Pharmaceutical Statistics

WILEY

MAIN PAPER

Bayesian Sample Size Calculation in Small n, Sequential Multiple Assignment Randomized Trials (snSMART)

Fang Fang[1] | Roy N. Tamura[2] | Thomas M. Braun[1] | Kelley M. Kidwell[1]

[1]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA | [2]Health Informatics Institute, University of South Florida, Tampa, Florida, USA

Correspondence: Kelley M. Kidwell (kidwell@umich.edu)

Received: 13 September 2023 | **Revised:** 31 August 2024 | **Accepted:** 9 December 2024

Funding: This work was supported by U.S. Food and Drug Administration, 75F40120C00195.

Keywords: Bayesian statistics | clinical trial | rare disease | repeated measures

ABSTRACT

A recent study design for clinical trials with small sample sizes is the small n, sequential, multiple assignment, randomized trial (snSMART). An snSMART design has been previously proposed to compare the efficacy of two dose levels versus placebo. In such a trial, participants are initially randomized to receive either low dose, high dose or placebo in stage 1. In stage 2, participants are re-randomized to either dose level depending on their initial treatment and a dichotomous response. A Bayesian analytic approach borrowing information from both stages was proposed and shown to improve the efficiency of estimation. In this paper, we propose two sample size determination (SSD) methods for the proposed snSMART comparing two dose levels with placebo. Both methods adopt the average coverage criterion (ACC) approach. In the first approach, the sample size is calculated in one step, taking advantage of the explicit posterior variance of the treatment effect. In the other two step approach, we update the sample size needed for a single-stage parallel design with a proposed adjustment factor (AF). Through simulations, we demonstrate that the required sample sizes calculated using the two SSD approaches both provide the desired power. We also provide an applet to allow for convenient and fast sample size calculation in this snSMART setting.

1 | Introduction

Patient recruitment to a placebo-controlled randomized clinical trial (RCT) is especially challenging in rare disease clinical trials because of the limited number of eligible study participants. 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) [1] is such a design that provides efficient treatment effect estimates in small samples. As a variation of a sequential, multiple assignment, randomized trial (SMART) design [2], an snSMART design is a multi-stage clinical trial design where there are at least two randomizations in sequence, often with subsequent randomization depending on response to initial treatment. The goal of an snSMART is to obtain more information on a small group of individuals through multiple stages to identify the best first stage intervention by incorporating information from multiple stages. The goal of a SMART design is to identify the most effective dynamic treatment regimen or tailored sequence of treatment across both stages. Hence, SMART designs are typically placebo-free and do not seek to provide evidence of a (dose of) drug relative to placebo. Fang et al. [3] developed an snSMART design and a Bayesian analytical approach with the aim of extracting information from two stages, from which they also demonstrated gains in efficiency of treatment effect estimates as compared to one-stage designs with placebo control.

The authors gratefully acknowledge the support from FDA's Advancing Regulatory Science Broad Agency Announcement Contract #75F40120C00195.

© 2025 John Wiley & Sons Ltd.

Sample size determination (SSD) is a decision feature that seeks to determine the number of human participants needed for an experiment. Frequentist SSD methods have existed for decades [4], and most of the classical frequentist parametric methods are based on power and size control (PSC) rules [5] to determine the required sample size. 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 [6] 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, Wolfson, and Berger [7] 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 [8] 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 [9, 10]. In addition to these performance based approaches, Lindley [11] 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 [12–14], evidence-based approaches [15], approaches using historical data [16, 17], and Bayesian-frequentist SSD approaches [18]. 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 required for repetitive simulations is considerably high despite increasing computational speed over the past few decades. When alternative explicit sample size calculation methods are available, these methods can be used more extensively than simulation-based methods. In the simulation-based approach by Wang and Gelfand (2002) [12], a framework with two distinct prior distributions was proposed: one for fitting in Bayesian analysis and another for sampling to generate data. This two-prior framework has also been adopted in other Bayesian SSD approaches [13, 16]. In this article, we employ a hybrid Bayes–frequentist approach, utilizing a sampling distribution with fixed parameter values and a single fitting prior for Bayesian inference with several considerations: (1) justification complexity for the chosen sampling prior, (2) mitigation of potential bias introduced by the sampling prior, and (3) advocacy for wider adoption in clinical trial practices. While our approach is a Bayesian conditional approach to SSD, it is important to note that many investigators are still navigating the incorporation of Bayesian concepts into conventional trial design. Hence, we aim for our method to help bridge this gap and facilitate a smoother integration of Bayesian approaches in practice.

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. Available SMART SSD methods [19] are not appropriate in this setting due to the very different objectives and models used for an snSMART. In traditional 2×2 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 [20]. 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 [21]. Researchers then construct frequentist or Bayesian SSD approaches based on these fundamental derivations [22, 23].

In this article, we address SSD for an snSMART of placebo, low and high doses, 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 in one step. 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 in two steps (i.e., we adjust by a deflation factor as opposed to the VIF used in cluster randomized controlled trials). The two-step approach explicitly quantifies the savings in sample size for a traditional one-stage trial due to the two-stage snSMART design and Bayesian analysis.

In Section 2.1, we review the study design and statistical model of interest. In Sections 2.2-2.3, we present the details of the two candidate SSD approaches. In Section 3, we conduct simulations to demonstrate the performance of our calculated sample size in an snSMART that analyzes data using the proposed joint stage Bayesian model from Fang et al. [3] The two SSD approaches proposed in Sections 2.2-2.3 are available through an interactive applet: https://umich-biostatistics.shinyapps.io/snSMART_doseversusplacebo_SSD/.

2 | Methods

2.1 | A Revisit of the Study Design and the Bayesian Model

We consider the snSMART design proposed in Fang et al. [3] (Figure 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

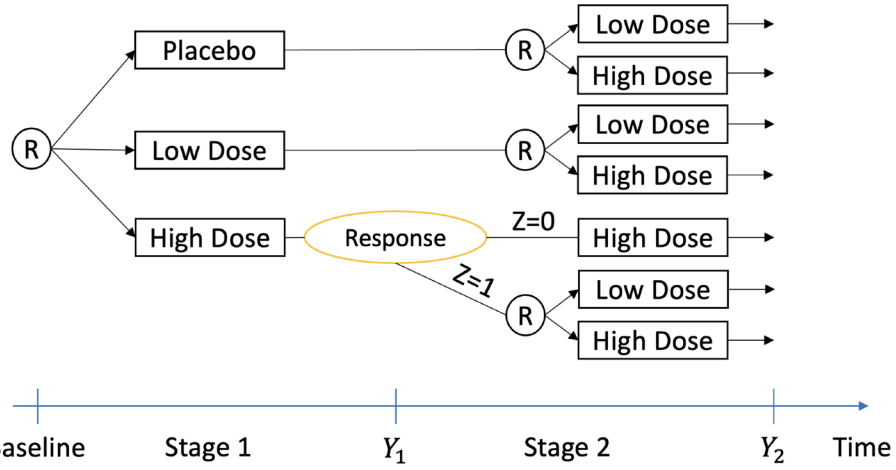


FIGURE 1 | Study design of the proposed snSMART. Participants are equally randomized (R) to one of the first stage treatment arms, placebo, low dose or high dose. At the end of stage 1, patients are re-randomized to their second stage treatment based on their response status. Outcomes are collected at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

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 Fang et al. [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 $Y_{i1} \sim N\left(m_1\left(T_{i1}\right), \sigma^2\right)$, in which $m_1\left(T_{i1}\right) = \sum_k \mu_k I\left(T_{i1} = k\right)$, for the three treatment arms $k \in \{P, L, H\}$. In stage 2, we adopt the conditional model $Y_{i2} \sim N\left(m_2\left(T_{i1}, T_{i2}, Y_{i1}\right), \sigma^2\right)$, in which $m_2\left(T_{i1}, T_{i2}, Y_{i1}\right) = \mu_{k'} I\left(T_{i2} = k'\right) + \alpha + \beta\left(Y_{i1} - m_1\left(T_{i1}\right)\right)$ and $k' \in \{L, H\}$.

Based on the Bayesian joint stage model construction, with a multivariate normal prior distribution of $\boldsymbol{\mu} = \left(\mu_P, \mu_L, \mu_H, \alpha\right)^T$, we obtain their posterior also following a multivariate normal distribution. See Appendix A for the derivation details.

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.

2.2 | 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. For the purpose of illustration, $\mu_L - \mu_P$ will be denoted as δ_L in the rest of this paper.

Like in Wei et al. [24], we follow the ACC rule [6] to determine the required sample size to guarantee a desired "significance level" of ϕ given 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

$$\int_{\boldsymbol{z}} \left\{ \int_{q-l/2}^{q+l/2} f\left(\delta_L | \boldsymbol{z}, N\right) d\delta_L \right\} f(\boldsymbol{z}) d\boldsymbol{z} \geq 1 - \phi,$$

where $\boldsymbol{z} = \left(\boldsymbol{z_1}, \boldsymbol{z_2}\right)$ denotes the integrated data from two stages of a trial; q denotes the posterior mean of δ_L. We choose length l as $\frac{2\delta_L}{1 - z_{1-\psi} / z_{1-\phi/2}}$, 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 δ_L.

2.3 | Two-Step Approach Based on an Adjustment Factor (AF)

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 traditional one-stage design that mimics the first stage of the snSMART. Given the proposed Baysian joint-stage

TABLE 1 | Presumed scenarios 1–10 for sample size calculations.

	Sample size calculation inputs								
	Re-randomization scheme	Within-subject correlations			Outcome parameters				
Scenarios	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**	0.5	0.45	0	2	4	0	4
		0.3:0.7 for non-responders							
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

Note: 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 2.1, where $\rho(Y_{i1}, Y_{i2}) = \beta / \sqrt{1+\beta^2}$ 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.

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

$$\mu_P \mid \boldsymbol{x_1}, \boldsymbol{y_1}, \kappa_P, \sigma^2 \sim N\left(\frac{\sum_{i \in I(P)} y_{i1} + \kappa_P \mu_0^P}{N/3 + \kappa_P}, \frac{\sigma^2}{N/3 + \kappa_P}\right)$$

$$\mu_L \mid \boldsymbol{x_1}, \boldsymbol{y_1}, \kappa_L, \sigma^2 \sim N\left(\frac{\sum_{i \in I(L)} y_{i1} + \kappa_L \mu_0^L}{N/3 + \kappa_L}, \frac{\sigma^2}{N/3 + \kappa_L}\right),$$

where $\boldsymbol{x_1}, \boldsymbol{y_1}$ 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 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 (ranging from 0 to 1) roughly quantifies the sample size savings from the two-stage design as compared to the conventional one-stage design. The AF is:

$$\mathrm{AF} = \int_{\boldsymbol{z}} \frac{\mathrm{Var}(\delta_L \mid \boldsymbol{z}) f(\boldsymbol{z})}{\mathrm{Var}(\delta_L \mid \boldsymbol{z_1}) f(\boldsymbol{z_1})} d\boldsymbol{z}$$

where the posterior variance of mean outcome difference from the stage 1 model is $\mathrm{Var}(\delta_L \mid \boldsymbol{z_1}) = \sigma^2 / (N/3 + \kappa_P) + \sigma^2 / (N/3 + \kappa_L)$, and its counterpart from the joint stage model $\mathrm{Var}(\delta_L \mid \boldsymbol{z})$ is derived in Section A given a fixed β. Given the complexity of integration,

TABLE 2 | Presumed scenarios 1, 11–15 for sample size calculations.

	Parameters in data generation					
Scenarios	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**

Note: 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 2.1, where $\rho(Y_{i1}, Y_{i2}) = \beta / \sqrt{1+\beta^2}$ 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.

we approximate the expected ratio of posterior variances by the average ratio of variances given a large number of random draws of hypothetical study data that follow the sample size calculation assumptions. Furthermore, in order to reduce the processing time, we approximate the AF using the ratio between the posterior variances computed given simulated exemplary data $\boldsymbol{z_0}$ that reflects the expected treatment allocation in two stages. This approximation by exemplary data is a similar idea as described in Shieh and O'Brien [25]. 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.

3 | Simulation Design

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. All of the scenarios considered are summarized in Tables 1 and 2 and described in more detail below.

3.1 | Main Analysis: Scenarios 1–10

To select the sample size for a hypothetical snSMART, researchers 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 SSD approaches.

We first define a standard input setting (Scenario 1 in Table 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 ($\rho(Y_{i1}, Y_{i2}) = 0.45$).

Given Scenario 1 as a benchmark, we examined 9 additional scenarios as specified in Table 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 SSD approaches. We then repeatedly simulate 6000 data sets of that calculated sample size from each SSD approach. We analyze each data set with the statistical method described in Section 2.1. The estimated power is computed as the proportion of the $100(1-\phi)\%$ posterior highest density interval (HDI) of low dose versus placebo not covering 0 among these Bayesian analyses among 6000 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 traditional one-stage design versus the proposed two-stage snSMART design, and (2) the sample size calculated using a conventional frequentist analytical 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 outcome parameters. Therefore, the effective sample size of the prior distribution of each mean parameter is $\kappa_k = \sigma^2 / \tau^2 = 4$ for $k = P, L, H$ [26], 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 scenarios. The shift parameter α follows a normal prior distribution with standard deviation $\tau_\alpha = 2$. The prior distributions for σ and β are $\text{Gamma}(4, 1)$ and $N(0, 1)$, respectively. We conduct all simulations using R (version 4.02).

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 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.

4 | Simulation Results

4.1 | Main Analysis: Scenarios 1–10

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 3 presents the calculated sample size using several approaches regarding the scenarios listed in Table 1. In Table 3, N_{freq} and N_{Bayes} are the sample sizes needed for a traditional 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 AF) accordingly. In general, sample sizes using both approaches (N_{B1} and N_{B2}) are extremely close, while the two-stage approach requires one

TABLE 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.

			Two-stage snSMART design				
	One-stage design		One-step approach		Two-step approach		
Scenario	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

Note: We consider Scenarios 1–10 as defined in Table 1 under two study designs. In a one-stage parallel design, N_{Freq} stands for the sample size calculated from the classical frequentist 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 one-step approach and two-step approach as in Fang et al. [3] 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 6000 simulated datasets compatible with each scenario.

more subject in almost all scenarios. Both SSD methods give reasonable sample size estimates when assumptions of the calculation hold.

In Scenario 1, we assume a moderate amount of true treatment effect of $\delta_L = \mu_L - \mu_P = 2$ that correctly reflects our assumptions (prior mean), and assume that σ and β are fixed and known. In a traditional 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 for a traditional one-stage design 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 (_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 ($\delta_L = \mu_L - \mu_P = 1$), 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 ($\delta_L = \mu_L - \mu_P = 4$) 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 traditional one-stage design are no longer needed.

In addition, the proposed SSD approaches account for various amounts of prior information. For instance, we applied noninformative 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 using 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, 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.

4.2 | Sensitivity Analysis: Scenarios 11–15

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.

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 δ_L, 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 SSD, 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, 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.

5 | Discussion

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 AF, as a smaller AF indicates more savings. Motivated by the VIF as defined in cluster randomized controlled trials, the AF 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 two approximations are used. Moreover, the two-step approach does not necessarily require the input of the values 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 somewhat more conservative estimations of the sample sizes.

Efficient sample size calculations for Bayesian analysis of novel study designs are greatly needed, because often simulations are difficult to perform and require computational time that is not available. 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.

An alternative way of doing power analysis is to utilize Markov Chain Monte Carlo (MCMC) procedures and draw inference from the constructed posterior samples. However, MCMC can be

TABLE 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.

			Two-stage snSMART design				
	One-stage design		One-step Approach		Two-step approach		
Scenario	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

Note: We consider Scenarios 1, 11–15 as defined in Table 2 under two study designs. In a one-stage parallel design, N_{Freq} stands for the sample size calculated from the classical frequentist 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 one-step approach and two-step approach as in Fang et al. [3]. 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 6000 simulated datasets compatible with each scenario.

computationally burdensome. 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 6000 simulations in each setting still required more than 6 h 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 repetitive simulations. Meanwhile, we avoided using 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 computational 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 designs [27] share several properties with an snSMART. For instance, both designs lead to a data structure with multiple stages (periods). Also, participants may receive more than one treatment in the study. Specifically, complete crossover designs in which each patient receives all treatments sequentially may not always be feasible, especially in trials with more than three treatments. As an alternative, similar to an snSMART design, patients in an incomplete crossover design only receive a subset of all treatment options, requiring less resources than a complete crossover design. Moreover, both snSMART and crossover designs apply statistical approaches that account for the 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. Besides the crossover designs, the sequential parallel designs involve multiple stages, often without treatment switching during the study. Widely used sample size calculation in sequential parallel designs relies on the control of family-wise type I error [28, 29] with certain assumptions of the joint distribution of the sequential test statistic. Different boundaries can be applied to make flexible interim decisions. Our Bayesian approach, on the other hand, can be considered as a Bayesian alternative with an explicit derivation of the posterior distribution of data from both stages to potentially help determine the sample size that guarantees the desired performance across stages.

To summarize, the SSD approach we proposed in this manuscript can be extended to meet a variety of purposes. For instance, it can be applied directly to find the appropriate Bayesian sample size for either a traditional one-stage design or a two-stage design as illustrated in this manuscript. And, the explicit posterior distributions for data from both stages have a great potential in further developing approaches to apply in clinical trials share similar properties, such as the crossover designs. In addition, it can be used to find the explicit posterior distributions for data from multiple stages under proper assumptions, which enables applications including but not restricted to SSD.

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, we show that misspecification can lead to some power loss. Identifying the sensitivity of the calculation to different magnitudes of the 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. 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.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

References

1. R. N. Tamura, J. P. Krischer, C. Pagnoux, et al., "A Small n Sequential Multiple Assignment Randomized Trial Design for Use in Rare Disease Research," *Contemporary Clinical Trials* 46 (2016): 48–51.

2. S. A. Murphy, "An Experimental Design for the Development of Adaptive Treatment Strategies," *Statistics in Medicine* 24 (2005): 1455–1481.

3. F. Fang, R. N. Tamura, T. M. Braun, and K. M. Kidwell, "Comparing Dose Levels to Placebo Using a Continuous Outcome in a Small n, Sequential, Multiple Assignment, Randomized Trial (snSMART)," *Statistics in Biopharmaceutical Research* 15 (2022): 1–8.

4. J. Cohen, "The Statistical Power of Abnormal-Social Psychological Research: A Review," *Journal of Abnormal and Social Psychology* 65 (1962): 145–153.

5. C. J. Adcock, "Sample Size Determination: A Review," *Journal of the Royal Statistical Society: Series D (The Statistician)* 46 (1997): 261–283.

6. C. J. Adcock, "A Bayesian Approach to Calculating Sample Sizes," *Journal of the Royal Statistical Society: Series D (The Statistician)* 37 (1988): 433–439.

7. L. Joseph, D. B. Wolfson, and R. D. Berger, "Sample Size Calculations for Binomial Proportions via Highest Posterior Density Intervals," *Journal of the Royal Statistical Society: Series D (The Statistician)* 44 (1995): 143–154.

8. L. Joseph and P. Belisle, "Bayesian Sample Size Determination for Normal Means and Differences Between Normal Means," *Journal of the Royal Statistical Society: Series D (The Statistician)* 46 (1997): 209–226.

9. A. E. Raftery, W. Gilks, S. Richardson, and D. Spiegelhalter, *Hypothesis Testing and Model* (New York: Chapman & Hall London, 1995).

10. R. Weiss, "Bayesian Sample Size Calculations for Hypothesis Testing," *Journal of the Royal Statistical Society: Series D (The Statistician)* 46 (1997): 185–191.

11. D. V. Lindley, "The Choice of Sample Size," *Journal of the Royal Statistical Society: Series D (The Statistician)* 46 (1997): 129–138.

12. F. Wang and A. E. Gelfand, "A Simulation-Based Approach to Bayesian Sample Size Determination for Performance Under a Given Model and for Separating Models," *Statistical Science* 17 (2002): 193–208.

13. S. K. Sahu and T. M. F. Smith, "A Bayesian Method of Sample Size Determination With Practical Applications," *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 169 (2006): 235–253.

14. Q. Fu, H. Hoijtink, and M. Moerbeek, "Sample-Size Determination for the Bayesian t Test and Welch's Test Using the Approximate Adjusted Fractional Bayes Factor," *Behavior Research Methods* 53 (2021): 139–152.

15. F. De Santis, "Statistical Evidence and Sample Size Determination for Bayesian Hypothesis Testing," *Journal of Statistical Planning and Inference* 124 (2004): 121–144.

16. F. De Santis, "Using Historical Data for Bayesian Sample Size Determination," *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 170 (2007): 95–113.

17. M. H. Chen, J. G. Ibrahim, P. Lam, A. Yu, and Y. Zhang, "Bayesian Design of Noninferiority Trials for Medical Devices Using Historical Data," *Biometrics* 67 (2011): 1163–1170.

18. P. Brutti, F. De Santis, and S. Gubbiotti, "Bayesian-Frequentist Sample Size Determination: A Game of Two Priors," *Metron* 72 (2014): 133–151.

19. A. Turchetta, E. E. M. Moodie, D. A. Stephens, and S. D. Lambert, "Bayesian Sample Size Calculations for Comparing Two Strategies in SMART Studies," *Biometrics* 79 (2022): 2489–2502.

20. S. C. Chow, J. Shao, H. Wang, and Y. Lokhnygina, *Sample Size Calculations in Clinical Research* (New York: Chapman and Hall/CRC, 2017).

21. K. Hemming, A. J. Girling, A. J. Sitch, J. Marsh, and R. J. Lilford, "Sample Size Calculations for Cluster Randomised Controlled Trials With a Fixed Number of Clusters," *BMC Medical Research Methodology* 11 (2011): 1–11.

22. H. Wang, S. C. Chow, and M. Chen, "A Bayesian Approach on Sample Size Calculation for Comparing Means," *Journal of Biopharmaceutical Statistics* 15 (2005): 799–807.

23. K. H. Zou, F. S. Resnic, A. S. Gogate, S. Ondategui-Parra, and L. Ohno-Machado, "Efficient Bayesian Sample Size Calculation for Designing a Clinical Trial With Multi-Cluster Outcome Data," *Biometrical Journal: Journal of Mathematical Methods in Biosciences* 45 (2003): 826–836.

24. B. Wei, T. M. Braun, R. N. Tamura, and K. M. Kidwell, "Sample Size Determination for Bayesian Analysis of Small n Sequential, Multiple Assignment, Randomized Trials (snSMARTs) With Three Agents," *Journal of Biopharmaceutical Statistics* 30 (2020): 1109–1120.

25. G. Shieh and R. G. O'Brien, *A Simpler Method to Compute Power for Likelihood Ratio Tests in Generalized Linear Models* (Alexandria, VA: Annual Joint Statistical Meetings of the American Statistical Association, 1998).

26. S. Morita, P. F. Thall, and P. Müller, "Determining the Effective Sample Size of a Parametric Prior," *Biometrics* 64 (2008): 595–602.

27. A. S. Hedayat and M. Yang, "Optimal and Efficient Crossover Designs for Comparing Test Treatments With a Control Treatment," *Annals of Statistics* 33 (2005): 915–943.

28. J. Whitehead, *The Design and Analysis of Sequential Clinical Trials*, vol. 53 (New York: John Wiley & Sons, 1997), 1564.

29. L. Cui, H. M. J. Hung, and S. J. Wang, "Modification of Sample Size in Group Sequential Clinical Trials," *Biometrics* 55 (1999): 853–857.

Appendix A

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:

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

Suppose we denote the N observed outcomes at the end of stage $j = 1, 2$ as $\boldsymbol{Y_j}$, that is, $\boldsymbol{Y_1} = \left(Y_{11}, Y_{21}, \ldots, Y_{N1}\right)^T$ and $\boldsymbol{Y_2} = \left(Y_{12}, Y_{22}, \ldots, Y_{N2}\right)^T$. Similarly, we denote treatment assignments in stage $j = 1, 2$ as $\boldsymbol{T_j}$ with $\boldsymbol{T_1} = \left(T_{11}, T_{21}, \ldots, T_{N1}\right)^T$ and $\boldsymbol{T_2} = \left(T_{12}, T_{22}, \ldots, T_{N2}\right)^T$. To help clarify the derivation, we define the $(N \times 4)$ matrix $\boldsymbol{X_1} = \left(I(\boldsymbol{T_1}=P)^T, I(\boldsymbol{T_1}=L)^T, I(\boldsymbol{T_1}=H)^T, \boldsymbol{0}^T\right)$ and the $(N \times 4)$ matrix $\boldsymbol{X_2} = \left(I(\boldsymbol{T_2}=P)^T, I(\boldsymbol{T_2}=L)^T, I(\boldsymbol{T_2}=H)^T, \boldsymbol{1}^T\right)$, where $I(\boldsymbol{T_j}=k)$ are indicators of receiving treatment k at stage j, $\boldsymbol{0}^T$ and $\boldsymbol{1}^T$ are vectors of length N. Denoting the parameters relevant to the expected values of outcomes as the (4×1) matrix $\boldsymbol{\mu} = \left(\mu_P, \mu_L, \mu_H, \alpha\right)^T$, then $\boldsymbol{Y} = \left(\boldsymbol{Y_1}, \boldsymbol{Y_2}\right)^T$ has a bivariate normal distribution with mean $\left(\boldsymbol{X_1\mu}, \boldsymbol{X_2\mu}\right)^T$ and a $(2N \times 2N)$ covariance matrix

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

We partition the inverse matrix of $\sum$ by four $N \times N$ square block matrices, that is, $\sum^{-1} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$.

We assume a priori that μ has a multivariate normal distribution with mean $\boldsymbol{\mu_0} = \left(\mu_0^P, \mu_0^L, \mu_0^H, \mu_0^\alpha\right)^T$ and covariance matrix Σ_0, where

$$\sum_{\boldsymbol{0}} = \begin{pmatrix} \tau_P^2 & 0 & 0 & 0 \\ 0 & \tau_L^2 & 0 & 0 \\ 0 & 0 & \tau_H^2 & 0 \\ 0 & 0 & 0 & \tau_\alpha^2 \end{pmatrix}$$

with prior effective sample size of the mean outcome parameters as $\kappa_k = \sigma^2 / \tau_k^2$ for treatment k and $\kappa_\alpha = \sigma^2 / \tau_\alpha^2$ for the shift parameter α. For known values $\boldsymbol{Y} = \boldsymbol{y} = \left(\boldsymbol{y_1}, \boldsymbol{y_2}\right)^T$, the posterior distribution of μ assuming β and σ are known is

$$\begin{aligned} p\left(\boldsymbol{\mu}\middle|\, \boldsymbol{Y}=\boldsymbol{y}, \sum\right) \quad &\propto \exp\left\{-\frac{1}{2}\left[\begin{pmatrix} \boldsymbol{y_1}-\boldsymbol{X_1\mu} \\ \boldsymbol{y_2}-\boldsymbol{X_2\mu} \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} \boldsymbol{y_1}-\boldsymbol{X_1\mu} \\ \boldsymbol{y_2}-\boldsymbol{X_2\mu} \end{pmatrix} + \left(\boldsymbol{\mu}-\boldsymbol{\mu_0}\right)^T \sum_{\boldsymbol{0}}^{-1} \left(\boldsymbol{\mu}-\boldsymbol{\mu_0}\right)\right]\right\} \\ &\propto \exp\{-\frac{1}{2}[\left(\boldsymbol{y_1}-\boldsymbol{X_1\mu}\right)^T \sum_{\boldsymbol{11}} \left(\boldsymbol{y_1}-\boldsymbol{X_1\mu}\right) + \left(\boldsymbol{y_2}-\boldsymbol{X_2\mu}\right)^T \sum_{\boldsymbol{12}} \left(\boldsymbol{y_1}-\boldsymbol{X_1\mu}\right) + \\ &\quad \left(\boldsymbol{y_1}-\boldsymbol{X_1\mu}\right)^T \sum_{\boldsymbol{21}} \left(\boldsymbol{y_2}-\boldsymbol{X_2\mu}\right) + \left(\boldsymbol{y_2}-\boldsymbol{X_2\mu}\right)^T \sum_{\boldsymbol{22}} \left(\boldsymbol{y_2}-\boldsymbol{X_2\mu}\right)]\} \\ &\propto \exp\{-\frac{1}{2}[\boldsymbol{\mu}^T\left(\boldsymbol{X_1}^T \sum_{\boldsymbol{11}} \boldsymbol{X_1} + \boldsymbol{X_2}^T \sum_{\boldsymbol{12}} \boldsymbol{X_1} + \boldsymbol{X_1}^T \sum_{\boldsymbol{21}} \boldsymbol{X_2} + \boldsymbol{X_2}^T \sum_{\boldsymbol{22}} \boldsymbol{X_2} + \sum_{\boldsymbol{0}}^{-1}\right)\boldsymbol{\mu} + \\ &\quad \boldsymbol{\mu}^T(2\boldsymbol{X_1}^T \sum_{\boldsymbol{11}} \boldsymbol{y_1} + \boldsymbol{X_2}^T \sum_{\boldsymbol{12}} \boldsymbol{y_1} + \boldsymbol{X_1}^T \sum_{\boldsymbol{12}} \boldsymbol{y_2} + \boldsymbol{X_1}^T \sum_{\boldsymbol{21}} \boldsymbol{y_2} + \boldsymbol{X_2}^T \sum_{\boldsymbol{21}} \boldsymbol{y_1} + \\ &\quad 2\boldsymbol{X_2}^T \sum_{\boldsymbol{22}} \boldsymbol{y_2} + 2\sum_{\boldsymbol{0}}^{-1} \boldsymbol{\mu_0})]\}. \end{aligned} \tag{A1}$$

Let $\boldsymbol{A} = \boldsymbol{X_1}^T \sum_{11} \boldsymbol{X_1} + \boldsymbol{X_2}^T \sum_{12} \boldsymbol{X_1} + \boldsymbol{X_1}^T \sum_{21} \boldsymbol{X_2} + \boldsymbol{X_2}^T \sum_{22} \boldsymbol{X_2} + \sum_{\boldsymbol{0}}^{-1}$ and $\boldsymbol{B} = (1/2)\left(2\boldsymbol{X_1}^T \sum_{11} \boldsymbol{Y_1} + \boldsymbol{X_2}^T \sum_{12} \boldsymbol{y_1} + \boldsymbol{X_1}^T \sum_{12} \boldsymbol{y_2} + \boldsymbol{X_1}^T \sum_{21} \boldsymbol{y_2} + \boldsymbol{X_2}^T \sum_{21} \boldsymbol{y_1} + 2\boldsymbol{X_2}^T \sum_{22} \boldsymbol{y_2} + 2\sum_{\boldsymbol{0}}^{-1} \boldsymbol{\mu_0}\right)$, then the posterior distribution can be written as

$$\begin{aligned} p\left(\boldsymbol{\mu}\middle|\, \boldsymbol{Y}=\boldsymbol{y}, \sum\right) \quad &\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}^T\boldsymbol{A}^{-1}\boldsymbol{\mu} - 2\boldsymbol{\mu}^T\boldsymbol{B}\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}-\boldsymbol{A}^{-1}\boldsymbol{B}\right)^T\boldsymbol{A}\left(\boldsymbol{m}-\boldsymbol{A}^{-1}\boldsymbol{B}\right)\right\} \end{aligned} \tag{A2}$$

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

$$\widehat{\boldsymbol{\mu}} = \frac{1}{2}\boldsymbol{V}\left(2\boldsymbol{X}_{\mathbf{1}}^{T}\sum_{\mathbf{11}}\boldsymbol{y}_{\mathbf{1}} + \boldsymbol{X}_{\mathbf{2}}^{T}\sum_{\mathbf{12}}\boldsymbol{y}_{\mathbf{1}} + \boldsymbol{X}_{\mathbf{1}}^{T}\sum_{\mathbf{12}}\boldsymbol{y}_{\mathbf{2}} + \boldsymbol{X}_{\mathbf{1}}^{T}\sum_{\mathbf{21}}\boldsymbol{y}_{\mathbf{2}} + \boldsymbol{X}_{\mathbf{2}}^{T}\sum_{\mathbf{21}}\boldsymbol{y}_{\mathbf{1}} + 2\boldsymbol{X}_{\mathbf{2}}^{T}\sum_{\mathbf{22}}\boldsymbol{y}_{\mathbf{2}} + 2\sum\nolimits_{\mathbf{0}}^{-\mathbf{1}}\boldsymbol{\nu}_{\mathbf{0}}\right)$$

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

$$\text{and } \mathbf{V} = \left(\boldsymbol{X}_{\mathbf{1}}^{T}\sum_{\mathbf{11}}\boldsymbol{X}_{\mathbf{1}} + \boldsymbol{X}_{\mathbf{2}}^{T}\sum_{\mathbf{12}}\boldsymbol{X}_{\mathbf{1}} + \boldsymbol{X}_{\mathbf{1}}^{T}\sum_{\mathbf{21}}\boldsymbol{X}_{\mathbf{2}} + \boldsymbol{X}_{\mathbf{2}}^{T}\sum_{\mathbf{22}}\boldsymbol{X}_{\mathbf{2}} + \boldsymbol{\mu}_{\mathbf{0}}^{-\mathbf{1}}\right)^{-1}$$

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

In the expression of $\boldsymbol{V}$, $N_{kk'} = \sum_{i=1}^{N} I(T_{i1} = k, T_{i2} = k')$, 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 proposed sample size calculation approaches.