stats

MDPI

Article

Adaptive Clinical Trials and Sample Size Determination in the Presence of Measurement Error and Heterogeneity

Hassan Farooq [1], Sajid Ali [1,*], Ismail Shah [1,2,*], Ibrahim A. Nafisah [3,†] and Mohammed M. A. Almazah [4,†]

[1] Department of Statistics, Quaid-i-Azam University, Islamabad 45320, Pakistan; farooqhassan949@gmail.com
[2] Department of Statistical Sciences, University of Padua, 35121 Padova, Italy
[3] Department of Statistics and Operations Research, College of Sciences, King Saud University, Riyadh 11451, Saudi Arabia; inafisah@ksu.edu.sa
[4] Department of Mathematics, College of Sciences and Arts (Muhyil), King Khalid University, Muhyil 61421, Saudi Arabia; mmalmazah@kku.edu.sa
* Correspondence: sajidali@qau.edu.pk (S.A.); ismail.shah@unipd.it (I.S.)
† These authors contributed equally to this work.

Abstract: Adaptive clinical trials offer a flexible approach for refining sample sizes during ongoing research to enhance their efficiency. This study delves into improving sample size recalculation through resampling techniques, employing measurement error and mixed distribution models. The research employs diverse sample size-recalculation strategies standard simulation, R1 and R2 approaches where R1 considers the mean and R2 employs both mean and standard deviation as summary locations. These strategies are tested against observed conditional power (OCP), restricted observed conditional power (ROCP), promising zone (PZ) and group sequential design (GSD). The key findings indicate that the R1 approach, capitalizing on mean as a summary location, outperforms standard recalculations without resampling as it mitigates variability in recalculated sample sizes across effect sizes. The OCP exhibits superior performance within the R1 approach compared to ROCP, PZ and GSD due to enhanced conditional power. However, a tendency to inflate the initial stage's sample size is observed in the R1 approach, prompting the development of the R2 approach that considers mean and standard deviation. The ROCP in the R2 approach demonstrates robust performance across most effect sizes, although GSD retains superiority within the R2 approach due to its sample size boundary. Notably, sample size-recalculation designs perform worse than R1 for specific effect sizes, attributed to inefficiencies in approaching target sample sizes. The resampling-based approaches, particularly R1 and R2, offer improved sample size recalculation over conventional methods. The R1 approach excels in minimizing recalculated sample size variability, while the R2 approach presents a refined alternative.

Keywords: adaptive clinical trials; group sequential design; measurement error in trials; promising zone; observed conditional power; sample size determination; restricted observed conditional power; mixture distribution

MSC: 62P10

check for updates

Academic Editor: Wei Zhu

Received: 16 March 2025
Revised: 23 April 2025
Accepted: 23 April 2025
Published: 25 April 2025

Citation: Farooq, H.; Ali, S.; Shah, I.; Nafisah, I.A.; Almazah, M.M.A. Adaptive Clinical Trials and Sample Size Determination in the Presence of Measurement Error and Heterogeneity. *Stats* **2025**, *8*, 31. https://doi.org/10.3390/stats8020031

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Clinical trials play a pivotal role in the evaluation of new medical treatments or interventions, providing reliable evidence for the safety and efficacy of treatments. However, traditional clinical trial designs can be inefficient and inflexible, requiring careful calculation of sample sizes based on accurate parameter assumptions. Incorrect parameter

assumptions can lead to sample sizes that are too small or too large, with significant ethical and economic consequences. Adaptive clinical trials (ACTs) offer a promising alternative, allowing for modifications to the trial's parameters based on accumulating data. This flexibility can maximize efficiency, enhance the trial's ability to answer research questions effectively and reduce the number of participants required.

One critical aspect of adaptive clinical trials is the sample size, which directly impacts the trial's statistical power and the ability to draw meaningful conclusions [1]. The traditional fixed-sample size designs often face many challenges, as the required sample size is typically determined before any data are collected, leading to inefficiencies and potential resource wastage. Sample size recalculation in adaptive clinical trials is an important area of research, as it can help to improve the efficiency and accuracy of clinical trials [2]. This research explores a new approach to sample size recalculation based on a measurement error model (MEM) and mixture distribution. The proposed approach generates treatment and control groups and adaptively adjusts the sample size based on interim analysis. The study investigates several approaches to recalculating sample size in adaptive clinical trials, with a focus on using a two-component mixture distribution to generate treatment and control groups. These approaches adaptively adjust the sample size based on interim analysis, ensuring that the trial is appropriately powered and resource-efficient.

Sample size determination is a critical aspect of clinical trial design. It is the process of estimating the number of participants needed to provide sufficient statistical power to detect a meaningful treatment effect. Inadequate sample sizes can lead to inconclusive or false-negative results, while excessively large sample sizes can waste resources. Calculating sample size typically involves using four parameters: Type-I error, power, control group assumptions (response rate and standard deviation) and predicted treatment impact. Measurement error models are statistical models used to account for measurement errors in data. In clinical trials, measurement errors can arise due to various factors, such as variability in assessments or instruments used to measure outcomes. These errors can lead to biased results and reduced statistical power. By accounting for measurement errors, researchers can better distinguish true treatment effects from noise, leading to more reliable and robust conclusions about the effectiveness of interventions. Moreover, the utilization of mixture-distribution models for the generation of treatment and control groups addresses the challenges associated with randomization, covariate balance and group heterogeneity, ultimately enhancing the robustness and credibility of clinical trial outcomes.

Several methods have been proposed to implement resampling in adaptive clinical trials, including Pocock [3] and O'Brien and Fleming [4] boundaries. These methods allow for sample size updates based on the observed interim effect while controlling the overall Type-I error rate of the study. Denne [5] noticed that sample size may be determined by one or more nuisance characteristics, which are typically unknown, to obtain a specific power at a predetermined absolute difference in mean response. Friede and Kieser [6] developed the internal pilot study design, which enables the sample size to be revised for a trial using the estimated variance discovered by interim analysis. Charles et al. [7] noticed that the primary goal of an a priori sample size calculation is to determine the minimum number of participants required to identify clinical significance. Hammouri et al. [8] used urn allocation and the O'Brien and Fleming multiple testing procedure to compare two treatments in clinical trials in a novel way.

Kieser and Friede [9] discussed two-stage techniques, where the variance is reestimated from a subsample and the sample size is modified as needed, which are appealing due to the uncertainty in the design step. They demonstrated analytically that the Type-I error rate of the t-test is not impacted by the use of straightforward, blind variance estimators for sample size recalculation. Harden and Friede [10] and Das et al. [11] analyzed

multi-center randomized clinical trials, which is crucial for evidence-based medicine. To address clustering in data, the authors' mixed models are used. The existing sample size-calculation methods only consider balanced treatment allocations, which may not be realistic. To overcome this, a new sample size-determination procedure is proposed for multi-center trials comparing two treatment groups [10]. The method incorporated random effects, allowing arbitrary sample sizes and assumed fixed block length block randomization. Through simulations, the proposed approach demonstrated its superiority over conventional methods, taking into account parameters such as block length and center heterogeneity.

Jennison and Turnbull [12] discussed the common practice of setting the sample size in clinical trials based on a specified treatment effect, disregarding the importance of detecting smaller but clinically significant effects. The proposed group sequential designs focused on reducing the expected sample size while maintaining sufficiency and considering the possibility of small treatment effects at the design stage. The methods are compared with Fisher's variance spending procedure and show potential advantages. Pritchett et al. [13] compared different types of sample size-recalculation (SSR) designs, including blinded SSR, unblinded SSR and conventional group sequential designs (GSDs). The study also presented statistical methods for unblinded and blinded SSR designs and highlighted the importance of controlling Type-I error rate and estimating the treatment effect accurately. Chakraborty and Gu [14] discussed the prevalent challenges posed by missing values and dropouts in longitudinal studies within medical and public health fields.

Boos and Brownie [15] introduced novel rank-based methods within the framework of mixed linear models for analyzing data from multisite clinical trials. Unlike current rank methods, the proposed procedures specifically assess a drug's main effect in the presence of a random drug-by-site (or investigator) interaction. Nagin and Odgers [16] highlighted the growing utilization of group-based trajectory models in clinical research for tracking symptom development and gauging diverse responses to clinical interventions. The review furnished a comprehensible overview of both group-based trajectory and growth mixture modeling, coupled with instances showcasing their clinical research applications. Deng et al. [17] discussed a novel two-stage multivariate Mendelian randomization method for investigating the causal effects of clinical factors on various outcomes, especially in cases of mixed correlated outcomes with different distributions. The conventional MR methodology focused on single outcomes, disregarding correlation structures, potentially resulting in reduced statistical power. The proposed method addressed this limitation by jointly analyzing multiple outcomes using genetic instrumental variables. Spanbauer and Sparapani [18] suggested precision medicine's transformative potential for clinical trials and subsequent treatment strategies. The study highlighted the relevance of modern machine learning techniques, particularly Bayesian additive regression trees (BART), for identifying distinct population segments and devising personalized treatment rules. Liang et al. [19] proposed an innovative mixed-effects varying-coefficient model to address measurement errors in covariates. Morgan and Elashoff [20] discussed the impact of measurement error in prognostic factors, often considered as covariates in clinical trials assessing treatment effects. Employing Weibull regression models and asymptotic theory, the study investigated the efficiency of treatment effect estimation when adjusting for a dichotomous and a continuous covariate affected by measurement error. The analysis revealed how such errors can diminish estimation efficiency.

Measurement error in covariates is a critical challenge in biomedical research, introducing bias and imprecision in exposure–outcome relationships. Wang et al. [21] discussed the challenges posed by measurement error in the context of generalized linear mixed models (GLMMs) for clustered data, where one predictor is afflicted by such error. Yang et al. [22]

introduced a corrected empirical likelihood approach for statistical inference in generalized linear measurement error models, encompassing Gaussian, Poisson and logistic regressions. Brakenhoff et al. [23] investigated the impact of measurement error in covariates within medical research, particularly their potential to introduce bias and imprecision in exposure–outcome relationships. Despite the acknowledged significance of this issue, the extent to which it was addressed in current research practices remains uncertain. Through a systematic review of general medicine and epidemiology literature, the study highlighted a lack of consideration for covariate measurement error in a majority of high-impact journal publications. This oversight makes it challenging for readers to assess the robustness of presented results. The research underscored the need for heightened awareness regarding the possible repercussions of measurement error and calls for guidance on employing correction methods. Measurement error, arising from inaccuracies in measurement instruments and data, poses a critical challenge to valid inferences in biomedical research. As medical datasets continue to expand, recognizing and addressing covariate measurement error becomes imperative for maintaining the integrity of research findings.

This research aims to improve the efficiency and accuracy of adaptive clinical trials by (i) proposing a new approach to generate treatment and control groups based on a measurement error and mixture model and (ii) investigate the application of different sample size-recalculation methods at interim analyses. The research is structured around a rigorous methodological framework. To achieve the objectives, a simulation-based study is conducted to evaluate the performance of the MEM in generating trial groups. Furthermore, various sample size-recalculation approaches are applied in interim analyses, including observed conditional power (OCP), restricted observed conditional power (ROCP), promising zone (PZ) and group sequential design (GSD). The effectiveness of these methods is assessed concerning statistical power, Type-I error control and overall trial efficiency.

The rest of the study is organized as follows: Section 2 discusses the measurement error and mixture models. Different methodologies to recalculate the sample size are discussed in Section 3. The performance evaluation criteria of sample size-calculation methods are discussed in Section 4. A simulation study to evaluate the performance assuming mixture and measurement error models is tabulated and discussed in Section 5 and Section 6, respectively. An illustrative study for alleviating pain among osteoarthritis patients is discussed in Section 7. Finally, Section 8 presents concluding and future remarks.

2. Measurement Error Models and Mixture Distribution

2.1. Measurement Error Model

Measurement error models (MEMs) are statistical techniques used to account for inaccuracies and uncertainties in the measurement process when analyzing data. In various research fields, measurements often contain errors that can distort the true relationships between variables. These errors can arise from a variety of sources, such as imperfect instruments, human error, environmental factors and inherent variability in the phenomenon being measured. MEMs help researchers address these issues by providing a framework to estimate the true relationships between variables while considering the impact of measurement errors. These models can be broadly categorized into two main types:

1. **Classical Measurement Error Models:** In classical MEMs, the error is assumed to be present in the independent (explanatory) variable. This type of model is commonly known as an errors-in-variables (EIV) model. It accounts for the fact that the observed values of the independent variable are subject to measurement errors, leading to biased and inconsistent parameter estimates if not properly addressed. Mathematically, a classical EIV model can be expressed as follows
   true relationship: $Y = \alpha + \beta * X + u$

observed relationship: $Y_{obs} = \alpha + \beta * X_{obs} + \epsilon$
where $X_{obs} = X + \eta$ is the observed (error-prone) value of X, ϵ is the error in the observed Y and η is normally distributed with zero mean and variance σ^2.

2. **Errors-in-Response or Dependent Variable Models:** These models focus on measurement errors in the dependent (response) variable. In this case, the observed responses are considered to be measured with errors. Errors-in-response models are less common but are used when the measurement error is primarily concentrated in the outcome variable. Mathematically,
true relationship: $Y = \alpha + \beta * X + u$
observed relationship: $Y_{obs} = Y + \epsilon$
where ϵ is the error in the observed response variable Y.

Measurement errors can significantly affect the reliability and validity of the estimates derived from statistical analyses. The key implications of measurement errors on estimates are the following. Measurement errors can introduce bias in parameter estimates, leading to incorrect conclusions about the relationships between variables. For example, if the true relationship between two variables is linear, measurement errors can make it appear non-linear or attenuate the observed relationship. The researchers may overestimate or underestimate the strength of associations, leading to misguided policy recommendations or interventions. Also, measurement errors can reduce the precision and efficiency of parameter estimates. The variability introduced by measurement errors can inflate standard errors, leading to wider confidence intervals and reduced statistical power. Inconsistent estimates occur when the magnitude and direction of bias change across different samples or settings. This can lead to difficulties in replicating research findings and generalizing results. Furthermore, measurement errors can distort hypothesis tests, leading to incorrect p-values and flawed decisions about statistical significance. This can result in both Type-I and Type-II errors.

2.2. Mixed Distribution Models

The mixed distribution model (MDM) is a powerful statistical technique that allows researchers to create well-balanced and comparable groups, taking into account the heterogeneity of the underlying data. This section outlines the process of utilizing the MDM for group assignment, its advantages and the steps involved in its implementation. The MDM is a sophisticated statistical approach that combines elements of probability distributions to create groups that are representative of the underlying population. It allows for the incorporation of various covariates and factors, ensuring that the treatment and control groups are not only randomized but also balanced using relevant characteristics. The MDM takes into consideration both continuous and categorical variables, accommodating the complexity of real-world data. The general form of a mixed distribution model can be expressed as follows:

$$f(x;\theta) = \sum_{i=1}^{k} \pi_i \cdot f_i(x;\theta_i)$$

where $f(x;\theta)$ represents the mixed distribution with parameters θ, k is the number of components in the mixture, π_i are the mixing proportions, satisfying $\sum_{i=1}^{k} \pi_i = 1$ and $f_i(x;\theta_i)$ represents the ith component distribution with parameters θ_i. For a two-component mixture $f(x;\theta) = \sum_{i=1}^{2} \pi_i \cdot f_i(x;\theta_i) = \pi_1 \cdot f_1(x;\theta_1) + \pi_2 \cdot f_2(x;\theta_2)$, which further can be written as $f(x;\theta) = \pi \cdot f_1(x;\theta_1) + (1-\pi) \cdot f_2(x;\theta_2)$ due to the fact $\sum_{i=1}^{2} \pi_i = 1$. The MDM ensures that treatment and control groups are balanced, thereby reducing the risk of confounding variables affecting the results. This balance enhances the internal validity of the clinical trial. Moreover, by accounting for the underlying distribution of the data, the MDM helps ensure that the generated groups accurately represent the population.

3. Methodology

In this section, a comprehensive overview of the research design, data-analysis techniques and any other pertinent procedures employed to ensure the validity and reliability of the study's findings is discussed.

We consider a clinical experiment that is two-armed, randomized and controlled. The n observations in treatment group T and control group C have a normal distribution with means μ^T and μ^C with a common variance σ^2, i.e., $X_i^T \sim N(\mu^T, \sigma^2), X_i^C \sim N(\mu^C, \sigma^2)$, $i = 1, 2, ..., n$. The current study is interested in investigating the one-sided superiority test problem, which can be stated as $H_0 : \mu^T - \mu^C \leq 0$ vs. $H_1 : \mu^T - \mu^C > 0$ which refers to a situation where high values of the endpoint are viewed favorably. We investigate an adaptive group sequential design with two stages, which is the most basic and widely used adaptive group sequential design. Consequently, we have to make two independent statistics,

$$T_i = \frac{\bar{X}_i^T - \bar{X}_i^C}{S_{\text{pooled},i}} \cdot \sqrt{\frac{n_i}{2}}, \tag{1}$$

where $i \in \{1, 2\}$ denote stages, $\bar{X_i^T}$ and $\bar{X_i^C}$ are the means of treatment and control groups, $S_{\text{pooled,i}}$ is the pooled standard deviation and n_i denote sample size per group in stage i with $n_1 + n_2 = n$. The well-known formula for S_{pooled} is given as:

$$S_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}, \tag{2}$$

where n_1 and n_2 are the sample sizes of the two groups, S_1 and S_2 are the standard deviations of the two groups. It must be noted that T_1 only includes data from the first stage, and T_2 only includes data from the second stage, both of which have an approximately normal distribution.

If the interim test statistic T_1 falls within the recalculation area (RA) given as $[q_{1-\alpha_0}; q_{1-\alpha_1})$, where α_0 denotes futility stopping bound for a one-sided p-value of stage one, α_1 refers to the local one-sided significance level and q are the respective quantiles of the normal distribution, the trial moves to the second stage. If $T_1 \geq q_{1-\alpha_1}$ the trial is stopped with an early rejection of the null hypothesis after the first stage, or accept the null hypothesis if $T_1 < q_{1-\alpha_0}$. After combining all observed data over two stages using an inverse normal combination test given as

$$T = \frac{w_1.T_1 + w_2.T_2}{\sqrt{w_1^2 + w_2^2}}, \tag{3}$$

where w_1 and w_2 are the weights, i.e., $w_1 = \sqrt{n_1}$ and $w_2 = \sqrt{n_2}$, T_1 and T_2 are two stochastically independent test statistics. If $T \geq q_{1-\alpha}$, where α refers to the local one-sided significance level for the final analysis, then the null hypothesis is rejected at the final analysis. For instance, local significance levels can be determined using the adjustments suggested by Pocock [3] or O'Brien and Fleming [4].

3.1. Methods for Recalculating Sample Size

Various approaches exist for adjusting sample sizes during an ongoing clinical trial. One straightforward method is the implementation of a group sequential design (GSD), where a fixed predetermined sample size is allocated for each stage. A more flexible variation of this concept is found in adaptive group sequential designs, where interim sample sizes can be determined based on the accumulating data. In the realm of adaptive group sequential designs, a prevalent strategy involves sample size adjustments aimed at

attaining a predefined conditional power value. This conditional power metric delineates the probability of accurately rejecting the null hypothesis, given the observed interim test statistic value and the cumulative sample size per group. The calculation of conditional power (CP) is contingent upon the true standardized treatment effect (Δ), which gauges the difference $(\mu^T - \mu^C)$ between means of the treatment and control groups divided by the shared standard deviation (σ), i.e., $\Delta = (\mu^T - \mu^C)/\sigma$. This dynamic approach to modifying sample sizes within adaptive designs holds substantial promise for enhancing trial efficiency and statistical robustness based on emerging data trends. Thus,

$$CP_\Delta(t_{1,n}) = \begin{cases} 0, \text{if the trial ends early due to futility,} \\ 1 - \Phi\left(q_{1-\alpha} \cdot \sqrt{\frac{w_1^2+w_2^2}{w_2^2}} - t_1 . \frac{w_1}{w_2} - \Delta . \sqrt{\frac{n_1}{2}} \cdot \sqrt{\frac{n-n_1}{n_1}}\right), \text{if the sample size is recalculated,} \\ 1, \text{if the trial is stopped early for efficacy.} \end{cases} \quad (4)$$

The subsequent subsections outline three distinct methods for recalculating the sample size by leveraging the observed conditional power. In this context, the formula incorporates the substitution of Δ which is replaced by the observed interim effect denoted as $t_1\sqrt{2/n_1}$. Alternatively, akin strategies entail the integration of an assumed effect for Δ, often termed as the anticipated conditional power. Notably, our current focus revolves around evaluating the impact of the resampling tool assuming MEM and MDM, which harmonizes seamlessly with existing recalculation approaches. Given this specific emphasis, we prioritize a comprehensive exploration of recalculation rules hinged on the observed conditional power, thereby omitting an exhaustive examination of varied recalculation strategies diverging primarily in their selection of Δ values. Within this framework, we exclusively investigate recalculation rules that draw from observed conditional power, while concurrently imposing an upper limit of n_{max} on the overall total sample size per group, a measure employed for practical viability considerations.

3.1.1. The Observed Conditional Power (OCP) Approach

For observed interim test statistics to fall in the recalculation area $[q_{1-\alpha_0}; q_{1-\alpha_1})$, it is necessary to make sure that we have the right number of people in our groups to catch any real effects. This idea is called conditional power, which tells us how likely we are to find something real. If this power $(1-\beta)$ is higher, it is better.

The basic concept is to find the smallest whole number $\tilde{n}$ that fits a special rule. This rule says that $\tilde{n}$ has to be greater than or equal to a specific value. This value comes from the equation

$$\tilde{n} \geq n_1 . \left(1 + \left(\frac{q_\beta - q_{1-\alpha} . \frac{\sqrt{w_1^2+w_2^2}}{w_2} + t_1 . \frac{w_1}{w_2}}{t_1}\right)^2\right). \quad (5)$$

Equation (5) helps us to know how many people we need in each group so that our study makes sense and we can obtain meaningful results. Using the OCP approach, we determine the overall sample size for each group as follows. If the interim test statistic t_1 falls within a recalculated area then the total sample size per group is the smaller value between a calculated quantity denoted as $\tilde{n}(t_1)$ and a predefined maximum value n_{max}. On the other hand, if t_1 does not fall within the recalculated area, then the total sample size per group remains fixed at n_1. This method ensures that we adapt the sample size based on the interim results and the level of confidence we seek in our study. This leads to the following formula

$$n_{\text{OCP}}(t_1) = \begin{cases} \min(\tilde{n}(t_1), n_{\max}), & \text{if } t_1 \in RA, \\ n_1, & \text{else.} \end{cases} \quad (6)$$

3.1.2. The Restricted Observed Conditional Power Approach

The approach known as the restricted observed conditional power (ROCP) shares similarities with the OCP method; however, as implied by its name, it comes with a specific restriction. An issue raised about the OCP approach centers on a particular scenario: when the Formula (5) indicates the need for larger sample sizes than the maximum value $n_{\max}$, the sample size is then capped at $n_{\max}$, irrespective of the potential conditional power achievable with this larger size. A viable approach might be to consider enlarging the sample size only if it ensures a minimum acceptable conditional power defined as $(1-\beta_{low}^{ROCP})$. This adjustment aims to strike a balance between sample size requirements and the attainable level of statistical confidence. As a result, the total sample size per group according to the ROCP approach is as follows

$$n_{\mathrm{ROCP}}(t_1) = \begin{cases} \min(\tilde{n}(t_1), n_{\max}), & \text{if } t_1 \in \mathrm{RA}, \text{and } CP(t_1, n_{\max}) \geq 1-\beta_{low}^{ROCP}, \\ n_1, & \text{else.} \end{cases} \tag{7}$$

3.1.3. The Promising Zone Approach

The innovative promising zone (PZ) approach, introduced by Mehta and Pocock [24], presents a distinct methodology. It commences with the determination of an initial total sample size, denoted as n_{ini}, for each group. Notably, this initial size is intentionally kept smaller than the maximum allowable total sample size, n_{max}, for each group. Additionally, the PZ approach establishes a predetermined lower threshold for the conditional power, represented as $1-\beta_{low}^{PZ}$. Importantly, it is worth highlighting that $1-\beta_{low}^{PZ}$ is not necessarily equal to $1-\beta_{low}^{ROCP}$, introducing flexibility based on specific requirements.

As the study progresses, the PZ approach facilitates sample size updates contingent on the observed interim test statistic t_1. These updates are governed by two potential pathways. First, if the recalculated sample size, referred to as $\tilde{n}$, adheres to the Formula (5), it follows the trajectory set by the initially proposed total sample size, n_{ini}. Alternatively, the sample size is restricted to the maximum value, n_{max}, per group. This adaptive approach optimally balances sample size requirements with the potential for achieving a robust conditional power, further enhancing the precision and reliability of our study outcomes. Consequently, the total sample size per group according to the promising zone (PZ) approach equals

$$n_{\mathrm{PZ}}(t_1) = \begin{cases} \min(\tilde{n}(t_1), n_{\max}), & \text{if } t_1 \in \mathrm{RA} \text{ and } 1-\beta_{\mathrm{PZ_low}} \leq CP(t_1, n_{ini}) < 1-\beta, \\ n_{ini}, & \text{if } t_1 \in \mathrm{RA} \text{ and } CP(t_1, n_{ini}) < 1-\beta_{low}^{PZ}, \\ & \text{or } t_1 \in \mathrm{RA} \text{ and } CP(t_1, n_{ini}) \geq 1-\beta, \\ n_1, & \text{else.} \end{cases} \tag{8}$$

4. Evaluating the Performance of Sample Size-Recalculation Rules

When our objective centers on enhancing the effectiveness of rules for recalculating sample sizes, it becomes essential to establish appropriate criteria for evaluating their performance. Among the common evaluation criteria are the average sample size and global power, both of which take on a stochastic nature within the framework of adaptive design. Recognizing the significance of not only assessing central tendencies but also incorporating measures of variability, we need a comprehensive perspective.

The evaluation of a sample size-recalculation rule can be approached from different angles. The global perspective examines the scenario before the trial's commencement, providing an average view of the two options: early trial termination or sample size recalculation at interim stages. However, this perspective presents challenges in interpreting a

combination of performance aspects associated with both, stopping early and recalculating sample sizes.

An alternative approach that prompts the researcher to consider how the sample size should be recalculated if the observed effect falls within the recalculation area during the interim analysis is a conditional perspective. Here, we assess the recalculation rules under the assumption that the trial continues past the interim point, where t_1 falls between $t_1 \in [q_{1-\alpha_0}; q_{1-\alpha_1})$, even though the specific value of t_1 remains unknown. This conditional perspective pertains to the recalculation area rather than a particular t_1 value. In this study, our focus is squarely on this conditional perspective, providing a comprehensive exploration that offers valuable insights into the optimization of sample size-recalculation methods. As a result, we explore ways to adjust sample sizes, looking at how well they perform based on the following criteria:

1. The expected conditional power, denoted as $E[CP_{\Delta}^{RA}]$.
2. The variability of the conditional power, represented as $Var[CP_{\Delta}^{RA}]$.
3. The anticipated conditional total sample size per group, expressed as $E[CN_{\Delta}^{RA}]$, which is the average size per group in the recalculation area.
4. The variability of the conditional total sample size per group, indicated by $Var[CN_{\Delta}^{RA}]$.

The assessment of performance measures encompasses a range of true standardized effect sizes, quantified by $\Delta = \frac{\mu^T - \mu^C}{\sigma}$. Notably, these evaluation criteria can be unified into a comprehensive performance metric, known as the conditional performance score (CS). While we provide an overview of this score's essential characteristics, the CS is composed of four distinct components: two components for evaluating the location and variability of the conditional power ($e_{CP}(\Delta)$ and $v_{CP}(\Delta)$) and two components for the location and variability of the conditional sample size ($e_{CN}(\Delta)$ and $v_{CN}(\Delta)$).

The fundamental concept underlying the location components is to compare expected values against predefined target values. In cases where the maximum allowed sample size is not greater than the corresponding fixed sample size and the effect size is non-zero, the initially planned power value of $1 - \beta$ serves as the target for the conditional power. Conversely, when circumstances differ, such as when the trial might not merit continuation to the second stage, the target values shift to the first stage's sample size n_1 and the global one-sided significance level α.

In terms of the variation components, the observed variation is juxtaposed against the maximum feasible variation within the specific context. Each of the four score components can assume values ranging from zero to one, permitting independent evaluation. Moreover, these components can be amalgamated into two sub-scores: the conditional power sub-score $SCP(\Delta)$ and the conditional sample size sub-score $SCN(\Delta)$, or be consolidated into a singular performance value, denoted by the conditional performance score (CPS). Mathematically, this score is calculated as

$$CPS(\Delta) = \frac{1}{2} \cdot [SCP(\Delta) + SCN(\Delta)]. \tag{9}$$

In evaluating all (sub-)scores and components, it is important to note that higher values are indicative of superior performance. We can give different levels of importance to the parts included in the two sub-scores by carefully deciding how much weight to assign to each. The components of $CPS(\Delta)$ can be derived as follows [25]. The conditional power sub-score is

$$SCP(\Delta) = \gamma_{\text{loc}} \cdot e_{_CP}(\Delta) + \gamma_{\text{var}} \cdot v_{_CP}(\Delta), \text{ with } \gamma_{\text{loc}} + \gamma_{\text{var}} = 1. \tag{10}$$

Here,γ_{loc} and γ_{var} represent the weights assigned to the location component e_{CP} and the variation component v_{CP}, respectively. A similar approach is taken for the conditional sample size sub-score. Here, we opt equal weighting of all components, which means $\gamma_{loc} = \gamma_{var} = 0.5$ while the results for $\gamma_{loc} = 0.2$ are also computed and depicted in Figures S20–S31 of the supplementary text. The fundamental concept underpinning the two location components, namely e_cp and e_n, is to assess and compare the calculated average conditional power and the calculated average conditional sample size against predefined target values. The target value for the sample size is specified as

$$N_{\text{target}} = \begin{cases} n^{fix}, & \text{if } n^{fix} \leq n_{max} \text{ and } \Delta \neq 0, \\ n_1, & \text{if } n^{fix} \geq n_{max} \text{ or } \Delta = 0, \end{cases} \tag{11}$$

where n^{fix} is required sample size in fixed study design. The target value for conditional power is as follows

$$CP_{\text{target}} = \begin{cases} 1-\beta, & \text{if } n^{fix} \leq n_{max} \text{ and } \Delta \neq 0, \\ \alpha, & \text{if } n^{fix} \geq n_{max} \text{ or } \Delta = 0, \end{cases} \tag{12}$$

where α is defined as global one-sided significance level. For the conditional sample size, the sub-score is defined as

$$SCP(\Delta) = \gamma_{\text{loc}}\left(1 - \frac{|E[CN_{\Delta}^{RA}(T_1)] - N_{\text{target}}|}{n_{\max} - n_1}\right) + \gamma_{\text{var}}\left(1 - \sqrt{\frac{\text{Var}\left(N_{\Delta}^{RA}(T_1)\right)}{\text{Var}_{\max}\left(N_{\Delta}^{RA}(T_1)\right)}}\right) \tag{13}$$

Thus, $\left(1 - \frac{|E[CN_{\Delta}^{RA}(T_1)] - N_{\text{target}}|}{n_{\max} - n_1}\right)$ = e_cp and $\left(1 - \sqrt{\frac{\text{Var}\left(N_{\Delta}^{RA}(T_1)\right)}{\text{Var}_{\max}\left(N_{\Delta}^{RA}(T_1)\right)}}\right)$ = v_cp, where γ_{loc} + γ_{var} =1. Both sub-scores, SCN and SCP, have a range of [0, 1]. These are determined by the degree to which certain conditions are met. Specifically, SCN is influenced by the closeness of the variation to its target value. Similarly, SCP is determined by the degree to which another set of conditions is met. Both SCN and SCP attain larger values when the respective variations are small, and the predefined target values are closely approached. Therefore, we can define the point-wise total conditional performance score as follows.

$$CN_{\text{target}} = \begin{cases} n_{\Delta}^{\text{fix}}, & \text{if } n_{\Delta}^{\text{fix}} \leq n_{\max} \text{ and } \Delta \neq 0 \\ n_1, & \text{if } n_{\Delta}^{\text{fix}} > n_{\max} \text{ or } \Delta = 0 \end{cases} \tag{14}$$

$$CP_{\text{target}} = \begin{cases} 1-\beta, & \text{if } n_{\Delta}^{\{\}\S} \leq n_{\max} \text{ and } \Delta \neq 0 \\ \alpha, & \text{if } n_{\Delta}^{\{\}\S} > n_{\max} \text{ or } \Delta = 0 \end{cases} \tag{15}$$

$$\text{SCN}(\Delta) = \gamma_{\text{loc}}\left(1 - \frac{\left|\mathbb{E}[\text{CN}_{\Delta}^{\text{RA}}(T_1)] - \text{CN}_{\text{target}}\right|}{n_{\max} - n_1}\right) + \gamma_{\text{var}}\left(1 - \sqrt{\frac{\text{Var}(\text{CN}_{\Delta}^{\text{RA}}(T_1))}{\text{Var}_{\max}(\text{CN}_{\Delta}^{\text{RA}}(T_1))}}\right) \tag{16}$$

Then,

$$CPS(\Delta) = \frac{1}{2}(SCP(\Delta) + SCN(\Delta)). \tag{17}$$

Resampling Approach for Recalculating Sample Size

To account for the potential variations in the interim effect, we may consider utilizing resampling as a technique to assess the fluctuation of a random variable. The application of the resampling approach is contingent upon the observed interim test statistic

falling within the designated recalculation area, indicating the proposition of a second stage. In this context, B test statistics are resampled from a normal distribution with the observed interim test statistic as the average and a standard deviation of one. It is important to note that the resampling procedure exclusively occurs if the observed interim test statistic aligns with the recalculation area, suggesting the feasibility of a subsequent stage. Consequently, all resampled test statistics, including those outside the recalculation area, contribute to the computation of the final value for the sample size of the second stage. This process unfolds as follows: for each of the B resampled test statistics, the second-stage sample size is reevaluated, resulting in an array of sample sizes denoted as $\tilde{n}_{(*),1}(t_1), \tilde{n}_{(*),2}(t_1), \tilde{n}_{(*),3}(t_1), \ldots, \tilde{n}_{(*),B}(t_1)$, where (*) signifies the index for the initial sample size-recalculation rule. It is important to acknowledge that some of these "recalculated" sample sizes may indeed correspond to the initial sample size n_1. In the final step, a comprehensive location metric summarizes the entire set of B sample sizes, ultimately determining the definitive value for the second-stage sample size. This methodology empowers us to effectively incorporate the variability of the interim effect into our decision-making process for sample size adjustment. In our exploration, we distinguish between two different ways:

1. The simpler approach involves setting the second stage sample size as the average of all the resampled sample sizes:

$$n_{(*)}^{R1}(t_1) = \frac{1}{B}\sum_{b=1}^{B}(\tilde{n}_{(*),i}(t_1)) \tag{18}$$

   We refer to this method as the R1 approach.
2. Considering that the initial sample size of the first stage can greatly influence the resampled sample sizes, we contemplate an alternative. Here, we compute the final second-stage sample size as the mean plus the standard deviation of the resampled sample sizes:

$$n_{(*)}^{R2}(t_1) = \frac{1}{B}\sum_{b=1}^{B}(\tilde{n}_{(*),i}(t_1)) + \frac{1}{B-1}\sqrt{\sum_{b=1}^{B}\left(\tilde{n}_{(*),i}(t_1) - \frac{1}{B}\sum_{b=1}^{B}(\tilde{n}_{(*),i}(t_1))\right)^2} \tag{19}$$

This inclusion of the standard deviation means that we tend to select larger sample sizes. We term this approach as the R2 method. It is worth noting that instead of incorporating the standard deviation, other measures of the distribution of resampled sample sizes (such as predefined quantiles) could also be used to achieve a similar effect. As such, the R2 approach serves as just one illustrative possibility within this context.

5. Simulation Study for Evaluating the Performance of Sample Size-Recalculation Approaches

To comprehensively assess the effectiveness of the various sample size-recalculation methods as outlined earlier, a comprehensive simulation study is conducted. These approaches are meticulously evaluated through specific performance measures, including the novel conditional performance score with parameter values of $\gamma_{loc} = \gamma_{var} = 0.5$.

In this simulation study, we work with groups of equal sizes, with the first stage consisting of $n_1 = 50$ participants. The initial second stage sample size per group is established as $n_2 = 50$, culminating in an initial total sample size of $n_{ini} = n_1 + n_2 = 100$. The maximum feasible sample size per group is set at four times the interim sample size n_1, resulting in $n_{max} = 200$. The weights for the inverse normal combination test are uniformly assigned as $w_1 = w_2 = \sqrt{50}$. We fixed the global one-sided significance level at $\alpha = 0.025$,

while the local significance levels are calculated according to Pocock [3] method, specifically, $\alpha_1 = \alpha = 0.0147$. Additionally, a futility bound of $\alpha_0 = 0.5$ is established.

A desired level of conditional power is set at $1 - \beta = 0.8$. For the ROCP, the lower bound for conditional power ($1 - \beta_{\text{low}}^{\text{ROCP}}$) is fixed at 0.6. Similarly, for the promising zone (PZ), a lower bound ($1 - \beta_{\text{low}}^{\text{PZ}}$) is fixed at 0.36, following the approach proposed by Mehta and Pocock [24]. To explore the performance of these designs across various scenarios, we considered a range of underlying true standardized treatment effects $\Delta \in \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5\}$. To ensure statistical robustness, each scenario underwent 10,000 simulation iterations. Notably, for the resampling methods, a total of $B = 5000$ samples are used. For the sake of comparison, a group sequential design (GS) employing $n_1 = n_2 = 50$ and employing the same decision boundaries as described above is also simulated and evaluated side by side. This extensive simulation endeavor facilitated a comprehensive exploration of the performance characteristics of the various sample size adjustment strategies under a wide array of circumstances.

For the simulation purpose, initially, we generated a treatment group from MEM and a control group from a normal distribution having $n = 50$, $\mu = 0.3$ and $\sigma = 1$. Also, we generated both treatment and control groups from MEM and run the simulation.

In the case of mixture distribution, we generated treatment and control groups of size $n = 50$ from the normal mixture. One distribution is taken as the standard normal distribution, i.e., N(0,1), while the other distribution is also taken as normal having $\mu = 0.5$ and $\sigma = 1$, i.e., N (0.5,1). However, the mixing proportion is chosen as 0.1,0.2 and 0.5, i.e., $\pi = 0.1,\ 0.2,\ 0.5$.

First, the results are discussed with the help of the conditional power score (CPS). A method with a higher value of the CPS is considered a better performer. In the standard sample size recalculation without resampling, one can see from Table 1 that group sequential design performed relatively well. The reason behind this is that there is no variation in recalculated sample sizes. If we access the performance of the R1 approach, this approach considers the mean as the summary location measure, and it performs better than the respective standard sample size-recalculation approach without resampling for CPS for all Δ, true standardized effect sizes. The reason behind the better performance of the R1 approach over the standard sample size-recalculation approach without resampling is that the resampling approach (R1 approach) reduces the variability in recalculated sample sizes for all Δ.

Furthermore, in the R1 approach, the OCP performed either better than the GS design or had a similar conditional performance score as compared to the ROCP and PZ (Table 1, columns 3 and 4) due to the better conditional power of the OCP (Tables 2 and 3). It is to be noted that there is an observable tendency towards increasing the initial stage's sample size, denoted as the mean of n_1 when utilizing sample size recalculation via the R1 approach. This phenomenon arises because, under the R1 approach, test statistics that fall outside the recalculated area (RA) could undergo resampling, even if the interim test statistic falls within RA. To address this concern, the R2 approach has been developed. This alternative approach relies on a distinct summary location measure, which is calculated as the mean of the resampled sample sizes along with its corresponding standard deviation. As it can be seen in Table 4 the R2 approach tends to move towards GS design due to n_{max} as the sample size boundary.

In the R2 approach, the ROCP performed well for almost all Δ for the CPS (Table 1, column 5) compared to the OCP and PZ approach. Overall, the R2 approaches secured a distinguished position against the original sample size recalculation without resampling, However, the GS design outperforms the different designs in the R2 approach for all Δ. Note that the CPS of respective designs in the R2 approach is worse than the R1 approach

for the same values of $\Delta \in (0.0, 0.1, 0.2, 0.3, 0.4, 0.5)$. The reason behind this worst behavior of the R2 approach is that the sample size-recalculation designs in the R2 approach do not tend towards the target values of sample size effectively, this can be seen in the worst values of conditional sample size sub-score SCN (Tables 3 and 4). The results also indicate that a larger mixing proportion tends to favor the R2 approach, while the R1 approach performs better when the mixing proportion is small. The results for mixing proportion $\pi = 0.2$ are given the supplementary text.

Overall, the results indicate that the R1 approach is the most effective method for improving CPS across various designs and effect sizes, while the R2 approach offers some advantages but does not consistently outperform the original methods or the R1 approach. The R1 method's ability to maintain a balanced sample size contributes to its superior performance, while the R2 approach's tendency to approach maximum sample sizes may hinder its effectiveness, particularly for smaller true underlying effect sizes.

Table 1. Conditional power score with both treatment and control generated from MDM with $\pi = 0.5$.

Δ	Design	Standard Simulation	R1 Approach	R2 Approach
0	OCP	0.378	0.4736	0.5082
0	ROCP	0.4255	0.6096	0.6604
0	PZ	0.4978	0.6511	0.6684
0	GSD	0.6741	0.7764	0.7764
0.1	OCP	0.378	0.4300	0.4654
0.1	ROCP	0.4255	0.5403	0.6174
0.1	PZ	0.4978	0.5954	0.6282
0.1	GSD	0.6741	0.7422	0.7422
0.2	OCP	0.378	0.3981	0.4306
0.2	ROCP	0.4255	0.4798	0.5820
0.2	PZ	0.4978	0.549	0.5943
0.2	GSD	0.6741	0.7105	0.7104
0.3	OCP	0.6242	0.6208	0.6923
0.3	ROCP	0.4072	0.3898	0.6232
0.3	PZ	0.5381	0.5272	0.6521
0.3	GSD	0.6173	0.6099	0.6099
0.4	OCP	0.5279	0.5522	0.6015
0.4	ROCP	0.5257	0.5443	0.6878
0.4	PZ	0.6069	0.6222	0.7001
0.4	GSD	0.7443	0.7562	0.7562
0.5	OCP	0.4685	0.5407	0.5836
0.5	ROCP	0.4662	0.5221	0.6644
0.5	PZ	0.5475	0.5923	0.6740
0.5	GSD	0.685	0.7212	0.7212

Table 2. Standard simulation with both treatment and control generated from MDM with $\pi = 0.5$.

Δ	Design	$\mathbb{E}[CP_\Delta^{RA}]$	$e_{CP}(\Delta)$	$Var(CP_\Delta^{RA})$	$v_{CP}(\Delta)$	$SCP(\Delta)$	$\mathbb{E}[CN_\Delta^{RA}]$	$e_{CN}(\Delta)$	$Var(CN_\Delta^{RA})$	$v_{CN}(\Delta)$	$SCN(\Delta)$	$CPS(\Delta)$
0	OCP	0.5421	0.4696	0.0924	0.3922	0.4309	167.0027	0.2200	1825.7502	0.4303	0.3251	0.3780
0	ROCP	0.4451	0.5691	0.1504	0.2244	0.3968	102.5864	0.6494	3087.8035	0.2591	0.4543	0.4255
0	PZ	0.4626	0.5512	0.1244	0.2946	0.4229	117.4977	0.5500	921.0116	0.5954	0.5727	0.4978
0	GSD	0.3869	0.6288	0.0897	0.4010	0.5149	100.0000	0.6667	0.0000	1.0000	0.8333	0.6741
0.1	OCP	0.5421	0.4696	0.0924	0.3922	0.4309	167.0027	0.2200	1825.7502	0.4303	0.3251	0.3780
0.1	ROCP	0.4451	0.5691	0.1504	0.2244	0.3968	102.5864	0.6494	3087.8035	0.2591	0.4543	0.4255
0.1	PZ	0.4626	0.5512	0.1244	0.2946	0.4229	117.4977	0.5500	921.0116	0.5954	0.5727	0.4978
0.1	GSD	0.3869	0.6288	0.0897	0.4010	0.5149	100.0000	0.6667	0.0000	1.0000	0.8333	0.6741
0.2	OCP	0.5421	0.4696	0.0924	0.3922	0.4309	167.0027	0.2200	1825.7502	0.4303	0.3251	0.3780
0.2	ROCP	0.4451	0.5691	0.1504	0.2244	0.3968	102.5864	0.6494	3087.8035	0.2591	0.4543	0.4255
0.2	PZ	0.4626	0.5512	0.1244	0.2946	0.4229	117.4977	0.5500	921.0116	0.5954	0.5727	0.4978
0.2	GSD	0.3869	0.6288	0.0897	0.4010	0.5149	100.0000	0.6667	0.0000	1.0000	0.8333	0.6741
0.3	OCP	0.5421	0.7355	0.0924	0.3922	0.5639	167.0027	0.9386	1825.7502	0.4303	0.6845	0.6242
0.3	ROCP	0.4451	0.6360	0.1504	0.2244	0.4302	102.5864	0.5092	3087.8035	0.2591	0.3842	0.4072
0.3	PZ	0.4626	0.6539	0.1244	0.2946	0.4743	117.4977	0.6086	921.0116	0.5954	0.6020	0.5381
0.3	GSD	0.3869	0.5764	0.0897	0.4010	0.4887	100.0000	0.4919	0.0000	1.0000	0.7450	0.6173
0.4	OCP	0.5421	0.7355	0.0924	0.3922	0.5639	167.0027	0.2200	1825.7502	0.4303	0.4919	0.5279
0.4	ROCP	0.4451	0.6360	0.1504	0.2244	0.4302	102.5864	0.9831	3087.8035	0.2591	0.6211	0.5257
0.4	PZ	0.4626	0.6539	0.1244	0.2946	0.4743	117.4977	0.8837	921.0116	0.5954	0.7395	0.6069
0.4	GSD	0.3869	0.5764	0.0897	0.4010	0.4887	100.0000	0.9996	0.0000	1.0000	0.9998	0.7443
0.5	OCP	0.5421	0.7355	0.0924	0.3922	0.5639	167.0027	0.3160	1825.7502	0.4303	0.3731	0.4685
0.5	ROCP	0.4451	0.6360	0.1504	0.2244	0.4302	102.5864	0.7454	3087.8035	0.2591	0.5023	0.4662
0.5	PZ	0.4626	0.6539	0.1244	0.2946	0.4743	117.4977	0.6460	921.0116	0.5954	0.6207	0.5475
0.5	GSD	0.3869	0.5764	0.0897	0.4010	0.4887	100.0000	0.7626	0.0000	1.0000	0.8813	0.6850

Table 3. R1 approach with both treatment and control generated from MDM with $\pi = 0.5$.

Δ	Design	$\mathbb{E}[CP_{\Delta}^{RA}]$	$e_{CP}(\Delta)$	$Var(CP_{\Delta}^{RA})$	$v_{CP}(\Delta)$	$SCP(\Delta)$	$\mathbb{E}[CN_{\Delta}^{RA}]$	$e_{CN}(\Delta)$	$Var(CN_{\Delta}^{RA})$	$v_{CN}(\Delta)$	$SCN(\Delta)$	$CPS(\Delta)$
0	OCP	0.2609	0.7580	0.0877	0.4077	0.5828	192.0012	0.0533	592.1704	0.6756	0.3644	0.4736
0	ROCP	0.1579	0.8637	0.0960	0.3803	0.6220	73.1073	0.8460	2388.8495	0.3484	0.5972	0.6096
0	PZ	0.1801	0.8409	0.0762	0.4480	0.6445	107.3459	0.6177	513.5704	0.6979	0.6578	0.6511
0	GSD	0.1493	0.8725	0.0470	0.5666	0.7196	100.0000	0.6667	0.0000	1.0000	0.8333	0.7764
0.1	OCP	0.3398	0.6771	0.1023	0.3604	0.5188	186.2969	0.0914	940.2896	0.5912	0.3413	0.4300
0.1	ROCP	0.2322	0.7875	0.1257	0.2909	0.5392	81.5523	0.7897	2810.1827	0.2932	0.5414	0.5403
0.1	PZ	0.2543	0.7648	0.1018	0.3618	0.5633	110.4791	0.5968	657.4211	0.6581	0.6275	0.5954
0.1	GSD	0.2096	0.8107	0.0647	0.4913	0.6510	100.0000	0.6667	0.0000	1.0000	0.8333	0.7422
0.2	OCP	0.4232	0.5916	0.1058	0.3496	0.4706	179.2309	0.1385	1335.0099	0.5129	0.3257	0.3981
0.2	ROCP	0.3178	0.6997	0.1460	0.2360	0.4678	90.8845	0.7274	3112.3824	0.2562	0.4918	0.4798
0.2	PZ	0.3361	0.6810	0.1195	0.3088	0.4949	113.4007	0.5773	775.1167	0.6288	0.6031	0.5490
0.2	GSD	0.2784	0.7401	0.0798	0.4352	0.5876	100.0000	0.6667	0.0000	1.0000	0.8333	0.7105
0.3	OCP	0.5090	0.7015	0.0992	0.3701	0.5358	170.7438	0.9636	1713.0396	0.4482	0.7059	0.6208
0.3	ROCP	0.4099	0.5999	0.1524	0.2194	0.4097	99.7227	0.4901	3166.8077	0.2497	0.3699	0.3898
0.3	PZ	0.4269	0.6174	0.1261	0.2897	0.4535	116.5629	0.6024	903.0076	0.5994	0.6009	0.5272
0.3	GSD	0.3556	0.5442	0.0889	0.4036	0.4739	100.0000	0.4919	0.0000	1.0000	0.7450	0.6099
0.4	OCP	0.5868	0.7814	0.0829	0.4243	0.6028	160.4308	0.5975	1987.2072	0.4057	0.5016	0.5522
0.4	ROCP	0.5027	0.6951	0.1419	0.2467	0.4709	106.8062	0.9550	2911.7071	0.2806	0.6178	0.5443
0.4	PZ	0.5168	0.7095	0.1194	0.3089	0.5092	118.4771	0.8772	931.1456	0.5932	0.7352	0.6222
0.4	GSD	0.4365	0.6272	0.0906	0.3980	0.5126	100.0000	0.9996	0.0000	1.0000	0.9998	0.7562
0.5	OCP	0.6525	0.8487	0.0617	0.5031	0.6759	150.6506	0.4250	2120.4296	0.3861	0.4055	0.5407
0.5	ROCP	0.5832	0.7776	0.1194	0.3090	0.5433	111.9700	0.6828	2608.7351	0.3190	0.5009	0.5221
0.5	PZ	0.5970	0.7918	0.1023	0.3604	0.5761	119.8581	0.6302	961.3278	0.5866	0.6084	0.5923
0.5	GSD	0.5105	0.7031	0.0843	0.4192	0.5612	100.0000	0.7626	0.0000	1.0000	0.8813	0.7212

Table 4. R2 approach with both treatment and control generated from MDM with $\pi = 0.5$.

Δ	Design	$\mathbb{E}[CP_{\Delta}^{RA}]$	$e_{CP}(\Delta)$	$Var(CP_{\Delta}^{RA})$	$v_{CP}(\Delta)$	$SCP(\Delta)$	$\mathbb{E}[CN_{\Delta}^{RA}]$	$e_{CN}(\Delta)$	$Var(CN_{\Delta}^{RA})$	$v_{CN}(\Delta)$	$SCN(\Delta)$	$CPS(\Delta)$
0	OCP	0.2759	0.7427	0.1058	0.3494	0.5460	197.8248	0.0145	30.7337	0.9261	0.4703	0.5082
0	ROCP	0.1992	0.8213	0.0739	0.4564	0.6389	119.5676	0.5362	167.0806	0.8277	0.6819	0.6604
0	PZ	0.2027	0.8178	0.0735	0.4577	0.6378	129.5780	0.4695	28.7307	0.9285	0.6990	0.6684
0	GSD	0.1493	0.8725	0.0470	0.5666	0.7196	100.0000	0.6667	0.0000	1.0000	0.8333	0.7764
0.1	OCP	0.3655	0.6508	0.1290	0.2816	0.4662	196.9414	0.0204	46.8099	0.9088	0.4646	0.4654
0.1	ROCP	0.2752	0.7434	0.0975	0.3756	0.5595	122.5617	0.5163	154.5922	0.8342	0.6752	0.6174
0.1	PZ	0.2786	0.7399	0.0966	0.3785	0.5592	130.6722	0.4622	25.9024	0.9321	0.6972	0.6282
0.1	GSD	0.2096	0.8107	0.0647	0.4913	0.6509	100.0000	0.6667	0.0000	1.0000	0.8333	0.7422
0.2	OCP	0.4617	0.5521	0.1391	0.2541	0.4031	195.5099	0.0299	72.7119	0.8863	0.4581	0.4306
0.2	ROCP	0.3594	0.6571	0.1136	0.3259	0.4915	125.3510	0.4977	130.9209	0.8474	0.6726	0.5820
0.2	PZ	0.3626	0.6538	0.1122	0.3300	0.4919	131.5675	0.4562	22.1661	0.9372	0.6967	0.5943
0.2	GSD	0.2784	0.7401	0.0798	0.4352	0.5876	100.0000	0.6667	0.0000	1.0000	0.8333	0.7105
0.3	OCP	0.5627	0.7566	0.1359	0.2627	0.5097	193.6949	0.8834	100.3212	0.8665	0.8749	0.6923
0.3	ROCP	0.4517	0.6428	0.1197	0.3080	0.4754	127.7445	0.6769	102.5951	0.8650	0.7709	0.6232
0.3	PZ	0.4544	0.6456	0.1179	0.3133	0.4794	132.2288	0.7068	18.5098	0.9426	0.8247	0.6521
0.3	GSD	0.3556	0.5442	0.0889	0.4036	0.4739	100.0000	0.4919	0.0000	1.0000	0.7460	0.6099
0.4	OCP	0.6591	0.8555	0.1199	0.3076	0.5815	191.4547	0.3907	122.8456	0.8522	0.6214	0.6015
0.4	ROCP	0.5450	0.7385	0.1150	0.3217	0.5301	129.6691	0.8026	70.1434	0.8883	0.8455	0.6878
0.4	PZ	0.5471	0.7406	0.1130	0.3277	0.5342	132.6528	0.7827	14.5122	0.9492	0.8659	0.7001
0.4	GSD	0.4365	0.6272	0.0906	0.3980	0.5126	100.0000	0.9996	0.0000	1.0000	0.9998	0.7562
0.5	OCP	0.7418	0.9403	0.0945	0.3851	0.6627	189.0689	0.1688	143.8981	0.8401	0.5045	0.5836
0.5	ROCP	0.6276	0.8232	0.0995	0.3691	0.5961	130.8612	0.5569	46.9952	0.9086	0.7327	0.6644
0.5	PZ	0.6293	0.8249	0.0976	0.3751	0.5999	132.7406	0.5444	13.1325	0.9517	0.7480	0.6740
0.5	GSD	0.5105	0.7031	0.0843	0.4192	0.5612	100.0000	0.7626	0.0000	1.0000	0.8813	0.7212

For the standard sample size recalculation, the GSD outperformed the rest (Table 5). This is due to the fact that there is no variation in recalculated sample sizes. The R1 approach performed well as compared to standard sample size recalculation for all Δ values in all designs. The reason behind the better performance of the R1 approach against standard simulation is that the resampling approach (R1 approach) reduces the variability in recalculated sample sizes for all Δ (Table 6 vs. Table 7). Furthermore, in the R1 approach, the ROCP and PZ have approximately the same CPS as the GSD, while the OCP has a smaller CPS than the GSD. This is due to a smaller conditional power of the OCP. While recalculating sample size using the R1 approach, this approach has a tendency to increase initial sample size, denoted as n_1 (7). This is due to the reason that if the test statistics fall outside the recalculation area, it could go through resampling, even if interim test statistics fall within the recalculated area (RA). To overcome this issue, the R2 approach is introduced, which relies on the mean as well as standard deviation of resampled sample sizes as the summary location measure. The effect of this can be seen in (8). The R2 approach tends to recalculate sample size up to the maximum allowed sample size n_{max}. Overall, the R2 approach has secured a distinct position against the R1 approach and standard simulation. However, the GSD in standard simulation outperformed various designs in the R2 approach for all Δ. The detailed simulation results are presented in Tables 6–8. The box plots for second-stage conditional power and recalculated sample size for standard simulation, R1 approach and R2 approach, are illustrated in Figures S1–S12, respectively.

Table 5. Conditional power score with both treatment and control generated from MDM with $\pi = 0.1$.

Δ	Design	Standard Simulation	R1 Approach	R2 Approach
0	OCP	0.3781	0.5479	0.3963
0	ROCP	0.4160	0.7264	0.5437
0	PZ	0.4902	0.6603	0.5554
0	GSD	0.6700	0.6700	0.6700
0.1	OCP	0.3781	0.5480	0.3963
0.1	ROCP	0.4160	0.7266	0.5440
0.1	PZ	0.4902	0.6603	0.5554
0.1	GSD	0.6700	0.6700	0.6700
0.2	OCP	0.3781	0.5478	0.3963
0.2	ROCP	0.4160	0.7266	0.5439
0.2	PZ	0.4902	0.6603	0.5553
0.2	GSD	0.6700	0.6700	0.6700
0.3	OCP	0.6284	0.6485	0.7173
0.3	ROCP	0.4213	0.5714	0.6488
0.3	PZ	0.4902	0.6156	0.6729
0.3	GSD	0.6231	0.6231	0.6231
0.4	OCP	0.5391	0.6759	0.5904
0.4	ROCP	0.5298	0.6984	0.6784
0.4	PZ	0.6114	0.7391	0.6911
0.4	GSD	0.7501	0.7501	0.7501
0.5	OCP	0.4797	0.6164	0.5310
0.5	ROCP	0.4703	0.7069	0.6190
0.5	PZ	0.5520	0.6797	0.6317
0.5	GSD	0.6908	0.6908	0.6908

Table 6. Standard simulation with both treatment and control generated from MDM with $\pi = 0.1$.

Δ	Design	$\mathbb{E}[CP_{\Delta}^{RA}]$	$e_{CP}(\Delta)$	$Var(CP_{\Delta}^{RA})$	$v_{CP}(\Delta)$	$SCP(\Delta)$	$\mathbb{E}[CN_{\Delta}^{RA}]$	$e_{CN}(\Delta)$	$Var(CN_{\Delta}^{RA})$	$v_{CN}(\Delta)$	$SCN(\Delta)$	$CPS(\Delta)$
0	OCP	0.5638	0.4474	0.0877	0.4079	0.4276	164.9150	0.2339	1871.2579	0.4233	0.3286	0.3781
0	ROCP	0.4717	0.5418	0.1469	0.2335	0.3876	105.5823	0.6295	3087.6155	0.2592	0.4443	0.4160
0	PZ	0.4862	0.5270	0.1218	0.3019	0.4145	118.4798	0.5435	952.6482	0.5885	0.5660	0.4902
0	GSD	0.4063	0.6090	0.0887	0.4044	0.5067	100.0000	0.6667	0.0000	1.0000	0.8333	0.6700
0.1	OCP	0.5638	0.4474	0.0877	0.4079	0.4276	164.9150	0.2339	1871.2579	0.4233	0.3286	0.3781
0.1	ROCP	0.4717	0.5418	0.1469	0.2335	0.3876	105.5823	0.6295	3087.6155	0.2592	0.4443	0.4160
0.1	PZ	0.4862	0.5270	0.1218	0.3019	0.4145	118.4798	0.5435	952.6482	0.5885	0.5660	0.4902
0.1	GSD	0.4063	0.6090	0.0887	0.4044	0.5067	100.0000	0.6667	0.0000	1.0000	0.8333	0.6700
0.2	OCP	0.5638	0.4474	0.0877	0.4079	0.4276	164.9150	0.2339	1871.2579	0.4233	0.3286	0.3781
0.2	ROCP	0.4717	0.5418	0.1469	0.2335	0.3876	105.5823	0.6295	3087.6155	0.2592	0.4443	0.4160
0.2	PZ	0.4862	0.5270	0.1218	0.3019	0.4145	118.4798	0.5435	952.6482	0.5885	0.5660	0.4902
0.2	GSD	0.4063	0.6090	0.0887	0.4044	0.5067	100.0000	0.6667	0.0000	1.0000	0.8333	0.6700
0.3	OCP	0.5638	0.7577	0.0877	0.4079	0.5828	164.9150	0.9247	1871.2579	0.4233	0.6740	0.6284
0.3	ROCP	0.4717	0.6633	0.1469	0.2335	0.4484	105.5823	0.5292	3087.6155	0.2592	0.3942	0.4213
0.3	PZ	0.4862	0.6781	0.1218	0.3019	0.4900	118.4798	0.6151	952.6482	0.5885	0.6018	0.4902
0.3	GSD	0.4063	0.5962	0.0887	0.4044	0.5003	100.0000	0.4919	0.0000	1.0000	0.7460	0.6231
0.4	OCP	0.5638	0.7577	0.0877	0.4079	0.5828	164.9150	0.5676	1871.2579	0.4233	0.4954	0.5391
0.4	ROCP	0.4717	0.6633	0.1469	0.2335	0.4484	105.5823	0.9632	3087.6155	0.2592	0.6112	0.5298
0.4	PZ	0.4862	0.6781	0.1218	0.3019	0.4900	118.4798	0.8772	952.6482	0.5885	0.7328	0.6114
0.4	GSD	0.4063	0.5962	0.0887	0.4045	0.5003	100.0000	0.9996	0.0000	1.0000	0.9998	0.7501
0.5	OCP	0.5638	0.7577	0.0877	0.4079	0.5828	164.9150	0.3299	1871.2579	0.4233	0.3766	0.4797
0.5	ROCP	0.4717	0.6633	0.1469	0.2335	0.4484	105.5823	0.7254	3087.6155	0.2592	0.4923	0.4703
0.5	PZ	0.4862	0.6781	0.1218	0.3019	0.4900	118.4798	0.6394	952.6482	0.5885	0.6140	0.5520
0.5	GSD	0.4063	0.5963	0.0887	0.4044	0.5003	100.0000	0.7626	0.0000	1.0000	0.8813	0.6908

Table 7. R1 approach with both treatment and control generated from MDM with $\pi = 0.1$.

Δ	Design	$\mathbb{E}[CP_{\Delta}^{RA}]$	$e_{CP}(\Delta)$	$Var(CP_{\Delta}^{RA})$	$v_{CP}(\Delta)$	$SCP(\Delta)$	$\mathbb{E}[CN_{\Delta}^{RA}]$	$e_{CN}(\Delta)$	$Var(CN_{\Delta}^{RA})$	$v_{CN}(\Delta)$	$SCN(\Delta)$	$CPS(\Delta)$
0	OCP	0.4993	0.5136	0.0969	0.3775	0.4455	129.9407	0.4671	155.9233	0.8335	0.6503	0.5479
0	ROCP	0.3276	0.6896	0.0653	0.4891	0.5894	79.6770	0.8022	31.8719	0.9247	0.8634	0.7264
0	PZ	0.4035	0.6118	0.0799	0.4347	0.5233	101.0799	0.6595	23.5975	0.9352	0.7974	0.6603
0	GSD	0.4063	0.6090	0.0887	0.4044	0.5067	100.0000	0.6667	0.0000	1.0000	0.8333	0.6700
0.1	OCP	0.4992	0.5136	0.0969	0.3776	0.4456	129.9226	0.4672	155.9836	0.8335	0.6503	0.5480
0.1	ROCP	0.3274	0.6898	0.0652	0.4894	0.5896	79.6364	0.8024	31.7193	0.9249	0.8637	0.7266
0.1	PZ	0.4035	0.6118	0.0799	0.4347	0.5233	101.0741	0.6595	23.6070	0.9352	0.7974	0.6603
0.1	GSD	0.4063	0.6090	0.0887	0.4044	0.5067	100.0000	0.6667	0.0000	1.0000	0.8333	0.6700
0.2	OCP	0.4994	0.5135	0.0969	0.3775	0.4455	129.9531	0.4670	156.3103	0.8333	0.6501	0.5478
0.2	ROCP	0.3275	0.6897	0.0652	0.4893	0.5895	79.6552	0.8023	31.7583	0.9249	0.8636	0.7266
0.2	PZ	0.4035	0.6118	0.0799	0.4348	0.5233	101.0739	0.6595	23.7079	0.9351	0.7973	0.6603
0.2	GSD	0.4063	0.6090	0.0887	0.4044	0.5067	100.0000	0.6667	0.0000	1.0000	0.8333	0.6700
0.3	OCP	0.4993	0.6915	0.0969	0.3775	0.5345	129.9278	0.6915	155.5575	0.8337	0.7626	0.6485
0.3	ROCP	0.3276	0.5155	0.0653	0.4890	0.5022	79.6738	0.3564	31.8717	0.9247	0.6406	0.5714
0.3	PZ	0.4035	0.5933	0.0799	0.4348	0.5140	101.0767	0.4991	23.6722	0.9351	0.7171	0.6156
0.3	GSD	0.4063	0.5962	0.0887	0.4044	0.5003	100.0000	0.4919	0.0000	1.0000	0.7460	0.6231
0.4	OCP	0.4993	0.6916	0.0969	0.3775	0.5345	129.9513	0.8007	155.6195	0.8337	0.8172	0.6759
0.4	ROCP	0.3275	0.5154	0.0652	0.4892	0.5023	79.6554	0.8640	31.7655	0.9249	0.8944	0.6984
0.4	PZ	0.4035	0.5933	0.0799	0.4347	0.5140	101.0715	0.9932	23.6154	0.9352	0.9642	0.7391
0.4	GSD	0.4063	0.5962	0.0887	0.4044	0.5003	100.0000	0.9996	0.0000	1.0000	0.9998	0.7501
0.5	OCP	0.4994	0.6916	0.0969	0.3774	0.5345	129.9593	0.5629	155.2517	0.8339	0.6984	0.6165
0.5	ROCP	0.3276	0.5155	0.0653	0.4891	0.5023	79.6695	0.8982	31.6864	0.9249	0.9116	0.7070
0.5	PZ	0.4035	0.5933	0.0799	0.4348	0.5141	101.0786	0.7554	23.7130	0.9351	0.8453	0.6797
0.5	GSD	0.4063	0.5962	0.0887	0.40444	0.5003	100.0000	0.7626	0.0000	1.0000	0.8813	0.6908

Table 8. R2 approach with both treatment and control generated from MDM with $\pi = 0.1$.

Δ	Design	$\mathbb{E}[CP_{\Delta}^{RA}]$	$e_{CP}(\Delta)$	$Var(CP_{\Delta}^{RA})$	$v_{CP}(\Delta)$	$SCP(\Delta)$	$\mathbb{E}[CN_{\Delta}^{RA}]$	$e_{CN}(\Delta)$	$Var(CN_{\Delta}^{RA})$	$v_{CN}(\Delta)$	$SCN(\Delta)$	$CPS(\Delta)$
0	OCP	0.6282	0.3814	0.1240	0.2958	0.3386	192.5196	0.0499	113.2262	0.8581	0.4540	0.3963
0	ROCP	0.5122	0.5003	0.1152	0.3211	0.4107	129.2916	0.4714	78.1722	0.8821	0.6768	0.5437
0	PZ	0.5146	0.4979	0.1133	0.3268	0.4124	132.6754	0.4488	15.2234	0.9480	0.6984	0.5554
0	GSD	0.4063	0.6090	0.0887	0.4044	0.5067	100.0000	0.6667	0.0000	1.0000	0.8333	0.6700
0.1	OCP	0.6282	0.3814	0.1240	0.2959	0.3386	192.5073	0.0499	113.3504	0.8581	0.4540	0.3963
0.1	ROCP	0.5120	0.5005	0.1152	0.3211	0.4108	129.2288	0.4718	77.7308	0.8825	0.6771	0.5440
0.1	PZ	0.5145	0.4979	0.1133	0.3268	0.4124	132.6560	0.4490	15.3417	0.9478	0.6984	0.5554
0.1	GSD	0.4063	0.6090	0.0887	0.4044	0.5067	100.0000	0.6667	0.0000	1.0000	0.8333	0.6700
0.2	OCP	0.6282	0.3814	0.1240	0.2958	0.3386	192.5183	0.0499	112.9620	0.8583	0.4541	0.3963
0.2	ROCP	0.5121	0.5004	0.1152	0.3212	0.4108	129.2657	0.4716	77.6728	0.8825	0.6770	0.5439
0.2	PZ	0.5146	0.4979	0.1133	0.3268	0.4124	132.6547	0.4490	15.5099	0.9475	0.6982	0.5553
0.2	GSD	0.4063	0.6090	0.0887	0.4044	0.5067	100.0000	0.6667	0.0000	1.0000	0.8333	0.6700
0.3	OCP	0.6282	0.8237	0.1240	0.2959	0.5598	192.5155	0.8913	112.9250	0.8583	0.8748	0.7173
0.3	ROCP	0.5122	0.7048	0.1153	0.3211	0.5129	129.2884	0.6872	78.1121	0.8822	0.7847	0.6488
0.3	PZ	0.5146	0.7072	0.1133	0.3268	0.5170	132.6620	0.7097	15.3983	0.9477	0.8287	0.6729
0.3	GSD	0.4063	0.5962	0.0887	0.4044	0.5003	100.0000	0.4919	0.0000	1.0000	0.7460	0.6231
0.4	OCP	0.6282	0.8238	0.1240	0.2958	0.5598	192.5257	0.3835	112.8584	0.8584	0.6209	0.5904
0.4	ROCP	0.5121	0.7048	0.1153	0.3210	0.5129	129.2641	0.8053	77.7168	0.8825	0.8439	0.6784
0.4	PZ	0.5145	0.7072	0.1133	0.3268	0.5170	132.6471	0.7827	15.4134	0.9477	0.8652	0.6911
0.4	GSD	0.4063	0.5962	0.0887	0.4044	0.5003	100.0000	0.9996	0.0000	1.0000	0.9998	0.7501
0.5	OCP	0.6282	0.8238	0.1240	0.2958	0.5598	192.5379	0.1457	112.3631	0.8587	0.5022	0.5310
0.5	ROCP	0.5121	0.70477	0.1152	0.3210	0.5129	129.2849	0.5674	77.5956	0.8826	0.7250	0.6190
0.5	PZ	0.5145	0.7072	0.1133	0.3269	0.5171	132.6613	0.5449	15.2100	0.9480	0.7464	0.6317
0.5	GSD	0.4063	0.5962	0.0887	0.4044	0.5003	100.0000	0.7626	0.0000	1.0000	0.8813	0.6908

6. Results from Measurement Error Model (MEM)

From Table 9, one can see that the GSD is the performance winner in standard simulation without resampling. This is due to the fact that recalculated sample sizes have no variation. Moreover, it is interesting to note that the CPS in standard simulation remained the same for $\Delta = \{0, 0.1, 0.2, 0.3\}$. In the R1 approach, the ROCP and PZ have approximately the same CPS, while the OCP has a smaller CPS value. Furthermore, the performance of the R1 approach is better against the standard simulation. This is due to decreased variation by the R1 approach in recalculated sample size. The performance of the R1 approach and R2 approach is the same for all Δ. This is mainly due to the generation of treatment and control groups from the MEM. The detailed simulation results are presented in Tables 10–12. The box plots for second stage conditional power and recalculated sample size for standard simulation, R1 approach and R2 approach are illustrated in Figures S13–S17, respectively.

Table 9. Conditional power score with both treatment and control generated from MEM.

Δ	Design	Standard Simulation	R1 Approach	R2 Approach
0	OCP	0.4355	0.6210	0.6210
0	ROCP	0.5436	0.7960	0.7960
0	PZ	0.5997	0.7329	0.7329
0	GSD	0.7475	0.7475	0.7475
0.1	OCP	0.4355	0.6209	0.6209
0.1	ROCP	0.5436	0.7960	0.7960
0.1	PZ	0.5997	0.7330	0.7329
0.1	GSD	0.7475	0.7475	0.7475
0.2	OCP	0.4355	0.6209	0.6209
0.2	ROCP	0.5436	0.7960	0.7960
0.2	PZ	0.5997	0.7329	0.7329
0.2	GSD	0.7475	0.7475	0.7475
0.3	OCP	0.6045	0.6170	0.6170
0.3	ROCP	0.3408	0.5380	0.5380
0.3	PZ	0.5050	0.5939	0.5940
0.3	GSD	0.5954	0.5954	0.5954
0.4	OCP	0.4776	0.6300	0.6300
0.4	ROCP	0.4677	0.6650	0.6650
0.4	PZ	0.5974	0.7095	0.7095
0.4	GSD	0.7223	0.7223	0.7223
0.5	OCP	0.4181	0.5706	0.5706
0.5	ROCP	0.4710	0.6888	0.6888
0.5	PZ	0.5380	0.6501	0.6501
0.5	GSD	0.6631	0.6631	0.6631

Table 10. Standard simulation results with both treatment and control generated from MEM.

Δ	Design	$\mathbb{E}[CP_{\Delta}^{RA}]$	$e_{CP}(\Delta)$	$Var(CP_{\Delta}^{RA})$	$v_{CP}(\Delta)$	$SCP(\Delta)$	$\mathbb{E}[CN_{\Delta}^{RA}]$	$e_{CN}(\Delta)$	$Var(CN_{\Delta}^{RA})$	$v_{CN}(\Delta)$	$SCN(\Delta)$	$CPS(\Delta)$
0	OCP	0.3319	0.6852	0.1003	0.3668	0.5260	187.4409	0.0837	872.6132	0.6062	0.3449	0.4355
0	ROCP	0.2240	0.7959	0.1226	0.2997	0.5478	81.2458	0.7917	2857.3641	0.2873	0.5395	0.5436
0	PZ	0.2453	0.7740	0.0987	0.3717	0.5729	110.4077	0.5973	666.4830	0.6558	0.6265	0.5997
0	GSD	0.2011	0.8194	0.0615	0.5039	0.6617	100.0000	0.6667	0.0000	1.0000	0.8333	0.7475
0.1	OCP	0.3319	0.6852	0.1003	0.3668	0.5260	187.4409	0.0837	872.6132	0.6062	0.3449	0.4355
0.1	ROCP	0.2240	0.7959	0.1226	0.2997	0.5478	81.2458	0.7917	2857.3641	0.2873	0.5395	0.5436
0.1	PZ	0.2453	0.7740	0.0987	0.3717	0.5729	110.4077	0.5973	666.4830	0.6558	0.6265	0.5997
0.1	GSD	0.2011	0.8194	0.0615	0.5039	0.6617	100.0000	0.6667	0.0000	1.0000	0.8333	0.7475
0.2	OCP	0.3319	0.6852	0.1003	0.3668	0.5260	187.4409	0.0837	872.6132	0.6062	0.3449	0.4355
0.2	ROCP	0.2240	0.7959	0.1226	0.2997	0.5478	81.2458	0.7917	2857.3641	0.2873	0.5395	0.5436
0.2	PZ	0.2453	0.7740	0.0987	0.3717	0.5729	110.4077	0.5973	666.4830	0.6558	0.6265	0.5997
0.2	GSD	0.2011	0.8194	0.0615	0.5039	0.6617	100.0000	0.6667	0.0000	1.0000	0.8333	0.7475
0.3	OCP	0.3319	0.5199	0.1003	0.3668	0.4434	187.4409	0.9251	872.6132	0.6062	0.7656	0.6045
0.3	ROCP	0.2240	0.4092	0.1226	0.2997	0.3544	81.2458	0.3669	2857.3641	0.2873	0.3271	0.3408
0.3	PZ	0.2453	0.4311	0.0987	0.3717	0.4014	110.4077	0.5613	666.4830	0.6558	0.6086	0.5050
0.3	GSD	0.2011	0.3858	0.0615	0.5039	0.4448	100.0000	0.4919	0.0000	1.0000	0.7450	0.5954
0.4	OCP	0.3319	0.5199	0.1003	0.3668	0.4434	187.4409	0.4174	872.6132	0.6062	0.5118	0.4776
0.4	ROCP	0.2240	0.4092	0.1226	0.2997	0.3544	81.2458	0.8746	2857.3641	0.2873	0.5810	0.4677
0.4	PZ	0.2453	0.4311	0.0987	0.3717	0.4014	110.4077	0.9310	666.4830	0.6558	0.7934	0.5974
0.4	GSD	0.2011	0.3858	0.0615	0.5039	0.4448	100.0000	0.9996	0.0000	1.0000	0.9998	0.7223
0.5	OCP	0.3319	0.5199	0.1003	0.3668	0.4434	187.4409	0.1797	872.6132	0.6062	0.3929	0.4181
0.5	ROCP	0.2240	0.4092	0.1226	0.2997	0.3544	81.2458	0.8877	2857.3641	0.2873	0.5875	0.4710
0.5	PZ	0.2453	0.4311	0.0987	0.3717	0.4014	110.4077	0.6933	666.4830	0.6558	0.6745	0.5380
0.5	GSD	0.2011	0.3858	0.0615	0.5039	0.4448	100.0000	0.7626	0.0000	1.0000	0.8813	0.6631

Table 11. R1 approach with both treatment and control generated from MEM.

Δ	Design	$\mathbb{E}[CP_{\Delta}^{RA}]$	$e_{CP}(\Delta)$	$Var(CP_{\Delta}^{RA})$	$v_{CP}(\Delta)$	$SCP(\Delta)$	$\mathbb{E}[CN_{\Delta}^{RA}]$	$e_{CN}(\Delta)$	$Var(CN_{\Delta}^{RA})$	$v_{CN}(\Delta)$	$SCN(\Delta)$	$CPS(\Delta)$
0	OCP	0.2674	0.7514	0.0835	0.4220	0.5867	134.2499	0.4383	91.6999	0.8723	0.6553	0.6210
0	ROCP	0.1566	0.8650	0.0426	0.5874	0.7262	75.0924	0.8327	57.6779	0.8987	0.8657	0.7960
0	PZ	0.2043	0.8161	0.0587	0.5155	0.6658	103.4696	0.6435	10.7147	0.9564	0.8000	0.7329
0	GSD	0.2011	0.8194	0.0615	0.5040	0.6617	100.0000	0.6667	0.0000	1.0000	0.8333	0.7475
0.1	OCP	0.2674	0.7514	0.0835	0.4221	0.5867	134.2495	0.4383	92.3015	0.8719	0.6551	0.6209
0.1	ROCP	0.1566	0.8650	0.0426	0.5874	0.7262	75.0968	0.8327	57.6619	0.8988	0.8657	0.7960
0.1	PZ	0.2042	0.8162	0.0587	0.5155	0.6658	103.4566	0.6436	10.6550	0.9565	0.8001	0.7330
0.1	GSD	0.2011	0.8194	0.0615	0.5040	0.6617	100.0000	0.6667	0.0000	1.0000	0.8333	0.7475
0.2	OCP	0.2674	0.7513	0.0835	0.4221	0.5867	134.2559	0.4383	92.4370	0.8718	0.6551	0.6209
0.2	ROCP	0.1566	0.8651	0.0425	0.5875	0.7263	75.0861	0.8328	57.7243	0.8987	0.8657	0.7960
0.2	PZ	0.2042	0.8162	0.0587	0.5155	0.6658	103.4633	0.6436	10.6586	0.9565	0.8000	0.7329
0.2	GSD	0.2011	0.8194	0.0615	0.5040	0.6617	100.0000	0.6667	0.0000	1.0000	0.8333	0.7475
0.3	OCP	0.2674	0.4537	0.0835	0.4221	0.4379	134.2269	0.7201	92.1377	0.8720	0.7961	0.6170
0.3	ROCP	0.1566	0.3401	0.0426	0.5873	0.4637	75.0970	0.3259	57.8272	0.8986	0.6123	0.5380
0.3	PZ	0.2042	0.3889	0.0587	0.5157	0.4523	103.4598	0.5150	10.7746	0.9562	0.7356	0.5939
0.3	GSD	0.2011	0.3858	0.0615	0.5039	0.4448	100.0000	0.4919	0.0000	1.0000	0.7460	0.5954
0.4	OCP	0.2675	0.4538	0.0835	0.4219	0.4379	134.2471	0.7721	91.7246	0.8723	0.8222	0.6300
0.4	ROCP	0.1565	0.3401	0.0425	0.5875	0.4638	75.0896	0.8336	57.4948	0.8989	0.8662	0.6650
0.4	PZ	0.2042	0.3889	0.0587	0.5157	0.4523	103.4654	0.9773	10.8016	0.9562	0.9667	0.7095
0.4	GSD	0.2011	0.3858	0.0615	0.5039	0.4448	100.0000	0.9996	0.0000	1.0000	0.9998	0.7223
0.5	OCP	0.2674	0.4537	0.0835	0.4221	0.4379	134.2260	0.5345	92.0832	0.8721	0.7033	0.5706
0.5	ROCP	0.1566	0.3401	0.0425	0.5875	0.4638	75.0960	0.9287	57.5004	0.8989	0.9138	0.6888
0.5	PZ	0.2042	0.3889	0.0587	0.5157	0.4523	103.4638	0.7395	10.7710	0.9562	0.8479	0.6501
0.5	GSD	0.2011	0.3858	0.0615	0.5039	0.4448	100.0000	0.7626	0.0000	1.0000	0.8813	0.6631

Table 12. R2 approach with both treatment and control generated from MEM.

Δ	Design	$\mathbb{E}[CP_{\Delta}^{RA}]$	$e_{CP}(\Delta)$	$Var(CP_{\Delta}^{RA})$	$v_{CP}(\Delta)$	$SCP(\Delta)$	$\mathbb{E}[CN_{\Delta}^{RA}]$	$e_{CN}(\Delta)$	$Var(CN_{\Delta}^{RA})$	$v_{CN}(\Delta)$	$SCN(\Delta)$	$CPS(\Delta)$
0	OCP	0.2674	0.7514	0.0835	0.4220	0.5867	134.2499	0.4383	91.6999	0.8723	0.6553	0.6210
0	ROCP	0.1566	0.8650	0.0426	0.5874	0.7262	75.0924	0.8327	57.6779	0.8987	0.8657	0.7960
0	PZ	0.2042	0.8161	0.0587	0.5155	0.6658	103.4696	0.6435	10.7147	0.9564	0.8000	0.7329
0	GSD	0.2011	0.8194	0.0615	0.5039	0.6617	100.0000	0.6667	0.0000	1.0000	0.8333	0.7475
0.1	OCP	0.2674	0.7514	0.0835	0.4221	0.5867	134.2495	0.4383	92.3015	0.8719	0.6551	0.6209
0.1	ROCP	0.1566	0.8650	0.0426	0.5874	0.7262	75.0968	0.8327	57.6619	0.8988	0.8657	0.7960
0.1	PZ	0.2042	0.8162	0.0587	0.5155	0.6658	103.4566	0.6436	10.6550	0.9565	0.8001	0.7329
0.1	GSD	0.2011	0.8194	0.0615	0.5039	0.6617	100.0000	0.6667	0.0000	1.0000	0.8333	0.7475
0.2	OCP	0.2674	0.7513	0.0835	0.4221	0.5867	134.2559	0.4383	92.4370	0.8718	0.6551	0.6209
0.2	ROCP	0.1566	0.8651	0.0425	0.5875	0.7263	75.0861	0.8328	57.7243	0.8987	0.8657	0.7960
0.2	PZ	0.2042	0.8162	0.0587	0.5155	0.6658	103.4633	0.6436	10.6586	0.9565	0.8000	0.7329
0.2	GSD	0.2011	0.8194	0.0615	0.5039	0.6617	100.0000	0.6667	0.0000	1.0000	0.8333	0.7475
0.3	OCP	0.2674	0.4537	0.0835	0.4221	0.4379	134.2269	0.7201	92.1377	0.8720	0.7961	0.6170
0.3	ROCP	0.1566	0.3401	0.0426	0.5873	0.4637	75.0970	0.3259	57.8272	0.8986	0.6123	0.5380
0.3	PZ	0.2042	0.3889	0.0587	0.5157	0.4523	103.4599	0.5150	10.7746	0.9562	0.7356	0.5940
0.3	GSD	0.2011	0.3858	0.0615	0.5040	0.4448	100.0000	0.4919	0.0000	1.0000	0.7460	0.5954
0.4	OCP	0.2675	0.4538	0.0835	0.4220	0.4379	134.2471	0.7721	91.7246	0.8723	0.8222	0.6300
0.4	ROCP	0.1565	0.3401	0.0425	0.5875	0.4638	75.0896	0.8336	57.4948	0.8989	0.8662	0.6650
0.4	PZ	0.2042	0.3889	0.0587	0.5157	0.4523	103.4654	0.9773	10.8016	0.9562	0.9667	0.7095
0.4	GSD	0.2011	0.3858	0.0615	0.5040	0.4448	100.0000	0.9996	0.0000	1.0000	0.9998	0.7223
0.5	OCP	0.2674	0.4537	0.0835	0.4221	0.4379	134.2260	0.5345	92.0832	0.8721	0.7033	0.5706
0.5	ROCP	0.1566	0.3401	0.0425	0.5875	0.4638	75.0960	0.9287	57.5004	0.8989	0.9138	0.6888
0.5	PZ	0.2042	0.3889	0.0587	0.5157	0.4523	103.4638	0.7395	10.7710	0.9562	0.8479	0.6501
0.5	GSD	0.2011	0.3858	0.0615	0.5040	0.4448	100.0000	0.7626	0.0000	1.0000	0.8813	0.6631

Table 13 lists the CPS of the standard, R1 and R2 approaches. In the standard sample size recalculation (Table 14), the GSD performs well against all other designs for all Δ. This is due to no variation in the recalculated sample size. Comparing the designs in the R1 approach (Table 15) with the GSD, one can see that the ROCP and PZ have approximately a similar performance as the GSD, while the OCP has a little small score. This is due to the better conditional power of the ROCP and PZ against the OCP. The R1 approach demonstrates competitive performance across different Δ values. It is expected that this approach reduces variability in recalculated sample sizes, contributing to consistent CPSs. The R2 approach might have some tendencies to converge toward the performance of the GSD design, possibly due to the sample size boundary imposed by n_max (Table 16). This behavior can be seen in the CPSs for certain scenarios. The CPS of the R1 approach is higher than the R2 approach. The detailed simulation results are presented in Tables 14–16. The box plots for the second stage conditional power and recalculated sample size for the standard simulation, the R1 approach and the R2 approach are illustrated in Figures S18 and S19, respectively.

Table 13. Conditional power score where only treatment is generated with MEM and control from a normal distribution.

Δ	Design	Standard Simulation	R1 Approach	R2 Approach
0	OCP	0.4797	0.6577	0.5135
0	ROCP	0.6176	0.8263	0.6648
0	PZ	0.6608	0.7650	0.6725
0	GSD	0.7794	0.7794	0.7794
0.1	OCP	0.4797	0.6576	0.5135
0.1	ROCP	0.6176	0.8263	0.6649
0.1	PZ	0.6608	0.7649	0.6723
0.1	GSD	0.7794	0.7794	0.7794
0.2	OCP	0.4797	0.6576	0.5135
0.2	ROCP	0.6176	0.8262	0.6648
0.2	PZ	0.6608	0.7649	0.6724
0.2	GSD	0.7794	0.7794	0.7794
0.3	OCP	0.6059	0.6162	0.6472
0.3	ROCP	0.3480	0.5372	0.5701
0.3	PZ	0.5159	0.5975	0.6142
0.3	GSD	0.5974	0.5974	0.5974
0.4	OCP	0.479	0.6292	0.5202
0.4	ROCP	0.4749	0.6640	0.6341
0.4	PZ	0.6205	0.7116	0.6434
0.4	GSD	0.7243	0.7243	0.7243
0.5	OCP	0.4196	0.5698	0.4608
0.5	ROCP	0.5081	0.6948	0.5748
0.5	PZ	0.5611	0.6522	0.5840
0.5	GSD	0.665	0.6650	0.6650

Table 14. Standard simulation where only treatment is generated with MEM and control from a normal distribution.

Δ	Design	$\mathbb{E}[CP_{\Delta}^{RA}]$	$e_{CP}(\Delta)$	$Var(CP_{\Delta}^{RA})$	$v_{CP}(\Delta)$	$SCP(\Delta)$	$\mathbb{E}[CN_{\Delta}^{RA}]$	$e_{CN}(\Delta)$	$Var(CN_{\Delta}^{RA})$	$v_{CN}(\Delta)$	$SCN(\Delta)$	$CPS(\Delta)$
0	OCP	0.2484	0.7708	0.0858	0.4141	0.5925	192.1952	0.0520	569.4070	0.6819	0.3669	0.4797
0	ROCP	0.1523	0.8694	0.0933	0.3890	0.6292	72.2506	0.8517	2302.4635	0.3602	0.6060	0.6176
0	PZ	0.1713	0.8500	0.0737	0.4570	0.6535	106.7504	0.6217	458.6497	0.7145	0.6681	0.6608
0	GSD	0.1426	0.8794	0.0458	0.5718	0.7256	100.0000	0.6667	0.0000	1.0000	0.8333	0.7794
0.1	OCP	0.2484	0.7708	0.0858	0.4141	0.5925	192.1952	0.0520	569.4070	0.6819	0.3669	0.4797
0.1	ROCP	0.1523	0.8694	0.0933	0.3890	0.6292	72.2506	0.8517	2302.4635	0.3602	0.6060	0.6176
0.1	PZ	0.1713	0.8500	0.0737	0.4570	0.6535	106.7504	0.6217	458.6497	0.7145	0.6681	0.6608
0.1	GSD	0.1426	0.8794	0.0458	0.5718	0.7256	100.0000	0.6667	0.0000	1.0000	0.8333	0.7794
0.2	OCP	0.2484	0.7708	0.0858	0.4141	0.5925	192.1952	0.0520	569.4070	0.6819	0.3669	0.4797
0.2	ROCP	0.1523	0.8694	0.0933	0.3890	0.6292	72.2506	0.8517	2302.4635	0.3602	0.6060	0.6176
0.2	PZ	0.1713	0.8500	0.0737	0.4570	0.6535	106.7504	0.6217	458.6497	0.7145	0.6681	0.6608
0.2	GSD	0.1426	0.8794	0.0458	0.5718	0.7256	100.0000	0.6667	0.0000	1.0000	0.8333	0.7794
0.3	OCP	0.2484	0.4343	0.0858	0.4141	0.4242	192.1952	0.8934	569.4070	0.6819	0.7876	0.6059
0.3	ROCP	0.1523	0.3357	0.0933	0.3890	0.3624	72.2506	0.3069	2302.4635	0.3602	0.3336	0.3480
0.3	PZ	0.1713	0.3552	0.0737	0.4570	0.4061	106.7504	0.5369	458.6497	0.7145	0.6257	0.5159
0.3	GSD	0.1426	0.3258	0.0458	0.5718	0.4488	100.0000	0.4919	0.0000	1.0000	0.7450	0.5974
0.4	OCP	0.2484	0.4343	0.0858	0.4141	0.4242	192.1952	0.3857	569.4070	0.6819	0.5338	0.4790
0.4	ROCP	0.1523	0.3357	0.0933	0.3890	0.3624	72.2506	0.8146	2302.4635	0.3602	0.5874	0.4749
0.4	PZ	0.1713	0.3552	0.0737	0.4570	0.4061	106.7504	0.9554	458.6497	0.7145	0.8349	0.6205
0.4	GSD	0.1426	0.3258	0.0458	0.5718	0.4488	100.0000	0.9996	0.0000	1.0000	0.9998	0.7243
0.5	OCP	0.2484	0.4343	0.0858	0.4141	0.4242	192.1952	0.1480	569.4070	0.6819	0.4149	0.4196
0.5	ROCP	0.1523	0.3357	0.0933	0.3890	0.3624	72.2506	0.9476	2302.4635	0.3602	0.6539	0.5081
0.5	PZ	0.1713	0.3552	0.0737	0.4570	0.4061	106.7504	0.7176	458.6497	0.7145	0.7161	0.5611
0.5	GSD	0.1426	0.3258	0.0458	0.5718	0.4488	100.0000	0.7626	0.0000	1.0000	0.8813	0.6650

Table 15. R1 approach where only treatment is generated with MEM and control from a normal distribution.

Δ	Design	$\mathbb{E}[CP_{\Delta}^{RA}]$	$e_{CP}(\Delta)$	$Var(CP_{\Delta}^{RA})$	$v_{CP}(\Delta)$	$SCP(\Delta)$	$\mathbb{E}[CN_{\Delta}^{RA}]$	$e_{CN}(\Delta)$	$Var(CN_{\Delta}^{RA})$	$v_{CN}(\Delta)$	$SCN(\Delta)$	$CPS(\Delta)$
0	OCP	0.1944	0.8262	0.0664	0.4848	0.6555	134.2621	0.4383	79.1251	0.8814	0.6598	0.6577
0	ROCP	0.1095	0.9134	0.0310	0.6481	0.7807	73.0042	0.8466	59.4450	0.8972	0.8719	0.8263
0	PZ	0.1459	0.8760	0.0445	0.5782	0.7271	103.8817	0.6408	6.8861	0.9650	0.8029	0.7650
0	GSD	0.1426	0.8794	0.0458	0.5718	0.7256	100.0000	0.6667	0.0000	1.0000	0.8333	0.7794
0.1	OCP	0.1945	0.8262	0.0664	0.4846	0.6554	134.2819	0.4381	79.2867	0.8813	0.6597	0.6576
0.1	ROCP	0.1095	0.9134	0.0310	0.6477	0.7805	73.0023	0.8467	59.3560	0.8973	0.8720	0.8263
0.1	PZ	0.1459	0.8760	0.0445	0.5781	0.7271	103.8963	0.6407	6.9131	0.9649	0.8028	0.7649
0.1	GSD	0.1426	0.8794	0.0458	0.5718	0.7256	100.0000	0.6667	0.0000	1.0000	0.8333	0.7795
0.2	OCP	0.1944	0.8262	0.0664	0.4847	0.6554	134.2665	0.4382	79.0940	0.8814	0.6598	0.6576
0.2	ROCP	0.1095	0.9133	0.0310	0.6477	0.7805	73.0069	0.8466	59.4505	0.8972	0.8719	0.8262
0.2	PZ	0.1459	0.8760	0.0445	0.5782	0.7271	103.8917	0.6407	6.9135	0.9649	0.8028	0.7650
0.2	GSD	0.1426	0.8794	0.0458	0.5718	0.7256	100.0000	0.6667	0.0000	1.0000	0.8333	0.7794
0.3	OCP	0.1945	0.3790	0.0664	0.4846	0.4318	134.2649	0.7204	79.7490	0.8809	0.8005	0.6162
0.3	ROCP	0.1095	0.2918	0.0310	0.6478	0.4697	72.9981	0.3119	59.3621	0.8973	0.6046	0.5372
0.3	PZ	0.1459	0.3291	0.0445	0.5783	0.4537	103.8943	0.5179	6.9698	0.9648	0.7413	0.5975
0.3	GSD	0.1426	0.3258	0.0458	0.5718	0.4488	100.0000	0.4919	0.0000	1.0000	0.7460	0.5974
0.4	OCP	0.1944	0.3789	0.0664	0.4848	0.4318	134.2287	0.7722	79.5930	0.8811	0.8266	0.6292
0.4	ROCP	0.1096	0.2918	0.0311	0.6473	0.4696	73.0094	0.8197	59.5854	0.8971	0.8584	0.6640
0.4	PZ	0.1459	0.3292	0.0445	0.5781	0.4536	103.9047	0.9743	6.9338	0.9649	0.9696	0.7116
0.4	GSD	0.1426	0.3258	0.0458	0.5718	0.4488	100.0000	0.9996	0.0000	1.0000	0.9998	0.7243
0.5	OCP	0.1944	0.3789	0.0664	0.4848	0.4318	134.2600	0.5342	79.3819	0.8812	0.7077	0.5698
0.5	ROCP	0.1095	0.2918	0.0310	0.6477	0.4697	73.0023	0.9426	59.5091	0.8971	0.9199	0.6948
0.5	PZ	0.14590	0.3291	0.0445	0.5782	0.4537	103.8885	0.7367	6.9205	0.9649	0.8508	0.6522
0.5	GSD	0.1426	0.3258	0.0458	0.5718	0.4488	100.0000	0.7626	0.0000	1.0000	0.8813	0.6650

Table 16. R2 approach where only treatment is generated with MEM and control from a normal distribution.

Δ	Design	$\mathbb{E}[CP_{\Delta}^{RA}]$	$e_{CP}(\Delta)$	$Var(CP_{\Delta}^{RA})$	$v_{CP}(\Delta)$	$SCP(\Delta)$	$\mathbb{E}[CN_{\Delta}^{RA}]$	$e_{CN}(\Delta)$	$Var(CN_{\Delta}^{RA})$	$v_{CN}(\Delta)$	$SCN(\Delta)$	$CPS(\Delta)$
0	OCP	0.2631	0.7558	0.1039	0.3552	0.5555	197.8316	0.0145	28.8411	0.9284	0.4714	0.5135
0	ROCP	0.1901	0.8306	0.0724	0.4618	0.6462	118.9021	0.5407	170.0101	0.8262	0.6834	0.6648
0	PZ	0.1934	0.8273	0.0721	0.4630	0.6451	129.2970	0.4714	28.8894	0.9283	0.6998	0.6725
0	GSD	0.1426	0.8795	0.0458	0.5718	0.7256	100.0000	0.6667	0.0000	1.0000	0.8333	0.7794
0.1	OCP	0.2631	0.7558	0.1040	0.3552	0.5555	197.8344	0.0144	28.6883	0.9286	0.4715	0.5135
0.1	ROCP	0.1901	0.8307	0.0724	0.4618	0.6462	118.8988	0.5407	169.4012	0.8265	0.6836	0.6649
0.1	PZ	0.1935	0.8272	0.0721	0.4629	0.6450	129.3287	0.4711	29.0120	0.9282	0.6997	0.6723
0.1	GSD	0.1426	0.8794	0.0458	0.5718	0.7256	100.0000	0.6667	0.0000	1.0000	0.8333	0.7794
0.2	OCP	0.2631	0.7558	0.1040	0.3552	0.5555	197.8328	0.01445	28.6835	0.9286	0.4715	0.5135
0.2	ROCP	0.1901	0.8307	0.0725	0.4617	0.6462	118.9015	0.5407	169.7098	0.8263	0.6835	0.6648
0.2	PZ	0.1934	0.8272	0.0721	0.4630	0.6451	129.3149	0.4712	28.9132	0.9283	0.6998	0.6724
0.2	GSD	0.1426	0.8794	0.0458	0.5718	0.7256	100.0000	0.6667	0.0000	1.0000	0.8333	0.7794
0.3	OCP	0.2631	0.4494	0.1040	0.3552	0.4023	197.8225	0.8559	28.8739	0.9284	0.8921	0.6472
0.3	ROCP	0.1901	0.3744	0.0724	0.4617	0.4181	118.8925	0.6179	169.5006	0.8264	0.7222	0.5701
0.3	PZ	0.1935	0.3779	0.0721	0.4630	0.4204	129.3271	0.6874	28.8702	0.9284	0.8079	0.6142
0.3	GSD	0.1426	0.3258	0.0458	0.5718	0.4488	100.0000	0.4919	0.0000	1.0000	0.7460	0.5974
0.4	OCP	0.2631	0.4493	0.1040	0.3552	0.4023	197.8144	0.3483	29.2819	0.9279	0.6381	0.5202
0.4	ROCP	0.1901	0.3745	0.0725	0.4615	0.4180	118.9108	0.8743	170.1555	0.8261	0.8502	0.6341
0.4	PZ	0.1935	0.3780	0.0722	0.4628	0.4204	129.3448	0.8047	29.0698	0.9281	0.8664	0.6434
0.4	GSD	0.1426	0.3258	0.0458	0.5718	0.4488	100.0000	0.9996	0.0000	1.0000	0.9998	0.7243
0.5	OCP	0.2631	0.4493	0.1039	0.3552	0.4023	197.8254	0.1105	28.8752	0.9284	0.5194	0.4608
0.5	ROCP	0.1901	0.3745	0.0725	0.4616	0.4181	118.8953	0.6367	169.9171	0.8262	0.7314	0.5748
0.5	PZ	0.1935	0.3779	0.0721	0.4629	0.4204	129.3121	0.5672	29.0699	0.9281	0.7477	0.5840
0.5	GSD	0.1426	0.3258	0.0458	0.5718	0.4488	100.0000	0.7626	0.0000	1.0000	0.8813	0.6650

Previously, we assumed $\gamma_{\text{loc}} = \gamma_{\text{var}} = 0.5$. To assess the effect of unequal weights, we assumed $\gamma_{\text{loc}} = 0.2$ and $\gamma_{\text{var}} = 0.8$ for the MDM with $\pi = 0.5$ and the resulting study is reported in Table 17 for four different design strategies (OCP, ROCP, PZ and GSD) across a range of Δ values from 0.0 to 0.5. The evaluation uses three different approaches, i.e., R1, R2 and standard. The table shows that the standard approach yields the lowest CPS across all designs and Δ values, underscoring the significant effect of unequal weighting on the location parameter. This indicates that the standard method is more sensitive to changes in location weight. In contrast, both the R1 and R2 approaches outperform the standard method in terms of CPS, suggesting greater robustness to variations in location weighting. Among the two, R1 generally performs better than R2, particularly in designs such as ROCP, PZ and GSD. However, for the OCP design, R1 and R2 produce comparable results, with R2 occasionally achieving slightly higher scores. For larger values of Δ, the performance of R2 improves when a smaller value of γ_{loc} is used, whereas R1 performs better with a larger γ_{loc}.

Table 17. Comparison of conditional power score for R1, R2 and standard approaches using $\gamma_{\text{loc}} = 0.2$ and MDM with $\pi = 0.5$.

Δ	Design	Standard Approach	R1 Approach	R2 Approach
0.0	OCP	0.3575	0.6679	0.5845
0.0	ROCP	0.6095	0.7897	0.6477
0.0	PZ	0.6293	0.7641	0.6816
0.0	GSD	0.7100	0.7789	0.7789
0.1	OCP	0.3113	0.6289	0.5410
0.1	ROCP	0.5015	0.7580	0.6084
0.1	PZ	0.5535	0.7271	0.6427
0.1	GSD	0.6617	0.7425	0.7425
0.2	OCP	0.2917	0.6089	0.5210
0.2	ROCP	0.4088	0.7409	0.5914
0.2	PZ	0.4898	0.7073	0.6246
0.2	GSD	0.6243	0.7231	0.7231
0.3	OCP	0.2905	0.6149	0.6163
0.3	ROCP	0.3083	0.6459	0.6011
0.3	PZ	0.3967	0.6534	0.6382
0.3	GSD	0.5195	0.6665	0.6665
0.4	OCP	0.3109	0.6430	0.5979
0.4	ROCP	0.3712	0.7099	0.6468
0.4	PZ	0.4304	0.7121	0.6700
0.4	GSD	0.5459	0.7264	0.7264
0.5	OCP	0.3481	0.6374	0.5988
0.5	ROCP	0.4207	0.7221	0.6459
0.5	PZ	0.4503	0.6993	0.6652
0.5	GSD	0.5631	0.7130	0.7130

7. An Illustrative Clinical Trial Example

In the context of a clinical trial conducted by Bowden and Mander [26], we investigate the effectiveness of treatment labeled as T and placebo labeled as P in alleviating pain among osteoarthritis patients. The trial aims to assess pain relief over 2-week period compared to the baseline. Pain relief levels are measured using the McGill pain scale Melzack and Torgerson [27], which ranges from 0 (indicating no pain) to 50 (indicating the highest pain level). In this trial, the dataset is generated through the MDM having both

distributions normal and mixing proportion $p = 0.5$. Moreover, the pain relief values may be subject to measurement errors and may follow a MEM.

To enhance comprehension of the proposed methods, we have adapted the original clinical trial design based on recommendations from Herrmann and Rauch [28]. Initially, a pilot study indicated the superiority of the new treatment. However, further evidence is needed to quantify its actual effect, considering both MEM and MDM situations. As a result, the following hypotheses are formulated:

$$H_0 : \mu^T_{\text{baseline}} - \mu^P_{\text{baseline}} \leq 0 \quad \text{vs.} \quad H_1 : \mu^T_{\text{baseline}} - \mu^P_{\text{2 weeks}} > 0. \tag{20}$$

Here, μ^T_{baseline} represents the expected pain relief after 2 weeks for the new treatment, while μ^P_{baseline} signifies the same for the standard treatment. Given the potential for measurement errors, the clinical trial employs an adaptive two-stage design, allowing adjustments to the sample size during an interim analysis. Specifically, $n_1 = n_2 = 50$ is chosen, and the maximum sample size is capped at $n_{\max} = 200$, recognizing the need for a larger sample size to account for the potential measurement inaccuracies and then utilized the results from Tables 2–4 and Tables 10–12. Additionally, the trial incorporates a binding futility stop bound ($\alpha_0 = 0.5$), a global significance level ($\alpha = 0.025$) and locally adjusted significance levels using the Pocock method, and the adjustments made to account for the possible deviations from a strictly normal distribution.

Suppose that during an interim analysis, an interim effect size of $\Delta = 0.2$ is observed, corresponding to an interim test statistic of $T_1 = 1$. The focus lies in assessing the conditional performance differences among the OCP, ROCP and PZ approaches, both with and without the R1 resampling approach while taking into account both the MEM and the MDM. The evaluation assigns equal weight to conditional performance score components, with the understanding that recalibrating the sample size should be driven by reasonable adjustments considering the complex data distribution.

While the primary emphasis is on the performance at $\Delta = 0.2$, neighboring effect sizes ($\Delta = 0.1$ and 0.3) are also considered by taking into account the mixed distribution performance metrics are presented in Tables 1–4. Moreover, for the measurement errors, the performance metrics are presented in Table 9, along with Tables 10–12. Without resampling and with an interim effect size of $\Delta = 0.2$, the OCP approach suggests a maximum sample size of 164. Conversely, the ROCP approach implies no need for an increased sample size or a second trial stage, taking into account the complexity of the data distribution. The PZ approach advocates adhering to the total sample size of 118, with potential adjustments for measurement errors and mixed distribution. However, upon implementing the R1 resampling approach, the trial continuation is recommended for all three approaches (OCP, ROCP, PZ), with total sample sizes ranging from at least 75 to a maximum of 130.

On comparing overall conditional performance, as quantified by the conditional performance score, the R1 approach outperforms the original approach across all three recalculation rules and the considered effect sizes, especially given the mixed distribution and measurement errors. This improvement primarily results from variance reduction in the conditional sample size and power due to the resampling approach, which is particularly relevant when dealing with complex data distributions (as illustrated in Tables 2 and 3). For effect sizes of 0.1 and 0.2, the ROCP R1 resampling approach exhibits the best performance, while for an effect size of 0.3, the OCP R1 resampling approach takes the lead. This shows the adaptability of these approaches to different effect sizes within the mixed distribution.

For those interested in global performance, the OCP R1 approach attains greater global power than the other ROCP and PZ R1 approaches across the considered effect sizes, due to larger sample sizes and robust statistical methods. In conclusion, the integration of resampling techniques accounts for potential measurement errors and the MDM, thereby

improving the reliability of the sample size-determination process, especially when dealing with complex and non-normal data distributions.

8. Conclusions

Integrating resampling techniques into established sample size-recalculation rules enhances the robustness of recalculation approaches, leading to significantly improved performance across various individual characteristics and conditional performance scores. This improvement is primarily due to the decreased variance of the conditional sample size and conditional power. It is important to note that reference values and the weighting scheme for the conditional performance score can also be adjusted. Additionally, it is observed that the CPS fluctuates around $\Delta = 0.3$ is a common trait of recalculation rules. This pattern emerges because, for small effects, increasing the sample size is not advantageous, while for medium effects and beyond, an increase becomes reasonable.

One could argue against increasing the sample size as the interim test statistic increases, and similarly against substantial jumps in the sample size function, as this implies a drastic change in sample size with minimal alterations in the test statistic. The resampling approach, where conventional rules exhibit issues, presents a compromise between these extremes. It can be contended that any recalculation rule with significant jumps is not inherently reasonable, and thus, the compromise offered by the resampling approach might not be optimal either. A general recommendation suggests configuring design settings to avoid these jumps, such as by setting a smaller maximal sample size n_{max} or a larger local significance level (α_1+α_2). While the resampling approaches surpass the original sample size-recalculation rules concerning the conditional performance score, it does not imply the resulting sample sizes are point-wise optimal. Instead, it mitigates the average risk of selecting a completely incorrect sample size, leading to favorable average outcomes. However, in specific cases, this approach might not be suitable. This characteristic is not a drawback for the resampling approach but is generally applicable to sample size-recalculation rules. Resampling-based sample size-recalculation rules provide a favorable approach to balancing the cost-benefit ratio. By reducing the average deviation from the ideal sample size, the method effectively navigates the costs and benefits of a study, achieving an optimal trade-off. Notably, the similarity between resampling procedures and sample size recalculation for group sequential designs is remarkable. To be more specific, the promising zone approach in conjunction with resampling closely approximates a group sequential design. This arises because the promising zone approach introduces significant sample size adjustments within a narrow range of interim effects, minimally impacting the smoothed sample size curve. This observation further supports the notion that group sequential designs hold a distinctive position among designs incorporating sample size recalculation. It is noticed from the results that a larger mixing proportion tends to favor the R2 approach in the presence of heterogeneity. In the case of measurement error, both the R1 and R2 approaches perform similarly when measurement error is assumed in both the treatment and control groups. However, when measurement error is present only in the treatment group, the R1 approach demonstrates better performance. The findings suggest that the R1 approach is the most reliable method for achieving optimal conditional performance scores, while the R2 approach, despite its improvements, does not consistently match the performance of the R1 method or the standard simulation.

Nevertheless, while sample size recalculation based on group sequential designs depends solely on the interim test statistic for early trial termination, integrating resampling into recalculation rules permits basing sample size adjustments on conditional power considerations. This integration effectively mitigates drastic fluctuations in sample size.

Consequently, resampling enhances the robustness of sample size-recalculation rules, effectively addressing the inherent randomness in observed interim test statistics.

Considering the resampling techniques outlined in Formulas (18) and (19), the current work can be extended to studies involving various types of endpoints, as long as the test statistics exhibit approximate normal distribution characteristics. For example, in studies with binary endpoints, one can readily apply the resampling approach by utilizing the normal approximation to the binomial. Similarly, to deal with time-to-event endpoints, one can explore the possibility of employing the resampling approach through the utilization of the logrank test within an adaptive design framework. As an alternative path to enhancing the efficacy of sample size recalculation, consider the development of a more direct approach. This approach could involve formulating a sample size-recalculation function that is specifically designed to optimize the conditional performance score. Also, one can explore the idea of implementing the alternative approach within a numerical-constrained optimization framework. This would involve setting up a mathematical optimization problem where the objective is to maximize the conditional performance score while adhering to certain constraints. It is worth mentioning that if the measurement error model is misspecified, it can have several significant consequences, depending on the context of the analysis. For instance, the resulting estimates may be biased and associated with large standard errors, leading to invalid statistical inference. This misspecification can also impact power analysis, mislead model selection criteria and ultimately result in poor predictive performance. Given the seriousness of these implications, the present computational framework can be studied in model misspecification scenarios.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/stats8020031/s1. The the additional figures and tables are given in the supplementary text. In particular, Section S1 tabulates the results assuming $\pi = 0.2$. Section S2 depicts the box plots of CPS for MDM and MEM.

Author Contributions: Conceptualization, H.F. and S.A.; methodology, H.F. and I.S.; software, H.F. and S.A.; validation, S.A., I.S. and M.M.A.A.; formal analysis, H.F. and S.A.; investigation, H.F. and S.A.; resources, I.S., I.A.N. and S.A.; data curation, S.A. and I.S.; writing—original draft preparation, H.F., S.A. and I.A.N.; writing—review and editing, S.A. and I.S.; visualization, H.F. and I.S.; supervision, S.A. and I.S.; project administration, S.A., I.S. and M.M.A.A.; funding acquisition, I.A.N. All authors have read and agreed to the published version of the manuscript.

Funding: The authors received no specific funding for the article.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Acknowledgments: The authors thanks the two anonymous reviewers for their constructive comments.

Conflicts of Interest: The authors declare no conflict of interest.

Notations

The following notations are used in this manuscript:

$\mathbb{E}[\mathrm{CN}_{\Delta}^{\mathrm{RA}}]$	Expected value of the conditional sample size in the RA.
$Var(CP_{\Delta}^{RA})$	Variance of the conditional power in the RA.
$e_{CN}(\Delta)$	Location component of the conditional sample size sub-score.
$v_{CN}(\Delta)$	Variation component of the conditional sample size sub-score.
$SCN(\Delta)$	Conditional sample size sub-score.

$\mathbb{E}[\mathrm{CP}_{\Delta}^{\mathrm{RA}}]$	Expected value of the conditional power in the RA.
$Var(CP_{\Delta}^{RA})$	Variance of the conditional power in the RA.
$e_{CP}(\Delta)$	Location component of the conditional power sub-score.
$v_{CP}(\Delta)$	Variation component of the conditional power sub-score.
$SCP(\Delta)$	Conditional power sub-score.
$CPS(\Delta)$	Final point-wise conditional score.

References

1. Chow, S.C.; Chang, M. Adaptive design methods in clinical trials—A review. *Orphanet J. Rare Dis.* **2008**, *3*, 11. [CrossRef] [PubMed]
2. Herrmann, C.; Kluge, C.; Pilz, M.; Kieser, M.; Rauch, G. Improving sample size recalculation in adaptive clinical trials by resampling. *Pharm. Stat.* **2021**, *20*, 1035–1050. [CrossRef] [PubMed]
3. Pocock, S.J. Group sequential methods in the design and analysis of clinical trials. *Biometrika* **1977**, *64*, 191–199. [CrossRef]
4. O'Brien, P.C.; Fleming, T.R. A Multiple Testing Procedure for Clinical Trials. *Biometrics* **1979**, *35*, 549–556. [CrossRef]
5. Denne, J.S. Sample size recalculation using conditional power. *Stat. Med.* **2001**, *20*, 2645–2660. [CrossRef]
6. Friede, T.; Kieser, M. A comparison of methods for adaptive sample size adjustment. *Stat. Med.* **2001**, *20*, 3861–3873. [CrossRef]
7. Charles, P.; Giraudeau, B.; Dechartres, A.; Baron, G.; Ravaud, P. Reporting of sample size calculation in randomised controlled trials. *BMJ* **2009**, *338*, b1732. [CrossRef]
8. Hammouri, H.; Ali, M.; Alquran, M.; Alquran, A.; Abdel Muhsen, R.; Alomari, B. Adaptive Multiple Testing Procedure for Clinical Trials with Urn Allocation. *Mathematics* **2023**, *11*, 3965. [CrossRef]
9. Kieser, M.; Friede, T. Simple procedures for blinded sample size adjustment that do not affect the type I error rate. *Stat. Med.* **2003**, *22*, 3571–3581. [CrossRef]
10. Harden, M.; Friede, T. Sample size calculation in multi-centre clinical trials. *BMC Med. Res. Methodol.* **2018**, *18*, 156. [CrossRef]
11. Das, S.; Mitra, K.; Mandal, M. Sample size calculation: Basic principles. *Indian J. Anaesth.* **2016**, *60*, 652. [CrossRef] [PubMed]
12. Jennison, C.; Turnbull, B.W. Mid-course sample size modification in clinical trials based on the observed treatment effect. *Stat. Med.* **2003**, *22*, 971–993. [CrossRef] [PubMed]
13. Pritchett, Y.L.; Menon, S.; Marchenko, O.; Antonijevic, Z.; Miller, E.; Sanchez-Kam, M.; Morgan-Bouniol, C.C.; Nguyen, H.; Prucka, W.R. Sample size re-estimation designs in confirmatory clinical trials—current state, statistical considerations, and practical guidance. *Stat. Biopharm. Res.* **2015**, *7*, 309–321. [CrossRef]
14. Chakraborty, H.; Gu, H. A Mixed Model Approach for Intent-to-Treat Analysis in Longitudinal Clinical Trials with Missing Values. *Psychiatry Res.* **2009**, *129*, 1–9. [CrossRef]
15. Boos, D.D.; Brownie, C. A rank-based mixed model approach to multisite clinical trials. *Biometrics* **1992**, *48*, 61–72. [CrossRef]
16. Nagin, D.S.; Odgers, C.L. Group-based trajectory modeling in clinical research. *Annu. Rev. Clin. Psychol.* **2010**, *6*, 109–138. [CrossRef]
17. Deng, Y.; Tu, D.; O'Callaghan, C.J.; Liu, G.; Xu, W. Two-stage multivariate Mendelian randomization on multiple outcomes with mixed distributions. *Stat. Methods Med. Res.* **2022**, *32*, 1543–1558 [CrossRef]
18. Spanbauer, C.; Sparapani, R. Nonparametric machine learning for precision medicine with longitudinal clinical trials and Bayesian additive regression trees with mixed models. *Stat. Med.* **2021**, *40*, 2665–2691. [CrossRef]
19. Liang, H.; Wu, H.; Carroll, R.J. The relationship between virologic and immunologic responses in AIDS clinical research using mixed-effects varying-coefficient models with measurement error. *Biostatistics* **2003**, *4*, 297–312. [CrossRef]
20. Morgan, T.M.; Elashoff, R.M. Effect of covariate measurement error in randomized clinical trials. *Stat. Med.* **1987**, *6*, 31–41. [CrossRef]
21. Wang, N.; Lin, X.; Gutierrez, R.G.; Carroll, R.J. Bias analysis and SIMEX approach in generalized linear mixed measurement error models. *J. Am. Stat. Assoc.* **1998**, *93*, 249–261. [CrossRef]
22. Yang, Y.; Li, G.; Tong, T. Corrected empirical likelihood for a class of generalized linear measurement error models. *Sci. China Math.* **2015**, *58*, 1523–1536. [CrossRef]
23. Brakenhoff, T.B.; Mitroiu, M.; Keogh, R.H.; Moons, K.G.; Groenwold, R.H.; van Smeden, M. Measurement error is often neglected in medical literature: A systematic review. *J. Clin. Epidemiol.* **2018**, *98*, 89–97. [CrossRef] [PubMed]
24. Mehta, C.R.; Pocock, S.J. Adaptive increase in sample size when interim results are promising: a practical guide with examples. *Stat. Med.* **2011**, *30*, 3267–3284. [CrossRef]
25. Herrmann, C.; Pilz, M.; Kieser, M.; Rauch, G. A new conditional performance score for the evaluation of adaptive group sequential designs with sample size recalculation. *Stat. Med.* **2020**, *39*, 2067–2100. [CrossRef]
26. Bowden, J.; Mander, A. A review and re-interpretation of a group-sequential approach to sample size re-estimation in two-stage trials. *Pharm. Stat.* **2014**, *13*, 163–172. [CrossRef]

27. Melzack, R.; Torgerson, W.S. On the language of pain. *J. Am. Soc. Anesthesiol.* **1971**, *34*, 50–59. [CrossRef]
28. Herrmann, C.; Rauch, G. Smoothing corrections for improving sample size recalculation rules in adaptive group sequential study designs. *Methods Inf. Med.* **2021**, *60*, 001–008. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.