2023

EARLY TERMINATION IN PHASE II CLINICAL TRIALS: ADMISSIBLE DESIGNS USING DECREASINGLY INFORMATIVE PRIORS

Chen Wang
Virginia Commonwealth University

EARLY TERMINATION IN PHASE II CLINICAL TRIALS: ADMISSIBLE DESIGNS USING DECREASINGLY INFORMATIVE PRIORS

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by
Chen Wang

Committee Members:

Roy T. Sabo, Ph.D., Associate Professor, Department of Biostatistics (Director)
Edward Boone, Ph.D., Department of Statistical Sciences and Operations Research
Nitai Mukhopadhyay, Ph.D., Department of Biostatistics
Robert A. Perera, Ph.D., Department of Biostatistics
Andrew S. Poklepovic, M.D., Department of Internal Medicine

Virginia Commonwealth University
Richmond, Virginia
August, 2023

Acknowledgements

I want to start by expressing my gratitude to my dissertation advisor, Dr. Roy Sabo, for his encouragement and guidance throughout my research in this program. He is always here guiding me on the right path, supporting me solving problem, praising any efforts I have ever made, and saving me from depression. I feel lucky, even blessed to be a student of Dr. Sabo. His great example of a leader and mentor will be haunting in my future professional life. Additionally, I want to give great thanks to my committee members - Dr. Edward Boone, Dr. Nitai Mukhopadhyay, Dr. Robert Perera, and Dr. Andrew Poklepovic - for their constructive feedback and guidance on this dissertation.

I also want to thank Dr. Leory Thacker for the experience and knowledge he provided in Biostatistical Consulting Lab. And thank all the Biostatistics department professors for their dedication to make me harvest in this program. Likewise for the very special colleagues in the Biostatistics department: Rebecca Rasnick, Dongho Shin, Atika Farzana Urmi, and Jing Zhang. Thank them for being here for me through these unforgettable years and for keeping me mentally healthy. We are not only taking courses together, but also conquer through all the frustrations and challenges in a closed friendship.

Thanks to my husband, mom, dad, mom in law, and dad in law for their continued yearly supports in helping me care for the kids, so I can devote myself to the study. Finally, special thanks to my adorable daughters - Erica Shao and Erin Shao; being their mom makes me strong.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract

EARLY TERMINATION IN PHASE II CLINICAL TRIALS: ADMISSIBLE DESIGNS USING DECREASINGLY INFORMATIVE PRIORS

By Chen Wang

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2023.

Director: Roy T. Sabo, Ph.D.,

Associate Professor, Department of Biostatistics

In Phase II clinical trials, the efficacy and futility of a new treatment are evaluated to determine whether it warrants further investigation in a larger-scale confirmatory Phase III trial. Ethically, Thall and Simon's Bayesian posterior probability design is commonly implemented in Phase II clinical trials to allow for an early termination in case of evidence of sufficient efficacy or a lack of futility based on posterior probability [1][2]; this in turn requires a pre-selected prior distribution based on known clinical opinion or historical information which is directly related to statistical decision making. Moreover, this Bayesian approach can result in an issue of inflating type I error rate by monitoring interim data to inform early termination decisions. Alternatively, Bayesian approach with the decreasingly informative prior (DIP), which is an informative yet skeptical prior, can be implemented to overcome the contentious prior selection and constrain the prior from providing evidence that would favor termination at early phase of a trial, but more adaption and informed by the observed data as more subjects are accrued, and consequently guarantee the control of the type I error rate.[3]

We apply the Bayesian DIP approach to one-parameter and two-parameter models of Phase II clinical trials, and aim to calculate the required smallest sample size, stopping decision cutoffs, the expected sample size, and the exact power and type I error rate, given an admissible target power and significant level. We facilitate the DIP construction by utilizing the prior effective sample size (ESS) and functionalize the prior ESS in terms of the nonaccrued sample size and center the prior distribution at some null values. For implementing the Bayesian DIP approach to two-parameter models, we extend the expected local-information-ratio (ELIR) approach, which is used for one-parameter models for determining prior ESS, for single-parameter in multivariate cases. Simulation comparing the performance of the standard Bayesian approach and DIP approach in both one-parameter and two-parameter models show that the DIP approach requires fewer patients when admissible designs are achieved; otherwise, the DIP approach controls the type I error and type II error rates with comparable or fewer sample size. We also build an R package, **BayesDIP**, that accommodates the admissible designs for both standard Bayesian approach and DIP approach for one-parameter and two-parameter models of Phase II clinical trials.

CHAPTER 1

INTRODUCTION

1.1 Background

Phase II clinical studies typically focus on determining whether a treatment has sufficient evidence of preliminary efficacy to warrant further investigation in a Phase III trial, or whether the investigation should be discontinued due to unacceptable safety or a lack of efficacy. These studies tend to be small, and data monitoring tends to occur continuously as subjects accrue. For ethical and financial considerations, Phase II trials often allow early termination for efficacy, safety or futility if early results are extreme. In the case that a trial is terminated early due to efficacy, the intervention will progress to Phase III trial sooner and more subjects will benefit from the treatment. In the case where evidence suggests the trial is unlikely to achieve its objectives, it can be terminated early for futility, as there is no reason to continue enrollment.[4],[5] While traditional frequentist methods provide stopping rules or termination guidance in Phase II trials at a small number of fixed time points, Bayesian methods (Thall and Simon [1],[2]) allow for more continual monitoring as well as the inclusion of prior or historical information, which may help improve decision making. Bayesian approaches are also more amenable to adaptive designs and complex modelings.[6]

1.1.1 Bayesian Approaches to Early Termination

In early termination research, several approaches are considered [7]: traditional frequentist approach (e.g., Pocock group sequential designs, O'Brien-Fleming alpha-spending function), Simon's two-stage design, Lee and Liu Bayesian predictive posterior probability design [8], and Thall and Simon Bayesian posterior probability design [1][2]. In a two-sample trial, the

Pocock group sequential design splits subjects into K equally-sized and sequential groups. For each group, an equal number of subjects are allocated to each treatment. After the kth group is observed, we calculate test statistic for the kth interim analysis. We decide to terminate the trial for efficacy if the test statistic is greater than a constant critical value; otherwise we continue to enroll the next group, and continue in this manner until (i) we terminate or (ii) the predetermined total number of subjects are accrued. Instead of using a constant critical value for all the K groups, the O'Brien-Fleming approach sets the test stopping boundaries progressively as more groups of data are collected. Simon's two-stage design is frequently used in single-sample designs with a binary outcome, splitting the sample into two stages. In stage I, a fixed n_1 subjects are accrued and observed, and if the number of responses (y_1) is low, we terminate the trial and claim the treatment is unpromising. If a sufficient number of responses is observed at stage I, an additional fixed n_2 subjects are accrued in stage II. At the end of the trial, if the total number of responses is low, then we claim the treatment is unpromising, otherwise we claim the treatment is promising.

Bayesian approaches allow the inclusion of prior or historical information, which may improve statistical decision making. These Bayesian approaches are also more amenable to adaptive designs and complex modeling.[6] Despite these benefits, the Bayesian approach can be subject to inflated type I error rates. These approaches can be based on either posterior probabilities or predictive posterior probabilities, the latter of which refer to the future observations of data when we have observed the enrolled subjects, though we will not investigate that possibility here.

In Thall and Simon's one-sample Bayesian approach with binary outcomes [1], suppose we have a likelihood function $f(y|\theta)$, with each $y_i \in \{0, 1\}$ representing a failure or a success, and prior distribution $\pi(\theta)$ as beta distribution $beta(a, b)$ for simplicity. Let θ_1 be a parameter(s) representing efficacy in a new treatment, and let θ_0 reflect null levels representing the boundary between an efficacious and non-efficacious treatment. Then, the hypotheses we

are testing are

$$H_0 : \theta_1 \leq \theta_0 + \delta_0 \tag{1.1}$$
$$H_1 : \theta_1 > \theta_0 + \delta_0$$

where δ_0 is a fixed targeted improvement for the new treatment to achieve (which could be 0). Note that these hypotheses assume that larger values of θ_1 are reflective of greater efficacy; we could simply switch the directions of the inequalities (and likely select a negative δ_0) if lower values imply greater efficacy. We also set predetermined upper and lower boundaries for the posterior probability, denoted as p_s and p_f respectively, representing the probabilistic thresholds needed to be met in order to terminate the trial for superiority or futility, respectively. Throughout the trial, we can decide to terminate for efficacy if the evidence is promising ($P(\theta_1 > \theta_0 + \delta_0|y) \geq p_s$) or terminate for futility if the evidence is unpromising ($P(\theta_1 > \theta_0 + \delta_0|y) \leq p_f$), and we continue the trial and enroll additional subjects if the evidence is inconclusive ($p_f < P(\theta_1 > \theta_0 + \delta_0|y) < p_s$). These probabilities can be estimated and the resulting decisions can be made after each new subject is enrolled and observed until the new treatment is determined as being either efficacious or futile, or when all the predetermined total number of subjects are recruited. The posterior probabilities could also be calculated after cohorts of patients are accrued and observed.

Suggested by Thall and Simon [1], the chosen prior distribution $\pi(\theta)$ should be formulated informatively to reflect a practically useful design. Based on the concentration parameter $c_e = a+b$ in the prior, it is recommended that $a = c_e(\theta_0+\delta_0/2)$ and $b = c_e(1-(\theta_0+\delta_0/2))$ so that the mean equals $\frac{a}{a+b} = \frac{c_e(\theta_0+\delta_0/2)}{c_e} = \theta_0 + \delta_0/2$, corresponding to the most pessimistic view that the efficacy in a new treatment is on average identical to the null levels and the most optimistic view that the new treatment provides the targeted improvement $\delta_0/2$. Thall and Simon also suggested the range of $c_e \in [2, 10]$ and discussed that for the same θ_0 and δ_0, larger values of c_e correspond to a narrower 90% probability interval (W_{90}) of the dispersion

of $\pi(\theta)$, and the prior with smaller W_{90} is highly localized around its mean and more informative, whereas the prior with larger W_{90} corresponds to a dispersed prior distribution and less informative.[1]

1.1.2 Prior Effective Sample Size

Prior selection in Bayesian approach is crucial, because it is possible to generate posterior distributions that are strongly influenced by the priors.[6] Knowing the prior effective sample size (ESS) facilitates prior selection as it indicates the amount of information contained in the prior in an intuitive value (i.e. number of subjects-worth of information). Though several approaches are available (Morita[9],[10]), ESS can be determined using the *expected local-information-ratio (ELIR)* approach for standard one-parameter exponential family (Neuenschwander[11]). Letting $i(p(\theta))$ and $i_F(\theta)$ be the information of the prior distribution $p(\theta)$ and the expected Fisher information for one information unit,

$$i(p(\theta)) = -\frac{d^2 log\, p(\theta)}{d\,\theta^2}, \quad i_F(\theta) = -E_{y|\theta}\Big\{\frac{d^2 log\, f(y|\theta)}{d\,\theta^2}\Big\} \tag{1.2}$$

the ESS is defined as the expected ratio of the prior information to the Fisher information.

$$ESS_{ELIR} = E_\theta\{r(\theta)\} = E_\theta\Big\{\frac{i(p(\theta))}{i_F(\theta)}\Big\} \tag{1.3}$$

In the exponential family,

$$f(y|\theta) = h(y)\, exp\Big(\sum_{i=1}^{k} \eta_i(\theta) t_i(y) - c(\theta)\Big)$$

the information ratio for the natural parameter η can be written as $r(\eta) = i(p(\eta))/i_F(\eta)$. If we rewrite the prior and observed distribution as

$$p(\eta) = exp\{n_0 m_0 \eta - n_0 M(\eta)\}, \quad f(y|\eta) = exp\{y\eta - M(\eta)\} \tag{1.4}$$

we can have

$$i(p(\eta)) = -\frac{d^2 log\, p(\eta)}{d\,\eta^2} = -\frac{d^2\,(n_0 m_0 \eta - n_0 M(\eta))}{d\,\eta^2} = \frac{n_0 d^2 M(\eta)}{d\,\eta^2}, \quad \text{and} \tag{1.5}$$

$$i_F(\eta) = -E_{y|\eta}\Big\{\frac{d^2 log\, f(y|\eta)}{d\,\eta^2}\Big\} = -E_{y|\eta}\Big\{\frac{d^2\,(y\eta - M(\eta))}{d\,\eta^2}\Big\} = \frac{d^2 M(\eta)}{d\,\eta^2}. \tag{1.6}$$

Thus, based on the Equation 1.3, it follows that $ESS = E(r(\eta)) = n_0$.

For example, consider a Poisson distribution with a Gamma(a,b) prior for the mean λ. The exponential family of Poisson distribution can be written as $f(y|\lambda) = \frac{1}{y!}\, exp\{y\, log\lambda\, -\lambda\}$, whose natural parameter is $\eta = log\lambda$ so that $\lambda = exp(\eta) = M(\eta)$. This can equivalently be written as $f(y|\eta) = \frac{1}{y!}\, exp\{y\,\eta - exp(\eta)\}$, from which we know $i_F(\eta) = exp(\eta)$ (Equation 1.6). The exponential family of Gamma prior distribution can be expressed as $p(\lambda|a,b) = \lambda^{-1}\, exp\{-b\lambda + a\, log\lambda + a\, logb - log\,\Gamma(a)\}$, and an equivalent form of the natural parameter expression is $p(\eta|a,b) = exp(\eta)^{-1}\, exp\{-b\, exp(\eta) + a\,\eta + a\, logb - log\,\Gamma(a)\}$. Then, we can calculate the prior information $i(p(\eta)) = b\, exp(\eta)$ by Equation 1.5, and $n_0 = b$ by Equation 1.3.

Some ESS_{ELIR} values for common one-parameter exponential families are provided by Neuenschwander[11], and we can refer to the quantities straightforwardly in our Aim 1. However, the method introduced by Neuenschwander[11] is restricted to single-parameter models. We will extend this method to multivariate models.

1.1.3 Decreasingly Informative Prior Approach

In practice, the Bayesian approach can be contentious when prior information is based mainly on subject matter experts.[6] To exchange this type of subjectivity for assumptions more directly related to statistical decision making in clinician studies and trials, the decreasingly informative prior (DIP) (Sabo[3], Donahue and Sabo[12]) is considered, where null skepticism is explicitly incorporated into the prior in a manner that decreases its prior effective sample

size (ESS) as subjects accrue. This approach was initially introduced in a Bayesian response-adaptive allocation method for the $beta - binomial$ model by Sabo[3], where an informative yet skeptical prior is parameterized in a way that centers the prior distribution around the mode p_0, resulting in a DIP distribution $p_i \sim Beta(1 + p_0(N - n), 1 + (1 - p_0)(N - n))$ for both treatment groups. The choice of $beta - binomial$ conjugate model with DIP yields a posterior distribution for efficacy rate p_i, as shown below:

$$\begin{aligned} y_i|p_i &\sim Binomial(n_i, p_i), \quad i = 1, 2 \\ p_i &\sim Beta(1 + p_0(N - n), 1 + (1 - p_0)(N - n)) \\ p_i|y_i &\sim Beta(1 + y_i + p_0(N - n), 1 + n_i - y_i + (1 - p_0)(N - n)) \end{aligned} \tag{1.7}$$

where y_i is the observed number of successes out of n_i accrued patients for the i^{th} group. Note that p_0 can be vaguely modeled with a hyperprior or a chosen value from many sources such as historical studies. The net effect of the DIP prior formulation is that it restricts response-based adaptation early in a trial (since it centers the posterior mean at a null value), gradually permitting more adaptation as the overall Bayesian model transfers the total effective sample size from the prior to the likelihood. In this way, the posterior distribution is increasingly informed by observed data and less by the prior information as subjects are accrued. Sabo[3] compared the performance of the DIP method with the functional exponent-based Thall and Wathen (TW) method[13]. Both methods achieve a gradual lead-in to response-adaptive allocation weights and behave similarly in terms of power, sample size and variability.

Donahue and Sabo[12] extended the DIP approach to continuous outcomes, primarily in the *normal* conjugate model with unknown mean and variance by parameterizing the prior ESS to equal the unobserved sample size as follows:

$$
\begin{aligned}
y_i|\mu,\sigma^2 &\sim N\left(\mu,\sigma^2\right), \quad i=1,2 \\
\mu|\sigma^2 &\sim N\left(\mu_0,\frac{\sigma^2}{N-n}\right) \\
\sigma^2 &\sim Inverse-Gamma\left(N-n,\sigma_0^2\right)
\end{aligned} \tag{1.8}
$$

where μ_0 is the prior mean, $N-n$ is both the ESS that μ_0 is based on and the prior degrees of freedom on which σ_0^2, the prior mean of σ^2, is based. Toward the end of the trial, the posterior estimate of μ_n approaches $\bar{y}$ and the posterior estimate of σ_n approaches $\frac{n-1}{n}s^2$. By comparing various DIP methods, including the ESS DIP, linear DIP (functionalizing the prior variance with $\sigma^2 = \frac{c(n+1)}{N+1}$), exponential DIP (functionalizing the prior variance with $\sigma^2 = \frac{c(n+1)}{N+1}exp(\frac{cn}{N-p})$), and frequentist response-adaptive in two allocation equations: the moment-based allocation equation proposed by Zhang and Rosenberger[14] and the effect-size mapping allocation equation defined by Bandyopadhyay and Bhattacharya[15], the ESS DIP was proved to increase the allocation weights steadily as the trial progresses and offer the lower variability early in the trial. In comparing the behaviors of the ESS DIP approach and the other approaches, including the frequentist response-adaptive approach, the Bayesian approach with non-informative priors, and the balanced design approach, the ESS DIP approach maintained increased numbers of total responses with lower variability and greater power.

1.2 Specific Aims and Proposed Methods

1.2.1 Aim 1: Implement the Decreasingly Informative Prior in Single-Parameter Phase II Models

In order to reduce the chance of erroneously adapting trials too early in Bayesian early termination methods, we propose implementing a skeptical but informative prior - decreasingly

informative prior (DIP) - in one-parameter statistical models for Phase II clinical trials. The null skepticism is directly functionalized into the prior in a manner that decreases its prior effective sample size (ESS) as subjects accrue. In this way, the posterior distribution is increasingly informed by observed data and less by the prior information as subjects are accrued. We hypothesize that under admissible designs (those with at least 80% power and no more than 5% type-I-error rates), the DIP approach will have fewer or comparable subjects accrued, similar power, and better-controlled type-I-error, compared to the conventional Bayesian priors.

1.2.2 Aim 2: Extend the Prior Effective Sample Size Derivation to Multivariate models and Implement Decreasingly Informative Prior Approach in Two-parameter Phase II Models

For early termination under admissible designs, we first will extend the one-dimensional ESS_{ELIR} method to multi-dimensional ESS_{ELIR} of a single-parameter to help functionalize the DIP for multivariate models in Phase II early termination trials, such as $Dirichlet-Multinomial$ model, $Normal-Inverse-Gamma$ model, and $Weibull$ model with unknown scale and shape parameters, etc.. We then consider implement the DIP for continuous outcomes (e.g., $Normal-Inverse-Gamma$ model) and survival outcomes (e.g., $Weibull$ model). We hypothesize that compared to the conventional Bayesian priors, the DIP approach will require fewer or comparable subjects, have similar power, and better-controlled type I error for admissible designs (those with at least 80% power and no more than 5% type-I-error rates).

1.2.3 Aim 3: Build an R Package

We will build an R package, *BayesDIP*, which has two functionalities. First, this package can help to determine the total planned sample size N to achieve the admissible designs;

Second, it can estimate the smallest sample size under admissible Phase II trials for the following distributions: *Bernoulli*, *Poisson*, and *Normal* with the DIP or a traditional Bayesian prior. The specific function of interest will depend on the type of outcomes and the interested parameters in efficacy. Users can input target power and type-I-error rates, as well as lower and upper boundaries for posterior probabilities (used in deciding whether to terminate), in order to achieve an admissible design. All functions output the smallest sample size necessary to obtain the admissible design, the exact power, and the exact type I error.

1.3 Dissertation Format

Each of the following chapters is written as individually distinct manuscripts. Therefore, they do not form a continuous narrative.

CHAPTER 2

EARLY TERMINATION IN SINGLE-PARAMETER MODEL PHASE II CLINICAL TRIAL DESIGNS USING DECREASINGLY INFORMATIVE PRIORS

This manuscript was published on International Journal of Clinical Trials on April 25, 2022.

2.1 Abstract

Background: To exchange the type of subjective Bayesian prior selection for assumptions more directly related to statistical decision making in clinician studies and trials, the decreasingly informative prior (DIP) is considered. We expand standard Bayesian early termination methods in one-parameter statistical models for Phase II clinical trials to include decreasingly informative priors (DIP). These priors are designed to reduce the chance of erroneously adapting trials too early by parameterize skepticism in an amount always equal to the unobserved sample size.

Method: We show how to parameterize these priors based on effective prior sample size and provide examples for common single-parameter models, include Bernoulli, Poisson, and Gaussian distributions. We use a simulation study to search through possible values of total sample sizes and termination thresholds to find the smallest total sample size (N) under admissible designs, which we define as having at least 80% power and no greater than 5% type I error rate.

Results: For Bernoulli, Poisson, and Gaussian distributions, the DIP approach requires fewer patients when admissible designs are achieved. In situations where type I error or power are not admissible, the DIP approach yields similar power and better-controlled type I error with comparable or fewer patients than other Bayesian priors by Thall and Simon.

Conclusions: The DIP helps control type I error rates with comparable or fewer patients, especially for those instances when increased type I error rates arise from erroneous termination early in a trial.

2.2 Introduction

Phase II clinical studies typically focus on determining whether a treatment has sufficient evidence of preliminary efficacy to warrant further investigation, such as in Phase III trials, or whether the investigation should be discontinued due to a lack of efficacy or safety. These studies tend to be small, and data monitoring tends to occur as subjects are accrued so that decisions on whether to stop the study early - for efficacy, safety, or futility - can be made as soon as possible, even before the planned end of the study.

While the traditional frequentist methods (e.g., Pocock group sequential designs, O'Brien-Fleming alpha-spending function, etc.) provide stopping rules or termination guidance in Phase II trials, Bayesian methods allow the inclusion of prior or historical information, which may help to improve decision making.[1],[2] The Bayesian approaches are also more amenable to adaptive designs and complex modelings.[6] Despite these benefits, the Bayesian approach can be subject to inflated type I error rates.

Further, the prior selection in Bayesian approach is crucial because it is possible to generate posterior distributions that are strongly influenced by the priors which is not desirable. In practice, the Bayesian approach can be contentious when prior information is based mainly on subject matter experts.[6] To exchange this type of subjectivity for assumptions more directly related to statistical decision making in clinician studies and trials, the decreasingly informative prior (DIP) is considered, where null skepticism is elicited into the prior in a manner that decreases its prior effective sample size (ESS) as subjects accrued.[3],[12],[16] In this way, the posterior distribution is increasingly informed by observed data and less by the prior information as subjects are accrued.

The goal of this paper is to develop and present the DIP approach based on the effective sample size for single-parameter models, include Bernoulli, Poisson, and Gaussian, and compare the DIP approach to the Thall and Simon's Bayesian approaches.[1],[2] The net effect of this DIP formulation is that it restricts response-based adaptation early in a trial, gradually permitting more adaptation as the overall Bayesian model transfers the total effective sample size from the prior to the likelihood. If applied to designs featuring early termination processes, this decreasingly informative prior could possibly help control type I error rates, especially for those instances when increased type I error rates arise from erroneous termination early in a trial.

This paper presents an alternative Bayesian approach to early termination in Phase II trials using DIP in single-parameter statistical models. Following a description of standard Bayesian early termination Phase II trial designs in Section 2.3.1, the rationale of DIP approach and the general model is detailed in Section 2.3.2. Examples of one-sample models, including Bernoulli, Poisson, and Gaussian distributions, are presented in 2.3.3. Simulation studies (Section 2.3.4) are used to compare the performance of the DIP approach with the standard Bayesian model, focusing on identifying admissible designs (those with at least 80% power and no more than 5% type-I error rates) and the minimum sample size that yields such designs. Section 2.5 concludes the paper with a discussion.

2.3 Statistical Methods

2.3.1 Standard Bayesian Early Termination Phase II Trials

In single-group Phase II studies and two-group Phase II trials, we often need to know if an experimental treatment is sufficiently efficacious relative to some threshold or the other treatment. Suppose we have a likelihood function $f(y|\theta)$ and prior distribution $\pi(\theta)$ for outcome vector y and scalar parameter θ. Let θ_1 be the parameter value representing efficacy in a new treatment, while θ_2 reflects either some null level representing the boundary between

an efficacious and non-efficacious treatment or the efficacy parameter in a comparison group. Then, the hypotheses we are testing are

$$H_0 : \theta_1 \leq \theta_2 + \delta_0,$$
$$H_1 : \theta_1 > \theta_2 + \delta_0 \tag{2.1}$$

where δ_0 is a fixed targeted improvement for the new treatment to achieve (which could be 0). Note that these hypotheses assume that larger values of θ_1 are reflective of greater efficacy; we could simply switch the directions of the inequalities if lower values imply greater efficacy. We also set upper and lower boundaries for the posterior probability, denoted as p_s and p_f respectively, representing the probabilistic thresholds needed to be met in order to terminate the trial for superiority or futility. Throughout the trial, we can decide to terminate for efficacy if the evidence is promising ($P(\theta_1 > \theta_2 + \delta_0|y) \geq p_s$) or terminate for futility if the evidence is unpromising ($P(\theta_1 > \theta_2 + \delta_0|y) \leq p_f$), and we continue the trial and enroll additional subjects if the evidence is inconclusive ($p_f < P(\theta_1 > \theta_2 + \delta_0|y) < p_s$). These probabilities can be estimated and the resulting decisions can be made after each new subject is enrolled and observed until the new treatment is determined as either efficacious or futile, or when all the predetermined total number of subjects are recruited. Note that posterior probabilities could be calculated after cohorts of patients are accrued and observed, though we will not investigate that possibility here.

2.3.2 Decreasingly Informative Prior

A decreasingly informative prior (DIP) is a skeptical prior that decreases in effective sample size (ESS) as a trial progresses. To that end, it incorporates both the predetermined total sample size and the current observed sample size in such a way that the unobserved sample size $N-n$ is made explicitly or approximately equal to the prior ESS in the prior distribution. The DIP is also parameterized in a way that centers the prior distribution at some value

or values that would reflect conditions of the null hypothesis (i.e. the new therapy is not efficacious). The basic steps for constructing a DIP are as follows:

1. Determine the prior ESS for a statistical model.
2. Functionalize the prior in terms of the observed sample size n and the planned sample size N (often $N - n$, the unobserved sample size) so that the prior ESS is N at the beginning of the trial and 0 at the end of the trial.
3. Center the prior distribution at some value reflecting the null hypothesis; which could come from a hyperprior.

Though several approaches are available, ESS can be determined using the *expected local-information-ratio* approach (Neuenschwander[11]). For example, given binary outcomes with response rate p and a prior $Beta(a, b)$, we know the mode of the prior is $\frac{a-1}{a+b-2}$, as well as the prior $ESS = a + b$. If we want the mode of the prior centered around p_0, the value from the null hypothesis, then we can set prior $ESS = a + b = N - n$ and $\frac{a-1}{a+b-2} = p_0$, and solve to get $a = 1 + p_0(N - n - 2)$ and $b = 1 + (1 - p_0)(N - n - 2)$. We could slightly alter the prior parameterization of a and b as $a = 1 + p_0(N - n)$ and $b = 1 + (1 - p_0)(N - n)$ so that the prior would be non-informative when the unobserved sample size is 0 at the end of the trial. Similarly, if we want the prior mean centered around a null value of p_0, then we let prior $ESS = a+b = N-n$ and $\frac{a}{a+b} = p_0$, and derive $a = p_0(N-n)$ and $b = (1-p_0)(N-n)$; we again set $a = 1 + p_0(N - n)$ and $b = 1 + (1 - p_0)(N - n)$ to make sure we have at least non-informative prior information when $n = N$ in the trial.

In clinical trials, the number of accrued subjects n is small at the beginning of a trial relative to the planned sample size N, making the prior ESS large (e.g. 2+N-n) and the DIP informative. Since the prior is skeptical and parameterized to reflect the conditions stated in the null hypothesis (with mean or mode set to θ_0), the resulting posterior distribution, with a low effective sample size n in the likelihood function, will be restricted from providing evidence in the form of posterior probabilities that would favor termination of the trial. As

the trial progresses, and the prior ESS is "transferred" to the likelihood via the increased observed sample size, the posterior distribution becomes increasingly more sensitive to the likelihood, and terminating the trial – if evidence for concluding as such is present – becomes more likely.

2.3.3 Examples

2.3.3.1 Bernoulli data with a Beta prior

Thall and Simon evaluate the efficacy of a new treatment based on *Bernoulli* outcomes where the interested parameter is the response rate.[1] In this case, we assume $\theta = p$ and we have the likelihood distribution $f(y|\theta) = f(y|p) = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}$. In the one-sample case, we temporarily assume a non-informative prior distribution $\pi(\theta) = \pi(p) \sim beta(1,1)$; we will relax this assumption in subsequent paragraphs. Let $\theta = p$ denote the response rate of the new treatment and $\theta_0 = p_0$ denote the null response rate (which could be taken as the standard or current rate), we can derive the posterior distribution of p based on the conjugate nature of the $beta-binomial$ pairing $p|y \sim beta(1+y, 1+n-y)$, where $y = \sum_{i=1}^{n} y_i$ is the total number of successes out of the n observed subjects in the trial.

Instead of assuming a non-informative prior, we could elicit an informative prior – as suggested by Thall and Simon – by setting the prior mean equal to $p_0 + \delta_0/2$ and selecting a value for concentration parameter c_e.[1] Thus, we can reparameterize the $beta(a,b)$ prior using $a = c_e(p_0 + \delta_0/2)$ and $b = c_e[1-(p_0+\delta_0/2)]$. Thall and Simon discuss several possible values, including low values of c_e (e.g., 2) representing a sparse prior distribution, and larger values c_e (e.g., 10) representing an informative prior distribution localized around its mean.[1] The posterior distribution of p for an informative prior is then given by $p|y \sim beta(a+y, b+n-y)$.

With a binary outcome and conjugate *beta* prior distribution, an informative and skeptical DIP can be specified as $beta(1+p_0(N-n), 1+(1-p_0)(N-n))$, as discussed in Section 2.3.2. Combining the DIP with the *binomial* likelihood function, the posterior distribution

of p is

$$p|y \sim beta(1 + p_0(N - n) + y, 1 + (1 - p_0)(N - n) + (n - y)).$$

At the beginning of the trial, n and y are small and the posterior distribution of p is more centered at the prior mode p_0. As n and y become larger, the accrued data become increasingly more important while the prior information is decreasing in importance.

2.3.3.2 Poisson data with a Gamma prior

If the outcome in a clinical trial is the number of the events per subject, then a *Poisson* distribution with *Gamma* prior is a plausible choice for likelihood. One choice of the standard Bayesian prior distribution is a Jeffreys' non-informative *Gamma* prior (limiting case) $\lambda \sim Gamma(0.5, 0.001)$ and the posterior is $\lambda|y \sim Gamma(0.5 + y, 0.001 + n)$. To apply a DIP for the one-sample case with *Poisson* outcomes with mean event rate $\theta = \lambda$ and a prior $Gamma(a, b)$, we know the null mean (λ_0) of the prior is $\frac{a}{b}$ and the prior $ESS = b$ (Neuenschwander[11]). If we want the prior centered around its null mean, then set prior ESS $b = N - n$ and $\frac{a}{b} = \lambda_0$, and get $a = \lambda_0(N - n)$ and $b = N - n$. Then, functionalize a and b as $a = 0.5 + \lambda_0(N - n)$ and $b = 0.001 + (N - n)$ so that the prior would be non-informative when the unobserved sample size is 0 at the end of the trial. Thus, the DIP model for the count outcomes is defined as follows:

$$\begin{aligned} y|\lambda &\sim Poisson(\lambda) \\ \lambda &\sim Gamma(0.5 + \lambda_0(N - n), 0.001 + (N - n)) \\ \lambda|y &\sim Gamma(0.5 + \lambda_0(N - n) + y, 0.001 + N) \end{aligned}$$

In the DIP approach, when more subjects n are accrued in the trial, the skewness of the posterior distribution will depend more on the observed data instead of the prior information.

2.3.3.3 Normal data with known variance

For outcomes that could be modeled with a *Normal* distribution with variance s^2 known, we have the likelihood function $f(y|\theta) \sim N(\theta, s^2)$ with a normal prior $\theta \sim N(\theta_0, \tau^2)$. In this case, the prior $ESS = s^2/\tau^2$ (Neuenschwander[11]). In a one-sample clinical trial, we assume $\theta = \mu$ is the new treatment mean and $\tau^2 = s^2/n_0$, where n_0 is the prior ESS with a null-mean μ_0. For the given likelihood $y|\mu \sim N(\mu, s^2)$ and prior $\mu \sim N(\mu_0, s^2/n_0)$, the posterior distribution of μ can be written as

$$\mu|s^2, y \sim N\left(\frac{n_0}{n_0+n}\mu_0 + \frac{n}{n_0+n}\bar{y}, \frac{s^2}{n_0+n}\right) \tag{2.2}$$

The value of prior parameters n_0 determines the level of information contained in the prior and the contribution of the null mean. If n_0 is small and s^2/n_0 is large, the prior distribution is dispersed and less informative; when n_0 is larger, the prior distribution will be more tightly centered around the null mean and become more informative.

For the DIP model, we set a skeptical prior (centered at μ_0) as initially informative with $n_0 = N - n$. This formulation allows the information contained within the posterior distribution of μ to shift from the skeptical prior at the beginning of the trial to the likelihood function as subjects accrued. The DIP posterior distribution of μ is

$$\mu|s^2, y \sim N\left(\frac{N-n}{N}\mu_0 + \frac{n}{N}\bar{y}, \frac{s^2}{N}\right)$$

As the posterior mean in the Bayesian model is a weighted average of the prior mean μ_0 and the sample mean $\bar{y}$, this DIP formulation will cause the posterior mean to approximate the prior mean μ_0 early in a trial, and will increasingly approximate $\bar{y}$ as subjects accrued.

2.3.4 Simulation Templates

The goals of our simulation studies are to identify the smallest possible sample size N among admissible designs and to compare the DIP approach with other Bayesian approaches. We

define an admissible design as having at least 80% power and no greater than 5% type I error rate. If there are no admissible designs based on power, we default to selecting the combination of parameters that yield the highest power and best-controlled type I error. If there are no admissible designs based on type I error, we default to selecting the combination of parameters that yield the lowest type I error and at least 80% power. We will explore *Bernoulli*-, *Poisson*- and *Normal*-distributed outcomes.

In these simulations, the observed outcome y_i for each subject in each trial is randomly simulated from the probability density or mass function $f(y_i|\theta)$, where θ is based off the population-level values assumed for that trial. Each subsequent subject is recruited until the trial is stopped (for futility or efficacy) or the planned sample size N is reached. In all one-sample cases, the upper (efficacy) and lower (futility) decision boundaries are set to $p_s \in (0.80, 0.99)$ and $p_f \in (0.01, 0.10)$ respectively, and the total sample size $N \in (10, 100)$. For simplicity, we assume the target threshold δ_0 equals 0. We simulate the observed data and estimate the power and type I error for each combination of p_s, p_f and the planned total sample size N. We then select the smallest total sample size under the admissible power and type I error. Type I error is measured as the proportion of trials where the null hypothesis is rejected under the null hypothesis (e.g. $\theta_1 = \theta_0$ for one-sample case), while power is measured as the proportion of trials where the null hypothesis is rejected under the alternative hypothesis (e.g. $\theta_1 > \theta_0$). Each parameter setting is repeated in 1000 simulated trials. All simulations are coded using R 1.4.1717.[17] The random samples are generated with the same seed.

For the *Bernoulli* outcome when $\theta = p$, we assume the treatment group with higher response rate is more efficacious, and consider several models: a non-informative prior $beta(1, 1)$, informative prior $beta(a, b)$ with several choices of prior information $a + b = 2, 6$ or 10, and a DIP skeptical prior, as illustrated in Section 2.3.3.1. In the one-sample case, we consider null response rates of $p_0 = 0.1, 0.3, 0.5$, or 0.7, with the actual response rate for the

new treatment response rate p_1 set at $p_1 = p_0 + \delta$, where we range $\delta \in (0, 0.05, 0.10, 0.15, 0.2)$. The outcome for each subject is randomly generated from $Bernoulli(p_1)$.

For the *Poisson* outcomes where $\theta = \lambda$, we assume the lower values of event rates imply improved efficacy, and consider a Jeffrey's non-informative prior (limiting case) $Gamma(0.5, 0.001)$ and a decreasingly informative prior (DIP) stated in Section 2.3.3.2. We set the null event rate as $\lambda_0 = 0.5$ or 5, and define the new treatment event rate as $\lambda_1 = \lambda_0 - \delta$. When $\lambda_0 = 0.5$, we set $\delta \in (0, 0.05, 0.10, 0.15, 0.2)$; when $\lambda_0 = 5$, we set $\delta \in (0, 0.5, 1, 1.5, 2)$. Each subject is randomly generated from $Poisson(\lambda_1)$.

For the *Normal* cases where $\theta = \mu$ with known variance, we consider Bayesian models with $n_0 = 2, 6$, or 10 for Equation 2.2, as well as a DIP case. We study low-variance and high-variance cases reflecting our assumptions about the known variability. For each template, we set the null mean as $\mu_0 = 100$ and expect lower values to imply improved efficacy; thus the new treatment mean is defined as $\mu_1 = \mu_0 - \delta$, where we set $\delta = 0, 5$, or 10 and set $\delta_0 = 0$ for simplicity; we consider the low variability with $s = 15$ and the high variability with $s = 30$. Each subject is randomly generated from $N(\mu_1, s^2)$.

2.4 Results

Table 1 shows the simulation results for one-sample *Bernoulli* cases with a low response rate ($p_0 = 0.1$). Compared with different standard Bayesian approaches, the DIP approach always has better-controlled type I error, with a comparative or lower sample size. For some cases in which the standard Bayesian approach cannot achieve the admissible design on type I error, such as the case $p_1 = 0.2$, the DIP approach not only controlled type I error, but also has the smallest planned sample size. Results are similar for other cases ($p_0 = 0.3, 0.5, 0.7$) (see Tables S.1, S.2, and S.3 in the Appendix B).

The simulation results for the one-sample *Poisson* cases are shown in Table 2. Compared to the non-informative Bayesian approach, the DIP approach performs better in controlling

Table 1. Simulation Results for Bernoulli Cases - One Sample ($p_0 = 0.1$)

Model	p_0	p_1	Sample Size[a]	Futility	Efficacy	Power	Type I Error[b]
DIP	0.1	0.15	94	0.01	0.89	0.808	0.239
Bayesian ($Beta(1,1)$)	0.1	0.15	96	0.02	0.94	0.802	0.294
Bayesian ($a+b=2$)	0.1	0.15	92	0.01	0.85	0.804	0.342
Bayesian ($a+b=6$)	0.1	0.15	100	0.01	0.87	0.810	0.295
Bayesian ($a+b=10$)	0.1	0.15	98	0.03	0.85	0.805	0.301
DIP	0.1	0.20	76	0.10	0.98	0.802	0.050
Bayesian ($Beta(1,1)$)	0.1	0.20	88	0.02	0.99	0.868	0.076
Bayesian ($a+b=2$)	0.1	0.20	85	0.01	0.99	0.816	0.050
Bayesian ($a+b=6$)	0.1	0.20	90	0.01	0.99	0.800	0.051
Bayesian ($a+b=10$)	0.1	0.20	86	0.07	0.98	0.820	0.050
DIP	0.1	0.25	42	0.06	0.98	0.843	0.050
Bayesian ($Beta(1,1)$)	0.1	0.25	38	0.03	0.99	0.811	0.056
Bayesian ($a+b=2$)	0.1	0.25	52	0.09	0.99	0.816	0.050
Bayesian ($a+b=6$)	0.1	0.25	43	0.01	0.97	0.816	0.050
Bayesian ($a+b=10$)	0.1	0.25	38	0.04	0.96	0.808	0.050
DIP	0.1	0.30	22	0.02	0.98	0.801	0.050
Bayesian ($Beta(1,1)$)	0.1	0.30	24	0.08	0.99	0.818	0.052
Bayesian ($a+b=2$)	0.1	0.30	25	0.08	0.97	0.821	0.050
Bayesian ($a+b=6$)	0.1	0.30	25	0.03	0.96	0.826	0.050
Bayesian ($a+b=10$)	0.1	0.30	25	0.08	0.95	0.810	0.050

[a] The Planned Sample Size

[b] Type I error is calculated under the null hypothesis $p_1 = p_0$

type I error. Additionally, when the effect size is large, the DIP approach has a lower sample size and better-controlled type I error rate.

Table 3 show the simulation results for one-sample *Normal* cases. In both low and high variability settings, compared with different standard Bayesian approaches, the DIP approach has lower or comparative sample size when type I error is controlled (0.05). When the admissible type I error rate cannot be achieved, the DIP approach has a better-controlled type I error and comparable sample size.

Table 2. Simulation Results for Poisson Cases

Model	λ_0	λ_1	Sample Size[a]	Futility	Efficacy	Power	Type I Error[b]
DIP	0.5	0.45	100	0.10	0.80	0.695	0.424
Bayesian	0.5	0.45	98	0.02	0.80	0.793	0.587
DIP	0.5	0.4	99	0.03	0.87	0.804	0.253
Bayesian	0.5	0.4	80	0.05	0.90	0.807	0.355
DIP	0.5	0.35	98	0.02	0.96	0.803	0.075
Bayesian	0.5	0.35	97	0.06	0.97	0.831	0.138
DIP	0.5	0.3	68	0.07	0.98	0.806	0.050
Bayesian	0.5	0.3	86	0.06	0.99	0.845	0.058
DIP	5	4.5	98	0.03	0.96	0.802	0.071
Bayesian	5	4.5	99	0.09	0.97	0.802	0.147
DIP	5	4	29	0.03	0.97	0.808	0.050
Bayesian	5	4	37	0.02	0.99	0.802	0.050
DIP	5	3.5	12	0.09	0.96	0.819	0.050
Bayesian	5	3.5	14	0.03	0.97	0.832	0.050
DIP	5	3	10	0.03	0.95	0.945	0.050
Bayesian	5	3	10	0.04	0.96	0.931	0.050

[a] The Planned Sample Size

[b] Type I error is calculated under the null hypothesis $\lambda_1 = \lambda_0$

2.5 Discussion

In summary, we introduced the rationale of the DIP, applied the DIP to three formulations of early termination phase II trial designs (*Poisson*, *Bernoulli*, and *Normal*), and compared the performance to the Bayesian approaches by Thall and Simon using simulation studies. The results show that, for the three distributions and across all one-sample settings, compared to the traditional Bayesian approaches by Thall and Simon, the DIP approach requires fewer patients when admissible designs are achieved.In the designs where type I error or power are not admissible, the DIP approach yields similar power and better-controlled type I error with comparable or fewer patients than Thall and Simon's Bayesian approaches[1][2]. We also extend the one-sample case to two-sample cases, and the results

Table 3. Simulation Results for Normal Cases with Known Variance

Model	μ_0	μ_1	s	Sample Size[a]	Futility	Efficacy	Power	Type I Error[b]
DIP	100	95	15	61	0.07	0.98	0.802	0.050
Bayesian ($\kappa_0 = 2$)	100	95	15	71	0.03	0.99	0.814	0.050
Bayesian ($\kappa_0 = 6$)	100	95	15	74	0.07	0.99	0.805	0.050
Bayesian ($\kappa_0 = 10$)	100	95	15	67	0.02	0.98	0.808	0.050
DIP	100	90	15	19	0.06	0.97	0.869	0.050
Bayesian ($\kappa_0 = 2$)	100	90	15	17	0.04	0.97	0.926	0.050
Bayesian ($\kappa_0 = 6$)	100	90	15	16	0.06	0.95	0.816	0.050
Bayesian ($\kappa_0 = 10$)	100	90	15	17	0.07	0.95	0.821	0.050
DIP	100	95	30	100	0.09	0.88	0.805	0.204
Bayesian ($\kappa_0 = 2$)	100	95	30	98	0.08	0.92	0.802	0.272
Bayesian ($\kappa_0 = 6$)	100	95	30	94	0.01	0.92	0.80	0.249
Bayesian ($\kappa_0 = 10$)	100	95	30	98	0.05	0.91	0.803	0.246
DIP	100	90	30	60	0.05	0.97	0.811	0.050
Bayesian ($\kappa_0 = 2$)	100	90	30	73	0.10	0.99	0.819	0.050
Bayesian ($\kappa_0 = 6$)	100	90	30	73	0.01	0.99	0.805	0.050
Bayesian ($\kappa_0 = 10$)	100	90	30	67	0.10	0.98	0.809	0.050

[a] The Planned Sample Size

[b] Type I error is calculated under the null hypothesis $\mu_1 = \mu_0$

are presented in the Appendix B. For two-sample cases, it is concluded that the DIP approach performed better than Thall and Simon's Bayesian approaches for moderate to large response rates, but performed poorly with low response rates and low effect sizes.

It should be noted that the focus of this study is on identifying the smallest sample size to achieve an admissible design, defined by the commonly used thresholds of at least 80% power and at most 5% type I error. Changing the minimum power and maximum type I error rate might change our findings and conclusions, though these values are conventional. We also ignored admissible designs with a larger sample sizes, forfeiting designs with possibly higher power or lower type I error. While our choices for the predetermined sample size (N) are limited within the admissible design as having at least 80% power and no greater than 5% type I error, the choices for parameters settings in the simulations are broadly and comprehensively considered to reflect realistic scenarios. For each parameter set, we also

investigated a non-informative and three informative models ($\kappa_0 = 2, 6, 10$) in comparison with the DIP model.

While we elicited the DIP in the way that is not based on any historical or optimistic prior, the researchers can still explore other subjective priors at the end of the trial to determine the robustness of their findings. We also motivated the DIP approach using conjugate examples: *poisson-gamma*, *beta-binomial*, and *normal-normal* models. We can easily extend this to other prior-likelihood combinations, particularly those that lead to non-conjugate or intractable posterior distributions using MCMC approaches. The key of the DIP approach with a non-conjugate prior is to parameterize the prior so that its effective sample size equals $N - n$, which may require numerical or simulation-based determination[9][11]. In future work, we plan to extend the single parameter DIP model to cases with two or more parameters.

CHAPTER 3

PRIOR EFFECTIVE SAMPLE SIZE OF A SINGLE PARAMETER ON MULTIVARIATE CASES

3.1 Abstract

We extend the expected local-information-ratio (ELIR) approach for determining respective prior effective sample size (ESS) for single parameters in multivariate cases. The extension satisfies the predictive consistency criterion, as exemplified through cases of the Dirichlet-Multinomial and Normal-Inverse-Gamma models.

3.2 Introduction

Neuenschwander et al.(2020) addressed the importance of knowing the prior effective sample size (ESS) in clinical trial design. They compared different information-based methods for obtaining prior ESS for one-parameter exponential families and proposed an expected local-information-ratio (ELIR) method which shows the predictive consistency by taking the expected values of the ratio of prior and Fisher information. This work exclusively focused on the one-parameter exponential family, leaving the multivariate case an open question.[11][18] In Discussion papers, Biswas and Angers considered the determinant of the prior and Fisher information matrices to define a predictively consistent ESS for the multivariate case[19], but Neuenschwander et al. argued that the determinant version fails the predictive consistency criterion in the case of two-sample Normal distributions.[18] Biswas and Angers also mentioned a trace-based approach but did not give an analytical formula.[19] Neuenschwander et al. also stated in the rejoinder that there was no discussion of how to obtain a predictively consistent ESS of a single parameter θ_j for multivariate case.[18]

In this paper, instead of focusing on the overall ESS of the entire parameter vector $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$, we focus on the respective ESS of each single parameter θ_j on multivariate case by naturally expanding the ELIR ESS (Neuenschwander et al.[11]) from the univariate case. The desirable property of predictive consistency criterion in multivariate settings are evaluated in the examples of Dirichlet-Multinomial distribution and Normal-Inverse-Gamma distribution. The general definition of ELIR ESS in the multivariate case is introduced in Section 3.3. Examples of Multinomial distribution with Dirichlet prior and Normal-Inverse-Gamma distribution are presented in Section 3.4. Simulations are used in obtaining the prior and expected posterior ESS of Normal-Inverse-Gamma model, given in Section 3.5. Section 3.6 concludes the paper with a brief discussion.

3.3 Methods

The approach by Neuenschwander et al. (expected local-information-ratio, ELIR) to determine ESS is appropriate for one-parameter cases and is defined as the expectation of the ratio of the prior information to the Fisher information, given by[11]

$$ESS_{ELIR} = E_\theta\{r(\theta)\} = E\left\{\frac{i(p(\theta))}{i_F(\theta)}\right\}. \tag{3.1}$$

Supposing we have a multivariate model with likelihood function $f(y|\theta)$ with a joint prior distribution $\pi(\theta)$ for parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_n)$, we can extend Equation 3.1 to matrix form and define the multivariate ESS_{ELIR} as

$$ESS_{ELIR} = E_{\boldsymbol{\theta}}\{r(\boldsymbol{\theta})\} = E_{\boldsymbol{\theta}}\{i(p(\boldsymbol{\theta}))\,(i_F(\boldsymbol{\theta}))^{-1}\} \tag{3.2}$$

where ESS_{ELIR} is a $n \times n$ matrix, $i(p(\theta_1, \theta_2, ..., \theta_n))$ is the information matrix of the joint prior distribution, and $i_F(\theta_1, \theta_2, ..., \theta_n)$ is the Fisher information matrix for single subject. Similarly, we extend Section 2.6 in Neuenschwander et al.[11] to decide the expected posterior

ESS matrix form for a sample size of N under the prior predictive distribution, which is given by

$$E\big\{\big[i(p(\theta_1,\theta_2,\ldots,\theta_n))+Ni_F(\theta_1,\theta_2,\ldots,\theta_n)\big](i_F(\theta_1,\theta_2,\ldots,\theta_n)^{-1})\big\} \tag{3.3}$$

where $\boldsymbol{\theta}=(\theta_1,\theta_2,...,\theta_n)$ follows the joint posterior distributions. Note that instead of providing one general ELIR ESS for the whole parameter vector, we instead will focus on providing the respective ELIR ESS for each separate parameter.

For comparison, we include possible alternatives for approximating ESS by the variance ratio (VR) and precision ratio (PR) approaches. Due to complexity of joint posterior and prior formulations, we apply the univariate approaches separately for each parameter. For example, the variance ratio and precision ratio for θ_1 are

$$ESS^{\theta_1}_{VR}=\frac{E\{i_F^{-1}(\theta_1|\theta_2,...,\theta_n)\}}{Var(\theta_1)},\quad ESS^{\theta_1}_{PR}=\frac{Var^{-1}(\theta_1)}{E\{i_F(\theta_1|\theta_2,...,\theta_n)\}}$$

where $i_F(\theta_1|\theta_2,...,\theta_n)$ is the conditional Fisher information for a single subject.

3.4 Examples

Here we discuss two multivariate examples.

3.4.1 Multinomial model with Dirichlet prior

We first consider a conjugate prior distribution: the Dirichlet-Multinomial model with $\boldsymbol{\theta}=(p_1,\ldots,p_k)$. The density function of the Multinomial model of one trial with k categories is

$$f(y|\boldsymbol{\theta})=f(y_1,\ldots,y_k|p_1,\ldots,p_k)=\prod_{j=1}^{k}p_j^{y_j} \tag{3.4}$$

where $p_k=1-\sum_{j=1}^{k-1}p_j$ and $y_k=1-\sum_{j=1}^{k-1}y_j$. The probability of each category is $0<p_j<1$, and $y_j=1$ if category j is observed, otherwise $y_j=0$. The Fisher information matrix of

$f(y_1, ...y_k|p_1, ..., p_k)$ for one information unit is given by

$$i_F(p_1, \ldots, p_k) = \begin{bmatrix} 1/p_1 & \ldots & 0 \\ & \ddots & \vdots \\ 0 & & 1/p_{k-1} \end{bmatrix} + \frac{1}{1 - \sum_{j=1}^{k-1} p_j} J_{k-1} \tag{3.5}$$

Thus, we can derive the inverse of Fisher information by Binomial Inverse Theorem and get

$$i_F(p_1, \ldots, p_k)^{-1} = \begin{bmatrix} p_1 & \ldots & 0 \\ & \ddots & \vdots \\ 0 & & p_{k-1} \end{bmatrix} - \begin{bmatrix} p_1 \\ \vdots \\ p_{k-1} \end{bmatrix} \begin{bmatrix} p_1, \ldots, p_{k-1} \end{bmatrix} \tag{3.6}$$

With a conjugate Dirichlet prior for $\boldsymbol{\theta} = (p_1, ..., p_k)$, which is given by

$$\pi(p_1, \ldots, p_k|\alpha_1, ..., \alpha_k) = \frac{\Gamma(\alpha_0)}{\prod_{j=1}^{k} \Gamma(\alpha_j)} \prod_{j=1}^{k-1} p_j^{(\alpha_j - 1)} (1 - \sum_{j=1}^{k-1} p_j)^{\alpha - \alpha_j - 1} \tag{3.7}$$

where α_j are independent and $\alpha = \sum_{j=1}^{k} \alpha_j$. We can get the information of the prior distribution, which is

$$i(p(p_1, \ldots, p_k)) = \begin{bmatrix} \frac{\alpha_1 - 1}{p_1^2} & \ldots & 0 \\ & \ddots & \vdots \\ 0 & & \frac{\alpha_{k-1} - 1}{p_{k-1}^2} \end{bmatrix} + \begin{bmatrix} \frac{\alpha - \alpha_1 - 1}{(1 - \sum_{j=1}^{k-1} p_j)^2} & \ldots & 0 \\ & \ddots & \vdots \\ 0 & & \frac{\alpha - \alpha_{k-1} - 1}{(1 - \sum_{j=1}^{k-1} p_{k-1})^2} \end{bmatrix} \tag{3.8}$$

Then, with Equation 3.2, the multivariate ESS_{ELIR} of Multinomial distribution with Dirichlet prior is defined as

$$ESS_{ELIR} = E\{i(p(p_1, \ldots, p_k))(i_F(p_1, \ldots, p_k)^{-1}\} \tag{3.9}$$

We can get the diagonal elements of the Equation 3.9, which is

$$\begin{aligned} ESS_{ELIR}^{p_j} &= E\left\{\left(\frac{\alpha_j-1}{p_j^2}+\frac{\alpha-\alpha_j-1}{(1-\sum_{j=1}^{k-1}p_j)^2}\right)\left(p_j-p_j^2\right)\right\} \\ &= E\left\{\frac{(\alpha_j-1)(1-p_j)}{p_j}+\frac{(\alpha-\alpha_j-1)p_j(1-p_j)}{(1-\sum_{j=1}^{k-1}p_j)^2}\right\} \end{aligned} \tag{3.10}$$

With known $E(p_j)=\frac{\alpha_j}{\alpha}$, $E(\frac{1-p_j}{p_j})=\frac{\alpha-\alpha_j}{\alpha_j-1}$, and $E\left(\frac{p_j}{1-p_j}\right)=\frac{\alpha_j}{\alpha-\alpha_j-1}$, we get

$$\begin{aligned} ESS_{ELIR}^{p_j} &= \alpha-\alpha_j+(\alpha-\alpha_j-1)E\left(\frac{p_j(1-p_j)}{(1-\sum_{j=1}^{k-1}p_j)^2}\right) \\ &\approx \alpha-\alpha_j+(\alpha-\alpha_j-1)E\left(\frac{p_j}{1-p_j}\right) \\ &= \alpha \end{aligned} \tag{3.11}$$

From Equation 3.3, we know the expected posterior ESS is

$$E\{[i(p_1,\ldots,p_k)+Ni_F(p_1,\ldots,p_k)](i_F(p_1,\ldots,p_k)^{-1})\} \tag{3.12}$$

and the diagonal elements of the posterior ESS is

$$\begin{aligned} &E\left\{\left[\left(\frac{\alpha_j-1}{p_j^2}+\frac{\alpha-\alpha_j-1}{(1-\sum_{j=1}^{k-1}p_j)^2}\right)+N\cdot\left(\frac{1}{p_j}+\frac{1}{1-\sum_{j=1}^{k-1}p_j}\right)\right]\left(p_j-p_j^2\right)\right\} \\ &= \alpha+N\cdot E\left(1-p_j+\frac{p_j(1-p_j)}{1-\sum_{j=1}^{k-1}p_j}\right) \\ &\approx \alpha+N \end{aligned} \tag{3.13}$$

After some algebra, we get the posterior ESS is approximate to $\alpha+N$. The ESS values for multinomial model are listed in Table 4 and it is predictively consistent.

3.4.2 Normal model with both mean and variance unknown

Consider a Normal distribution with both mean μ and variance σ^2 unknown. The density function is $f(y|\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}exp\{-\frac{1}{2\sigma^2}(y-\mu)^2\}$, with potential priors $\mu|\sigma^2\sim N(\mu_0,\sigma^2/\kappa_0)$

Table 4. Multivariate expression for ESS for Multinomial model

Parameter	Method	Prior ESS	Expected posterior ESS
p_j	ELIR	α	$\alpha + N$
	VR	α	$\alpha + N$
	PR	α	$\alpha + N$

Note: $\alpha = \sum_{j=1}^{k} \alpha_j$

and $\sigma^2 \sim \Gamma^{-1}(\nu_0/2, \nu_0\sigma_0^2/2)$, where μ_0, σ_0^2, κ_0 and ν_0 can be interpreted as the mean, variance and sample sizes from a set of prior observations, and where $\nu_0\sigma_0^2$ can be thought of as prior sums of squares.[20] The Fisher information matrix of $f(y|\mu, \sigma^2)$ for one information unit is given by

$$i_F(\mu, \sigma^2) = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2(\sigma^2)^2} \end{bmatrix} \tag{3.14}$$

with inverse

$$(i_F(\mu, \sigma^2))^{-1} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2(\sigma^2)^2 \end{bmatrix}. \tag{3.15}$$

The joint prior distribution (μ, σ^2) is

$$\pi(\mu, \sigma^2) = \left(2\pi\frac{\sigma^2}{\kappa_0}\right)^{-1/2} \frac{\left(\frac{\nu_0\sigma_0^2}{2}\right)^{\nu_0/2}}{\Gamma(\frac{\nu_0}{2})}(\sigma^2)^{-\frac{\nu_0}{2}-1} exp\left(-\frac{1}{2\sigma^2}\left[\kappa_0(\mu-\mu_0)^2 + \nu_0\sigma_0^2\right]\right), \tag{3.16}$$

which yields joint information matrix

$$i(p(\mu, \sigma^2)) = \begin{bmatrix} \frac{\kappa_0}{\sigma^2} & \frac{k_0(\mu-\mu_0)}{(\sigma^2)^2} \\ \frac{k_0(\mu-\mu_0)}{(\sigma^2)^2} & -\frac{1}{2(\sigma^2)^2} - (\frac{\nu_0}{2}+1)\frac{1}{(\sigma^2)^2} + \frac{1}{(\sigma^2)^3}\left[\kappa_0(\mu-\mu_0)^2 + \nu_0\sigma_0^2\right] \end{bmatrix}. \tag{3.17}$$

Based on Equation 3.2, the multivariate ESS_{ELIR} of Normal distribution is

$$ESS_{ELIR} = E\big\{r(\mu,\sigma^2)\big\} = E\big\{i(p(\mu,\sigma^2)\left(i_F(\mu,\sigma^2)\right)^{-1}\big\}. \tag{3.18}$$

By plugging in Equations 3.15 and 3.17, we have

$$\begin{aligned} ESS_{ELIR} &= E\left\{ \begin{bmatrix} \frac{\kappa_0}{\sigma^2} & \frac{k_0(\mu-\mu_0)}{(\sigma^2)^2} \\ \frac{k_0(\mu-\mu_0)}{(\sigma^2)^2} & -\frac{1}{2(\sigma^2)^2} - (\frac{\nu_0}{2}+1)\frac{1}{(\sigma^2)^2} + \frac{1}{(\sigma^2)^3}\left[\kappa_0(\mu-\mu_0)^2 + \nu_0\sigma_0^2\right] \end{bmatrix} \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2(\sigma^2)^2 \end{bmatrix} \right\} \\ &= E\left\{ \begin{bmatrix} \boldsymbol{\kappa_0} & 2\kappa_0(\mu-\mu_0) \\ \kappa_0(\mu-\mu_0)/\sigma^2 & \boldsymbol{-\nu_0 - 3 + \frac{2}{\sigma^2}\left[\kappa_0(\mu-\mu_0)^2 + \nu_0\sigma_0^2\right]} \end{bmatrix}, \right\} \end{aligned} \tag{3.19}$$

where $1/\sigma^2 \sim \Gamma(\nu_0/2, \nu_0\sigma_0^2/2)$ and $\mu \sim N(\mu_0, \sigma^2/\kappa_0)$. We identify the ELIR ESS for each parameter by examining the diagonal elements of the expectation matrix. In this case, the prior $ESS^{\mu}_{ELIR} = \kappa_0$ and $ESS^{\sigma^2}_{ELIR} = E\{-\nu_0 - 3 + \frac{2}{\sigma^2}\left[\kappa_0(\mu-\mu_0)^2 + \nu_0\sigma_0^2\right]\}$. From Equation 3.3, we know the expected posterior ESS is

$$\begin{aligned} &E\{[i(p(\mu,\sigma^2)) + Ni_F(\mu,\sigma^2)](i_F(\mu,\sigma^2))^{-1}\} \\ =&E\left\{ \begin{bmatrix} \frac{\kappa_0}{\sigma^2} + \frac{N}{\sigma^2} & \frac{k_0(\mu-\mu_0)}{(\sigma^2)^2} \\ \frac{k_0(\mu-\mu_0)}{(\sigma^2)^2} & -\frac{1}{2(\sigma^2)^2} - (\frac{\nu_0}{2}+1)\frac{1}{(\sigma^2)^2} + \frac{1}{(\sigma^2)^3}\left[\kappa_0(\mu-\mu_0)^2 + \nu_0\sigma_0^2\right] + \frac{N}{2(\sigma^2)^2} \end{bmatrix} \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2(\sigma^2)^2 \end{bmatrix} \right\} \\ =&E\left\{ \begin{bmatrix} \boldsymbol{\kappa_0 + N} & 2\kappa_0(\mu_p-\mu_0) \\ \kappa_0(\mu_p-\mu_0)/\sigma_p^2 & \boldsymbol{N - \nu_0 - 3 + \frac{2}{\sigma_p^2}\left[\kappa_0(\mu_p-\mu_0)^2 + \nu_0\sigma_0^2\right]} \end{bmatrix} \right\} \end{aligned} \tag{3.20}$$

where (μ_p, σ_p^2) follows the joint posterior distribution. For this example we have $ESS^{\mu}_{ELIR} = E\{\kappa_0 + N\}$ and $ESS^{\sigma^2}_{ELIR} = E\{N - \nu_0 - 3 + \frac{2}{\sigma_p^2}\left[\kappa_0(\mu_p-\mu_0)^2 + \nu_0\sigma_0^2\right]\}$. The ESS values for each method are listed in Table 5.

3.5 Simulation and results

With Equations 3.19 and 3.20, it is straightforward to decide the prior and the expected posterior ELIR ESS of the mean μ – the prior $ESS^{\mu}_{ELIR} = \kappa_0$ and the expected posterior

Table 5. Multivariate expression for ESS for Normal model with both mean and variance unknown

Parameter	Method	Prior ESS	Expected posterior ESS
μ	ELIR	κ_0	$\kappa_0 + N$
	VR	$\frac{\kappa_0\nu_0}{\nu_0-2}$	$\frac{\kappa_n\nu_n\sigma_n^2}{(\nu_n-2)\sigma_p^2}$
	PR	κ_0	$\frac{\kappa_n\sigma_n^2}{\sigma_p^2}$
σ^2	ELIR	$E\{-\nu_0 - 3 + \frac{2}{\sigma^2}\left[\kappa_0(\mu-\mu_0)^2 + \nu_0\sigma_0^2\right]\}$	$E\{N - \nu_0 - 3 + \frac{2}{\sigma_p^2}\left[\kappa_0(\mu_p-\mu_0)^2 + \nu_0\sigma_0^2\right]\}$
	VR	$\nu_0 - 2$	$\nu_n - 2$
	PR	$\frac{(\nu_0-2)^2(\nu_0-4)}{\nu_0(\nu_0+2)}$	$\frac{(\nu_n-2)^2(\nu_n-4)}{\nu_n(\nu_n+2)}$

Note: $\nu_n\sigma_n^2 = \nu_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 N}{\kappa_n}(\bar{y}-\nu_0)^2$ (posterior sum of square)

$ESS^{\sigma^2}_{ELIR} = \kappa_0 + N$, respectively. However, for the prior and the expected posterior ELIR ESS of the variance σ^2, the mean of 10,000 simulations are used to determine the expectation. Each simulated data set of size $N = (10, 50, 100, 1000)$ is generated from the sampling model and used to generate the respective posterior distribution (μ_p, σ_p^2), from which the ESS for each parameter are obtained. Table 6 shows the prior ESS, the expected posterior ESS, and the difference between the expected posterior ESS and N, with different settings of sample size of prior observations for mean and variance – κ_0 and ν_0 – respectively.

We see from Table 6 that for both mean μ and variance σ^2, the expected posterior ELIR ESS is the sum of prior ELIR ESS and N, especially as N increases (e.g., $N > 10$). Only the ELIR ESS seems to satisfy the predictive consistency criterion for the two parameters μ and σ^2 simultanuesouly. The VR ESS seems to be predictively consistent on the variance σ^2, but not predictively consistent on the mean μ, while the PR ESS is only consistent (short of a small constant bias) for μ.

Table 6. Prior ESS and expected posterior ESS - N for planned sample size $N = 10, 50, 100,$ and 1000: for normal data with mean and variance unknown.

		Prior ESS	Expected Posterior ESS				Expected Posterior ESS - N			
Parameter	Method	$N = 0$	$N = 10$	$N = 50$	$N = 100$	$N = 1000$	$N = 10$	$N = 50$	$N = 100$	$N = 1000$
$\kappa_0 = 1, \nu_0 = 1$										
μ	ELIR	1	11	51	101	1001	1	1	1	1
	VR	-1	11	51	101	1001	1	1	1	1
	PR	1	9	49	99	999	-1	-1	-1	-1
σ^2	ELIR	0	9	50	100	1000	-1	0	0	0
	VR	-1	9	49	99	999	-1	-1	-1	-1
	PR	-1	4	42	91	991	-6	-8	-9	-9
$\kappa_0 = 2, \nu_0 = 5$										
μ	ELIR	2	12	52	102	1002	2	2	2	2
	VR	3	12	52	102	1002	2	2	2	2
	PR	2	10	50	100	1000	0	0	0	0
σ^2	ELIR	4	12	54	104	1004	2	4	4	4
	VR	3	13	53	103	1003	3	3	3	3
	PR	0	7	46	95	995	-3	-4	-5	-5
$\kappa_0 = 4, \nu_0 = 5$										
μ	ELIR	4	14	54	104	1004	4	4	4	4
	VR	**7**	14	54	104	1004	4	4	4	4
	PR	4	12	52	102	1002	2	2	2	2
σ^2	ELIR	4	12	53	104	1004	2	3	4	4
	VR	3	13	53	103	1003	3	3	3	3
	PR	0	7	46	95	995	-3	-4	-5	-5

ELIR = expected local-information-ratio method, VR = variance ratio method, PR = precision ratio method

3.6 Discussion

Our straightforward multivariate extension of the original univariate ELIR ESS approach requires a holistic approach across parameters, focusing on each as a separate part of a vector-valued whole, rather than attempting to determine a single ESS for the entire prior. This approach maintains predictive consistency, in both the cases of Multinomial distribution and Normal distribution. The results show that the ELIR ESS of each parameter in multivariate cases inherits the predictive consistency property as the ELIR ESS in original univariate cases, whereas neither the VR ESS or the PR ESS have this property for every parameter in the models.

It should be noted that in comparisons, instead of obtaining an exact value of the VR ESS and the PR ESS, we approximate the VR ESS and the PR ESS for each parameter separately; as a limitation, we are not sure if these approximates lead to under- or over-estimates of ESS. We also do not make use of the off-diagonal values from the ESS_{ELIR} matrices, though as our focus is on particular parameters this may not be a limitation. From our simulations we notice that the predictive consistency does not change with prior sample sizes (κ_0 and ν_0), nor is the approach affected after a modicum of observed sample size is attained, as the predictive consistency of the ELIR ESS on the variance parameter (σ^2) is satisfied when the sample size of N is enough large (e.g., $N > 10$).

CHAPTER 4

EARLY TERMINATION IN TWO-PARAMETER MODEL PHASE II CLINICAL TRIAL DESIGNS USING DECREASINGLY INFORMATIVE PRIORS

4.1 Abstract

Decreasingly informative priors (DIP) have been previously used in one-parameter models for early termination Phase II clinical trials.[21] In this paper, we extend the DIP approach to two-parameter models for $Normal$ and $Weibull$ data, by functionalizing the prior effective sample size (ESS) of each parameter to equal the unobserved sample size and center the distribution around skeptical assumptions reflecting the null hypothesis. Simulated studies comparing the performance of the DIP with standard Bayesian priors show that the DIP approach needs fewer sample size with lower variability under admissible designs and helps to control type I error rates.

4.2 Introduction

Phase II clinical trials typically aim to determine if there is sufficient evidence of preliminary efficacy for a treatment to warrant futher investigation, such as in Phase III trials, or whether the investigation should be stopped early due to a lack of efficacy or safety. These trials are often small and data monitoring usually happens as subjects are enrolled, allowing decisions on whether to stop the trials early for efficacy, safety or futility to be made as soon as possible, even before the planned end of the study. Thall and Simon proposed Bayesian methods for early termination Phase II clinical trial [1][2], but the prior or historical information is required and crucial because the resulting posterior distributions are strongly influenced by

the priors. One way to exchange the subjectivity of prior information is through the use of a decreasingly informative prior (DIP), which incorporates a skeptical prior in a manner that decreases its prior effective sample size (ESS) as subjects accrue. [3][12] In this way, as the number of subjects increases, the posterior distributions is increasingly informed by the observed data and less by the prior information. The DIP is also centered around skeptical assumptions about treatment efficacy, which restricts termination early in the trial, gradually allowing for more sensitivity to the observed data and thus terminating the trial becomes more likely. The DIP approach has previously been implemented in one-parameter models for the early termination Phase II trials.[21] In one-parameter models, we investigated the performance of the DIP approach by comparing with different standard Bayesian priors using simulation studies and found that the DIP approach requires fewer subjects with lower variability under admissible designs. Further, the DIP could help control type I error rates, especially for those instances when increases type I error rates arise from erroneous termination early in a trial.[21]

In this paper, we extend research by considering two-parameter models to implement the DIP approach for continuous outcomes and survival outcomes. The Bayesian framework for early termination Phase II trials, as well as the rationale of the DIP approach with examples, are described in Section 4.3. Simulation studies are conducted in Section 4.4 to compare the performance of the DIP approach with the standard Bayesian priors. Section 4.5 concludes the paper with a discussion. Throughout this paper we focus on one-sample clinical trials.

4.3 Statistical Methods

In early termination Phase II trial, for one-parameter models, we assumed that we have a likelihood function $f(y|\theta)$ and prior distribution $\pi(\theta)$ for outcome vector y and scalar parameter θ.[21] For multi-parameter models, suppose we have a multivariate likelihood function

$f(y|\boldsymbol{\theta})$ and prior distribution $\pi(\boldsymbol{\theta})$ for vector-based $\boldsymbol{\theta} = \{\theta_1, \theta_2, ..., \theta_n\}$ and the prior distribution $\pi(\boldsymbol{\theta})$, which can be a joint prior distribution such as $\pi(\boldsymbol{\theta}) = \pi(\theta_1|\theta_2, ..., \theta_n)\pi(\theta_2, ..., \theta_n)$ or independent prior distributions such as $\pi(\boldsymbol{\theta}) = \pi(\theta_1)\pi(\theta_2)...\pi(\theta_n)$. Let function $m(\boldsymbol{\theta})$ be the value representing efficacy in the experimental treatment, while $m(\boldsymbol{\theta_0})$ reflects some null level representing the boundary between an efficacious and non-efficacious treatment. Then the hypotheses we are interested in are:

$$\begin{aligned} H_0 &: m(\boldsymbol{\theta}) \leq m(\boldsymbol{\theta_0}) + \delta_0 \\ H_1 &: m(\boldsymbol{\theta}) > m(\boldsymbol{\theta_0}) + \delta_0 \end{aligned} \tag{4.1}$$

where δ_0 is a fixed improvement that the new treatment is intended to achieve (which could be 0) in efficacy. Note that these hypotheses assume that larger values of $m(\boldsymbol{\theta})$ are reflective of more efficacy; if lower values imply more efficacy, we could simply reverse the directions of the inequalities. We can decide to terminate the trial if the evidence is promising ($P(m(\boldsymbol{\theta}) > m(\boldsymbol{\theta_0})|\pi, N, n) \geq p_s$), or unpromising ($P(m(\boldsymbol{\theta}) > m(\boldsymbol{\theta_0})|\pi, N, n) \leq p_f$), where p_s and p_f are the probabilities thresholds needed to terminate the trial for superiority or futility, π reflects the prior information we assume, N is the predetermined sample size for the new treatment, n is the total number of recruited subjects. These probabilities can be estimated and the resulting decisions can be made after each new subject is enrolled and observed until the experimental treatment is determined as either efficacious or futile, or when all the predetermined total number of subjects for any of the two treatments are recruited.

4.3.1 Decreasingly Informative Prior

For models in early termination Phase II clinical trial designs, in the traditional Bayesian framework, the prior or historical information are necessary for the prior selection. Instead of requiring prior or historical information, we propose implementing a Bayesian method incorporating a natural lead-in skeptical prior - decreasingly informative prior (DIP), which

is parameterized in such a way that the unobserved sample size $N - n$ is made explicitly or approximately equal to the prior effective sample size (ESS) in the prior distribution $\pi(\boldsymbol{\theta})$. In clinical trials, the number of accrued subjects n is small at the beginning of a trial relative to the predetermined sample size N, making the prior ESS predominate and the DIP informative. Additionally, the DIP is parameterized in a way that the prior distribution $\pi(\boldsymbol{\theta})$ is centered at the value or set of values that would reflect conditions of the null hypothesis in Equation 4.1. Therefore, the DIP approach will constrain the resulting posterior distribution early in the trial from providing evidence in the form of posterior probabilities that would favor termination of the trial. As more subjects are enrolled, the observed data dominates the posterior, allowing the posterior distribution becomes more sensitive to the likelihood function instead of the DIP and thus terminating the trial becomes more likely. The following are the steps for Bayesian models incorporating the DIP.

1. Determine the prior ESS for each parameter of the multivariate likelihood function.
2. For each parameter, functionalize the prior in terms of the unobserved sample size $N - n$ so that the prior ESS is N at the beginning of the trial and 0 at the end of the trial.
3. Center each prior distribution at some value reflecting the null hypothesis to constrain termination early in a trial.

In this manuscript, we default to use *expected local-information-ratio (ELIR)* to determine the prior ESS.[11] We extended Neuenschwander's ELIR approach to two-parameter approach and showed that it maintains predictive consistency for two-parameter models in Chapter 3.

4.3.2 Examples

4.3.2.1 Normal data with both mean and variance unknown

For outcomes that could be modeled with a $Normal$ distribution, the vector-valued $\boldsymbol{\theta} = (\mu, \sigma^2)$ and the likelihood function $f(y|\boldsymbol{\theta}) = f(y|\mu, \sigma^2) \sim N(\mu, \sigma^2)$. We here assume the variance σ^2 is unknown, then a natural choice would be the conjugate joint prior distribution

$$\pi(\boldsymbol{\theta}) = \pi(\mu, \sigma^2) = \pi(\mu|\sigma^2)\pi(\sigma^2)$$

Suppose we have priors $\pi(\mu|\sigma^2) \sim N(\mu_0, \sigma^2/\kappa_0)$ and $\pi(\sigma^2) \sim \Gamma^{-1}(\nu_0/2, \nu_0\sigma_0^2/2)$, which can be written as

$$\pi(\mu, \sigma^2) = \left(2\pi\frac{\sigma^2}{\kappa_0}\right)^{-1/2} \frac{\left(\frac{\nu_0\sigma_0^2}{2}\right)^{\nu_0/2}}{\Gamma(\frac{\nu_0}{2})}(\sigma^2)^{-\frac{\nu_0}{2}-1} exp\left(-\frac{1}{2\sigma^2}\left[\kappa_0(\mu-\mu_0)^2 + \nu_0\sigma_0^2\right]\right) \tag{4.2}$$

where the prior mean μ_0 is based on κ_0 prior obervations from $Normal$ distribution, σ_0^2 can be interpreted as the sample variance from ν_0 prior observations from $Inverse-Gamma$ distribution, and $\nu_0\sigma_0^2$ can be thought of as prior sums of squares.[22] Based on Bayesian theory, the posterior distribution of the treatment mean μ can be written as

$$\mu|\sigma^2, y \sim N\left(\frac{\kappa_0}{\kappa_0+n}\mu_0 + \frac{n}{\kappa_0+n}\bar{y}, \frac{\sigma^2}{\kappa_0+n}\right) \tag{4.3}$$

and the posterior distribution of σ^2 can be derived from $\sigma^2|y \sim \Gamma^{-1}(\nu_n/2, \nu_n\sigma_n^2/2)$, where $\nu_n = \nu_0 + n$ and $\nu_n\sigma_n^2 = \nu_0\sigma_0^2 + (n-1)s_y^2 + \frac{\kappa_0 n}{\kappa_0+n}(\bar{y}-\mu_0)^2$, which can be considered as posterior sums of squares, and s_y^2 is the observed sample variance.[22]

The values of prior parameters κ_0 and ν_0 determine the level of information contained in the prior and the contribution of the prior mean. If κ_0 is small and σ^2/κ_0 is large, the prior $Normal$ distribution is dispersed and less informative; as κ_0 increases, the prior $Normal$ distribution will be more tightly centered around mean and become more informative.

For the DIP model in the *Normal* case, the prior ESS is given by

$$ESS_{ELIR} = E\left\{ \begin{bmatrix} \boldsymbol{\kappa_0} & 2\kappa_0(\mu-\mu_0) \\ \kappa_0(\mu-\mu_0)/\sigma^2 & \boldsymbol{-\nu_0 - 3 + \frac{2}{\sigma^2}\left[\kappa_0(\mu-\mu_0)^2 + \nu_0\sigma_0^2\right]} \end{bmatrix} \right\} \tag{4.4}$$

We identify the ESS for μ and σ^2 by examining the diagonal elements of the expectation matrix, which are $ESS^{\mu}_{ELIR} = \kappa_0$ and $ESS^{\sigma^2}_{ELIR} = E\{-\nu_0 - 3 + \frac{2}{\sigma^2}\left[\kappa_0(\mu-\mu_0)^2 + \nu_0\sigma_0^2\right]\}$. Next, we will set a skeptical prior centered at μ_0 for the parameter μ and a skeptical prior centered at σ_0^2 for the parameter σ^2 as initially informative with

$$E\{\kappa_0\} = N - n \tag{4.5}$$

$$E\{-\nu_0 - 3 + \frac{2}{\sigma^2}\left[\kappa_0(\mu-\mu_0)^2 + \nu_0\sigma_0^2\right]\} = N - n \tag{4.6}$$

These formulations allow the information contained within the posterior distribution of μ and σ^2 to shift from the skeptical prior at the beginning of the trial to the likelihood function as subjects accrue. By simplified approximation with $\mu = \mu_0$ and $\sigma^2 = \sigma_0^2$, the formulations 4.5 and 4.6 can be written as

$$\kappa_0 = N - n \tag{4.7}$$

$$\nu_0 = N - n + 3 \tag{4.8}$$

With this construction, we have the DIP posterior distribution of the treatment mean μ, which is

$$\mu|\sigma^2, y \sim N\left(\frac{N-n}{N}\mu_0 + \frac{n}{N}\bar{y}, \frac{\sigma^2}{N}\right) \tag{4.9}$$

and the posterior distribution of the nuisance parameter σ^2 can be derived from $\sigma^2|y \sim \Gamma^{-1}(\nu_n/2, \nu_n\sigma_n^2/2)$, where $\nu_n = \nu_0 + n$ and $\nu_n\sigma_n^2 = \nu_0\sigma_0^2 + (n-1)s_y^2 + \frac{\kappa_0 n}{\kappa_0+n}(\bar{y}-\mu_0)^2$, with $\nu_0 = N - n + 3$ and $\kappa_0 = N - n$ in the DIP approach.

In this way, as the posterior mean in the Bayesian model is a weighted average of the

prior mean μ_0 and the sample mean $\bar{y}$, this DIP formulation will cause the posterior mean to approximate the skeptical prior mean μ_0 early in the trial, and will increasingly approximate $\bar{y}$ as subjects accrue.

4.3.2.2 Weibull data with both shape and rate parameters unknown

In survival data, outcomes are often modeled with distribution $f(y|\boldsymbol{\theta}) = f(y|\lambda, \gamma) \sim Weibull(\lambda, \gamma)$, parameterized by vector-based $\boldsymbol{\theta} = (\lambda, \gamma)$, where λ is the rate parameter and γ is the shape parameter. Assume that both the parameters λ and γ are unknown, and consider the prior distribution

$$\begin{aligned} \pi(\boldsymbol{\theta}) = \pi(\lambda, \gamma) &= \pi(\lambda)\pi(\gamma) \\ &\propto \lambda^{a-1} e^{-b\lambda} \gamma^{c-1} e^{-d\gamma} \end{aligned} \tag{4.10}$$

where $\lambda \sim \Gamma(a, b)$ and $\gamma \sim \Gamma(c, d)$. We are interested in the hypotheses

$$\begin{aligned} H_0 &: m_1 \leq m_0 + \delta_0 \\ H_1 &: m_1 > m_0 + \delta_0 \end{aligned} \tag{4.11}$$

where m_1 is the median survival time for the new treatment and m_0 is the median survival time for the null-level treatment, which can be calculated as $m = \{\lambda^{-1} log(2)\}^{1/\gamma}$, and δ_0 is the fixed targeted improvement for the treatment to achieve. Since there are no conjugate choices, the posterior distribution of λ and γ can be calculated using Markov Chain Monte Carlo (MCMC) approaches, for instance, with R (RJAGS package).

For the DIP model in the *Weibull* case, the prior ESS is determined using the *ELIR* approach and functionalize the prior with $N - n$, which is given by

$$ESS^{\lambda}_{ELIR} = a - 1 = N - n \tag{4.12}$$

$$ESS^{\gamma}_{ELIR} = E\Big\{ \frac{c-1}{1 + \lambda\gamma^2 E\left[y^{\gamma-1} + \gamma(\gamma-1)y^{\gamma-2}\right]} \Big\} = N - n \tag{4.13}$$

Based on the DIP construction steps, we then center the prior mean at the null hypothesis value m_0 for both the parameters λ and γ, that are $a/b = m_0$ and $c/d = m_0$, . Thus, the DIP formulations for the rate parameter $\lambda \sim \Gamma(a, b)$ can be written as

$$\begin{aligned} a &= N - n + 1 \\ b &= (N - n + 1)/m_0, \text{ where } m_0 = \{\lambda_0^{-1} log(2)\}^{1/\gamma_0} \end{aligned} \tag{4.14}$$

And for the shape parameter $\gamma \sim \Gamma(c, d)$, we want the shape parameter c on the Gamma prior to be as large as enough so that $\pi(\gamma)$ could be approximate $Normal$ distribution at the beginning of the trial. With that, we can simplify Equation 4.13 so that the DIP prior for $\gamma \sim \Gamma(c, d)$ can be derived as

$$\begin{aligned} c &= N - n + 1 \\ d &= (N - n + 1)/m_0, \text{ where } m_0 = \{\lambda_0^{-1} log(2)\}^{1/\gamma_0} \end{aligned} \tag{4.15}$$

In the DIP approach, as more subjects n accrued in the trial, the posterior distribution of the median survival time m will depend more on the observed data whose size n is related to the prior.

4.4 Simulation Studies

The goals of our simulation studies are to compare the smallest possible sample size N among admissible designs, the correspondent end-of-trial sample size n and its standard deviation, power, type I error, and the percentage of end-of-trial for futility between the DIP and the standard Bayesian approaches. We define the admissible design as having at least 80% power and no greater than 5% type I error. If there are no admissible designs based on power, we default to selecting the combination of parameters that yield the highest power and best-controlled type I error. If there are no admissible designs based on type I error, we default to selecting the combination of parameters that yield the lowest type I error and at least 80%

power. Power is measured as the proportion of trials where the null hypothesis is rejected under the alternative hypothesis, while type I error is measured as the proportion of trials where the null hypothesis is rejected under the null hypothesis. All simulations are coded using R for at least $10,000$ simulated trials per parameter setting.

4.4.1 Simulation Template

We will explore both $Normal-$ and $Weibull-$ distributed outcomes. The observed data y_i for each subject in each trial is randomly simulated from the probability density function $f(y_i|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is based off the population-level values assumed for that trial. Each subsequent subject is recruited until the trial is stopped for futility or efficacy or the maximum N is reached. To identify the smallest possible sample size among the admissible design, we set the total sample size $N \in (10, 100)$, the threshold probabilities $p_f \in (0.01, 0.10)$ and $p_s \in (0.80, 0.99)$, both in increments of 0.01. For simplicity, we assume the target threshold $\delta_0 = 0$. With the simulated observed data, we estimate the power, type I error, end-of-trial sample size and its standard deviation, and the percentage of stopping the trial for futility for each combination of p_s, p_f and N. Then, we report the smallest total sample size N under those combinations (p_s, p_f, and N) yielding admissible power and type I error rate, with the corresponding estimated characteristics. An indicator of whether the null hypothesis was rejected, the number of patients recruited, and the reason the trial was stopped early are recorded for each simulated trial.

In $Normal$ case, we consider Bayesian models with $\nu_0 = 1$ and $\kappa_0 = 2, 6$, or 10. For both standard Bayesian and the DIP approaches, we study four different cases reflecting our assumptions about prior variability: (1) highly informative prior (σ_0^2) with low observed response variability (σ_1^2), (2) minimally informative prior (σ_0^2) with high observed response variability (σ_1^2), (3) overestimation of the prior variability ($\sigma_0^2 \gg \sigma_1^2$), and (4) underestimation of the prior variability ($\sigma_0^2 \ll \sigma_1^2$). We look at these cases as they represent plausible mistakes

that the statisticians could make when eliciting a prior. For each template, we set the null mean as $\mu_0 = 100$ and expect lower values to imply improved efficacy; thus the new treatment mean is defined as $\mu_1 = \mu_0 - \delta$, where δ with the observed variability (σ_1^2) determine the standardized small, medium and large effect sizes, noted as $d = \delta/\sigma_1$. We consider highly informative prior with $\sigma_0 = \sigma_1 = 12$ and $\delta \in \{2, 5, 10\}$, the minimally informative prior with $\sigma_0 = \sigma_1 = 30$ and $\delta \in \{5, 10, 20\}$, the overestimating prior variability with $\sigma_0 = 30, \sigma_1 = 12$ and $\delta \in \{2, 5, 10\}$, and the underestimating prior variability with $\sigma_0 = 15, \sigma_1 = 30$ and $\delta \in \{5, 10, 20\}$. Each subject is randomly generated from $N(\mu_1, \sigma_1^2)$.

In *Weibull* case, we study three different cases reflecting our assumptions about hazard function: monotonically decreasing hazard function ($\gamma_0 = \gamma_1 = 0.5$), constant hazard function ($\gamma_0 = \gamma_1 = 1$), and monotonically increasing hazard function ($\gamma_0 = \gamma_1 = 2$). Here, the rate parameters are set with $\lambda_0 = 0.69$ and $\lambda_1 \in \{0.29, 0.49, 0.69\}$ for those three cases. Each survival time (t_i) is randomly generated from $Weibull(\lambda_1, \gamma_1)$. To make the survival analysis realistic, we simulate a censoring time $c_i \sim Poisson(\eta)$, and cap the survival time (t) and the censoring time (c) with a maximum observation time (T). The parameters η and T are determined by a 10% censoring rate and are different for each case. We set the traditional Bayesian models with *Gamma* priors for the shape (γ) and rate (λ) parameters centered on the null values.

4.4.2 Simulation Results

The simulation results for *Normal* cases with the first template - highly informative prior (σ_0^2) and low observed response variability (σ_1^2) - are shown in Table 7. Comparing the DIP to several Bayesian priors, we can see that when the effect sizes are at medium ($\delta = 5$) or high ($\delta = 10$) level, the DIP approach requires lower planned sample size as well as lower expected sample size to achieve an admissible design on power and type I error; when the effect size is low ($\delta = 2$), the DIP approach has comparable planned sample size, higher expected sample

size but with lower standard deviation. Though at the low effect size ($\delta = 2$), both the DIP and the Bayesian priors cannot achieve the admissible design on type I error, the DIP approach can control type I error better, compared to other priors.

Table 7. Simulation Results for Normal Distribution (Mean and Variance Unknown) - Low Prior Variance and Low Observed Response Variability

Model	μ_0	μ_1	σ_0	σ_1	N[a]	n (SD)[b]	Futility (%[c])	Efficacy	Power	Type I Error[d]
DIP	100	90	12	12	10	10 (0)	0.02 (0)	0.95	0.817	0.050
Bayesian ($\kappa_0 = 2$)	100	90	12	12	12	11 (0.8)	0.02 (0)	0.96	0.829	0.050
Bayesian ($\kappa_0 = 6$)	100	90	12	12	36	14 (4.5)	0.06 (0)	0.98	0.998	0.050
Bayesian ($\kappa_0 = 10$)	100	90	12	12	12	11 (0.8)	0.10 (0)	0.91	0.837	0.050
DIP	100	95	12	12	40	29 (8.4)	0.03 (0)	0.97	0.816	0.050
Bayesian ($\kappa_0 = 2$)	100	95	12	12	49	29 (14.0)	0.05 (0.35%)	0.99	0.811	0.050
Bayesian ($\kappa_0 = 6$)	100	95	12	12	83	36 (19.6)	0.03 (0)	0.99	0.956	0.050
Bayesian ($\kappa_0 = 10$)	100	95	12	12	51	31 (13.2)	0.10 (0.31%)	0.98	0.839	0.050
DIP	100	98	12	12	95	58 (26.1)	0.08 (0.73%)	0.87	0.803	0.235
Bayesian ($\kappa_0 = 2$)	100	98	12	12	100	45 (34.5)	0.01 (0.70%)	0.92	0.807	0.320
Bayesian ($\kappa_0 = 6$)	100	98	12	12	99	47 (33.8)	0.01 (0.30%)	0.91	0.804	0.305
Bayesian ($\kappa_0 = 10$)	100	98	12	12	96	45 (31.7)	0.06 (2.50%)	0.90	0.802	0.296

[a] The planned sample size

[b] The expected sample size (standard deviation)

[c] The percentage of stopping the trial for futility

[d] Type I error is calculated under the null hypothesis $\mu_1 = \mu_0$

The results with the second template - minimally informative prior (σ_0^2) and high observed response variability (σ_1^2) are shown in Table S.5 in the supplemental materials and have the same conclusions. Furthermore, we notice that in this case, some standard Bayesian priors at the medium effect size ($\delta = 10$) cannot achieve admissible designs on type I error; however, the DIP approach can get the admissible design on power and type I error, with lower planned sample size and comparable expected sample size with a smaller variability.

For the *Normal* case with overestimating prior in Table S.6 (see supplemental materials), the DIP approach performs better at the medium effect size ($\delta = 5$) compared to the other Bayesian priors, with lower planned sample size and comparable expected sample size with a smaller variability under the admissible designs. Similarly, in the *Normal* case of underestimating prior in Table S.7 (see supplemental materials), the DIP approach works better when the effect sizes are at medium ($\delta = 10$) or high ($\delta = 20$) levels.

Table 8 shows the simulation results for *Weibull* cases. We can see that in the case of the decreasing hazard rates ($\gamma_0 = \gamma_1 = 0.5$), the DIP approach has smaller maximum planned sample sizes, and similar recruited sample sizes with lower standard deviation to terminate the trial. The type I error in both the DIP and Bayesian approaches are not controlled well. With the constant hazard rates ($\gamma_0 = \gamma_1 = 1$), that is the special case of *Weibull* distribution - *Exponential* distribution, and the inverse of the rate parameter represents the mean time until the event occurs. The results show that the DIP approach works well and has a larger power compared to the standard Bayesian approach when the average survival time of two groups are large. However, the DIP shows inferior to the standard Bayesian approach when the average survival time of two groups are small. In the case of the increase hazard rates ($\gamma_0 = \gamma_1 = 2$), which is the most common case among oncology studies, the DIP works as well as the standard Bayesian approach; importantly, the DIP performs slightly better in the situation of rate differences are small, with a smaller maximum planned sample size and a higher power.

4.5 Discussion

In conclusion, we applied the DIP approach to multivariate models and took the examples of two-parameter models in early termination Phase II clinical trial designs with simulations to evaluate the performance of the DIP, by comparing it with the standard Bayesian posterior probability approach. For both *Normal* and *Weibull* cases, we considered different scenarios

Table 8. Simulation Results for Weibull Distribution

Model	γ_0	γ_1	λ_0	λ_1	N[a]	n (SD)[b]	Futility (%[c])	Efficacy	Power	Type I Error[d]
DIP	0.5	0.5	0.69	0.29	20	19 (1.2)	0.09 (0.3%)	0.99	0.81	0.23
Bayesian	0.5	0.5	0.69	0.29	29	16 (5.5)	0.10 (0.2%)	0.99	0.84	0.16
DIP	0.5	0.5	0.69	0.49	23	21 (2.6)	0.09 (5%)	0.95	0.81	0.39
Bayesian	0.5	0.5	0.69	0.49	26	20 (5.7)	0.10 (0.2%)	0.99	0.82	0.39
DIP	1	1	0.69	0.29	10	10 (0)	0.02 (0)	0.93	0.86	0.05
Bayesian	1	1	0.69	0.29	10	10 (0)	0.02 (0)	0.93	0.80	0.05
DIP	1	1	0.69	0.49	41	32 (8.0)	0.04 (9.8%)	0.98	0.85	0.08
Bayesian	1	1	0.69	0.49	26	20 (5.8)	0.06 (0.2%)	0.99	0.82	0.17
DIP	2	2	0.69	0.29	10	10 (0)	0.01 (0)	0.89	1	0.05
Bayesian	2	2	0.69	0.29	10	10 (0)	0.01 (0)	0.94	1	0.05
DIP	2	2	0.69	0.49	10	10 (0)	0.10 (0)	0.90	0.82	0.05
Bayesian	2	2	0.69	0.49	12	10 (0.9)	0.07 (0)	0.95	0.80	0.05

[a] The planned sample size

[b] The expected sample size (standard deviation)

[c] The percentage of stopping the trial for futility

[d] Type I error is calculated under the null hypothesis $\mu_1 = \mu_0$

that reflect a broad scope of the real-world trials. The results show that, in some but not all cases, such as *Normal* case with high effect sizes, the DIP performs better, including fewer maximum planned patients, fewer expected patients with lower variability, better-controlled type I error with admissible power. These results are consistent with the previous research on the DIP approach in one-parameter models. Noteworthy is the fact that the results of *Weibull* case do not indicate that the DIP approach is superior to the standard Bayesian approach. While the performance of the DIP in *Weibull* case with increasing hazard functions appears to be acceptable, more research is needed to confirm that the DIP

is more efficient than the standard Bayesian approach and leads to ethical and statistical benefits. For example, changing the shape and rate parameters in $Weibull$ distributions might change our findings and conclusions.

While we introduced the rationale of the DIP approach in general multivariate models in Section 4.3.1, there are limitations that we only considered two-parameter models as examples in this study. Future work could apply the DIP approach to $Multinomial - Dirichlet$ model for early termination designs of multiple outcomes.[23] In particular, as for the simulation-based designs, the computation time cannot be ignored. As we mentioned before, MCMC approaches with RJAGS package are used to calculate the single or multiple unknown parameters for the non-conjugate models, including, for example, $Weibull$ case. The compiling time of RJAGS package is dependent on the number of interested parameters, iterations, chains, and the "burn-in" period, which is much longer than we expected. It is suggested to set a small number of simulations, iterations, and chains to test the parameter settings, especially for the multiple unknown parameter cases; once the appropriate range of parameter values are determined, the number of simulations and iterations can be increased to estimate more precise results.

CHAPTER 5

BAYESDIP: AN R PACKAGE FOR BAYESIAN EARLY TERMINATION IN PHASE II CLINICAL TRIALS USING DECREASINGLY INFORMATIVE PRIORS

5.1 Abstract

In Phase II clinical trials, Thall and Simon's Bayesian posterior probability design is commonly used to monitor the data and decide the trial stopped for sufficient efficacy or a lack of futility associated with a new treatment. However, this Bayesian approach depends on a pre-selection prior and may inflate type I error rate and raise the risk of erroneously terminating the trials too early. In this paper, we primarily implement Sabo's decreasingly informative prior to provide an informative yet skeptical prior on Bayesian early termination Phase II trials, that can overcome the challenges of the standard Bayesian priors. We present an R package, **BayesDIP**, that allows flexibility with standard Bayesian priors or the decreasingly informative prior for Bayesian early termination Phase II trials to accommodate power analysis or expected sample size calculation given target significant level, power and stopping decision cutoffs. The relevant statistical theory, models, and examples for each distribution using **BayesDIP** are discussed in this paper.

5.2 Introduction

For ethical and economical considerations, a Phase II trial is typically conducted by examining the interim data to decide whether a new treatment warrants further investigation in a Phase III trial, or whether the investigation should be stopped early for futility or lack of efficacy. Thall and Simon proposed a Bayesian posterior probability design with binary out-

comes for defining termination rules in terms of efficacy and futility in Phase II trials[1][2], but this standard Bayesian approach relies on a pre-selected prior and may raise an issue of inflating type I error rate and a risk of erroneously adapting the trails too early.

Extending from the standard Bayesian approach, Sabo developed a skeptical prior distribution, which incorporates the nonaccrued data and functionalizes the prior to be centered around at some value reflecting the null hypothesis.[3] The proposed DIP distribution is just functional and does not rely on any prior or historical information. In early termination of Phase II trials, this informative yet skeptical prior distribution restricts the posterior distribution from providing evidence that would favor termination of the trial during early period of a trial, and with more data observed, the prior distribution becomes decreasingly informative and the posterior distribution becomes increasingly more sensitive to the likelihood so that terminating the trial becomes more likely. Wang et.al applied the DIP to admissible designs of early termination Phase II trials for single-parameter models, include *Bernoulli*, *Normal* and *Poisson* distribution, where the simulation results showed that the DIP requires fewer patients when admissible designs are achieved and when there are no designs attaining the target power, the DIP yields similar power and better-controlled type I error with comparable or fewer patients, compared with the standard Bayesian priors.[21]

Currently, two packages (**ph2bayes**[24] and **ph2bye**[25]) are available in R to perform the Bayesian early termination rules in Phase II clinical trials. The **ph2bayes**[24] package implements Thall and Simon's Bayesian posterior probability design[1] and Lee and Liu's Bayesian predictive probability design[8] to determine the stopping decision cutoffs for single-arm Phase II trials for binary outcomes. The package **ph2bye**[25] determines sample size and stopping decision cutoffs for single-arm Phase II trials based on Bayesian posterior/predictive probabilities. Until now, there exists no R package providing a comprehensive toolkit capable of the power analysis or calculating the expected sample size. Such a comprehensive program would allow for use of Thall and Simon's standard Bayesian priors or Sabo's DIP given the

significant level, power and stopping decision cutoffs, when the outcomes have a *Bernoulli*, *Normal* or *Poisson* distribution. The **BayesDIP** package described below fills this gap.

This paper describes the theoretical formulas of both the DIP and standard Bayesian approaches for early termination rules during Phase II trials in Section 5.3. Different functions with a variety of data distributions as well as input parameters, output and results are detailed in Section 5.3. Example applications of clinical trials in how to implement and customize the functions are provided in Section 5.4. Section 5.5 summaries the paper with a discussion.

5.3 BayesDIP: R Package Description

The functions in this package can be used to determine the minimum planned sample size N achieving the admissible design or to determine the expected sample size n with its standard deviation, the exact power and the exact type-I-error (e.g. given the planned sample size N and the stopping boundaries). The specific function of interest will be determined by the type of outcome variable and the data distribution. All functions return an object of list where the users can extract detailed information about the designs they are interested in.

Note that all functions conduct simulations to estimate the power and type-I-error, where power is measured as the proportion of trials when the null hypothesis is rejected under the alternative hypothesis, while type-I-error is measured as the proportion of trials where the null hypothesis is rejected under the null hypothesis.

5.3.1 Binomial

Testing one-sample or two-sample response rates are common in clinical trials when the outcome follows a binomial distribution. Let y_i is a binary response from the i-the subject with response rate p, $i = 1, \ldots, n$. We have the likelihood function $f(y|p) = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}$.

The hypothesis we are testing is:

$$H_0 : p_1 \leq p_0 + \delta_0,$$

$$H_1 : p_1 > p_0 + \delta_0$$

where p_1 is the response rate of the new treatment, p_0 is the null hypothesized value, and δ_0 is a fixed targeted improvement of the new treatment to achieve (which could be 0). Note that this hypothesis assumes that larger response rate implies greater effcacy; we could simply switch the directions of the inequalities if lower values imply greater efficacy. Based on the conjugate natural of $Beta - Binomial$ model, the standard posterior distribution of p for an informative prior $Beta(a, b)$ is:

$$p|y \sim Beta(a + y, b + n - y)$$

where $y = \sum_{i=1}^{n} y_i$ is the total number of successes out of the n observed subjects in the trial. Assume that an informative and skeptical DIP can be specified as

$$Beta(1 + p_0(N - n), 1 + (1 - p_0)(N - n))$$

the posterior distribution of p for the DIP is

$$p|y \sim beta(1 + p_0(N - n) + y, 1 + (1 - p_0)(N - n) + (n - y)).$$

In the DIP approach, the posterior distribution of p is more centered at the prior mode p_0 early in the trial. As more data observed and unobserved sample size $N - n$ becomes small, this skeptical prior becomes less informative and the posterior distribution of p becomes more sensitive to the observed data instead of the prior; thus, terminating the trial becomes more likely.

We can easily extend the one-sample case to two-sample case. Here, null distributions – centered at the null hypothesized value – could be incorporated through a hyper-prior.

The DIP for parameters in each group would still be functionalized with $N - n$ equalling the effective sample size, but the parameter in question would be given its own probability distribution reflecting the possible values under the null. Supposed that in two-sample case, we assume y_{ij} be a binary response from the i-th subject in the j-th group, $i = 1, \ldots, n_j, j = 1, 2$. For the standard Bayesian model, the resulting posterior with an informative prior $p_j \sim beta(a, b)$ is given by:

$$p_j | y_{ij} \sim Beta(a + y_j, b + n_j - y_j)$$

where y_j is the total number of successes out of the n_j observed subjects for the j-th group in the trial. For the DIP approach, the resulting posterior is given by:

$$\begin{aligned} y_{ij} | p_j &\sim Binomial(n_j, p_j), i = 1, 2 \\ p_j &\sim Beta(1 + p_0(N_j - n_j), 1 + (1 - p_0)(N_j - n_j)) \\ p_0 &\sim Beta(1, 1) \\ p_j | y_{ij} &\sim Beta(1 + p_0(N_j - n_j) + y_j, 1 + (1 - p_0)(N_j - n_j) + (n_j - y_j)) \end{aligned}$$

where N_j is the total planned sample size for the j-th group.

We have developed four functions in this package for the cases where the binary outcomes are assumed follow a *Bernoulli* distribution.

The **OneSampleBernoulli** function is used for one-sample *Bernoulli* distribution and returns an object including the expected sample size n with its standard deviation, the exact power and the exact type-I-error. This function uses algorithm described above. The arguments for **OneSampleBernoulli** are as follows:

```
OneSampleBernoulli(prior, N = 100, p0, p1, d = 0, ps = 0.95, pf =
  0.05, alternative = c("less","greater"), seed = 202209, sim =
  5000)
```

A list of length 3 containing the distributional information for the **prior**. The first element is a number specifying the type of the prior. Options are

1. DIP;

2. Beta(a,b), where a = shape, b = scale

If the prior is specified as option 1 (DIP), the second and the third elements of the list are set as 0. Otherwise, the second and the third elements of the list are the parameters a and b, respectively. The total planned sample size **N** has a default value of 100. The null response rate **p0**, which could be taken as the standard or historical rate, and the response rate of the new treatment **p1** must be specified. The target improvement, which can be considered as the minimal clinically meaningful difference, is set with **d** and default to 0. The efficacy (upper) and futility (lower) decision boundaries are indicated as **ps** and **pf**, respectively, with 0.95 and 0.05 as the default value. The argument **alternative** specifies the alternative hypothesis as either "less" (lower values imply greater efficacy) or "greater" (larger values imply greater efficacy). The observed data are simulated with a default **seed** and a default value of 5000 iterations indicated as **sim** to estimate the posterior probability (e.g., $P(p_1 > p_0 + \delta_0|y)$) for early termination decision.

The **OneSampleBernoulli.Design** function is used for one-sample *Bernoulli* distribution in determining the minimum planned sample size to achieve an admissible design and returns the minimum planned sample size N, the expected sample size n with its standard deviation, the exact power and the exact type-I-error. This function uses algorithm described above and iterates over the specified searching sample size to identify the minimum sample size under the admissible design. The arguments for **OneSampleBernoulli.Design** are as follows:

```
OneSampleBernoulli.Design(prior, nmin = 10, nmax=100, p0, p1, d = 0,
    ps, pf, power=0.8, t1error=0.05, alternative = c("less","greater
   "), seed = 202209, sim = 1000)
```

A list of length 3 containing the distributional information for the **prior**. The first element is a number specifying the type of the prior. Options are

1. DIP;

2. Beta(a,b), where a = shape, b = scale

If the prior is specified as option 1 (DIP), the second and the third elements of the list are set as 0. Otherwise, the second and the third elements of the list are the parameters a and b, respectively. The minimum and maximum searching sample size are set with **nmin** and **nmax** and default to 10 and 100, respectively. The null response rate **p0**, which could be taken as the standard or historical rate, and the response rate of the new treatment **p1** must be specified. The target improvement, which can be considered as the minimal clinically meaningful difference, is set with **d** and default to 0. The efficacy (upper) and futility (lower) decision boundaries are indicated as **ps** and **pf**, respectively. The admissible design is specified with **power** and **t1error** and we default to achieve an admissible design which has 80% power and 5% type-I-error. If there are no admissible design based on this type-I-error, then we default to output the designs with the lowest type-I-error and at least the user-defined (e.g., 80%) power. The argument **alternative** specifies the alternative hypothesis as either "less" (lower values imply greater efficacy) or "greater" (larger values imply greater efficacy). The observed data are simulated with a default **seed** and a default value of 1000 iterations indicated as **sim** to estimate the posterior probability (e.g., $P(p_1 > p_2 + \delta_0|y)$) for early termination decision.

The function **TwoSampleBernoulli** is used for two-sample *Bernoulli* distribution in calculating the expected sample size with its standard deviation, the exact power and the exact type-I-error. This function uses algorithm for two-sample case described above and applies equal allocation between two treatment groups, with the same arguments as the function **OneSampleBernoulli** shown as follows, where the parameter **N** defaults to 200 and the response rate of the new treatment **p1** and the response rate of the compared-group treatment **p2** must be specified. See **OneSampleBernoulli** for more details.

```
TwoSampleBernoulli(prior, N = 200, p1, p2, d = 0, ps = 0.95, pf =
   0.05, alternative = c("less","greater"), seed = 202209, sim =
   5000)
```

The **TwoSampleBernoulli.Design** function is designed to determine the minimum total planned sample size under an admissible design. The arguments are as follows, same with the function **OneSampleBernoulli.Design**, where the maximum searching sample size defaults to 200 and the response rate of the new treatment **p1** and the compared-group treatment **p2** must be specified. See **OneSampleBernoulli.Design** for more details.

```
TwoSampleBernoulli.Design(prior, nmin = 10, nmax=200, p1, p2, d = 0,
    ps, pf, power=0.8, t1error=0.05, alternative = c("less","greater
   "), seed = 202209, sim = 1000)
```

5.3.2 Normal

When the data has a normal distribution, we are interested in comparing two sample means with the hypothesis:

$$H_0 : \mu_1 \geq \mu_0 + \delta_0$$

$$H_1 : \mu_1 < \mu_0 + \delta_0$$

where μ_1 is the mean value of the new treatment, μ_0 is the null mean value and δ_0 is the fixed target improvement of the new treatment to achieve (which could be 0). Here, we expect lower mean values imply improved efficacy and we could simply switch the directions of the inequalities if higher mean values imply greater efficacy.

With the case of variance known, let y_i be a continuous response from the i-th subject

with mean μ and variance s^2, $i = 1, \ldots, n$, the standard posterior distribtuion is given by:

$$\begin{aligned} y_i|\mu &\sim Normal\left(\mu, s^2\right) \\ \mu &\sim Normal\left(\mu_0, \frac{s^2}{n_0}\right) \\ \mu|s^2, y_i &\sim Normal\left(\frac{n_0}{n_0+n}\mu_0 + \frac{n}{n_0+n}\bar{y}, \frac{s^2}{n_0+n}\right) \end{aligned} \tag{5.1}$$

where s^2 is the known variance, μ_0 is the null mean value, which could be considered as the prior mean from the standard or historical data, and n_0 is a prior parameter indicating the level of information contained in the prior and the contribution of the null mean. Small value of n_0 provides a dispersed and less informative prior distribution; otherwise, when n_0 is large, a more informative prior distribution will be more tightly centered around the null mean.

With a skeptical and informative DIP, the resulting posterior distribution in the case of known variance is as follows:

$$\begin{aligned} y_i|\mu &\sim Normal\left(\mu, s^2\right) \\ \mu &\sim Normal\left(\mu_0, \frac{s^2}{N-n}\right) \\ \mu|s^2, y_i &\sim Normal\left(\frac{N-n}{N}\mu_0 + \frac{n}{N}\bar{y}, \frac{s^2}{N}\right) \end{aligned} \tag{5.2}$$

where N is the planned sample size and $N - n$ is the unobserved sample size in the trial.

With the case of variance unknown, let y_i be a continuous response from the i-th subject with mean μ and variance σ^2, $i = 1, \ldots, n$, the standard posterior distribution of μ is given

by[22]:

$$
\begin{aligned}
y_i|\mu,\sigma^2 &\sim Normal\left(\mu,\sigma^2\right)\\
\mu &\sim Normal\left(\mu_0,\frac{\sigma^2}{\kappa_0}\right)\\
\sigma^2 &\sim \Gamma^{-1}\left(\frac{\nu_0}{2},\frac{\nu_0\sigma_0^2}{2}\right)\\
\mu|\sigma^2,y_i &\sim Normal\left(\frac{\kappa_0}{\kappa_0+n}\mu_0+\frac{n}{\kappa_0+n}\bar{y},\frac{\sigma^2}{\kappa_0+n}\right)
\end{aligned}
\tag{5.3}
$$

where σ^2 is the unknown variance which is taken as a nuisance parameter, μ_0 is the null mean value, which can be taken as the prior mean from the standard or historical data, κ_0 is a prior parameter indicating the level of the information contained in the prior mean, σ_0^2 indicates the sample variance from ν_0 prior observations, and $\nu_0\sigma_0^2$ can be thought of as prior sums of squares. Small κ_0 value results in a dispersed and less informative prior distribution of μ; otherwise, as κ_0 increases, the prior distribution of μ will be more informative. The posterior distribution of the nuisance parameter σ^2 in Equation 5.3 can be derived from $\sigma^2|y \sim \Gamma^{-1}(\nu_n/2,\nu_n\sigma_n^2/2)$, where $\nu_n=\nu_0+n$ and $\nu_n\sigma_n^2=\nu_0\sigma_0^2+(n-1)s_y^2+\frac{\kappa_0 n}{\kappa_0+n}(\bar{y}-\mu_0)^2$, which can be considered as posterior sums of squares, and s_y^2 is the observed sample variance.[22]

In the case of unknown variance, the DIP posterior distribution of μ can be written as:

$$
\begin{aligned}
y_i|\mu,\sigma^2 &\sim Normal\left(\mu,\sigma^2\right)\\
\mu &\sim Normal\left(\mu_0,\frac{\sigma^2}{N-n}\right)\\
\sigma^2 &\sim \Gamma^{-1}\left(\frac{N-n+3}{2},\frac{(N-n+3)\sigma_0^2}{2}\right)\\
\mu|\sigma^2,y_i &\sim Normal\left(\frac{N-n}{N}\mu_0+\frac{n}{N}\bar{y},\frac{\sigma^2}{N}\right)
\end{aligned}
\tag{5.4}
$$

where N is the planned sample size, $N-n$ is the unobserved sample size, and σ_0^2 indicates the sample variance from prior observations. The posterior distribution of the nuisance parameter σ^2 in Equation 5.4 can be derived from $\sigma^2|y \sim \Gamma^{-1}(\nu_n/2,\nu_n\sigma_n^2/2)$, where $\nu_n =$

$N+3$ and $\nu_n\sigma_n^2 = (N-n+3)\sigma_0^2 + (n-1)s_y^2 + \frac{(N-n)n}{N}(\bar{y}-\mu_0)^2$, which can be considered as posterior sums of squares, and s_y^2 is the observed sample variance.[22]

For the DIP approach in Equation 5.2 and 5.4, the posterior distribution of μ is a weighted average of the prior mean μ_0 and the sample mean $\bar{y}$; thus, the posterior mean μ will approximate the prior mean μ_0 early in the trial and will increasingly approximate $\bar{y}$ as more data observed.

In this package, we offer four functions for the cases where the outcomes follow a *Normal* distribution.

The **OneSampleNormal1** function is used for one-sample *Normal* distribution with known variance, and returns the expected sample size with its standard deviation, the exact power and the exact type-I-error. This function uses the algorithm described above in Equation 5.1 and 5.2. The arguments for **OneSampleNormal1** are as follows:

```
OneSampleNormal1(prior, N = 100, mu0, mu1, var, d = 0, ps = 0.95, pf
    = 0.05, alternative = c("less","greater"), seed = 202209, sim =
   5000)
```

A list of length 2 containing the distributional information for the **prior**. The first element is a number specifying the type of the prior. Options are

1. DIP;

2. Normal(μ_0, s^2/ν_0), where μ_0 is the prior mean, and s^2 is the known variance.

If the prior is specified as option 2 (standard prior), the second element of the list is the parameter ν_0; otherwise, we can set any value to the second element of the list. The planned sample size **N** defaults to 100. The null mean value **mu0**, which could be taken as the standard or historical mean, and the mean value of the new treatment **mu1** must be specified. The known variance is indicated as **var**. The target improvement, which can be considered as the minimal clinically meaningful difference, is set with **d** and default to 0. The efficacy (upper) and futility (lower) decision boundaries are indicated as **ps** and **pf**, and default to

0.95 and 0.05, respectively. The argument **alternative** specifies the alternative hypothesis as either "less" (lower values imply greater efficacy) or "greater" (larger values imply greater efficacy). The observed data are simulated with a default **seed** and a default value of 5000 iterations indicated as **sim** to estimate the posterior probability (e.g., $P(\mu_1 < \mu_0 + \delta_0|y)$) for early termination decision.

The function **OneSampleNormal1.Design** is useful for a study design to identify the minimum planned sample size under an admissible design, when the outcomes follow a *Normal* distribution with variance known. This function uses algorithm described above in Equation 5.1 and 5.2 and iterates over a range of planned sample size. The arguments for **OneSampleNormal1.Design** are as follows:

```
OneSampleNormal1.Design(prior, nmin = 10, nmax=100, mu0, mu1, var, d
    = 0, ps, pf, power=0.8, t1error=0.05, alternative = c("less","
   greater"), seed = 202209, sim = 1000)
```

A list of length 2 containing the distributional information for the **prior**. The first element is a number specifying the type of the prior. Options are

1. DIP;

2. Normal(μ_0, s^2/ν_0), where μ_0 is the prior mean, and s^2 is the known variance.

If the prior is specified as option 2 (standard prior), the second element of the list is the parameter ν_0; otherwise, we can set any value to the second element of the list. The minimum and maximum searching sample size are set with **nmin** and **nmax** and default to 10 and 100, respectively. The null mean value **mu0**, which could be taken as the standard or historical mean, and the mean value of the new treatment **mu1** must be specified. The known variance is indicated as **var**. The target improvement, which can be considered as the minimal clinically meaningful difference, is set with **d** and default to 0. The efficacy (upper) and futility (lower) decision boundaries are indicated as **ps** and **pf**, respectively. The admissible design is specified with power and t1error and we default to achieve an

admissible design which has 80% power and 5% type-I-error. If there are no admissible design based on this type-I-error, then we default to output the designs with the lowest type-I-error and at least the user-defined (e.g., 80%) power. The argument **alternative** specifies the alternative hypothesis as either "less" (lower values imply greater efficacy) or "greater" (larger values imply greater efficacy). The observed data are simulated with a default **seed** and a default value of 1000 iterations indicated as **sim** to estimate the posterior probability (e.g., $P(\mu_1 < \mu_0 + \delta_0|y)$) for early termination decision.

The function **OneSampleNormal2** is designed for one-sample *Normal* distribution with variance unknown, and returns the expected sample size with its standard deviation, the exact power and the exact type-I-error. This function uses the algorithm described above in Equation 5.3 and 5.4 and applies the same arguments with the function **OneSampleNormal1** but adds an argument of **var0** indicates the prior sample variance σ_0^2 in Equation 5.3 and 5.4. See **OneSampleNormal1** for more details.

```
OneSampleNormal2(prior, N = 100, mu0, mu1, var0, var, d = 0, ps =
   0.95, pf = 0.05, alternative = c("less","greater"), seed =
   202209, sim = 5000)
```

The **OneSampleNormal2.Design** function is used for a study design to determine the minimum planned sample size under an admissbile design, assuming the outcomes follow a *Normal* distribution with variance unknown. This function uses algorithm described above in Equation 5.3 and 5.4, with the same arguments as the function **OneSampleNormal1.Design** but adds an argument of **var0** indicates the prior sample variance σ_0^2 in Equation 5.3 and 5.4. See **OneSampleNormal1.Design** for more details.

```
OneSampleNormal2.Design(prior, nmin = 10, nmax=100, mu0, mu1, var0,
   var, d = 0, ps, pf, power=0.8, t1error=0.05, alternative = c("
   less","greater"), seed = 202209, sim = 1000)
```

5.3.3 Poisson

If the outcome in a clinical trial is the number of events occurring in a fixed interval time, length of stay in hospital, or other count measurements, then a *Poisson* distribution with *Gamma* prior is a plausible choice for early termination models. Supposed that we are interested in test the *Possion* rates and the hypothesis is as follows:

$$H_0 : \lambda_1 \geq \lambda_0 + \delta_0$$

$$H_1 : \lambda_1 < \lambda_0 + \delta_0$$

where λ is the rate as well as the mean value of the new treatment, λ_0 is the null event rate (null mean value) and δ_0 is the fixed target improvement of the new treatment to achieve (which could be 0). Here, we expect lower values of the event rates imply improved efficacy and we could simply switch the directions of the inequalities if higher values of the rates imply greater efficacy.

Let y_i be the number of events from i-th subject with rate λ, $i = 1, \ldots, n$, the standard posterior distribution of λ is:

$$y_i|\lambda \sim Poisson(\lambda)$$

$$\lambda \sim \Gamma(a, b)$$

$$\lambda|y_i \sim \Gamma(a + y, b + n)$$

where $y = \sum_{i=1}^{n} y_i$ is the total number of events from the n observed subjects in the trial. The constructed DIP posterior distribution of λ is defined as follows:

$$y_i|\lambda \sim Poisson(\lambda)$$

$$\lambda \sim Gamma(0.5 + \lambda_0(N - n), 0.001 + (N - n))$$

$$\lambda|y_i \sim Gamma(0.5 + \lambda_0(N - n) + y, 0.001 + N)$$

where λ_0 is the null mean value or the null event rate , which could be taken as the standard

or historical event rate, N is the planned sample size, and $N - n$ is the unobserved sample size in the trial. In the DIP approach, with more subjects observed in the trial, the skewness of the posterior distribution will be determined more by the observed data rather than the prior distribution.

The function **OneSamplePoisson** is designed to estimate the expected sample size with its standard deviation, the exact power and the exact type-I-error for one-sample case where the outcomes follow a *Poisson* distribution. The algorithm of this function is described above. The arguments for **OneSamplePoisson** are as follows:

```
OneSamplePoisson(prior, N = 100, m0, m1, d = 0, ps = 0.95, pf =
   0.05, alternative = c("less","greater"), seed = 202209, sim =
   5000)
```

A list of length 3 containing the distributional information for the **prior**. The first element is a number specifying the type of the prior. Options are

1. DIP;

2. Gamma(a,b), where a = shape, b = scale

If the prior is specified as option 2 (standard *Gamma* prior), the second and the third elements of the list are the parameters a and b, respectively; otherwise, the second and the third elements of the lists can be set as NULL or any values. The total planned sample size **N** has a default value of 100. The null event rate **m0**, which could be taken as the standard or historical rate, and the event rate of the new treatment **m1** must be specified. The target improvement, which can be considered as the minimal clinically meaningful difference, is set with **d** and default to 0. The efficacy (upper) and futility (lower) decision boundaries are indicated as **ps** and **pf**, respectively, with 0.95 and 0.05 as the default value. The argument **alternative** specifies the alternative hypothesis as either "less" (lower values imply greater efficacy) or "greater" (larger values imply greater efficacy). The observed data are simulated

with a default **seed** and a default value of 5000 iterations indicated as **sim** to estimate the posterior probability (e.g., $P(\lambda_1 < \lambda_0 + \delta_0|y)$) for early termination decision.

The function **OneSamplePoisson.Design** is designed to estimate the minimum planned sample size necessary to obtain an admissible design for one-sample case where the outcomes follow a *Poisson* distribution. This function uses algorithm described above and iterates over a range of sample size to identify the minimum sample size achieving the admissible design. The arguments for **OneSamplePoisson.Design** are as follows:

```
OneSamplePoisson.Design(prior, nmin=10, nmax=100, m0, m1, d = 0, ps,
    pf, power=0.8, t1error=0.05, alternative=c("less","greater"),
   seed = 202209, sim = 1000)
```

A list of length 3 containing the distributional information for the **prior**. The first element is a number specifying the type of the prior. Options are

1. DIP;

2. Gamma(a,b), where a = shape, b = scale

If the prior is specified as option 2 (standard *Gamma* prior), the second and the third elements of the list are the parameters a and b, respectively; otherwise, the second and the third elements of the lists can be set as NULL or any values. The minimum and maximum searching sample size are set with **nmin** and **nmax** and default to 10 and 100, respectively. The null event rate **m0** and the event rate of the new treatment **m1** must be specified. The target improvement, which can be considered as the minimal clinically meaningful difference, is set with **d** and default to 0. The efficacy (upper) and futility (lower) decision boundaries are indicated as **ps** and **pf**, respectively. The admissible design is specified with power and t1error and we default to achieve an admissible design which has 80% power and 5% type-I-error. If there are no admissible design based on this type-I-error, then we default to output the designs with the lowest type-I-error and at least the user-defined (e.g., 80%) power. The argument **alternative** specifies the alternative hypothesis as either "less" (lower values

imply greater efficacy) or "greater" (larger values imply greater efficacy). The observed data are simulated with a default **seed** and a default value of 1000 iterations indicated as **sim** to estimate the posterior probability (e.g., $P(\lambda_1 < \lambda_0 + \delta_0|y)$) for early termination decision.

5.4 Example Application of BayesDIP

In this section, we provide two examples implemented with **BayesDIP**. One example demonstrates the use of examining the performance of the posterior probability; The other example shows how to perform the minimum planned sample size calculations with a target power and significant level.

Simmons et al.[26] describe a Phase II study assessing whether vitamin C helps improve recovery from Bone Marrow Transformation (BMT). In this study, A total of 40 patients received intravenous (IV) vitamin C. Based on the binary outcomes of cytomegalovirus (CMV) infection, which is an important risk factor associated with BMT, the study aims at determining whether receiving IV vitamin C can reduce the CMV infection rate to less than 0.5. The hypotheses for the study are as follows:

$$\begin{aligned} H_0 &: \text{the CMV infection rate } \geq 0.5 \\ H_1 &: \text{the CMV infection rate } < 0.5 \end{aligned} \tag{5.5}$$

In this example, the true CMV infection rate (p_1) in this study is considered as 0.275, and the null CMV infection rate (p_0) is 0.5. We set the lower (futility) and upper (efficacy) decision boundaries as default to 0.05 and 0.95, respectively. To determine the expected sample size for early terminating the trial and examine the performance of the posterior probability, we use **OneSampleBernoulli** to estimate the expected sample size with its standard deviation, the power, the type-I-error and the probabilities of reaching the efficacy and futility boundaries. We apply both the DIP approach and the standard Bayesian approach to the study. The default values of the simulation seed and the number of simulations are used.

The output is given below:

```
library(BayesDIP)

# DIP approach
DIP.result <- OneSampleBernoulli(list(1,0,0), N=40, p0=0.5, p1
   =0.275, alternative = "less")
DIP.result$power
[1] 0.9352
DIP.result$type_I_error
[1] 0.086
DIP.result$expected_sample_size
[1] 23.6
DIP.result$expected_sample_size_std
[1] 8.12
DIP.result$the_prob_futility
[1] 2e-04

# Standard non-informative prior Beta(1,1)
Bayes.result <- OneSampleBernoulli(list(2,1,1), N=40, p0=0.5, p1
   =0.275, alternative = "less")
Bayes.result$power
[1] 0.956
Bayes.result$type_I_error
[1] 0.176
Bayes.result$expected_sample_size
[1] 16.2
Bayes.result$expected_sample_size_std
[1] 8.6
Bayes.result$the_prob_futility
[1] 0.001
```

The obtained result indicates that the DIP approach stops the trial at the sample size of 24 ($sd = 8.1$) to reach a 95% efficacious boundary and demonstrate receiving IV vitamin C can reduce the CMV infection rate to less than 0.5, with significant level of 0.086 and power of 0.9352; while the non-informative Bayesian prior requires 17 ($sd = 8.6$) subjects, with significant level of 0.176 and power of 0.956.

Again using data from the Simmons et al.[26] study, we use the number of CD56+ natural killer (NK) at day 30 as a continuous outcome, with the goal to assess whether receiving IV vitamin C can improve the amount of the CD56+NK cells from the null mean

200 (with known variance $s^2 = 100$). The hypothesis for the study is:

$$\begin{aligned} H_0 &: \text{the amount of the CD56+NK cells } \leq 250 \\ H_1 &: \text{the amount of the CD56+NK cells } > 250 \end{aligned} \tag{5.6}$$

In this example, we use a range of sample size 10 to 50 in **OneSampleNormal1.Design** and fine-tune through iterations to determine the minimum sample size with target power equal to 0.8 and significant level equal to 0.05. Assumed the true mean value of the amount of the CD56+NK cells from the Simmons et.al[26] study is 294, with known variance 100, and the null mean value is 250. The posterior probability boundaries are set to 0.95 and 0.05 for efficacy and futility, respectively. The target power, the significant level, the simulation seed and the number of simulations are defaulted. Note that since simulations and iterations are used within the function **OneSampleNormal1.Design**, computing time is dependent upon the number of trials in the simulation and the tuned sample size. It is suggested to provide an appropriate range of the sample sizes (e.g., $[10, 50]$) for a large number of simulations (e.g., sim $= 1000$). The output for this example is displayed below.

```
# DIP approach
DIP.design <- OneSampleNormal1.Design(list(1,0), nmin=10, nmax=50,
   mu0=250, mu1=294, var=100, ps=0.95, pf=0.05, alternative = "
   greater")
DIP.design$planned_samle_size
[1] 10
DIP.design$power
[1] 1
DIP.design$type_I_error
[1] 0.057
DIP.design$expected_sample_size
[1] 10

# Standard prior Normal(250, 100/5)
Bayes.design <- OneSampleNormal1.Design(list(2,5), nmin=10, nmax=50,
    mu0=250, mu1=294, var=100, ps=0.95, pf=0.05, alternative = "
   greater")
Bayes.design$planned_sample_size
[1] 16
```

```
Bayes.design$power
[1] 1
Bayes.design$type_I_error
[1] 0.050
Bayes.design$expected_sample_size
[1] 10
```

The results indicate that the DIP approach requires fewer planned sample size to achieve an admissible design, compared to a less informative Bayesian prior.

5.5 Summary and Discussion

The **BayesDIP** package was developed to perform study designs capable of calculating the minimum planned sample size and the expected sample size given admissible significant level, power and stopping rule cutoffs for Bayesian early termination Phase II clinical trials when the outcomes have a *Binomial*, *Normal*, or *Poisson* distribution. The package incorporates Bayesian early termination rules with an innovative DIP approach to achieve the admissible designs with smaller or comparable sample sizes, without relying on historical data or optimistic prior.[21][3][16][12] For a comprehensive use of the package, we also provide user-defined standard Bayesian priors in the package. As for a simulation-based package, users can customize the designs with options to set the priors, to specify the assumptions for the hypothesis test, to tune the stopping boundaries, to give the target power and significant level, and to set the number of trials in the simulation. Ten functions are provided in the package and allow the user to study observed data with *Binomial*, *Normal*, and *Poisson* distributions, with one-sample or two-sample cases. The models behind the DIP and the standard Bayesian priors are specified and described in details, and practical applications are illustrated by clinical study examples. The **BayesDIP** provides an comprehensive, all-in-one solution for study designs for common distributions encountered in Phase II clinical trials. We possible develop study designs for *Weibull* distribution and *Multinomial* distribution in the future.

5.6 Computational Details

The results in this paper were obtained using R.4.0.3. R itself and R packages used are available from the CRAN at https://CRAN.R-project.org/.

CHAPTER 6

DISCUSSION

We have presented the rationale of the decreasingly informative prior in the Bayesian framework and applied it to featuring early termination designs for Phase II clinical trials. The decreasingly informative prior is a functionalized prior distribution that centers the prior at skeptical null values and equates the prior ESS in terms of the unaccrued sample size, which refrains the termination at early stage of a trial and allows the posterior distribution to be increasingly informed by the likelihood function and less by the skeptical prior information as subjects are accrued. Our first aim applied DIP in Bayesian early termination Phase II trial designs for single-parameter models, where the outcomes could be modeled with a *Bernoulli*, *Poisson*, or *Normal* distribution. We compared the sample size under admissible designs, which is defined by the target of at least 80% power and at most 5% type I error rate, between the DIP approach and the traditional Bayesian approaches by Thall and Simon for one-sample and two-sample cases. In one-sample cases, the DIP performed better in terms of smaller sample size and control of type I error under admissible designs for different settings. In two-sample *Bernoulli* cases, we concluded that the DIP performed better for moderate to large response rates. Before utilizing the DIP in Bayesian early termination designs for two-parameter models in Aim 2, we introduced the calculation of prior ELIR ESS of single parameter on multivariate cases, which was extended from the univariate ELIR ESS approach proposed by Neuenschwander et al[11]. After that, we incorporated the DIP to Bayesian designs for two-parameter models, where the response was assumed to follow a *Normal* distribution with both mean and variance unknown, or a *Weibull* distribution with both scale and shape parameters unknown. By comparing with tradtional Bayesian

approaches, we found that the DIP approach performed better in *Normal* cases when the effect sizes are medium or high. Aim 3 introduced an R package **BayesDIP** that was developed for an comprehensive and integrated solution for common distributions illustrated in Aim 1 and Aim 2. The package includes ten functions that are capable of calculating sample size by customizing the thresholds of admissible designs and stopping decision cutoffs for Bayesian early termination Phase II trials.

One contribution of our research is to establish a decreasingly informative prior for Bayesian early termination trials, and this prior distribution avoids the necessity for a traditional prior selection which is based mainly on clinical opinions or historical data. The key of the DIP approach is to obtain the prior effective sample size and the mean, median or mode of the prior distribution, and then setting up the simultaneous equations or simulations to derive the parameters of the DIP distribution. The feature that the DIP is centered at null value early in the trial prevents the trail terminated erroneously too early. Our proposed DIP approach can be used in a broad scope of models for Bayesian early termination Phase II trial designs, including but not limited to *Bernoulli*, *Normal*, *Poisson*, *Multinomial*, and *Weibull* models. Generally, we found that compared to the traditional Bayesian approach, the DIP approach employs similar or smaller sample size to reach the desired 80% power and control a better type I error, which benefits in cutting down on the trial duration and financial costs and addresses the issue of potentially inflated type I error rates in Bayesian continuous monitoring designs.

We investigated multiple scenarios for each model with a range of effect sizes, which reflect a broad scope of realistic scenarios. For some scenarios with low effect sizes, it is difficult to attain a design with at least 80% power and a control of 5% type I error, within a reasonable sample size for Phase II trials. In these scenarios, the DIP approach does not present a significant advantage over the standard Bayesian approach. To achieve a reliable application of the DIP approach in power analysis for Phase II trial designs, we recommend

employing the DIP at medium or high effect sizes. Further, we recommend utilizing our R package **BayesDIP** to see the subtle changes prompted by the choice of the DIP or the regular Bayesian priors. Since simulation is used within each function, in part, computational time depends on the number of trials in the simulation. It is suggested that a small value of trials be used, such as 100, to determine an initial range for the sample size (e.g., N) and stopping decision cutoffs p_s and p_f. The efficacious threshold p_s is suggested to be set to a low value if we know there are only a few, such as 20, subjects available. When the sample size cannot be increased, we can also adjust the decision boundaries p_s and p_f to find an admissible design. Once an appropriate parameters are determined, the number of trials should be increased, such as 1000, to output more stable and accurate results.

In addition to Thall and Simon posterior probability design, the decreasingly informative prior can also be used in Lee and Liu predictive posterior probability design [8]. The logic of functionalizing the DIP is similar except that we have to consider the current observed data in the trial. Additionally, the DIP approach can also be applied to Bayesian sequential monitoring designs with multiple outcomes, including, for example, an efficacy and a toxicity outcome [23] with $Beta - Binomial$ model, which is the special case of $Dirichlet - Multinomial$ model, is used to track the posterior probability of efficacy and toxicity independently. In this case, the functionalization of the DIP is based on the prior effective sample size of the single parameter of the multivariate models, see Chapter 3. Furthermore, other future considerations include:

- Extension of the DIP approach in Bayesian adaptive designs of hierarchical models
- Maintain and optimize the published R package **BayesDIP**, including adding more functions, for example, $Weibull$ distribution with time-to-event outcomes.
- Consider methods to reduce execution time, such as compiling the code using C/C++.
- Create an R shiny application that provides an user-friendly platform to implement

the DIP in Bayesian sequential monitoring designs of Phase II trials.

Appendix A

R CODE RELEVANT TO CHAPTER 2

One-sample Normal distribution (variance known) with DIP approach

```
# prior parameters
mu0<-100
sig<-15

for( N in seq(from=10, to=100, by=1) ){
  for(mu1 in seq(from=90, to=100, by=5)){
    for (p_f in seq(from=0.01, to=0.10, by=0.01) ){
      for (p_s in seq(from=0.80, to=0.99, by=0.01) ){
        s<-1000
        cat1s<-0
        cat1f<-0
        for(k in seq(from=1,to=s,by=1)){
          y<-NULL
          j<-0
          cat<-0
          cats<-0
          catf<-0
          pp_stop<-0.5

          while(cat==0){
            j<-j+1
            y<-append(y,rnorm(1,mu1,sig))
            y_sd<-sd(y)
            y_mean<-mean(y)
            n<-length(y) # number of accrued subjects

            #Early Termination Trigger (With DIP)
            if(j>=10){
              mu1_s<-rnorm(1000,((N-n)*mu0+n*y_mean)/N,sig/sqrt(N))
              mu1_s[is.na(mu1_s)]<-mean(y)
              pp_stop<-sum(mu1_s<mu0-d)/1000
            }
            if(pp_stop>=p_s){cats<-1}
            if(pp_stop<p_f){catf<-1}
            cat<-cats+catf
            if(j==N){cat<-1}
          }
```

```
          #Calculation of posterior probability of efficacy
          if(cats==1){cat1s<-cat1s+1}
          if(cats==0){cat1s<-cat1s}
          if(catf==1){cat1f<-cat1f+1}
          if(catf==0){cat1f<-cat1f}
        }
        power<-append(power,cat1s/s)
        N_v<-append(N_v, N)
        ss_v<-append(ss_v,j)
        p_f_v<-append(p_f_v,p_f)
        p_s_v<-append(p_s_v,p_s)
        mu0_v<-append(mu0_v,mu0)
        mu1_v<-append(mu1_v,mu1)
        sig_v<-append(sig_v,sig)
        data.normalDIP1 <- cbind("DIP", mu1_v, mu0_v, sig_v, N_v, ss_v, p_f_v, p_s_v, power)
      }
    }
  }
}
```

Two-sample Bernoulli distribution with DIP approach

```
for( N in seq(from=100, to=200, by=1) ){
  for (p_1 in seq(from=p_2, to=0.15, by=0.05)) {
    for (p_f in seq(from=0.01, to=0.10, by=0.01) ){
      for (p_s in seq(from=0.80, to=0.99, by=0.01) ){
        N1 <- N/2 # planned sample size for group 1
        N2 <- N/2 # planned sample size for group 2
        s=1000
        cat1s<-0
        cat1f<-0
        for (k in seq(from=1, to=s, by=1)) {
          y<-NULL
          Group<-NULL
          r<-0.5 # equal allocation
          j<-0
          cat<-0
          cats<-0
          catf<-0
          pp_stop<-0.5
          while(cat==0)
          {
            j<-j+1
            u<-runif(1,min = 0,max = 1)
            if(u<=r){
```

```
            Group=append(Group,1)
            y<-append(y,rbinom(1,1,p_1))}
          if(u>r){
            Group=append(Group,0)
            y<-append(y,rbinom(1,1,p_2))}
          Matd<-as.data.frame(cbind(y,Group))
          y1<-Matd$y[which(Matd$Group==1)]
          y2<-Matd$y[which(Matd$Group==0)]
          n1<-length(y1)
          n2<-length(y2)
          sn1 <- length(y1[y1==1]) # number of successes
          sn2 <- length(y2[y2==1])
          if(sn1>0 & sn2>0)
          {
            p_0 <- rbeta(1000,1,1) # hyper-prior
            p1_s<-rbeta(1000,1+sum(y1)+p_0*(N1-n1),1+(n1-sum(y1))+(1-p_0)*(N1-n1))
            p2_s<-rbeta(1000,1+sum(y2)+p_0*(N2-n2),1+(n2-sum(y2))+(1-p_0)*(N2-n2))
            p1_s[is.na(p1_s)]<-sum(y1)/n1
            p2_s[is.na(p2_s)]<-sum(y2)/n2
            pp_stop<-sum(p1_s>p2_s+d)/1000
          }
          if(pp_stop>=p_s){cats<-1}
          if(pp_stop<p_f){catf<-1}
          cat<-cats+catf
          if(j==N){cat<-1}
        }
        if(cats==1){cat1s<-cat1s+1}
        if(cats==0){cat1s<-cat1s}
        if(catf==1){cat1f<-cat1f+1}
        if(catf==0){cat1f<-cat1f}
      }
      p_1_v<-append(p_1_v,p_1)
      p_2_v<-append(p_2_v,p_2)
      power<-append(power,cat1s/s)
      N_v<-append(N_v, N)
      ss1_v <- append(ss1_v, n1)
      ss2_v <- append(ss2_v, n2)
      p_f_v<-append(p_f_v,p_f)
      p_s_v<-append(p_s_v,p_s)
      data.dip1 <- cbind(p_1_v, p_2_v, N_v, ss1_v, ss2_v, p_f_v, p_s_v, power)
    }
  }
 }
}
```

Appendix B

CHAPTER 2 SUPPLEMENTAL MATERIALS

Table S.1. Simulation Results for Bernoulli Cases - One Sample ($p_0 = 0.3$)

Model	p_0	p_1	Sample Size[a]	Futility	Efficacy	Power	Type I Error[b]
DIP	0.3	0.35	98	0.09	0.80	0.792	0.410
Bayesian ($Beta(1,1)$)	0.3	0.35	77	0.03	0.83	0.802	0.480
Bayesian ($a+b=2$)	0.3	0.35	98	0.02	0.83	0.800	0.453
Bayesian ($a+b=6$)	0.3	0.35	97	0.02	0.80	0.803	0.472
Bayesian ($a+b=10$)	0.3	0.35	96	0.04	0.81	0.809	0.460
DIP	0.3	0.40	100	0.03	0.95	0.804	0.097
Bayesian ($Beta(1,1)$)	0.3	0.40	98	0.07	0.97	0.806	0.174
Bayesian ($a+b=2$)	0.3	0.40	96	0.01	0.96	0.806	0.167
Bayesian ($a+b=6$)	0.3	0.40	100	0.07	0.95	0.802	0.156
Bayesian ($a+b=10$)	0.3	0.40	97	0.01	0.95	0.819	0.145
DIP	0.3	0.45	65	0.05	0.97	0.828	0.050
Bayesian ($Beta(1,1)$)	0.3	0.45	75	0.04	0.99	0.828	0.057
Bayesian ($a+b=2$)	0.3	0.45	72	0.02	0.99	0.813	0.052
Bayesian ($a+b=6$)	0.3	0.45	79	0.06	0.99	0.812	0.050
Bayesian ($a+b=10$)	0.3	0.45	76	0.03	0.98	0.830	0.050
DIP	0.3	0.50	36	0.07	0.97	0.808	0.050
Bayesian ($Beta(1,1)$)	0.3	0.50	44	0.05	0.99	0.819	0.050
Bayesian ($a+b=2$)	0.3	0.50	47	0.02	0.99	0.812	0.050
Bayesian ($a+b=6$)	0.3	0.50	48	0.07	0.99	0.806	0.050
Bayesian ($a+b=10$)	0.3	0.50	40	0.05	0.97	0.805	0.050

[a] The Planned Sample Size

[b] Type I error is calculated under the null hypothesis $p_1 = p_0$

Table S.2. Simulation Results for Bernoulli Cases - One Sample ($p_0 = 0.5$)

Model	p_0	p_1	Sample Size[a]	Futility	Efficacy	Power	Type I Error[b]
DIP	0.5	0.55	94	0.08	0.80	0.767	0.399
Bayesian ($Beta(1,1)$)	0.5	0.55	96	0.05	0.82	0.802	0.480
Bayesian ($a+b=2$)	0.5	0.55	84	0.03	0.82	0.800	0.499
Bayesian ($a+b=6$)	0.5	0.55	95	0.03	0.83	0.801	0.450
Bayesian ($a+b=10$)	0.5	0.55	92	0.04	0.81	0.800	0.486
DIP	0.5	0.60	98	0.06	0.93	0.806	0.124
Bayesian ($Beta(1,1)$)	0.5	0.60	97	0.02	0.96	0.816	0.189
Bayesian ($a+b=2$)	0.5	0.60	98	0.03	0.96	0.809	0.196
Bayesian ($a+b=6$)	0.5	0.60	97	0.07	0.95	0.803	0.178
Bayesian ($a+b=10$)	0.5	0.60	90	0.01	0.94	0.800	0.169
DIP	0.5	0.65	68	0.04	0.97	0.810	0.050
Bayesian ($Beta(1,1)$)	0.5	0.65	82	0.09	0.99	0.806	0.056
Bayesian ($a+b=2$)	0.5	0.65	74	0.06	0.99	0.811	0.056
Bayesian ($a+b=6$)	0.5	0.65	79	0.02	0.99	0.812	0.051
Bayesian ($a+b=10$)	0.5	0.65	87	0.09	0.99	0.805	0.050
DIP	0.5	0.70	36	0.07	0.96	0.804	0.050
Bayesian ($Beta(1,1)$)	0.5	0.70	48	0.02	0.99	0.819	0.050
Bayesian ($a+b=2$)	0.5	0.70	48	0.04	0.99	0.830	0.051
Bayesian ($a+b=6$)	0.5	0.70	49	0.08	0.99	0.812	0.050
Bayesian ($a+b=10$)	0.5	0.70	45	0.04	0.98	0.822	0.050

[a] The Planned Sample Size

[b] Type I error is calculated under the null hypothesis $p_1 = p_0$

Table S.3. Simulation Results for Bernoulli Cases - One Sample ($p_0 = 0.7$)

Model	p_0	p_1	Sample Size[a]	Futility	Efficacy	Power	Type I Error[b]
DIP	0.7	0.75	100	0.06	0.80	0.769	0.373
Bayesian ($Beta(1,1)$)	0.7	0.75	81	0.05	0.82	0.804	0.443
Bayesian ($a+b=2$)	0.7	0.75	98	0.02	0.88	0.803	0.450
Bayesian ($a+b=6$)	0.7	0.75	88	0.09	0.85	0.804	0.459
Bayesian ($a+b=10$)	0.7	0.75	99	0.02	0.85	0.801	0.434
DIP	0.7	0.80	100	0.04	0.95	0.815	0.075
Bayesian ($Beta(1,1)$)	0.7	0.80	97	0.05	0.97	0.808	0.131
Bayesian ($a+b=2$)	0.7	0.80	98	0.05	0.98	0.800	0.148
Bayesian ($a+b=6$)	0.7	0.80	100	0.05	0.98	0.803	0.124
Bayesian ($a+b=10$)	0.7	0.80	94	0.05	0.97	0.803	0.114
DIP	0.7	0.85	50	0.07	0.96	0.816	0.050
Bayesian ($Beta(1,1)$)	0.7	0.85	70	0.01	0.99	0.850	0.050
Bayesian ($a+b=2$)	0.7	0.85	62	0.09	0.99	0.859	0.073
Bayesian ($a+b=6$)	0.7	0.85	63	0.02	0.99	0.834	0.050
Bayesian ($a+b=10$)	0.7	0.85	66	0.01	0.99	0.816	0.050
DIP	0.7	0.90	24	0.06	0.95	0.823	0.050
Bayesian ($Beta(1,1)$)	0.7	0.90	37	0.08	0.99	0.830	0.050
Bayesian ($a+b=2$)	0.7	0.90	27	0.10	0.99	0.802	0.050
Bayesian ($a+b=6$)	0.7	0.90	38	0.03	0.99	0.851	0.050
Bayesian ($a+b=10$)	0.7	0.90	29	0.10	0.97	0.809	0.050

[a] The Planned Sample Size

[b] Type I error is calculated under the null hypothesis $p_1 = p_0$

Table S.4. Simulation Results for Bernoulli Cases - Two Samples ($p_2 = 0.5$)

Model	p_1	p_2	Sample Size[a]	Futility	Efficacy	Power	Type I Error[b]
DIP	0.55	0.5	199	0.02	0.80	0.651	0.380
Bayesian ($Beta(1,1)$)	0.55	0.5	200	0.01	0.83	0.805	0.589
Bayesian ($a+b=2$)	0.55	0.5	191	0.01	0.82	0.802	0.604
Bayesian ($a+b=6$)	0.55	0.5	196	0.02	0.80	0.811	0.585
Bayesian ($a+b=10$)	0.55	0.5	196	0.02	0.80	0.787	0.572
DIP	0.60	0.5	174	0.09	0.82	0.809	0.319
Bayesian ($Beta(1,1)$)	0.60	0.5	196	0.03	0.91	0.804	0.390
Bayesian ($a+b=2$)	0.60	0.5	190	0.05	0.90	0.804	0.402
Bayesian ($a+b=6$)	0.60	0.5	187	0.05	0.89	0.819	0.374
Bayesian ($a+b=10$)	0.60	0.5	194	0.01	0.89	0.808	0.342
DIP	0.65	0.5	190	0.05	0.92	0.807	0.115
Bayesian ($Beta(1,1)$)	0.65	0.5	193	0.01	0.97	0.809	0.182
Bayesian ($a+b=2$)	0.65	0.5	193	0.02	0.97	0.800	0.181
Bayesian ($a+b=6$)	0.65	0.5	195	0.02	0.96	0.802	0.148
Bayesian ($a+b=10$)	0.65	0.5	197	0.02	0.96	0.803	0.139
DIP	0.70	0.5	166	0.04	0.96	0.810	0.050
Bayesian ($Beta(1,1)$)	0.70	0.5	168	0.01	0.99	0.823	0.066
Bayesian ($a+b=2$)	0.70	0.5	166	0.02	0.99	0.810	0.071
Bayesian ($a+b=6$)	0.70	0.5	175	0.01	0.99	0.802	0.050
Bayesian ($a+b=10$)	0.70	0.5	193	0.05	0.99	0.800	0.051

[a] The Planned Sample Size

[b] Type I error is calculated under the null hypothesis $p_1 = p_2$

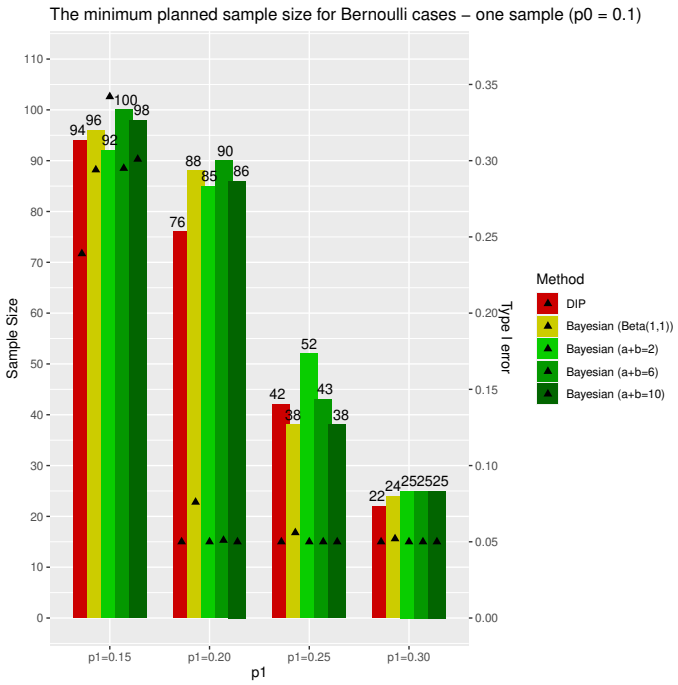

Fig. S.1. Simulation Results for Bernoulli Cases - One Sample ($p_0 = 0.1$)

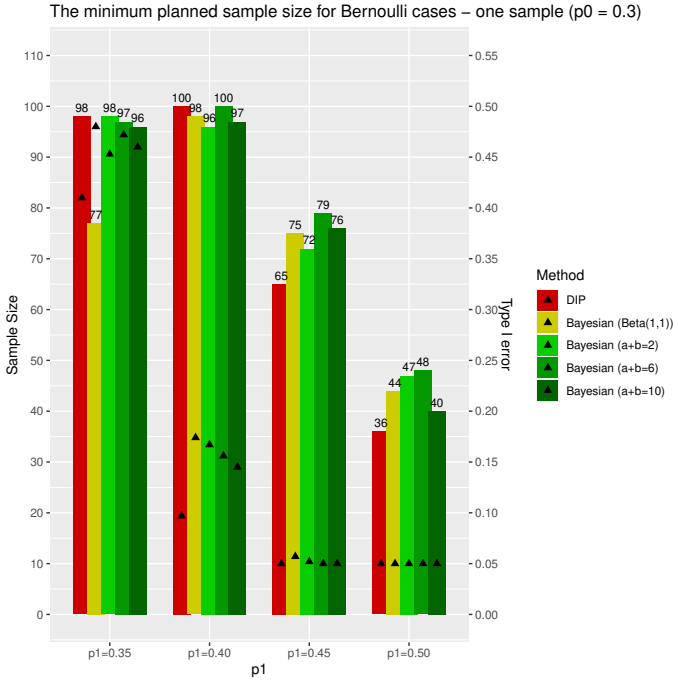

Fig. S.2. Simulation Results for Bernoulli Cases - One Sample ($p_0 = 0.3$)

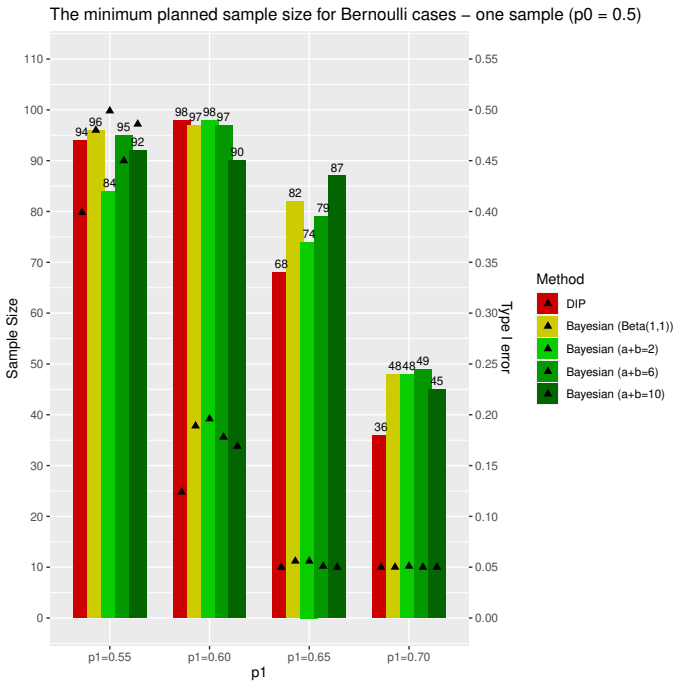

Fig. S.3. Simulation Results for Bernoulli Cases - One Sample ($p_0 = 0.5$)

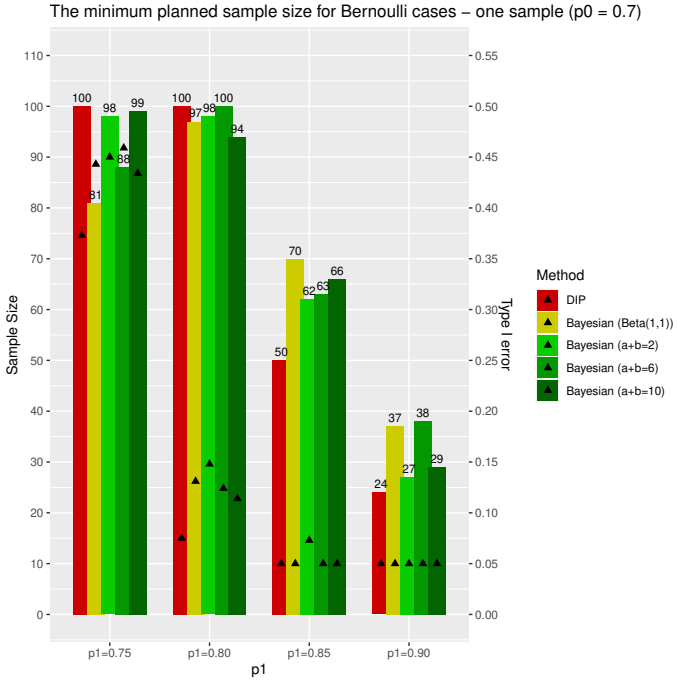

Fig. S.4. Simulation Results for Bernoulli Cases - One Sample ($p_0 = 0.7$)

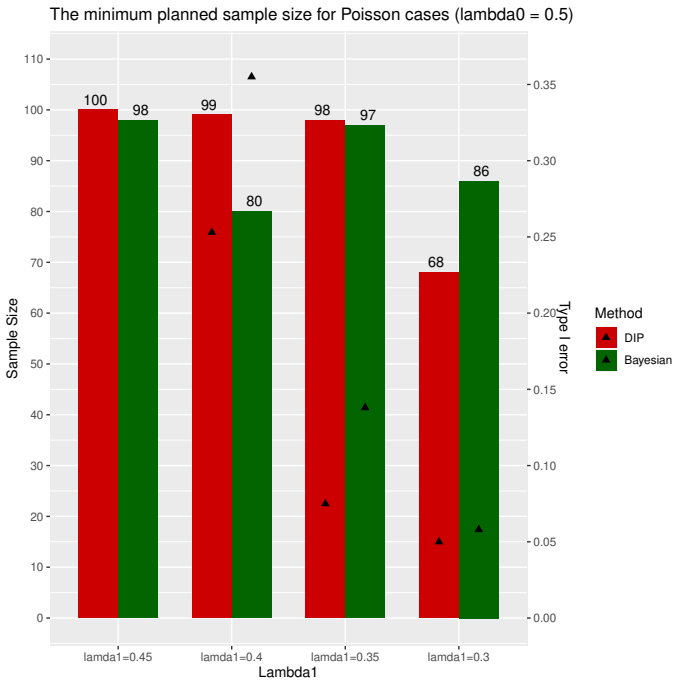

Fig. S.5. Simulation Results for Poisson Cases ($\lambda_0 = 0.5$)

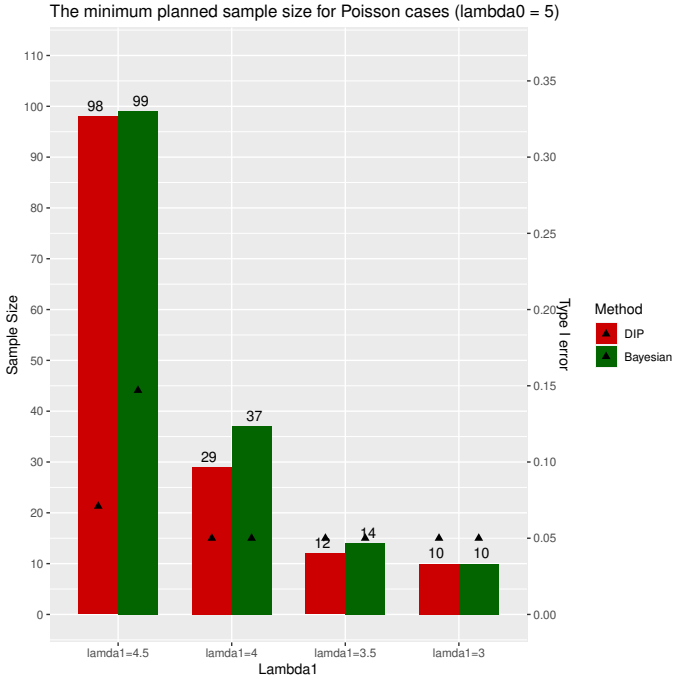

Fig. S.6. Simulation Results for Poisson Cases ($\lambda_0 = 5$)

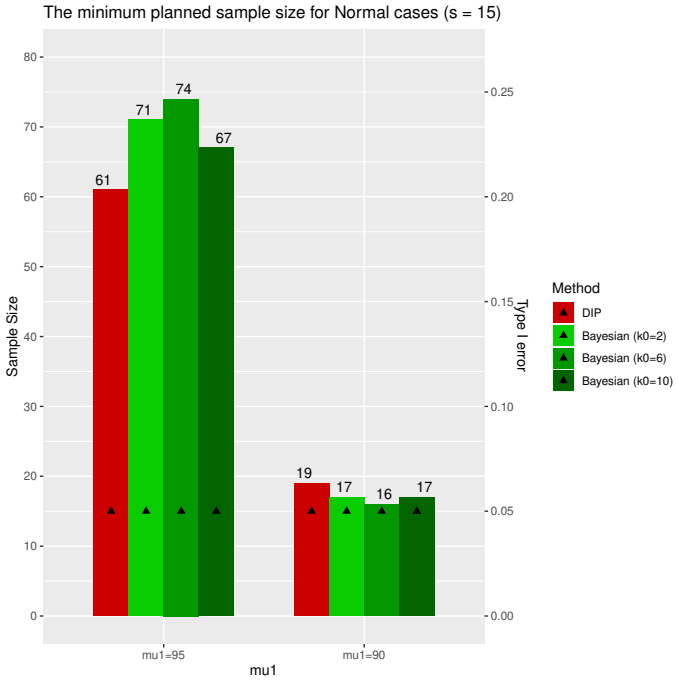

Fig. S.7. Simulation Results for Normal Cases with Known Variance ($s = 15$)

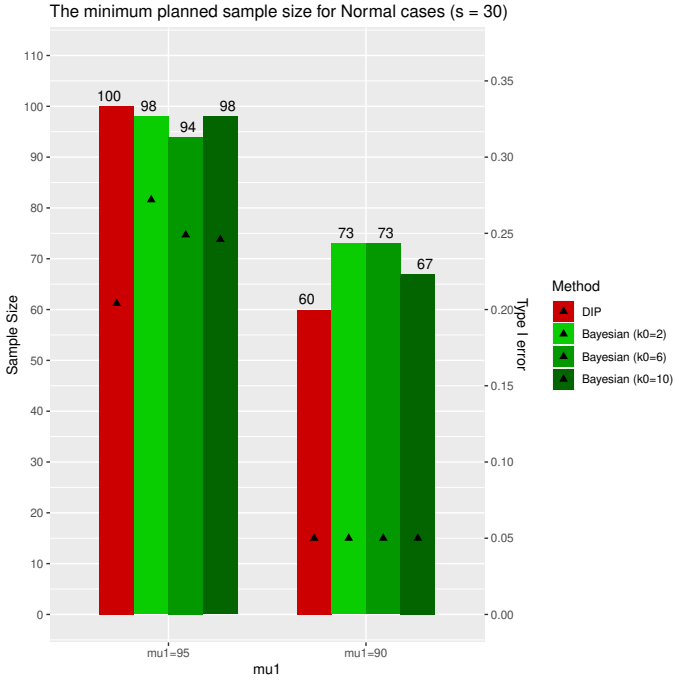

Fig. S.8. Simulation Results for Normal Cases with Known Variance ($s = 30$)

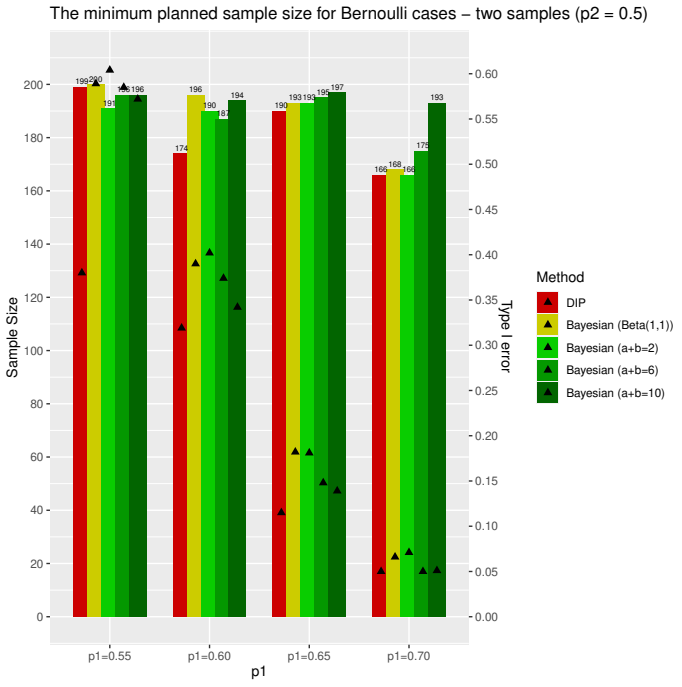

Fig. S.9. Simulation Results for Bernoulli Cases - Two Samples ($p_0 = 0.5$)

Appendix C

R CODE RELEVANT TO CHAPTER 3

Assess Predictive Consistency - Variance Ratio and Precision Ratio methods

```
# Monte Carlo sampling approach to calculate the posterior
rm(list = ls())
set.seed(202210)
nu0 = 10
k0 = 10
s20 = 0.01
mu0 = 1.9
N = 1000
s21 <- 1/rgamma(1, nu0/2, nu0*s20/2)
mu1 <- rnorm(1, mu0, sqrt(s21/k0))
y<-rnorm(N, mean=mu1, sd=sqrt(s21)) # observed data
n<-length(y)
s2<-var(y)
ybar<-mean(y)
kn <- k0+n
nun <- nu0+n
s2n<-(1/nun)*(nu0*s20+(n-1)*s2+k0*n*(ybar-mu0)^2/(kn))
s2.post<-1/rgamma(10000,nun/2, s2n*nun/2)
mu.post<-rnorm(10000,(k0*mu0+n*ybar)/kn,sqrt(s2.post/kn))
mu.p <- mean(mu.post)
s2.p <- mean(s2.post)

# ------------------------------Calculate ESS-----------------------------------
# VR
VR.mu = (kn*nun*s2n) / ( (nun - 2)*s2.p )
VR.mu.prior = (k0*nu0) / (nu0 - 2)
print(round(c(VR.mu.prior, VR.mu), 0))

VR.s2 = nun - 2
VR.s2.prior = nu0 - 2
print(round(c(VR.s2.prior, VR.s2), 0))

# PR
PR.mu = (kn*s2n) / s2.p
PR.mu.prior = k0
print(round(c(PR.mu.prior, PR.mu), 0))
```

```
PR.s2 = ( (nun - 2)^2*(nun - 4) ) / ( nun * (nun + 2) )
PR.s2.prior = ( (nu0 - 2)^2*(nu0 - 4) ) / ( nu0 * (nu0 + 2) )
print(round(c(PR.s2.prior, PR.s2), 0))
```

Assess Predictive Consistency - ELIR method

```
# Monte Carlo sampling approach to calculate the # Monte Carlo sampling approach to calculate t
set.seed(202210)
rm(list = ls())
nu0 = 1
k0 = 1
s20 = 0.01
mu0 = 1.9
N = 50

elir <- NULL
elir.pr <- NULL
n.sim <- 10000

for (ss in 1:n.sim) {
  s21 <- 1/rgamma(1, nu0/2, nu0*s20/2)
  mu1 <- rnorm(1, mu0, sqrt(s21/k0))
  y<-rnorm(N, mean=mu1, sd=sqrt(s21)) # observed data

  n<-length(y)
  s2<-var(y)
  ybar<-mean(y)
  kn <- k0+n
  nun <- nu0+n
  s2n<-(1/nun)*(nu0*s20+(n-1)*s2+k0*n*(ybar-mu0)^2/(kn))
  s2.post<-1/rgamma(10000,nun/2, s2n*nun/2)
  mu.post<-rnorm(10000,(k0*mu0+n*ybar)/kn,sqrt(s2.post/kn))
  mu.p <- mean(mu.post)
  s2.p <- mean(s2.post)


  # ------------------------------Calculate Expected ESS----------------------------------

  # ELIR
  ELIR.pr = -nu0 - 3 + 2 * k0  * (mu1 - mu0)^2 / s21 + 2*nu0*s20/s21 # prior ESS
  ELIR.s2 = N - nu0 - 3 + 2 * k0 * (mu.p - mu0)^2 / s2.p + 2*nu0*s20/s2.p # posterior ESS
  elir.pr = append(elir.pr, ELIR.pr)
  elir <- append(elir, ELIR.s2)

}
```

```
# Expected ELIR
round(mean(elir.pr), 0)
round(mean(elir), 0)
```

Appendix D

R CODE RELEVANT TO CHAPTER 4

One-sample Normal distribution (both mean and variance unknown) with DIP approach

```
# Low prior variance and Low observed response variability
sig0<-12 #prior sample sd
sig1<-12 #observed sample sd
d<-0
# simulation size
n.sim = 10000
# mean
meanAll = rbind(c(100,90),
                c(100,95),
                c(100,98),
                c(100,100))
mean.labels = apply(meanAll, 1, function(e)
  paste("mu0=", e[1], " mu1=", e[2]))
# planned N
Nall = seq(from=10, to=100, by=1)
# tune parameter
pfAll = seq(from=0.01, to=0.10, by=0.01)
psAll = seq(from=0.80, to=0.99, by=0.01)
outAll = matrix(ncol=11, nrow=nrow(rbind(meanAll))*length(psAll)*length(pfAll)*length(Nall))
colnames(outAll) = c("mu0", "mu1", "sig0", "sig1", "N",
                     "p_s", "p_f", "power", "n", "SD", "futility")
check = 0
for( j.mean in 1:nrow(meanAll)){
  for(j.N in 1:length(Nall)){
    for(j.ps in 1:length(psAll)){
      for(j.pf in 1:length(pfAll)){
     check = check+1
     mu0 = meanAll[j.mean, 1]
     mu1 = meanAll[j.mean, 2]
     N = Nall[j.N]
     p_s = psAll[j.ps]
     p_f = pfAll[j.pf]
        v<-1000
        cat1s<-0
        cat1f<-0
        sample_size<-NULL
        mu1_s<-NULL
```

```
s1_inv<-NULL
s1<-NULL
for(i in 1:n.sim){
  y<-NULL
  j<-0
  cat<-0
  cats<-0
  catf<-0
  pp_stop<-0.5
  while(cat==0){
    j<-j+1
    y<-append(y,rnorm(1,mu1,sig1))
    y_sd<-sd(y)
    y_mean<-mean(y)
    n1<-length(y)
    #Early Termination Trigger (With Bayesian Prior)
    if(j>=10){
      k0<-N-j #DIP
      nu0<-3+(N-j) #DIP - nuisance parameter
      kn<-k0+j
      nun<- nu0+j
      sig1_n<-(1/nun)*(nu0*sig0^2+(j-1)*y_sd^2+k0*(j)*(1/kn)*(y_mean-mu0)^2)
      s1_inv<-rgamma(v,nun/2,nun*sig1_n/2)
      s1<-1/s1_inv
      mu1_s<-rnorm(v,(k0*mu0+j*y_mean)/kn,sqrt(s1/kn))
      mu1_s[is.na(mu1_s)]<-mean(y)
      pp_stop<-sum(mu1_s<mu0-d)/v
    }
    if(pp_stop>=p_s){cats<-1}
    if(pp_stop<p_f){catf<-1}
    cat<-cats+catf
    if(j==N){cat<-1}
  }
  #Calculation of posterior probability of efficacy
  if(cats==1){cat1s<-cat1s+1}
  if(cats==0){cat1s<-cat1s}
  if(catf==1){cat1f<-cat1f+1}
  if(catf==0){cat1f<-cat1f}
  sample_size = append(sample_size, j)
}
outAll[check,c("power")] = cat1s/n.sim
outAll[check,c("n")] = round(mean(sample_size),1.)
outAll[check,c("SD")] = round(sd(sample_size),2.)
outAll[check,c("futility")] = cat1f/n.sim
outAll[check,c("mu0")] = mu0
```

```
        outAll[check,c("mu1")] = mu1
        outAll[check,c("sig0")] = sig0
        outAll[check,c("sig1")] = sig1
        outAll[check,c("N")] = N
        outAll[check,c("p_s")] = p_s
        outAll[check,c("p_f")] = p_f
     }
    }
  }
}
```

One-sample Weibull distribution (both shape and rate unknown) with DIP approach

```
# Here the parameters
rate0 <- 0.69
rate1 <- 0.29
shape0 <- 0.5 # common shape with different scale parameters
shape1 <- 0.5
nsim <- 1000
rp <- 1
jags.file <- paste("wb_",shape0,"_",rate1,"_",rp,".txt",sep = "")

library(coda)
library(rjags)

#Define the model
cat("model
{
    for(i in 1:j){
    is.censoredp[i] ~ dinterval(t[i], t.cens[i])
    t[i] ~ dweib(shape, (1/rate)^shape)
    }
    rate ~ dgamma(a,b)
    shape ~ dgamma(a,b)
}", file=jags.file
)

# the median survival time for the Weibull distribution
m0 <- (1/rate0*log(2))^(1/shape0)
maxT<-4  #maximum time

N_v<-NULL
shape0_v<-NULL
shape1_v<-NULL
rate0_v<-NULL
rate1_v<-NULL
```

```
p_f_v<-NULL
p_s_v<-NULL
power_v<-NULL
fut_stop<-NULL
ss_v <- NULL # sample size
std_v <- NULL # standard deviation

for(N in seq(from=10, to=100, by=1) ){
    for (p_f in seq(from=0.01, to=0.10, by=0.01) ){
      for (p_s in seq(from=0.80, to=0.99, by=0.01) ){

        s<-nsim
        cat1s<-0
        cat1f<-0
        check<-0
        sample_size<-NULL
        datalist<-list()

        for(k in seq(from=1,to=s,by=1)){
          check<-check+1

          surt<-NULL
          surt.cens<-NULL

          t<-NULL # survival time to be passed into JAGS
          t.cens<-NULL # censoring time to be passed into JAGS

          j<-0
          cat<-0
          cats<-0
          catf<-0
          pp_stop<-0.5

          sample_size <- NULL

          while(cat==0){

            # enroll subjects
            j<-j+1
            surt<-append(surt, rweibull(1,shape=shape1,scale=1/rate1))

            if((j>=10&j<=N)){

              n1<-length(surt)
              surt.cens<-rpois(n1,2) # censoring time - predetermined
```

```
      surt.cens[surt.cens >= maxT] <- maxT # right censoring

      t <- surt
      t[surt >= surt.cens] <- NA # censored observations are NA in the survival times

      t.cens <- surt.cens
      t.cens[surt < surt.cens] <- 0 # If a failure was observed

      # data<-cbind(t, t.cens)

      #JAGS
      a <- N-j+1
      b <- (N-j+1)/m0
      datalist <- list(t=t, t.cens = t.cens, j=j, a=a, b=b)
      jagsModel <- jags.model(file=jags.file, data=datalist, n.chains=1, n.adapt=500)
      codaSample <- coda.samples(jagsModel, variable.names = c("shape", "rate"),
      n.iter=2000)
      rate.s<-as.vector(unlist(codaSample[,1]))
      shape.s<-as.vector(unlist(codaSample[,2]))

      m1 <- (1/rate.s*log(2))^(1/shape.s)
      pp_stop<-mean(ifelse(m1>m0,1,0))
    }

    if(pp_stop>=p_s){cats<-1}
    if(pp_stop<p_f){catf<-1}
    cat<-cats+catf
    if(j==N){cat<-1}
  }

  #Calculation of posterior probability of efficacy
  if(cats==1){cat1s<-cat1s+1}
  if(cats==0){cat1s<-cat1s}
  if(catf==1){cat1f<-cat1f+1}
  if(catf==0){cat1f<-cat1f}
  sample_size = append(sample_size, j)
}

shape0_v<-append(shape0_v,shape0)
shape1_v<-append(shape1_v,shape1)
rate0_v<-append(rate0_v,rate0)
rate1_v<-append(rate1_v,rate1)
p_f_v<-append(p_f_v,p_f)
p_s_v<-append(p_s_v,p_s)
power_v<-append(power_v,cat1s/nsim)
```

```
        fut_stop<-append(fut_stop, cat1f/nsim)
        N_v<-append(N_v, N)
        ss_v <- append(ss_v, mean(sample_size))
        std_v <- append(std_v, sd(sample_size))

      }
    }
}
```

Appendix E

CHAPTER 4 SUPPLEMENTAL MATERIALS

Table S.5. Simulation Results for Normal Cases (Mean and Variance Unknown) - High Prior Variance and High Observed Reponse Variability

Model	μ_0	μ_1	σ_0	σ_1	N[a]	n (SD)[b]	Futility (%[c])	Efficacy	Power	Type I Error[d]
DIP	100	80	30	30	19	14 (3.3)	0.09 (0)	0.97	0.860	0.050
Bayesian ($\kappa_0 = 2$)	100	80	30	30	43	17 (8.0)	0.05 (0)	0.99	0.985	0.050
Bayesian ($\kappa_0 = 6$)	100	80	30	30	19	13 (3.5)	0.10 (0)	0.96	0.833	0.050
Bayesian ($\kappa_0 = 10$)	100	80	30	30	21	14 (4.0)	0.05 (0)	0.95	0.867	0.050
DIP	100	90	30	30	66	45 (14.2)	0.08 (0.06%)	0.97	0.841	0.051
Bayesian ($\kappa_0 = 2$)	100	90	30	30	72	41 (22.0)	0.01 (0)	0.99	0.806	0.059
Bayesian ($\kappa_0 = 6$)	100	90	30	30	81	46 (23.3)	0.09 (1.11%)	0.99	0.823	0.050
Bayesian ($\kappa_0 = 10$)	100	90	30	30	67	40 (18.7)	0.10 (0.93%)	0.98	0.804	0.056
DIP	100	95	30	30	99	60 (27.5)	0.01 (0.02%)	0.87	0.806	0.236
Bayesian ($\kappa_0 = 2$)	100	95	30	30	100	45 (34.5)	0.01 (0.65%)	0.92	0.803	0.324
Bayesian ($\kappa_0 = 6$)	100	95	30	30	97	46 (33.0)	0.02 (0.86%)	0.91	0.801	0.302
Bayesian ($\kappa_0 = 10$)	100	95	30	30	96	45 (31.8)	0.05 (1.96%)	0.90	0.800	0.299

[a] The planned sample size

[b] The expected sample size (standard deviation)

[c] The percentage of stopping the trial for futility

[d] Type I error is calculated under the null hypothesis $\mu_1 = \mu_0$

Table S.6. Simulation Results for Normal Cases (Mean and Variance Unknown) - Overestimating Prior

Model	μ_0	μ_1	σ_0	σ_1	N[a]	n (SD)[b]	Futility (%[c])	Efficacy	Power	Type I Error[d]
DIP	100	90	30	12	34	20 (3.7)	0.08 (0)	0.92	0.999	0.050
Bayesian ($\kappa_0 = 2$)	100	90	30	12	10	10 (0)	0.09 (0)	0.89	0.823	0.050
Bayesian ($\kappa_0 = 6$)	100	90	30	12	21	12 (3.0)	0.10 (0)	0.94	0.969	0.050
Bayesian ($\kappa_0 = 10$)	100	90	30	12	22	13 (3.2)	0.09 (0)	0.93	0.977	0.050
DIP	100	95	30	12	45	36 (6.3)	0.08 (0)	0.93	0.876	0.050
Bayesian ($\kappa_0 = 2$)	100	95	30	12	100	37 (21.0)	0.03 (0)	0.99	0.982	0.050
Bayesian ($\kappa_0 = 6$)	100	95	30	12	69	34 (16.6)	0.10 (0.26%)	0.98	0.929	0.050
Bayesian ($\kappa_0 = 10$)	100	95	30	12	55	32 (13.5)	0.04 (0)	0.97	0.882	0.050
DIP	100	98	30	12	100	76 (19.6)	0.01 (0.01%)	0.82	0.801	0.202
Bayesian ($\kappa_0 = 2$)	100	98	30	12	98	47 (33.1)	0.01 (0.30%)	0.91	0.801	0.297
Bayesian ($\kappa_0 = 6$)	100	98	30	12	99	48 (32.7)	0.03 (0.68%)	0.90	0.804	0.284
Bayesian ($\kappa_0 = 10$)	100	98	30	12	99	49 (32.2)	0.04 (0.62%)	0.89	0.805	0.279

[a] The planned sample size

[b] The expected sample size (standard deviation)

[c] The percentage of stopping the trial for futility

[d] Type I error is calculated under the null hypothesis $\mu_1 = \mu_0$

Table S.7. Simulation Results for Normal Cases (Mean and Variance Unknown) - Underestimating Prior

Model	μ_0	μ_1	σ_0	σ_1	N[a]	n (SD)[b]	Futility (%[c])	Efficacy	Power	Type I Error[d]
DIP	100	80	15	30	16	12 (2.5)	0.09 (0.07%)	0.98	0.807	0.055
Bayesian ($\kappa_0 = 2$)	100	80	15	30	34	16 (7.1)	0.02 (0)	0.99	0.957	0.050
Bayesian ($\kappa_0 = 6$)	100	80	15	30	71	20 (8.7)	0.03 (0)	0.99	0.999	0.050
Bayesian ($\kappa_0 = 10$)	100	80	15	30	25	15 (5.0)	0.01 (0)	0.96	0.919	0.050
DIP	100	90	15	30	64	39 (16.9)	0.10 (0.60%)	0.98	0.827	0.063
Bayesian ($\kappa_0 = 2$)	100	90	15	30	71	38 (21.9)	0.10 (2.13%)	0.99	0.804	0.064
Bayesian ($\kappa_0 = 6$)	100	90	15	30	85	46 (24.6)	0.09 (1.14%)	0.99	0.844	0.050
Bayesian ($\kappa_0 = 10$)	100	90	15	30	90	50 (24.8)	0.09 (0.69%)	0.99	0.861	0.052
DIP	100	95	15	30	100	50 (31.1)	0.09 (2.82%)	0.90	0.804	0.268
Bayesian ($\kappa_0 = 2$)	100	95	15	30	98	42 (33.2)	0.03 (2.69%)	0.92	0.802	0.318
Bayesian ($\kappa_0 = 6$)	100	95	15	30	98	44 (33.0)	0.04 (2.23%)	0.91	0.802	0.313
Bayesian ($\kappa_0 = 10$)	100	95	15	30	99	43 (32.1)	0.09 (5.08%)	0.90	0.801	0.305

[a] The planned sample size

[b] The expected sample size (standard deviation)

[c] The percentage of stopping the trial for futility

[d] Type I error is calculated under the null hypothesis $\mu_1 = \mu_0$

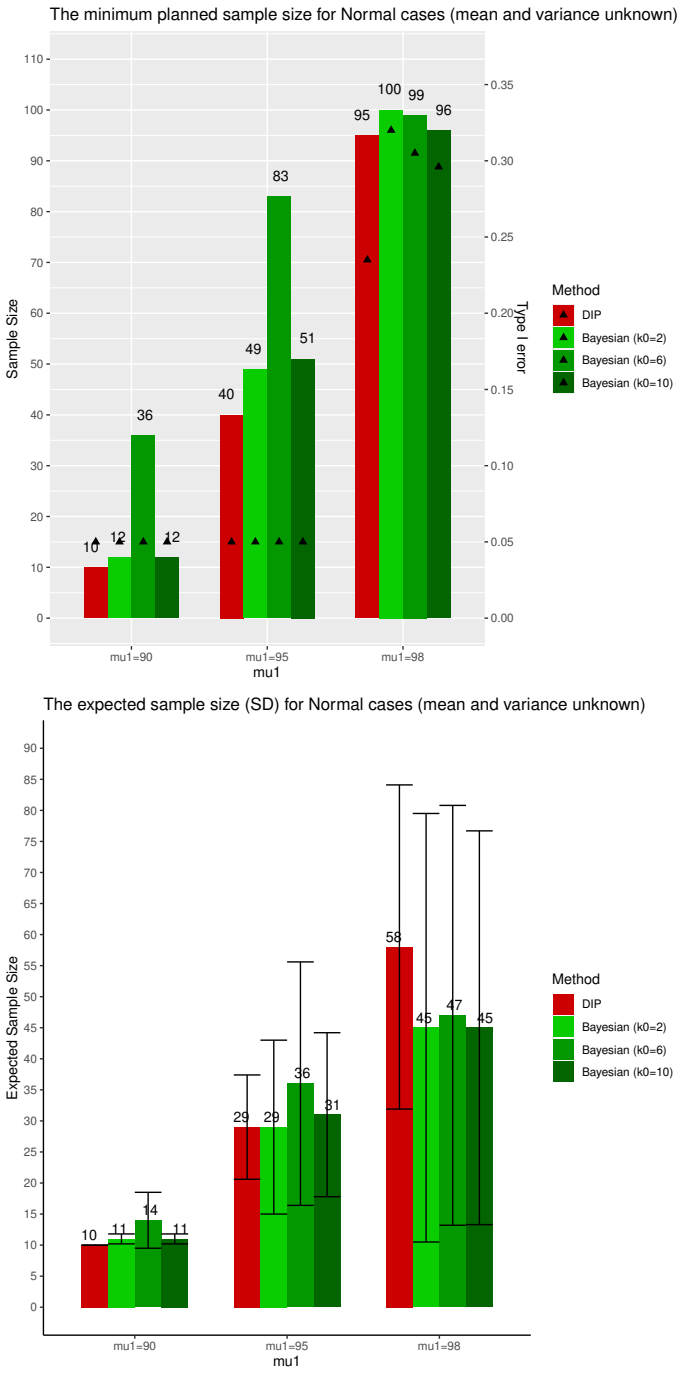

Fig. S.10. Simulation Results for Normal Distribution (Mean and Variance Unknown) - Low Prior Variance and Low Observed Response Variability

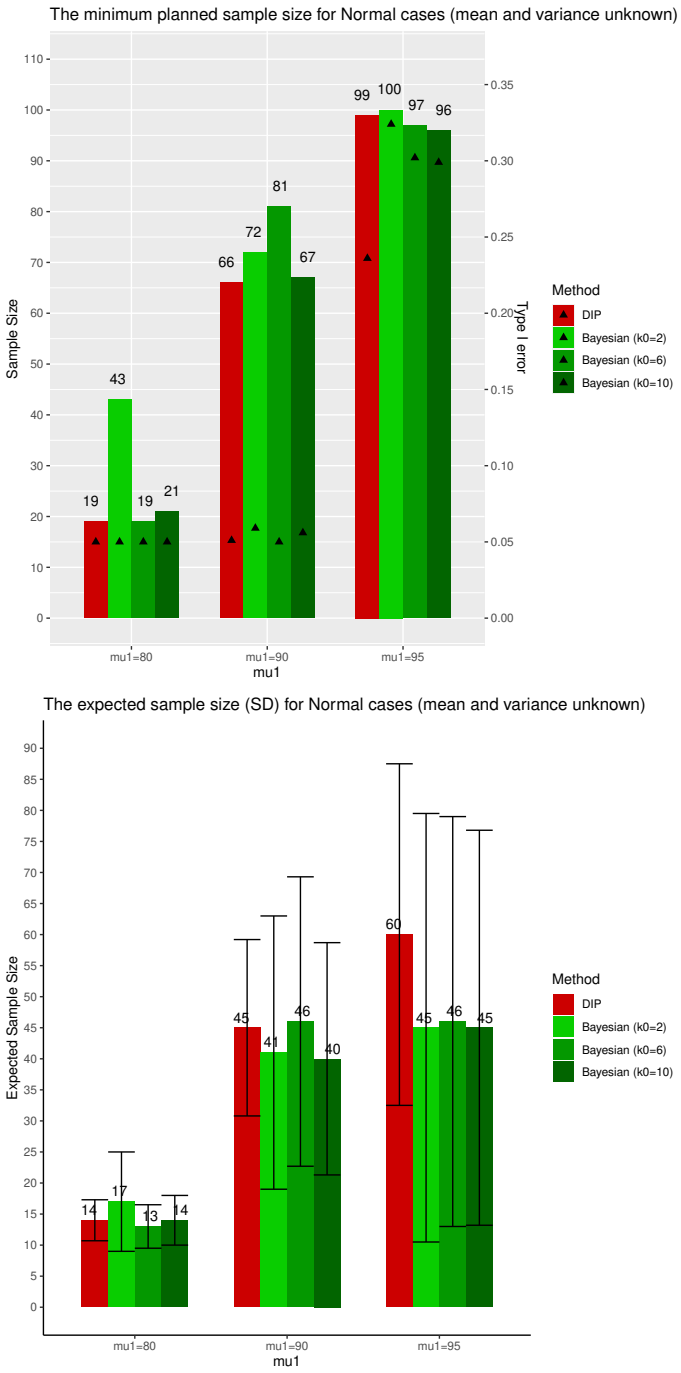

Fig. S.11. Simulation Results for Normal Distribution (Mean and Variance Unknown) - High Prior Variance and High Observed Response Variability

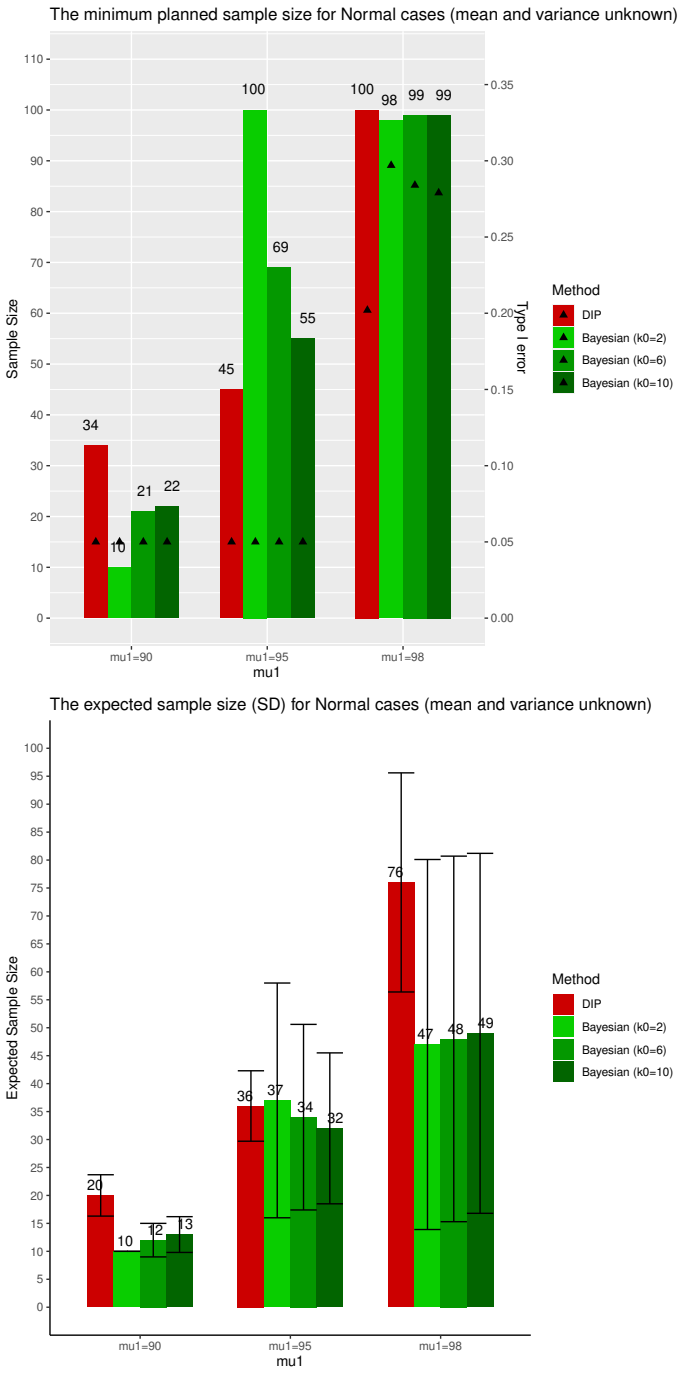

Fig. S.12. Simulation Results for Normal Distribution (Mean and Variance Unknown) - Over-estimating Prior

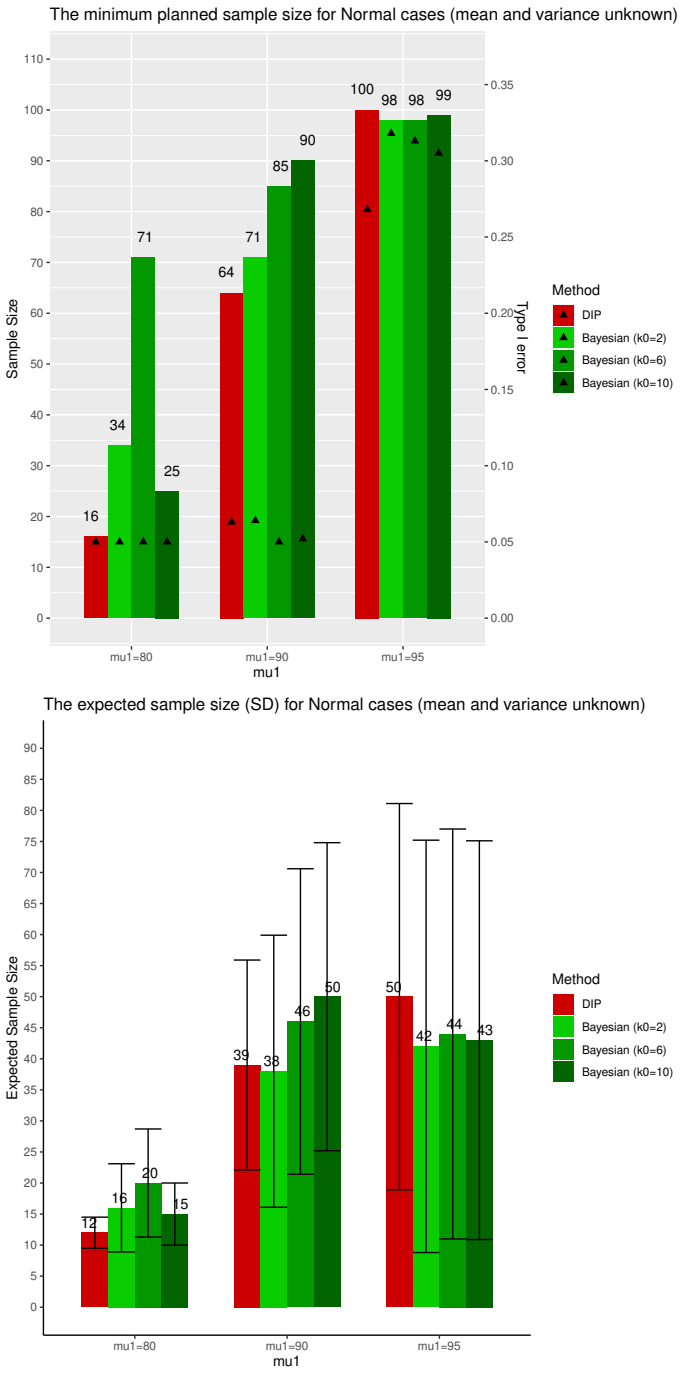

Fig. S.13. Simulation Results for Normal Distribution (Mean and Variance Unknown) - Underestimating Prior

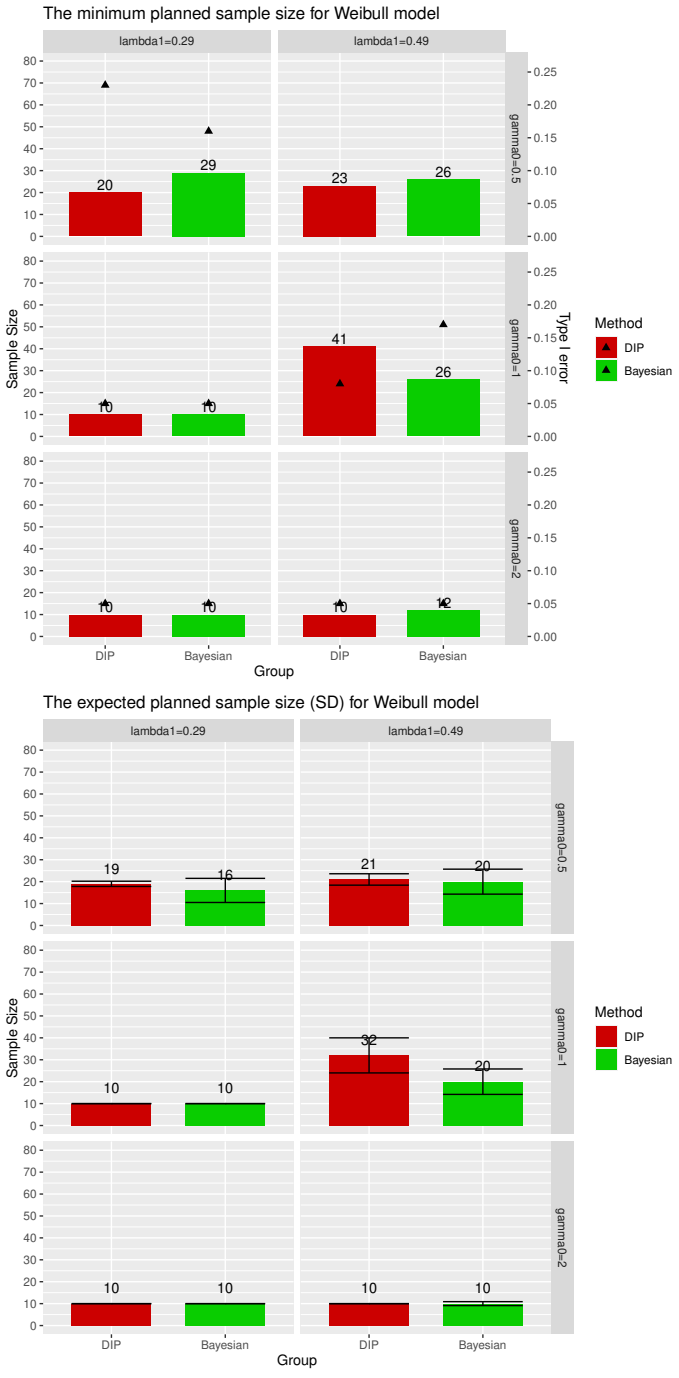

Fig. S.14. Simulation Results for Weibull Distribution

Appendix F

R CODE RELEVANT TO CHAPTER 5

One-sample Bernoulli distribution

```
# ------------------------------------------------------------------------------
#     prior --> list containing the information for prior
#          [[1]] - the prior distribution type:
#                1 - DIP
#                2 - Beta(a,b)
#          [[2]] - a: first parameter of the Beta distribution
#          [[3]] - b: second parameter of the Beta distribution
# ------------------------------------------------------------------------------

#' One sample Bernoulli model
#'
#' For a given planned sample size, the efficacy and futility boundaries,
#' return the power, the type I error, the expected sample size and its
#' standard deviation, the probability of reaching the efficacy and futility boundaries.
#'
#'
#' @param prior A list of length 3 containing the distributional information of the prior.
#' The first element is a number specifying the type of prior. Options are
#' \enumerate{
#' \item DIP ;
#' \item Beta(a,b), where a = shape, b = scale}
#' The second and third elements of the list are the parameters a and b, respectively.
#' @param N The planned sample size.
#' @param p0 The null response rate, which could be taken as the standard or historical rate.
#' @param p1 The response rate of the new treatment.
#' @param d The target improvement (minimal clinically meaningful difference).
#' @param ps The efficacy boundary (upper boundary).
#' @param pf The futility boundary (lower boundary).
#' @param alternative less (lower values imply greater efficacy) or greater (larger
#' values imply greater efficacy).
#' @param seed The seed for simulations.
#' @param sim The number of simulations.
#' @return A list of the arguments with method and computed elements
#' @examples
#' # with traditional Bayesian prior Beta(1,1)
#' OneSampleBernoulli(list(2,1,1), N = 100, p0 = 0.3, p1 = 0.5, d = 0.05,
#'                    ps = 0.98, pf = 0.05, alternative = "greater",
```

```
#'                     seed = 202210, sim = 10)
#' # with DIP
#' OneSampleBernoulli(list(1,0,0), N = 100, p0 = 0.3, p1 = 0.5, d = 0.05,
#'                     ps = 0.98, pf = 0.05, alternative = "greater",
#'                     seed = 202210, sim = 10)
#' @importFrom stats rbeta rbinom rgamma rnorm rpois
#' @export OneSampleBernoulli

OneSampleBernoulli <- function(prior, N = 100, p0, p1, d = 0,
                               ps = 0.95, pf = 0.05,
                               alternative = c("less", "greater"), seed = 202209, sim = 5000)
{
  alternative <- match.arg(alternative)
  # Define the inputs
  if(prior[[1]] == 1){
     prior[[2]] <- NA
     prior[[3]] <- NA}
  ## N limit
  if(!is.null(N) && (!is.numeric(N) || N <= 0 ))
    stop("N must be positive number and greater than 10")
  ## p0 limit
  if(!is.null(p0) && (!is.numeric(p0) || (p0 < 0 | p0 > 1)))
    stop("p0 must be numeric in [0,1]")
  ## p1 limit
  if(!is.null(p1) && (!is.numeric(p1) || (p1 < 0 | p1 > 1)))
    stop("p1 must be numeric in [0,1]")
  ## d limit
  if(!is.null(d) && (!is.numeric(d) || (d < 0 | d > abs(p1-p0))))
    stop("d must be numeric in [0, |p1-p0|]")
  ## efficacy boundary limit
  if(!is.null(ps) && (!is.numeric(ps) || (ps < 0.8 | ps > 1)))
    stop("ps (efficacy boundary) must be numeric in [0.8,1]")
  ## futility boundary limit
  if(!is.null(pf) && (!is.numeric(pf) || (pf < 0 | pf > 0.2)))
    stop("pf (futility boundary) must be numeric in [0,0.2]")
  ## set.seed
  if(!is.numeric(seed))
    stop("seed must be numeric")
  ## number of simulation
  if(!is.numeric(sim))
    stop("simulation number  must be numeric")

  set.seed(seed)

  # Functions to calculate the posterior
```

```
Bernoulli <- function(a,b,y){posterior<-rbeta(1000, a+sum(y), b+(length(y)-sum(y)))}
Bernoulli.DIP <- function(p0, y, N){
  j<-length(y)
  posterior<-rbeta(1000,1+sum(y)+p0*(N-j),1+(j-sum(y))+(1-p0)*(N-j))
}

# Simulated Data
# calculate power
n.enrolled <- NULL
cat1s <- 0
cat1f <- 0
for (k in 1:sim) {
  y.data<-NULL
  j<-0
  cat<-0
  cats<-0
  catf<-0
  pp_stop<-0.5
  while(cat == 0)
  {
    j<-j+1
    y.data<-append(y.data,rbinom(1,1,p1))
    if(j>=10)
    {
      if (prior[[1]] == 2){
          p1_s<-Bernoulli(a = prior[[2]], b = prior[[3]], y = y.data)
      }else if (prior[[1]] == 1){
          p1_s <- Bernoulli.DIP(p0, y = y.data, N = N)
      }

      if (alternative == "greater"){
        pp_stop<-sum(p1_s>p0+d)/length(p1_s)
      }else if (alternative == "less"){
        pp_stop<-sum(p1_s<p0-d)/length(p1_s)
      }

    }
    if(pp_stop>=ps){cats<-1}
    if(pp_stop<pf){catf<-1}
    cat<-cats+catf
    if(j==N){cat<-1}
  }
  if(cats==1){cat1s<-cat1s+1}
  if(cats==0){cat1s<-cat1s}
  if(catf==1){cat1f<-cat1f+1}
```

```
  if(catf==0){cat1f<-cat1f}

  # Recruited Sample Size
  n.enrolled <- append(n.enrolled, j)
}
ss <- round(mean(n.enrolled), digits = 1)
sd <- round(sd(n.enrolled), digits = 2)
fut.rate <- cat1f/sim
power <- cat1s/sim


# calculate type I error
cat1s <- 0
cat1f <- 0
for (k in 1:sim) {
  y.data<-NULL
  j<-0
  cat<-0
  cats<-0
  catf<-0
  pp_stop<-0.5
  while(cat == 0)
  {
    j<-j+1
    y.data<-append(y.data,rbinom(1,1,p0)) # under the null hypothesis p1 = p0
    if(j>=10)
    {
      if (prior[[1]] == 2){
        p1_s<-Bernoulli(a = prior[[2]], b = prior[[3]], y = y.data)
      }else if (prior[[1]] == 1){
        p1_s <- Bernoulli.DIP(p0, y = y.data, N = N)
      }

      if (alternative == "greater"){
        pp_stop<-sum(p1_s>p0+d)/length(p1_s)
      }else if (alternative == "less"){
        pp_stop<-sum(p1_s<p0-d)/length(p1_s)
      }

    }
    if(pp_stop>=ps){cats<-1}
    if(pp_stop<pf){catf<-1}
    cat<-cats+catf
    if(j==N){cat<-1}
  }
```

```
    if(cats==1){cat1s<-cat1s+1}
    if(cats==0){cat1s<-cat1s}
    if(catf==1){cat1f<-cat1f+1}
    if(catf==0){cat1f<-cat1f}
  }
  t1error <- cat1s/sim

  # Outputs
  if (prior[[1]] == 1) {method = "DIP"
  } else if (prior[[1]] == 2) {method=paste("Beta(",prior[[2]],",",prior[[3]], ")",sep="")}

  z <- list(method = method, power = power, type_I_error = t1error,
            expected_sample_size = ss, expected_sample_size_std = sd,
            the_prob_efficacy = power, the_prob_futility = fut.rate)
  z
}
```

One-sample Bernoulli distribution - Trial Design

```
# ------------------------------------------------------------------------------
#
#     prior --> list containing the information for prior
#          [[1]] - the prior distribution type:
#                1 - DIP
#                2 - Beta(a,b)
#          [[2]] - a: first parameter of the Beta distribution
#          [[3]] - b: second parameter of the Beta distribution
#
#
# ------------------------------------------------------------------------------

#' One sample Bernoulli model - Trial Design
#'
#' Calculate the minimum planned sample size under an admissible design.
#' The users decide the power and type-I-error, and pick the efficacy and futility boundaries.
#' If there are no admissible design based on controlled type-I-error, then default to output
#' the designs with the lowest type-I-error and at least the user-defined (e.g. 80\%) power.
#'
#'
#' @param prior A list of length 3 containing the distributional information of the prior.
#' The first element is a number specifying the type of prior. Options are
#' \enumerate{
#' \item DIP ;
#' \item Beta(a,b), where a = shape, b = scale}
```

```
#' The second and third elements of the list are the parameters a and b, respectively.
#' @param nmin The start searching sample size
#' @param nmax The stop searching sample size
#' @param p0 The null response rate, which could be taken as the standard or historical rate.
#' @param p1 The response rate of the new treatment.
#' @param d The target improvement (minimal clinically meaningful difference).
#' @param ps The efficacy boundary (upper boundary).
#' @param pf The futility boundary (lower boundary).
#' @param power The power to achieve.
#' @param t1error The controlled type-I-error.
#' @param alternative less (lower values imply greater efficacy) or greater (larger
#' values imply greater efficacy).
#' @param seed The seed for simulations.
#' @param sim The number of simulations.
#' @return A list of the arguments with method and computed elements.
#' @examples
#' \donttest{
#' # with traditional Bayesian prior Beta(1,1)
#' OneSampleBernoulli.Design(list(2,1,1), nmin = 10, nmax=100, p0 = 0.3, p1 = 0.5, d = 0,
#'                     ps = 0.98, pf = 0.02, power = 0.80, t1error=0.05,
#'                     alternative = "greater", seed = 202210, sim = 10)
#' # with DIP
#' OneSampleBernoulli.Design(list(1,0,0), nmin = 10, nmax=100, p0 = 0.3, p1 = 0.5, d = 0,
#'                     ps = 0.98, pf = 0.02, power = 0.80, t1error=0.05,
#'                     alternative = "greater", seed = 202210, sim = 10)
#' }
#' @importFrom stats rbeta rbinom rgamma rnorm rpois
#' @export OneSampleBernoulli.Design

OneSampleBernoulli.Design <- function(prior, nmin = 10, nmax = 100, p0, p1, d = 0,
                                ps, pf, power = 0.8, t1error = 0.05,
                                alternative = c("less", "greater"), seed = 202209, sim = 1000)
  {
  alternative <- match.arg(alternative)
  # Define the inputs
  if(prior[[1]] == 1){
    prior[[2]] <- NA
    prior[[3]] <- NA}
  ## nmin limit
  if(!is.null(nmin) && (!is.numeric(nmin) || nmin < 10 || nmin >= nmax))
    stop("nmin must be positive number and at least 10")
  ## nmax limit
  if(!is.null(nmax) && (!is.numeric(nmax) || nmax <= nmin || nmax >= 200))
    stop("nmax must greater than 'nmin' and less than 200")
  ## p0 limit
```

```
  if(!is.null(p0) && (!is.numeric(p0) || (p0 < 0 | p0 > 1)))
    stop("p0 must be numeric in [0,1]")
  ## p1 limit
  if(!is.null(p1) && (!is.numeric(p1) || (p1 < 0 | p1 > 1)))
    stop("p1 must be numeric in [0,1]")
  ## d limit
  if(!is.null(d) && (!is.numeric(d) || (d < 0 | d > abs(p1-p0))))
    stop("d must be numeric in [0, |p1-p0|]")
  ## efficacy boundary limit
  if(!is.null(ps) && (!is.numeric(ps) || (ps < 0.8 | ps > 1)))
    stop("ps (efficacy boundary) must be numeric in [0.8,1]")
  ## futility boundary limit
  if(!is.null(pf) && (!is.numeric(pf) || (pf < 0 | pf > 0.2)))
    stop("pf (futility boundary) must be numeric in [0,0.2]")
  ## power limit
  if(!is.null(power) && (!is.numeric(power) || (power < 0 | power > 1)))
    stop("power must be numeric in [0,1]")
  ## t1error limit
  if(!is.null(t1error) && (!is.numeric(t1error) || (t1error < 0 |t1error > 1)))
    stop("type-I-error must be numeric in [0,1]")
  ## set.seed
  if(!is.numeric(seed))
    stop("seed must be numeric")
  if(!is.numeric(sim))
    stop("simulation number must be numeric")

  set.seed(seed)

  # Functions to calculate the posterior
  Bernoulli <- function(a,b,y){posterior<-rbeta(1000, a+sum(y), b+(length(y)-sum(y)))}
  Bernoulli.DIP <- function(p0, y, N){
    j<-length(y)
    posterior<-rbeta(1000,1+sum(y)+p0*(N-j),1+(j-sum(y))+(1-p0)*(N-j))
  }

# Simulated Data
# calculate N that can achieve the power
N_v <- NULL
power_v <- NULL
n_v <- NULL
sd_v <- NULL
for (N in seq(from=nmin, to=nmax, by=1)){
    cat1s <- 0
    cat1f <- 0
    n.enrolled <- NULL
```

```
for (k in 1:sim) {
  y.data <- NULL
  j <- 0
  cat <- 0
  cats <- 0
  catf <- 0
  pp_stop <- 0.5
  while(cat == 0){
    j <- j+1
    y.data <- append(y.data, rbinom(1,1,p1))
    if(j>=10)
    {
      if (prior[[1]] == 2){
        p1_s<-Bernoulli(a = prior[[2]], b = prior[[3]], y = y.data)
      }else if (prior[[1]] == 1){
        p1_s <- Bernoulli.DIP(p0, y = y.data, N = N)
      }

      if (alternative == "greater"){
        pp_stop<-sum(p1_s>p0+d)/length(p1_s)
      }else if (alternative == "less"){
        pp_stop<-sum(p1_s<p0-d)/length(p1_s)
      }
    }
    if(pp_stop>=ps){cats<-1}
    if(pp_stop<pf){catf<-1}
    cat<-cats+catf
    if(j==N){cat<-1}
  }
  if(cats==1){cat1s<-cat1s+1}
  if(cats==0){cat1s<-cat1s}
  if(catf==1){cat1f<-cat1f+1}
  if(catf==0){cat1f<-cat1f}

  # Recruited Sample Size
  n.enrolled <- append(n.enrolled, j)
}
power.cal <- cat1s/sim

jitter <- 0.01
if (power.cal >= power-jitter){
N_v <- append(N_v, N)
power_v <- append(power_v, power.cal)
n_v <- append(n_v, round(mean(n.enrolled), 0))
sd_v <- append(sd_v, round(sd(n.enrolled), 1))
```

```
    }
    result1 <- cbind(N_v, power_v, n_v, sd_v)
} # End of power calculation

if (is.null(result1)){
message("Suggest: please adjust your input values!")
stop(paste("No sample size in the range [",nmin,",",nmax,"] can achieve ",power*100,"% power",
sep=""))
}

# calculate type I error
nmin1 <- N_v[which.min(N_v)] # start minimum sample size in calculation of exact type I error
N_v <- NULL
t1error_v <- NULL
for (N in seq(from=nmin1, to=nmax, by=1)){
  cat1s <- 0
  cat1f <- 0
  for (k in 1:sim) {
    y.data <- NULL
    j <- 0
    cat <- 0
    cats <- 0
    catf <- 0
    pp_stop <- 0.5
    while(cat == 0){
      j <- j+1
      y.data <- append(y.data, rbinom(1,1,p0)) # under the null hypothesis p1 = p0
      if(j>=10)
      {
        if (prior[[1]] == 2){
          p1_s<-Bernoulli(a = prior[[2]], b = prior[[3]], y = y.data)
        }else if (prior[[1]] == 1){
          p1_s <- Bernoulli.DIP(p0, y = y.data, N = N)
        }

        if (alternative == "greater"){
          pp_stop<-sum(p1_s>p0+d)/length(p1_s)
        }else if (alternative == "less"){
          pp_stop<-sum(p1_s<p0-d)/length(p1_s)
        }
      }
      if(pp_stop>=ps){cats<-1}
      if(pp_stop<pf){catf<-1}
      cat<-cats+catf
      if(j==N){cat<-1}
```

```
    }
    if(cats==1){cat1s<-cat1s+1}
    if(cats==0){cat1s<-cat1s}
    if(catf==1){cat1f<-cat1f+1}
    if(catf==0){cat1f<-cat1f}
  }
  t1error.cal <- cat1s/sim
  N_v <- append(N_v, N)
  t1error_v <- append(t1error_v, t1error.cal)
  result2 <- cbind(N_v, t1error_v)
} # End of Type-I-error calculation

# Outputs
if (!is.null(result1) & !is.null(result2)){
   result <- merge(result1, result2, by=c("N_v"))
   final <- as.data.frame(result)

   # select the lowest/best-controlled type I error
   final$diff <- abs(final$t1error_v - t1error)
   final <- final[order(final$diff, final$t1error_v, final$power_v, final$N_v), ]
   ff <- final[1,]
   planN <- ff$N_v
   exact.power <- ff$power_v
   exact.t1 <- ff$t1error_v
   ss <- ff$n_v
   sd <- ff$sd_v
   if (prior[[1]] == 1) {method = "DIP"
   } else if (prior[[1]] == 2) {method=paste("Beta(",prior[[2]], ",", prior[[3]], ")", sep="")
   }
   z <- list(method = method, planned_sample_size = planN,
             efficacy_boundary = ps, futility_boundary = pf,
             exact_power = exact.power, exact_type_I_error = exact.t1,
             expected_sample_size = ss, expected_sample_size_std = sd)
   z
 } # End of Outputs
}
```

REFERENCES

[1] Peter F. Thall and Richard Simon. "Practical Bayesian Guidelines for Phase IIB Clinical Trials". In: *Biometrics* 50.2 (1994), pp. 337–349. DOI: 10.2307/2533377.

[2] Peter F. Thall and Richard Simon. "A Bayesian Approach to Establishing Sample Size and Monitoring Criteria for Phase II Clinical Trials". In: *Controlled Clinical Trials* 15.6 (1994), pp. 463–481. ISSN: 0197-2456. DOI: 10.1016/0197-2456(94)90004-3.

[3] Roy T. Sabo. "Adaptive Allocation for Binary Outcomes Using Decreasingly Informative Priors". In: *Journal of Biopharmaceutical Statistics* 24.3 (2014), pp. 569–578. DOI: 10.1080/10543406.2014.888441.

[4] Scott M. Berry et al. *Bayesian Adaptive Methods for Clinical Trials*. CRC Press, 2011. ISBN: 9781439825518.

[5] U.S. Department of Health, Human Services Food, and Drug Administration. *Guidance for Clinical Trial Sponsors: Establishment and Operation of Clinical Trial Data Monitoring Committees*. 2006. URL: https://www.fda.gov/media/75398 (visited on 03/2006).

[6] U.S. Department of Health, Human Services Food, and Drug Administration. *Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials*. 2010.

[7] Guosheng Lin. *Clinical Trial Design: Bayesian and Frequentist Adaptive Methods*. John Wiley & Sons, Incorporated, 2012.

[8] J Jack Lee and Diane D Liu. "A predictive probability design for phase II cancer clinical trials". In: *Clinical Trials* 5.2 (2008), pp. 93–106. DOI: 10.1177/1740774508089279.

[9] Satoshi Morita, Peter F. Thall, and Peter Muller. "Determining the effective sample size of a parametric prior". In: *Biometrics* 64 (2008), pp. 595–602. DOI: 10.1111/j.1541-0420.2007.00888.x.

[10] Satoshi Morita, Peter F. Thall, and Peter Müller. "Prior Effective Sample Size in Conditionally Independent Hierarchical Models". In: *Bayesian Analysis* 7.3 (2012), pp. 591–614. DOI: 10.1214/12-BA720. URL: https://doi.org/10.1214/12-BA720.

[11] Beat Neuenschwander et al. "Predictively consistent prior effective sample sizes". In: *Biometrics* 76 (2020), pp. 578–587. DOI: 10.1111/biom.13252.

[12] Erin Donahue and Roy T. Sabo. "A natural lead-in approach to response-adaptive allocation for continuous outcomes". In: *Pharmaceutical Statistics* (2021).

[13] Peter F. Thall and J. Kyle Wathen. "Practical Bayesian adaptive randomisation in clinical trials". In: *European Journal of Cancer* 43.5 (2007), pp. 859–866. DOI: 10.1016/j.ejca.2007.01.006.

[14] Lanju Zhang and William F. Rosenberger. "Response-adaptive randomization for clinical trials with continuous outcomes". In: *Biometrics* 62 (2006), pp. 562–569. DOI: 10.1111/j.1541-0420.2005.00496.x.

[15] Uttam Bandyopadhyay and Rahul Bhattacharya. "On a Class of Optimal Type Covariate Adjusted Response Adaptive Allocations for Normal Treatment Responses". In: *Biometrics* 62 (2006), pp. 562–569. DOI: 10.1111/j.1541-0420.2005.00496.x.

[16] Roy T. Sabo and Ghalib Bello. "Optimal and lead-in adaptive allocation for binary outcomes: a comparison of Bayesian methodologies". In: *Communications in Statistics - Theory Methods* 46.6 (2017), pp. 2823–2836. DOI: 10.1080/03610926.2015.1053929.

[17] Alexey Stukalov Martyn Plummer and Matt Denwood. *RJAGS: Bayesian Graphical Models Using MCMC*. 2022. URL: https://CRAN.R-project.org/package=rjags.

[18] Beat Neuenschwander et al. "Rejoinder to "Predictively consistent prior effective sample sizes," by Beat Neuenschwander, Sebastian Weber, Heinz Schmidli, and Anghony O'Hagan". In: *Biometrics* 76 (2020), pp. 602–605. DOI: 10.1111/biom.13245.

[19] Atanu Biswas and Jean-Francois Angers. "Discussion on "Predictively consistent prior effective sample sizes," by Beat Neuenschwander, Sebastian Weber, Heinz Schmidli, and Anghony O'Hagan". In: *Biometrics* 76 (2020), pp. 591–594. DOI: 10.1111/biom.13247.

[20] Peter D. Hoff. *A First Course in Bayesian Statistical Methods*. Springer, 2009.

[21] Chen Wang et al. "Early termination in single-parameter model phase II clinical trial designs using decreasingly informative priors". In: *Int J Clin Trials* 9.2 (2022), pp. 107–117. DOI: 10.18203/2349-3259.ijct20221110.

[22] Peter D. Hoff. *A First Course in Bayesian Statistical Methods*. Springer, 2009. ISBN: 978-0-387-92299-7.

[23] Peter F. Thall, Richard M. Simon, and Elihu H. Estey. "Bayesian sequential monitoring designs for single-arm clinical trials with multiple outcomes". In: *Statistics in Medicine* 14.4 (1995), pp. 357–379. DOI: 10.1002/sim.4780140404.

[24] Kengo Nagashima. "ph2bayes: Bayesian Single-Arm Phase II Designs". In: (2018). URL: https://CRAN.R-project.org/package=ph2bayes.

[25] Yanlin Zhu and Rui Qin. "ph2bye: Phase II Clinical Trial Design Using Bayesian Methods". In: (2016). URL: https://CRAN.R-project.org/package=ph2bye.

[26] Gary Simmons et al. "Safety and Tolerability of Intra-Venous Ascorbic Acid in Allogeneic Hematopoietic Cell Transplant Recipients: A Matched Historical Control Study". In: *Blood* 136 (2020), pp. 29–30. DOI: 10.1182/blood-2020-138543.

VITA

Chen Wang received her Bachelor of Science Degree in Geographic Information System from Northwest University, Xi'an, China in 2010, and her Master of Science degree in Statistics from University of Delaware in 2013. Prior to beginning her time at Virginia Commonwealth University (VCU), Chen worked as a Biostatistician for Parexel, where she was responsible for statistical analysis on several Phase II/III clinical trials. During her time at VCU, she worked as a research assistant at Biostatistical Consulting Lab (BCL), where she consulted with investigators on study designs and data analysis.