Bayesian Risk Optimisation in Clinical Portfolios

Alkiviadis Triantafyllidis

MSc in Data Science & Statistics

University of Bath

September 2024

Bayesian Risk Optimisation in Clinical Portfolios

submitted by

Alkiviadis Triantafyllidis

for the degree of MSc in Data Science & Statistics of the

University of Bath

September 2024

Declaration

This dissertation is submitted to the University of Bath in accordance with the requirements of the degree of MSc in Data Science & Statistics in the Department of Computer Science. No portion of the work in this thesis has been submitted in support of an application for any other degree or qualification of this or any other university or institution of learning. Except where specifically acknowledged, it is the work of the author.

Generative AI statement

Generative AI was used for the means of checking grammatical and syntactical structure of the document. At all times tools were asked to make recommendations and not unsupervised amendments to the text. A brief and concise conclusion was developed and checked using such tools to ensure that key takeaways were included and not omitted by mistake due to familiarity gained from the extensive amount of time spent developing the document.

Signature of Author.....

Abstract

This thesis explores a risk-driven approach to optimising clinical trial portfolios through Bayesian optimization. It addresses the challenges faced by pharmaceutical companies in managing uncertainty during drug development, particularly in clinical trials, which are resource-intensive and prone to failure. By adopting a Bayesian framework, this work enables dynamic updates to trial designs as new data becomes available, allowing for early decision-making and more efficient resource allocation. The thesis integrates advanced statistical methods and Mixed Integer Linear Programming (MILP) to compute optimal strategies, minimising recruited patients and costs while maximising the overall utility of the portfolio. Through simulated data, the research demonstrates how these methods can guide better decision-making in clinical programme portfolios, balancing risk and reward for pharmaceutical sponsors.

This thesis guides the reader through a carefully formulated sequence of examples to naturally bring together more complex concepts. Use of non-sensitive simulated data provides the advantage of developing thought-provoking examples to effectively convey fundamentals and motivation for the thesis. We contribute to the research area by delivering abstract metrics such as programme utility and portfolio yield intuitively.

Contents

1 Introduction

Drug development programs are traditionally long and expensive, often requiring a decade or more to progress from initial discovery to the point where a drug is brought to market (DiMasi, Grabowski and Hansen, 2016). Pharmaceutical companies face lengthy development timelines but also significant uncertainty at every stage. The greatest source of uncertainty lies in the outcomes of clinical trials, as it is difficult to predict whether a treatment will perform as anticipated. Even if a treatment is approved, there is still uncertainty regarding its market performance, influenced by external factors such as therapeutic area, competition or demand (Antonijevic, 2014).

The Blockbuster drug era refers to the period between 1995–2015, when Big Pharma companies would make minor adjustments to treatments used in treating diseases affecting huge populations in order to extend their patent life. Blockbuster drugs reach at least $1 Billion in sales, which in 2005 accounted for 60% of the $245 billion in sales of the 10 leading pharmaceutical companies (Malik, 2007). By targeting diseases such as diabetes or depression, pharmaceutical companies with extensive resources had vast yields. Brute resource allocation for these companies was essentially invested without significant risk. However, post 2015, there has been a demographic shift in the research landscape. Companies acknowledged that certain product categories are saturated and the possibility of developing billion dollar treatments within them is diminishing (Collier, 2011). Jørgensen (2008) characterised this as an unrealised opportunity of targeting subpopulations with broad categories, as these treatments were created using a suboptimal generalised effect and did not consider diverse patient biologics. Additionally, diseases such as cancer and Alzheimer's are not fully understood and breakthroughs are yet to be discovered.

Efficacy refers to the evidence required to establish a significant new treatment (O'Hagan, Stevens and Campbell, 2005). Jørgensen (2008) claims that the move away from Blockbuster drugs is a product of their low efficacy rates. Governments worldwide increasingly aimed to reduce healthcare costs and restrict funding for medical treatments that provide minimal improvement over existing therapies. In 2012, major perscription drug insurer Pharmacy Benefit Managers (PBMs), "began refusing coverage for many newly approved drugs when cheaper alternatives were available" (Agha, Kim and Li, 2022). As a result, between 2007 and 2011, 25% of US drug sales lost patent protection. Of those, Blockbuster drugs accounted for 90% due to competition from cheaper innovative treatments (Malik, 2007). The Food & Drug Administration (FDA) now requires comprehensive evidence to approve marketing of a treatment (Food and Drug Administration and others., 1998). This increased competition has called for efficient trial designs. Sponsors now recognize the opportunity in increasing the efficacy of treatments by developing treatments for smaller groups (Jørgensen, 2008), shifting from "bigger is better" models to lean, focused operations. By incorporating modeling and simulation into the drug development process, targeted or "stratified" treatments can be identified efficiently, reducing trial and error and improving efficacy rates (Gautam and Pan, 2016).

Drug development programs follow a series of stages that must be completed before a treatment can be brought to market. Clinical trials are one of these stages in the programme and consist of a sequence of confirmatory trials monitored by a regulatory body such as the European Medicines Agency (EMA). The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) is the European organisation that maintains a variety of guidelines for the pharmaceutical industry. Topic E9 in the handbook consists of statistical principles suggesting best practices in conducting clinical trials (ICH Steering Committee and others., 1998). In this work we adapt the requirements for confirmatory trials submitted to the EMA, and follow the handbook guidelines to design them. In modeling the treatment effect, we adapt a Bayesian framework which offers the advantage of updating our distributional model as data becomes available. The FDA acknowledges the use of a Bayesian framework in clinical trial design and has established specific guidelines for its implementation (Guidance, 2010). Implementing Bayesian designs in multiple decision problems allows for early stopping in programmes. Both stopping a trial due to sufficient evidence that a treatment is effective or because of negative outcomes in which case it would be futile to continue, minimizes the number of patients required. This not only yields a cost-effective solution to the sponsor, but has also certain ethical benefits. The case of Pallmann et al. (2017), is an example of a trial in emergency medicine that could only be made feasible with an innovative Bayesian design. A comprehensive resource we use to implement such designs includes the work of Jennison and Turnbull (1999).

Clinical trials present high failure rates of up to 60% in the years 2011-2012 (Thomas, 2016). Inherently, pharmaceutical companies with a restricted budget may prefer to spread risk across multiple programmes. This results in a sponsor with individual priorities, managing several programmes simultaneously. A challenge in developing a portfolio is that decisions made in one program often affect others, as limited resources must be distributed among them. As a result, sponsors must carefully choose which programmes to continue and which to abandon. The utility of a programme largely depends on the uncertainty of the treatment effect. Under some distributional assumptions we are able to define the expected yield of a clinical programme. The standard metric used in industry is the eNPV (Wiklund, Thorn, Götte, Hacquoil, Saint-Hilary and Carlton, 2024) which incorporates development costs, market value and other parameters to return a comparable metric between programmes. This quantity differs from a traditional NPV as it is an expectation. Accounting for the underlying stochasticity of clinical programmes, this becomes a fitting measure for such applications. Thereafter, computing the optimal portfolio in different contexts is achieved using a flexible Mixed Integer Linear Programming (MILP) algorithm with a choice of constraints.

We acknowledge that this is a broad reseach area where scientists have produced extensive research and sophisticated quantitative methods on. The originality of this work is to guide the reader through a carefully formulated sequence of examples to naturally bring together more complex methods. In this project, we work with simulated data that

would otherwise consist of sensitive patient information. An advantage of doing so is the flexibility of developing thought-provoking examples to effectively convey fundamentals and motivation for the thesis. Thus, we contribute to the research area by delivering abstract metrics such as programme utility and portfolio yield intuitively and with clear applicability.

This thesis is structured into several key sections to guide the reader through the research process. Section 2 begins with an introduction to the fundamentals of clinical trials and the role of uncertainty in drug development. In this section we define core programme specific parameters and introduce how hypothesis testing is used by regulating bodies and companies to design trials. To maintain a manageable workload we also introduce how to compute key quantities through simulation when their analytical solution is otherwise complicated. In Section 3, we formalize what we mean by sequential decisions and group sequential designs. These concepts have great significance both on the trajectory of a single programme but also on a portfolio level. In Section 4, we introduce two integral estimation methods and discuss their use and differences. In Section 5, we define the utility function required to evaluate different programmes. We discuss why this is an expectation as well as some major influencing parameters. Section 6 consists of details on the Bayesian optimization framework and optimal risk averse designs. Using a Bayesian paradigm, we are able to update optimisation methods as information becomes available. This motivates the re-assessment of whether a sponsor wishes to continue developing a treatment. Having covered the moving parts of a programme, in Section 7 we combine the previous sections to introduce portfolio optimisation stragies to select an optimal set of programmes. In Section 8, we simulate interesting scenarios to apply the methodology introduced. We evaluate different strategies and discuss implications of different sponsor priorities.

2 Clinical trials

2.1 Trial phases

Fo a treatment to get approved by regulating bodies, ir must be tested in different ways known as phases. Each phase has individual objectives and can be summarised by different questions posed towards the candidate treatment.

Phase I - Is the drug safe? A small number of human patients are recruited to evaluate whether the drug is safe to proceed into phases with larger sample sizes.

Phase II - Does the drug have a significant effect over the current standard? This phase recruits a moderate number of patients to determine what dose produces a significant effect. Although results in this phase are not submitted to a regulatory body, the sponsor must be confident to some level that the candidate treatment has a significant effect over the current standard.

Phase III - Is the new treatment substantially more effective than the current standard? This phase typically recruits the largest number of patients to provide substantial probabilistic evidence that the candidate treatment outperforms the current standard. The results of this phase are submitted to the regulatory body for approval. This is known as a confirmatory trial and consists of a frequentist hypothesis test. How confident the results are is at the discretion of the sponsor and ultimately determines the value of a treatment when it reaches the market.

Phase IV - Does the approved drug continue to be safe and effective? This phase does not have any associated recruiting but monitors patients that have been assigned treatments using the marketed drug.

The success rate and reason for failure in Phases I-III through the years 2011-2012 are reported in Table 1. The most common reason for failure in Phases II & III is efficacy, demonstrating a statistically significant effect. A better understanding of which programmes fail the hypothesis test is both controllable and a primary focus of this work, especially when there are finite resources to be distributed. Safety in clinical trials consists primarily of the Data Monitoring Committees who receive a set of unblinded data on which they assess concurrently (CHMP, 2004). If at any point there are safety violations, the trial ceases immediately. Although slightly different, we discuss ethical considerations and strategies used for the early stopping of programmes that arise due to harmful results in Section 3.2. Implementing group sequential designs inherently complement such considerations and end harmful trials early.

		Reason for failure	
Phase	Success Rate	Safety	Efficacy
Phase I	63.2%	–	–
Phase II	30.7%	22%	59%
Phase III	58.1%	35%	52%

Table 1: Success rate per phase (Thomas, 2016) and reason for failure in respective phase (Arrowsmith and Miller, 2013). Years including 2011 and 2012.

In the available time of this work we limit the scope to Phases II & III in which a sponsor is most flexible to apply established methods to maximize the return of their investment. Programmes in general have variable lengths, with some lasting multiple years (DiMasi, Grabowski and Hansen, 2016). For our purpose, we do not take into consideration phase lengths and suppose that decisions are made pre Phase II or III and both phases last an equal amount of time. The sponsor is then free to make programme and design choices in a series of sequential decisions, detailed in Section 3. Different programmes in this work are denoted by P. Phases are made up of confirmatory trials and are denoted by t. A given programme can consist of multiple trial variants in the optimisation process.

2.2 Treatment effects

The objective of a clinical trial is to establish whether the candidate treatment delivers a substantial effect over the current standard. We define the current standard as the control treatment, and the candidate as the experimental treatment. A total of n patients is chosen according to their suitability in benefitting from these treatments. We let n_x and n_y represent the number of subjects assigned to the experimental and control group respectively, via the ratio n_x/n_y. The randomization process is a crucial component in clinical trials, however we assume that assignment is sufficiently randomised for the purpose of this work. Such methods are studied by established scientists in industry with examples including but not limited to the work of Tygstrup, Lachin and Juhl (1982) and Lachin, Matts and Wei (1988).

A core assumption of the basis of this work are the effects X and Y, representing the experimental and control treatment observations, derived respectively from the following distributions $X_i \sim N(\mu_x,\ \sigma_x^2)$ and $Y_i \sim N(\mu_y,\ \sigma_y^2)$. The obervations of the overall treatment effect can then be written $\theta_i \sim N(\mu_x - \mu_y,\ \sigma_x^2 + \sigma_y^2)$. Since their true values are unknown, in practice we must work with an estimate of the treatment effect, $\hat{\theta}$. To achieve this, the candidate and control treatments are administered to subjects and the effects are recorded in $X \in \{X_1, \ldots, X_{n_x}\}$ and $Y \in \{Y_1, \ldots, Y_{n_y}\}$. The treatment effect is estimated by $\hat{\theta} = \bar{X} - \bar{Y}$, where $\bar{X} = \frac{1}{n_x}\sum_{i=1}^{n_x} X_i$ and $\bar{Y} = \frac{1}{n_y}\sum_{i=1}^{n_y} Y_i$. Assuming a balanced design $n_x = n_y = n$, and observations for both treatments to have constant and equal variance $\sigma_x = \sigma_y = \sigma$, the sampling distribution for the overall treatment effect is,

$$\hat{\theta} \sim N\left(\theta,\ \frac{4\sigma^2}{n}\right).$$

Toy example

Let t_1 denote a confirmatory trial satisfying the assumptions above, where the candidate and control treatments are $X_i \sim N(30,\ 45^2)$ and $Y_i \sim N(0,\ 45^2)$ respectively. The estimated treatment effect θ_1 for this trial is sampled,

$$\hat{\theta}_1 \sim N\left(30,\ \frac{8100}{n}\right).$$

2.3 Hypothesis testing

To evaluate the effect the candidate treatment has over the current standard we perform a frequentist hypothesis test. This is a statistical inference technique of two hypotheses representing a parameter value in a distribution. In our case this is the mean θ of the Gaussian distributed treatment effect. We define the null hypothesis as the treatment having no effect over the current standard, $H_0 : \theta = 0$ and the alternative hypothesis as the treatment

having a chosen effect δ over the control treatment $H_1 : \theta = \delta$. We aim to evaluate the overlap of these distributions in order to inform a decision to either reject or not reject H_0 based on the following error rates:

Type I - The probability of rejecting the null hypothesis when it is true.

Type II - The probability of rejecting the null hypothesis when the alternative hypothesis is false.

To formalize a decision, we choose a significance level α which represents the allowable Type I error probability. Consequently, this corresponds to some Type II error for a fixed n and δ. A critical value is derived to quantify the maximum overlap of these distributions according to the chosen significance level. It is possible that two distributions would overlap on both tails, in which case we might consider a two-tailed test including the error of negative effects of a treatment. Given the symmetry of the Gaussian distribution, the allowable error is split equally between the tails. Most existing literature is concerned with these types of tests, however these methods are easily adapted to a one-tailed approach as covered by Jennison and Turnbull (1999).

Conducting a test

The objective in this work is to determine a positive treatment effect $\hat{\theta} > 0$. Assuming we are testing positive treatment effects, the Type I error under the lower tail would be negligible. For this reason we use a test on upper tail as suggested from ICH best practices (ICH Steering Committee and others., 1998), evaluating:

$$H_0 : \hat{\theta} < 0 \qquad vs \qquad H_1 : \hat{\theta} > 0 \ .$$

To evaluate these parameter values one derives a test statistic through collecting some observations, in our case the sample mean treatment effect $\hat{\theta}$. To facilitate computation, it is convenient to standardize this statistic to a multiple of σ away from the mean of a standard normal distribution. The expression $Z = \frac{\sqrt{n}\hat{\theta}}{2\sigma}$ can be used assuming constant variance in observations. Using this notation, the Type I error for a given configuration is given by $P_{\theta=0}(\text{Reject } H_0) = \alpha$. Defining Φ as the cumulative standard normal distribution facilitates computations. The quantity $P(Z > Z_{1-\alpha}) = 1 - \Phi(Z)$ is known as the p-value. Evaluating whether the p-value is lower than the prescribed significance α, $P_{\theta=0}(\text{Reject } H_0) < \alpha$, concludes a test:

$$\text{Conclusion} = \begin{cases} \text{Reject } H_0 & \text{if } 1 - \Phi(Z) < \alpha \\ \text{Fail to reject } H_0 & \text{otherwise} \end{cases} .$$

2.4 Power analysis

2.4.1 Effect of sample size

Another useful quantity is the power of an experiment and is defined as the probability of correctly rejecting the control treatment as the superior drug for a chosen δ,

$$P_{\theta=\delta}(\text{Reject } H_0) = 1 - \beta \ . \tag{1}$$

In order to have confidence in the conclusion of the confirmatory trial, the sponsor would like the power to satisfy a target threshold. We later show how this directly affects the profitability of a programme. Assuming the variance of observations remains constant, we can solve for the sample size required for a target power requirement,

$$\begin{aligned}
1 - \beta &= P\left(Z > \frac{\delta}{SE} - z_{1-\alpha}\right) \\
1 - \beta &= 1 - P\left(Z < \frac{\delta}{SE} - z_{1-\alpha}\right) \\
\Phi^{-1}\left\{1 - P\left(Z < \frac{\delta}{SE} - z_{1-\alpha}\right)\right\} &= \Phi^{-1}\left\{1 - \beta\right\} \\
\frac{\delta}{SE} - z_{1-\alpha} &= z_{1-\beta} \\
z_{1-\beta} + z_{1-\alpha} &= \frac{\sqrt{n}\delta}{2\sigma^2}
\end{aligned} \tag{2}$$

$$\text{Sample size} = n = 4\sigma^2 \left(\frac{\Phi^{-1}\left\{1-\alpha\right\} + \Phi^{-1}\left\{1-\beta\right\}}{\delta}\right)^2 \ . \tag{3}$$

Toy example

In t_1, suppose we would like to establish a significant treatment effect $\delta_1 = 20$. The significance level enforced by a regulatory body is $\alpha_1 = 0.025$ and the sponsor's target power is $1 - \beta_1 = 0.90$. These are the industry standard levels (ICH Steering Committee and others., 1998) and will be used for Phase III trials globally in simulations in later sections. Using Equation 3 we compute the analytical sample size,

$$n_1 = 180 \left(\frac{\Phi^{-1}\left\{0.975\right\} + \Phi^{-1}\left\{0.90\right\}}{20}\right)^2 = 214 \ . \tag{4}$$

Since $\hat{\theta}_1$ accounts for both treatments, it suffices to simulate a set of observations $\{\theta_1, \ldots, \theta_{107}\}$. Suppose one arbitrary simulation returns a sample mean effect of $\bar{\theta} = 23.59$. The p-value deriving from this test statistics results in $p = 6.281 * 10^{-05}$, which concludes rejecting the null hypothesis.

2.4.2 Power for an increasing sample size

One of the main design choices we control when optimising a Phase III program is the sample size N_3. It follows from (2) that we can plot a power curve to demonstrate how power is affected by increasing the sample size. We plot this result for t_1 in Figure 1. It is clear that the gradient is steep for lower choices for N and decays gradually noticeably above $N_1 = 100$, tending asymptotically towards 1. Drawing a reference at $N_1 = 214$ we find the analytical solution computed in (4). At this point, power is $1 - \beta = 0.9$ and shows that any additional resources beyond this would offer comparatively marginal gains. Understanding how power changes for a range of values of

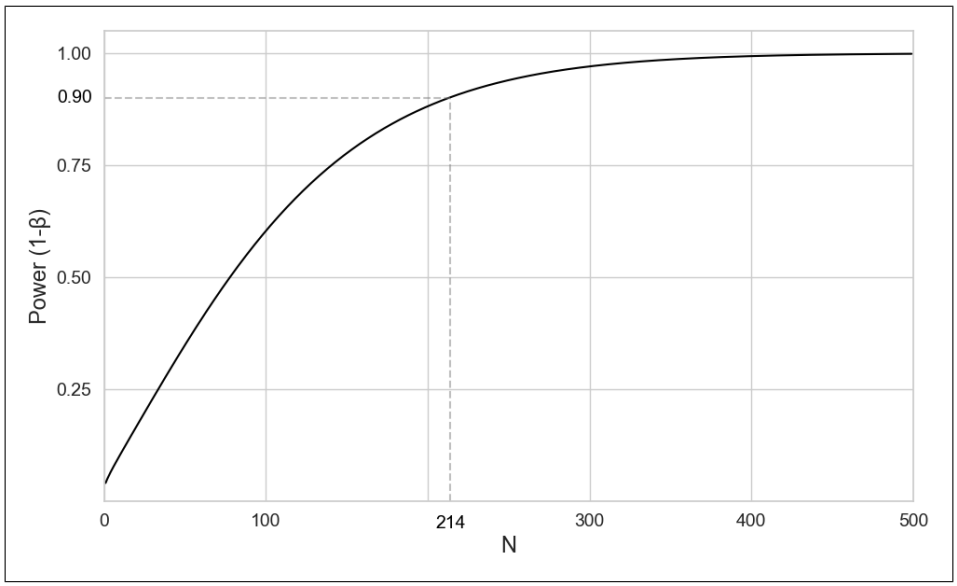

Figure 1: Power $(1 - \beta)$ for an increasing sample size N_3. Specific for the confirmatory trial configuration t_1, $\theta_1 = 30$, $\sigma_1 = 45$, $\delta_1 = 20$, $\alpha_1 = 0.025$.

sample size is a beneficial result for a sponsor. We show in Section 2.5 how the power directly affects the utility of a program. Applying constraints such as a maximum budget available means there is a limit on how many patients can be recruited, consequently limiting the achievable power of a trial. These quantities become of use in Section 7 where a limited number of trials is selected according to their expected return and the resources allocated to them.

2.4.3 Experimental method to compute power

Being able to estimate power through simulation is a useful process for instances where it is computationally demanding. By simulating m trials, the power estimate $1 \hat{-} \beta$ corresponds to the proportion of times H_0 was rejected. Therefore, the random variable R which corresponds to rejecting H_0 can be modelled by the Binomial

distribution and respective standard error,

$$R \sim Bin(m,\ 1 \hat{-} \beta),$$
$$SE_{Bin} = \sqrt{\frac{(1 \hat{-} \beta)\hat{\beta}}{m}}.$$

These properties illustrate that we can target a desired precision for the simulated power estimate by controlling the number of iterations. Table 2 shows the effect the choice of m has on estimating power for a selection of different treatment effects and same hypothesis parameters configured for t_1. Increasing m effectively decreases the standard error in all cases with the biggest improvements in smaller treatment effects.

θ	Theoretical power	m	Estimated power	Standard Error
15	0.684	500	0.728	0.020
	0.684	1500	0.691	0.012
	0.684	7500	0.691	0.005
20	0.902	500	0.936	0.011
	0.902	1500	0.903	0.008
	0.902	7500	0.895	0.004
25	0.982	500	0.986	0.005
	0.982	1500	0.987	0.003
	0.982	7500	0.980	0.002

Table 2: Power estimate and standard error computed by simulating m trials for different treatment effects θ_1. For trial t_1 with configuration $\sigma_1 = 45$, $\delta_1 = 20$, $\alpha_1 = 0.025$ and $\beta_1 = 0.10$.

2.5 Assurance

In Section 6.1 we will introduce the Bayesian paradigm and how it will be used in the context of this study. For now, lets simply assume $\pi(\theta)$ is a prior distribution which represents our belief about how the treatment effect θ is distributed. The conditional probability for power in this case is given by,

$$\nu(\theta) = P(\text{Reject } H_0|\theta) = \Phi\left\{\frac{\sqrt{n}\theta}{2\sigma^2} - \Phi^{-1}\left\{1 - a\right\}\right\}.$$

A quantity we are interested in is the unconditional probability that a trial will reject the null hypothesis H_0 (O'Hagan, Stevens and Campbell, 2005), known as assurance and Probability of Success (POS). Some literature

on clinical portfolio optimisation, use the same acronym (POS) to quantify the success of a programme considering technical and regulatory hurdles through the entire timeline of the approval process (Wiklund et al., 2024). The POS therefore should be differentiated from assurance in the way that it will be used in Section 7. In the case of a single experiment or trial, POS is used to express the probability of concluding an efficacy objective, whereas in a portfolio setting POS is used to assess the feasibility of executing an efficient programme to produce a competitive treatment in a broader scope.

Since the prior expresses some uncertainty about the fixed but unknown true value of the treatment effect, we can marginalise out θ to compute the quantity,

$$\mathbb{E}_{\pi(\theta)}[\nu(\theta)] = \int_{-\infty}^{\infty} \pi(\theta)\nu(\theta)\ d\theta\ . \tag{5}$$

In the duration of trials we have a prior distribution $\pi_1(\theta)$ that informs the sampling of the treatment effect θ, details of which should not concern us at the moment and are covered in Section 6. Using Bayesian methods we can update this prior belief to account for some observed data, therefore it is imortant to understand how different priors affect the assurance.

Toy example

To illustrate the effect priors of the form $N(\theta, \sigma^2)$ have on the assurance of a trial, consider the configuration t_1 proceeding to Phase III with $N_3 = 214$ recruited patients. Table 3 demonstrates the effect a selection of priors have on the assurance $E_{\pi(\theta)}[\nu(\theta)]$, computed using Simpson's Composite Rule. The inner workings of this method are introduced in Section 4.1. For now r can be thought to control the accuracy of the estimate. The configuration $r = 16$ yields evaluations within 10^{-6} (Jennison and Turnbull, 1999) of their true values. This is a sufficiently accurate result for our use case, and will be used as such throughout this work. As the significant treatment effect in t_1 is $\delta_1 = 20$, Table 3 shows that as θ increases beyond δ the expected power improves. The decaying effect of assurance as variance increases is best illustrated in the case where $\theta = \delta = 20$.

3 Sequential decision-making

3.1 Sequential decisions in clinical trials

In conducting a trial, besides the final analysis that concludes Phase III we are able to make a decision after having observed some data in Phase II, known as a sequential decision. The benefit of using such as design is to allow for early stopping which has various ethical, montoring and financial intentions. It may be the case that a treatment presents harmful effects to patients, in which case the trial should cease immediately. If there is overwhelming

θ	σ	$E_{\pi(\theta)}[\nu(\theta)]$
15	10	0.599
	30	0.538
	50	0.523
30	10	0.937
	30	0.721
	50	0.639
45	10	0.997
	30	0.859
	50	0.743

Table 3: Effect of different combinations of treatment effect θ_1 and σ_1 on assurance. For trial t_1 with configuration $\delta_1 = 20$, $\alpha_1 = 0.025$ and $\beta_1 = 0.10$. Evaluated numerically using Simpson's Composite Rule.

proof that the null hypothesis is false at an interim analysis the experiment stops due to efficacy, and vice versa for futility (Lewis, 2023). In general, a sponsor would like to recruit the least patients possible. Continuing after efficacy increases recruitment cost with no additional gain from marketing the drug. The case of futility is equally important to the sponsor for limitting his losses if a treatment is unlikely to be approved. To perform such a test, one uses the framework introduced in Section 2.3, with the only amendment of introducing the significance b. If results are poor and the test statistic yields a Type I Error greater than the significance level b. The test is concluded for futility if $P_{\theta=0}(\text{Do not reject } H_0) > b$.

3.2 Group sequential designs

3.2.1 Canonical distribution for test statistics

This family of experiments builds on sequential decisions to consider a number of interim analyses in the duration of an experiment where a hypothesis test can be performed on accumulated data. This is different to simple sequential decision-making in our context as it is used to perform Phase II or III trials.

To define a group sequential test it is useful to define the canonical distribution for test statistics, following the practice of Jennison and Turnbull (1999). In doing so we define the following parameters at interim analysis k: the cumulative sample size n_k, Fisher information $I_k = n_k/(2\sigma^2)$, MLE of treatment effect $\hat{\theta}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} X_i - Y_i$ and

$\hat{\theta}_k \sim N(\theta,\ I_k^{-1})$. Letting $Z_k = \hat{\theta}_k\sqrt{I_k}$, the properties are:

$$\begin{aligned} &(Z_1,\ \ldots, Z_k) \text{ is multivariate normal,} \\ &\mathbb{E}(Z_k) = \theta\sqrt{I_k},\ \text{for } k = 1,\ \ldots,\ K, \\ &Cov(Z_{k_1},\ Z_{k_2}) = \sqrt{I_{k_1}/I_{k_2}} \text{ for } k_1 < k_2. \end{aligned} \tag{6}$$

3.2.2 Error spending

Repeated significance testing on accumulated data requires a strategy to manage the nominal significance level across interim analyses. If the nominal α was assigned at each interim analysis, the final significance realised would surpass the intended level increasingly as $k > 1$. The work of (Armitage, McPherson and Rowe, 1969, Table 2) shows through numerical techniques and simulation that multiple looks at data in an experiment with nominal two-sided $\alpha = 0.02$ would result in ~ 3 times the significance with $k = 5$, and ~ 4.5 with $k = 10$.

Jennison and Turnbull (1999) define the functions $f(t)$ and $g(t)$ that determine the amount of the nominal α and β respectively to be spent at each interim analysis. In a maximum information design we have a target maximum I_{max} after which there will be a definite conclusion, therefore $I_K = I_{max}$. At an interim analysis k of maximum K analyses, t is defined as the fraction of maximum observation observed, $t = I_k/I_K$. The expressions are used to spend up to reaching the nominal level by:

$$\begin{aligned} f(t) &= minimum\left\{\alpha t^{\rho},\ \alpha\right\} \\ g(t) &= minimum\left\{\beta t^{\rho},\ \beta\right\} \end{aligned}$$

The above ensures satisfying the nominal totals of α and β when information is observed in unpredictable sizes. In our case we define each analysis to have equal amounts of information, $I_1 = I_2 = \cdots = I_K$. Pampallona and Tsiatis (1994) have generalised an error spending approach in what they have defined as the power family. The rejection and acceptance boundaries b_k and a_k,

$$\begin{aligned} b_k &= c^R_{WT}(K,\ \alpha,\ \Delta)(k/K)^{\Delta-1/2} \\ a_k &= \delta\sqrt{I_k} - c^{R'}_{WT}(K,\ \alpha,\ \Delta)(k/K)^{\Delta-1/2}. \end{aligned}$$

An interim analysis follows the following decision rule $d_k(Z_k)$:

$$d_k(Z_k) = \begin{cases} \text{if } Z_k > b_k, \text{ Stop. Reject } H_0 \text{ for efficacy} \\ \text{if } a_k < Z_k < b_k, \text{ Continue with group } k+1 \\ \text{if } Z_k < a_k, \text{ Stop. Do not reject } H_0 \text{ for futility} \end{cases}$$

The choice of Δ controls the shape of the boundaries. Other established error spending approaches can be defined as special cases of the power family i.e. Pocock (1977) with $\Delta = 1/2$, and O'Brien and Fleming (1979) with $\Delta = 0$. Different boundary shapes control how easy it is to accept or reject a hypothesis as the experiment progresses. O' Brien and Flemming boundaries assign a small α in first analyses, making it difficult to reject the null hypothesis. This can be considered a responsible choice for ethical reasons such as mitigating the effect of randomness in a small sample size and going to market with a harmful or weaker treatment. Pocock assigns a constant significance level for each interim analysis resulting in a narrower linear boundary with a larger portion of the nominal significance spent in early analyses. This effect is opposite to O'Brien boundaries and makes it easier to reject the null hypothesis due to randomness of a small sample. The difference between the two boundaries is depicted in an arbitrary trial in Figure 2. We need not concern ourselves with the details of this trial but only with the test statistic and boundary shape affecting the trial outcomes. In the case of the O'Brien Flemming, the trial continues up to the 3rd interim analysis to conclude for futility. For the same test statistic, the trial concludes in the first interim analysis in the case of Pocock. Whereas O'Brien Flemming boundaries need more evidence in early analyses to conclude a trial which can be ethically correct, one could also argue that the trial concluded for futility prematurely in the case of Pocock. This is ultimately up to the sponsor to decide and carefully evaluate the effect different realizations of the trial have both on the recruited patients and programme outcome (Jennison and Turnbull, 1999).

3.2.3 Early stopping

The probability of concluding an experiment at analysis k is then given via the following expressions,

$$P(\text{Reject } H_0 \text{ at } k \mid \theta) = P(a_1 < Z_k < b_1, \ \ldots, a_{k-1} < Z_{k-1} < b_{k-1}, Z_k > b_k)$$
$$P(\text{Do not reject } H_0 \text{ at } k \mid \theta) = P(a_1 < Z_k < b_1, \ \ldots, a_{k-1} < Z_{k-1} < b_{k-1}, Z_k < a_k) \ ,$$

which can be evaluated to compute the expected sample size. It is the case that group sequential designs explicitly lower the expected sample size. Jennison and Turnbull (1999) details this process using numerical integration techniques as in 4.1 for root solving calculations and fixed treatment effect θ. Peck (2020) extends this to a Bayesian paradigm as in our case. Integrating the analytical solution would require more time than is available in

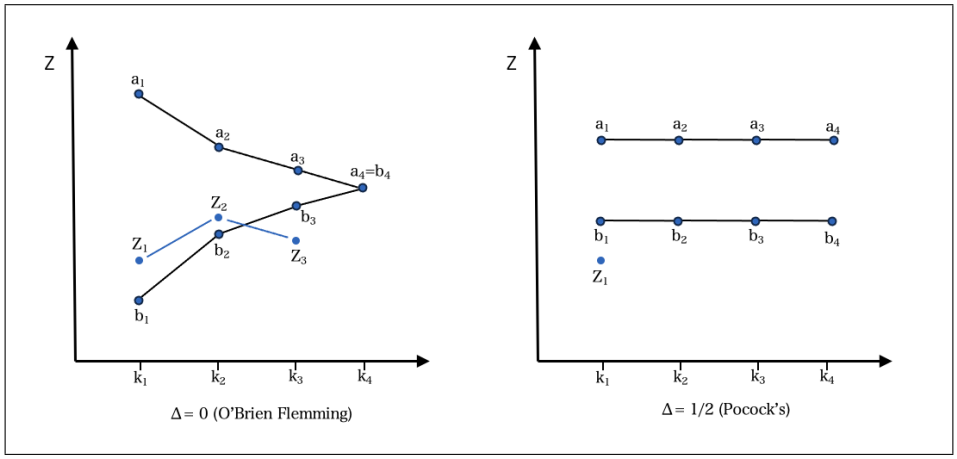

Figure 2: Early stopping in an arbitrary group sequential trial with $k = 4$. Different realisations for comparing $\Delta = 0$ (O' Brien & Flemming) and $\Delta = 1/2$ (Pocock) boundaries.

our case. Recognising the benefit of such designs however, we compute the acceptance and rejection boundaries using the process described above and integrate the R package 'MAMS' (Jaki, Pallmann and Magirr, 2019) to simulate N trials. Finally, we can sufficiently estimate the assurance and expected sample size using the process described in Section 2.4.3.

4 Integral estimation

4.1 Simpson's rule

An integral we frequently evaluate includes the density of some Gaussian distribution $\phi \sim N(\mu,\ \sigma^2)$ over the real number line $\mathbb{R}$. The integral has the form

$$I = \int_{\mathbb{R}} f(x)\phi(x)\ dx \tag{7}$$

which given its asymptotic nature, a solution exists only approximately using numerical integration. We will adapt the approach of (Jennison and Turnbull, 1999, Chapter 19), where the integral is broken up into a set of sub-intervals by setting initial points x logarithmically spaced towards the tails and densely around the mean of ϕ, controlled by the parameter r. A larger r will give better precision for additional computational cost.

$$x_i = \begin{cases} \mu + \sigma\left(-3 - 4\log(\frac{r}{i})\right) & \text{when } i = 1,\ \ldots\ ,\ r-1 \\ \mu + \sigma\left(-3 + 3(\frac{i-r}{2r})\right) & \text{when } i = r,\ \ldots\ ,\ 5r \\ \mu + \sigma\left(3 + 4\log(\frac{r}{6r-i})\right) & \text{when } i = 5r+1,\ \ldots\ ,\ 6r-1 \end{cases}$$

Working better with smooth functions, Simpson's rule is a numerical integration technique that fits a second order polynomial between pairs of odd numbered indices $x_i \in \{1,\ 3,\ \ldots,\ m\}$. The quadratic is then evaluated over the range of the two indices for all sub-intervals. To improve the sensitivity of this method and to allow for less smooth functions, we opt into using Composite Simpson's rule. This approach uses $N = 2m - 1$ points by setting the midpoints of the initial integration points as the even indices, $z_i = x_{(i+1)/2}\ for\ i\ \in \{1,\ 3,\ \ldots,\ N\}$, and $z_i = (z_{i-1} + z_{i+1})/2\ for\ i\ \in \{2,\ 4,\ \ldots,\ N-1\}$ as odd indices. To evaluate the entire integral, we evaluate the sub-intervals between to odd numbered indices summation over the population. It can be shown that the corresponding cumulative density is given by,

$$\int_{z_{i-1}}^{z_{i+1}} f(x)\phi(x)\ dx = \frac{d}{6}f(z_{i-1})\phi(z_{i-1}) + \frac{4d}{6}f(z_i)\phi(z_i) + \frac{d}{6}f(z_{i+1})\phi(z_{i+1}). \tag{8}$$

The weights and approximation to the full integral are as in (9) and (10), yielding an error of magnitude $O(N^{-4})$:

$$w_i = \begin{cases} \frac{1}{6}(z_3 - z_1) & \text{when } i = 1 \\ \frac{2}{6}(z_{i+2} - z_{i-2}) & \text{when } i = 3,\ 5,\ \ldots\ ,\ N-2 \\ \frac{4}{6}(z_{i+1} - z_{i-1}) & \text{when } i = 2,\ 4,\ \ldots\ ,\ N-1 \\ \frac{1}{6}(z_N - z_{N-2}) & \text{when } i = N \end{cases} \tag{9}$$

$$\int_{\mathbb{R}} f(x)\phi(x)\ dx \approx \sum_{i=1}^{N} w_i f(z_i)\phi(z_i) \tag{10}$$

Jennison and Turnbull (1999) compute the standard error for equations of the form (5) with $r = 16$ to be within 10^{-6} of their true values. It is also stipulated that halving the magnitude of r reduces the accuracy by one decimal place. We use the configuration $r = 16$ throughout this thesis, providing a high degree of accuracy for the purpose of the research.

4.2 Monte-Carlo

An estimate to the integral (7) is also attainable through the Monte-Carlo method (Robert et al., 2004) whith error $O(N^{-1/2})$. This method is often preferred when estimating high dimensional problems to save computation at the expense of increasing the error. In our case we have a smooth function for which Simpson's rule works sufficiently well. However, the 'MAMS' package used to compute group sequential designs uses Monte-Carlo integration to estimate the integral (7). This adds a layer of variation which will be discussed in due course.

Monte-Carlo integration works by taking N random observations from ϕ,

$$x_1, \ \ldots, x_N \overset{iid}{\sim} \phi \ ,$$

and using the following summation to approximate over the entire number line,

$$\int_{\mathbb{R}} f(x)\phi(x) \ dx \approx \frac{1}{N} \sum_{i=1}^{N} f(x_i) \ .$$

Using the sample mean $\bar{f}$, the sample variance is given by $\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (f(x_i) - \bar{f})^2$. The resulting standard error is given by,

$$SE_{MC} = \frac{\sigma}{\sqrt{N}}$$

To demonstrate the effect different sampling sizes N have on the accuracy of Monte Carlo integration, we estimate $E_{\pi(\theta)}[\nu(\theta)]$ for different choices of N and configuration t, $\delta = 20$, $\alpha = 0.025$, $\beta = 0.10$, $N_3 = 214$ and $\theta_1 \sim N(30, 30)$. From the results shown in Table 4 we see that the decrease of standard error decays for larger values of N. In earlier Section 2.5, Table 3, we used Simpson's rule and computed $E_{\pi(\theta)}[\nu(\theta)] = 0.721$, which is within 10^{-6} of the true value. Using Monte Carlo integration, the estimate converges to a sufficiently accurate estimate of the same value for $N = 50000$ and above. Additional accuracy beyond this choice is marginal and will result in longer computation for large problems.

N	$E_{\pi(\theta)}[\nu(\theta)]$	SE
100	0.676	3.311
200	0.734	1.978
500	0.741	1.360
1000	0.699	0.956
3000	0.730	0.541
10000	0.719	0.298
50000	0.721	0.134
100000	0.722	0.095
200000	0.721	0.067

Table 4: Effect different number of simulations N has on the standard error of estimating $E_{\pi(\theta)}[\nu(\theta)]$ in Phase III using Monte Carlo integration. For configuration t: $\delta = 20$, $\alpha = 0.025$, $\beta = 0.10$, $N_3 = 214$ and $\theta \sim N(30, 30)$.

5 Programme utility

In clinical trials, sponsors want to objectively evaluate a financial model corresponding to each individual programme. Utility and risk functions offer such a solution which can be used interchangeably with an appropriate preceding sign. The prevailing preference to monitor the performance of investments is the eNPV (Wiklund, Thorn, Götte, Hacquoil, Saint-Hilary and Carlton, 2024) (Antonijevic, 2014). It is designed to return the overall utility of an investment, accounting development costs, time, and other miscellaneous risks such as inflation (Brigham and Houston, 2013). Besides being an informative value in of itself, it also offers a comparable metric between decisions in the course of the programme. One can include infinite design choices. In this work it is important to maintain a relevant financial model and keep the computational cost manageable without losing valuable descriptors. The key parameters we consider in our financial model effectively model market value, true effect of treatment, trial specific costs, and experiment design for the purpose of this thesis.

Source	Parameter	Description	Value
Treatment	G	The monetary gain to the sponsor given treatment effect δ, if a treatment successfully reaches the market.	$G \in \mathbb{R}^+$
	$\zeta(\theta)$	The new treatment should be substantially better than the current standard. The true difference from the target difference δ will affect the market value.	$\zeta(\theta) \in (0,\ 2)$
Trial design	c_0	General setup cost.	$c_0 \in \mathbb{R}^+$
	c_2	Phase II setup cost.	$c_2 \in \mathbb{R}^+$
	c_3	Phase III setup cost.	$c_3 \in \mathbb{R}^+$
	N_2	The number patients recruited in Phase II.	$N_2 \in \mathbb{N}$
	γ_2	The cost of recruiting patients in Phase II.	$\gamma_2 \in \mathbb{R}^+$
	N_3	The number patients recruited in Phase III.	$N_3 \in \mathbb{N}$
	γ_3	The cost of recruiting patients in Phase III.	$\gamma_3 \in \mathbb{R}^+$
	c_{int}	Cost of conducting an interim analysis.	$c_i nt \in \mathbb{R}^+$
	K	Total number of interim analyses.	$K \in \mathbb{N}$

Table 5: Details of the utility function (NPV).

Put together, the utility function takes the following form,

$$U(\theta) = G\zeta(\theta)\mathbb{1}_{Market} - c_0 - (c_2 + N_2\gamma_2)\mathbb{1}_{Phase\ II} - (c_3 + N_3\gamma_3)\mathbb{1}_{Phase\ III} + c_{int}K\ , \tag{11}$$

where $\mathbb{1}$ is an indicator variable with a value of 1 for each of: a programme having a successful outcome, incurring a setup cost for Phases II and III. For the portfolio problem, we simplify the cost of conducting only Phase II of a programme to $C_2 = c_0 + (c_2 + N_2\gamma_2)\mathbb{1}_{Phase\ II}$, and Phase III of a programme to $C_3 = (c_3 + N_3\gamma_3)\mathbb{1}_{Phase\ III} + c_{int}K$.

We do not need to include the interim cost in C_2 as it is absorbed by the trial setup term c_2.

To simulate how different observed treatment effects affect the market value of a programme we use the expression,

$$\zeta(\theta) = 2/1 + \exp^{-S(\theta-\delta)} \tag{12}$$

Using this expression, the market value tends asymptotically to 2 times the expected market value as the true effect tends to infinity, and zero times as the effect tends to negative infinity. The parameter S controls the sensitivity of the sigmoid function for changes in θ. Figure 3 demonstrates how different values of S behave for a range of true treatment effects. A high value of S would increase/decrease the market value with greater sensitivity over a lower value for S. In practice, treatments in different therapeutic areas will have different values for S. For example, a small change in effect of a drug used to treat cancer would have a more sensitive multiplier than a simple painkiller. Since the programmes that are worth pursuing have treatment effect greater than δ, the programmes with negative observed effects (below δ) have greater downside. A neutral effect of these treatments already has a positive multiplier through $\zeta(\theta)$ and any additional positive effects tends asymptotically towards 2.

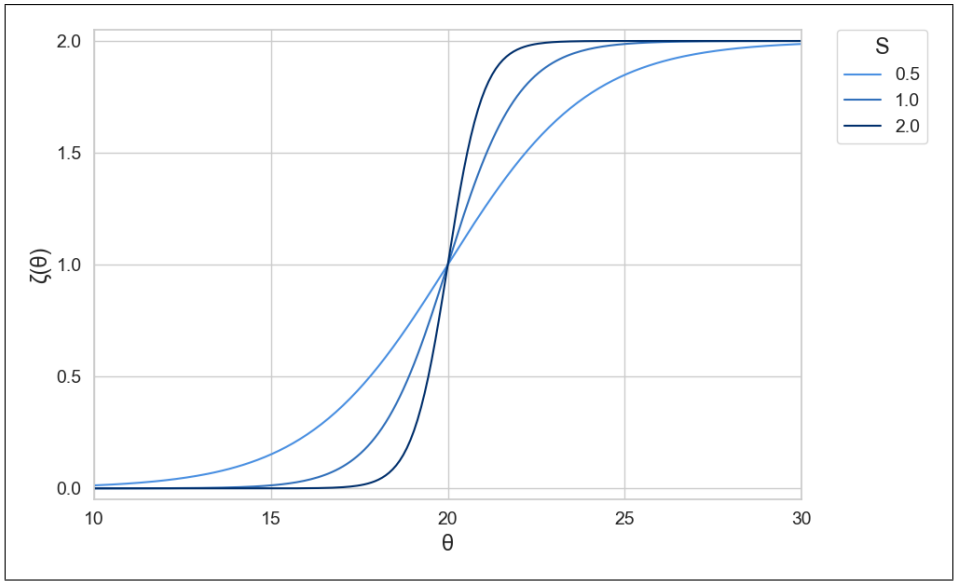

Figure 3: Multiple $\zeta(\theta)$ of the market value at specified δ. In the context of configuration t_1 with $\sigma_1 = 45$, $\delta_1 = 20$, $\alpha_1 = 0.025$, $\beta_1 = 0.10$ for a range of true treatment effect θ.

A quantity we are interested in to evaluate different programmes is the expected NPV (eNPV). Provided the treatment effect has sampling distribution $\pi(\theta)$ we can calculate it by marginalising out θ,

$$E_{\pi(\theta)}[U(\theta)] = \int_{-\infty}^{\infty} \pi(\theta)U(\theta)\ d\theta\ . \tag{13}$$

A programme is originally designed in Phase II considering some initial distributional assumptions of the treatment effect. After Phase II has been conducted we have some data concerning the true treatment effect. In Section 6 we show how the prior is affected by the accrual of new information to yield a posterior. It follows that the final utility of a programme is affected both by the assurance (5) computed using the posterior, and the expectation of the true treatment effect as in (12). Having considered Phase II observations.

It is important to clarify the design choices within programmes. The sample size N_2 is the first design choice and given it includes the smaller sample size in practice, we opt into choosing the design using the fixed sample size calculation (3). The choice of N_3 is either a fixed sample in the case of $K = 1$ interim analyses or the choice of maximum sample size N_{3max} for the case $K > 1$. From the maximum sample size, we derive the expected sample size to be substituted as N_3 in the utility function. To manage the computational time in examples we consider a maximum of $K = 2$ analyses.

Toy example

Consider two variants of programme P_1 denoted $P_{1.1}$ and $P_{1.2}$. These have common parameters: $G = 1000000,\ S = 0.7,\ c_0 = 80000,\ c_2 = 30000,\ c_3 = 50000,\ N_2 = 50,\ \gamma_2 = 200,\ N_3 = 214,\ \gamma_3 = 800,\ c_{int} = 2000,\ K = 1$. The only difference between the programmes is the prior at the start of Phase III, $\pi_{P_{1.1}}(\theta) \sim N(25,\ 20)$ and $\pi_{P_{1.2}}(\theta) \sim N(30,\ 15)$. Using (13) we calculate $E_{P_{1.1}}[U(\theta)] \sim$ £1.08m and $E_{P_{1.2}}[U(\theta)] \sim$ £1.40m. The eNPV in the case of $trial_2$ is clearly superior. The larger mean and smaller variance alone increases the eNPV by $\sim 30\%$ from the otherwise identical t_1.

Now consider the case where $\pi_{P_{1.1}} = \pi_{P_{1.2}} = N(30,\ 15)$. Figure 4 demonstrates the effect different sensitivies S would have on the eNPV of otherwise identical programmes. Lower sensitivities have greater differences in eNPV for an increasing sample size. As the sensitivity increases note the eNPV achieves higher quantities. Due to the nature of $\zeta(\theta)$, the treatment value can only achieve up to twice the initial researched market value. Rationally, a sponsor should not opt into pursuing a treatment that has negative treatment effects. If the treatment effect θ were close to the programmes δ it might have some small eNPV, however, increasing the sample size for a treatment with a negative prior mean would result in a negative and decreasing eNPV. The key takeway from this demonstration is that programmes with otherwise similar parameters but with higher sensitivity will have greater differences in eNPV when adjusting the choice of sample size below the maximal solution.

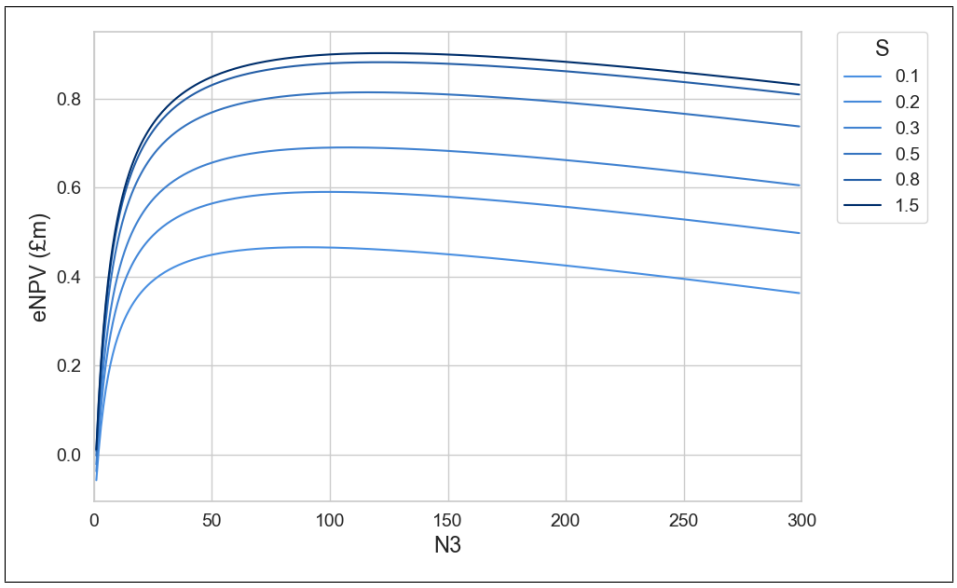

Figure 4: Effect of varying sensitivity on programmes $P_{1.1}$ and $P_{1.2}$. The programmes have common parameters $G = 1000000,\ c_0 = 80000,\ c_2 = 30000,\ c_3 = 50000,\ N_2 = 50,\ \gamma_2 = 200,\ N_3 = 214,\ \gamma_3 = 800,\ c_{int} = 2000,\ K = 1$.

6 Bayesian methods

6.1 State of nature

Although we use Frequentist techniques it does not limit us and can use a Bayesian paradigm to frame several aspects. The primary difference is that instead of a parameter having a fixed value, we say that it is sampled from a state of nature commonly denoted by Θ (Berger, 2013). In context, we now define some uncertainty about the static treatment effect θ by assigning it the sampling distribution $\Theta \sim \pi(\theta)$ known as the prior. Priors are manifested through some degree of consensus of subjective but informed clinical opinions (Dolan, Bordley and Mushlin, 1986), details of which extend beyond this work.

6.2 Posterior distribution

As the names suggest, the posterior distribution updates the prior post observing some data using Bayes' rule. The probability of observing some data x conditional to the state of nature $f(x|\theta)$, is known as the likelihood of the data. Consider the joint density function of the likelihood and prior distributions $h(x, \theta) = f(x|\theta)\pi(\theta)$. Since θ is continuous, integrating $h(x, \theta)$ over the state of nature θ evaluates to the marginal unconditional density $m(x)$

(Berger, 2013), known as the evidence or $P(Data)$.

$$h(x,\theta) = f(x|\theta)\pi(\theta)$$
$$m(x) = \int_{\theta} f(x|\theta)\pi(\theta)\ d\theta$$
$$\pi(\theta|x) = \frac{h(x,\theta)}{m(x)}$$

When optimising, we neglect normalising constants since they do not influence the parameter choice that maximize a given argument. Since $m(x)$ evaluates to a scalar quantity the posterior is proportional to the following,

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta)\ .$$

Often times, deriving the posterior is technically difficult. Some combinations of likelihoods and priors have no exact solution and have to be evaluated numerically, whereas certain others yield a posterior belonging to the same distributional family as the prior. This result can be shown algebraically and are known as conjugate priors (Berger, 2013). Using conjugate families simplifies a vast computational burden that would otherwise require using Markov Chain Monte Carlo algorithms to be evaluated efficiently. Such families have been explored in depth by Raiffa and Schlaifer (2000). A trial assumption we have made is that the variance is known and remains constant, therefore, the likelihood of the observed data x must also be Gaussian with known variance σ^2. In this case, with Gaussian prior $\pi(\theta) \sim N(\mu_0, \sigma_0^2)$ the posterior also belongs to the Gaussian family and can be computed via,

$$\pi(\theta|x) \sim N\left(\frac{1}{\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}}\left(\frac{\mu_0}{\sigma_0^2}+\frac{n\bar{x}}{\sigma^2}\right), \left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)^{-1}\right) \tag{14}$$

This expression becomes useful on completion of Phase II. Having observed some data x_2 we can update the distribution $\pi_{Ph2}(\theta)$ to obtain $\pi_{Ph3}(\theta|x_2)$. The posterior is then accepted as the new prior $\pi_{Ph3}(\theta)$ for any future calculations on the efficacy of continuing a program such as the assurance illustrated in 5.

6.3 Bayes' Optimal Decision

The need for decision science arises to formalize actions when there are uncertain outcomes on a measurable objective. With infinite data there is not much unknown. In some cases, it may be too abstract or may have too many parameters, however, by quantifying trade-offs, choices are supported by probability and expected value. Evaluating a decision to one single action is known as a deterministic rule. A randomised rule extends this by choosing a rule via some probability. There exist a multitude of well known decision rules, yet the one we will use

to evaluate a trials expected utility is Bayes' Optimal Decision (Ghosh, 1988), which opts for the action with the greatest expected utility,

$$A^* = \arg\max_A \int_\theta \pi(\theta)\ G(\theta)\ d\theta$$

Optimal Phase III design

At the end of Phase II we use Bayes' Optimal Decision to choose the decision $d_1(x_2) = A_3$, $A_3 \in \Omega_{A_3}$, corresponding to the different Phase III designs one can opt for by configuring the parameters in Section 5. The decision d_1 is made when data x_2 has been observed, hence we can update the prior to be used in Phase III calculations, $\pi_3(\theta) = \pi_2(\theta|x_2)$. The integral of the joint distribution $\pi_3(\theta)G(A_3, \theta)$ over all possible values of θ, returns the expected gain $E[G(A_3, \Theta)]$ of a design option.

$$\begin{aligned} E[G(A_3, \Theta)] &= \int_\theta \pi_2(\theta) \int_{x_2} f(x_2|\theta) G(A_3, \theta)\ dx_2\ d\theta \\ &= \int_{x_2} \int_\theta \pi_2(\theta) f(x_2|\theta) G(A_3, \theta)\ d\theta\ dx_2 \\ &= \int_{x_2} m(x_2) \left[\int_\theta \pi_2(\theta|x_2) G(A_3, \theta)\ d\theta \right]\ dx_2 \end{aligned}$$

Since $\int_{x_2} m(x_2)\ dx_2$ is scalar quantity, the optimal Phase III design d_1 simplifies to the following expression:

$$d_1^* = \arg\max_{d_1} \int_\theta \pi_3(\theta)\ G(A_3, \theta)\ d\theta \tag{15}$$

It is the case that the utility function (11) is convex for a varying sample size N_3. This means there is a single design configuration for Phase III that will yield a global maximum eNPV for specific programme with fixed number of interim analyses K. This is different to the maximum achievable power. Increasing the sample size will always yield greater power, however we showed in Figure 1 how beyond a certain point gains become marginal. In Figure 5, we plot the eNPV for different choices of N_3 in the configuration of t_1 defined in previous sections with $k = 1$ interim analyses. In this specific instance we find that $N_3 = 123$ gives the maximum eNPV of £0.90m. However, we also find that $N_3 = 80$ and $N_3 = 180$ both give £0.89m which is less than 2% of the maximum. Since N_3 is ultimately a design choice, a sponsor can take such information into consideration. In a portfolio of multiple drugs, a decrease in the maximum yielding sample size of one programme can translate into an increase in the total eNPV due to the greater effect of increasing efforts in another programme.

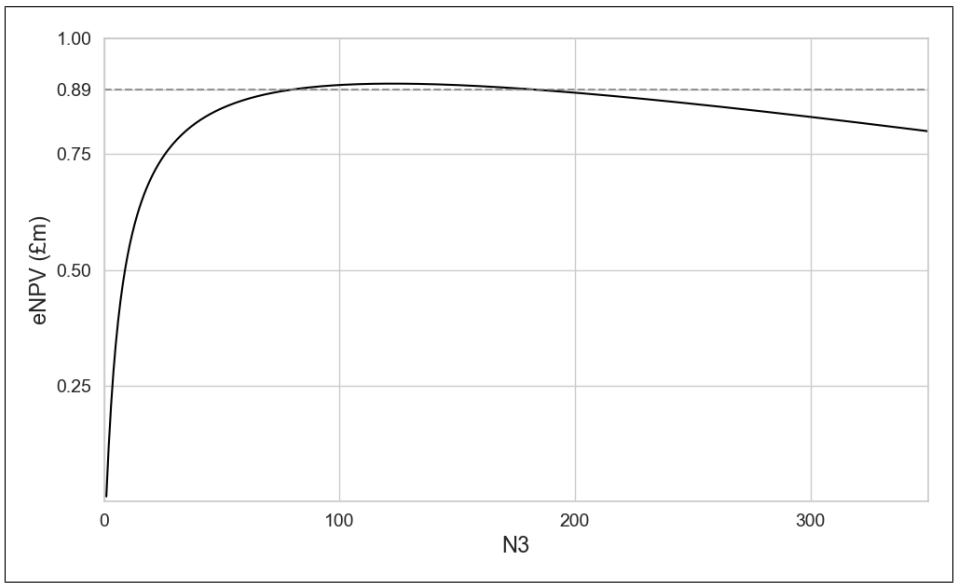

Figure 5: eNPV in £m for increasing N_3 in programme P_1, $G = 1000000,\ S = 0.7,\ c_0 = 80000,\ c_2 = 30000,\ c_3 = 50000,\ N_2 = 50,\ \gamma_2 = 200,\ N_3 = 214,\ \gamma_3 = 800,\ c_{int} = 2000,\ K = 1$. Reference line at sub-maximal eNPV to demonstrate marginal effect of differences in N_3.

Efficient searching

In convex cases, there are elegant techniques to reduce computation, however, most traditional methods are defined for continuous functions. Although our case is discrete, have an ordered set of choices for $N_3 \in \mathbb{N}$ which provides an opportunity for efficient searching. To find the optimal sample size N_3 for $k = 1$ interim analyses, and the maximum sample size for cases $k > 1$, we use a custom algorithm inspired by traditional 2-dimensional gradient descent and binary search to reduce the computation. This algorithm draws inspiration from the Golden Section Search (GSS) method (Press, 2007). Although GSS computes derivatives, it is still defined for a continuous function. The adapted algorithm for discrete functions has computational complexity $O(log(N))$ for a fixed K and works as follows:

1 Define search boundaries and design options - Set the lower and upper boundaries N_{min} and N_{max}. We typically have an initial estimate of how many patients to recruit which can help inform the initial search area. If we are uncertain we can set the initial boundaries to an exhaustive range or other restrictions i.e. budget. For the $k = 1$ case, we define a unit increasing array corresponding to the total sample size. For the cases $k > 1$ we define an array in increments of 4 up to a multiple of the fixed sample size design ($k = 1$) in order to allow for integer assignments to treatments within groups. The algorithm operates using list indices and thus remains unaffected.

2 Evaluate eNPV for midpoint and neighbours - Compute the utility of the midpoint with sample size $N_{mid} = (N_{min} + N_{max})//2$, as well as for the left and right neighbours N_{mid-1} and N_{mid+1}. All eNPV

evaluations are stored to avoid unnecessary re-computation in future iterations.

3 Update next midpoint - Evaluate in which direction the eNPV increases and set the next midpoint in that direction, i.e. if the utility of the left neighbor is greater set $N_{mid_2} = (N_{min} + N_{mid_1})//2$ and the new upper, provided there will be another iteration, to $N_{max_2} = N_{mid_1}$.

4 Repeat until converged - The algorithm repeats steps 2 & 3 until N_{mid_i} yields the maximum utility.

Consider the programme with configuration $G = 1000000,\ S = 0.7,\ c_0 = 80000,\ c_2 = 30000,\ c_3 = 50000,\ N_2 = 50,\ \gamma_2 = 200,\ N_3 = 214,\ \gamma_3 = 800,\ c_{int} = 2000$, now with $K = 1$ and $\pi_{Ph3}(\theta) \sim N(30,\ 15)$ in Phase III. In Table 6 we initialise the lower and upper boundaries for t_1 to search over 600 choices for N_3. The algorithm converges in 4 iterations and 16 evaluations of eNPV, and finds that a sample of $N_3 = 229$ yields the maximum.

Iteration	N_3	eNPV
1	341	1358456
	342	1358023
	343	1357588
2	191	1381501
	192	1381830
	193	1382148
3	266	1383290
	267	1383086
	268	1382877
4	228	1387304
	229	1387308
	230	1387302

Table 6: Custom search algorithm illustration for computing the optimal design in Phase III. For P_1, $G = 1000000,\ S = 0.7,\ c_0 = 80000,\ c_2 = 30000,\ c_3 = 50000,\ N_2 = 50,\ \gamma_2 = 200,\ N_3 = 214,\ \gamma_3 = 800,\ c_{int} = 2000$ and $\pi_{Ph3}(\theta) \sim N(30, 15)$, $K = 1$. The eNPV is displayed in £ and decimal places are preserved in order to demonstrate the maximal solution.

Optimal solution in group sequential Phase III trials

Recall that for Phase III trials we use the 'MAMS' package to compute the assurance. This package uses Monte-Carlo integration to estimate the integral (7), and the method introduced in Section 2.4.3 to estimate the assurance. Both of these add a layer of variation in calculating the eNPV for designs with $K > 2$. We simulate the eNPV in

P_1 with $K = 2$ for an increasing maximum sample size N_{3max} and plot the results in Figure 6. In this simulation, the curve is relatively smooth, however, there is noticeable variation which can result in inaccurate optimisation results. When computing the maximal eNPV design, the algorithm may recommend an N_{3max} that in fact does not yield the global maximum for the specific programme. Additionally, the results may no longer be convex, in which case the algorithm may find itself in local maxima, again returning a contextually erroneous result, i.e. N_{mid} finding itself at $N_{3max} = 656$ in any state of the algorithm (Figure 6). To resolve this issue, we must compute the expected sample size analytically which as previously discussed is a potential direction for future work.

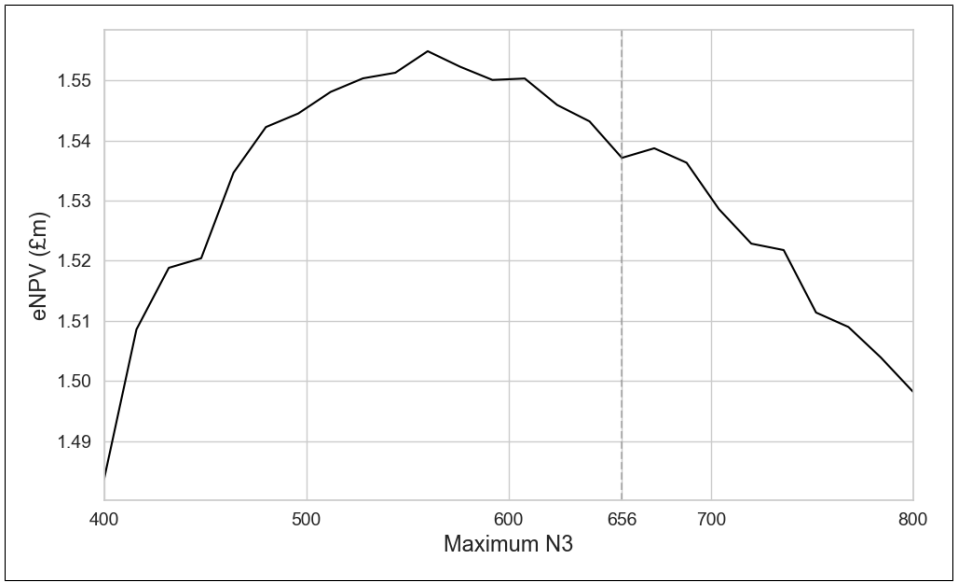

Figure 6: The eNPV in £m for increasing N_3. Programme configuration P_1, $G = 1000000$, $S = 0.7$, $c_0 = 80000$, $c_2 = 30000$, $c_3 = 50000$, $N_2 = 50$, $\gamma_2 = 200$, $N_3 = 214$, $\gamma_3 = 800$, $c_{in.t} = 2000$, $K = 2$. Reference line $N_3 = 656$ demonstrates issue of algorithm trap in local minima.

7 Portfolio optimisation

7.1 Programme selection

We have introduced the components required to optimise a single programme. Although we can refer to this as a portfolio with one available programme, it yields no additional complexity. With unlimited funds, a sponsor could design exhaustive Phase III trials for all programmes. In Figure 2.4.3 we showed how the decaying gain in power does not yield an efficient strategy and would deplenish eNPV with unnecessary subject recruitment. It follows that typically, a sponsor will have to make a choice out of a group of programmes to proceed with under certain constraints. In the simplest case, this could be a maximum budget that cannot be exceeded. However, we are free to apply any constraints according to other strategies. For example, we might wish to assign a lower initial budget

in anticipation of a promising programme at a later date.

In Section 6.2, we introduced how new data can be used to obtain a posterior or informed distribution. This means that the eNPV and optimal design of a programme must be re-caclulated. Concequently, the decision controlling the portfolio will accept updated metrics to select programmes. A portfolio with interim decision points can also result in a significant budget surplus due to early stopping of trials. Strategies that complement designs with multiple decision points are described in detail in Section 7.1.3. Using such a strategy, the sponsor would assign the budget progressively rather than all at once in the beginning of a trial.

7.1.1 Ranked assignment

Recall that we arrived at optimal Phase III trials via decision d_1. Suppose now the sponsor has available set of P programmes, $P \in \Omega_P$. Each programme has a corresponding prior for sampling in Phase II. In the simplest case, we define $d_2(P)$ to be the choice of programmes designed with maximum eNPV (15). Patel and Ankolekar (2014) suggest a straightforward approach for assigning resources. Assuming each progamme is optimally designed and the eNPV is available, the approach simply ranks the programmes by eNPV and allocates budget in descending order to depletion. One disadvantage of this approach is that it can opt into selecting fewer high return programmes with more expensive development costs over multiple smaller yields with smaller development costs but greater total eNPV. Such an instance is demonstrated in Table 7. Suppose a sponsor had available £0.6m to assign to a portfolio of trials. Using ranked assignment, the sponsor would select programmes A and F, with total eNPV of £1.75m. However, under the same budget constraint, one could choose programmes C, E and F, which would yield a total eNPV of £2.1m. Clearly, the latter would be the most beneficial, however, it is not always possible for a sponsor to pursue an unlimited number of trials. Other miscellaneous factors, such as the empoloyees required per trial, can influence or even dictate a decision.

Programme	Total cost	eNPV
1	0.50	1.30
2	0.35	1.20
3	0.30	1.00
4	0.25	0.80
5	0.20	0.65
6	0.10	0.45

Table 7: Arbitrary available programme pool with corresponding total cost required for a maximal design. Cost and eNPV displayed in £m.

7.1.2 Constrained optimization

To guarantee that decision d_2 yields the global maximum total eNPV and avoid the issue of ranked assignment, we use constrained optimisation. Linear Programming (LP) was invented to "optimise linear objective functions subject to a finite number of linear equality and inequality constraints" (Karloff, 2008). Following LP, Integer Programming (IP) was introduced to adapt LP for discrete parameters and functions. Early adaptations of IP include the widely known Travelling Salesman Problem (Applegate, 2006), and has demonstrated a plethora of flexible applications since. Combining LP and IP is known as Mixed Integer Linear Programming (MILP) which can include both integer and continuous parameters. Patel et al. (2013) implement a variety of MILP methods in the context of clinical trials from which inspiration in this section.

The 0-1 Knapsack Problem is a case of IP that selects items of some weight up to a maximum total weight. Items can be selected only once and in order to maximize an objective function (Du and Pardalos, 1998). Letting Z^* equal the indices of programmes one needs to select to maximize total eNPV, we write

$$Z^* = \arg\max_{Z} \left[\sum_i eNPV_i * Z_i \right]$$
$$\text{Subject to: } \sum_i d_i Z_i \leq B \quad \& \quad \sum_i Z_i \leq 1 \ ,$$

where the index i represents an individual programme, d_i the development cost, $Z_i \in \{0,\ 1\}$ denoting which programmes are selected, and B the available budget. The constraint $\sum_i Z_i \leq 1$ ensures that a programme can be selected only once, and $\sum_i d_{ij} Z_{ij} \leq B$ that the total development cost cannot exceed the available budget.

In our case, the sponsors objective is to maximize the total eNPV of a portfolio. Although we can compute the design of parameters N_3 and K that yield the maximum eNPV, it is important to note that it is not always the ideal strategy. In Section 6.3 we demonstrated how sample sizes can have little effect on eNPV near the optimal choice. To guarantee that the design of each programme selected in the portfolio leads to the portfolio maximum, we must search over a set of designs j specific to each programme i. Specific choices of N_3 and K may yield sub-maximal individual eNPV, but lead to a larger portfolio total. This is written,

$$Z^* = \arg\max_{Z} \left[\sum_i \sum_j eNPV_{ij} * Z_{ij} \right]$$
$$\text{Subject to: } \sum_i \sum_j d_{ij} Z_{ij} \leq B \quad \& \quad \sum_i \sum_j Z_{ij} \leq 1 \ .$$

The index j denotes the respective design specific to each individual programme i. From these indices, d_{ij} represents

the development cost of each design, $Z_{ij} \in \{0,\ 1\}$ denoting which specific design of a programme is selected and B is the available budget. The constraint $\sum_i \sum_j Z_{ij} \leq 1$ ensures that a programme and specific design can be selected only once, and $\sum_i \sum_j d_{ij} Z_{ij} \leq B$ that the total development cost cannot exceed the available budget. For simplification purposes, we define Ψ to be the set of programme names selected in the optimal solution.

Here, we have defined the simple case of a sponsor applying a budget constraint not to be exceeded by the total development cost of a portfolio. Extending this basic algorithm with additional constraints is not a troublesome process. i.e. one could introduce $\sum_i T_i \leq 1$ where T represents whether a treatment belongs to a specific therapeutic area. In this way we can restrict the algorithm in selecting a diverse portfolio. Patel et al. (2013) use a variety of Integer Programming (IP) solutions to optimise Phase III development portfolios with eNPV as the utility metric. A valuable addition in their research is extending to trial schedules in the design choice. For simplicity, we will assume that all programmes are available at each decision point and the interim portfolio evaluation is conducted when all programmes originally selected have completed Phase II. Stochastic modelling such as simulating optimistic/pessimistic scenarios at interim decision points is used to inform decisions. For example, a sponsor could decide to increase efforts in a programme that outperformed expectations or stop a programme early which would free up budget.

Algorithms & Efficiency

To find the optimal solution obtained from N items consider the following cases:

> **Case 0: Cannot include -** This is the base case, if the development cost of the N^{th} programme is greater than the available budget, then it cannot be included.
>
> **Case 1: Exclude -** The maximum obtainable eNPV of $N-1$ programmes by excluding the N^{th}.
>
> **Case 2: Include -** The total eNPV of the N^{th} programme and maximum eNPV obtainable from the remaining $N-1$ programmes and budget.

The algorithm iteratively evaluates these cases until all combinations have been explored. For the exhaustive algorithm, the time complexity is $O(2^N)$. Computation time grows exponentially and whereas this approach is feasible for small programme pools, it is recommended to use more sophisticated algorithms to improve performance. The main adaptation we make is memoisation, which caches expensive computations known as states, and reuses the result in future calls that reuse the same state (Baka, 2017). In practice, the total eNPV of different programme combinations is stored in an $N*W$ matrix. Solving the optimisation problem recursivley using memoisation becomes a dynamic programming solution (Bellman, 1966). Using this approach, the matrix allows us to look up states in

constant time, resulting in $O(N * W)$ time complexity.

With freedom of choice over N_3 we can limit the search of sub-maximal designs to reduce computation. For example, the sponsor may wish to set a threshold N_3 specific to each programme in order to mitigate the randomness of a small sample size. In our algorithm, we set the lower search limit to a sensible value for sample size corresponding to 80% of the maximum eNPV found in Section 6.3. We know that any additional recruiting beyond the optimal choice point will only increase development costs, so it suffices to be set as the upper limit.

We run the optimisation sequentially with the above specification and the resulting time in seconds for portfolios with increasing number of programmes as reported in Table 8. The process does not complete within 3600 seconds (10 hours) for portfolios with more than 12 programmes. To keep computation time reasonable, we will use a pool of 8 specified programmes when developing simulations.

Number of programmes	Total designs	Computing time (s)
4	68	0.19
5	87	0.75
6	128	6.49
7	169	16.55
8	188	35.57
9	208	92.28
10	227	160.05
11	242	528.59
12	256	1543.93

Table 8: Computation time for deriving the optimal portfolio selection for an increasing number of programmes. Total designs include all sub-maximal eNPV yielding N_3 designs. The optimisation is run sequentially using a 2.3 GHz 8-Core Intel Core i9 processor.

7.1.3 Real Option Valuation

In finance, a call option is an asset class which gives you the right to buy an asset at a specified price and future date Kodukula and Papudesu (2006). These are different to Real Options which instead of a financial asset can represent some tangible entity, in our case trials. In the case of financial options, the transaction is valid by paying a smaller fee at present. Although their definitions differ, a trial can be thought of as a financial asset. Modelling our problem as such, offers the option but not the obligation to take action in a programme in a future stage. Sequentially, this means choosing to invest in Phase II of a programme provides the option to later invest in Phase III. This brings forth the notion of spreading a fraction of the total budget over several Phase II programmes without committing to

their Phase III development costs. Assuming all the available programmes Phases are synchronised, once Phase II has been conducted we are free to re-evaluate the portfolio. The main difference to financial options is that instead of the option to buy an asset, we have a set of actions available for each programme (Trigeorgis, 1996):

Option to abandon - After completing a Phase II, a sponsor has the option to abandon a programme if the observed treatment effets are poor. The value of this option limits risk by concluding efforts in likely unsuccessful drug. Aside from being unsuccessful we mentioned how the market value is affected by the true performance of the treatment. If a treatment underperforms, the posterior will return a fraction of the market value as defined in the utility function (11). This happens as a lower mean and higher variance both increase the sample required for target power, but also a lower mean effect affects the market value through $\zeta(\theta)$.

Option to expand - If a treatment at the end of Phase II outperforms the expected effect. A sponsor has the option to expand, or increase efforts in the programme. Decisions could include increasing the sample size to achieve a greater power if for example the mean treatment effect is higher than expected but with larger variance. A case like this may require an optimal Phase III with a larger sample size increasing the development cost.

Option to differ - Deferral is an important option to a sponsor when they would like to delay the start of a clinical trial until there is more clarity. This could be when other programmes in the portfolio have completed. If a sponsor belives that another programme in the portfolio will be more expensive than anticipated they may wish to preserve budget. Alternatively if a programme cannot proceed due to budget, the sponsor may wish to delay until there are realizations in other programmes that free up budget.

In essence, Real Options evaluate the value of keeping options open in uncertain environments. We adapt this to create strategies recognising the flexibility of making non-binding initial programme investments. Further investment now depends on achieving specific milestones in the programme as well as informing decisions with observed data. Unplanned variation in budget is an important consideration. In Section 8, we perform possible re-optimisation with observed information subject to some conditions and evaluate different strategies leading to the motivation behind stochastic programming.

8 Simulations & Discussion

8.1 Programme pool & Sponsor resources

We have introduced a non-exhaustive variety of methods that are used in industry. In this section, we explore several thought-provoking examples of different programme realisations using our optimisation methods. The catalogue of available programmes and their respective parameter configurations used throughout these examples are listed in Table 9. To effectively demonstrate the effect of different strategies, we assign an available budget of £2m unless otherwise stated.

Significance levels are global to maintain coherence and comparability between programmes. We set target levels $\alpha_{Ph2} = 0.05$, $\beta_{Ph2} = 0.2$ and $\alpha_{Ph3} = 0.025$, $\beta_{Ph3} = 0.1$ for Phases II & III respectively. All programmes that are designed with a group sequential Phase III are configured with O'Brien Flemming boundaries ($\Delta = 0$). The futility significance for these programmes is set $b = \alpha = 0.025$. Individual parameter configurations were formulated in an effort to make programmes as diverse as possible. Programmes T to AI represent lower cost trials.

8.2 Portfolio with a single decision point

For this demonstration we consider the case where a sponsor would like to spend the entirety of the budget at once. This single decision point is defined before the start of Phase II for each programme. All Phase III trials for these programmes are configured with a group sequential design, $K = 2$. We assume that in this case the programme is carried out in its entirety and there is no sequential decision after Phase II. Restricting that the data collected in Phase II cannot be used in the confirmatory Phase III trial, we say $\pi_{Ph2}(\theta) = \pi_{Ph3}(\theta)$. The optimal Phase II design has been computed using the analytical sample size calculation (3) and design of Phase III follows the adapted Golden Search method introduced in Section 6.3. Recall that this results in the sample size to maximize the eNPV of a single programme but not the entire portfolio necessarily. We begin by choosing a diverse pool of similar cost category programmes consisting of those in Table 10.

Using these configurations, we compute the eNPV for each programme in the pool. The results are reported in Table 11. For the rest of this section we will summarise programme information to relevant quantities such as development costs per phase and eNPV.

As with all examples in this section we use an available budget of £2.00m. Using ranked assignment, we select $\Psi_1 = \{\text{E, P, H}\}$ amounting to a total eNPV of £7.46m with an initial investment of £1.63m. Applying the simple MILP algorithm without considering sub-maximal designs we compute the selection $\Psi_2 = \{\text{E, H, M, R }\}$ yielding total a eNPV of £8.96m with a total investment of £1.98m required. Using ranked assignment, the decision rule

	Treatment			Market value		Trial Design						
Programme	θ	σ	δ	G (£m)	S	c_0 (£k)	c_2 (£k)	γ_2 (£)	N_2	c_3 (£k)	γ_3 (£)	K (£k)
A	30	45	20	1.00	0.80	80	30	600	126	50	640	20
B	25	40	18	1.50	0.90	90	35	750	124	55	720	25
C	20	35	15	1.20	1.00	85	32	900	136	53	600	22
D	28	50	22	1.80	0.30	95	31	840	128	60	680	27
E	35	60	25	2.00	0.40	100	40	1050	144	65	760	30
F	22	48	19	1.60	0.50	88	37	825	158	52	696	24
G	27	43	17	1.40	0.70	86	31	675	160	48	648	26
H	33	55	23	1.70	0.30	92	38	960	142	62	736	28
I	19	39	16	1.30	0.20	83	35	795	148	56	692	23
J	32	50	21	1.85	0.40	97	36	870	142	59	712	27
K	26	42	20	1.55	0.30	91	34	825	110	57	664	25
L	21	37	18	1.45	1.30	89	33	840	106	54	624	24
M	29	46	22	1.70	0.30	93	35	900	110	60	672	26
N	31	53	24	1.90	1.20	98	37	990	122	61	712	29
O	23	38	16	1.25	0.40	84	31	735	140	53	664	25
P	34	60	26	2.10	0.30	105	41	1110	132	66	768	32
Q	20	36	17	1.35	0.20	87	33	780	112	52	680	24
R	25	44	19	1.65	0.40	92	34	870	134	57	720	26
S	28	48	21	1.75	0.30	95	36	930	130	61	696	27
T	22	40	16	0.55	1.30	41	21	480	156	31	416	17
U	20	37	14	0.54	0.80	39	21	450	174	31	400	16
V	24	43	17	0.57	1.20	42	22	480	160	32	424	17
W	26	45	19	0.59	1.30	44	24	495	140	33	440	18
X	22	40	15	0.52	0.30	41	21	465	176	31	416	17
Y	23	42	16	0.55	0.40	42	22	480	172	31	420	16
Z	19	36	13	0.51	0.50	40	20	450	190	30	408	16
AA	21	39	15	0.53	0.70	41	21	465	168	31	416	17
AB	25	44	18	0.57	0.30	43	23	495	148	33	432	18
AC	23	41	16	0.55	0.40	42	21	480	164	31	416	17
AD	18	35	13	0.50	0.50	39	20	450	180	29	408	16
AE	21	38	14	0.52	0.70	41	21	465	184	31	416	17
AF	24	43	17	0.55	1.10	42	22	480	160	32	424	17
AG	20	37	14	0.51	1.00	40	21	450	174	30	400	16
AH	26	45	19	0.58	1.40	44	24	510	140	34	440	18
AI	23	42	16	0.55	1.60	42	21	480	172	31	416	17

Table 9: Available programme catalogue. Treatment represents the prior distribution π_{Ph2} originally available. N_2 has been designed using the analytical sample size calculation (3).

Programme	Treatment			Market value		Trial Design							
	δ	θ	σ	G	S	c_0	c_2	γ_2	N_2	c_3	γ_3	N_3	K
E	25	35	60	2.00	0.40	100	40	1050	144	65	760	184	30
F	19	22	48	1.60	0.50	88	37	825	158	52	696	406	24
H	23	33	55	1.70	0.30	92	38	960	142	62	736	182	28
I	16	19	39	1.30	0.20	83	35	795	148	56	692	188	23
M	22	29	46	1.70	0.30	93	35	900	110	60	672	142	26
O	16	23	38	1.25	0.40	84	31	735	140	53	664	178	25
P	26	34	60	2.10	0.30	105	41	1110	132	66	768	170	32
R	19	25	44	1.65	0.40	92	34	870	134	57	720	174	26

Table 10: Programme pool of programmes with similar development costs. Market value G displayed in £m, all trial design costs c are displayed in £k and patient recruitment costs γ in £.

Programme	Total cost	eNPV
E	0.56	2.65
F	0.64	1.94
H	0.52	2.26
I	0.47	1.03
M	0.43	2.11
O	0.44	1.55
P	0.55	2.55
R	0.48	1.93

Table 11: Total cost and market value in £m for the maximum eNPV designs.

has mis-allocated resources, to the extent that $\sim 20\%$ of the budget remains unused. Using the simple Knapsack solution, the algorithm employs nearly the entire budget to obtain a solution with 20% greater total eNPV.

Finally, we apply the MILP algorithm that considers sub-maximal design and report the results in Table 12. The optimal choice using this method is $\Psi_3 = \{\text{E, M, P, R }\}$. The algorithm in this case has opted into lower sample sizes in programmes M, P, and R, with the main difference being swapping programme H with program P. Although programme P has a significantly lower sample size than its maximum eNPV design, it offers an overall larger contribution to the portfolio. The optimal choice Ψ_3 obtains an eNPV of £9.18m with £1.99m development costs. Comparing the optimal portfolio with the design that maximizes the eNPV for each programme in the selection we see that the maximal design is not an affordable choice with a total cost of £2.02m. We also see that decreasing the

sample size in P by 40 ($\sim 12\%$) only changes the assurance by 0.04 ($\sim 5.5\%$), resulting in a difference of £0.15m ($\sim 6\%$) in eNPV.

	Maximal design				Optimal design			
Programme	N_3	Assurance	Total cost	eNPV	N_3	Assurance	Total cost	eNPV
E	368	0.80	0.56	2.65	368	0.80	0.56	2.65
M	284	0.76	0.43	2.11	272	0.75	0.43	2.09
P	340	0.74	0.55	2.55	300	0.70	0.54	2.40
R	340	0.77	0.47	2.04	340	0.77	0.47	2.04
Ψ_3	–	–	2.02	9.35	–	–	1.99	9.18

Table 12: Comparing the N_3 design choice for programmes in the optimal portfolio Ψ_3 with the choice that maximizes the eNPV. Recruited patients quantities are not comparable between programmes and are omitted in the portfolio total. Total development costs and eNPV are displayed in £m. Portfolio N_3 and assurance totals are ommitted as they are non-comparable metrics.

8.3 Portfolio with multiple decision points

In practice, different trial realisations and true treatment effects are of great importance to the sponsor. In most cases, a sponsor will allow for multiple decision points in a portfolio of programmes to re-evaluate at interim points in time. For the following simulations recall that C_2 and C_3 is the cost incurred by the sponsor in order to proceed with the development of a treatment up to Phase II and Phase III trials respectively. In this section, we limit the programmes from Table 9 to a pool of 8 drugs chosen for their lower development costs comparatively with other programmes. All designs are configured with $K = 1$ to ensure exact analytical expectation. Using these programmes, we simulate different neutral, positive and negative effects in Phase II which are used to update the prior used in the design of Phase III trials, $\pi_{Ph3}(\theta)$. Using the respective distributions, we compute the Phase II and III designs that yield the maximum eNPV. All programmes have common C_2 which represents the Phase II prior used in the initial decision in the portfolio. The costs for the neutral, positive and negative cases are C_3, C_{3P} and C_{3N} respectively.

Consider the programme pool in Table 13 which is selected for their lower development costs comparative to others. For the neutral case, we let $\pi_{Ph2}(\theta) = \pi_{Ph3}(\theta) = \pi(\theta)$. With this true effect, the costs C_2 and C_3 simply reflect the optimal solution as designed in Section 8.2, where the prior in Phase III was not perimitted to use data from the previous phase. However, when a programme has negative results, the analytical solution to achieve the target power requires additional resources. Specifically the sponsor must recruit more patients than originally anticipated,

increasing C_3. Similarly, in the case of positive results, the cost of C_3 for the target power decreases, as the posterior would now require a smaller sample size.

	Neutral effect				Positive effect			Negative effect		
Programme	$\pi(\theta)$	C_2	C_3	eNPV	$\pi_{Ph3}(\theta)$	C_{3P}	eNPV	$\pi_{Ph3}(\theta)$	C_{3N}	eNPV
C	$N(20, 49^2)$	0.24	0.27	0.87	$N(24, 40^2)$	0.19	1.09	$N(17, 54^2)$	0.32	0.53
F	$N(22, 68^2)$	0.26	0.35	1.01	$N(26, 54^2)$	0.24	1.41	$N(19, 75^2)$	0.41	0.10
K	$N(26, 59^2)$	0.22	0.25	1.36	$N(31, 48^2)$	0.19	1.61	$N(22, 65^2)$	0.30	0.93
N	$N(31, 75^2)$	0.26	0.29	1.69	$N(37, 60^2)$	0.22	1.97	$N(26, 82^2)$	0.34	1.21
P	$N(34, 85^2)$	0.29	0.34	1.84	$N(41, 68^2)$	0.25	2.14	$N(29, 93^2)$	0.40	1.41
Y	$N(23, 59^2)$	0.15	0.23	0.27	$N(28, 48^2)$	0.16	0.39	$N(20, 65^2)$	0.27	0.15
AC	$N(23, 58^2)$	0.14	0.21	0.29	$N(28, 46^2)$	0.15	0.41	$N(20, 64^2)$	0.25	0.17
AH	$N(26, 64^2)$	0.14	0.20	0.34	$N(31, 51^2)$	0.14	0.46	$N(22, 70^2)$	0.24	0.21

Table 13: Simulated priors for respective neutral, positive and negative effects on a pool of programmes. C_2 only exists for the neutral effect of each programme as it is derived from the initial prior available to the sponsor. Costs C_3 are derived from the updated distribution $\pi_{Ph3}(\theta)$ thefore affecting the optimal sample size and eNPV. All development costs denoted C and eNPV are displayed in £m.

With multiple decision points, we adjust the original budget to a fraction of the available budget. Having to take the final portfolio decision at a later date, a sensible approach is to ensure there is sufficient capital remaining to invest in the entire original selection. For example, the development costs for the MILP solution in the first decision are now C_2. In this case, a budget of £2m would simply select all programmes, however, the total cost required if the programmes have neutral results is £3.84m, nearly twice the available budget.

MILP extension

A sensible approach is to ensure that there is enough budget remaining after the selection of Phase II programmes so that, under neutral results, there is enough liquidity to proceed with Phase III trials for the entire portfolio. If we let the parameter B_R represent the remaining budget after the initial selection, we can simply introduce the constraint $B_R < C_{3Total}$ in the MILP solution, where C_{3Total} is the total Phase III costs for the portfolio Ψ. Applying the updated MILP algorithm we arrive at the optimal portfolio $\Psi_4 = \{$K, N, P, AH $\}$, with total eNPV of £5.24m and development cost £1.99m.

This is a small scale example, however, consider the case where all programmes have negative true effects after conducting Phase II. Using Table 13 we can compute the capital required for the worst case scenario. For Phase II, the sponsor would need an initial investment of £0.90m. For Phase III, even though the original investment

expectation was £1.09m, negative results in all programmes would require an investment of £1.27m. To manage such outcomes, the sponsor could limit the available budget in the MILP for liquidity in future decisions. Case scenarios are useful here as a sponsor can decide whether they want to prepare for the worst or make higher risk investments. We discuss an extension to this notion in Section 10.

8.4 Managing budget

Another valid strategy for the sponsor is to spread a fraction of the original budget over several Phase II programmes. This is done whilst explicitly aware that proceeding with the entire portfolio in Phase III is not affordable. In this section, we consider a lower cost programme pool with an available budget of £1.5m. For simplicity, we configure the programmes with $K = 1$ for analytical solutions and design both Phases II & III using the fixed sample size calculation (3) with $\pi_{Ph2}(\theta) = \pi_{Ph3}(\theta) = \pi(\theta)$. Let decision d_1 be the first of an indefinite number of sequential decisions available to the sponsor $d \in \mathbb{N}$. The initial portfolio is obtained using the MILP algorithm with a restricted budget of £1m to allow for future liquidity. The selected programmes in d_1 are $\Psi_5 = \{\text{AA, AB, AC, AD, AE, AF, AG, AH, AI}\}$ according to £0.99m of development cost and £2.55m of eNPV. Once Phase II has been conducted, the portfolio is up for re-evaluation using the updated information available under d_2 in Table 14, and available budget £0.51m.

	d_1				d_2		
Programme	$\pi(\theta)$	C_2	C_3	eNPV	$\pi_{Ph3}(\theta)$	C_3	eNPV
AA	$N(21, 55^2)$	0.141	0.083	0.336	$N(18, 61^2)$	0.078	0.252
AB	$N(25, 62^2)$	0.139	0.088	0.390	$N(30, 50^2)$	0.096	0.456
AC	$N(23, 58^2)$	0.142	0.084	0.367	$N(28, 46^2)$	0.093	0.430
AD	$N(18, 49^2)$	0.140	0.078	0.294	–	–	–
AE	$N(21, 54^2)$	0.148	0.082	0.327	$N(18, 59^2)$	0.082	0.269
AF	$N(24, 61^2)$	0.141	0.085	0.366	$N(29, 49^2)$	0.093	0.429
AG	$N(20, 52^2)$	0.139	0.080	0.320	–	–	–
AH	$N(26, 64^2)$	0.139	0.090	0.402	$N(31, 51^2)$	0.098	0.471
AI	$N(23, 59^2)$	0.146	0.085	0.360	$N(20, 65^2)$	0.079	0.291

Table 14: Prior and respective costs of low cost programme pool for sequential decisions d_1 and d_2. Decision d_1 represents the pool and maximal eNPV design considering the prior does not change in Phase III. Decision d_2 contains information on the maximal eNPV design considering different realisations and affected prior after d_1. All costs denoted C and eNPV are shown in £m with 3dp to preserve the detail of low development cost.

Real options analysis

At decision d_2, the sponsor now has a set of options in each programme: abandon, expand or differ. Taking a closer look at individual programme performance in d_2, AA, AE and AI, have performed poorly with significant drops in eNPV of up to $\sim 33\%$ in AH. In these cases, the sponsor can dictate individual decisions on an Ad Hoc basis. For example, one might find that these investments are no longer worthwhile and abandon development or deffer until other programs have concluded. Should one wish to expand, there are now increased C_3 costs to achieve the target power and must increase efforts in these programmes. On the other hand, programmes AB, AC, AF and AH have outperformed the original expectation. Naturally, a sponsor would like to keep developing these, since now the eNPV is $\sim 15\%$ greater accross these programmes. Interestingly, C_3 costs in the well performing programmes have increased in d_2 as this is the choice that maximises the eNPV. In this case, if one wanted to achieve the same eNPV as the neutral case in d_1, the required C_3 cost would be significantly lower.

As portfolios scale, choosing which programmes to expand or abandon is not always feasible. We apply the MILP algorithm to obtain the optimal portfolio using the updated C_3 costs in d_2. The resulting portfolio is $\Psi_6 = \{$AB, AC, AF, AH, AI$\}$, costing £0.46m of the available £0.51m, with an eNPV of £2.08m. The programmes that performed well in d_2 computed higher C_3 costs (Table 14). This is a good opportunity to apply the MILP algorithm considering sub-maximal designs. Doing so, we obtain the optimal portfolio $\Psi_7 = \{$AA, AB, AC, AE, AF, AH, AI$\}$, costing £0.51m with an eNPV of £2.53m. The flexibility of this approach is clear. Loosening the maximal design and allowing for design differences at the unit level, the entirety of the budget is used to maximise the portfolio total for a $\sim 22\%$ increase.

Details summarising the maximal and optimal portfolio solution are summarised in Table 15. In programmes AA and AE the optimal solution is the maximal design. However, in the rest of the programmes, reducing the sample size has little effect on the eNPV. In AH, a 40% decrease in sample size results in only a 3.4% decrease in eNPV. In addition, without considering sub-maximal designs, the portfolio total development cost would be £0.57m, thus exceeding the available budget. This strategy is only comparable with positive and negative simulations. If the programmes had neutral results, the optimal portfolio would be the same as that calculated in d_1.

Finally, having deffered programmes AD and AG in d_1 we have the option of including these in the decision d_2. Doing so, we obtain $\Psi_8 = \{$AB, AC, AD, AF, AG, AH, AI$\}$, including both of the defferred programmes. This solution uses the entire budget for an eNPV of £2.65m. Including these in the portfolio would lead to d_3 when the programmes conclude Phase II which means we should, as before, allow for some liquidity at this point. Alternatively, the sponsor could defer these until other programmes have concluded and there is additional budget available. This sequence of decisions evaluating Real Options continues indefinitely $d \in \mathbb{N}$.

	Maximal design			Optimal design		
Programme	N_3	C_3	eNPV	N_3	C_3	eNPV
AA	112	0.078	0.252	112	0.078	0.252
AB	105	0.078	0.451	79	0.067	0.440
AC	131	0.086	0.429	79	0.064	0.414
AE	122	0.082	0.269	122	0.082	0.269
AF	88	0.069	0.418	78	0.065	0.412
AH	135	0.093	0.471	81	0.070	0.455
AI	115	0.079	0.291	113	0.079	0.291
Ψ_7	–	0.57	2.58	–	0.51	2.53

Table 15: Comparing the N_3 design choice for programmes in the optimal portfolio Ψ_7 with the choice that maximizes the eNPV. Recruited patients quantities are not comparable between programmes and are omitted in the portfolio total. Development costs C and eNPV are displayed in £m.

9 Conclusions

This thesis addresses the challenge of optimising risk in pharmaceutical clinical trial portfolios using Bayesian methods. With the decline of the blockbuster drug model, pharmaceutical companies face growing pressure to make innovative, data-driven decisions under uncertainty. By applying Bayesian optimization, this work provides a framework that allows decision-makers to incorporate prior knowledge and continuously update predictions as trials progress, leading to more informed and adaptable strategies.

The methodology focused on improving trial design through power analysis and sample size optimization, ensuring robust results while managing costs. By integrating risk-reward trade-offs, the approach moved beyond individual trials, allowing for comprehensive portfolio-level optimization. This helped maximize potential returns while mitigating the risks of costly failures.

The results of this thesis focused on demonstrating how Bayesian optimization could improve decision-making in clinical trial portfolios. By calculating expected yield under different distributional assumptions using the eNPV, the work showed that companies could assess the viability of clinical programs more accurately. The optimization was carried out using a Mixed Integer Linear Programming (MILP) algorithm, allowing for flexibility in resource allocation and portfolio management across multiple clinical programmes.

The simulations demonstrate the effectiveness of the proposed Bayesian framework in optimizing portfolios with a variety of resource constraints and multiple interim decision points. Through a series of simulated scenarios,

the benefit of considering sub-maximal design Phase III designs was demonstrated. Also, a case was made for the flexibility offered by incorporating multiple decision points. Spreading budget in early decisions to diversify investments provides the option but not the obligation to sequentially reallocate resources to the most promising programs. Ultimately, the results validate the benefit of dynamic portfolio optimization in an effort to protect the sponsor against unanticipated programme performance, and manage costs within constrained resources.

A key contribution of this thesis lies in delivering abstract metrics like programme utility and portfolio yield intuitively and with clear applicability. Through clear examples with simulated data, the work bridged Bayesian concepts with real-world scenarios, providing a practical guide for researchers and industry professionals. In the broader research landscape, this thesis serves as an intuitive introduction to the use of Bayesian approaches in clinical trial management to offer flexible and adaptive solutions.

10 Future work

Some extensions have been mentioned in brief alongside the presented content. In this section, we introduce fitting extensions in greater detail.

Group sequential calculation

In section 3.2, we introduced the expression that needs to be evaluated to compute the expected sample size of a group sequential design analytically. Jennison and Turnbull (1999) details this process using numerical integration techniques as in 4.1 for root solving calculations with a static treatment effect θ, whereas (Peck, 2020, Chapter 3.B Appendices) extends this to a Bayesian paradigm. This would be the first item in any future work.

Future work that is of great interest includes developing a solution to otimising the entire group sequential design. What this means essentially is computing the optimal decision at each interim analysis using the same approach as in 6.3. When the problem is small, we are able to compute this decision for a sequence of decisions. In problems where there are multiple decision points, non-conjugate priors, portfolio of drgus and other possibilities, computing this decision requires sophisticated programming. This naturally comes at a great computational expense. A paramount attribute of canonical distribution (6) is that the test statistics form a Markov Chain. This property facilitates efficient storing of computations to transition to the next state by considering both the current state and the eNPV of the sequence of future states (Neumann, 1993). Peck (2020) details his solution with the objective of knowing the optimal decision for all interim states by using Bellman's dynamic programming equation to solve a high-dimensional problem by breaking it down into a series of sub-problems and solving from the final state recursively (Bellman, 1966). Paired with a Markov Chain Monte Carlo algorithm with certain performance boosting features

such as splines and coupling, Peck extends this solution to a portfolio of programmes.

Another noteworthy direction that uses much of this additional literature is sample size re-estimation. Such solutions use interim results in sequential designs to re-evaluate the required sample at interim points in time. Jennison and Turnbull (1999) serves as a complete handbook in implementing such designs. Updating the maximum sample size at interim decisions would be an additional consideration in the optimisation model.

Portfolio risk level

In our problem formulation we consider the costs of a programme to be known. There is no uncertainty around what recruitment, setup, interim costs will be when the time comes to invest in a programme. Farid et al. (2021) formulate a solution using Chance Constrained Programming (CCP) which is a stochastic optimization approach to handle the uncertainties in the cost. CCP ensures that budget constraints are respected with a certain probability level. The method includes Monte Carlo simulations to estimate the costs and revenues under uncertainty and selects a portfolio selection via a MILP algorithm similar to this thesis. A solution of this type would yield Value-at-Risk metrics and offer a simple way of selecting portfolios for a sponsor.

Stochastic simulation

In Section 8.4 we introduced the notion of an indefinite number of sequential decisions d available to the sponsor. At each decision, we simulate multiple market indicators as we did in Section 8.4. By simulating quality of results, not only in 3 categories as in our case, we can fine tune the optimization model to evaluate portfolio performace over a vast size of market conditions. This allows the sponsor to select programmes that perform well over a variety of different outcomes. Rogers, Maranas and Ding (2005) introduces such programming methods in the context of clinical trials and is implemented by Graham, Jaki and Harbron (2020) in order to compare with another portfolio optimisation technique 'Project Scheduling'. Incorporating programme lifetime and phase duration adds a dynamic layer of optimising with a parallel agenda of purely maximising eNPV of portfolio designs.

To fully implement this stochastic optimization approach, future research must include:

Extensive Scenario Simulation - Expanding beyond the current categorization, we must simulate trial outcomes under more detailed and variable market conditions. This could mean simulating an exhaustive set of indicators for each Phase realisation. If at a decision point d_i we had N programmes and considered Q indicators for each, we would need to simulate a total of $i * N * Q$ states of programmes (Graham, Jaki and Harbron, 2020).

Dynamic Programming - Each of the simulated states are computed sequentially according to the previous

decision. These are not necessarily computed on true observed data and can be simulated according to some sponsor expectation of how the market demographic will move (Colvin and Maravelias, 2008). For this we will most likely need to familiarize with previously mentioned Bellman's dynamic programming central equation Bellman (1966) to handle the increased computational load, particularly for larger portfolios with numerous parallel and sequential decision points.

Advanced utility function - We would need to enrich the utility function used in this work with more sophisticated performance indicators and risk measures that account for time-varying factors like market competition, patent life, and shifting regulatory landscapes. The Probability of Technical and Regulatory Success (PTRS) lays emphasis the need to demonstrate both safety and efficacy, but also run a feasible program and produce a competitive product (Wiklund et al., 2024). The main challenge here is to maintain a modern and relevant model.

By addressing these aspects, future work will enable the development of a more flexible, dynamic optimization model that better reflects the uncertainties and strategic goals inherent in a realistic clinical programme. Throughout this work we introduced an achievable example in the time available. Extending beyond this toy application offers the benefit of improving the real-world applicability of the multitude of considerations one has to make.

A Software

All programming components with mentioned exceptions, have been written manually and are available GitHub (hovjr, 2024). This includes custom classes to set up clinical programmes, numerical integration methods, custom search algorithms, code for producing plots and the various portfolio simulations. The data used is also available for reproduction of results.

References

Agha, L., Kim, S. and Li, D., 2022. Insurance design and pharmaceutical innovation. *American economic review: Insights*, 4(2), pp.191–208.

Antonijevic, Z., 2014. *Optimization of pharmaceutical r&d programs and portfolios: design and investment strategy*. Springer.

Applegate, D.L., 2006. *The traveling salesman problem: a computational study*, vol. 17. Princeton university press.

Armitage, P., McPherson, C. and Rowe, B., 1969. Repeated significance tests on accumulating data. *Journal of the royal statistical society: Series a (general)*, 132(2), pp.235–244.

Arrowsmith, J. and Miller, P., 2013. Phase ii and phase iii attrition rates 2011-2012. *Nature reviews drug discovery*, 12(8), pp.569–570.

Baka, B., 2017. *Python data structures and algorithms*. Packt Publishing Ltd.

Bellman, R., 1966. Dynamic programming. *science*, 153(3731), pp.34–37.

Berger, J.O., 2013. *Statistical decision theory and bayesian analysis*. Springer Science & Business Media.

Brigham, E.F. and Houston, J.F., 2013. *Fundamentals of financial management*. South-Western Cengage Learning.

CHMP, T.T., 2004. Committee for medicinal products for human use (chmp). *Group*.

Collier, R., 2011. Bye, bye blockbusters, hello niche busters.

Colvin, M. and Maravelias, C.T., 2008. A stochastic programming approach for clinical trial planning in new drug development. *Computers & chemical engineering*, 32(11), pp.2626–2642.

DiMasi, J.A., Grabowski, H.G. and Hansen, R.W., 2016. Innovation in the pharmaceutical industry: new estimates of r&d costs. *Journal of health economics*, 47, pp.20–33.

Dolan, J.G., Bordley, D.R. and Mushlin, A.I., 1986. An eualuation of clinicians' subjective prior probability estimates. *Medical decision making*, 6(4), pp.216–223.

Du, D. and Pardalos, P.M., 1998. *Handbook of combinatorial optimization*, vol. 4. Springer Science & Business Media.

Farid, M., Chaudhry, A., Ytterstad, M. and Wiklund, S.J., 2021. Pharmaceutical portfolio optimization under cost uncertainty via chance constrained-type method. *Journal of mathematics in industry*, 11(1), p.3.

Food and Drug Administration and others., 1998. Guidance for industry: E9 statistical principles for clinical trials. *Food and drug administration: Rockville, maryland, usa*.

Gautam, A. and Pan, X., 2016. The changing model of big pharma: impact of key trends. *Drug discovery today*, 21(3), pp.379–384.

Ghosh, M., 1988. Stastical decision theory and bayesian analysis.

Graham, E., Jaki, T. and Harbron, C., 2020. A comparison of stochastic programming methods for portfolio level decision-making. *Journal of biopharmaceutical statistics*, 30(3), pp.405–429.

Guidance, F.D., 2010. Adaptive design clinical trials for drugs and biologics. *Biotechnol law rep*, 29(2), p.173.

hovjr, 2024. Bayesian-risk-optimisation. GitHub. Available from: https://github.com/hovjr/Bayesian-Risk-Optimisation. [Accessed 09 September 2024].

ICH Steering Committee and others., 1998. Statistical principles for clinical trials (e9). *International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. international conference on harmonisation.*

Jaki, T., Pallmann, P. and Magirr, D., 2019. The r package mams for designing multi-arm multi-stage clinical trials. *Journal of statistical software*, 88, pp.1–25.

Jennison, C. and Turnbull, B.W., 1999. *Group sequential methods with applications to clinical trials.* CRC Press.

Jørgensen, J.T., 2008. Are we approaching the post-blockbuster era? pharmacodiagnostics and rational drug development. *Expert review of molecular diagnostics*, 8(6), pp.689–695.

Karloff, H., 2008. *Linear programming.* Springer Science & Business Media.

Kodukula, P. and Papudesu, C., 2006. *Project valuation using real options: a practitioner's guide.* J. Ross Publishing.

Lachin, J.M., Matts, J.P. and Wei, L., 1988. Randomization in clinical trials: conclusions and recommendations. *Controlled clinical trials*, 9(4), pp.365–374.

Lewis, F.I., 2023. An introduction to group sequential methods: planning and multi-aspect optimization. *arxiv preprint arxiv:2303.01040.*

Malik, N., 2007. Has the era of blockbuster drugs come to an end? Available from: www.biopharminternational.com/view/has-era-blockbuster-drugs-come-end. [Accessed 13 April 2024].

Neumann, K., 1993. Dynamic programming basic concepts and applications. *Optimization in planning and operation of electric power systems: Lecture notes of the svor/asro tutorial thun, switzerland, october 14–16, 1992.* Springer, pp.31–56.

O'Brien, P.C. and Fleming, T.R., 1979. A multiple testing procedure for clinical trials. *Biometrics*, pp.549–556.

O'Hagan, A., Stevens, J.W. and Campbell, M.J., 2005. Assurance in clinical trial design. *Pharmaceutical statistics: The journal of applied statistics in the pharmaceutical industry*, 4(3), pp.187–201.

Pallmann, P., Jaki, T., Maclennan, G., Campbell, M. and Jansen, J., 2017. A bayesian group-sequential design in emergency care: the uk-reboa trial. Available from: `https://doi.org/10.13140/RG.2.2.30528.12807`.

Pampallona, S. and Tsiatis, A.A., 1994. Group sequential designs for one-sided and two-sided hypothesis testing with provision for early stopping in favor of the null hypothesis. *Journal of statistical planning and inference*, 42(1-2), pp.19–35.

Patel, N.R. and Ankolekar, S., 2014. *Optimization of pharmaceutical r&d programs and portfolios: design and investment strategy*. Springer.

Patel, N.R., Ankolekar, S., Antonijevic, Z. and Rajicic, N., 2013. A mathematical model for maximizing the value of phase 3 drug development portfolios incorporating budget constraints and risk. *Statistics in medicine*, 32(10), pp.1763–1777.

Peck, R., 2020. *Optimal decision making in drug development*. Ph.D. thesis. University of Bath.

Pocock, S.J., 1977. Group sequential methods in the design and analysis of clinical trials. *Biometrika*, 64(2), pp.191–199.

Press, W.H., 2007. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press.

Raiffa, H. and Schlaifer, R., 2000. *Applied statistical decision theory*, vol. 78. John Wiley & Sons.

Robert, C.P., Casella, G., Robert, C.P. and Casella, G., 2004. Monte carlo optimization. *Monte carlo statistical methods*, pp.157–204.

Rogers, M.J., Maranas, C.D. and Ding, M., 2005. Valuation and design of pharmaceutical r&d licensing deals. *Aiche journal*, 51(1), pp.198–209.

Thomas, D.W., 2016. Clinical development success rates 2006–2015. *Bio industry anal.*, 1, p.16.

Trigeorgis, L., 1996. *Real options: Managerial flexibility and strategy in resource allocation*. MIT press.

Tygstrup, N., Lachin, J.M. and Juhl, E., 1982. *The randomized clinical trial and therapeutic decisions*. Marcel Dekker Inc.

Wiklund, S.J., Thorn, K., Götte, H., Hacquoil, K., Saint-Hilary, G. and Carlton, 2024. Going beyond probability of success: Opportunities for statisticians to influence quantitative decision-making at the portfolio level. *Pharmaceutical statistics*.