 and $\bar{X}$ becomes

$$\theta_2 \mid n_2, \bar{X} \sim \mathcal{N}\left(\frac{\mu_1 n_1 + \bar{X} n_2}{n_1 + n_2}, \frac{4\sigma_2^2}{n_1 + n_2}\right).$$

After the phase II trial has concluded, the next step for the sponsor is to decide if it should try to get the new treatment approved for market authorisation by a RA. It is assumed that the RA requires that one-sided statistical significance is shown for efficacy in two independent trials of the same size. It is also assumed that the RA demands a specific size for the phase III trials. For example, this size may be determined by the conventional frequentist requirements on the type I and type II errors, or perhaps by the necessity to gather enough data for a safety evaluation. Hence, for this example, the phase III sample size is not chosen by the sponsor, but instead fixed at a given value n_3. The sponsor only decides on the value of the binary variable d, with $d = 1$ ($d = 0$) corresponding to a GO (NO GO) decision.

Just as for phase II, the two phase III trials are assumed to be parallel-group trials with equal randomisation, each containing $n_3/2$ patients per group. They are assumed to result in sample mean estimates $\bar{Y}_1$ and $\bar{Y}_2$ of the incremental efficacy, which, given the true incremental efficacy mean θ_3 of the target population, are independent and normally distributed. Moreover, reflecting the fact that the efficacy for the patients recruited to the phase II trial may not be the

same as the one for the target population, θ_3 is assumed to be normally distributed with mean θ_2 and variance τ^2. Hence, the distributional assumptions connecting phase II to phase III are

$$\bar{Y}_1 \perp\!\!\!\perp \bar{Y}_2 \mid \theta_3, \quad \bar{Y}_1, \bar{Y}_2 \mid \theta_3 \sim \mathcal{N}\left(\theta_3, \frac{4\sigma_3^2}{n_3}\right), \quad \theta_3 \mid \theta_2 \sim \mathcal{N}\left(\theta_2, \tau^2\right),$$

where σ_3^2 denotes the sample variance per individual phase III response (assumed known).

The utility for the sponsor of choosing n_2, observing $\bar{x}$ in phase II, making the GO or NO GO decision d, and observing $\bar{y_1}$ and $\bar{y_2}$ as outcomes in the phase III trial is defined as

$$u(n_2, \bar{x}, d, \bar{y_1}, \bar{y_2}) = d\left(V\mathbb{I}\left\{\text{RA accepts}\right\} - c_3\right) - c_2 n_2,$$

where V is the market value of introducing the new treatment, c_3 is the total cost of performing the two phase III trials and c_2 is the marginal cost of increasing the sample size of the phase II trial. $\mathbb{I}$ denotes the indicator function for the event given as argument. The expected utility of choosing n_2, observing $\bar{x}$ and choosing d is

$$\bar{u}(n_2, \bar{x}, d) = d\left(V\mathbb{E}\left[\mathbb{I}\left\{\text{RA accepts}\right\} | n_2, \bar{x}\right] - c_3\right) - c_2 n_2. \tag{3.7}$$

By assumption, the RA accepts the new treatment if and only if the null hypothesis of no incremental efficacy can be rejected at the one-sided level α in both of the phase III trials. Since $\bar{Y}_1$ and $\bar{Y}_2$ are assumed to be independent given θ_3, the conditional expectation $\mathbb{E}\left[\mathbb{I}\left\{\text{RA accepts}\right\} | \theta_3\right]$ equals

$$\mathbb{P}\left(\bar{Y}_1 > z_\alpha \sqrt{\frac{4\sigma_3^2}{n_3}} \middle| \theta_3\right) \mathbb{P}\left(\bar{Y}_2 > z_\alpha \sqrt{\frac{4\sigma_3^2}{n_3}} \middle| \theta_3\right) = \Phi\left(\frac{\sqrt{n_3}\theta_3}{2\sigma_3} - z_\alpha\right)^2.$$

By maximising over d and subsequently taking the expectation with respect to the prior predictive distribution for $\bar{X}$, the backward induction is completed and the expected utility for the sponsor of choosing a sample size n_2 and continue optimally may be written as

$$\tilde{u}(n_2) = \mathbb{E}\left[\max\left(0, V\mathbb{E}\left[\Phi\left(\frac{\sqrt{n_3}\theta_3}{2\sigma_3} - z_\alpha\right)^2 \middle| n_2, \bar{X}\right] - c_3\right)\right] - c_2 n_2.$$

In the equation above, note that the inner expectation is computed with respect to the posterior distribution of θ_3 given n_2 and $\bar{X}$,

$$\theta_3 \mid n_2, \bar{X} \sim \mathcal{N}\left(\frac{\mu_1 n_1 + \bar{X} n_2}{n_1 + n_2}, \frac{4\sigma_2^2}{n_1 + n_2} + \tau^2\right),$$

and that the outer expectation is computed with respect to the prior predictive distribution for $\bar{X}$,

$$\bar{X} \sim \mathcal{N}\left(\mu_1, \frac{4\sigma_2^2}{n_1} + \frac{4\sigma_2^2}{n_2}\right).$$

Eq. (3.7) provides the basis for studying the expected utility for the sponsor as a function of the phase II sample size. It is also interesting to consider how the choice of n_2 affects the probability of a GO decision in phase III. From Eq. (3.7), it immediately follows that a GO decision ($d = 1$) will be optimal if and only if the phase II trial results in an estimate $\bar{x}$ such that

$$\mathbb{E}\left[\mathbb{I}\left\{\text{RA accepts}\right\}|n_2, \bar{x}\right] > \frac{c_3}{V}.$$

Hence, the probability of a GO decision for a specific sample size n_2 is

$$\mathbb{P}\left(\mathbb{E}\left[\Phi\left(\frac{\sqrt{n_3}\theta_3}{2\sigma_3} - z_\alpha\right)^2 \middle| n_2, \bar{X}\right] > \frac{c_3}{V}\right).$$

Figure 3.2 shows $\bar{u}(n_2)$ and the probability of a phase III GO decision, computed numerically over a grid for n_2 for a specific combination of parameter values. Two cases are shown, corresponding to a large ($V = 1000$) and a small ($V = 200$) market. Note that the expected utility for a small market size is negative for all sample sizes considered, implying that the optimal decision for the sponsor is to abandon the project in phase II.

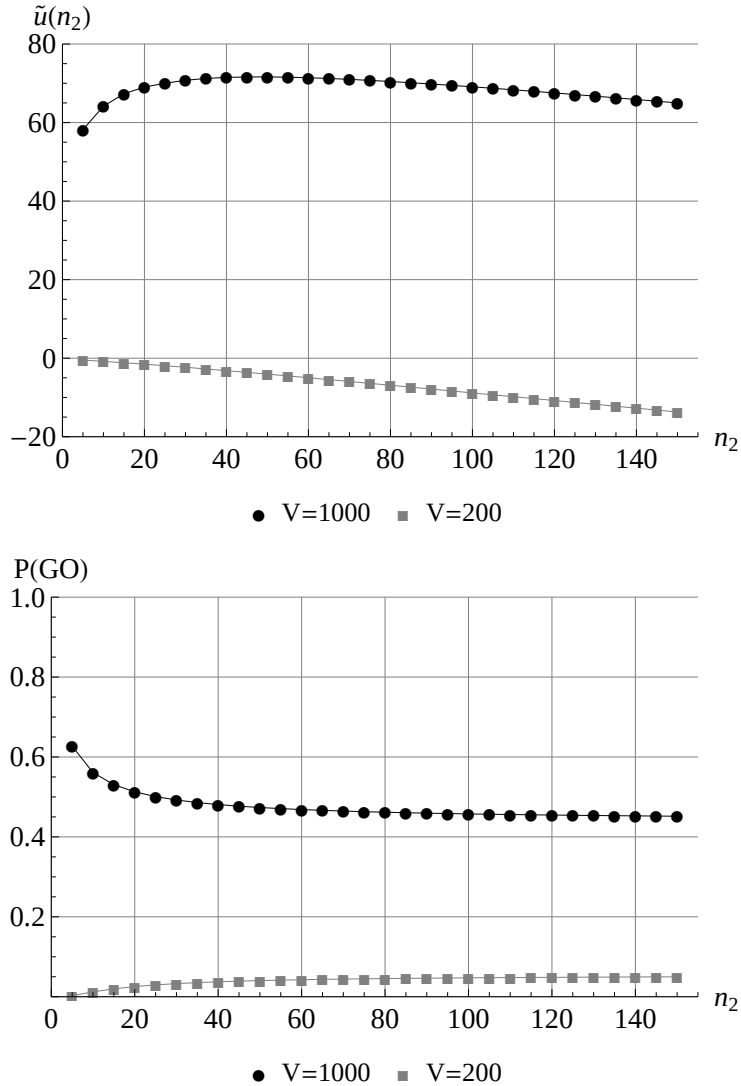

Figure 3.2: Plot of $\tilde{u}(n_2)$ and the probability of a phase III GO decision over the grid $n_2 = \{5, 10, \ldots, 150\}$. Parameter values: $\mu_1 = 0$, $n_1 = 10$, $\sigma_2 = \sqrt{1/10}$, $c_2 = 0.1$, $c_3 = 100$, $n_3 = 1000$, $\sigma_3 = 2$, $\tau = 0.2$, $\alpha = 0.025$.

Chapter 4

Multiple testing procedures

Consider a statistical model in which the probability distribution of an observable (and possibly multi-variate) random variable X is determined by the value of a parameter θ. Assume that θ belongs to some set Θ, commonly referred to as the *parameter space*. In this setting, a subset H of Θ is referred to as a *null hypothesis* about the value of θ, and the complement H^c is called the *alternative hypothesis*. A *test* associated with a null hypothesis is then defined using a binary function, T say, which maps the value X into a value $T(X)$ that determines whether or not the null hypothesis is rejected.

Given a hypothesis $H \subseteq \Theta$ and an associated test T, the *type I error* is defined as the (maximum) probability of falsely rejecting H. Formally,

$$\text{type I error} = \sup_{\theta \in H} \mathbb{P}_\theta \left(T(X) \text{ rejects } H \right).$$

A statistical test is said to control the type I error at the level α if the error is less than or equal to α. When multiple hypotheses are tested in an experiment, the question then arises as to how to properly generalise the concept of a type I error for a single hypothesis.

Let $\{H_1, \ldots, H_m\}$ be a family of $m \geq 1$ hypotheses. For any value of the parameter θ, let $I(\theta) \subseteq \{1, \ldots, m\}$ be the index set corresponding to all hypotheses containing θ, $I(\theta) = \{i \mid \theta \in H_i\}$. Any fixed value of θ defines a subset of true hypotheses, $M_{I(\theta)} = \{H_i \mid i \in I(\theta)\}$, containing $m_{I(\theta)} = |M_{I(\theta)}|$ elements. In this setting of multiple hypotheses, a test function T maps an outcome X of the experiment into a subset of $\{H_1, \ldots, H_m\}$, the elements of which correspond to the individual hypotheses that are rejected by the test.

Having a fixed test procedure T in mind, let R (an observable random variable) be the total number of hypotheses which are rejected by the test and let V (an unobservable random variable) be the number of true hypotheses which

are rejected. There are several alternative definitions that may be used when generalising the type I error rate to the testing of several hypotheses (Bretz et al., 2011, Chapter 2). The *Per Comparison-Error Rate* (PCER) is defined as the expected number of true hypotheses rejected per comparison,

$$\text{PCER} = \mathbb{E}_\theta \left[V \right] / m.$$

The *False Discovery Rate* (FDR) is defined as the expected proportion of falsely rejected hypotheses among the rejected hypotheses (V/R is defined as 0 for $R = 0$),

$$\text{FDR} = \mathbb{E}_\theta \left[V/R \mid R > 0 \right] \mathbb{P}_\theta \left(R > 0 \right).$$

The *Family-Wise Error Rate* (FWER) is defined as the probability that at least one hypothesis is falsely rejected,

$$\text{FWER} = \mathbb{P}_\theta \left(V > 0 \right).$$

In the context of confirmatory clinical trials, a common contemporary regulatory requirement is that the FWER is controlled at a given significance level α (Bretz et al., 2011, p. 13). This is also the generalisation of type I error to multiple hypotheses used in this thesis.

A procedure which ensures control of the error rate in a situation involving the testing of multiple hypotheses is referred to as a *Multiple Testing Procedure* (MTP). A MTP is said to control the FWER in the *weak* sense if the rate is controlled only under the *global null hypothesis*, which is defined as the intersection of all null hypotheses in the family. Hence, there is weak FWER control if

$$\sup_{\theta \in \cap_{i=1}^{m} H_i} \mathbb{P}_\theta \left(V > 0 \right) \leq \alpha.$$

Control of the FWER is said to be *strong* if the type I error is controlled for any configuration of true and false hypotheses, which may be expressed as

$$\sup_{\theta \in \cup_{i=1}^{m} H_i} \mathbb{P}_\theta \left(V > 0 \right) \leq \alpha.$$

In many situations of multiple testing, natural individual tests and corresponding *unadjusted p-values* may be associated with each null hypothesis H_i, $i = 1, \ldots, m$. These tests would then control the type I error rate for a single hypothesis and are often used as a basis for the construction of a MTP. An *adjusted p-value* can then be associated with each individual hypothesis, being defined in such a way that a direct comparison with the overall FWER significance level α is possible.

The closed testing principle (Marcus et al., 1976) can be used to construct a MTP given that tests for rejection have been defined for all possible intersection

hypotheses $H_I = \cap_{i \in I} H_i$, $I \subseteq \{1, \ldots, m\}$. An elementary hypothesis H_i is then rejected by the overall test (i.e., by the MTP) if H_I can be rejected for all I containing i. It can be shown that, if the tests for the intersections control the error rate at level α, then so will the overall MTP.

One of the simplest and most well-known MTP is the Bonferroni procedure, which may serve to illustrate the general philosophy behind such methods. Suppose that individual tests $T_1, \ldots, T_m$ have been defined for the hypotheses, with corresponding unadjusted p-values $p_1, \ldots, p_m$. The Bonferroni procedure is then implemented by rejecting H_i if $p_i \leq \alpha/m$, $1 \leq i \leq m$. That the FWER is controlled in the strong sense at the level α follows directly by the Bonferroni inequality, since $\mathbb{P}_\theta\left(V > 0\right)$ may be written as

$$\mathbb{P}_\theta\left(\cup_{i \in I(\theta)} \{p_i \leq \alpha/m\}\right) \leq \sum_{i \in I(\theta)} \mathbb{P}_\theta\left(p_i \leq \alpha/m\right) \leq m_{I(\theta)} \left(\frac{\alpha}{m}\right) \leq \alpha. \tag{4.1}$$

4.1 The Spiessens-Debois test

Since no distributional assumptions are used in the proof of Eq. (4.1), the Bonferroni MTP is completely general and may be applied in any situation. However, this generality comes at a price. More specific procedures that are only applicable if certain distributional assumptions are placed on the test statistics often lead to higher power. One such procedure, which is employed in paper II, have been proposed by Spiessens and Debois (2010). They refer to it as the *general bivariate normal method*.

We now consider the construction of a MTP controlling the FWER in the strong sense by combining the closed testing principle with the method of Spiessens and Debois. The context is that of a clinical trial with two analyses, namely, an overall analysis of the efficacy in the total population and a subgroup analysis which only considers a subset of the population. The two test statistics corresponding to the overall and subgroup analyses are assumed to follow a joint, bivariate normal distribution. Let Z_1 and Z_2 denote the standardised test statistics used to test the null hypotheses H_1 and H_2 of a zero treatment effect in the total population and the subgroup, respectively. The global null hypothesis that there is no treatment effect in the total population nor in the subgroup is defined as $H_{12} = H_1 \cap H_2$. The test constructed will be one-sided, in the sense that it builds on one-sided tests of H_1 and H_2.

It may be shown (Jennison and Turnbull, 2000) that under H_{12},

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \sqrt{\lambda} \\ \sqrt{\lambda} & 1 \end{bmatrix}\right), \tag{4.2}$$

where $\lambda = I_S/I_T$ is the fraction of information[1] in the subgroup relative to the total population.

Consider now the construction of a test for the rejection of the overall null hypothesis H_{12} which controls the type I error at the level α. If H_{12} is true, then it is unlikely that extreme values will be observed for Z_1 or Z_2. Hence, a reasonable test is one which rejects H_{12} if either $Z_1 > z_{\alpha_1}$ or $Z_2 > z_{\alpha_2}$, where z_{α_1} and z_{α_2} are the critical values corresponding to some significance levels α_1 and α_2. Suppose that the level α_1 is fixed at some specific value. Then in order to control the FWER at an overall significance level α, the other significance level α_2 must be chosen in such a way that

$$\mathbb{P}_\theta \left(Z_1 > z_{\alpha_1} \text{ or } Z_2 > z_{\alpha_2} \right) = \alpha$$

for $\theta \in H_{12}$. Since

$$\begin{aligned} \mathbb{P}_\theta \left(Z_1 > z_{\alpha_1} \text{ or } Z_2 > z_{\alpha_2} \right) &= \mathbb{P}_\theta \left(Z_1 > z_{\alpha_1} \right) + \mathbb{P}_\theta \left(Z_1 \leq z_{\alpha_1} \text{ and } Z_2 > z_{\alpha_2} \right) \\ &= \alpha_1 + \mathbb{P}_\theta \left(Z_1 \leq z_{\alpha_1} \text{ and } Z_2 > z_{\alpha_2} \right), \end{aligned}$$

this is equivalent to finding a value of α_2 satisfying

$$\mathbb{P}_\theta \left(Z_1 \leq z_{\alpha_1} \text{ and } Z_2 > z_{\alpha_2} \right) = \alpha - \alpha_1. \tag{4.3}$$

Using the explicit density in Eq. (4.2), it can be seen that solving Eq. (4.3) is equivalent to solving

$$\int_{-\infty}^{z_{\alpha_1}} \Phi \left(\frac{z_{\alpha_2} - \sqrt{\lambda} z_1}{\sqrt{1-\lambda}} \right) \phi(z_1) \, \mathrm{d}x = 1 - \alpha. \tag{4.4}$$

Given a selected value of α_1, the solution of Eq. (4.4) provides a value of α_2 such that the test which rejects the intersection hypothesis H_{12} if $Z_1 > z_{\alpha_1}$ or $Z_2 > z_{\alpha_2}$ controls the type I error at level α. Clearly, the canonical univariate tests which rejects H_1 if $Z_1 > z_\alpha$ and H_2 if $Z_2 > z_\alpha$ also controls the type I error rate for rejection the of H_1 and H_2, respectively. Hence, by the closed testing principle, the following MTP used in paper II controls the FWER in the strong sense:

1. Specify values of α and α_1.
2. Reject H_1 if $Z_1 > z_\alpha$ (unadjusted rejection of H_1) and $Z_1 > z_{\alpha_1}$ or $Z_2 > z_{\alpha_2}$ (unadjusted rejection of H_{12}).
3. Reject H_2 if $Z_2 > z_\alpha$ (unadjusted rejection of H_2) and $Z_1 > z_{\alpha_1}$ or $Z_2 > z_{\alpha_2}$ (unadjusted rejection of H_{12}).

[1] Information is here defined as the reciprocal of the variance of the respective parameter.

The use of the correlation $\sqrt{\lambda}$ between the statistics Z_1 and Z_2 in the construction of the closed Spiessens-Debois test implies that the power to reject either H_1 or H_2 is strictly greater than the corresponding power for the Bonferroni test. To see this, fix $\alpha > 0$ and let α_1 satisfy $0 < \alpha_1 < \alpha$. The power for the Bonferroni test is

$$\mathbb{P}_\theta \left(Z_1 > z_{\alpha_1} \text{ or } Z_2 > z_{\alpha_2} \right),$$

where $\alpha_2 = \alpha - \alpha_1$. The power for the Spiessens-Debois test is given by the same expression, since the event to reject H_1 or H_2 may be written as

$$Z_1 > z_\alpha \text{ and } (Z_1 > z_{\alpha_1} \text{ or } Z_2 > z_{\alpha_2}) \text{ or } Z_2 > z_\alpha \text{ and } (Z_1 > z_{\alpha_1} \text{ or } Z_2 > z_{\alpha_2}) =$$
$$Z_1 > z_{\alpha_1} \text{ or } (Z_1 > z_\alpha \text{ and } Z_2 > z_{\alpha_2}) \text{ or } Z_2 > z_{\alpha_2} \text{ or } (Z_1 > z_{\alpha_1} \text{ and } Z_2 > z_\alpha) =$$
$$Z_1 > z_{\alpha_1} \text{ or } Z_2 > z_{\alpha_2}.$$

However, for the Spiessens-Debois test, α_2 is defined by Eq. (4.3). Since the event $Z_1 \leq z_{\alpha_1}$ and $Z_2 > z_{\alpha_2}$ is trivially a proper subset of the event $Z_2 > z_{\alpha_2}$, it follows that α_2 must satisfy $\alpha_2 > \alpha - \alpha_1$ if it solves the equation. Hence, the power to reject either of the two hypotheses is strictly greater for the Spiessens-Debois test.

Chapter 5

Some basic tools from analysis

This chapter briefly describes two results from elementary analysis that are applied in paper I.

5.1 The implicit function theorem

In paper I we are faced with the problem of maximising a certain objective function (an expected utility) with respect to a price variable. In addition to the price variable, the objective function also depends on a set of parameters and it is of interest to analyse how the optimal price depends on these parameters. The implicit function theorem (see, for example, Rudin (1976, Chapter 9)) is used in paper I to establish that the optimal price function is continuously differentiable, given that this holds for the objective function that is maximised. Moreover, the theorem provides formulas for computing the partial derivatives of the optimal price function with respect to the parameters. For reference, a version of the theorem will now be stated.

Let f be a continuously differentiable function mapping points $(\mathbf{x}, y) \in \mathbb{R}^{n+1}$ into $\mathbb{R}$ and suppose that the point $(\mathbf{a}, b) = (a_1, \ldots, a_n, b) \in \mathbb{R}^{n+1}$ satisfies $f(\mathbf{a}, b) = 0$. Then, if $(\partial f/\partial y)(\mathbf{a}, b) \neq 0$, there exists an open set U containing $\mathbf{a}$, an open set V containing b, and a unique continuously differentiable function $g : U \to V$ such that the local graph of g, $\{(\mathbf{x}, g(\mathbf{x})) \mid \mathbf{x} \in U\}$, coincides with the local level set $\{(\mathbf{x}, y) \in U \times V \mid f(\mathbf{x}, y) = 0\}$. Moreover, the partial derivative of g with respect to the component x_i in the point $\mathbf{a}$ is given

by

$$\frac{\partial g}{\partial x_i}(\mathbf{a}) = -\frac{\partial f}{\partial x_i}(\mathbf{a}, b) \Big/ \frac{\partial f}{\partial y}(\mathbf{a}, b).$$

5.2 The envelope theorem

A general result that is often useful when analysing how changes to parameters influence the optimised value of an objective function is the envelope theorem (Varian, 1992, Chapter 27). There are many versions of this theorem, but the result will only be stated here under rather strong smoothness assumptions. The theorem is used in paper I to analyse how the optimal expected utility responds to changes in various parameters.

Suppose that f is a differentiable function mapping points $(\mathbf{x}, y) \in \mathbb{R}^{n+1}$ into $\mathbb{R}$, and consider the maximisation problem

$$f^*(\mathbf{x}) \equiv \max_y f(\mathbf{x}, y).$$

Assume further that the maximising argument $y^*(\mathbf{x})$ is differentiable for $\mathbf{x}$ in some region U of interest and that $(\mathbf{x}, y^*(\mathbf{x}))$ corresponds to an interior global optimum for $\mathbf{x} \in U$. The envelope theorem then states that

$$\frac{\partial}{\partial x_i}\left(f^*(\mathbf{x})\right) = \frac{\partial f}{\partial x_i}(\mathbf{x}, y)\Big|_{y=y^*(\mathbf{x})}. \tag{5.1}$$

Hence, changes in the optimal value of the objective function may be analysed in terms of partial derivatives of the original objective function. With the stated assumptions, this result is easily shown. Since $f^*(\mathbf{x}) = f(\mathbf{x}, y^*(\mathbf{x}))$, the left hand side of Eq. (5.1) is

$$\frac{\partial}{\partial x_i}\left(f(\mathbf{x}, y^*(\mathbf{x}))\right) = \frac{\partial f}{\partial x_i}(\mathbf{x}, y^*(\mathbf{x})) + \frac{\partial f}{\partial y}(\mathbf{x}, y^*(\mathbf{x}))\frac{\partial y^*}{\partial x_i}(\mathbf{x}).$$

But from the assumptions of differentiability and an interior optimum, $(\partial y^*/\partial x_i)(\mathbf{x}) = 0$ and the result follows.

Chapter 6

Summary of papers

6.1 Paper I: Late-Stage Pharmaceutical R&D and Pricing Policies under Two-Stage Regulation

Paper I presents a BDT framework for investigating R&D incentives for the pharmaceutical industry in the presence of two exogenous regulatory stages. In Stage 0, a commercial sponsor deliberates on whether to run a phase III trial and, if it decides to go ahead, selects the sample size of the trial. The trial results in an estimate of the incremental effectiveness, which is denoted by x. Upon trial completion, a RA in charge of granting access to a market considers the evidence produced by the trial. Approval for marketing is granted if the sample size is large enough and the new treatment shows superiority to the current standard alternative at a one-sided level of statistical significance. In Stage 1, a price is proposed by the sponsor for the new treatment. When this price is combined with the effectiveness estimate provided by the trial, it determines the Incremental Cost-Effectiveness Ratio (ICER) upon which a HCI bases its reimbursement decision. The sponsor's optimal Stage 1 pricing policy depends on the estimate x, which is a random variable from the perspective of Stage 0. The optimal policy for the sponsor over both stages is found using backward induction.

From the perspective of the sponsor, the value of the HCI's Maximum Willingness to Pay (WTP) for a unit increase in effectiveness is uncertain and is modelled using a continuous random variable W. It is assumed that W belongs to a location-scale family of random variables, implying that any member can be uniquely characterised in terms of a pair (m, s), where m is the expected value

(or location parameter) of W and the scale, s, can be considered a measure of how uncertain the sponsor is about the HCI's WTP. We identify three ranges for the uncertainty parameter s, in which increases in uncertainty have different effects. In the 'low uncertainty' range, increases in s result in lower optimal prices, lower optimal expected profits and a smaller optimal trial sample size. In the 'high uncertainty' range, the situation is reversed: greater uncertainty leads to higher prices, higher expected profits and a larger trial sample size. For 'intermediate uncertainty', prices are increasing, expected profits decreasing and sample size decreasing in the degree of uncertainty. Hence, for the range of intermediate uncertainty, a smaller value for s benefits the sponsor, the HCI and the patients.

The framework is applied to a recent NICE appraisal of mannitol dry powder for treating cystic fibrosis. The status of cystic fibrosis as a rare disease means that the R&D decision could potentially be considered to be a marginal project, that is, one with a market size that is close to the minimum population size required for the investment to be deemed profitable. Within the context of this application, we briefly consider how the RA parameters defining the one-sided significance level and the minimum sample size required for marketing authorisation impact the minimum size of the target population that the sponsor requires in order to expect a positive profit when acting optimally.

6.2 Paper II: Optimising Trial Designs for Targeted Therapies

Paper II is concerned with pivotal clinical trials in which the efficacy of a treatment is tested in an overall population and/or in a pre-specified subpopulation defined by a binary biomarker. A BDT framework is used to derive optimised trial designs by maximising utility functions. The optimisation is done from two perspectives: from the viewpoint of a commercial sponsor and from the viewpoint of a public health decision maker.

For both perspectives, three different types of trial designs are considered. These are referred to as as the *classical design*, the *stratified design* and the *enrichment design*. The classical design makes no use of the biomarker status and only tests for a treatment effect in the full population. This is done using a standard, parallel-group trial with equal group sizes. The stratified design also recruits patients from the full population, but the biomarker status of each patient is determined and the treatment effect is tested in the full population and in the subpopulation. This implies that the stratified design may lead to approval in either the full population or in the subpopulation only, which necessitates an appropriate control of the FWER. Such control is implemented

using the closed Spiessens-Debois test. In the enrichment design, patients are screened for their biomarker status and only biomarker positive patients are included in the trial. The use of the biomarker test in the stratified and enrichment designs implies that a fixed cost must be paid to develop the screening procedure. Moreover, a marginal cost must be paid for each patient screened. These biomarker costs are not present for the classical design. By comparing the optimal expected utilities for these three design types, the framework allows us to assess when it is favourable to determine the biomarker status of the patients in a clinical trial and when it is actually more efficient to disregard the biomarker and to proceed with a classical trial design.

The sample size is optimised for each of the three design types. Moreover, for the stratified design, the two significance levels defining the Spiessens-Debois test are also optimised. This optimisation is done with respect to a prior that encodes the pre-trial knowledge about the efficacy of the treatment by means of a two dimensional distribution on the true effect sizes in the full population and the subpopulation. The considered utility functions account for the different costs of the design types as well as the expected benefit when demonstrating efficacy in the different subpopulations.

Examples of trial designs obtained by numerical optimisation are presented for both perspectives. We find that the optimal type of design depends sensitively on the various parameters of the framework. A parameter of particular interest is the prevalence of the biomarker positive patients in the total target population, and we consider the impact of this parameter in detail.

Chapter 7

Discussion

In both of the papers that constitute this thesis, the objective is to characterise the optimal trial design for a Bayesian decision maker in a fixed regulatory environment. Although simplifying assumptions were made, the models used for market approval and reimbursement were defined so as to correspond to current standard practice. Within such a setting, it is possible to draw some conclusions about the impact of changing the parameters defining the regulatory structure. For example, in paper I, it was possible to study how the minimum market size required for a sponsor's GO decision depends on the value of the one-sided significance level used by the RA.

Given that the optimal behaviour of the sponsor can be computed for a fixed regulatory environment, it is natural to go one step further and ask if it is possible to change the rules for market approval and reimbursement so as to improve the expected utility for the patient population. There has been much recent work in the health economics literature investigating the performance of alternative regulatory schemes. Babar (2015, Chapter 21) provides an introduction to the ideas behind some different pricing policies and reviews some of the recent contributions to the literature. It is noted there that the main objective of any such policy is to provide a good trade-off between the value for money for the payer (static efficiency) and the need to provide enough incentives for R&D investments by the pharmaceutical industry (dynamic efficiency). An interesting extension of the work in paper I would be to specify a model for the connection between the sponsor's level of appropriation of the total welfare resulting from adopting a new treatment (i.e., its profit) and its willingness and capability to invest in future innovation efforts. Provided that a social welfare function is also specified, such an expanded model could then be used as a basis for the search of an optimal set of rules for market approval and reimbursement within some space of reasonable regulatory schemes.

Some work in this direction is described by Jena and Philipson (2008), which analyse the use of cost-effectiveness methods by payers to decide on reimbursement for a new treatment. They mention the ICER threshold employed by NICE in the U.K. as one example and note that such schemes are closely related to other forms of supply side price regulation (e.g., price-controls and rate-of-return regulations). They argue that such methods are at least implicitly concerned with maximising the surplus available to consumers upon acceptance of the new treatment at the cost of a reduced profit for the pharmaceutical company responsible for the innovation. They also stress that their major point is not that existing incentives are either too high or too low, but that any policy based on cost-effectiveness for treatment adoption and reimbursement do have implications for dynamic welfare which should be considered when evaluating it. Their emphasis lies on the pricing part of the regulation. Hence, it seems worthwhile to expand their analysis by also including an explicit model for the clinical trial stage, as done in paper I.

Shifting focus from pricing schemes to the rules determining under what conditions reimbursement takes place, Levaggi et al. (2016) compare the performance of regulatory acceptance of a new treatment based on standard ICER threshold comparisons versus that of performance-based risk-sharing agreements. The idea behind the latter mechanism is that the firm will not be paid in full if the observed effectiveness of the treatment falls below a certain level after market introduction. Relative to the work by Jena and Philipson (2008), there is not only focus on policy performance with respect to static and dynamic efficiency, but also an effort to determine which of the alternative regulatory mechanisms that is to be preferred. Although they do not specify a social welfare function and proceed with optimisation of regulatory parameters, their work may possibly be extended in this direction. Their model involves three stages for the firm's product: (1) discovery, (2) development, and (3) commercialisation. During the discovery stage, which corresponds to scientific efforts prior to clinical trials, the effectiveness of the drug is assumed to evolve as a geometric Brownian motion. The development stage, comprising phase I, II and III trials, is modelled in terms of a single probability of failure. However, it would be interesting to investigate to what extent an expanded policy specification could be handled that also contains parameters such as, for example, the minimum required sample size and a threshold for the type I error.

The contributions discussed above are primarily concerned with analysing the relationship between payer reimbursement and the incentives for the industry to invest in pharmaceutical R&D. Le Deley et al. (2012) instead asks the question: what are the appropriate values for the regulatory parameters pertaining to evaluation of the statistical evidence, such as the type I error and sample size, if the goal is to maximise the cumulative therapeutic benefit over a

time horizon comprising several successive new candidates for approval? Working within the context of cancer trials, they use a simulation study to reach the conclusion that expected survival benefits over a 15-year time horizon are maximised when individual trial sample size is smaller and type I error rates larger than the conventional values. Their simulation approach is appealing, and it would be interesting to investigate the effects of extending their model with a reimbursement stage after each trial.

It should also be pointed out that a simple model of a market situation consisting of two players, the profit-maximising firm performing the research and the trial on the one hand, and the regulator deciding on market approval and reimbursement on the other, can be placed within the principal-agent framework of game-theoretic mechanism design (see, for example, Baron (1989, Chapter 24) or Fudenberg and Tirole (1991, Chapter 7)). In this framework, the regulator (i.e., the principal) first chooses a mechanism, which consists of the specification of a message space and a function mapping every possible message into a collection of values for the regulatory parameters and the actions for the agent. This mechanism is presented to the agent (i.e., the firm), which chooses a message from those available so as to maximise its expected profits. The main complicating factor facing the principal in this setting is that the agent may be better informed about certain parameters of the decision process that will determine the overall social welfare. In other words, the agent has private information about some parameters, θ say, while the principal can only describe its knowledge about θ in terms of a probability distribution. Typical examples in the economics literature of what θ might represent are, for example, different types of costs, such as the marginal production cost or the cost of performing a clinical trial. In the special context of regulation of pharmaceutical firms, another piece of private information of interest for the principal would be the beliefs held by the agent about the efficacy of a new treatment prior to confirmatory clinical trials. By the revelation principle, the only message set that the principal needs to consider is the one that consists of all possible values of θ, and it may without loss of generality restrict the optimisation to mechanisms for which the agent truthfully reports the value of θ.

Although the specification of a utility function followed by a full BDT analysis has been promoted by several authors (see, e.g., Lindley (1997), Claxton (1999) or Berry (2006)), scepticism regarding the practicality of the approach can also be found in the literature. Some important points are given by Armitage (1985). He notes that it is an over-simplification to treat the result of a phase III trial only as a trigger for a subsequent treatment decision. The trial data also has a value in itself, since it adds to the total body of scientific knowledge. Another point is connected to the ethics of a clinical trials. In typical patient horizon models, health benefits are valued equally for both recruited

and post-trial patients. Although acceptable from a utilitarian viewpoint, such a valuation is not consistent with traditional medical ethics, which dictates that the interests of the current patient is paramount. He further argues that procedures for discounting the benefits of future patients introduce a degree of arbitrariness into the model and therefore makes it less persuasive. Moreover, it is difficult to find a value for the patient horizon itself, since this involves estimating the time until a new and better treatment option arrives. The issues put forth by Armitage are all important, but they hardly provide enough reasons for abandoning the BDT approach. Rather, they call for a more precise definition of the utility function, able to capture all gains and costs that the decision maker deems relevant, together with an appropriately specified probabilistic model for the uncertainty surrounding the true value of the patient horizon.

7.1 Conclusion

The work presented in this thesis can be summarised as being an application of the BDT framework to two specific problems of clinical trial optimisation. In paper I, the decision problem has two stages. A choice of sample size for a confirmatory trial is followed by a price choice for the new treatment if there is RA approval for market authorisation. The result of main interest is that the optimal strategy of the sponsor depends on how uncertain it is about the willingness of the HCI to provide reimbursement. In particular, the model implies that reducing such uncertainty will in some situations lead to both higher expected profits for the sponsor and lower proposed prices (hence, a smaller budget impact for the HCI). By keeping the model structure relatively simple, we were able to obtain many results regarding the impact of parameter values on the optimal strategy using mathematical analysis. From the viewpoint of a commercial sponsor, an extended model incorporating more development stages (phase I and II trials) and a more realistic specification of the regulatory rules used by the RA and the HCI would certainly be of interest. Analytic results may be hard to derive in such a generalised model, but it will still be possible to evaluate it using numerical computations. The RA and HCI rules were taken as exogenous and were not optimised. As indicated in the discussion, there exists numerous approaches to the problem of finding a regulatory environment that is optimal from a societal perspective given that drug development is done by rational, economic agents.

I paper II, the decision problem consists of a single stage. A sponsor, commercial or otherwise, selects an optimal trial design from a class of alternatives so as to maximise its expected utility. The class of designs is partitioned into classical, stratified and enrichment designs, where the latter two incorporates a

biomarker test. The main conclusion reached was that the optimal design depends heavily on the perspective of the sponsor (commercial or public health), the prior distribution used for the effect sizes and the prevalence of the subgroup of biomarker positive patients. In particular, it was found that there are situations in which it is optimal to disregard the biomarker test and go for a classical design. A natural exstension of the work is to enlarge the class of trial designs so as to also encompass adaptive designs with interim decisions. Adding too many stages will make accurate computations using the backward induction algorithm prohibitively expensive, but a single interim analysis giving a two stage design seems tractable.

Bibliography

Armitage, P. (1985, April). The search for optimality in clinical trials. *International Statistical Review 53*(1), 15–24.

Babar, Z.-U.-D. (Ed.) (2015). *Pharmaceutical Prices in the 21st Century* (First ed.). Switzerland: Springer International Publishing.

Baron, D. P. (1989). *Handbook of Industrial Organization* (First ed.), Volume 2. The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK: Elsevier Ltd.

Bernardo, J. M. and A. F. Smith (1994). *Bayesian Theory* (First ed.). 111 River Street, Hoboken, NJ 07030, USA: John Wiley & Sons.

Berry, D. A. (2006, January). Bayesian clinical trials. *Nature Reviews Drug Discovery 5*, 27–36.

Berry, S. M., B. P. Carlin, J. J. Lee, and P. Müller (2011). *Bayesian Adaptive Methods for Clinical Trials* (First ed.). 6000 Broken Sound Parkway NW, Suite 300,Boca Raton, FL 33487-2742: CRC Press, Taylor & Francis Group.

Biomarkers Definitions Working Group (2001, March). Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics 69*(3), 89–95.

Bretz, F., T. Hothorn, and P. Westfall (2011). *Multiple Comparisons Using R* (First ed.). 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742: Taylor & Francis Group.

Chow, S.-C. and J.-P. Liu (2014). *Design and Analysis of Clinical Trials* (Third ed.). Hoboken, New Jersey: John Wiley & Sons.

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

DeGroot, M. H. (1970). *Optimal Statistical Decisions* (First ed.). Hoboken, New Jersey: John Wiley & Sons.

DiMasi, J. A., H. G. Grabowski, and R. W. Hansen (2016, May). Innovation in the pharmaceutical industry: New estimates of r&d costs. *Journal of Health Economics 47*, 20–33.

EMA (2016, April). History of ema. Available at `http://www.ema.europa.eu/ema/index.jsp?curl=pages/about_us/general/general_content_000628.jsp&mid=WC0b01ac058087addd`.

Fisher, R. A. (1946). *Statistical Methods for Research Workers* (Tenth ed.). Edinburgh: Oliver and Boyd.

Fisher, R. A. (1956). *Statistical Methods and Scientific Inference*. Edinburgh: Oliver and Boyd.

Fudenberg, D. and J. Tirole (1991). *Game Theory* (First ed.). Cambridge, Massachusetts: MIT Press.

ICH (2016, April). Statistical principles for clinical trials (e9). Available at `http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf`.

Jena, A. B. and T. J. Philipson (2008, September). Cost-effectiveness analysis and innovation. *Journal of Health Economics 27*(5), 1224–1236.

Jennison, C. and B. W. Turnbull (2000). *Group Sequential Methods with Applications to Clinical Trials* (First ed.). 2000 N.W. Corporate Blvd., Boca Raton, Florida 33431: Chapman & Hall/CRC.

Kaitin, K. I. and J. A. DiMasi (2011). Pharmaceutical innovation in the 21st century: New drug approvals in the first decade, 2000-2009. *Clinical Pharmacology & Therapeutics 89*(2), 183–188.

Le Deley, M.-C., K. V. Ballman, J. Marandet, and D. Sargent (2012, June). Taking the long view: how to design a series of phase iii trials to maximize cumulative therapeutic benefit. *Clinical Trials 9*(3), 283–292.

Levaggi, R., M. Moretto, and P. Pertile (2016). The dynamics of pharmaceutical regulation and r&d investments. *Journal of Public Economic Theory*. (to appear).

Lindley, D. V. (1997, July). The choice of sample size. *The Statistician 46*(2), 129–138.

Marcus, R., E. Peritz, and K. R. Gabriel (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika 63*(3), 655–660.

Ondra, T., A. Dmitrienko, T. Friede, A. Graf, F. Miller, N. Stallard, and M. Posch (2016). Methods for identification and confirmation of targeted subgroups in clinical trials: A systematic review. *Journal of Biopharmaceutical Statistics 26*(1), 99–119.

Raiffa, H. and R. Schlaifer (1961). *Applied Statistical Decision Theory* (First ed.). Boston: President and Fellows of Harvard College.

Rudin, W. (1976). *Principles of mathematical analysis* (Third ed.). New York: McGraw-Hill.

Savage, L. J. (1972). *The Foundations of Statistics* (Second ed.). 180 Varick Street, New York, N.Y. 10014: Dover Publications.

Senn, S. (2007). *Statistical Issues in Drug Development* (Second ed.). The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England: John Wiley & Sons.

Speigelhalter, D. J., K. R. Abrams, and J. P. Myles (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation* (First ed.). The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England: John Wiley & Sons.

Spiessens, B. and M. Debois (2010, November). Adjusted significance levels for subgroup analyses in clinical trials. *Contemporary Clinical Trials 31*(6), 647–656.

Stallard, N., F. Miller, S. Day, S. W. Hee, J. Madan, S. Zohar, and M. Posch (2016). Determination of the optimal sample size for a clinical trial accounting for the population size. *Biometrical Journal*. (to appear).

Varian, H. R. (1992). *Microeconomic Analysis* (Third ed.). 500 Fifth Avenue, New York, N.Y. 10110: W. W. Norton & Company.

Ziegler, A., A. Koch, K. Krockenberger, and A. Grohennig (2012, October). Personalized medicine using dna biomarkers: a review. *Human Genetics 131*(10), 1627–1638.