Bayesian Approaches to Compare Dose Levels From Small N Sequential Multiple Assignment Randomized Trials (snSMART)

by

Fang Fang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in the University of Michigan
2023

Doctoral Committee:

Professor Thomas M. Braun, Co-Chair
Associate Professor Kelley M. Kidwell, Co-Chair
Assistant Professor Rahul Ladhania
Research Professor Matthew J. Schipper

Fang Fang
sqrfang@umich.edu
ORCID iD: 0000-0002-7089-3591

ACKNOWLEDGMENTS

I would like to express my appreciation to the following individuals for their invaluable support, guidance, and encouragement during the completion of this dissertation.

First and foremost, I am deeply grateful to my advisors, *Dr. Kelley Kidwell*, *Dr. Thomas Braun* and *Dr. Roy Tamura*, for their unwavering support, guidance, humor, and patience throughout this journey. Their expertise and insights have been invaluable to me, and I am truly grateful for the opportunity to work with such remarkable mentors. Moreover, I consider myself lucky to have them not only as my academic guides but also as my life mentors. Their advice has been critical in helping me navigate every intersection in graduate school. I am convinced that I would have ended up in a ditch without their support, particularly in the moments when life threw me curveballs. Thank you, *Dr. Kidwell*, *Dr. Braun*, and *Dr. Tamura*, I will always remember your influence in my life and career.

I would like to thank the incredible members of my dissertation committee, *Dr. Matthew Schipper* and *Dr. Rahul Ladhania*, for their valuable feedback, suggestions, and support. I feel fortunate to have had such a talented, thoughtful, and supportive group of scholars on my committee. Your contributions have been essential to the completion of this dissertation.

I would like to extend my heartfelt thanks to *Dr. Kirsten Herold*, whose expertise in academic writing, attention to detail, and constructive criticism have improved the clarity and coherence of my writing, saving me from my own clumsy writing skills.

I would like to extend my gratitude to all staff in the Department of Biostatistics for their help and support throughout my graduate studies.

I would also like to thank my talented lab mates and friends for their love and moral support throughout this process. Their encouragement has been a constant source of motivation. This journey has been challenging but also incredibly rewarding with you by my side.

I would like to express my deep gratitude to my parents, who have always believed in me and loved me unconditionally. Their love and support have transcended physical distance and remained a constant source of strength and motivation for me.

Finally, I would like to acknowledge the funds received from *National Center for Advancing Translational Sciences*, *Patient-Centered Outcomes Research Institute (PCORI)* and *FDA's Broad Agency Announcement (BAA)*, which have been instrumental in supporting my research.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

TABLE

ABSTRACT

Placebo-controlled randomized clinical trials (RCT) have been regarded as the gold standard for guiding the registry of new drugs. However, such trials in rare diseases are invariably small in sample size due to the small number of patients with disease and to enroll in studies with a placebo arm. A small n, sequential, multiple assignment, randomized trial (snSMART) is a design that include placebo and dose levels of a drug to facilitate efficient treatment effect estimates.

In the first project, we propose a new snSMART design that investigates the response rates of a drug tested at a low and high dose compared with placebo. Patients are randomized to an initial treatment (stage 1). In stage 2, patients are rerandomized, depending on their initial treatment and their response to that treatment in stage 1, to either the same or a different dose of treatment, with no placebo option. We present a Bayesian approach where information is borrowed between stage 1 and stage 2 to determine the efficacy of the active treatment. We compare our approach to standard methods using only stage 1 data and a log-linear Poisson model that uses data from both stages where parameters are estimated using generalized estimating equations. We observe that our method has smaller root-mean-square-error (rMSE) and 95% credible interval widths than standard methods in the tested scenarios.

In the second project, we extend the previous snSMART design with a binary endpoint to a design that considers a continuous outcome. Data from both stages are used to determine the marginal efficacy of the dose levels of the active treatment via a Bayesian model. We also compare the approach with the standard model based solely upon the data from stage 1 and evaluate different prior distributions in model fitting. We demonstrate that the joint stage Bayesian estimators have smaller rMSE and narrower credible intervals.

In the third project, we aim to determine the sample size needed to achieve a prespecified significance level and desired statistical power for an snSMART. We focus on the design and Bayesian analytical approach proposed in the second project to detect the difference in the mean outcome between low dose and placebo. We adjust a one stage sample size calculation to account for the extra information extracted from the two-stage design. We also provide an efficient approach to calculate sample size in one step. Through simulations, we demonstrate that the required sample sizes calculated using the two SSD methods both provide the desired power. Both approaches are available in an RShiny App to disseminate the methods to rare disease investigations.

CHAPTER 1

Introduction

A rare disease is defined as a disease that affects fewer than 200,000 people in the United States (107th Congress, 2002). Taken together, there are more than 7,000 rare diseases that affect over 30 million people in the United States (NIH, 2023).

Unfortunately, many of these diseases have been neglected for a long time due to limited resources and attention (Schieppati et al., 2008). The Orphan Drug Act was passed in 1983 in an effort to promote the development of treatment of rare diseases which raised awareness of this public-health issue (Herder, 2017). Despite the efforts in the past few decades, more work is needed. It is recorded that only 552 (7%) of rare diseases have an approved drug on the market having an orphan designation, leaving 93% of rare diseases without an approved treatment, causing a considerable unmet need for many patients (NORD, 2021).

The limited number of individuals affected by these diseases presents a challenge for the development of treatments. Approval of any drug is based on the same requirements for evidence of effectiveness, regardless of the size of the diseased population. While randomized clinical trials (RCTs) are utilized to demonstrate the strongest scientific evidence of an effective treatment, these trials are often difficult or impossible in rare disease settings because they require a large number of subjects. As a result, RCTs involving rare diseases often have reduced power when compared to studies of diseases that are not rare (Gupta et al., 2011). To combat these issues, Tamura et al. (2016) previously proposed a small n, sequential, multiple assignment, randomized trial (snSMART) design to investigate three active treatments for a rare disease. Here, we propose variations of the snSMART design that focuses on a single drug and placebo.

An snSMART is a variation of a SMART design (Lavori and Dawson, 2000; Murphy, 2005) that is specifically intended for small samples. In a SMART, patients are randomized to at least two sequential interventions in such a way that the second intervention assignment depends on the patient's response to the first intervention. The goal of a classic SMART is often to develop effective dynamic treatment regimens (DTRs) that specify an initial treatment for a patient followed by subsequent treatment, that is tailored by the response to the initial treatment (Robins, 1986;

Murphy, 2003). In contrast, rather than to identify sequences of treatments tailored to an individual, the stages in an snSMART are used to garner more information from a smaller set of subjects. In other words, snSMARTs are not designed with the goal of developing or estimating the effects of DTRs. Instead, the goal of an snSMART is to efficiently use data across the two stages of the trial to find a single superior treatment or dose of treatment in a small sample of individuals.

Indeed, there have been previous examples of repurposing well-developed trial designs to address novel goals. For example, randomized discontinuation trials have been studied as an alternative phase II design in oncology (Rosner et al., 2002). In addition, randomized discontinuation trials have been modified using SMART designs in order to answer a wider variety of clinical questions (Almirall et al., 2012). Researchers have also considered some enhanced crossover designs in the rare diseases spectrum to address the concerns about the unnecessary exposure to placebo or treatment of high toxicity. For example, Makubate and Senn (2010) and Nason and Follmann (2010) both discussed designs that allow for discontinuation from the study according to the absorbed binary endpoints after a subject receives the first treatment. Honkanen et al. (2001) introduced an alternative design that consists of an initial randomized placebo-controlled stage, a randomized withdrawal stage for subjects who respond, and a third randomized stage for placebo non-responders who subsequently respond to treatment.

In Chapter 2, we introduce an snSMART design that compares varying doses of a study treatment to placebo with a focus on the first stage treatment effect. In this snSMART design, patients are initially randomized to one of three treatment groups: low dose, high dose, or placebo. They are then followed for a pre-specified period. At the end of the stage 1, a binary endpoint is measured, based on which the patients are re-randomized to either receive low dose or high dose. The binary endpoint is measured again after stage 2. We propose a Bayesian model to jointly model the data from both stages and estimate the treatment effect of each dose level at stage 1. We perform simulation studies to evaluate the accuracy and efficiency of our approach and compare it with other alternative methods. This project has been published in *Statistics in Medicine* (Fang et al., 2021).

In Chapter 3, we extend the application of the snSMART design with a binary outcome to a design with a continuous outcome. A binary response that is different from the continuous endpoint is used to determine the re-randomization in the second stage. The trial design is similar to the one in Chapter 2. We propose a joint stage Bayesian model to determine the efficacy of the low dose and high dose of the treatment, and compare the proposed approach with standard models based on data from stage 1. We also evaluate the impact of different prior settings on Bayesian estimates. This work has been published in *Statistics in Biopharmaceutical Research* (Fang et al., 2022).

In Chapter 4, we present Bayesian sample size determination (SSD) methods based on the snS-

MART design proposed in Chapter 3. We propose two SSD methods for the proposed snSMARTs with a continuous outcome to detect the desired difference in the mean outcome of the low dose and placebo. Both methods adopt the average coverage criterion (ACC) approach. In one approach, we take advantage of the explicit posterior variance of the treatment effect from the joint model. The other method updates the sample size needed for a single-stage parallel design with a proposed adjustment factor (AF). Simulations evaluate the power of detecting the desired difference given the required sample sizes calculated using the two SSD methods. The sample size methods have been tranlated vis an RShiny app to disseminate the methods.

CHAPTER 2

Bayesian Methods to Compare Response Rates of Dose Levels with Placebo in a Small N, Sequential, Multiple Assignment, Randomized Trial

2.1 Introduction

Designing clinical trials is crucial for evaluating the efficacy and safety of novel drugs. For registry clinical trials, in many situations, there is only a single, novel drug of interest and the objective is to determine efficacy of that drug. As an example, the Vasculitis Clinical Research Consortium was recently interested in testing a novel drug for patients suffering from granulomatosis with polyangitis (GPA) or microscopic polyangitis (MPA), forms of vasculitis characterized by inflammation of the blood vessels. The binary endpoint of the study was remission after three weeks of therapy. It was assumed, however, that an effective drug would have to be taken for longer than three weeks in practice. The trial needed to be placebo controlled and the investigators were interested in novel designs that could potentially increase the power of detecting a drug effect. Given that vasculitis is a rare disease, it was also necessary that the trial design was appropriate for small sample sizes.

In the snSMART design of Tamura et al. (2016), three unique, active experimental treatments were compared. We propose extending this design to a three-arm trial comparing placebo to low and high doses of one experimental treatment (Figure 2.1). In such a trial, patients are initially randomized at stage 1 to either receive placebo, low dose, or high dose with equal likelihood. Patients receive this treatment for a pre-specified amount of time, at which time their binary response status is ascertained. In stage 2, patients are re-randomized to either the same or a different dose of treatment depending on their initial treatment and their response to that treatment. Specifically, patients who received placebo at stage 1 are re-randomized to receive either low dose or high dose at stage 2, regardless of their stage 1 response. This is advantageous for patients because it means that everyone enrolled in the trial will receive an active treatment by stage 2, even if they were initially randomized to placebo. All patients who initially received low dose, regardless of their

response status, are re-randomized to either low dose or high dose. In the case of patients who responded to low dose, this re-randomization is appropriate because it allows patients to either receive a higher dose of the drug that is already effective for them or to continue receiving low dose. Receiving low dose again is advantageous for patients because they continue receiving a drug they respond to and advantageous for the trial because we gain more information about the response rate to low dose in stage 2 of the study. All patients who did not respond to high dose receive high dose again in stage 2, whereas patients who initially responded to high dose are re-randomized to receive either high dose again or low dose. In this design, the primary objective is to estimate the difference in the probability of response to treatment between low dose and placebo and between high dose and placebo.

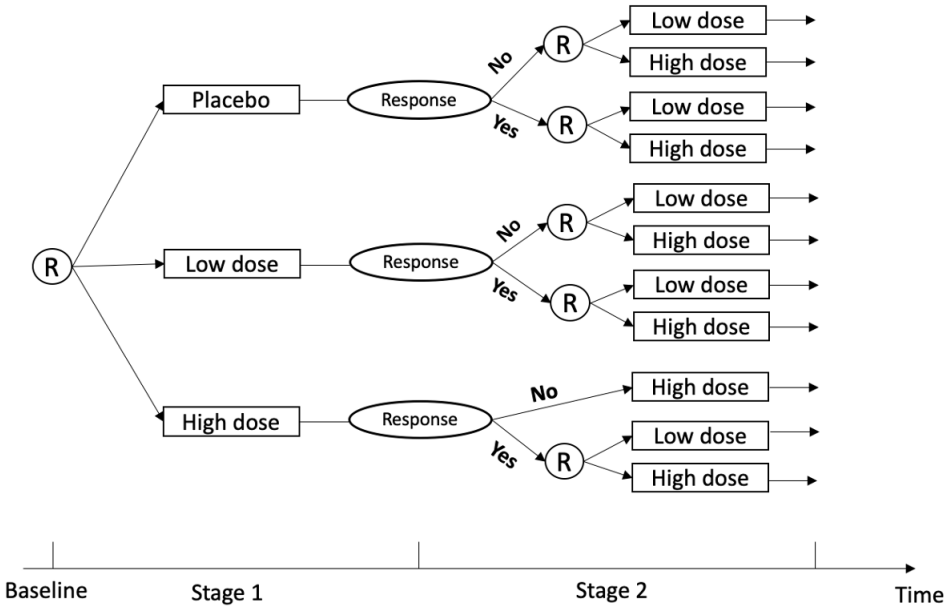

Figure 2.1: Study design of the proposed snSMART with a binary endpoint. Participants are randomized (R) to one of the first stage treatment arms, placebo, low dose or high dose equally (1:1:1). At the end of stage 1, patients are re-randomized to their second stage treatment based on their response status. Outcomes are collected at the end of stage 1 and stage 2.

Compared to other rare diseases and clinical trial designs, this snSMART design is advantageous for three reasons. First, this design allows for the comparison of treatment against placebo, which is necessary to demonstrate efficacy of an experimental treatment. Second, this design allows for the comparison of more than one dosage level of a drug, so that a lower, less toxic dose may be shown to be efficacious as opposed to investigating only a high dose. Third, individuals who respond to treatment in stage 1 may continue their original dose or may increase or decrease

dose. All participants receive an active treatment at some point in the trial. In addition, those who receive a low dose or high dose of the drug will continue to receive the drug at some level in both stages because there are no participants randomized to placebo in stage 2. These factors may improve patient engagement and recruitment, which is a challenge in the study of rare diseases. Wei et al. (2018) demonstrated efficiency gains of the previous snSMART design compared to a one stage design, but such advantages have not yet been confirmed for this setting.

In Section 2.2, we propose Bayesian and frequentist methods to analyze data for the primary efficacy analysis of the proposed snSMART design by borrowing information across patients and between trial stages. In the Bayesian model, we incorporate expert opinion and experience by using mildly informative prior distributions that are more flexible than those considered in Wei et al. (2018). In Section 2.3, we assess the influence of the prior distributions through simulation. We compare the Bayesian model to a frequentist model that also jointly models the response rates across the two stages of the snSMART. Both models are compared to models using only stage 1 data to illustrate the potential efficiency gain of the two-stage design. In Section 2.4, we complete this chapter with a discussion.

2.2 Methods

2.2.1 Bayesian Joint Stage Model

For each subject $i = 1, \ldots, N$, stage of the snSMART j = 1, 2 and treatment k = P, L, H, where N denotes the sample size, P denotes placebo, L denotes low dose, and H denotes high dose, let Y_{ijk} be the observed binary response outcome where 1 corresponds to "response" and 0 corresponds to "no response" to treatment. The stage 1 outcome and the stage 2 outcome given the stage 1 outcome are each modeled as Bernoulli random variables. The stage 1 response rate for treatment k is denoted as π_k. The stage 2 response rate for stage 1 responders to treatment k who receive treatment k′ in stage 2 is equal to $\beta_{1k}\pi_{k'}$. For non-responders to treatment k in stage 1 who receive treatment k* in stage 2, the stage 2 response rate is equal to $\beta_{0k}\pi_{k^*}$. Thus we have six unique linkage parameters that link stage 1 response to stage 2 response. Our proposed Bayesian joint stage model (BJSM) is as follows:

$$Y_{i1k}|\pi_k \sim Bernoulli(\pi_k) \tag{2.1}$$

$$Y_{i2k'}|\pi_k, \beta_{1k}, Y_{i1k} = 1 \sim Bernoulli(\beta_{1k}\pi_{k'}) \tag{2.2}$$

$$Y_{i2k^*}|\pi_k, \beta_{0k}, Y_{i1k} = 0 \sim Bernoulli(\beta_{0k}\pi_{k^*}) \tag{2.3}$$

Assumptions and prior distributions for the parameters are based on clinician input. Here, we

incorporate prior knowledge about disease and current treatments and assume that an ineffective treatment has a response rate of 15% and thus use an informative prior Beta(3, 17) for π_P. It is a setting similar to the GPA/MPA example mentioned in the Introduction. For the effect of low and high dose, we allow for a weak tendency for the drug response rates to be greater than the effect of placebo and assume that the logarithm of treatment effect ratio follows a Gaussian prior distribution N(μ, σ^2), i.e., $\log(\pi_L/\pi_P) \sim N(0.2, 100)$ and $\log(\pi_H/\pi_P) \sim N(0.2, 100)$. Note that $E(\pi_L/\pi_P) = e^{0.2} \approx 1.2$ under the proposed prior setting.

Wei et al. (2018) assumed that the linkage parameters (i) did not depend on the initial treatment and that, (ii) $\beta_0 \leq 1$ and (iii) $\beta_1 > 1$. Here, since both responders and non-responders are re-randomized and we are testing for a potential dose-response relationship between treatment arms, these previous assumptions are not appropriate. In our simulations, instead of assuming the Beta and Pareto priors used by Wei et al. (2018), we consider Gamma priors so that the linkage parameters can span the positive real line.

2.2.2 Log-linear Poisson Joint Stage Model

The log-linear Poisson joint stage model (LPJSM) presented in Wei et al. (2018) is slightly adjusted for our design. The LPJSM jointly models the stage 1 and stage 2 outcomes with a log link for interpretability. The LPJSM is shown below where there is a one-to-one correspondence to the parameters in the Bayesian model in Equations (2.1) - (2.3). Let Y_{ij} be the response of subject i in stage j ($j = 1, 2$), where $I(k_{ij} = k)$ is the indicator function for treatment k = P, L, H for subject i in stage j, then the LPJSM is as follows:

$$
\begin{aligned}
log(P(Y_{i1})) =& \alpha_1 I(k_{i1} = P) + \alpha_2 I(k_{i1} = L) + \alpha_3 I(k_{i1} = H) \\
log(P(Y_{i2})) =& \alpha_1 I(k_{i2} = P) + \alpha_2 I(k_{i2} = L) + \alpha_3 I(k_{i2} = H) + \alpha_4 I(k_{i1} = P, Y_{i1} = 0) + \\
& \alpha_5 I(k_{i1} = P, Y_{i1} = 1) + \alpha_6 I(k_{i1} = L, Y_{i1} = 0) + \alpha_7 I(k_{i1} = L, Y_{i1} = 1) + \\
& \alpha_8 I(k_{i1} = H, Y_{i1} = 0) + \alpha_9 I(k_{i1} = H, Y_{i1} = 1)
\end{aligned}
$$

Here we have nine estimated coefficients where α_1, α_2 and α_3 represent the log response rates of placebo, low and high dose. Coefficients α_4 - α_9 correspond to the six linkage parameters in the Bayesian model. The Poisson family is used to model the variance of the outcome to overcome convergence problems with log-binomial models in small samples (Williamson et al., 2013). The parameters are estimated via generalized estimating equations assuming an independent correlation structure. The variance of the LPJSM is corrected through robust sandwich estimators.

2.3 Simulations

In our simulations, we first assume that our drug of interest is ineffective and consider trials in the null scenario, i.e., the response rate of placebo, low dose and high dose are all equal (Scenario 1, P = L = H). Under the assumption that the drug of interest is effective, we consider three additional scenarios. In scenario 2, a dose-response relationship occurs, i.e. higher dose relates to higher treatment effect (response rates such that P $<$ L $<$ H). In scenario 3, no dose response occurs between low and high dose, so that the response rate of P $<$ L = H. Lastly, we consider an unlikely, but possible setting in scenario 4 where no dose response occurs and low dose is effective but high dose is not, so that the response rate of P = H $<$ L. We selected the Gamma(2, 2) prior for all linkage parameters, understanding this allows for positive probability for $\beta_{1k}\pi_{k'}$ and $\beta_{0k}\pi_{k*}$ to be greater than 1. We chose Gamma(2,2) as the prior for all $\beta_{1k}\pi_{k'}$ and $\beta_{0k}\pi_{k*}$ for three reasons: (i) simplicity, (ii) the distribution ranges from 0 to 3 for most of the random draws, which serves as a restriction to the prior distributions of the linkage parameters, and (iii) the distribution is centered at 1 with variance equal to 0.5, which allows for flexibility of the prior distribution of the linkage parameters to be below or above 1. This third property allows stage 1 responders the possibility to worsen in the second stage if they decrease dose and stage 1 non-responders the possibility to respond if they increase dose. See Table 1 for the scenarios and priors we used in simulations.

In the data generating process, we simulated 2,000 realizations per scenario under the four settings in Table 2.1. For each realization, $N/3$ subjects were assigned to each treatment arm in stage 1, with a total sample size N. Responses to stage 1 were computed as random Bernoulli variables with the proposed response rates under different scenarios (Table 2.1, column 3-6). Subjects were then re-randomized equally to their stage 2 treatment based on their stage 1 treatment and stage 1 response. Stage 2 responses were computed using formulae (2) and (3) under the different scenarios. We compared bias, root mean-square error (rMSE), coverage rates and widths of the 95% credible/confidence intervals (CIs) between the proposed BJSM, LPJSM, a Bayesian method using only the first stage data (BFSM), and a maximum likelihood method (FSMLE) using only the first stage data.

The 95% CI for BJSM and BFSM are the narrowest intervals that include 95% of the posterior distribution of π_k, while the 95% CI for LPJSM and FSMLE are the asymptotic, normal-approximation 95% confidence intervals. The R package rjags was used to generate the posterior distributions of π_k, β_{1k}, and β_{0k}, and the R package gee was used to estimate the parameters defined in LPJSM.

BJSM Prior for All Scenarios	Response Rates/ Linkage Parameters	Scenarios			
		P=L=H	P<L<H	P< L=H	P=H<L
Beta(3,17)	π_P	0.15	0.15	0.15	0.15
$\log(\pi_L/\pi_P) \sim N(0.2, 100)$	π_L	0.15	0.25	0.4	0.4
$\log(\pi_H/\pi_P) \sim N(0.2, 100)$	π_H	0.15	0.35	0.4	0.15
Gamma(2,2)	β_{0P}	0.9	0.9	0.9	0.9
Gamma(2,2)	β_{1P}	1.3	1.3	1.3	1.3
Gamma(2,2)	β_{0L}	0.8	0.8	0.8	0.8
Gamma(2,2)	β_{1L}	1.2	1.2	1.2	1.2
Gamma(2,2)	β_{0H}	0.7	0.7	0.7	0.7
Gamma(2,2)	β_{1H}	1.1	1.1	1.1	1.1

Table 2.1: Scenarios and priors for the simulation settings. π_k is the first stage response rate for treatment k, $k = P, L, H$, where P=placebo, L=low dose, H=high dose. β_{1k} is the linkage parameter for first stage responders who receive treatment k in stage 1. β_{0k} is the linkage parameter for first stage non-responders who receive treatment k in stage 1. Simulations are done under 4 scenarios: (i) P=L=H (low and high dose are both as effective as placebo), (ii) P<L<H (placebo is less effective than low dose, and low dose is less effective than high dose), (iii) P<L=H (low dose and high dose are equally effective, and they are more effective than placebo), and (iv) P=H<L (low dose is effective but high dose is not). BJSM prior setting (column 1) is where we use Gamma(2, 2) for all linkage parameters to relax the restriction of priors.

2.3.1 Results

In this section, we present simulation results for the snSMART design in Figure 2.1 with sample sizes of N=90. Results for N=300 and N=45 can be found in the Tables 2.4-2.5.

For all scenarios, Table 2.2 gives the bias and rMSE for estimators of the stage 1 response rates for placebo, low dose, and high dose. In the null scenario (scenario 1, P = L = H), we note that BJSM, BFSM, and FSMLE provide estimators of the difference in response rates and of individual response rates with small bias. While the estimators for the difference in response rates in LPJSM is comparable to the other methods, we see that the bias in the point estimates of π_P, π_L, and π_H is much larger than the other methods. This is likely because there are few patients that respond to treatment in the trial. We also note that BJSM estimators have the smallest rMSEs out of all methods.

For scenario 2 (P $<$ L $<$ H), there is, on average, low to no bias for the response rate estimators for each dose level. Looking specifically at the estimation of the placebo response rate, we see that there is no bias, on average, for BJSM, BFSM, and FSMLE. In the LPJSM method, the smallest bias is in the estimator of π_H. This is likely due to the large number of participants in the trial that receive high dose. The estimators of π_P and π_L likely have higher bias in the LPJSM because there are fewer patients that receive placebo and low dose in the trial. The estimator of π_H, however, has

the largest bias in the Bayesian methods (BJSM and BFSM). The bias of the high dose response rate estimate is likely large because the true value of π_H in scenario 2 is 0.35, which is relatively far from the prior mean (0.183) for stage 1 response rates. In comparison, in BJSM and BFSM the estimator of π_L has less bias than that of π_H, presumably because the true value of π_L is 0.25, which is closer to 0.183. Looking at rMSE, we observe that the BJSM method estimators of π_P, π_L, and π_H have the lowest rMSE out of the estimators we compared. While the FSMLE approach has very low bias, it tends to have the largest rMSE out of the compared methods because it only models first-stage outcomes. When we consider the estimators for the difference between placebo and low and high dose response rates, we see that all methods provide estimators with small bias. The estimators for the difference in response rates for placebo vs. low and high dose of the BJSM have the smallest rMSEs out of all methods.

In scenario 3 (P $<$ L = H), we see small bias for the response rate estimators, specifically for the LPJSM and FSMLE methods. Looking at the BJSM and BFSM results, we see that the bias is negligible for the estimators of π_P, but slightly larger for estimators of π_L and π_H. In contrast, the LPJSM estimators of π_L and π_H have negligible bias, but the estimator of π_P is slightly higher. As in scenario 2, the bias in the Bayesian methods is due to the difference between the true value of the parameters (0.40) and the prior mean (0.183). We expect to see larger bias in the estimation of π_P from the LPJSM because few patients are randomized to placebo. The BJSM provides response rate estimators for placebo, low dose, and high dose with the smallest rMSEs out of all methods. Again, we see that BFSM and FSMLE have larger rMSE than the joint stage modeling procedures. The results of scenario 3 for the bias and rMSE of the estimators of response rate differences are similar to that of scenario 2.

In scenario 4 (P = H $<$ L), we once again see similar patterns in bias and rMSE of response rate estimators to scenarios 2 and 3. Again, estimators for the difference in response rate estimators generally have small bias, and the BJSM estimators have the smallest rMSE out of all methods.

It should be noted that across all four scenarios, the efficiency gain observed using joint stage modeling approaches, compared to BFSM and FSMLE, is not large for the estimators of π_P. We see little efficiency gain using joint stage approaches because no one is randomized to placebo treatment in stage 2 of the design. As such, first stage methods are comparable to joint stage methods in estimating π_P. We do, however, see modest efficiency gains using joint stage modeling approaches for the estimators of π_L - π_P and π_H - π_P. Since the estimation of π_L - π_P and π_H - π_P is typically of greater interest than the estimation of π_P, the efficiency gains we observe represent an advantage of using BJSM procedures.

Across all scenarios, we see that the bias of LPJSM estimators of π_P is large compared to the other LPJSM response rate estimators, and compared to the estimators of π_P for other methods. This increased bias likely stems from the low number of patients receiving placebo. Since there are

few people in the placebo treatment arm, and none in stage 2 of the study, there is less information to estimate π_P, leading to more bias. In larger samples (see Table 2.4), we see negligible bias for the LPJSM estimator of π_P, which supports our explanation that the bias observed in Table 2.2 is due to a low sample size.

		BJSM		LPJSM		BFSM		FSMLE	
Scenario		Bias	rMSE	Bias	rMSE	Bias	rMSE	Bias	rMSE
(1) P = L = H	π_P	-0.001	0.039	0.018	0.063	-0.001	0.039	-0.001	0.065
	π_L	-0.003	0.048	0.022	0.062	-0.005	0.062	0.000	0.064
	π_H	-0.007	0.043	0.020	0.062	-0.005	0.062	0.000	0.064
	π_L - π_P	-0.003	0.062	0.004	0.081	-0.005	0.073	0.001	0.091
	π_H - π_P	-0.006	0.058	0.002	0.087	-0.005	0.073	0.000	0.091
(2) P $<$ L $<$ H	π_P	0.000	0.039	0.009	0.063	0.000	0.039	0.000	0.065
	π_L	-0.005	0.057	0.014	0.068	-0.011	0.078	-0.003	0.080
	π_H	-0.013	0.064	-0.001	0.075	-0.013	0.084	-0.001	0.860
	π_L - π_P	-0.005	0.070	0.004	0.092	-0.011	0.087	-0.003	0.102
	π_H - π_P	-0.013	0.074	0.011	0.101	-0.012	0.093	-0.001	0.108
(3) P $<$ L = H	π_P	0.000	0.040	0.007	0.064	0.000	0.040	-0.001	0.067
	π_L	-0.009	0.066	0.000	0.076	-0.013	0.087	0.000	0.089
	π_H	-0.012	0.065	0.000	0.074	-0.015	0.087	-0.003	0.089
	π_L - π_P	-0.009	0.077	-0.006	0.097	-0.013	0.096	0.001	0.112
	π_H - π_P	-0.012	0.076	-0.006	0.099	-0.015	0.096	-0.002	0.111
(4) P = H $<$ L	π_P	-0.001	0.039	0.014	0.063	-0.001	0.039	-0.001	0.065
	π_L	-0.011	0.068	-0.023	0.081	-0.011	0.086	0.002	0.088
	π_H	-0.003	0.044	0.030	0.060	-0.004	0.063	0.001	0.065
	π_L - π_P	-0.010	0.078	-0.037	0.104	-0.010	0.094	0.002	0.109
	π_H - π_P	-0.003	0.059	0.016	0.083	-0.003	0.075	0.002	0.093

Table 2.2: Simulated bias and root-mean-square error (rMSE) for the estimators of π_k. π_k is the stage 1 response rate for treatment k, $k = P, L, H$, where P = placebo, L = low dose, and H = high dose. Scenarios are given in Table 1. Four modeling approaches: Bayesian joint stage modeling (BJSM), log-linear Poisson joint stage modeling (LPJSM), Bayesian first stage modeling (BFSM), and first stage maximum likelihood estimation (FSMLE) are compared. The sample size per treatment arm is 30.

Table 2.3 presents the 95% CI width and coverage rates (CR). Here, we see that the BJSM methods has smaller average 95% CI width than the LPJSM, BFSM, and FSMLE methods. In addition, the CR is around the target 95% for the BJSM in all tested scenarios.

When a sample size of N = 300 is used, we see similar results (see Table 2.4). Overall, we observe smaller bias across all settings when N = 100 in each arm. Interestingly, there is still

Scenario		BJSM		LPJSM		BFSM		FSMLE	
		CR	Width	CR	Width	CR	Width	CR	Width
(1) P = L = H	π_P	0.98	0.187	0.97	0.261	0.99	0.187	0.94	0.245
	π_L	0.93	0.183	0.95	0.229	0.87	0.221	0.95	0.246
	π_H	0.93	0.171	0.95	0.228	0.87	0.221	0.95	0.246
(2) P $<$ L $<$ H	π_P	0.98	0.187	0.96	0.254	0.99	0.187	0.94	0.246
	π_L	0.94	0.225	0.94	0.263	0.88	0.280	0.94	0.304
	π_H	0.94	0.256	0.94	0.298	0.92	0.317	0.91	0.335
(3) P $<$ L = H	π_P	0.98	0.186	0.96	0.252	0.98	0.187	0.94	0.245
	π_L	0.95	0.267	0.94	0.296	0.94	0.327	0.93	0.346
	π_H	0.94	0.261	0.95	0.296	0.94	0.327	0.94	0.344
(4) P = H $<$ L	π_P	0.99	0.187	0.98	0.258	0.99	0.187	0.94	0.245
	π_L	0.96	0.282	0.92	0.301	0.94	0.328	0.94	0.336
	π_H	0.93	0.169	0.96	0.213	0.88	0.222	0.94	0.247

Table 2.3: Simulated width and 95% coverage rate (CR) for the estimators of π_k. π_k is the true stage 1 response rate for the treatment k, $k = P, L, H$, where P = placebo, L = low dose, and H = high dose. Scenarios are given in Table 1. Four modeling approaches: Bayesian joint stage modeling (BJSM), log-linear Poisson joint stage modeling (LPJSM), Bayesian first stage modeling (BFSM), and first stage maximum likelihood estimation (FSMLE) are compared. The sample size per treatment arm is 30.

an efficiency gain when using BJSM methods in larger sample sizes, as the BJSM response rate estimators have smaller rMSEs than the response rate estimators from the LPJSM approach. In addition, for small samples, N = 45, under the null setting where we assume a spontaneous response rate of 30% or 40% for placebo, low and high doses of the experimental therapy, we again observe efficiency gains when using BJSM methods (see Table 2.5). The BJSM response rate estimators also have smaller bias than the LPJSM and BFSM methods.

2.3.2 Sensitivity to Priors

In addition to the prior setting presented in Section 2.3.1, we also explored other prior settings to evaluate the robustness of the BJSM method. First, we adjusted the mean of the prior distribution for $\log(\pi_L/\pi_P)$ and $\log(\pi_H/\pi_P)$. While we settled on a mean of 0.2 to be conservative, we also tested mean values of 0.3, 0.4, and 0.5, and found that our results were largely unchanged in the null and dose-response scenarios (see Table 2.6-2.7). Second, we adjusted the center of the prior distribution for π_P. In our presented results, the mean of the prior distribution for π_P was equal

Scenario		BJSM		LPJSM		BFSM		FSMLE	
		CR	Width	CR	Width	CR	Width	CR	Width
(1) P = L = H	π_P	0.96	0.124	0.94	0.139	0.96	0.124	0.93	0.138
	π_L	0.94	0.112	0.95	0.121	0.93	0.135	0.93	0.139
	π_H	0.95	0.108	0.95	0.121	0.94	0.134	0.94	0.138
(2) P $<$ L $<$ H	π_P	0.95	0.125	0.94	0.138	0.95	0.125	0.93	0.139
	π_L	0.95	0.133	0.95	0.144	0.95	0.165	0.95	0.170
	π_H	0.95	0.156	0.93	0.164	0.95	0.182	0.94	0.186
(3) P $<$ L = H	π_P	0.96	0.124	0.94	0.138	0.96	0.124	0.94	0.139
	π_L	0.95	0.157	0.94	0.163	0.94	0.188	0.94	0.191
	π_H	0.95	0.156	0.94	0.163	0.93	0.188	0.94	0.191
(4) P = H $<$ L	π_P	0.96	0.125	0.94	0.139	0.96	0.125	0.93	0.139
	π_L	0.95	0.170	0.90	0.166	0.94	0.188	0.93	0.186
	π_H	0.94	0.101	0.90	0.117	0.93	0.134	0.93	0.138

Table 2.4: Simulated width and 95% coverage rate (CR) for the estimators of π_k. π_k is the true stage 1 response rate for the treatment k, $k = P, L, H$, where P = placebo, L = low dose, and H = high dose. Scenarios are given in Table 1. Four modeling approaches: Bayesian joint stage modeling (BJSM), log-linear Poisson joint stage modeling (LPJSM), Bayesian first stage modeling (BFSM), and first stage maximum likelihood estimation (FSMLE) are compared. The sample size per treatment arm is 100.

to the true value of π_P in all scenarios. In Tables 2.8-2.9, we present simulations for the null and dose-response scenarios with prior distributions of Beta(2,18) and Beta(4,16), and means of 0.10 and 0.20, respectively, for π_P. While in these simulations, we did find that our estimates of the placebo response rate were more biased than in the results in Section 2.3.1, our estimation of response rates for low and high doses were unchanged. Coverage rate and credible interval width estimates were also unchanged in our sensitivity analyses (See Tables 2.10-2.13). Additionally, even when the mean of our prior distribution for the placebo response rate did not match the true value of π_P, BJSM was still more efficient than the LPJSM and first-stage methods (see Tables 2.12-2.13). Based on these additional analyses, we conclude that our method is generally robust to the choice of mean for all prior distributions of π_P, $\log(\pi_L/\pi_P)$, and $\log(\pi_H/\pi_P)$.

2.4 Discussion

In this chapter, we adapted the Baysian method (BJSM) for use in a different snSMART design where low and high doses of a single experimental therapy are compared to placebo. Due to dose

Scenario		BJSM		LPJSM		BFSM		FSMLE	
		Bias	rMSE	Bias	rMSE	Bias	rMSE	Bias	rMSE
(1) P = L = H	π_P	-0.002	0.047	0.012	0.116	-0.084	0.098	0.001	0.120
$\pi = 0.3$	π_L	-0.010	0.086	0.031	0.108	-0.021	0.110	0.003	0.121
	π_H	-0.015	0.075	0.018	0.100	-0.018	0.111	0.002	0.120
	π_L - π_P	-0.008	0.098	0.019	0.157	0.062	0.135	0.002	0.175
	π_H - π_P	-0.012	0.085	0.006	0.152	0.066	0.138	0.002	0.168
(2) P = L = H	π_P	0.000	0.053	0.007	0.123	-0.141	0.151	0.004	0.128
$\pi = 0.4$	π_L	-0.015	0.087	0.012	0.107	-0.027	0.119	-0.003	0.130
	π_H	-0.018	0.084	0.006	0.105	-0.023	0.120	-0.001	0.124
	π_L - π_P	-0.015	0.100	0.005	0.162	0.114	0.172	-0.007	0.182
	π_H - π_P	-0.018	0.100	-0.001	0.158	0.118	0.176	-0.005	0.175

Table 2.5: Simulated bias and root-mean-square error (rMSE) for the estimators of π_k under null scenarios with different spontaneous response rate. π_k is the true stage 1 response rate for the treatment k, $k = P, L, H$, where P = placebo, L = low dose, and H = high dose. Under null scenario, we assume $\pi_P = \pi_L = \pi_H = \pi$. Four modeling approaches: Bayesian joint stage modeling (BJSM), log-linear Poisson joint stage modeling (LPJSM), Bayesian first stage modeling (BFSM), and first stage maximum likelihood estimation (FSMLE) are compared. The sample size per treatment arm is 15.

comparison and the stage 2 re-randomization strategy, our design required novel methods that use six linkage parameters to share information on the response rates from both stages of the trial. In this setting, the BJSM yields accurate estimators that are easy to interpret in a clinical setting. Our proposed method was compared to three other methods via simulation. Through simulation, we demonstrated that BJSM estimators are the most efficient of the methods presented.

An advantage of the BJSM method, is that it provides estimates of π_P, π_L, and π_H, even when the true response rates were low. In our simulation scenarios, we noted convergence issues for the LPJSM method, specifically under scenario 1, where all treatments have true response rates of 0.15. In this scenario, there were instances where no response outcomes were observed for a given stage 1 treatment. Thus, there would be no responders to re-randomize in stage 2; all stage 2 re-randomization would occur through the non-responder arm of that treatment. This low probability of response caused failures in convergence for the LPJSM method, but good estimation with low bias was still possible using the BJSM.

Interestingly, in simulations with large true response rates or large sample size, LPJSM performs better than BFSM (smaller rMSE) in terms of the estimation of each individual response rate. However, LPJSM performs worse than BFSM in many scenarios in terms of the difference between the response rates of different dosage levels. This is likely due to our assumption of a

prior distribution on the ratio of response rates in the Bayesian methods, which implicitly places correlation among response rates. No such correlation structure is assumed with LPJSM.

Another strength of the BJSM method is its robustness under different prior settings. As discussed in Section 2.3.2, the BJSM method remained efficient regardless of the center of the response rate estimator prior distributions. Additionally, the bias of the response rate estimators for low and high dose remained low in all tested scenarios for the BJSM, even when the mean of the prior distribution for π_P no longer matched the true mean in the simulation scenarios. This robustness is particularly important for trials investigating drugs in rare diseases, as there may be little previous data to guide prior distribution selection.

Our first formulation of the BJSM model had eleven, rather than six, linkage parameters. These parameters corresponded to the eleven unique paths through which a participant could follow in the trial. We found that, while this model still produced response rate estimators with small bias and with increased efficiency compared to other tested methods, these advantages were not substantial. By limiting the model to only six linkage parameters, we were able to retain small bias and gains in efficiency, while using a simpler model. These efficiency gains were present for estimators of π_L - π_P and π_H - π_P. These difference estimators are generally of greater interest than individual response rates in clinical trials. As such, the efficiency gains we observe represent an advantage of using Bayesian joint stage modeling procedures. This model could be expanded if investigators wanted a different bias-variance trade-off than shown here.

The efficiency gains of the BJSM are still relevant for clinical trials with larger sample sizes (see Table 2.4). A trial design that reduces rMSE would also reduce the total number of patients that need to enroll in the trial, and therefore results in a shorter duration of the trial. As such, this snSMART design may be appropriate not just in rare disease research, but also in time-sensitive research like emerging infectious diseases. Similarly, efficiency gains of the BJSM remain for clinical trials with even smaller sample size. Simulations with only N = 45 patients (N = 15 per arm) showed that the BJSM remains efficient and estimates response rates with low bias, even as sample size decreases (see Table 2.5).

A limitation that results from the proposed prior distribution settings and model assumptions is that the posterior distributions for the linkage parameters and π_k allow for $\beta_{1k}\pi_{k'} > 1$. However, we did not draw any samples where $\beta_{1k}\pi_{k'} > 1$ in our simulations. Thus, it is unlikely that this limitation would be a problem in clinical settings, unless the treatment under consideration has a high response rate. Another limitation of our design is we assume there are no carryover effects of the stage 1 treatment in stage 2 of the study. We note, however, that our trial design allows for investigators to implement a washout period between stage 1 and stage 2 of the study if there was concern with carryover effects. Our future work will include modifications to our method to account for carryover effects.

Future directions for this work include adapting the BJSM to continuous outcomes. We are also examining models with fewer unique linkage parameters to see if we can improve efficiency of the BJSM method without much increased bias. In addition, future work can construct sample size calculations based on the BJSM for snSMART designs. These sample size calculations will aid in the extension of snSMART designs for primary efficacy analysis when more than two dose levels of a drug are compared. We note that our study design allows for a customizable randomization scheme in stage 2. While balanced re-randomization was applied in our simulations, future work could consider unbalanced or/and stratified randomization in stage 2 within responders and non-responders.

Scenario		BJSM		BFSM	
		Bias	rMSE	Bias	rMSE
(1) P = L = H	π_P	-0.001	0.039	-0.001	0.039
$\mu = 0.2$	π_L	-0.003	0.048	-0.005	0.062
	π_H	-0.007	0.043	-0.005	0.062
	π_L - π_P	-0.003	0.062	-0.005	0.073
	π_H - π_P	-0.006	0.058	-0.005	0.073
(2) P = L = H	π_P	-0.001	0.039	-0.001	0.039
$\mu = 0.3$	π_L	-0.003	0.048	-0.005	0.062
	π_H	-0.007	0.043	-0.005	0.062
	π_L - π_P	-0.003	0.062	-0.005	0.073
	π_H - π_P	-0.006	0.058	-0.005	0.073
(3) P = L = H	π_P	-0.001	0.039	-0.001	0.039
$\mu = 0.4$	π_L	-0.003	0.048	-0.005	0.062
	π_H	-0.007	0.043	-0.005	0.062
	π_L - π_P	-0.003	0.062	-0.005	0.073
	π_H - π_P	-0.006	0.058	-0.005	0.073
(4) P = L = H	π_P	-0.001	0.039	-0.001	0.039
$\mu = 0.5$	π_L	-0.003	0.048	-0.005	0.062
	π_H	-0.007	0.043	-0.005	0.062
	π_L - π_P	-0.003	0.062	-0.005	0.073
	π_H - π_P	-0.006	0.058	-0.005	0.073

Table 2.6: Simulated bias and root-mean-square error (rMSE) under the "null" scenario for the estimators when assuming different prior mean for $log(\pi_L/\pi_P)$ and $log(\pi_L/\pi_P)$, i.e. $E(log(\pi_L/\pi_P)) = \mu$, given $\mu = 0.2, 0.3, 0.4, 0.5$. Results are presented for two modeling approaches: Bayesian joint stage model (BJSM) and Bayesian first stage modeling (BFSM). π_k is the true stage 1 response rate for the treatment k, $k = P, L, H$, where P = placebo, L = low dose, and H = high dose. The sample size per treatment arm is 30.

Scenario		BJSM		BFSM	
		Bias	rMSE	Bias	rMSE
(1) P $<$ L $<$ H	π_P	0.000	0.039	0.000	0.039
$\mu = 0.2$	π_L	-0.005	0.057	-0.011	0.078
	π_H	-0.013	0.064	-0.013	0.084
	π_L - π_P	-0.005	0.069	-0.011	0.087
	π_H - π_P	-0.013	0.074	-0.012	0.093
(2) P $<$ L $<$ H	π_P	0.000	0.039	0.000	0.039
$\mu = 0.3$	π_L	-0.005	0.057	-0.011	0.078
	π_H	-0.013	0.064	-0.013	0.084
	π_L - π_P	-0.005	0.069	-0.011	0.087
	π_H - π_P	-0.013	0.074	-0.012	0.093
(3) P $<$ L $<$ H	π_P	0.000	0.039	0.000	0.039
$\mu = 0.4$	π_L	-0.005	0.057	-0.011	0.078
	π_H	-0.013	0.064	-0.013	0.084
	π_L - π_P	-0.005	0.070	-0.011	0.087
	π_H - π_P	-0.013	0.074	-0.012	0.093
(4) P $<$ L $<$ H	π_P	0.000	0.039	0.000	0.039
$\mu = 0.5$	π_L	-0.005	0.057	-0.011	0.078
	π_H	-0.013	0.064	-0.013	0.084
	π_L - π_P	-0.005	0.070	-0.011	0.087
	π_H - π_P	-0.013	0.074	-0.012	0.093

Table 2.7: Simulated bias and root-mean-square error (rMSE) under the dose-response scenario for the estimators when assuming different prior mean for $log(\pi_L/\pi_P)$ and $log(\pi_L/\pi_P)$, i.e. $E(log(\pi_L/\pi_P)) = \mu$, given $\mu = 0.2, 0.3, 0.4, 0.5$. Results are presented for two modeling approaches: Bayesian joint stage model (BJSM) and Bayesian first stage modeling (BFSM). π_k is the true stage 1 response rate for the treatment k, $k = P, L, H$, where P = placebo, L = low dose, and H = high dose. The sample size per treatment arm is 30.

Scenario		BJSM		BFSM	
		Bias	rMSE	Bias	rMSE
(1) P = L = H	π_P	-0.001	0.039	-0.001	0.039
$E(\pi_P) = 0.15$	π_L	-0.003	0.048	-0.005	0.062
	π_H	-0.007	0.043	-0.005	0.062
	π_L - π_P	-0.003	0.062	-0.005	0.073
	π_H - π_P	-0.006	0.058	-0.005	0.073
(2) P = L = H	π_P	-0.021	0.044	-0.021	0.044
$E(\pi_P) = 0.1$	π_L	-0.003	0.048	-0.005	0.062
	π_H	-0.007	0.043	-0.005	0.062
	π_L - π_P	0.017	0.064	0.015	0.075
	π_H - π_P	0.014	0.059	0.015	0.075
(3) P = L = H	π_P	0.019	0.043	0.019	0.043
$E(\pi_P) = 0.2$	π_L	-0.003	0.048	-0.005	0.062
	π_H	-0.007	0.043	-0.005	0.062
	π_L - π_P	-0.023	0.066	-0.025	0.077
	π_H - π_P	-0.026	0.063	-0.025	0.077

Table 2.8: Simulated bias and root-mean-square error (rMSE) under the null scenario for the estimators when assuming different shape and scale parameter values for the placebo prior distributions. Three different prior settings are presented: $\pi_P \sim Beta(3, 17)$, $\pi_P \sim Beta(2, 18)$ and $\pi_P \sim Beta(4, 16)$, corresponding to the placebo prior mean $E(\pi_P) = (0.15, 0.1, 0.2)$, respectively. Results are presented for two modeling approaches: Bayesian joint stage model (BJSM) and Bayesian first stage modeling (BFSM). π_k is the true stage 1 response rate for the treatment k, $k = P, L, H$, where P = placebo, L = low dose, and H = high dose. The sample size per treatment arm is 30.

Scenario		BJSM		BFSM	
		Bias	rMSE	Bias	rMSE
(1) P $<$ L $<$ H	π_P	0.000	0.039	0.000	0.039
$E(\pi_P) = 0.15$	π_L	-0.005	0.057	-0.011	0.078
	π_H	-0.013	0.064	-0.013	0.084
	π_L - π_P	-0.005	0.069	-0.011	0.087
	π_H - π_P	-0.013	0.074	-0.012	0.093
(2) P $<$ L $<$ H	π_P	-0.020	0.044	-0.020	0.044
$E(\pi_P) = 0.1$	π_L	-0.005	0.057	-0.011	0.078
	π_H	-0.013	0.064	-0.013	0.084
	π_L - π_P	0.015	0.071	0.009	0.087
	π_H - π_P	0.007	0.073	0.008	0.092
(3) P $<$ L $<$ H	π_P	0.020	0.044	0.020	0.044
$E(\pi_P) = 0.2$	π_L	-0.005	0.057	-0.011	0.078
	π_H	-0.013	0.064	-0.013	0.084
	π_L - π_P	-0.025	0.074	-0.031	0.091
	π_H - π_P	-0.033	0.080	-0.032	0.098

Table 2.9: Simulated bias and root-mean-square error (rMSE) under the dose-response scenario for the estimators when assuming different shape and scale parameter values for the placebo prior distributions. Three different prior settings are presented: $\pi_P \sim Beta(3, 17)$, $\pi_P \sim Beta(2, 18)$ and $\pi_P \sim Beta(4, 16)$, corresponding to the placebo prior mean $E(\pi_P) = (0.15, 0.1, 0.2)$, respectively. Results are presented for two modeling approaches: Bayesian joint stage model (BJSM) and Bayesian first stage modeling (BFSM). π_k is the true stage 1 response rate for the treatment k, $k = P, L, H$, where P = placebo, L = low dose, and H = high dose. The sample size per treatment arm is 30.

Scenario		BJSM		BFSM	
		CR	Width	CR	Width
(1) P = L = H	π_P	0.984	0.187	0.986	0.187
$\mu = 0.2$	π_L	0.925	0.183	0.872	0.221
	π_H	0.934	0.171	0.869	0.221
(2) P = L = H	π_P	0.985	0.187	0.986	0.187
$\mu = 0.4$	π_L	0.926	0.183	0.872	0.221
	π_H	0.934	0.171	0.880	0.221
(3) P = L = H	π_P	0.985	0.187	0.986	0.187
$\mu = 0.4$	π_L	0.926	0.183	0.872	0.221
	π_H	0.931	0.171	0.872	0.221
(4) P = L = H	π_P	0.984	0.187	0.986	0.187
$\mu = 0.5$	π_L	0.926	0.183	0.870	0.222
	π_H	0.933	0.171	0.870	0.221

Table 2.10: Simulated width and 95% coverage rate (CR) under the "null" scenario for the estimators when assuming different prior mean for $log(\pi_L/\pi_P)$ and $log(\pi_L/\pi_P)$, i.e. $E(log(\pi_L/\pi_P)) = \mu$, given $\mu = 0.2, 0.3, 0.4, 0.5$. Results are presented for two modeling approaches: Bayesian joint stage model (BJSM) and Bayesian first stage modeling (BFSM). π_k is the true stage 1 response rate for the treatment k, $k = P, L, H$, where P = placebo, L = low dose, and H = high dose. The sample size per treatment arm is 30.

Scenario		BJSM		BFSM	
		CR	Width	CR	Width
(1) P $<$ L $<$ H	π_P	0.984	0.187	0.988	0.187
$\mu = 0.2$	π_L	0.942	0.225	0.882	0.280
	π_H	0.942	0.225	0.924	0.317
(2) P $<$ L $<$ H	π_P	0.984	0.187	0.986	0.187
$\mu = 0.4$	π_L	0.942	0.225	0.880	0.280
	π_H	0.942	0.225	0.925	0.317
(3) P $<$ L $<$ H	π_P	0.984	0.187	0.988	0.187
$\mu = 0.4$	π_L	0.942	0.225	0.884	0.280
	π_H	0.941	0.256	0.926	0.316
(4) P $<$ L $<$ H	π_P	0.985	0.187	0.986	0.187
$\mu = 0.5$	π_L	0.942	0.225	0.882	0.280
	π_H	0.942	0.256	0.926	0.316

Table 2.11: Simulated width and 95% coverage rate (CR) under the dose-response scenario for the estimators when assuming different prior mean for $log(\pi_L/\pi_P)$ and $log(\pi_L/\pi_P)$, i.e. $E(log(\pi_L/\pi_P)) = \mu$, given $\mu = 0.2, 0.3, 0.4, 0.5$. Results are presented for two modeling approaches: Bayesian joint stage model (BJSM) and Bayesian first stage modeling (BFSM). π_k is the true stage 1 response rate for the treatment k, $k = P, L, H$, where P = placebo, L = low dose, and H = high dose. The sample size per treatment arm is 30.

Scenario		BJSM		BFSM	
		CR	Width	CR	Width
(1) P = L = H	π_P	0.984	0.187	0.986	0.187
$E(\pi_P) = 0.15$	π_L	0.925	0.183	0.872	0.221
	π_H	0.934	0.171	0.869	0.221
(2) P = L = H	π_P	0.950	0.174	0.950	0.174
$E(\pi_P) = 0.1$	π_L	0.927	0.183	0.876	0.221
	π_H	0.934	0.171	0.878	0.221
(3) P = L = H	π_P	0.991	0.198	0.992	0.198
$E(\pi_P) = 0.2$	π_L	0.927	0.183	0.880	0.221
	π_H	0.934	0.171	0.880	0.221

Table 2.12: Simulated width and 95% coverage rate (CR) under the "null" scenario for the estimators when assuming different shape and scale parameter values for the placebo prior distributions. Three different prior settings are presented: $\pi_P \sim Beta(3, 17)$, $\pi_P \sim Beta(2, 18)$ and $\pi_P \sim Beta(4, 16)$, corresponding to the placebo prior mean $E(\pi_P) = (0.15, 0.1, 0.2)$, respectively. Results are presented for two modeling approaches: Bayesian joint stage model (BJSM) and Bayesian first stage modeling (BFSM). π_k is the true stage 1 response rate for the treatment k, $k = P, L, H$, where P = placebo, L = low dose, and H = high dose. The sample size per treatment arm is 30.

Scenario		BJSM		BFSM	
		CR	Width	CR	Width
(1) P < L < H	π_P	0.984	0.187	0.988	0.187
$E(\pi_P) = 0.15$	π_L	0.942	0.225	0.882	0.280
	π_H	0.942	0.225	0.924	0.317
(2) P < L < H	π_P	0.942	0.174	0.952	0.174
$E(\pi_P) = 0.1$	π_L	0.939	0.225	0.884	0.280
	π_H	0.941	0.256	0.926	0.317
(3) P < L < H	π_P	0.986	0.198	0.987	0.198
$E(\pi_P) = 0.2$	π_L	0.940	0.225	0.884	0.280
	π_H	0.940	0.256	0.926	0.317

Table 2.13: Simulated width and 95% coverage rate (CR) under the dose-response scenario for the estimators when assuming different shape and scale parameter values for the placebo prior distributions. Three different prior settings are presented: $\pi_P \sim Beta(3, 17)$, $\pi_P \sim Beta(2, 18)$ and $\pi_P \sim Beta(4, 16)$, corresponding to the placebo prior mean $E(\pi_P) = (0.15, 0.1, 0.2)$, respectively. Results are presented for two modeling approaches: Bayesian joint stage model (BJSM) and Bayesian first stage modeling (BFSM). π_k is the true stage 1 response rate for the treatment k, $k = P, L, H$, where P = placebo, L = low dose, and H = high dose. The sample size per treatment arm is 30.

CHAPTER 3

Bayesian Methods to Compare Dose Levels to Placebo in a Small N, Sequential, Multiple Assignment, Randomized Trial (snSMART) with a Continuous Outcome

3.1 Introduction

The dearth of available treatments of rare diseases suggests that more clinical studies are needed (NORD, 2021). Although natural history and registry studies are important and can provide critical information guiding treatment development, randomized clinical trials (RCTs) are generally regarded as providing the strongest scientific evidence for treatment efficacy (Grimes and Schulz, 2002). However, rare disease trials are more likely than non-rare disease trials to be single arm (63.0% versus 29.6%) and non-randomized (64.5% versus 36.1%), and studies of rare diseases often have reduced power relative to studies of non-rare diseases (Bell and Smith, 2014). This differential occurs for a variety of reasons, including the lack of sufficient numbers of individuals required for a RCT, as well as participant reluctance to be part of RCTs, especially those with a placebo arm. It is therefore critical to develop novel clinical trial designs and analyses that can maximize information from these small trials.

An snSMART design has been previously proposed to compare the efficacy of three unique, potentially active treatments for a rare disease (Tamura et al., 2016; Wei et al., 2018), and we later extended it to studies in which one of the arms was placebo rather than an active agent (Fang et al., 2021). In these approaches, efficiency gains were demonstrated over a single-stage design in which each participant is randomized only once to a treatment, while also maintaining very little bias for the treatment effect estimates. However, these methods assumed that patient outcomes were binary, while registration trials of new drugs often collect a continuous measure of response, which can provide more statistical power than the dichotomized binary outcomes (Donner and

Eliasziw, 1994; Bhandari et al., 2002). Thus, in this chapter, we present an snSMART design that incorporates comparison of two dose levels of a drug to placebo when the primary outcome is continuous. The methods required for analysis of an snSMART with continuous outcomes are not simple extensions of the methods developed to analyze an snSMART with binary outcomes. Instead, the new design incorporates new features that require the development of different models and prior distribution considerations.

Our design and methods are motivated by examples in lymphangioleiomyomatosis (LAM), which is a progressive, cystic, rare lung disease in women. Although no curative treatment exists, sirolimus at 2 mg/day has been shown to be effective in preventing loss of forced expiratory volume (FEV_1) in women who have already demonstrated a loss of FEV_1 (McCormack et al., 2011). Currently, a placebo controlled trial of sirolimus at a lower dose of 1 mg/day is being conducted in women with early stage LAM and normal FEV_1 (MILED trial: ClinicalTrials.gov Identifier NCT03150914). The prototype snSMART design proposed in this research would be an excellent candidate design to simultaneously evaluate both a low and high dose for sirolimus or any other investigational drug in early stage LAM.

In Section 3.2, we conceptualize the snSMART design and develop a Bayesian approach that borrows information across both stages for the primary analysis of comparing the mean outcome of dose levels versus placebo. In Section 3.3, we conduct simulations to assess the bias and efficiency of our approach in a variety of settings and compare our proposed method to a similar method using only stage 1 data. We present concluding thoughts in Section 3.4.

3.2 Methods

3.2.1 Design

Figure 3.1 conceptualizes the proposed snSMART design comparing dose levels to placebo. The specific goal of this design is to estimate the differences in the first stage mean outcome between low dose and placebo and between high dose and placebo. Patients are equally randomized to placebo, low dose or high dose in stage 1 as the initial treatment and then are re-randomized in stage 2 conditional on the stage 1 treatment and a dichotomized response variable collected at the end of stage 1. This response variable can be a dichotomized version of the continuous stage 1 outcome, or a variable that differs altogether from the stage 1 outcome. An important feature of our design is that placebo is only a treatment option in stage 1. By doing so, every patient enrolled in the trial will be treated with a dose level of treatment by the end of stage 2, even when initially assigned to receive placebo in stage 1. This re-randomization scheme is also advantageous for participants receiving low dose in stage 1 because they are equally likely to receive a higher dose

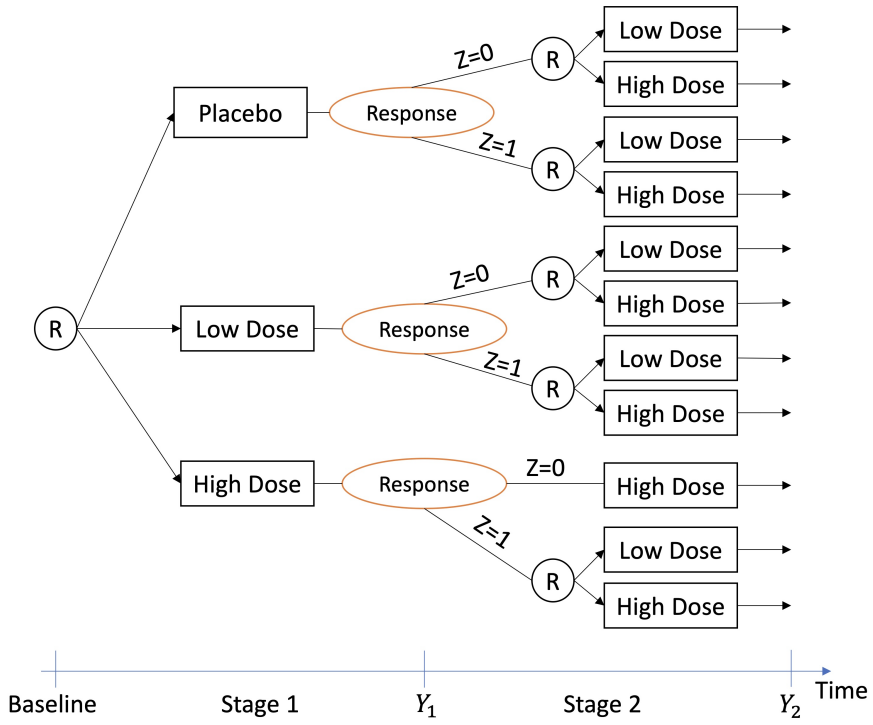

Figure 3.1: Study design of the proposed snSMART with a continuous endpoint. Participants are randomized (R) to one of the first stage treatment arms, placebo, low dose or high dose equally (1:1:1). At the end of stage 1, patients are re-randomized to their second stage treatment based on their response status. Outcomes are collected at the end of stage 1 and stage 2.

of the drug that is effective for them at a lower dose or to remain at the low dose. Receiving low dose in both stages may be advantageous because participants receive two administrations of a dose that could be effective for them, and the additional data helps us collect more information about the efficacy of low dose. Non-responders to high dose in stage 1 receive high dose again in stage 2, while patients who responded to high dose are re-randomized to either continue receiving high dose or receive low dose. This design allows high-dose responders to receive a lower dose later in stage 2 because a lower dose may still be efficacious and less toxic.

Using the setting of LAM as an example, women with normal lung function, defined as greater than or equal 70% predicted FEV_1, are enrolled and randomized at baseline to either placebo, 1 mg/day, or 2 mg/day of sirolimus. The primary endpoint is defined as change in FEV_1 from baseline, with increasingly negative changes indicating worsening lung function. After one year of follow-up for the primary endpoint, women are randomized to either 1 mg/day or 2 mg/day of sirolimus and are followed for an additional year. With regard to the response variable used to determine second stage treatment randomization, non-responders are defined as those women whose FEV_1 declined by 10% or more in stage 1 (McCormack et al., 2011; Taveira-DaSilva et al., 2004) . Non-responders to 2 mg/day of sirolimus are assigned to continue using the same dose level during stage 2, and all other participants are randomized between 1 mg/day and 2 mg/day.

3.2.2 Model

The trial will enroll a total of N subjects, each of whom will be observed for a continuous outcome at the end of each of two stages. For subject $i = 1, \ldots, N$ observed in stage $j = 1, 2$, we let T_{ij} be the assigned treatment and Y_{ij} denote the observed continuous outcome. A binary indicator (1=yes; 0=no) of response, Z_i, is measured at the end of S tage 1 and is used for re-randomization as illustrated in Figure 3.1.

In stage 1, we adopt the conditional model $Y_{i1} \sim N(\mu(T_{i1}), \sigma^2)$, in which $\mu(T_{i1}) = \sum_k \mu_k I(T_{i1} = k)$, where the three treatment arms are denoted by $k \in \{P, L, H\}$. In stage 2, we adopt the conditional model $Y_{i2} \sim N(\theta(T_{i1}, T_{i2}, Y_{i1}), \sigma^2)$, in which $\theta(T_{i1}, T_{i2}, Y_{i1}) = \mu_{k'} I(T_{i2} = k') + \alpha + \beta(Y_{i1} - \mu(T_{i1}))$ and $k' \in L, H$. Thus, the mean outcome after stage 2 for a treatment is the mean outcome that would have been observed for that treatment after stage 1, but augmented by a shift parameter α and a proportional residual response $[Y_{i1} - \mu(T_{i1})]$ from stage 1. Here, α serves as a parameter that allows for increases or decreases in treatment response in stage 2 relative to stage 1. We refer to β as a linkage parameter because it links an individual's residual in stage 1 to their mean outcome in stage 2, which implicitly induces a within-subject correlation of $\beta/\sqrt{1+\beta^2}$ between Y_{i1} and Y_{i2}.

3.2.3 Prior Distributions of Parameters

We incorporate prior knowledge by first assuming that the placebo mean has a $N(\mu_0, \eta^2)$ prior distribution, where the value of η quantifies how confident we are about the placebo having mean μ_0. This information would likely be informed by natural history studies, drug registries or expert opinions. For the mean outcomes of participants assigned to low and high dose, we use normal prior distributions with prior means $\mu_0 + \gamma_1$ and $\mu_0 + \gamma_2$, respectively, each with variance η^2. In practice, one can assume different prior standard deviations for the mean outcomes of all three treatment arms. We also assume σ, the standard deviation of the stage 1 and 2 outcomes, has a $Gamma(\theta_1, \theta_2)$ prior distribution.

The elicitation of prior distributions from experts' opinions is not always straightforward due to the scarcity of historic information or may not be representative of current data. Thus, for the prior distributions for the mean outcomes, we also consider a mixture prior approach that mixes an informative prior with a non-informative one. Suppose we denote two proper probability densities of a parameter x as $f_1(x)$ and $f_2(x)$, then $f(x) = \sum_{k=1}^{2} w_i f_i(x)$ is the density of a mixture distribution that consists of the two component densities with weights w_1 and w_2, respectively, given that $w_1 + w_2 = 1$. We assume such a mixture distribution for the prior distribution of each mean parameter of interest.

3.2.4 Computations of Posterior Distributions and Considerations

Based on the conditional joint stage model construction in Section 3.2.2, the observed outcomes from the two stages jointly follow a bivariate normal distribution as

$$Y_{i1}, Y_{i2}|_{T_{i1}=k, T_{i2}=k', \mu_P, \mu_L, \mu_H, \alpha, \beta} \overset{i.i.d}{\sim} BVN\left(\begin{pmatrix} \mu_k \\ \mu_{k'} + \alpha \end{pmatrix}, \begin{pmatrix} \sigma^2 & \beta\sigma^2 \\ \beta\sigma^2 & (\beta^2+1)\sigma^2 \end{pmatrix}\right).$$

The marginal expected outcome observed at the end of stage 2 is the mean outcome of the treatment received at stage 2 as if it were received in stage 1, but adjusted by the shift parameter α. The marginal variance of samples collected in stage 2 is influenced by the linkage parameter and becomes $(\beta^2 + 1)\sigma^2$. Meanwhile, the linkage parameter induces a within-subject correlation of $\beta/\sqrt{(1+\beta^2)}$ between Y_{i1} and Y_{i2}. Having the likelihood constructed based on the model assumption as above, the posterior distribution of all unknown parameters $\theta = (\mu_P, \mu_L, \mu_H, \alpha, \beta)$ given the observed data can then be computed by incorporating the prior distributions as specified in Section 3.2.3. We draw samples from the posterior distribution generated via Monte Carlo Markov Chain algorithm and we take the posterior means of each parameter as their Bayesian estimators.

Including the shift parameter α in the proposed model adds in flexibility and allows researchers to account for potential resistance to the drug given historic knowledge of the drug. The linkage

parameter results in more variability in stage 2 observations, but also helps in efficiency gains in estimating the mean outcomes of low dose and high dose by inducing within-patient correlation. The estimation of placebo is impacted by stage 2 data and therefore likely to be slightly biased, but our simulations suggest that the contrast between drug levels and placebo remain essentially unbiased and less variable than stage 1 estimators.

3.3 Simulations

3.3.1 Data Generation and Parameter Values

Via simulation, we examine the performance of our approach in three scenarios for a trial that will enroll a total of N participants. For each simulation, we assign $N/3$ participants equally to receive placebo, low and high doses of the study treatment in stage 1. The observed continuous outcomes collected after stage 1 are randomly drawn from a normal distribution with standard deviation $\sigma = 25$ and with means being drawn from the simulation scenarios to be specified later in this section. The response variable Z is set to 1 for any stage 1 outcome greater than or equal to 0 (e.g. No change or increase in FEV_1 from baseline). The participants who receive high dose in stage 1 with $Z = 0$ remain at the same dose level in stage 1, while the others are re-randomized with equal probabilities to receive low or high dose of the study treatment in stage 2. We set values of α and β for the conditional mean stage 2 outcomes as specified below. Finally, the continuous endpoints at the end of stage 2 are sampled from a normal distribution with the same standard deviation of $\sigma = 25$.

In terms of the parameter values for μ_P, μ_L and μ_H, we consider three scenarios. In Scenario 1, we simulate stage 1 outcomes under a "null" setting in which placebo, low and high doses all have the same mean outcome. We have selected $\mu_P = \mu_L = \mu_H = -60$, which is motivated by the results presented in McCormack et al. (2011). We then examine the performance of our approach in two alternative settings. In Scenario 2, we have a dose-response pattern among the three arms, with mean stage 1 outcomes increasing with the dose. Specifically, we have selected $\mu_P = -75$, $\mu_L = 0$, and $\mu_H = 25$. In Scenario 3, we examine a setting in which the low and high doses are equally effective, so that $\mu_P = -75$ and $\mu_L = \mu_H = 0$. All other parameter values used in these two scenarios are equal to those used in Scenario 1.

As the benchmark simulations, we implemented the three data generating scenarios described above with a total sample size of $N = 60$, $\alpha = 0$ and $\beta = 1$ that yields a conditional correlation of 0.71 between the primary outcomes in the two stages. We repeated our simulations for these three scenarios by first using smaller sample sizes ($N = 30$ and $N = 15$) to assess if smaller numbers of participants could still lead to acceptable operating characteristics. We also conducted a set of

simulations in which outcomes are no longer normally distributed by adding a skewed error to each individual mean. The skewed error is built by sampling values from a Gamma distribution with mean 2σ and standard deviation σ that are then shifted by 2σ so that the error terms are centered at 0 with a skewness of 0.28. To address the potential impact of discrepancy between our assumed prior means and true parameter values, we performed additional simulations in which $\alpha = -2$ and $\beta = 1$ and in which $\alpha = 0$ and $\beta = 0.5$. Because our model assumes values for α and β that are constant regardless of treatment received in stage 1, we performed additional simulations in which one of the two parameters, or both parameters, have values that are different across the three treatment arms.

3.3.2 Prior Distributions

In our Bayesian analysis, we examine two different sets of prior distributions. In the first set, we assume $\mu_P \sim N(-75, 25^2)$, $\mu_L \sim N(0, 25^2)$, and $\mu_H \sim N(25, 25^2)$, so that the means in the three prior distributions match the actual values used to simulate data in Scenario 2, i.e. the mean responses in stage 1 increase with dose. We refer to this as our "optimistic" set of priors. We also assume $\sigma \sim Gamma(25, 1)$, $\alpha \sim N(0, 2)$, and $\beta \sim N(0, 1)$. In the second set of prior distributions, which we refer to as a mixture set of priors, we mix the "optimistic" prior of each treatment mean parameter with a non-informative prior with equal weights ($w_1 = w_2 = 0.5$). The non-informative prior for each mean parameter is selected as a normal distribution with a mean of zero and a variance of 1000. This mixture allows for a more flexible shape of the prior distributions and indicates less certainty in the treatment effect. The prior distributions for σ, α, and β remain the same.

3.3.3 Results

All simulations were done using R, version 4.0.3. Each scenario was simulated 2,000 times and draws from the posterior distributions of all parameters were generated using Markov Chain Monte Carlo via the R library rjags. The posterior mean is used as a point estimate and we compute the highest posterior density (HPD) regions that include 95% of the posterior draws. Each scenario is summarized by the mean bias, root mean-square error (rMSE), and the coverage rate and width of the 95% credible intervals (CI) of the three stage 1 means, as well as the stage 1 mean differences between placebo and each of low and high doses. As a comparator, we also fit our models based solely on the data from the first stage. Results can be found in Tables 3.1 - 3.4. Simulation codes are available at `https://github.com/sqrfang/snSMART_PLH`.

In Table 3.1, we present the simulation results under Scenarios 1-3 with the optimistic prior setting. Across all three scenarios, the joint stage model outperforms the first stage model with

Scenario	Parameter	Bias		rMSE		CR		Width	
		JS	FS	JS	FS	JS	FS	JS	FS
1	μ_P	0.79	-0.64	4.23	5.34	0.96	0.96	17.12	21.66
	μ_L	2.46	2.96	4.86	5.97	0.92	0.93	16.78	21.74
	μ_H	3.16	4.33	5.28	6.82	0.90	0.89	16.97	21.82
	$\mu_L - \mu_P$	1.67	3.60	5.10	8.17	0.95	0.94	19.97	30.72
	$\mu_H - \mu_P$	2.37	4.97	5.42	8.85	0.94	0.92	20.50	30.78
2	μ_P	0.00	0.09	4.15	5.31	0.96	0.96	17.07	21.52
	μ_L	0.06	0.01	3.96	5.16	0.96	0.96	16.02	21.53
	μ_H	0.20	0.16	4.05	5.22	0.95	0.96	16.05	21.53
	$\mu_L - \mu_P$	0.07	-0.08	4.65	7.31	0.97	0.96	19.51	30.45
	$\mu_H - \mu_P$	0.21	0.07	4.70	7.29	0.96	0.95	19.60	30.44
3	μ_P	0.21	0.09	4.16	5.31	0.96	0.96	17.06	21.53
	μ_L	0.38	0.01	4.04	5.16	0.96	0.96	16.27	21.53
	μ_H	0.86	1.37	4.18	5.40	0.94	0.95	16.36	21.54
	$\mu_L - \mu_P$	0.16	-0.08	4.71	7.31	0.96	0.96	19.65	30.45
	$\mu_H - \mu_P$	0.64	1.28	4.78	7.40	0.96	0.95	19.89	30.46

Table 3.1: Bias, root mean-square error (rMSE), coverage rates (CR) and the width of 95% credible intervals for the posterior mean treatment effects under Scenarios 1-3 (see Section 3.2.2) with total sample size $N = 60$, using the optimistic prior setting. μ_P,μ_L and μ_H are the mean outcome for placebo, low dose and high dose in stage 1, respectively. JS=joint stage model, FS=first stage model.

comparably lower rMSE and narrower 95% credible intervals in estimating the mean outcomes of each dose level and their differences with placebo. In Scenario 1, μ_L and μ_H are overestimated in both the joint stage model and the first stage model, affecting the bias of their differences with μ_P as well. In Scenario 2, both joint stage estimators and first stage estimators for each dose level have low bias because we used priors that are centered around the true parameter values. As for Scenario 3, the prior mean of μ_H is far from the true value, while the prior means of μ_P and μ_L are correctly specified. We also see that the simulated coverage rates of the joint stage estimators for μ_L and μ_H are slightly below 95% under Scenario 1. However, the coverage rates for the difference in mean stage 1 outcomes are stable at around 95%, which is important because oftentimes decisions are made on the difference in means.

Table 3.2 consists of the same scenarios as shown in Table 1, but using the mixture prior setting. In Scenario 1, we observe slightly less biased estimates of μ_L and μ_H as well as their differences between placebo, as compared to the same models in Table 3.1, and we observe similarly biased estimates of parameters for mean outcomes under Scenarios 2 and 3. The differences in the widths of the 95% HPD regions and rMSEs for the estimates between Table 3.1 and Table 3.2 are neg-

Scenario	Parameter	Bias		rMSE		CR		Width	
		JS	FS	JS	FS	JS	FS	JS	FS
1	μ_P	0.26	-0.63	4.21	5.35	0.96	0.96	17.18	21.61
	μ_L	1.73	2.57	4.65	5.96	0.94	0.93	17.02	22.09
	μ_H	1.41	1.60	4.82	6.20	0.93	0.93	17.69	23.22
	$\mu_L - \mu_P$	1.47	3.20	5.07	8.13	0.95	0.94	20.04	30.93
	$\mu_H - \mu_P$	1.15	2.23	5.17	8.15	0.96	0.94	20.81	31.74
2	μ_P	0.00	0.09	4.15	5.31	0.96	0.96	17.07	21.54
	μ_L	0.07	0.01	3.96	5.17	0.96	0.96	16.02	21.53
	μ_H	0.20	0.16	4.05	5.23	0.95	0.96	16.05	21.53
	$\mu_L - \mu_P$	0.07	-0.08	4.65	7.32	0.96	0.96	19.51	30.45
	$\mu_H - \mu_P$	0.21	0.07	4.70	7.29	0.96	0.95	19.61	30.45
3	μ_P	0.21	0.09	4.16	5.31	0.96	0.96	17.07	21.54
	μ_L	0.37	0.01	4.05	5.17	0.96	0.96	16.28	21.54
	μ_H	0.85	1.36	4.18	5.41	0.95	0.95	16.36	21.56
	$\mu_L - \mu_P$	0.16	-0.08	4.71	7.32	0.96	0.96	19.66	30.46
	$\mu_H - \mu_P$	0.64	1.26	4.78	7.41	0.96	0.95	19.90	30.48

Table 3.2: Bias, root mean-square error (rMSE), coverage rates (CR) and the width of 95% credible intervals for the posterior mean treatment effects under Scenarios 1-3 (see Section 3.2.2) with total sample size $N = 60$, using the mixture prior setting. μ_P,μ_L and μ_H are the mean outcome for placebo, low dose and high dose in stage 1, respectively. JS=joint stage model, FS=first stage model.

ligible in Scenarios 2 and 3, while estimates in Table 3.2 have similar width and smaller rMSE because of less bias. Overall, we still see the efficiency gain in the estimates of all mean outcome parameters when using the joint stage estimators, which carries through to the differences between low dose versus placebo, and high dose versus placebo. The coverage rates of all parameters under all scenarios are around 95%.

Scenario	Parameter	Bias		rMSE		CR		Width	
		JS	FS	JS	FS	JS	FS	JS	FS
Optimistic priors									
1	μ_P	1.47	-1.37	5.79	7.11	0.96	0.96	24.05	30.26
	μ_L	4.90	6.05	7.51	9.26	0.89	0.90	23.66	30.55
	μ_H	5.61	8.10	8.14	10.85	0.86	0.85	23.72	30.91
	$\mu_L - \mu_P$	3.43	7.42	7.66	12.34	0.94	0.92	28.55	43.12
	$\mu_H - \mu_P$	4.14	9.47	8.12	13.94	0.92	0.88	28.83	43.42
2	μ_P	0.05	0.06	5.59	7.00	0.97	0.96	23.83	29.68
	μ_L	0.36	0.32	5.43	7.00	0.96	0.96	22.29	29.68
	μ_H	-0.07	-0.01	5.56	7.15	0.95	0.96	22.27	29.68
	$\mu_L - \mu_P$	0.31	0.26	6.67	9.86	0.96	0.96	27.56	41.98
	$\mu_H - \mu_P$	-0.12	-0.07	6.67	10.19	0.96	0.96	27.59	41.97
3	μ_P	0.44	0.06	5.61	7.00	0.97	0.96	23.82	29.70
	μ_L	0.91	0.32	5.61	7.00	0.95	0.96	22.68	29.71
	μ_H	1.19	2.29	5.73	7.51	0.95	0.95	22.65	29.76
	$\mu_L - \mu_P$	0.47	0.26	6.75	9.86	0.96	0.96	27.81	42.01
	$\mu_H - \mu_P$	0.75	2.23	6.78	10.43	0.96	0.95	27.91	42.05
Mixture priors									
1	μ_P	0.51	-1.33	5.72	7.12	0.97	0.96	24.18	30.03
	μ_L	3.60	5.25	7.02	9.18	0.91	0.91	24.24	31.39
	μ_H	2.40	3.11	7.30	9.66	0.91	0.91	25.48	34.26
	$\mu_L - \mu_P$	3.09	6.58	7.62	12.18	0.93	0.92	28.71	43.53
	$\mu_H - \mu_P$	1.89	4.44	7.68	12.53	0.94	0.93	29.58	45.68
2	μ_P	0.04	0.06	5.59	7.01	0.97	0.96	23.84	29.69
	μ_L	0.36	0.32	5.43	7.01	0.96	0.96	22.29	29.70
	μ_H	-0.08	-0.01	5.56	7.15	0.95	0.96	22.28	29.70
	$\mu_L - \mu_P$	0.31	0.26	6.67	9.87	0.96	0.96	27.56	42.00
	$\mu_H - \mu_P$	-0.12	-0.07	6.67	10.19	0.96	0.96	27.59	41.99
3	μ_P	0.44	0.06	5.61	7.01	0.97	0.96	23.83	29.72
	μ_L	0.91	0.32	5.61	7.01	0.95	0.96	22.68	29.73
	μ_H	1.17	2.26	5.74	7.52	0.95	0.95	22.67	29.81
	$\mu_L - \mu_P$	0.47	0.27	6.75	9.87	0.96	0.96	27.81	42.04
	$\mu_H - \mu_P$	0.73	2.20	6.78	10.44	0.96	0.95	27.92	42.09

Table 3.3: Bias, root mean-square error (rMSE), coverage rates (CR) and the width of 95% credible intervals for the posterior mean treatment effects under Scenarios 1-3 (see Section 3.2.2) with total sample size N = 30, using the optimistic prior setting or mixture prior setting. μ_P,μ_L and μ_H are the mean outcome for placebo, low dose and high dose in stage 1, respectively. JS=joint stage model, FS=first stage model.

Scenario	Parameter	Bias		rMSE		CR		Width	
		JS	FS	JS	FS	JS	FS	JS	FS
Optimistic priors									
1	μ_P	2.39	-2.64	7.87	9.40	0.97	0.97	34.19	42.03
	μ_L	8.79	11.07	11.75	14.31	0.84	0.86	33.46	43.10
	μ_H	10.59	15.84	13.13	18.35	0.79	0.73	33.42	44.21
	$\mu_L - \mu_P$	6.40	13.72	11.48	18.80	0.92	0.89	41.64	60.60
	$\mu_H - \mu_P$	8.20	18.49	12.42	22.71	0.91	0.82	41.54	61.58
2	μ_P	0.25	0.10	7.50	9.18	0.97	0.97	33.36	40.15
	μ_L	0.35	0.10	7.40	9.22	0.96	0.98	30.74	40.16
	μ_H	-0.10	0.29	7.36	9.34	0.96	0.96	30.59	40.15
	$\mu_L - \mu_P$	0.11	-0.01	9.17	13.03	0.96	0.97	39.49	56.78
	$\mu_H - \mu_P$	-0.35	0.19	8.96	13.29	0.97	0.96	39.41	56.78
3	μ_P	0.91	0.11	7.56	9.17	0.97	0.97	33.33	40.24
	μ_L	1.18	0.10	7.65	9.20	0.96	0.98	31.29	40.24
	μ_H	2.35	4.50	7.71	10.37	0.95	0.95	31.06	40.41
	$\mu_L - \mu_P$	0.28	-0.01	9.25	13.01	0.96	0.97	39.82	56.92
	$\mu_H - \mu_P$	1.45	4.40	9.16	14.01	0.96	0.95	39.72	57.04
Mixture priors									
1	μ_P	0.90	-2.51	7.80	9.46	0.97	0.97	34.40	41.40
	μ_L	6.62	9.33	10.88	13.98	0.89	0.88	34.85	45.18
	μ_H	5.09	7.53	11.29	15.78	0.89	0.88	37.40	51.74
	$\mu_L - \mu_P$	5.72	11.84	11.42	18.27	0.93	0.90	42.17	61.61
	$\mu_H - \mu_P$	4.19	10.04	11.27	19.64	0.94	0.89	43.40	66.68
2	μ_P	0.24	0.10	7.51	9.19	0.97	0.97	33.40	40.21
	μ_L	0.35	0.09	7.40	9.23	0.96	0.97	30.75	40.22
	μ_H	-0.10	0.29	7.37	9.37	0.96	0.96	30.62	40.20
	$\mu_L - \mu_P$	0.11	-0.01	9.17	13.05	0.96	0.97	39.52	56.86
	$\mu_H - \mu_P$	-0.35	0.19	8.97	13.32	0.97	0.96	39.44	56.86
3	μ_P	0.90	0.11	7.57	9.19	0.97	0.97	33.37	40.30
	μ_L	1.17	0.09	7.66	9.23	0.96	0.98	31.32	40.30
	μ_H	2.31	4.43	7.72	10.40	0.95	0.95	31.13	40.57
	$\mu_L - \mu_P$	0.27	-0.01	9.26	13.04	0.96	0.97	39.86	56.99
	$\mu_H - \mu_P$	1.41	4.32	9.17	14.05	0.96	0.95	39.76	57.18

Table 3.4: Bias, root mean-square error (rMSE), coverage rates (CR) and the width of 95% credible intervals for the posterior mean treatment effects under Scenarios 1-3 (see Section 3.2.2) with total sample size N = 15, using the optimistic prior setting. μ_P,μ_L and μ_H are the mean outcome for placebo, low dose and high dose in stage 1, respectively. JS=joint stage model, FS=first stage model.

We then repeated the simulations in Scenarios 1-3 using the two prior settings, but with total sample sizes of 30 and 15. The simulation results are presented in Table 3.3 for N = 30 and Table 3.4 for $N = 15$. In Scenario 1, as the sample size decreases, we observe increases in bias and rMSE for all estimators and slightly lower coverage rates than expected. In Scenarios 2 and 3, the bias using optimistic priors with smaller total sample size is comparable to the bias with $N = 60$, but estimates from smaller sample size have greater rMSE and width of the 95% credible intervals. Nevertheless, the joint stage model provides estimators with more efficiency and comparable or less bias, than the first stage model.

3.3.4 Sensitivity to Error Distributions

We built skewed error distributions for the observed outcomes in simulated data by sampling from a Gamma distribution with 2σ as the mean and σ as the standard deviation, then taking 2σ out of each sample. We repeat the simulated scenarios shown in Table 3.1 with the modified error distribution and present the results in Table 3.5.

In general, we observe very similar accuracy and efficiency of our models comparing the two data generation approaches. The difference in mean outcomes for dose levels versus placebo still have comparable bias when being compared to the reference scenarios. Notice that the similarity is due to the skewed error term being built in a way that mimics a normal distributed error with skewness of 0.28. Higher bias is likely to be observed when the error term has higher skewness.

3.3.5 Sensitivity to Shift and Linkage Parameter

Tables 3.6-3.7 demonstrate that joint stage model results are impacted by the discrepancy between the true values of α and β and their respective prior distributions. Specifically, in Table 3.6, although deviation of α from its prior mean leads to more biased estimators for the mean response of all treatment arms, there is no substantial change in their efficiency. In Table 3.7, we see that the true value of β impacts the efficiency of our estimates by altering the within-subject correlations in the outcomes, such that there is greater rMSE when correlation is lower.

Last, in Table 3.8, in which our assumption of constant values for α and β is violated, we find increased bias for the treatment means and slightly lower coverage rates when β actually varies by treatment arm. Bias and efficiency differ from ideal settings when assumptions about α and β are violated, but bias remains low in general and using data from both stages in a joint stage model is more efficient than using data from only the first stage. Thus, our methods are relatively robust to the prior distributions chosen for α and β and using a parsimonious model even if there are treatment specific α and β values.

Scenario		Bias		rMSE		CR		Width	
		JS	FS	JS	FS	JS	FS	JS	FS
1	μ_P	0.71	-0.79	4.24	5.35	0.95	0.95	17.13	21.59
	μ_L	2.39	2.77	4.99	6.14	0.91	0.93	16.79	21.67
	μ_H	2.88	4.18	5.25	6.95	0.90	0.89	16.92	21.76
	$\mu_L - \mu_P$	1.68	3.57	5.22	8.29	0.94	0.93	20.04	30.62
	$\mu_H - \mu_P$	2.17	4.97	5.53	9.08	0.93	0.92	20.46	30.69
2	μ_P	-0.08	-0.06	4.15	5.33	0.96	0.95	17.07	21.46
	μ_L	0.03	-0.16	4.09	5.28	0.95	0.95	16.02	21.46
	μ_H	-0.13	0.02	4.07	5.27	0.95	0.96	16.01	21.46
	$\mu_L - \mu_P$	0.12	-0.10	4.80	7.45	0.95	0.95	19.55	30.35
	$\mu_H - \mu_P$	-0.05	0.08	4.86	7.54	0.95	0.95	19.57	30.36
3	μ_P	0.14	-0.06	4.16	5.34	0.96	0.95	17.07	21.47
	μ_L	0.32	-0.16	4.20	5.28	0.95	0.95	16.34	21.47
	μ_H	0.56	1.23	4.17	5.49	0.95	0.95	16.38	21.48
	$\mu_L - \mu_P$	0.18	-0.10	4.86	7.45	0.95	0.95	19.74	30.36
	$\mu_H - \mu_P$	0.42	1.29	4.94	7.65	0.95	0.95	19.93	30.37

Table 3.5: Bias, root mean-square error (rMSE), coverage rates (CR) and the width of 95% credible intervals for the posterior mean treatment effects under Scenarios 1-3 (see Section 3.2.2) with Gamma-distributed individual error, using the optimistic prior setting. μ_P, μ_L and μ_H are the mean outcome of having placebo, low dose and high dose in stage 1, respectively. JS=joint stage model, FS=first stage model.

3.4 Discussion

In this chapter, we propose an snSMART design that compares dose levels to placebo when the outcome is continuous and describe how this model can be applied to a trial studying treatment for the rare disease LAM. We propose a Bayesian model to obtain the stage 1 mean outcomes of each dose level and test its performance across various assumptions via simulations. We conclude that the proposed model that incorporates data from both stages provides results with low bias and high efficiency compared to a similar model that uses only first stage data.

The endpoint of interest at the end of stage 1 and 2 in the proposed snSMART design is continuous, but the design requires a binary outcome variable for re-randomization to second stage treatment. In our example and the simulations, we defined the response variable, Z, as the dichotomization of the primary continuous outcome collected at the end of stage 1. In reality, this response variable can actually be any binary outcome that is reasonable to investigators, having either a strong or weak correlation to the primary outcome. Allowing this variable to differentiate from the primary outcome is especially helpful when the measurement of the primary outcome is expensive or time-consuming and could hold up the second stage treatment assignment. The binary response variable is not directly used in our proposed model because it is less informative than the actual continuous outcome observed at the end of stage 1 which would be readily available at the end of the trial for analysis.

Scenario		Bias		rMSE		CR		Width	
		JS	FS	JS	FS	JS	FS	JS	FS
1	μ_P	0.12	-0.64	4.93	5.34	0.96	0.96	19.90	21.65
	μ_L	2.18	2.96	4.71	5.97	0.93	0.93	16.99	21.74
	μ_H	2.56	4.33	4.55	6.82	0.92	0.89	15.65	21.82
	$\mu_L - \mu_P$	2.06	3.60	6.15	8.17	0.94	0.94	23.92	30.71
	$\mu_H - \mu_P$	2.44	4.97	6.17	8.85	0.95	0.92	23.57	30.79
2	μ_P	0.02	0.09	4.92	5.31	0.96	0.96	19.85	21.53
	μ_L	0.01	0.01	3.79	5.16	0.97	0.96	15.75	21.53
	μ_H	0.21	0.16	3.79	5.22	0.96	0.96	15.56	21.53
	$\mu_L - \mu_P$	-0.02	-0.08	5.61	7.31	0.96	0.96	23.33	30.44
	$\mu_H - \mu_P$	0.19	0.07	5.63	7.29	0.96	0.95	23.30	30.44
3	μ_P	0.14	0.09	4.92	5.31	0.96	0.96	19.84	21.53
	μ_L	0.20	0.01	3.91	5.16	0.97	0.96	16.19	21.54
	μ_H	0.81	1.37	3.84	5.41	0.95	0.95	15.58	21.55
	$\mu_L - \mu_P$	0.06	-0.08	5.67	7.31	0.96	0.96	23.53	30.46
	$\mu_H - \mu_P$	0.67	1.28	5.65	7.40	0.96	0.95	23.38	30.47

Table 3.6: Bias, root mean-square error (rMSE), coverage rates (CR) and the width of 95% credible intervals for the posterior mean treatment effects under Scenarios 1-3 (see Section 3.3.2) with $\alpha = 0$ and $\beta = 0.5$, using the optimistic prior setting. μ_P, μ_L and μ_H are the mean outcome of having placebo, low dose and high dose in stage 1, respectively. JS=joint stage model, FS=first stage model.

In the simulated trials with very small total sample sizes (< 30), we observe greater bias as the prior means deviate farther away from the actual values. We also observe slightly lower coverage rates for the credible intervals for low and high dose. However, well-specified prior distributions can substantially improve the accuracy and efficiency. Investigators should exercise extra caution in the choice of the prior distributions for the means when the total sample size is small (≤ 30).

Under all primary simulation scenarios, we tested two sets of prior distributions, including an optimistic one and one with a mixture prior, both assuming normal priors as the basic elements for all three parameters that represent the mean outcomes for three dose levels. The normal distribution is most intuitive since it is well-known among researchers and the parameters are easily elicited from experts. Morita et al. (2008) demonstrated that the effective sample size for a normal prior distribution is the ratio of the sampling variance to the normal prior variance when sampling variance is known. This helps in determining the values of the hyperparameters used in prior distributions. For the mixture prior distribution, we assumed equal weights for the informative and non-informative element distributions although we could consider unequal weights or even adaptively determining the weights. For example, researchers may consider incorporating ways such as an automatic prior elicitation method (Egidi et al., 2021) to choose the weight given the observed data. We also found that the estimates using the mixture priors are robust especially when the prior means misspecify the true mean outcomes. The mixture prior is associated with less bias and better

Scenario		Bias		rMSE		CR		Width	
		JS	FS	JS	FS	JS	FS	JS	FS
1	μ_P	1.40	-0.64	4.39	5.34	0.95	0.96	17.11	21.66
	μ_L	2.44	2.96	4.85	5.97	0.92	0.93	16.79	21.73
	μ_H	2.56	4.33	4.96	6.82	0.91	0.89	17.01	21.81
	$\mu_L - \mu_P$	1.05	3.60	4.93	8.17	0.96	0.94	19.94	30.71
	$\mu_H - \mu_P$	1.17	4.97	5.02	8.85	0.96	0.92	20.48	30.79
2	μ_P	0.58	0.09	4.19	5.31	0.96	0.96	17.08	21.52
	μ_L	-0.22	0.01	3.96	5.16	0.96	0.96	16.03	21.52
	μ_H	-0.15	0.16	4.07	5.22	0.95	0.96	16.10	21.52
	$\mu_L - \mu_P$	-0.80	-0.08	4.72	7.31	0.97	0.96	19.52	30.44
	$\mu_H - \mu_P$	-0.73	0.07	4.76	7.29	0.96	0.95	19.63	30.44
3	μ_P	0.80	0.09	4.23	5.31	0.96	0.96	17.07	21.53
	μ_L	0.19	0.01	4.02	5.16	0.96	0.96	16.26	21.53
	μ_H	0.42	1.37	4.15	5.41	0.95	0.95	16.41	21.55
	$\mu_L - \mu_P$	-0.61	-0.08	4.74	7.31	0.96	0.97	19.65	30.45
	$\mu_H - \mu_P$	-0.38	1.28	4.77	7.40	0.96	0.95	19.92	30.46

Table 3.7: Bias, root mean-square error (rMSE), coverage rates (CR) and the width of 95% credible intervals for the posterior mean treatment effects under Scenarios 1-3 (see Section 3.3.2) with $\alpha = -2$ and $\beta = 1$, using the optimistic prior setting. μ_P, μ_L and μ_H are the mean outcome of having placebo, low dose and high dose in stage 1 respectively. JS=joint stage model, FS=first stage model.

coverage under the null scenario where the treatment groups are equivalent.

When this study design is used in practice, we note a few additional considerations. When describing our design, we assume equal rerandomization in stage 2 for all patients except for the non-responders to high dose in stage 1. In our proposed design, questions might arise whether it is ethical to allow the low-dose responders to receive a higher dose level in stage 2. While toxicity may be a concern, a higher dose may result in increased efficacy; for patients with severe and fatal rare diseases such as certain types of cancers, the potential improvements in mitigation or curative effects from taking a higher dose level may outweigh the risks of the accompanied potential toxicity. However, in practice, researchers might consider reasonable unequal rerandomization adapting to response measured at the end of stage 1. For example, low dose responders may continue to receive low dose while all low dose non-responders could be assigned to high dose. When using a dichotomous variable Z to restrict rerandomization for any group, we suggest this Z variable is a quickly accessible response variable measured at the end of stage 1. A washout period can be included in our design, similar to a standard crossover design if a large carryover effect is expected. Our method can allow for a carryover effect by assuming varied α parameters among different stage 1 treatments in the Bayesian model. However, the model and estimation are most appropriate when little to no carryover effect is expected.

Similarly, when analyzing the data from this design, we should note some additional considera-

Scenario 1	Parameter	Bias		rMSE		CR		Width	
		JS	FS	JS	FS	JS	FS	JS	FS
$\alpha_P = \alpha_L = \alpha_H = 0$	μ_P	0.05	0.09	5.17	5.31	0.96	0.96	20.70	21.53
$\beta_P = \beta_L = \beta_H = 0.3$	μ_L	-0.02	0.01	3.67	5.16	0.97	0.96	15.37	21.52
	μ_H	0.21	0.16	3.63	5.22	0.96	0.96	15.11	21.53
	$\mu_L - \mu_P$	-0.07	-0.08	5.98	7.31	0.96	0.96	24.65	30.44
	$\mu_H - \mu_P$	0.16	0.07	5.99	7.29	0.96	0.95	24.56	30.45
$\alpha_P = 0, \alpha_L = -1, \alpha_H = -2$	μ_P	-0.10	0.09	5.17	5.31	0.96	0.96	20.72	21.53
$\beta_P = \beta_L = \beta_H = 0.3$	μ_L	-0.44	0.01	3.69	5.16	0.96	0.96	15.38	21.53
	μ_H	-0.19	0.16	3.65	5.22	0.96	0.96	15.13	21.53
	$\mu_L - \mu_P$	-0.34	-0.08	5.99	7.32	0.96	0.96	24.66	30.44
	$\mu_H - \mu_P$	-0.10	0.07	6.00	7.29	0.96	0.95	24.59	30.45
$\alpha_P = \alpha_L = \alpha_H = 0$	μ_P	0.05	0.09	5.62	0.93	0.93	0.96	20.42	21.52
$\beta_P = 0.01, \beta_L = 0.5, \beta_H = 1$	μ_L	0.56	0.01	3.90	0.97	0.97	0.96	16.22	21.53
	μ_H	-0.34	0.16	3.98	0.96	0.96	0.96	16.03	21.52
	$\mu_L - \mu_P$	0.51	-0.08	6.65	0.93	0.93	0.96	24.00	30.45
	$\mu_H - \mu_P$	-0.39	0.07	6.67	0.92	0.92	0.95	23.98	30.44
$\alpha_P = 0, \alpha_L = -1, \alpha_H = -2$	μ_P	-0.13	0.09	5.62	5.31	0.93	0.96	20.44	21.52
$\beta_P = 0.01, \beta_L = 0.5, \beta_H = 1$	μ_L	0.21	0.01	3.86	5.16	0.97	0.96	16.22	21.53
	μ_H	-0.61	0.16	4.03	5.22	0.95	0.96	16.06	21.52
	$\mu_L - \mu_P$	0.34	-0.08	6.63	7.31	0.92	0.96	24.00	30.45
	$\mu_H - \mu_P$	-0.48	0.07	6.69	7.29	0.92	0.95	24.01	30.44

Table 3.8: Bias, root mean-square error (rMSE), coverage rates (CR) and the width of 95% credible intervals for the posterior mean treatment effects under Scenario 1 (see Section 3.3.2) with varying α or β by stage 1 treatment arm, using the optimistic prior setting. μ_P, μ_L and μ_H are the mean outcome of having placebo, low dose and high dose in stage 1 respectively. JS=joint stage model, FS=first stage model. $\alpha_P, \alpha_L, \alpha_H$ are the shift parameter for patients receiving placebo, low dose and high dose in stage 1. Similarly, $\beta_P, \beta_L, \beta_H$ are the respetive linkage parameter for each treatment arm.

tions. When building the joint stage Bayesian model, we assume the same conditional variance for the observations from the two stages. In reality, this "equal variance" assumption can be relaxed as long as there is evidence to support other assumptions. In choosing the prior distributions for the parameters, our case simulations mainly focus on how to deal with the misspecification of the mean outcome parameters. In addition, for the shift parameter α, we always assume its mean to be around 0, implying little carryover effect nor stage effect exists. The prior variance of the α reflects our belief in the possible range of treatment effect shiftings in stage 2. Since β has the same sign as the correlation coefficient within a patient's observations, it is very likely to be positive. Having the prior distribution of β centered around 0 is to account for the situation when extremely weak correlation exists between observations across two stages. Moreover, although α and β might vary among different stage 1 treatments, our simulations show that our model was still quite robust to such circumstances.

One limitation of our Bayesian model is the potential model misspecification due to little prior

information. For example, in Section 3.3.3, we detect non-negligible bias when using the optimistic priors (implying belief in strong treatment effects) to analyze data generated under Scenario 1 where none of the three treatments were efficacious. As an alternative, we apply the mixture prior setting in which we assume equal weights for the optimistic and non-informative element distributions to analyze the same data. We found that the estimates using the mixture priors have smaller bias and better coverage under Scenario 1 than the estimates using the optimistic priors. Under the other two scenarios, the mixture priors exhibit little bias while sacrificing a small amount of efficiency, as compared to the optimistic priors. Thus, we believe the utilization of a flexible type of prior distribution can help accommodate the lack or misspecification of prior information. Another limitation of our design is the potential loss to follow-up for some patients due to the longer duration of snSMART. As opposed to the benchmark one-stage parallel arm design, snSMARTs allow patients to receive more than one dose level or treatment throughout their follow-up and collect more observations, all of which require an extended study period. Potential compliance issues are inevitable in such studies. Nevertheless, we believe that restricting placebo to be a treatment option only for stage 1 may potentially improve patients' retention and compliance and thus may offset the loss due to the extension of the follow-up period. As a final limitation, we note the risk of very high or very low observed response rates to stage 1 treatments. For example, the stage 1 response rates are extremely low (less than 5%) for all patients under Scenario 1 where all three treatments have equal low mean outcome. Since the response rate only impacts the rerandomization of those patients who receive high dose in stage 1, only a small number of patients are impacted. Thus, an unequal allocation does not make a great impact to the analysis nor the decision making since our major objective in this study is to examine the efficacy of each dose level instead of identifying the best treatment regimen. However, if randomization is more restricted in the second stage (i.e. second stage treatment depends on response for more than those initially assigned to high dose), bias and efficiency may be detrimentally affected.

Our future work includes exploring methods that estimate and compare the mean outcome of the embedded treatment regimens in the proposed two-stage design and constructing a way to calculate the sample size required for this snSMART design with regards to our proposed model.

CHAPTER 4

Bayesian Sample Size Determination Methods for Continuous Outcomes in a Small N, Sequential, Multiple Assignment, Randomized Trial (snSMART) Comparing Two Dose Levels with Placebo

4.1 Introduction

At the experimental design stage, determining sample size is very important, and it is a function of multiple factors including the study population size, effect size of interest, and available resources. Patient recruitment to a placebo-controlled randomized clinical trial (RCT) is especially challenging in rare disease clinical trials as the number of eligible and willing patients is limited. Thus, novel clinical trial designs that reduce the required sample size while efficiently estimating treatment effects are in great demand. The small n, sequential, multiple assignment, randomized trial (snSMART) (Tamura et al., 2016) is such a design that provides efficient treatment effect estimates in small samples.

Sample size determination (SSD) is the decision made about the number of human participants needed for an experiment. Frequentist SSD methods have existed for decades (Cohen, 1962), and most of the classical frequentist parametric methods follow power and size control (PSC) rules (Adcock, 1997) to determine the required sample size. Such methods involve the use of a test statistic from the observed data (i.e. t-statistic) to construct the corresponding power analysis. Given a hypothesis test, statistical power is defined as the probability of correctly rejecting the null hypothesis. Since the determination of sample size is made prior to the collection of data, investigators have to choose a single specific value of the effect size of interest and the variability of the effect based on previous knowledge or data.

Bayesian SSD approaches handle prior information about the parameters of interest differently than frequentist approaches. Bayesian approaches account for uncertainty in unknown parameters

through the use of prior distributions and determine the required sample size through a variety of performance criteria. Adcock (1988) first proposed the Average Coverage Criterion (ACC) rule, which determines sample size by controlling the average coverage rate of posterior credible intervals with fixed length. Joseph et al. (1995) developed the Average Length Criterion (ALC) method, choosing sample size by controlling the average length of posterior credible intervals with fixed coverage rate. Joseph and Belisle (1997) proposed the Worst Outcome Criterion (WOC) rule, which is a more conservative approach that controls both the expected coverage rate and the posterior interval length.

Methods have also been developed to account for hypothesis testing or model selection using the Bayes factor (Raftery et al., 1995; Weiss, 1997). In addition to these performance based approaches, Lindley (1997) argued that SSD is a decision problem and proposed a decision theoretic approach with utility theory. More recent work includes simulation-based Bayesian SSD approaches (Wang and Gelfand, 2002; Sahu and Smith, 2006; Fu et al., 2021), evidence-based approaches (De Santis, 2004), approaches using historical data (De Santis, 2007; Chen et al., 2011) and Bayesian-frequentist SSD approaches (Brutti et al., 2014). Simulation-based approaches are often implemented when using a complicated hierarchical Bayesian model since simulations do not require approximating the posterior distributions with explicit expressions. However, the computational burden caused by recursive simulations is considerably high despite increasing computational speed over the past few decades. When other explicit sample size calculation methods are available, these methods can be disseminated more broadly than simulation-based methods.

Most of the SSD methods we have described are concerned with the comparison of two population means with known sample variance or applications in linear regression. Applying these SSD methods to trials involving repeated measurements in multiple stages requires further consideration and derivation. In traditional 2x2 crossover designs, patients are randomized to receive a sequence of two study treatments. Inference is based on the comparison of effects within each patient removing the intersubject variability, simplifying the sample size calculation (Chow et al., 2017). In cluster randomized controlled trials, researchers take advantage of the intra-cluster correlation (ICC) coefficient and update the sample size derived for independent data with the estimated variance inflation factor (VIF) and the average cluster size (Hemming et al., 2011). Researchers then construct frequentist or Bayesian SSD approaches based on these fundamental derivations (Wang et al., 2005; Zou et al., 2003).

In Chapter 3, we developed an snSMART design and a Bayesian analytical approach with the aim of extracting information from two stages, from which we also demonstrated gains in efficiency. In this chapter, we address SSD for an snSMART of placebo, low and high dose, applying the ACC to the difference between the treatment effect of low dose and placebo. First, we derive the posterior distribution of the difference between mean outcomes of low dose versus placebo.

We then investigate two alternative ways of calculating sample size for the proposed snSMART. The first approach is to directly apply the existing Bayesian SSD method to the posterior difference as the outcome. The second method is motivated by cluster randomized controlled trials. We calculate the sample size needed for the first stage, and adjust this sample size to account for the extra information contained in the second stage (i.e. a deflation factor as opposed to the variance inflation factor used in cluster randomized controlled trials).

In Section 4.2.1, we review the study design and statistical model of interest. In Section 4.2.2 - 4.2.4 we present the details of the two candidate SSD approaches. In Section 4.3, we conduct simulations to demonstrate the performance of our calculated sample size in an snSMART that analyzes data using our proposed joint stage Bayesian model.

4.2 Methods

4.2.1 A Revisit of the Study Design and the Bayesian Model

We consider the snSMART design proposed in Chapter 3 (Figure 4.1). A total of N patients are enrolled in the trial. In stage 1, $N/3$ patients are randomly assigned to each of placebo, low dose and high dose of the study treatment, respectively. At the end of stage 1, a binary endpoint is measured for each participant, which determines the reallocation of treatment in stage 2. Patients receiving high dose in stage 1 who do not respond will continue to receive high dose in stage 2. All other participants are randomized to either low dose or high dose of the study drug in stage 2 with equal probability. The continuous endpoint of primary interest is collected at the end of both stages.

To estimate the first stage treatment effects, we adopt the Bayesian joint stage model proposed in 3. For patient $i = 1, \ldots, N$ in stage $j = 1, 2$, we denote T_{ij} as the treatment assigned and denote Y_{ij} as the observed continuous outcome. $Z_i = 0, 1$ is the binary intermediate endpoint measured at the end of stage 1 that is used for re-randomization. The conditional model for stage 1 data is written as $Y_{i1} \sim N(m_1(T_{i1}), \sigma^2)$, in which $m_1(T_{i1}) = \sum_k \mu_k I(T_{i1} = k)$, for the three treatment arms $k \in \{P, L, H\}$. In stage 2, we adopt the conditional model $Y_{i2} \sim N(m_2(T_{i1}, T_{i2}, Y_{i1}), \sigma^2)$, in which $m_2(T_{i1}, T_{i2}, Y_{i1}) = \mu_{k'} I(T_{i2} = k') + \alpha + \beta(Y_{i1} - m_1(T_{i1}))$ and $k' \in L, H$.

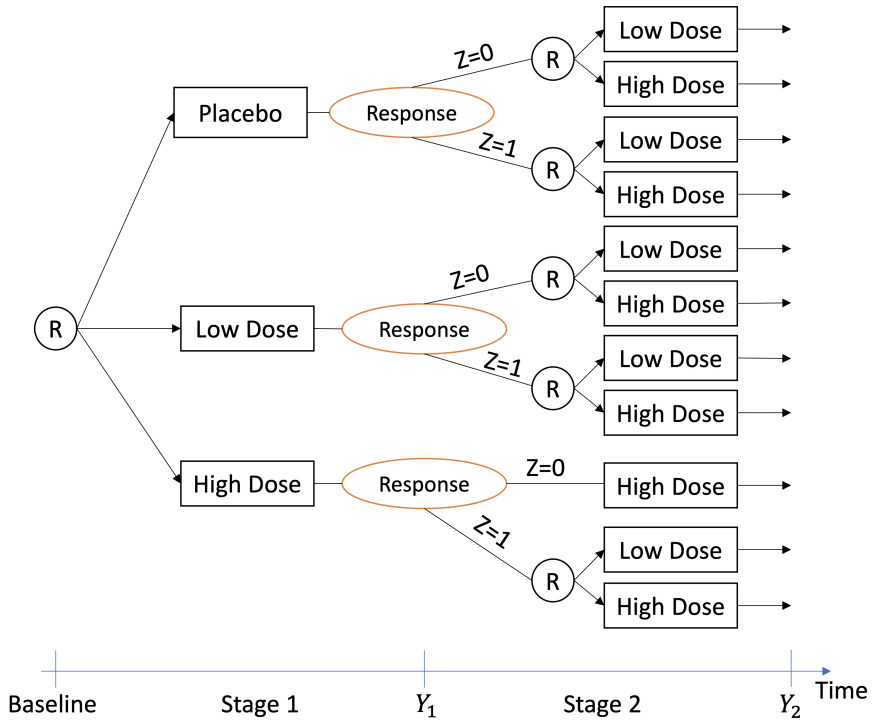

Figure 4.1: Study design of the proposed snSMART with a continuous endpoint. Participants are equally randomized (R) to one of the first stage treatment arms, placebo, low dose or high dose. At the end of stage 1, patients are re-randomized to their second stage treatment based on their response status. Outcomes are collected at the end of stage 1 and stage 2.

4.2.2 Derivation of the Posterior Variance-Covariance Matrix

Based on the Bayesian joint stage model construction, the observed outcomes from the two stages are assumed to jointly follow a bivariate normal distribution as:

$$Y_{i1}, Y_{i2}|_{T_{i1}=k, T_{i2}=k', \mu_k, \mu'_k, \alpha, \beta} \overset{i.i.d}{\sim} BVN\left(\begin{pmatrix} \mu_k \\ \mu_{k'} + \alpha \end{pmatrix}, \begin{pmatrix} \sigma^2 & \beta\sigma^2 \\ \beta\sigma^2 & (\beta^2+1)\sigma^2 \end{pmatrix} \right).$$

Suppose we denote the N observed outcomes at the end of stage $j = 1, 2$ as $\boldsymbol{Y_j}$, i.e., $\boldsymbol{Y_1} = (Y_{11}, Y_{21}, ..., Y_{N1})^T$ and $\boldsymbol{Y_2} = (Y_{12}, Y_{22}, ..., Y_{N2})^T$. Similarly, we denote treatment assignments in stage $j = 1, 2$ as T_j, i.e., $\boldsymbol{T_1} = (T_{11}, T_{21}, ..., T_{N1})^T$, $\boldsymbol{T_2} = (T_{12}, T_{22}, ..., T_{N2})^T$. To help demonstrate the derivation, we define $\boldsymbol{X_1} = (I(\boldsymbol{T_1} = P), I(\boldsymbol{T_1} = L), I(\boldsymbol{T_1} = H), \boldsymbol{0})$ and $\boldsymbol{X_2} = (I(\boldsymbol{T_2} = P), I(\boldsymbol{T_2} = L), I(\boldsymbol{T_2} = H), \boldsymbol{1})$, where $I(\boldsymbol{T_j} = k)$ are indicators of receiving treatment k at stage j, $\boldsymbol{0}$ and $\boldsymbol{1}$ are single-value vectors of length N (total number of participants). Denoting the parameters relevant to the expected values of outcomes as $\boldsymbol{\mu} = (\mu_P, \mu_L, \mu_H, \alpha)^T$, the likelihood of the observed data is specified as:

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y_1} \\ \boldsymbol{Y_2} \end{pmatrix} \sim MVN\left(\begin{pmatrix} \boldsymbol{X_1\mu} \\ \boldsymbol{X_2\mu} \end{pmatrix}, \boldsymbol{\Sigma} \right),$$

where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & \dots & 0 & \beta\sigma^2 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \sigma^2 & 0 & \dots & \beta\sigma^2 \\ \beta\sigma^2 & \dots & 0 & (1+\beta^2)\sigma^2 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \beta\sigma^2 & 0 & \dots & (1+\beta^2)\sigma^2 \end{pmatrix}$$

, with dimensions $2N \times 2N$. We partition the inverse matrix of $\boldsymbol{\Sigma}$ by four $N \times N$ square block matrices, i.e., $\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma_{11}} & \boldsymbol{\Sigma_{12}} \\ \boldsymbol{\Sigma_{21}} & \boldsymbol{\Sigma_{22}} \end{pmatrix}$.

Suppose we fit the model with independent normal prior distributions on unknown parameters as $\boldsymbol{\mu} \sim MVN(\boldsymbol{\mu_0}, \boldsymbol{\Sigma_0})$, where

$$\boldsymbol{\mu_0} = \begin{pmatrix} \mu_0^P \\ \mu_0^L \\ \mu_0^H \\ \mu_0^\alpha \end{pmatrix}, \text{and } \boldsymbol{\Sigma_0} = \begin{pmatrix} \tau_P^2 & 0 & 0 & 0 \\ 0 & \tau_L^2 & 0 & 0 \\ 0 & 0 & \tau_H^2 & 0 \\ 0 & 0 & 0 & \tau_\alpha^2 \end{pmatrix}$$

with prior effective sample size of the mean outcome parameters as $\kappa_k = \sigma^2/\tau_k^2$ for treatment $k \in P, L, H$ and $\kappa_\alpha = \sigma^2/\tau_\alpha^2$ for the shift parameter α. The posterior distribution of μ assuming β

and σ are known is

$$\begin{aligned} p(\boldsymbol{\mu}|\boldsymbol{Y}, \boldsymbol{\Sigma}) &\propto \exp\{-\frac{1}{2}[\begin{pmatrix} Y_1 - X_1\mu \\ Y_2 - X_2\mu \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} Y_1 - X_1\mu \\ Y_2 - X_2\mu \end{pmatrix} + (\mu - \mu_0)^T\Sigma_0^{-1}(\mu - \mu_0)]\} \\ &\propto \exp\{-\frac{1}{2}[(Y_1 - X_1\mu)^T\Sigma_{11}(Y_1 - X_1\mu) + (Y_2 - X_2\mu)^T\Sigma_{12}(Y_1 - X_1\mu) + \\ &\qquad (Y_1 - X_1\mu)^T\Sigma_{21}(Y_2 - X_2\mu) + (Y_2 - X_2\mu)^T\Sigma_{22}(Y_2 - X_2\mu)]\} \\ &\propto \exp\{-\frac{1}{2}[\mu^T(X_1^T\Sigma_{11}X_1 + X_2^T\Sigma_{12}X_1 + X_1^T\Sigma_{21}X_2 + X_2^T\Sigma_{22}X_2 + \Sigma_0^{-1})\mu + \\ &\qquad \mu^T(2X_1^T\Sigma_{11}Y_1 + X_2^T\Sigma_{12}Y_1 + X_1^T\Sigma_{12}Y_2 + X_1^T\Sigma_{21}Y_2 + X_2^T\Sigma_{21}Y_1 + \\ &\qquad 2X_2^T\Sigma_{22}Y_2 + 2\Sigma_0^{-1}\mu_0)]\}. \end{aligned} \tag{4.1}$$

Let $\boldsymbol{A} = X_1^T\Sigma_{11}X_1 + X_2^T\Sigma_{12}X_1 + X_1^T\Sigma_{21}X_2 + X_2^T\Sigma_{22}X_2 + \Sigma_0^{-1}$ and $\boldsymbol{B} = (1/2)(2X_1^T\Sigma_{11}Y_1 + X_2^T\Sigma_{12}Y_1 + X_1^T\Sigma_{12}Y_2 + X_1^T\Sigma_{21}Y_2 + X_2^T\Sigma_{21}Y_1 + 2X_2^T\Sigma_{22}Y_2 + 2\Sigma_0^{-1}\mu_0)$, then the posterior distribution can be written as

$$\begin{aligned} p(\boldsymbol{\mu}|\boldsymbol{Y}, \boldsymbol{\Sigma}) &\propto \exp\{-\frac{1}{2}(\mu^T A^{-1}\mu - 2\mu^T B)\} \\ &\propto \exp\{-\frac{1}{2}(\mu - A^{-1}B)^T A(\mu - A^{-1}B)\}. \end{aligned} \tag{4.2}$$

Thus, the posterior distribution of $\boldsymbol{\mu}$ is multivariate normal with posterior mean $\hat{\boldsymbol{\mu}} = \boldsymbol{A^{-1}B}$ and posterior variance $\boldsymbol{V} = \boldsymbol{A^{-1}}$. Specifically, given the observed data,

$$\hat{\boldsymbol{\mu}} = \frac{1}{2}V(2X_1^T\Sigma_{11}Y_1 + X_2^T\Sigma_{12}Y_1 + X_1^T\Sigma_{12}Y_2 + X_1^T\Sigma_{21}Y_2 + X_2^T\Sigma_{21}Y_1 + 2X_2^T\Sigma_{22}Y_2 + 2\Sigma_0^{-1}\mu_0)$$

$$= V \begin{pmatrix} \{\sum_{i:T_{i1}=P}[(1+\beta^2)Y_{i1} - \beta Y_{i2}] + \mu_0^P\kappa_P\}/\sigma^2 \\ \{\sum_{i:T_{i1}=L}[(1+\beta^2)Y_{i1} - \beta Y_{i2}] + \sum_{i:T_{i2}=L}[-\beta Y_{i1} + Y_{i2}] + \mu_0^L\kappa_L\}/\sigma^2 \\ \{\sum_{i:T_{i1}=H}[(1+\beta^2)Y_{i1} - \beta Y_{i2}] + \sum_{i:T_{i2}=H}[-\beta Y_{i1} + Y_{i2}] + \mu_0^H\kappa_H\}/\sigma^2 \\ \{\sum_{i=1}^N(-\beta Y_{i1} + Y_{i2}) + \mu_0^\alpha\kappa_\alpha\}/\sigma^2 \end{pmatrix}$$

and $\boldsymbol{V} = (X_1^T\Sigma_{11}X_1 + X_2^T\Sigma_{12}X_1 + X_1^T\Sigma_{21}X_2 + X_2^T\Sigma_{22}X_2 + \Sigma_0^{-1})^{-1}$

$$= \begin{pmatrix} \frac{(1+\beta^2)N_{P+}+\kappa_P}{\sigma^2} & \frac{-\beta N_{PL}}{\sigma^2} & \frac{-\beta N_{PH}}{\sigma^2} & \frac{-\beta N_{P+}}{\sigma^2} \\ \frac{-\beta N_{PL}}{\sigma^2} & \frac{(1+\beta^2)N_{L+}+N_{+L}-2\beta N_{LL}+\kappa_L}{\sigma^2} & \frac{-\beta(N_{LH}+N_{HL})}{\sigma^2} & \frac{N_{+L}-\beta N_{L+}}{\sigma^2} \\ \frac{-\beta N_{PH}}{\sigma^2} & \frac{-\beta(N_{LH}+N_{HL})}{\sigma^2} & \frac{(1+\beta^2)N_{H+}+N_{+H}-2\beta N_{HH}+\kappa_H}{\sigma^2} & \frac{N_{+H}-\beta N_{H+}}{\sigma^2} \\ \frac{-\beta N_{P+}}{\sigma^2} & \frac{N_{+L}-\beta N_{L+}}{\sigma^2} & \frac{N_{+H}-\beta N_{H+}}{\sigma^2} & \frac{N_{P+}+N_{L+}+N_{H+}+\kappa_\alpha}{\sigma^2} \end{pmatrix}^{-1}.$$

In the expression of $\boldsymbol{V}$, $N_{kk'} = \sum_{i=1}^{N} I(T_{i1} = k, T_{i2} = k')$ for $k \in \{P, L, H\}$ and $k' \in \{L, H\}$, and $N_{k+} = \sum_{i=1}^{N} I(T_{i1} = k)$, $N_{+k} = \sum_{i=1}^{N} I(T_{i2} = k)$. For instance, N_{PL} is the number of participants who receive placebo in stage 1 and are re-randomized to low dose in stage 2, and N_{+L} represents the total number of participants who receive low dose in stage 2. Before data collection, we approximate the expected number $N_{kk'}$ according to the prespecified randomization scheme and obtain the approximate posterior variance-covariance matrix of the mean outcome parameters $\boldsymbol{\mu}$. Given the multivariate normality of $\boldsymbol{\mu}$, the posterior distribution of any weighted linear combination of the three parameters is normal. We deploy this property for the sample size calculation approaches described in the following sections.

4.2.3 One-step Approach Based on the Joint Stage Posterior Distribution

Since the purpose of our design is to demonstrate the efficacy of dose levels against placebo, our hypotheses of interest are based on $\mu_L - \mu_P$ and $\mu_H - \mu_P$. For SSD, we consider the difference between the mean outcomes of low dose and placebo as it will provide a conservative sample size assuming the lower dose is not more efficacious than the higher dose.

Like in Wei et al. (2020), we follow the ACC rule (Adcock, 1988) to determine the required sample size to guarantee a desired "significance level" of ϕ and a desired "power" of ψ. We refer to "significance level" as one minus the coverage rate of the posterior credible interval of the outcome parameter of interest. We define "power" as the probability of rejecting the null from the $1 - \phi$ credible interval with the proposed Bayesian analysis when a prespecified treatment effect is present. For a fixed length l, the determined sample size is chosen as the smallest integer N that

ensures

$$\int_x \{\int_{z-l/2}^{z+l/2} f(\mu_L - \mu_P|x, N)d(\mu_L - \mu_P)\}f(x)dx \geq 1 - \phi,$$

where $x = (x_1, x_2)$ denotes the integrated data from two stages of a trial, where $x_1 = (X_1, Y_1)$ and $x_2 = (X_2, Y_2)$; z denotes the posterior mean of $\mu_L - \mu_P$. We choose length l as $\frac{2(\mu_L - \mu_P)}{1 - z_{1-\psi}/z_{1-\phi/2}}$, such that the sample size guarantees a power of ψ with the credible interval of length l in order to detect the treatment effect as specified in the prior mean of $\mu_L - \mu_P$.

4.2.4 Two-step Approach Based on an Adjustment Factor

In this approach, we focus on quantifying the separate benefits of (i) having two stages in an snSMART and (ii) using a Bayesian analysis as compared to a single stage RCT using a frequentist analysis. That is, we quantify sample size savings from the two-stage study design and contribution from the prior distributions used in analysis.

The first step of this approach is to calculate the sample size needed for a standard one-stage design that uses only the first stage of the trial. Given the proposed Baysian joint-stage model and a normal likelihood with normal conjugate priors for the mean of each dose level, the posterior distribution of μ_P and μ_L are derived as

$$\mu_P|x_1 \sim N(\frac{\sum_{i\in I(P)} Y_{i1} + \kappa_P \mu_0^P}{N/3 + \kappa_P}, \frac{\sigma^2}{N/3 + \kappa_P})$$

$$\mu_L|x_1 \sim N(\frac{\sum_{i\in I(L)} Y_{i1} + \kappa_L \mu_0^L}{N/3 + \kappa_L}, \frac{\sigma^2}{N/3 + \kappa_L}),$$

where x_1 denotes the data collected until the end of the first stage. Similarly, we apply the ACC rule for SSD.

The second step updates the sample size computed in the first step with an adjustment factor (AF). The AF assumes a fixed β and is defined as the expected ratio of the posterior variance derived from the joint stage model over the posterior variance derived from the first stage model. The AF is:

$$AF = \int_x \frac{Var(\mu_L - \mu_P|x)f(x)}{Var(\mu_L - \mu_P|x_1)f(x_1)}dx,$$

where the posterior variance of mean outcome difference from the stage 1 model is $Var(\mu_L - \mu_P|x_1) = \sigma^2/(N/3 + \kappa_P) + \sigma^2/(N/3 + \kappa_L)$, and its counterpart from the joint stage model $Var(\mu_L - \mu_P|x)$ is derived in Section 4.2.2 given a fixed β. Given the complexity of integration, we approximate the expected ratio of posterior variances by the average ratio of variances given a large number of random draws of the study data that follow the sample size calculation assump-

tions. Furthermore, in order to reduce the processing time, we approximate the AF using the ratio between the posterior variances based upon exemplary data from one and two stages that reflects the expected treatment allocation in two stages. This approximation by exemplery data is a similar idea as described in Shieh and O'Brien (1998). Through simulations, we demonstrate that the approximation through exemplary data provides an AF value very close to the one computed from approximating the expected ratio through simulations, and significantly reduces the computation time (results not shown). Then, the total sample size needed is the product of the AF and the number of patients needed to satisfy the ACC criterion using the posterior distribution from stage 1 data.

4.3 Simulations

In this section, we perform simulations regarding a hypothetical snSMART that compares low and high dose of a drug with placebo to illustrate the performance of our proposed SSD approaches.

4.3.1 Main Analysis: Scenarios 1-10

To select the sample size for a hypothetical snSMART trial, researchers need to specify parameter values related to features that best describe the study. In the proposed snSMART design, the first-stage treatment assignment is fixed, while the allocations of second-stage treatments are influenced by the response rate and the probability of participants being assigned to either low or high dose in stage 2. The variability in observed outcomes and the underlying correlation between observations from each stage are also critical factors that affect the precision in estimation. Finally, the mean outcome parameters μ_P, μ_L and μ_H determine the magnitude of the effect size, and thus, determine the desired sample size. Based on all these factors, we examine scenarios with varying inputs for the two proposed sample size calculations.

We first define a standard input setting (Scenario 1 in Table 4.1). In this scenario, the response rate to high dose in stage 1 is 0.6, meaning that 40% of the participants who received high dose in stage 1 continue to receive the high dose in stage 2. We do not set constraints based on the response rate to placebo or low dose, because all participants who receive placebo or low dose in stage 1 are rerandomized to low or high dose with equal probabilities, no matter how they respond to stage 1 treatment. We assume an ineffective placebo ($\mu_P = 0$), a moderately efficacious low dose ($\mu_L = 2$), and a high dose with greatest efficacy ($\mu_H = 4$). The standard deviation of the stage 1 outcome Y_{i1} is $\sigma = 4$. We set the true value of the shift parameter α to be 0, meaning that the treatment effects are not shifted in stage 2. We assume the linkage parameter β to be 0.5, indicating a moderate amount of within-subject correlation ($\rho(Y_{i1}, Y_{i2}) = 0.45$).

Scenarios	Sample Size Calculation Inputs								
	Re-randomization scheme		Within-subject Correlations		Outcome Parameters				
	π_Z	$R_L : R_H$	β	$\rho(Y_{i1}, Y_{i2})$	μ_P	μ_L	μ_H	α	σ
1	0.6	0.5:0.5	0.5	0.45	0	2	4	0	4
2	0.6	0.5:0.5	0.5	0.45	0	**1**	4	0	4
3	0.6	0.5:0.5	0.5	0.45	0	**3**	4	0	4
4	0.6	0.5:0.5	0.5	0.45	0	2	**3**	0	4
5	0.6	0.5:0.5	0.5	0.45	0	2	4	**-1**	4
6	0.6	0.5:0.5	0.5	0.45	0	2	4	0	**3**
7	**0.2**	0.5:0.5	0.5	0.45	0	2	4	0	4
8	0.6	**0.7:0.3 for responders 0.3:0.7 for non-responders**	0.5	0.45	0	2	4	0	4
9	0.6	0.5:0.5	**0**	**0**	0	2	4	0	4
10	0.6	0.5:0.5	**1**	**0.71**	0	2	4	0	4

Table 4.1: Presumed scenarios 1-10 for sample size calculations. π_Z represents the response rate of observing $Z = 1$ for patients on high dose in stage 1. $R_L : R_H$ represents the ratio of the probabilities for patients who are re-randomized to low and high dose in stage 2. β is the linkage parameter as defined in Section 4.2.1, whereas $\rho(Y_{i1}, Y_{i2}) = \beta/\sqrt{1+\beta^2}$ is the within-subject correlations between observations in two stages. For treatment $k \in P, L, H$, μ_k denotes the true value the mean outcome parameter. σ stands for the standard deviation of the observed outcomes. α is the true value of the shift parameter. Scenario 1 is the benchmark scenario. Any variation of the other settings from Scenario 1 is highlighted in bold.

Given Scenario 1 as a benchmark, we examined 9 additional scenarios as specified in Table 4.1 varying the inputs of key elements of the study, including effect sizes (Scenarios 2-4), value of the shift parameter α (Scenario 5), standard deviation of the observations σ (Scenario 6), randomization schemes (Scenarios 7-8), and values of the linkage parameter β which determines the within-subject correlation (Scenarios 9-10). All inputs match the true underlying parameters used to generate the simulated data.

Under each setting, we calculate the sample size needed to satisfy the pre-specified criteria of the chosen operating characteristics (coverage rate and power) using the two Bayesian approaches proposed in Fang et al. (2022). We then repeatedly simulate 6,000 data sets of that calculated sample size from each SSD approach. We analyze each data set with the statistical method described in Section 4.2.1. The estimated power is computed as the proportion of the $100(1-\phi)\%$ posterior HDI of low dose versus placebo not covering 0 among these Bayesian analyses among 6,000 repeats, which indicates that we expect the true underlying power to be within plus or minus 0.01 of the 95% confidence level of the simulated power. We quantify sample size savings by comparing 1) the sample size needed for a one-stage design versus the proposed two-stage snSMART design, and 2) the sample size calculated using a conventional frequentist method versus the proposed Bayesian methods. In all scenarios, we assume that the input parameter values are correctly specified and are equal to the prior means. The prior standard deviation is set to $\tau = 2$ for all mean

Scenarios	Parameters in Data generation					
	μ_P	μ_L	μ_H	α	σ	β
1	0	2	4	0	4	0.5
11	0	2	**3**	0	4	0.5
12	0	**1**	4	0	4	0.5
13	0	2	4	**-1**	4	0.5
14	0	2	4	0	**6**	0.5
15	0	2	4	0	4	**0.3**

Table 4.2: Presumed scenarios 1, 11-15 for sample size calculations. π_Z represents the response rate of observing $Z = 1$ for patients on high dose in stage 1. $R_L : R_H$ represents the ratio of the probabilities for patients who are re-randomized to low and high dose in stage 2. β is the linkage parameter as defined in Section 4.2.1, whereas $\rho(Y_{i1}, Y_{i2}) = \beta/\sqrt{1+\beta^2}$ is the within-subject correlations between observations in two stages. For treatment $k \in P, L, H$, μ_k denotes the true value the mean outcome parameter. σ stands for the standard deviation of the observed outcomes. α is the true value of the shift parameter. Scenario 1 is the benchmark scenario. Any variation of the other settings from Scenario 1 is highlighted in bold.

outcome parameters. Therefore, the effective sample size of the prior distribution of each mean parameter is $\kappa_k = \sigma^2/\tau^2 = 4$ for $k = P, L, H$ (Morita et al., 2008), providing information as if from four extra patients besides the ones enrolled in the trial through the Bayesian analysis. The estimated power is computed with the simulated data of the required size for each scenario. The shift parameter α follows a normal prior distribution with standard deviation $\tau_\alpha = 2$. The prior distributions for σ and β are $Gamma(4, 1)$ and $N(0, 1)$, respectively. We conduct all simulations using R (version 4.02).

4.3.2 Sensitivity Analysis: Scenarios 11-15

As a sensitivity analysis, we apply the study inputs from Scenario 1 and further explore the performance of our SSD approaches. In Table 4.2, we define scenarios that have (1) misspecification in either the mean outcome parameter or the shift parameter α (Scenarios 11-13), (2) misspecified variance of the observed outcome (Scenario 14), and (3) misspecified linkage parameter β which determines the within-subject correlation (Scenario 15). In these scenarios, the design parameters used in the two SSD methods differ from the data generation parameters, and the size of each simulated data set aligns with the computed sample size from the SSD process. Similar to the main analysis, we analyze the generated data and evaluate the performance of our methods in terms of the simulated power.

Scenario	One-stage design		Two-stage snSMART design				
			One-step Approach		Two-step approach		
	N_{Freq}	N_{Bayes}	N_{B1}	Simulated power	AF	N_{B2}	Simulated power
1	50	46	31	0.814	0.693	32	0.823
2	197	194	133	0.804	0.687	134	0.814
3	22	18	12	0.830	0.707	13	0.838
4	50	46	31	0.812	0.693	32	0.824
5	50	46	31	0.814	0.693	32	0.824
6	28	26	18	0.808	0.694	19	0.820
7	50	46	31	0.813	0.690	32	0.824
8	50	46	31	0.816	0.695	32	0.829
9	50	46	38	0.806	0.847	39	0.809
10	50	46	20	0.804	0.459	22	0.850

Table 4.3: The estimated sample size that ensures the desired power at 0.8 and the coverage rate at 0.9 and corresponding estimated power to demonstrate low dose effectiveness is given. We consider Scenarios 1-10 as defined in Table 4.1 under two study designs. In a one-stage parallel design, N_{Freq} stands for the sample size calculated from the classical frequentest approach, and N_{Bayes} stands for the sample size computed using ACC approach as if there is only one-stage of data. Under the two-stage snSMART design, N_{B1} and N_{B2} are the sample sizes computed using the proposed ACC approach and two-step approach as in Section 4.2. AF is the adjustment factor used in the two-step approach. Simulated power is computed as the proportion of the 95% posterior HDI not covering 0 among the Bayesian analyses conducted on 6,000 simulated datasets compatible with each scenario.

4.4 Results

4.4.1 Main Analysis: Sample Size Calculation Inputs Match the Truth

Suppose we are interested in finding the minimum sample size needed to guarantee the desired power of 0.8 and the coverage rate of 0.9. Table 4.3 presents the calculated sample size using several approaches regarding the scenarios listed in Table 4.1. In Table 4.3, N_{freq} and N_{Bayes} are the sample sizes needed for a one-stage design computed using the classic frequentist formula and the Bayesian ACC approach, respectively. Considering a two-stage snSMART design instead, N_{B1} and N_{B2} are the sample sizes computed using the one-step and two-step approach (based on the adjustment factor AF) accordingly. In general, sample sizes using both approaches (N_{B1} and N_{B2}) are extremely close, while the two-stage approach requires one more subject in almost all scenarios. Both SSD methods give reasonable sample size estimations when assumptions of the calculation hold.

In Scenario 1, we assume a moderate amount of true treatment effect of $\mu_L - \mu_P = 2$ that correctly reflects our assumptions (prior mean), and assume that σ and β are fixed and known. In a one-stage design analyzed by the classical frequentist method, we need 50 patients in each initial treatment arm to achieve the prespecified power. A Bayesian model that uses only the data

from stage 1 reduces the required sample size by 4 patients per arm, which is the prior effective size of the prior distribution for each mean outcome parameter (μ_P, μ_L and μ_H). The two-stage snSMART design further reduces the required sample size from 46 to 31 or 32 ($\sim$65%-70% to the original size in a one-stage design). Both approaches proposed for the two-stage snSMART design yield a sufficient sample size to detect the difference with 80% power.

Scenarios 2-4 explore different magnitudes of responses, assuming the true responses are identical to the ones used for SSD. With a smaller presumed treatment effect in Scenario 2 ($\mu_L - \mu_P = 1$), we require a much larger sample size as compared to Scenario 1. The proportion of participants saved from the usage of a two-stage design still holds, such that there is substantial reduction in the number of participants required (from 190+ to 130+). Similarly, a larger presumed treatment effect as shown in Scenario 3 ($\mu_L - \mu_P = 4$) leads to a smaller sample size while exhibiting a similar level of reduction in sample size from the one-stage design. Since the sample size consideration is based on the difference in treatment effects between low dose and placebo, changes in μ_H and μ_P (Scenario 4) and the shift parameter α (Scenario 5) make no difference in the required sample size as long as all treatment effects are correctly specified. Lower variability in the observed outcomes (Scenario 6) makes the difference more easily detected and thus fewer participants are required.

Scenarios 7-8 adjust prespecified factors that are relevant to the treatment allocation in stage 2, including changing the expected intermediate response rate for patients receiving high dose in stage 1 from 60% to 20% and varying the re-randomization probabilities to low and high dose in stage 2 given the response, while holding all other factors the same as Scenario 1. There are minor changes in AF values but not enough to affect the sample size in this case. We observe sufficient estimated powers for both scenarios.

Scenarios 9-10 are designed to examine the impact of the correlation between observations from two stages on the power at extreme values. When $\beta = 0$ (Scenario 4), two within-subject observations are completely independent of each other. As compared to Scenario 1, more participants are needed in the trial to achieve the expected power, while still decreasing the required sample size to 85% of the size for a one-stage design. Assuming $\beta = 1$ induces a high correlation in the observations with correlation coefficient of 0.71. Given such a high correlation, we observe a substantial shrinkage in the required sample size, and more than half of the participants required in a one-stage design are no longer needed.

In addition, the proposed SSD approaches account for various amounts of prior information. For instance, we applied non-informative prior distributions for all mean outcome parameters by changing the prior standard deviation τ from 2 to 100. Naturally, the sample size required for a one-stage design analyzed through the Bayesian method is the same as the conventional frequentist sample size. With both SSD approaches, 34 participants are expected in a two-stage design, which shows that the reduction in the sample size is induced from the design with two stages. Moreover,

Scenario	One-stage design		Two-stage snSMART design				
			One-step Approach		Two-step approach		
	N_{Freq}	N_{Bayes}	N_{B1}	Simulated power	AF	N_{B2}	Simulated power
1	50	46	31	0.814	0.693	32	0.823
11	50	46	31	0.822	0.693	32	0.830
12	50	46	31	0.378	0.693	32	0.382
13	50	46	31	0.799	0.693	32	0.810
14	50	46	31	0.555	0.693	32	0.569
15	50	46	31	0.774	0.693	32	0.783

Table 4.4: The estimated sample size that ensures the desired power at 0.8 and the coverage rate at 0.9 and corresponding estimated power to demonstrate low dose effectiveness in one-stage parallel design setting and two-stage snSMART design. We consider Scenarios 1, 11-15 as defined in Table 4.2 under two study designs. In a one-stage parallel design, N_{Freq} stands for the sample size calculated from the classical frequentest approach, and N_{Bayes} stands for the sample size computed using ACC approach as if there is only one-stage data. Under the two-stage snSMART design, N_{B1} and N_{B2} are the sample size computed using the proposed ACC approach and two-step approach as in Section 4.2. AF is the adjustment factor used in the two-step approach. Simulated power is computed as the proportion of the 95% posterior HDI not covering 0 among the Bayesian analyses conducted on 6,000 simulated datasets compatible with each scenario.

we tried to place a concentrated mass on the prior distribution of $\alpha = 0$ with an extremely small prior standard deviation at 0.01. Such a prior distribution is shown to not only reduce the uncertainty in the estimation of α alone, but also impact the estimations of the other mean outcome parameters of interest. Fewer participants are required as compared to Scenario 1 to retain the desired power.

4.4.2 Sensitivity Analysis: Misspecifications

We also explored additional settings with varying prior distributions and parameter values used in generating the simulated data. Results can be found in Table 4.4.

Scenario 11 displays the case when the presumed value of μ_H used for SSD and prior mean is slightly higher than the underlying parameter value in data generation. The misspecification of μ_H not only induces biased estimation of μ_H but also impacts the estimation of $\mu_L - \mu_P$, and thus, the simulated power. In this scenario, the simulated power is slightly above 0.8.

Scenario 13 explores the case when α is overestimated in the prior distribution for analysis. Since the point estimation of α is not a deciding factor in sample size determination, the required sample size stays the same as Scenario 1. Given the calculated sample size, the estimated power decreased by a small amount, but the confidence interval of power still includes 80%.

Scenarios 12, 14 and 15 focus on the misspecification of a few key components that determine the effect size size, including the overestimation of μ_L, underestimation of σ, and overestimation

of β. All three scenarios show great loss in simulated power, which is expected.

4.5 Discussion

In this chapter, we proposed two SSD approaches tailored for the snSMARTs proposed for comparing two dose levels with placebo. We examined the accuracy of both approaches via simulations and found them to provide sufficient sample sizes that fulfill certain user-defined criteria.

Although the two SSD approaches both provide adequate sample sizes that meet specific requirements, they are distinct from each other in several ways. The one-step approach directly applies the ACC rule using the posterior distribution of the expected treatment difference based on the two-stage data. In contrast, the two-step approach updates the desired sample size using the ACC rule for a one-stage design and intuitively quantifies the sample size savings from adding a second stage through the adjustment factor, as smaller AF indicates more savings. Motivated by the variance inflation factor as defined in cluster randomized controlled trials, the adjustment factor for an snSMART sample size calculation quantifies the sample size savings of both the multi-stage, multi-treatment design, and the prior information deployed in the Bayesian analysis. In general, with the same inputs, the two-step approach is slightly more conservative than the one-step approach, as the rounding procedure for sample size is conducted twice. Moreover, the two-step approach does not necessarily require the input of the value of μ_H and α, which may obscure the power loss due to the biased estimation caused by model misspecification. In that sense, the one-step approach is likely to provide the minimum sample size required when models are correctly specified. The two-step approach, in contrast, relies less on the model assumptions and gives intuitive quantification of the contribution of the second-stage data and conservative estimations of the sample sizes.

Although sample size calculation is anything but new, efficient sample size calculations for Bayesian analysis of novel study designs are greatly needed. We proposed two tailored SSD methods and established an explicit posterior distribution for the snSMART design comparing two dose levels to placebo with a continuous outcome. A well-chosen closed-form posterior distribution allows us to apply the fundamental SSD criterion under Bayesian framework without additional approximations, which enhances the accuracy of the calculation, allowing for ease of use for investigators in future applications. The two SSD approaches proposed in Sections and 4.2.3-4.2.4 are available through an interactive applet: https://sqrfang.shinyapps.io/snSMART_dosevp_SSD/.

An alternative way of doing power analysis without using an explicit form of the posterior distribution within a complex Bayesian modeling framework is to utilize Markov chain Monte Carlo (MCMC) procedures and draw inference from the constructed posterior samples. However,

MCMC creates an extensive burden on the computational environment. In our study, we simulated data and used MCMC to evaluate the power under certain settings. Although performed on a high performance computing (HPC) cluster, the time cost for 6,000 simulations in each setting still took more than 6 hours for a sample size of around 30 per arm. In contrast, our SSD approaches utilize the explicit expression of the posterior distribution, thus saving the need for recursive simulations. Meanwhile, we avoided extensive grid search and minimized the number of repeated loops in our program, reducing the required computational time to calculate a sample size down to seconds, significantly increasing the computation efficiency as compared to simulation-based SSD approaches that depend on MCMC.

Furthermore, our approaches can be adapted to other multi-stage, multi-treatment designs with repeated measures. In particular, crossover design shares several properties with snSMART. For instance, both designs lead to a data structure with multiple stages (periods). Participants may receive more than one treatment in the study. Both studies apply statistical approaches that account for treatment effect, period effect and carryover effect. Potentially, a similar SSD approach will be helpful in crossover studies that are analyzed under a Bayesian setting.

We acknowledge a few limitations in applying our SSD approaches and offer suggestions to mitigate problems. First, the elicitation of parameter values used for SSD is not always clear. In Section 4.4, we show that misspecification can lead to some power loss. Identifying the sensitivity of the calculation to different magnitudes of effect size of low dose versus placebo and the within-subject correlation are of most importance since other parameters are shown to have limited impact on the sample size. To deal with the uncertainty in prior beliefs on the expected treatment response, researchers can consider applying more than one possible value of the desired treatment response and take the average sample size needed among all attempts. Similarly, investigators may struggle to specify the correlation in observations between two stages as it affects the variance of the estimations. We can assume a positive correlation to narrow down the range of the coefficient of correlation and pick a few values for the corresponding β in SSD. Investigators can assign reasonable weights to each guess of the parameter values and compute a weighted average of the desired sample size. Another typical concern in designing studies with a complex structure is participant retention. So far, we have shown that our approaches perform well under the ideal case where there are no intermediate dropouts. From the study design perspective, we believe the proposed snSMART design can improve participant engagement by avoiding repeated exposure to inactive interventions. However, patient dropouts from clinical trials are inevitable. Normally, one can obtain the adjusted sample size by inflating the required number of participants by one minus the expected dropout rate. Further investigation is required to identify other approaches to handle missingness in snSMARTs. As such, our future work includes investigating the extension to other multi-stage, multi-treatment clinical trial designs, and exploring approaches to handle

missing data.

CHAPTER 5

Summary and Future Work

Motivated by previous work discussing small n, sequential, multiple assignment, randomized trials (snSMARTs) with multiple active treatments (Wei et al., 2018), in this dissertation, we focused on Bayesian approaches for snSMARTs evaluating different dose levels of one drug with a placebo control. This work aims to address the need for clinical trials in rare diseases with limited treatment options and to provide drug efficiency evidence when compared to a placebo for drug approval.

Chapter 2 presents an snSMART design comparing two dose levels to placebo with a binary outcome. In contrast to Wei et al. (2018), which compares three active treatments, we tested a single drug with two dose levels compared to placebo. The purpose of this trial design is to estimate the first stage treatment effect. We adjusted the Bayesian Joint Stage Model (BJSM) proposed by Wei et al. (2018) by making the prior distributions for the parameters more flexible to capture a dose response. We conducted simulations and demonstrated the efficiency gain compared to the log-linear Poisson joint stage model (LPJSM), a Bayesian model that only uses the first stage data (BFSM), and a frequentist model using only the first stage data (FSMLE).

In Chapter 3, we extended the snSMART in Chapter 2 to incorporate a continuous outcome. We built a joint stage Bayesian model that accounts for the within-subject correlation using a linkage parameter and evaluated its performance via simulations. We compared our model to standard methods that use the first stage data and demonstrated its advantages in improving efficiency. We also assessed the influence of mixed prior distributions on the Bayesian model.

Based on the trial design and statistical model proposed in Chapter 3, we proposed two Bayesian sample size determination (SSD) approaches in Chapter 4 to enhance applications of snSMARTs in clinical practice. We derived a closed-form posterior distribution of the treatment effect parameters under the Bayesian framework. Given that, we adopted a general Bayesian SSD method proposed by Adcock (1988, 1997) and developed two customized Bayesian SSD approaches. The first approach directly used the average coverage criterion (ACC) rule to calculate the required sample size for a two-stage trial in one step. The second approach quantifies the contribution of sample savings from the two-stage design as opposed to a one-stage balanced design using an ad-

justment factor (AF). We demonstrated that the power resulting from the calculated sample sizes were guaranteed for both approaches.

Future work may consider extending the existing design that consists dose levels and placebo to involve interim analysis that informs adaptations. Possible adaptations include early termination of the study and claiming ineffectiveness, dropping one dose early, or dynamically adapting the treatment allocation rules. Such an extension can potentially add flexibility to the trial, benefiting patients and improving recruitment. In addition, we considered trials with a single primary outcome (either binary or continuous) in our work, however, for some diseases, researchers are interested in multiple outcomes that have comparable importance, possibly with different data types. More sophisticated approaches are required in that case to handle multiple outcomes in an snSMART.

In conclusion, this dissertation presents innovative Bayesian methods for snSMARTs in addressing the challenge of limited patients in rare disease studies. There is great potential in applying this type of trial design and Bayesian analysis across many rare diseases. We hope the snSMART design can lead to more approved drugs for patients with a rare disease and ultimately, improved patient outcomes.

BIBLIOGRAPHY

107th Congress (2002). Rare diseases act of 2002. *Pub L 107–280*.

Adcock, C. (1988). A Bayesian approach to calculating sample sizes. *Journal of the Royal Statistical Society: Series D (The Statistician) 37*(4-5), 433–439.

Adcock, C. (1997). Sample size determination: a review. *Journal of the Royal Statistical Society: Series D (The Statistician) 46*(2), 261–283.

Almirall, D., S. N. Compton, M. A. Rynn, J. T. Walkup, and S. A. Murphy (2012). SMARTer discontinuation trial designs for developing an adaptive treatment strategy. *Journal of Child and Adolescent Psychopharmacology 22*(5), 364–374.

Bell, S. and C. Smith (2014). A comparison of interventional clinical trials in rare versus non-rare diseases: an analysis of ClinicalTrials.gov. *Orphanet journal of rare diseases 9*(1), 1–11.

Bhandari, M., H. Lochner, and P. Tornetta (2002). Effect of continuous versus dichotomous outcome variables on study power when sample sizes of orthopaedic randomized trials are small. *Archives of orthopaedic and trauma surgery 122*(2), 96–98.

Brutti, P., F. De Santis, and S. Gubbiotti (2014). Bayesian-frequentist sample size determination: a game of two priors. *Metron 72*(2), 133–151.

Chen, M.-H., J. G. Ibrahim, P. Lam, A. Yu, and Y. Zhang (2011). Bayesian design of noninferiority trials for medical devices using historical data. *Biometrics 67*(3), 1163–1170.

Chow, S.-C., J. Shao, H. Wang, and Y. Lokhnygina (2017). *Sample size calculations in clinical research*. chapman and hall/CRC.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: a review. *The Journal of Abnormal and Social Psychology 65*(3), 145.

De Santis, F. (2004). Statistical evidence and sample size determination for Bayesian hypothesis testing. *Journal of statistical planning and inference 124*(1), 121–144.

De Santis, F. (2007). Using historical data for Bayesian sample size determination. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 170*(1), 95–113.

Donner, A. and M. Eliasziw (1994). Statistical implications of the choice between a dichotomous or continuous trait in studies of interobserver agreement. *Biometrics*, 550–555.

Egidi, L., F. Pauli, and N. Torelli (2021). Avoiding prior–data conflict in regression models via mixture priors. *Canadian Journal of Statistics 50*(2), 491–510.

Fang, F., K. Hochstedler, R. Tamura, T. Braun, and K. Kidwell (2021). Bayesian methods to compare dose levels with placebo in a small n, sequential, multiple assignment, randomized trial. *Statistics in Medicine 40*(4), 963–977.

Fang, F., R. Tamura, T. M. Braun, and K. M. Kidwell (2022). Comparing dose levels to placebo using a continuous outcome in a small n, sequential, multiple assignment, randomized trial (snSMART). *Statistics in Biopharmaceutical Research*, 1–8.

Fu, Q., H. Hoijtink, and M. Moerbeek (2021). Sample-size determination for the Bayesian t test and Welch's test using the approximate adjusted fractional Bayes factor. *Behavior Research Methods 53*(1), 139–152.

Grimes, D. and K. Schulz (2002). An overview of clinical research: the lay of the land. *The lancet 359*(9300), 57–61.

Gupta, S., M. E. Faughnan, G. A. Tomlinson, and A. M. Bayoumi (2011). A framework for applying unfamiliar trial designs in studies of rare diseases. *Journal of Clinical Epidemiology 64*(10), 1085–1094.

Hemming, K., A. J. Girling, A. J. Sitch, J. Marsh, and R. J. Lilford (2011). Sample size calculations for cluster randomised controlled trials with a fixed number of clusters. *BMC medical research methodology 11*(1), 1–11.

Herder, M. (2017). What is the purpose of the orphan drug act? *PLoS medicine 14*(1), e1002191.

Honkanen, V. E. A., A. F. Siegel, J. P. Szalai, V. Berger, B. M. Feldman, and J. N. Siegel (2001). A three-stage clinical trial design for rare disorders. *Statistics in Medicine 20*(20), 3009–3021.

Joseph, L. and P. Belisle (1997). Bayesian sample size determination for normal means and differences between normal means. *Journal of the Royal Statistical Society: Series D (The Statistician) 46*(2), 209–226.

Joseph, L., D. B. Wolfson, and R. D. Berger (1995). Sample size calculations for binomial proportions via highest posterior density intervals. *Journal of the Royal Statistical Society: Series D (The Statistician) 44*(2), 143–154.

Lavori, P. W. and R. Dawson (2000). A design for testing clinical strategies: biased adaptive within-subject randomization. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 163*(1), 29–38.

Lindley, D. V. (1997). The choice of sample size. *Journal of the Royal Statistical Society: Series D (The Statistician) 46*(2), 129–138.

Makubate, B. and S. Senn (2010). Planning and analysis of cross-over trials in infertility. *Statistics in Medicine 29*(30), 3203–3210.

McCormack, F., Y. Inoue, J. Moss, L. Singer, C. Strange, K. Nakata, A. Barker, J. Chapman, M. Brantly, J. Stocks, et al. (2011). Efficacy and safety of sirolimus in lymphangioleiomyomatosis. *New England Journal of Medicine 364*(17), 1595–1606.

Morita, S., P. F. Thall, and P. Müller (2008). Determining the effective sample size of a parametric prior. *Biometrics 64*(2), 595–602.

Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*(2), 331–355.

Murphy, S. A. (2005). An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine 24*(10), 1455–1481.

Nason, M. and D. Follmann (1/9/2010). Design and analysis of crossover trials for absorbing binary endpoints. *Biometrics 66*(3), 958–965.

NIH (2023). FAQs About Rare Diseases — Genetic and Rare Diseases Information Center (GARD) – an NCATS Program.

NORD (2021). Orphan drugs in the united states: An examination of patents and orphan drug exclusivity. Technical report, National Organization for Rare Disorders.

Raftery, A. E., W. Gilks, S. Richardson, and D. Spiegelhalter (1995). *Hypothesis testing and model*. Chapman & Hall London.

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling 7*(9-12), 1393–1512.

Rosner, G. L., W. Stadler, and M. J. Ratain (2002). Randomized discontinuation design: application to cytostatic antineoplastic agents. *Journal of Clinical Oncology 20*(22), 4478–4484.

Sahu, S. and T. Smith (2006). A Bayesian method of sample size determination with practical applications. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 169*(2), 235–253.

Schieppati, A., J.-I. Henter, E. Daina, and A. Aperia (2008). Why rare diseases are an important medical and social issue. *The Lancet 371*(9629), 2039–2041.

Shieh, G. and R. O'Brien (1998). A simpler method to compute power for likelihood ratio tests in generalized linear models. In *Annual Joint Statistical Meetings of the American Statistical Association*.

Tamura, R. N., J. P. Krischer, C. Pagnoux, R. Micheletti, P. C. Grayson, Y.-F. Chen, and P. A. Merkel (2016). A small n sequential multiple assignment randomized trial design for use in rare disease research. *Contemporary Clinical Trials 46*, 48–51.

Taveira-DaSilva, A. M., M. P. Stylianou, C. J. Hedin, O. Hathaway, and J. Moss (2004). Decline in lung function in patients with lymphangioleiomyomatosis treated with or without progesterone. *Chest 126*(6), 1867–1874.

Wang, F. and A. E. Gelfand (2002). A simulation-based approach to bayesian sample size determination for performance under a given model and for separating models. *Statistical Science 17*(2), 193–208.

Wang, H., S.-C. Chow, and M. Chen (2005). A Bayesian approach on sample size calculation for comparing means. *Journal of Biopharmaceutical Statistics 15*(5), 799–807.

Wei, B., T. M. Braun, R. N. Tamura, and K. Kidwell (2020). Sample size determination for Bayesian analysis of small n sequential, multiple assignment, randomized trials (snSMARTs) with three agents. *Journal of Biopharmaceutical Statistics 30*(6), 1109–1120.

Wei, B., T. M. Braun, R. N. Tamura, and K. M. Kidwell (2018). A Bayesian analysis of small n sequential multiple assignment randomized trials (snSMARTs). *Statistics in Medicine 37*(26), 3723–3732.

Weiss, R. (1997). Bayesian sample size calculations for hypothesis testing. *Journal of the Royal Statistical Society: Series D (The Statistician) 46*(2), 185–191.

Williamson, T., M. Eliasziw, and G. H. Fick (2013). Log-binomial models: exploring failed convergence. *Emerging Themes in Epidemiology 10*(1), 14.

Zou, K. H., F. S. Resnic, A. S. Gogate, S. Ondategui-Parra, and L. Ohno-Machado (2003). Efficient Bayesian sample size calculation for designing a clinical trial with multi-cluster outcome data. *Biometrical Journal: Journal of Mathematical Methods in Biosciences 45*(7), 826–836.