Measurement Error in Clinical Trials and sample Size Determination

By

Hassan Farooq

Department of Statistics

Faculty of Natural Sciences

Quaid-i-Azam University, Islamabad

2023

Measurement Error in Clinical Trials and sample Size Determination

By

Hassan Farooq

A THESIS SUBMITTED IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF PHILOSOPHY IN STATISTICS

Supervised By

Dr. Sajid Ali

Department of Statistics

Faculty of Natural Sciences

Quaid-i-Azam University, Islamabad

2023

CERTIFICATE

Measurement Error in Clinical Trials and Sample Size Determination

By

Hassan Farooq

(Reg. No. 02222113012)

A THESIS SUBMITTED IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF M.PHIL. IN

STATISTICS

We accept this thesis as conforming to the required standards

1. _____
Dr. Sajid Ali
(Supervisor)

2. _____
Dr. Muhammad Adil
(External Examiner)

3 _____
Prof. Dr. Ijaz Hussain
(Chairman)

DEPARTMENT OF STATISTICS
QUAID-I-AZAM UNIVERSITY
ISLAMABAD, PAKISTAN
2023

Declaration

I "Hassan Farooq" hereby solemnly declare that this thesis titled, "Measurement Error in Clinical Trials and Sample Size determination".

- This work was done wholly in candidature for a degree of M.Phil Statistics at Quaid-i-Azam University Islamabad.
- Where I got help from the published work of others, this is always clearly stated.
- Where I have quoted from the work of others, the source is always mentioned. Except of such quotations, this thesis is entirely my own research work.
- Where the thesis is based on work done by myself jointly with my supervisor, I have made clear exactly what was done by others and what I have suggested

Dated:_____ Signature:_____

Dedication

I am feeling great honor and pleasure to dedicate this research work to

My beloved Parents

Whose endless affection, prayers and wishes have been a great source of comfort for me during my whole education period and my life

Acknowledgments

Foremost, I express my heartfelt gratitude and reverence to Allah Almighty, The Creator of the Universe, for His blessings and guidance. I extend profound respect and appreciation to the Holy Prophet Hazrat Muhammad (SAW), whose life and teachings continue to inspire and enlighten us with the essence of Islam.

I am deeply indebted to my esteemed supervisor, Dr. Sajid Ali, whose unwavering guidance and support have been invaluable throughout my research journey. His insightful advice, particularly in relation to my thesis, has been instrumental in shaping this work. I am truly fortunate to have benefited from his leadership and expertise, and I extend my sincerest thanks to him.

My sincere thanks and admiration go to the distinguished professors Prof. Dr. Ijaz Hussain (Chairman), Dr. Yousaf Shad, Dr. Abdul Haq, Prof. Dr. Javid Shabbir, Dr. Ismail Shah, and Dr. Manzoor Khan. Their imparted knowledge and steadfast guidance have been a beacon throughout my research endeavors.

I am profoundly grateful to my parents, Abid Farooq and Riffat Jabeen, for their boundless love, support, and encouragement that have been constant pillars in my life. Their unwavering belief in me has made this accomplishment possible. Furthermore, I would like to extend my heartfelt gratitude to my sisters for their unwavering support and encouragement. I am truly blessed to have them by my side. Thank you for being an integral part of my life.

I also wish to extend my gratitude to my friends and classmates whose assistance and cooperation have enriched my academic journey.

In conclusion, the support and encouragement of these individuals have been integral to the completion of this thesis. I am humbled and thankful for their contributions.

Abstract

Adaptive clinical trials offer a flexible approach to refining sample sizes during ongoing research, enhancing trial efficiency. This study delves into improving sample size recalculation through resampling techniques, employing measurement error and mixed distribution models. The core inquiry addresses the potency of resampling in enhancing sample size recalculation and evaluates the impact of measurement error and mixed distribution models on clinical trial efficacy. 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) on data generated from measurement error and mixed distribution models.

The key findings indicate that the R1 approach, capitalizing on mean as a summary location, notably 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 R2 approach demonstrates robust performance across most effect sizes, although GSD retains superiority within 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.

In conclusion, resampling-based approaches, particularly R1 and R2, offer improved sample size recalculation over conventional methods. R1 approach excels in minimizing recalculated sample size variability, while R2 approach incorporating both mean and standard deviation, presents a refined alternative. However, challenges in precisely approaching target sample sizes under certain conditions indicate avenues for further refinement. This research contributes to the optimization of adaptive clinical trials, enhancing their efficiency and reliability.

Contents

List of Tables

Chapter 1

Introduction

Clinical trials play a pivotal role in the evaluation of new medical interventions, aiming to provide reliable evidence for the safety and efficacy of treatments. With the increasing complexity of medical research and the demand for more efficient trial designs, adaptive clinical trials have gained significant attention. Adaptive designs allow for modifications to the study's parameters based on accumulating data, maximizing efficiency, and enhancing the trial's ability to answer research questions effectively.

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. The traditional fixed-sample size designs often face challenges, as the required sample size is typically determined before any data is collected, leading to inefficiencies and potential resource wastage.

This thesis delves into the realm of sample size recalculation in the context of adaptive clinical trials. Specifically, it focuses on a novel approach based on a measurement error model (MEM) to generate treatment and control groups and explores various sample size recalculation strategies that adaptively adjust the sample size based on interim analysis. Furthermore, the thesis delves into an intricate investigation of diverse strategies tailored for the recalibration of sample size. These strategies elegantly adapt the trial's sample size based on interim analyses, ensuring that the trial remains appropriately powered and resource-efficient. By seamlessly integrating the generation of treatment and control groups through a two-component mixture distribution and application of sample size recalculation approaches that recalculate sample size based on interim analysis.

1.1 Background

Adaptive clinical trials are a type of clinical trial design that allows for modifications to trial parameters based on interim data analysis while the trial is ongoing. Unlike traditional fixed designs, where the trial parameters (such as sample size, treatment regimen, or patient population) are predetermined and remain fixed throughout the trial, adaptive trials permit adjustments based on accumulated data. This flexibility can lead to more efficient and informative trials.

Advantages of Adaptive Trials over Traditional Fixed Designs:

1. **Efficiency:** Adaptive trials can lead to more efficient resource utilization by adjusting the trial parameters based on emerging evidence, which can reduce the number of patients needed to achieve a meaningful outcome.

2. **Ethical Considerations:** Adaptive trials can stop or modify a trial if interim results show that a treatment is significantly beneficial or harmful, potentially sparing patients from unnecessary exposure to ineffective or harmful interventions.

3. **Learning and Decision-Making:** Adaptive trials provide opportunities for learning from interim results, leading to more informed decisions about treatment strategies and patient populations.

4. **Increased Probability of Success:** Adaptive designs can increase the probability of successfully identifying effective treatments or interventions, as adjustments can be made based on accumulating evidence.

Sample Size Determination in Clinical Trials:

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.

Limitations of Fixed-Sample Size Approaches:

1. **Resource Utilization:** Fixed-sample size designs may lead to the enrolment of more participants than necessary, wasting resources and potentially exposing more patients to experimental treatments.

2. **Statistical Power:** Insufficient sample sizes can result in low statistical power, reducing the ability of the trial to detect a true treatment effect.

3. **Type I Error:** Fixed designs may increase the risk of Type-I error (false-positive results) if the predetermined sample size is too small.

1.2 Measurement Error Models and Generating Treatment and Control Groups

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.

In the context of generating treatment and control groups, MEMs can be used to improve the accuracy of treatment effect estimates. 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.

In summary, adaptive clinical trials offer flexibility and efficiency compared to traditional fixed designs. Sample size determination is crucial for ensuring trials have adequate statistical power. Fixed-sample size approaches have limitations in terms of resource utilization and statistical power. MEMs are important for minimizing bias and improving the accuracy of treatment effect estimates in clinical trials.

1.3 Objective

- The primary objective of this thesis is twofold:
    1. To propose a measurement error model-based approach for creating treatment and control groups in adaptive clinical trials.
    2. To investigate the application of different sample size recalculation methods at interim analyses.
- By implementing these strategies, we aim to achieve the following goals:
    1. Enhance the efficiency and accuracy of adaptive clinical trials.
    2. Contribute to the advancement of evidence-based medical research.

1.4 Methodology

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 at 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.

Significance

This thesis holds great significance for both the statistical and medical research communities. The proposed MEM approach offers a valuable contribution to the literature on adaptive clinical trial designs. By generating groups more accurately, it promises to enhance the internal validity and precision of treatment effect estimates. Moreover, The investigation of sample size recalculation methods addresses a crucial issue in adaptive trials, as it allows for more informed and data-driven decisions during the course of the study. 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. As clinical research continues to evolve, the integration of mixture distribution models holds the promise of enhancing the validity and impact of clinical trial outcomes, ultimately leading to improved patient care and medical decision-making. The thesis is organized as follows: Chapter 2 provides a comprehensive review of the research motivation behind the study, literature on adaptive clinical trials, sample size determination, MEMs and mixture distribution model. Chapter 3 outlines the various sample size recalculation methods and their implementation in adaptive trials. Also, details the proposed MEM and mixture distribution model -based approach for group generation. Moreover, we present the results and analyses from the simulation-based study. Finally, Chapter 4 offers concluding remarks, discusses the implications of the findings, and outlines potential areas for future research.

With a deep commitment to statistical rigor and scientific inquiry, this thesis aims to contribute to the advancement of adaptive clinical trial designs and promote the adoption of efficient and flexible methodologies in medical research.

Chapter 2

Research Motivation and Literature Review

Adaptive clinical trial designs have gained prominence due to their potential to improve efficiency, flexibility, and ethical considerations in drug development. These trials allow for modifications in study design, including sample size recalculation, based on interim analyses or accumulating data. However, the accuracy and reliability of estimated sample sizes can significantly impact study validity and subsequent decision-making.

The motivation behind this research is to refine and optimize existing sample size re-estimation techniques used in adaptive clinical trials. Although various methods are employed, there is always a room for improvement in terms of precision, robustness, and efficiency. The current approaches may rely on simplified assumptions or outdated statistical models, leading to suboptimal adaptations and potential biases.This research aims to address these limitations and contribute to advancing sample size re-estimation techniques. By developing novel statistical methods, we seek to enhance precision, ensuring adequate power while controlling Type-I error rates.

This research also aims to tackle practical challenges associated with sample size adaptations in adaptive trials. The challenges include accounting for uncertainty, managing logistics and costs, handling missing data, and considering recruitment rates and timelines. By addressing these issues, we can provide reliable and practical guidelines for conducting adaptive trials with optimal sample size recalculations.

The outcomes of this research will have far-reaching implications for clinical research and drug development. More accurate and robust sample size recalculation techniques will enable precise and efficient adaptive trial design, leading to better-informed decisions about treatment efficacy and safety.

Ultimately, the findings will contribute to advancing statistical methodology

in adaptive trial designs, providing a foundation for future studies. Improved precision in sample size adaptation will empower researchers, optimize resource allocation, and accelerate the development of safe and effective therapies.

By addressing research gaps and challenges in sample size re-estimation methods for adaptive clinical trials, this study aims to make a significant contribution to fostering innovation, efficiency, and reliability in the drug development process.

2.1 Literature review

Herrmann et al. (2021) noticed the importance of proper sample size calculations in clinical trials cannot be overstated. Choosing the wrong parameter assumptions can lead to underpowered or overpowered studies, which can have serious consequences for patients, researchers, and healthcare providers. They noticed that adaptive group sequential study designs have emerged as a promising approach to dealing with planning uncertainties in clinical trials. These designs allow for sample size updates during an ongoing trial based on observed interim effects, which can help improve the efficiency and accuracy of the study. One particular approach to adaptive group sequential study designs is resampling. Resampling involves repeatedly drawing samples from the observed data and using these samples to estimate the distribution of the test statistic under the null hypothesis. This approach can help deal with uncertainty related to the observed interim effect in adaptive clinical trials, which can be particularly useful in situations where the observed effect is smaller than expected or the variance is larger than anticipated.

Several methods have been proposed for implementing resampling in adaptive clinical trials, including the Pocock and O'Brien-Fleming 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 (2001) noticed that sample size may be determined by one or more nuisance characteristics, which are typically unknown, in order to obtain a specific power at a predetermined absolute difference in mean response. It has been extensively researched how to recalculate the sample size from an internal pilot using estimates of these characteristics. The majority of these strategies overlook the fact that information on the relevant parameter from this internal pilot will affect the final test statistic's result. To preserve the likelihood of rejecting the null hypothesis at the conclusion of the investigation under the prespecified absolute difference in mean response conditional on the data, the authors offered an approach that requires recalculating the target sample size.

Friede and Kieser (2001) developed the internal pilot study design, which enables the sample size to be revised over the course of a trial using the estimated

variance discovered by interim analysis. The treatment assignment first must be unblinded. There should be some benefit of this design over blindfolded sample size recalculation processes to justify the disclosure of the treatment code, as unblinding of an ongoing study should be avoided if possible. In this study, they contrasted a number of sample size recalculation methods that include and exclude unblinding.

Charles et al. (2009) noticed that the primary goal of an a priori sample size calculation is to determine the minimum number of participants required to identify a treatment effect that is clinically significant. Some claim that large studies, which subject too many participants to the new medicine, and small, underpowered trials, which might not produce meaningful findings, should be avoided.

Calculating sample size typically involves using four parameters: Type-I error, power, control group assumptions (response rate and standard deviation), and predicted treatment impact.

Kieser and Friede (2003) discussed two-stage techniques, where the variance is reestimated from a subsample and the sample size is modified as needed, are appealing due to the uncertainty in the design step. From a regulatory perspective, it is crucial to maintain blindness and the capacity to estimate or manage the Type-I error rate. Several recommendations for sample size adjustment methods in the context of t-tests have recently been made. Sadly, none of these approaches meet both of these demands. 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 (2018) analyzed multi-centre randomized clinical trials, which is crucial for evidence-based medicine, providing advantages like accelerated recruitment and increased result generalizability. To address clustering in data, mixed models are used. However, 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-centre trials comparing two treatment groups. 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 centre heterogeneity. It is important to note that unbalanced treatment allocation can lead to power loss. Therefore, the proposed approach ensured accurate sample size determination and improved study planning, addressing potential limitations in previous methodologies.

Das et al. (2016) emphasized the significance of multi-centre randomized controlled clinical trials in evidence-based medicine, highlighting advantages such as accelerated recruitment, and increased result generalizability. To address potential

clustering in the data, the study employed mixed models. However, existing sample size calculation methods for mixed models typically assume balanced treatment allocations, which may not be feasible in practice. In response to this limitation, the paper introduced a novel sample size determination procedure for multi-centre trials, incorporating random effects and accommodating arbitrary sample sizes. Through simulations, the proposed method demonstrated its superiority over conventional approaches, taking into account factors like block length and centre heterogeneity. Furthermore, the study highlighted the impact of unbalanced treatment allocation on power loss if centre heterogeneity is overlooked during planning. The proposed approach aimed to ensure accurate sample size determination and improved study planning for multi-centre trials, providing valuable insights for researchers in the field of clinical research.

Chow and Chang (2008) discussed the adaptive design methods in clinical research, which have gained popularity due to their flexibility and efficiency based on accrued data. They can be categorized into prospective, concurrent (ad hoc), and retrospective adaptive designs. However, concerns arise about the deviation of the patient population and control of the overall Type-I error rate after adaptations, potentially leading to trials that fail to address intended scientific questions. Despite these concerns, adaptive designs are valued for reflecting medical practice, ethical considerations, and providing flexibility and efficiency in clinical development. Industry groups have proposed strategies to address these issues and ensure the validity and integrity of the trials. By understanding and implementing adaptive design methods, researchers can efficiently identify clinical benefits and increase the success of clinical development.

Jennison and Turnbull (2003) 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. It addresses situations where weak evidence of a positive treatment effect is obtained in an interim stage, leading to a desire to modify the design for increased power to detect smaller 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 shown potential advantages. However, the study cautioned that the flexibility to redesign an experiment mid-course may come at a substantial cost in terms of the required number of observations to correct the initial design.

Pritchett et al. (2015) compared different types of sample size recalculation (SSR) designs, including blinded SSR, unblinded SSR, and conventional group sequential designs (GSD). Operational logistics for implementing SSR designs

are discussed, along with recommendations for final data analysis and reporting. Uncertainties in confirmatory study designs can be mitigated by SSR, which helps avoid underpowered studies and potential failure of a compound in later development stages. The study presented statistical methods for unblinded and blinded SSR designs and highlights the importance of controlling Type-I error rate and accurately estimating the treatment effect. The advantages of unblinded SSR over GSD are discussed, and it is emphasized that SSR should be prespecified and adaptive by design. Case studies using SSR designs have shown promising results, and the appropriate application of SSR in drug development is expected to continue growing as industry, regulators, and academia gain experience and knowledge in this area.

Chakraborty and Gu (2009) discussed the prevalent challenges posed by missing values and dropouts in longitudinal studies within medical and public health fields. They focused on Intent-to-Treat (ITT) analysis as a key method for analyzing controlled clinical trials. Missing values, stemming from non-completion of follow-up per protocol, introduce complexities in ITT analysis. The study investigated this issue through simulation studies, compared various ad hoc strategies with the linear mixed model approach. Results indicated that, particularly for studies with high missing value rates, the mixed model approach imputation is more potent. The study emphasized the significance of missing data in longitudinal studies, impacting dataset balance, information loss, and introducing potential bias. The types of missing data mechanisms are delineated as missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). The ITT analysis is advocated for unbiased treatment effect estimation, irrespective of deviations. The study objective was to provide recommendations for ITT analysis with missing values, focused on the power and size of different methods in longitudinal design.

Boos and Brownie (1992) introduced novel rank-based methods within the framework of mixed linear models for analyzing data from multisite clinical trials. Unlike current rank methods, the newly proposed procedures specifically assess a drug's main effect in the presence of a random drug by site (or investigator) interaction. Corresponding procedures are also outlined for the fixed-effects scenario, with comparisons drawn against existing methods. The rationale for assuming random investigator effects is explained. Clinical trials often involve multiple investigators across different sites, randomly assigning subjects to new or standard drugs. The study addressed the challenge of inferring drug effects for the study population and broader target populations. Randomization-based tests, like the van Elteren test, offer validity based on random allocation, while fixed-effects analysis is suitable for broader population inferences. The mixed model, with random investigator and drug by investigator effects, is pertinent for

larger clinician populations. Nonrandom selection of investigators could lead to bias, affecting estimates and variance components. The proposed mixed-model analysis provided a more realistic estimate of variability for evaluating drug effects while addressing biases introduced by investigator selection. The study outlined rank-based alternatives for mixed-model ANOVA F tests and extends them to the fixed-effects scenario, illustrated through a motivating example.

Nagin and Odgers (2010) 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. The study underscored challenges linked to these models' implementation and proposed initial guidelines for researchers to adhere to when presenting model outcomes. Prospective avenues for group-based modeling are explored, including leveraging trajectory models to enable causal inference in cases where random treatment assignment isn't feasible. Overall, the study provided insight into the evolving role of group-based trajectory models in clinical research, their methodological intricacies, and their potential for advancing causal inference in challenging contexts.

Deng et al. (2022) discussed a novel two-stage multivariate Mendelian randomization method (MRMO) for investigating 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 MRMO addressed this limitation by jointly analyzing multiple outcomes using genetic instrumental variables. It was designed to handle both measured and unmeasured confounders. By applying the gradient descent algorithm and the adaptive sum of powered score (aSPU) test, the method demonstrated enhanced power in testing the overall hypothesis while controlling Type-I errors. The study showcased simulation experiments and a clinical application involving colorectal cancer patients to underscore the method's benefits over univariate MR analysis. This approach offered a promising advancement in clinical research by enabling the comprehensive assessment of causal relationships between clinical factors and complex mixed outcomes, aiding in hypothesis generation and identification of potential associations.

Spanbauer and Sparapani (2021) suggested precision medicine's transformative potential for clinical trials and subsequent treatment strategies. Traditionally, trials aim to uncover universal treatments, yet this approach might overlook varying treatment effects across population subsets. 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. The study introduced novel BART extensions tailored to precision medicine's unique inferential needs. These extensions encompass random effects for longitudinal data and subject clustering within medical centers. An innovative interaction detection prior is integrated to discern treatment heterogeneity, linked to patient characteristics. These advancements coalesce within the mixedBART framework. The study showcased simulation studies and real randomized clinical trial applications, illustrating precision medicine's potential by harnessing BART's predictive prowess to optimize treatment decisions.

Liang et al. (2003) discussed the dynamic relationship between virologic and immunologic responses in AIDS clinical trials, specifically analyzing plasma HIV RNA copies (viral load) and CD4+ cell counts. An innovative mixed-effects varying-coefficient model is proposed, addressing measurement errors in covariates and capturing the evolving interplay between these markers during antiviral treatments. The study, centered on the AIDS Clinical Trials Group's (ACTG 315) data, uncovers a time-dependent inverse association between viral load and CD4+ T cell counts during treatment initiation, followed by a gradual recovery after 8 weeks. The model accommodated varying associations among individuals and provided insights into monitoring virologic and immunologic markers longitudinally, crucial for AIDS clinical studies. The work introduced a tailored approach to unravel the intricate relationship between viral load and CD4+ cell counts, offered a comprehensive understanding of treatment dynamics and implications for precision medicine.

Morgan and Elashoff (1987) 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. A real-world application involved a clinical trial with advanced lung cancer patients illustrated the findings. By elucidating the effects of measurement error on prognostic factors and subsequent treatment effect estimation, the study provided valuable insights into refining the design and interpretation of clinical trials, particularly when considering the interplay between measurement errors and covariate adjustments.

Wang et al. (1998) 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. Focused on additive and normally distributed measurement error, coupled with a normally distributed error-prone predictor, the research revealed that the observed data adhere to a GLMM framework, albeit with distinct fixed and random effects structures and parameter restrictions.

This divergence lead to biases in common GLMMs when measurement error is disregarded. The investigation employed the SIMEX method for parameter estimation, offered a novel approach free from assumptions about unobservable predictors' structure. Illustrated through simulations and an empirical example involved advanced lung cancer patients, the study underscored the importance of accounting for measurement error and provided a comprehensive understanding of its impact on parameter estimation in clustered data analysis.

Yang et al. (2015) introduced a corrected empirical likelihood approach for statistical inference in generalized linear measurement error models, encompassing Gaussian, Poisson, and logistic regressions. By leveraging the corrected score function's moment identities, the method mitigated the adverse impact of measurement error on parameter estimation. The empirical log-likelihood ratio's asymptotic distribution was established as a Chi-squared distribution under certain regularity conditions, facilitated the derivation of the maximum empirical likelihood estimator for the regression parameter. Confidence intervals for specific components of the regression parameter were constructed using partial profile empirical likelihood. The proposed approach's efficacy was demonstrated through simulation studies and the analysis of real data from the ACTG 175 study. The corrected empirical likelihood method offered a robust means of handling measurement error-induced bias and uncertainty in generalized linear models, presented a practical solution for medical research scenarios where covariates may be subject to inaccuracies.

Brakenhoff et al. (2018) 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 make 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.

Chapter 3

Methodology

This chapter serves as a vital component in understanding the research process undertaken to address the research questions and objectives. In this section, a comprehensive overview of the research design, data collection methods, data analysis techniques, and any other pertinent procedures employed to ensure the validity and reliability of the study's findings are discussed

The methodology chosen for this research is a critical determinant of the study's rigor and its ability to draw meaningful conclusions from the collected data. It outlines the systematic approach adopted to gather and interpret information, thereby shedding light on how the research objectives will be achieved.

This chapter is organized as follows: First, the research philosophy and approach are discussed to establish the overarching framework guiding the study. Second, the research design is elaborated upon, including the type of investigation, the selection of participants or samples, and the rationale behind these decisions. Third, the methods employed for data collection are detailed, emphasizing their alignment with the research design and the chosen philosophical stance. Fourth, an overview of the data analysis techniques is provided, illustrating how the collected data will be processed, interpreted, and synthesized into meaningful outcomes.

In short, this chapter offers a road map for the entire research process, showcasing the logical sequence of decisions and actions that have been carefully planned to ensure the validity and reliability of the study's outcomes. By transparently detailing the research methodology, this thesis aims to establish a strong foundation for readers to assess the credibility and robustness of the findings presented in the subsequent chapters.

3.1 The research problem

We contemplate 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 and a same variance of σ^2.

$$X_i^T \sim N(\mu^T, \sigma^2), \quad i = 1, 2, ..., n$$
$$X_i^C \sim N(\mu^C, \sigma^2), \quad i = 1, 2, ..., n$$

We investigate the one-sided superiority test problem throughout this chapter.

$$H_0 : \mu^T - \mu^C \leq 0 \quad \text{vs} \quad H_1 : \mu^T - \mu^C > 0 \tag{3.1}$$

therefore referring 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,

$$T_i = \frac{\bar{X}_i^T - \bar{X}_i^C}{S_{pooled,i}} . \sqrt{\frac{n_i}{2}}, \tag{3.2}$$

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

$$S_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}, \tag{3.3}$$

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 to a futility stopping bound for one sided p-value of stage one, α_1 refers to the local one-sided significance level and q are the respective quantiles of 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 null hypothesis after the first stage, or accept the null hypothesis if $T_1 < q_{1-\alpha_0}$. After combining the all observed data over two stages by means of inverse normal combination test,as given as

$$T_{1+2} = \frac{w_1.T_1 + w_2.T_2}{\sqrt{w_1^2 + w_2^2}}, \tag{3.4}$$

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 $T_{1+2} \geq q_{1-\alpha_{1+2}}$, where α_{1+2} 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 or O'Brien and Fleming.

3.2 Methods for recalculating sample size

Various approaches exist for adjusting sample sizes during the course of 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 is contingent upon the true standardized treatment effect (σ), 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.

$$CP_{\Delta}(t_{1,n}) = \begin{cases} 0, \text{if the trial ends early due to futility,} \\ 1 - \Phi\left(q_{1-\alpha_{1+2}} \cdot \sqrt{\frac{w_1^2+w_2^2}{w_2^2}} - t_1 . \frac{w_1}{w_2} - \Delta . \sqrt{\frac{n_1}{2}} . \sqrt{\frac{n-n_1}{n_1}}\right), \\ \text{if the sample size is recalculated,} \\ 1, \text{if the trial is stopped early for efficacy.} \end{cases} \tag{3.5}$$

In the subsequent subsections, we 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 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 proposed resampling tool, 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 nmax on the overall total sample size per group, a measure employed for practical viability considerations.

3.2.1 The observed conditional power approach (OCP)

For observed interim test statistics to fall in the recalculation area $[q_{1-\alpha_0}; q_{1-\alpha_1})$, we want 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-β) is higher, it's better.

The basic concept is: we are doing some calculations 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

$$\tilde{n} \geq n_1 . \left(1 + \left(\frac{q_\beta - q_{1-\alpha_{1+2}} . \frac{\sqrt{w_1^2+w_2^2}}{w_2} + t_1 . \frac{w_1}{w_2}}{t_1} \right)^2 \right) \tag{3.6}$$

This equation helps us know how many people we need in each group so that our study makes sense and we can get meaningful results.

In line with 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 nmax. On the other hand, if t_1 does not falls within 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. The corresponding formula looks like

$$n_{\text{OCP}}(t_1) = \begin{cases} \min(\tilde{n}(t_1), n_{\max}), & \text{if } t_1 \in RA, \\ n_1, & \text{else.} \end{cases} \tag{3.7}$$

3.2.2 The restricted observed conditional power approach

The approach known as 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 observed conditional power (OCP) approach centers on a particular scenario: when the formula (3.6) 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. In light of this, a viable approach might be to consider enlarging the sample size only if it ensures a minimum acceptable conditional power denoted 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 ROCP approach is as follows

$$n_{\text{ROCP}}(t_1) = \begin{cases} \min(\tilde{n}(t_1), n_{\max}), & \text{if } t_1 \in \text{RA}, \\ & \text{and } CP(t_1, n_{\max}) \geq 1 - \beta_{low}^{ROCP}, \\ n_1, & \text{else.} \end{cases} \tag{3.8}$$

3.2.3 The promising zone approach (PZ)

The innovative promising zone (PZ) approach, introduced by Mehta and Pocock (2011), 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's 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 (3.6), 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 promising zone (PZ) approach equals

$$n_{\text{PZ}}(t_1) = \begin{cases} \min(\tilde{n}(t_1), n_{\max}), & \text{if } t_1 \in \text{RA and } 1 - \beta_{\text{PZ\_low}} \leq CP(t_1, n_{\text{ini}}) < 1 - \beta, \\ n_{\text{ini}}, & \text{if } t_1 \in \text{RA and } CP(t_1, n_{\text{ini}}) < 1 - \beta_{low}^{PZ}, \\ & \text{or } t_1 \in \text{RA and } CP(t_1, n_{\text{ini}}) \geq 1 - \beta, \\ n_1, & \text{else.} \end{cases} \tag{3.9}$$

3.3 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 for 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 is the conditional perspective, which prompts the researcher to consider how the sample size should be recalculated in the event that the observed effect falls within the recalculation area during the interim analysis. 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 thesis, 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, marked as $E[CN_{\Delta}^{RA}]$, which is the average size per group when we're 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 $n1$ 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 0 to 1, 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 CS. Mathematically, this score is calculated as

$$CS(\Delta) = \frac{1}{2} \cdot [SCP(\Delta) + SCN(\Delta)] \tag{3.10}$$

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. This adds more depth and details to how we evaluate things, for example, for conditional power sub-score

$$SCP(\Delta) = \gamma_{\text{loc}} \cdot e_{\text{CP}}(\Delta) + \gamma_{\text{var}} \cdot v_{\text{CP}}(\Delta), \text{ with } \gamma_{\text{loc}} + \gamma_{\text{var}} = 1, \tag{3.11}$$

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. For the purposes of this thesis, we opt for an equal weighting of all components, which means $\gamma_{loc} = \gamma_{var} = 0.5$.

3.4 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 1.

It's 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:

$$n_{(*)}^{R1}(t_1) = \frac{1}{B}\sum_{b=1}^{B}(\tilde{n}_{(*),i}(t_1)) \tag{3.12}$$

   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:

$$n_{(*)}^{R2}(t_1) = \frac{1}{B}\sum_{b=1}^{B}(\tilde{n}_{(*),i}(t_1)) + \frac{1}{B-1}\sqrt{\sum_{b=1}^{B}\left(\tilde{n}_{(*),i}(t_1) - \frac{1}{B}\sum_{b=1}^{B}(\tilde{n}_{(*),i}(t_1))\right)^2} \tag{3.13}$$

This inclusion of the standard deviation means that we tend to select larger sample sizes. We term this approach 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.

3.5 Simulation study for evaluating the performance of sample size recalculation approaches

To comprehensively assess the effectiveness of the various sample size recalculation methods outlined earlier, a simulation study was undertaken. These approaches were meticulously evaluated through specific performance measures, including the novel conditional performance score (3.10), with parameter values of $\gamma_{loc} = \gamma_{var} = 0.5$.

In this simulation study, we adhered to the design specifications detailed in Section 3.1. This involved working with groups of equal size, with the first stage consisting of $n_1 = 50$ participants. The initial second stage sample size per group was 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 was set at four times the interim sample size n_1, resulting in $n_{max} = 200$. The weights for the inverse normal combination test were uniformly assigned as $w_1 = w_2 = \sqrt{50}$. Our chosen global one-sided significance level was $\alpha = 0.025$, while the local significance levels were calculated according to the Pocock method, specifically $\alpha_1 = \alpha_{1+2} = 0.0147$. Additionally, a futility bound of $\alpha_0 = 0.5$ was established.

A desired level of conditional power was set at $1 - \beta = 0.8$. For the ROCP, the lower bound for conditional power ($1 - \beta_{\text{low}}^{\text{ROCP}}$) was held at 0.6. Similarly, for the promising zone (PZ), a lower bound ($1 - \beta_{\text{low}}^{\text{PZ}}$) was fixed at 0.36, following the approach proposed by Mehta and Pocock (2011). 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 were 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 was 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.

3.5.0.1 Generation of treatment and control groups

1. **Measurement error model**
   MEMs are statistical tools 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.

   Measurement error models 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:

   (a) **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,
   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 and ϵ is the error in the observed Y.

   (b) **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 error. 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.

Impact of Measurement Error on the Reliability of Estimates:

Measurement errors can significantly affect the reliability and validity of the estimates derived from statistical analyses. The key implications of measurement errors on estimates include:

(a) Bias: 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.

(b) Efficiency Loss: 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.

(c) Inconsistency: 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.

(d) Incorrect Hypothesis Testing: 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.

(e) Misinterpretation of Relationships: Measurement errors can lead to misinterpretation of the true relationships between variables. Researchers may overestimate or underestimate the strength of associations, leading to misguided policy recommendations or interventions.

For simulation purpose, initially we generated treatment group from MEM and control group from normal distribution having n=50, μ= 0.3 and σ= 1, i.e., control=rnorm(50,0.3,1). Also, we generated both treatment and control groups from MEM and run the simulation.

2. **Mixed distribution models:**
We discuss the methodology employed to generate treatment and control groups for simulation using a mixed distribution model (MDM). The 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 with respect to 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 $\sum_{i=1}^{k} \pi_i = 1$.
- $f_i(x;\theta_i)$ represents the ith component distribution with parameters θ_i.

Advantages of MDM

- Enhanced Balance: 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.
- Sample Representativeness: By accounting for the underlying distribution of the data, the MDM helps ensure that the generated groups accurately represent the population.

For simulation purpose, we generated treatment and control groups of length n=50 from MDM. One distribution is taken as standard normal distribution,i.e., N(0,1), while, the other distribution is also taken as normal having μ= 0.5 and σ= 1, i.e., N (0.5,1). However, the mixing proportion is chosen 0.1, 0.5, i.e., p=0.1, 0.5.

3.6 Results from mixed distribution model

In this section, we discuss the results with the help of conditional power score (CPS). The criteria to access the performance is that which approach has higher conditional performance score than its respective approach is a better performer. In the standard sample size recalculation without resampling, we can see that

group sequential design performed well. The reason behind this is that there is no variation in recalculated sample sizes. If we access the performance of R1 approach, this approach consider mean as summary location measure, performs better than respective standard sample size recalculation approach without resampling with respect to CPS for all delta (Δ), true standardized effect sizes. The reason behind the better performance of R1 approach over standard sample size recalculation approach without resampling is that resampling approach (R1 approach) reduces the variability in recalculated sample sizes for all Δ.

Furthermore, in R1 approach the OCP performed either better than the GS design or have similar conditional performance score as compared to the ROCP and PZ(3.1, column 3 and 4). The reason behind this is better conditional power of the OCP (3.2 and 3.3). It is to be noted that there is an observable tendency towards increasing the initial stage's sample size, denoted as n1, when utilizing sample size recalculation via the R1 approach (3.3). This phenomenon arises because, under the R1 approach, test statistics that falls outside the recalculated area (RA) could undergo resampling, even if the interim test statistic actually falls within RA. To address this concern R2 approach has been developed. This alternative approach (R2 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 3.4 that R2 approach has tendency to move towards GS design due to n_{max} as sample size boundary.

In R2 approach, the ROCP performed well for almost all delta with respect to the CPS (3.1, column 5) compared to the OCP and PZ approach. Overall, the R2 approaches secured distinguished position against original sample size recalculation without resampling, However, the GS design outperforms the different designs in R2 approach for all Δ. It is to be noted that CPS of respective designs in R2 approach is worse than R1 approach for same value of $\Delta \in (0.0, 0.1, 0.2, 0.3, 0.4, 0.5)$. The reason behind this worst behaviour of R2 approach is that sample size recalculation designs in 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 (Table 3.3 and Table 3.4).

Table 3.1: Conditional power score when both treatment and control generated from MDM with p.mix=0.5

Delta	Design	Standard Simulation	R1 Approach	R2 approach
0	OCP	0.3780	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	0.3780	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	0.3780	0.3981	0.4306
0.2	ROCP	0.4255	0.4798	0.5820
0.2	PZ	0.4978	0.5490	0.5943
0.2	GSD	0.6741	0.7105	0.7105
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	0.6850	0.7212	0.7212

Table 3.2: Standard simulation when both treatment and control generated from MDM with p.mix=0.5

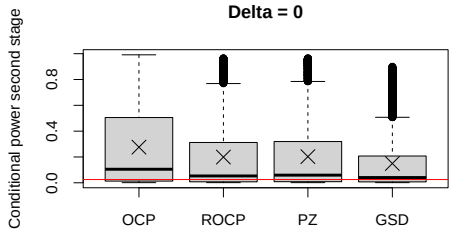
Δ	Design	mean_cp	e_cp	var_cp	v_cp	score_cp	mean_n	e_n	var_n	v_n	score_n	score_cond
0	OCP	0.5421	0.4696	0.0924	0.3922	0.4309	167.0027	0.2200	1825.7502	0.4303	0.3251	0.3780
0	ROCP	0.4451	0.5691	0.1504	0.2244	0.3968	102.5864	0.6494	3087.8035	0.2591	0.4543	0.4255
0	PZ	0.4626	0.5512	0.1244	0.2946	0.4229	117.4977	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 3.3: R1 Approach when both treatment and control generated from MDM with p.mix=0.5

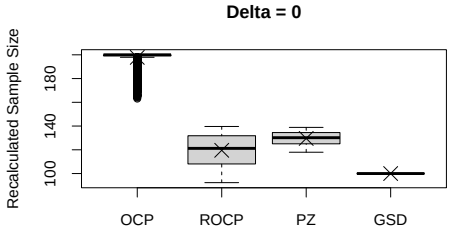
Delta	Metric	mean_cp	e_cp	var_cp	v_cp	score_cp	mean_n	e_n	var_n	v_n	score_n	score_cond
0	OCP	0.2609	0.7580	0.0877	0.4077	0.5828	192.0012	0.0533	592.1704	0.6756	0.3644	0.4736
0	ROCP	0.1579	0.8637	0.0960	0.3803	0.6220	73.1073	0.8460	2388.8495	0.3484	0.5972	0.6096
0	PZ	0.1801	0.8409	0.0762	0.4480	0.6445	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 3.4: R2 Approach when both treatment and control generated from MDM with p.mix=0.5

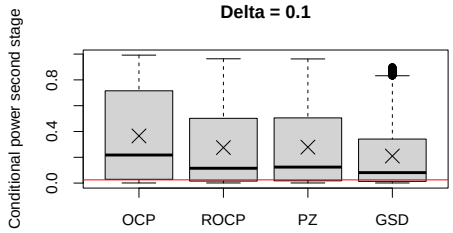
Delta	Metric	mean_cp	e_cp	var_cp	v_cp	score_cp	mean_n	e_n	var_n	v_n	score_n	score_cond
0	OCP	0.2759	0.7427	0.1058	0.3494	0.5460	197.8248	0.0145	30.7337	0.9261	0.4703	0.5082
0	ROCP	0.1992	0.8213	0.0739	0.4564	0.6389	119.5676	0.5362	167.0806	0.8277	0.6819	0.6604
0	PZ	0.2027	0.8178	0.0735	0.4578	0.6378	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	0.6510	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	0.6000	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



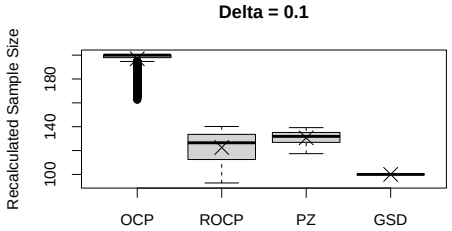
(a) Plot 1



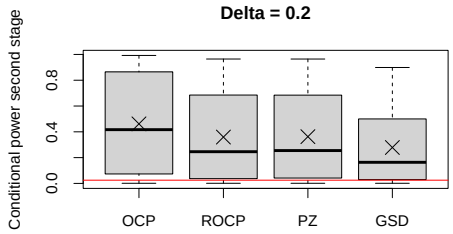
(b) Plot 2



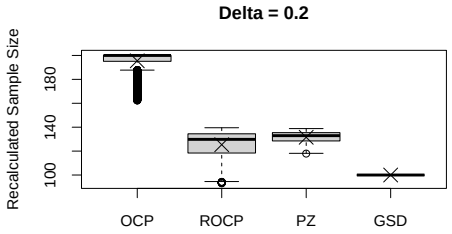
(c) Plot 3



(d) Plot 4

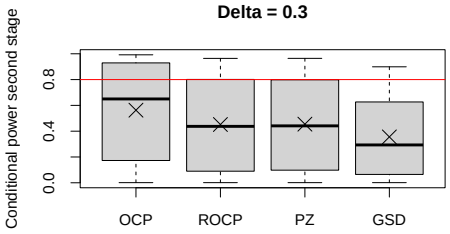


(e) Plot 5

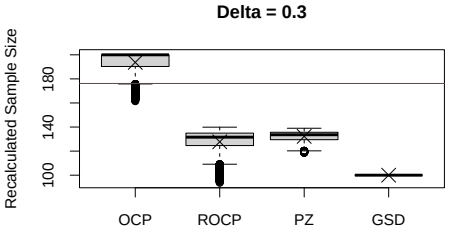


(f) Plot 6

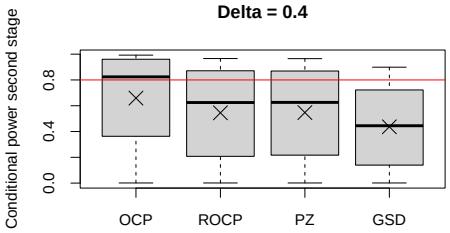
Figure 3.1: Standard simulation when both treatment and control generated from MDM with p.mix=0.5



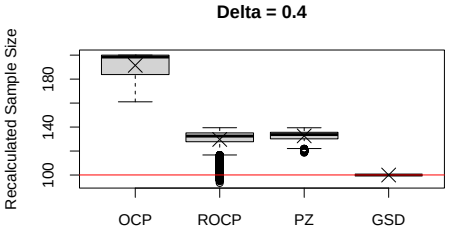
(a) Plot 7



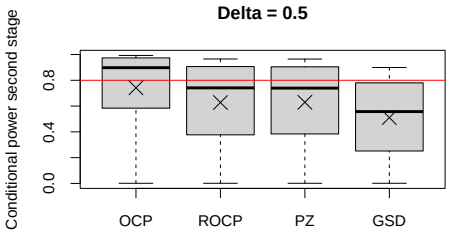
(b) Plot 8



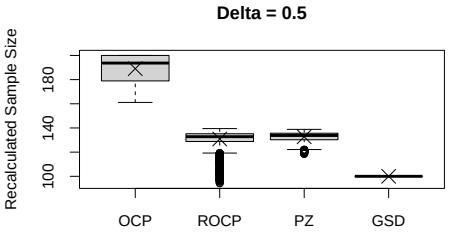
(c) Plot 9



(d) Plot 10

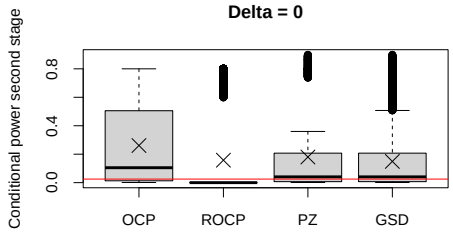


(e) Plot 11

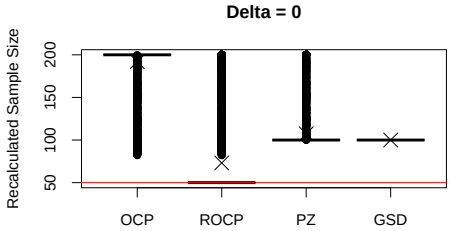


(f) Plot 12

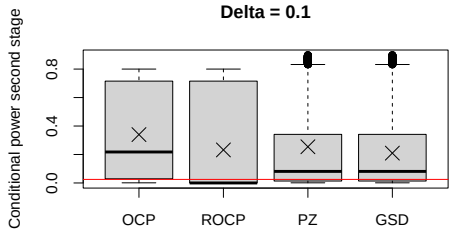
Figure 3.2: Standard simulation when both treatment and control generated from MDM with p.mix=0.5



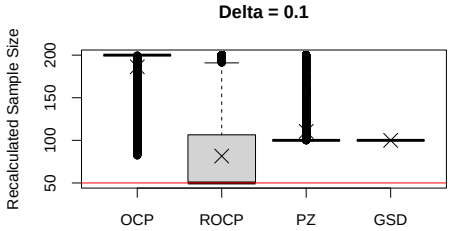
(a) Plot 1



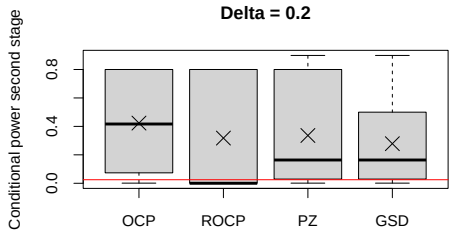
(b) Plot 2



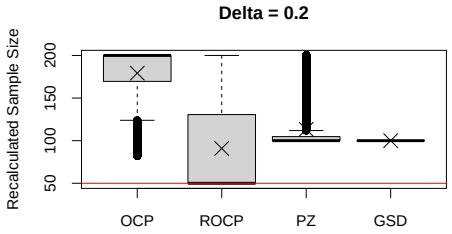
(c) Plot 3



(d) Plot 4

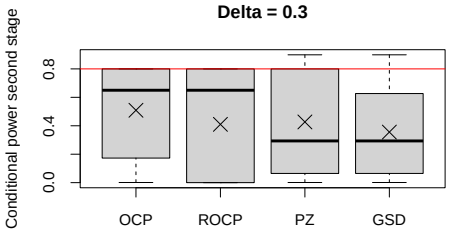


(e) Plot 5

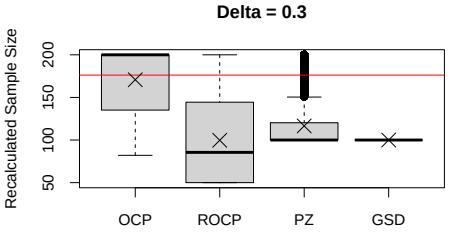


(f) Plot 6

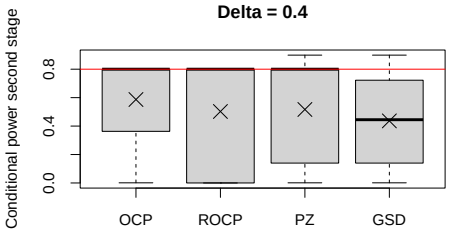
Figure 3.3: R1 Approach when both treatment and control generated from MDM with p.mix=0.5



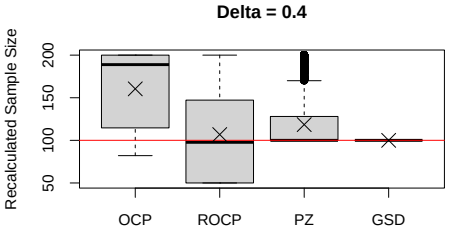
(a) Plot 7



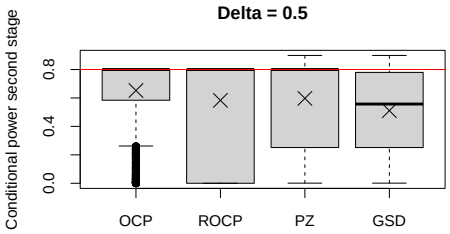
(b) Plot 8



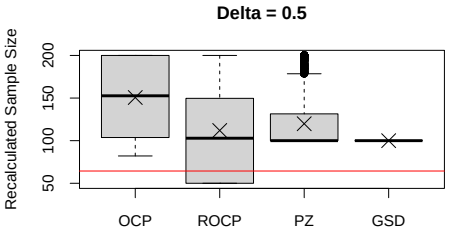
(c) Plot 9



(d) Plot 10

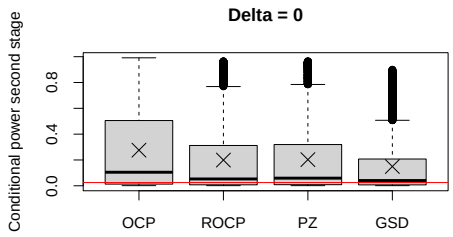


(e) Plot 11

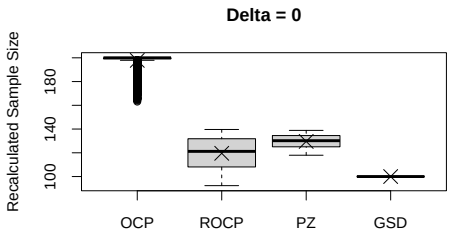


(f) Plot 12

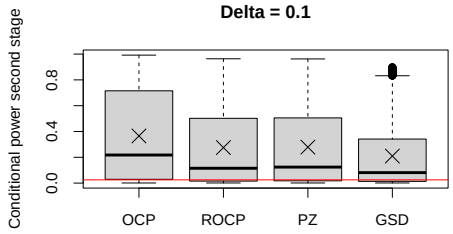
Figure 3.4: R1 Approach when both treatment and control generated from MDM with p.mix=0.5



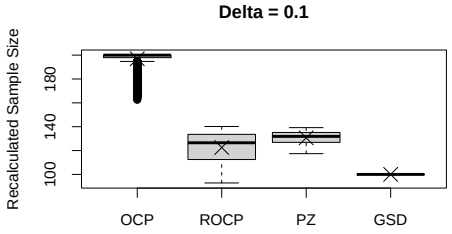
(a) Plot 1



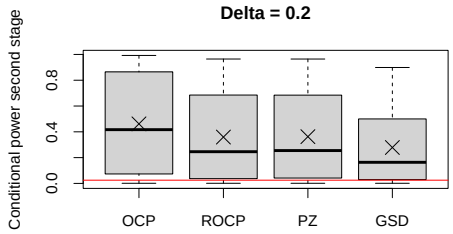
(b) Plot 2



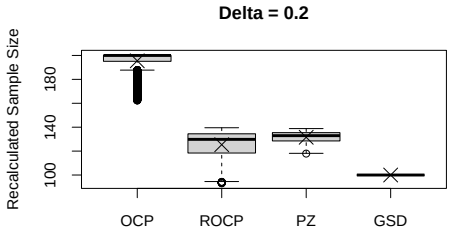
(c) Plot 3



(d) Plot 4

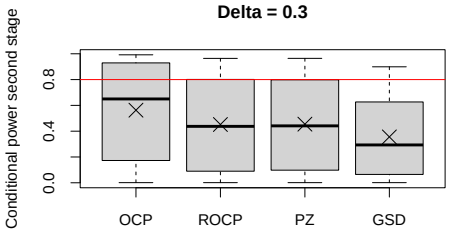


(e) Plot 5

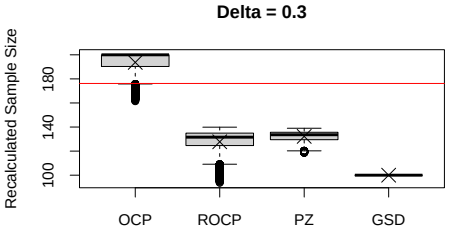


(f) Plot 6

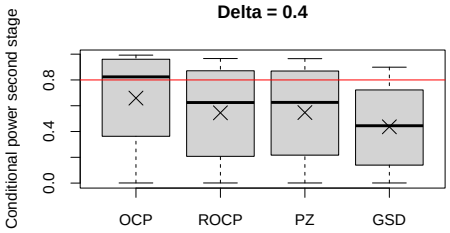
Figure 3.5: R2 Approach when both treatment and control generated from MDM with p.mix=0.5



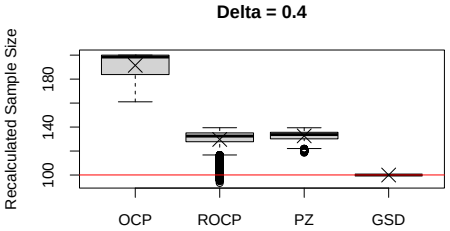
(a) Plot 7



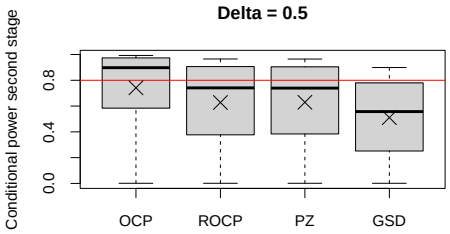
(b) Plot 8



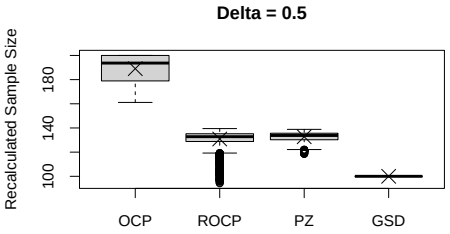
(c) Plot 9



(d) Plot 10



(e) Plot 11



(f) Plot 12

Figure 3.6: R2 Approach when both treatment and control generated from MDM with p.mix=0.5

Table 3.5: Conditional power score when both treatment and control generated from MDM with p.mix=0.1

Delta	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

In standard sample size recalculation, GSD is the performance winner (3.5, Column 3). This is due to the reason that there is no variation in recalculated sample sizes. R1 approach performed well as compared to standard sample size recalculation for all delta values in all designs. The reason behind better performance of R1 approach against standard simulation is that resampling approach (R1 approach) reduces the variability in recalculated sample sizes for all Δ (3.7). Furthermore, in R1 approach, the ROCP and PZ have approximately same CPS as GSD. While, the OCP have little less CPS than GSD. This is due to less conditional power of OCP. While recalculating sample size using R1 approach ,this approach has tendency to increase initial sample size, denoted as n_1 (3.7). This is due the reason that if the test statistics falls outside the recalculation area (RA), could go through resampling, even if interim test statistics falls within recalculated area (RA). To overcome this issue R2 approach is introduced ,this approach relies on mean as well as standard deviation of resampled sample sizes as summary location measure. The effect of this can be seen in (3.8). The R2 aprach

tends to recalculate sample size upto maximum allowed sample size n_{max}. Overall, R2 approach has secured distinct position against R1 approach and standard simulation. However, GSD in standard simulation outperformed various designs in R2 approach for all Δ. The detailed simulation results are presented in (3.6), (3.7), (3.7). The box plots for second stage conditional power and recalculated sample size for standard simulation, R1 approach and R2 approach are illustrated in (3.7, 3.8), (3.9,3.10), (3.11,3.12) respectively.

Table 3.6: Standard simulation when both treatment and control generated from MDM with p.mix=0.1

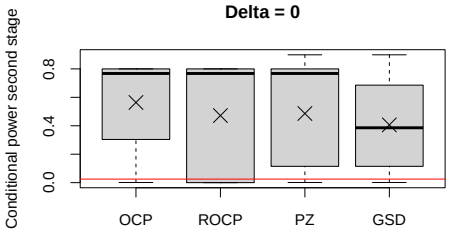
Delta	Metric	mean_cp	e_cp	var_cp	v_cp	score_cp	mean_n	e_n	var_n	v_n	score_n	score_cond
0	OCP	0.5638	0.4474	0.0877	0.4079	0.4276	164.9150	0.2339	1871.2579	0.4233	0.3286	0.3781
0	ROCP	0.4717	0.5418	0.1469	0.2335	0.3876	105.5823	0.6295	3087.6155	0.2592	0.4443	0.4160
0	PZ	0.4862	0.5270	0.1218	0.3019	0.4145	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	0.4044	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	0.5962	0.0887	0.4044	0.5003	100.0000	0.7626	0.0000	1.0000	0.8813	0.6908

Table 3.7: R1 Approach when both treatment and control generated from MDM with p.mix=0.1

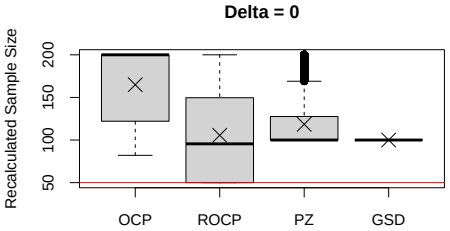
Delta	Metric	mean_cp	e_cp	var_cp	v_cp	score_cp	mean_n	e_n	var_n	v_n	score_n	score_cond
0	OCP	0.4993	0.5136	0.0969	0.3775	0.4455	129.9407	0.4671	155.9233	0.8335	0.6503	0.5479
0	ROCP	0.3276	0.6896	0.0653	0.4891	0.5894	79.6770	0.8022	31.8719	0.9247	0.8634	0.7264
0	PZ	0.4035	0.6118	0.0799	0.4347	0.5233	101.0800	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	0.6164
0.5	ROCP	0.3276	0.5155	0.0653	0.4891	0.5023	79.6695	0.8982	31.6864	0.9249	0.9116	0.7069
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	0.4044	0.5003	100.0000	0.7626	0.0000	1.0000	0.8813	0.6908

Table 3.8: R2 approach when both treatment and control generated from MDM with p.mix=0.1

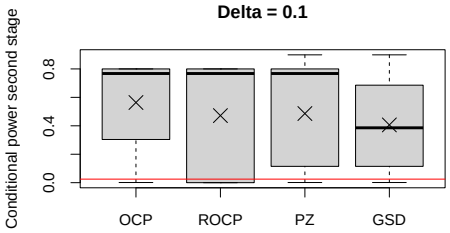
Delta	Metric	mean_cp	e_cp	var_cp	v_cp	score_cp	mean_n	e_n	var_n	v_n	score_n	score_cond
0	OCP	0.6282	0.3814	0.1240	0.2958	0.3386	192.5196	0.0499	113.2262	0.8581	0.4540	0.3963
0	ROCP	0.5122	0.5003	0.1152	0.3211	0.4107	129.2916	0.4714	78.1722	0.8821	0.6768	0.5437
0	PZ	0.5146	0.4979	0.1133	0.3268	0.4124	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	0.0500	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	0.7048	0.1152	0.3211	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



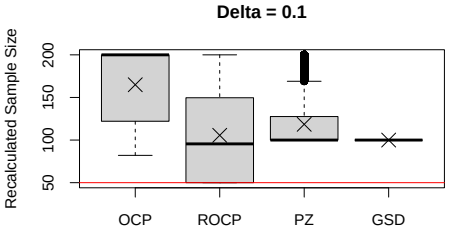
(a) Plot 1



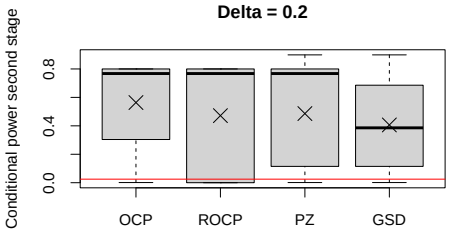
(b) Plot 2



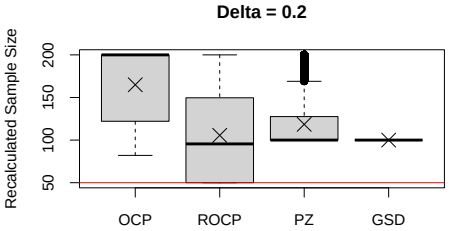
(c) Plot 3



(d) Plot 4

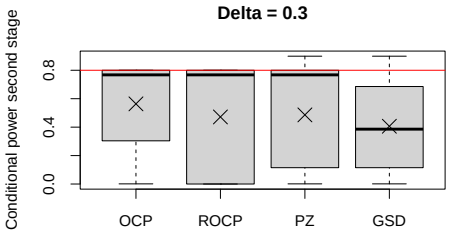


(e) Plot 5

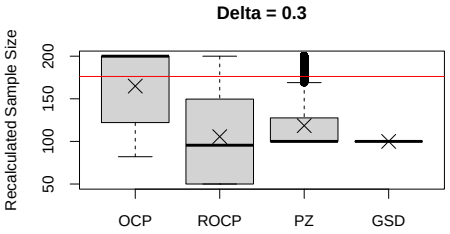


(f) Plot 6

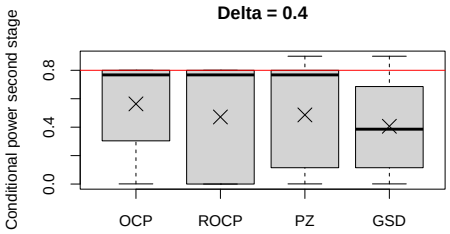
Figure 3.7: Standard simulation when both treatment and control generated from MDM with p.mix=0.1



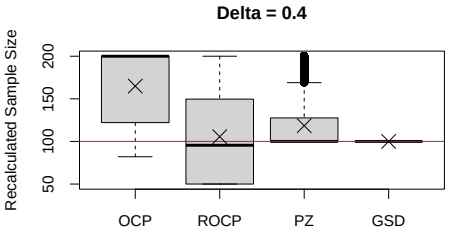
(a) Plot 1



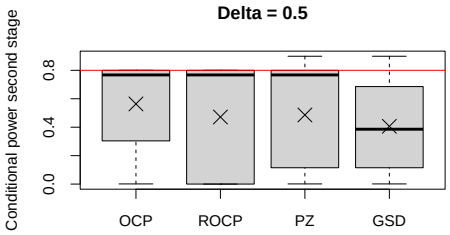
(b) Plot 2



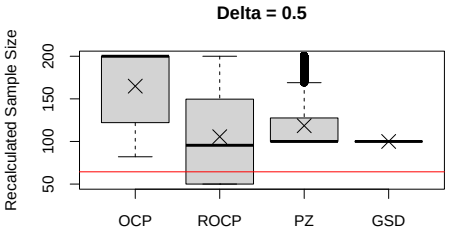
(c) Plot 3



(d) Plot 4

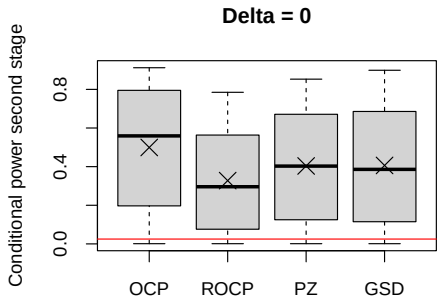


(e) Plot 5

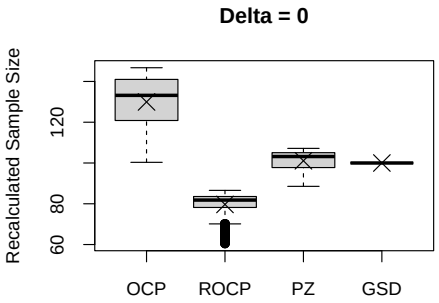


(f) Plot 6

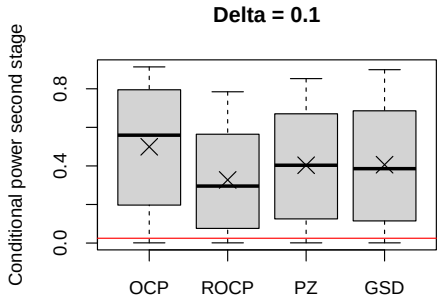
Figure 3.8: Standard simulation when both treatment and control generated from MDM with p.mix=0.1



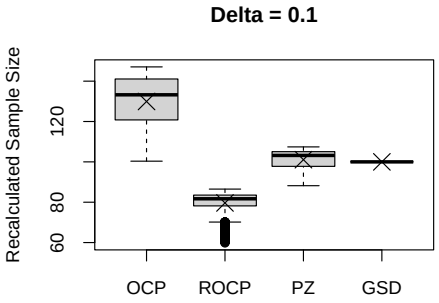
(a) Plot 1



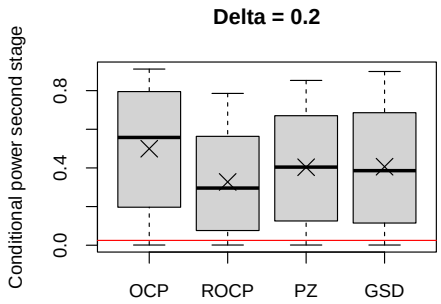
(b) Plot 2



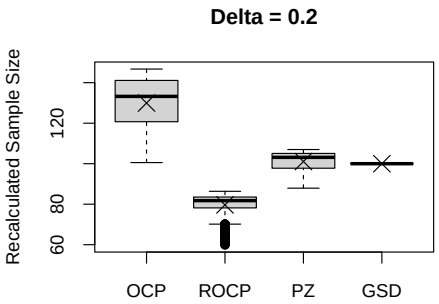
(c) Plot 3



(d) Plot 4

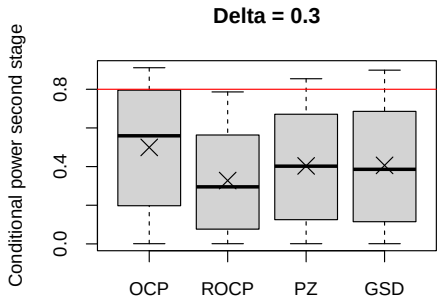


(e) Plot 5

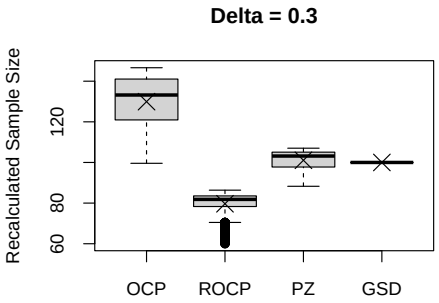


(f) Plot 6

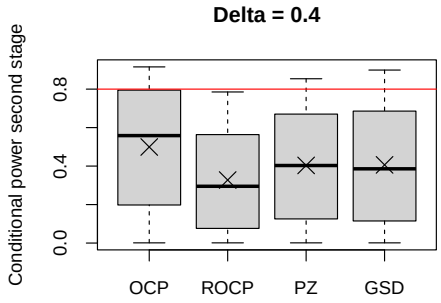
Figure 3.9: R1 Approach when both treatment and control generated from MDM with p.mix=0.1



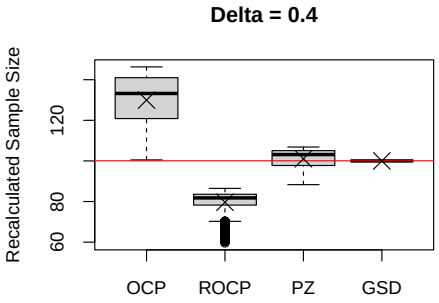
(a) Plot 1



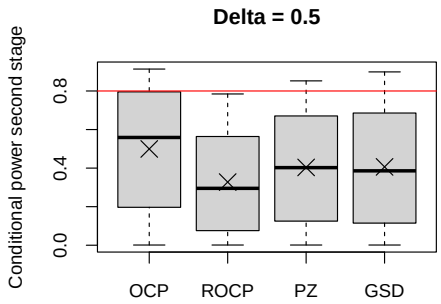
(b) Plot 2



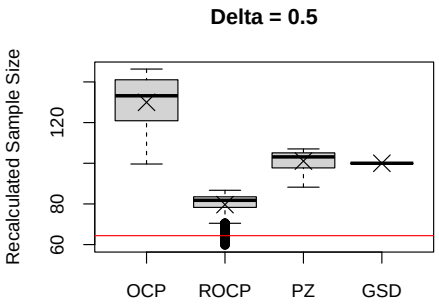
(c) Plot 3



(d) Plot 4

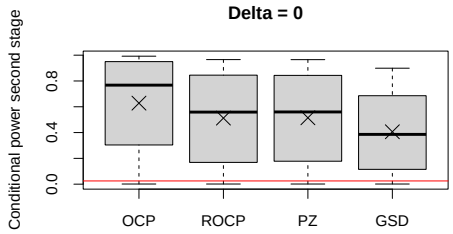


(e) Plot 5

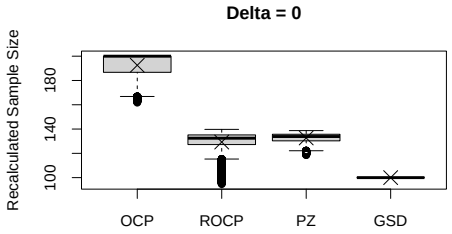


(f) Plot 6

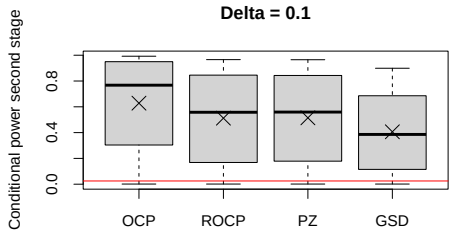
Figure 3.10: R1 Approach when both treatment and control generated from MDM with p=0.1



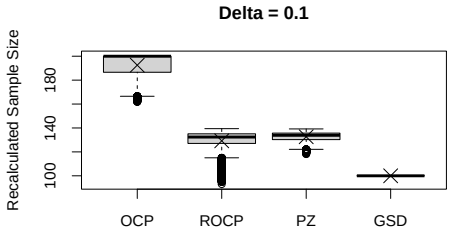
(a) Plot 1



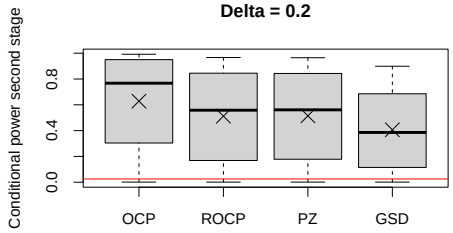
(b) Plot 2



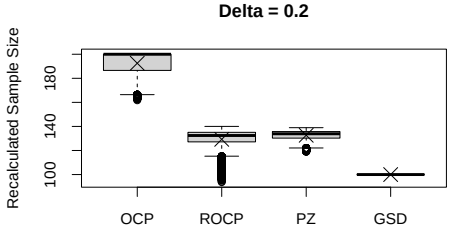
(c) Plot 3



(d) Plot 4

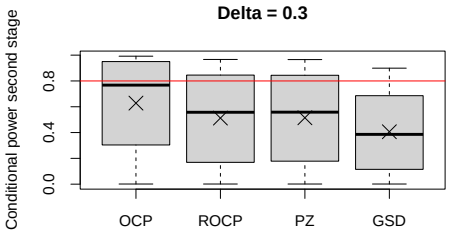


(e) Plot 5

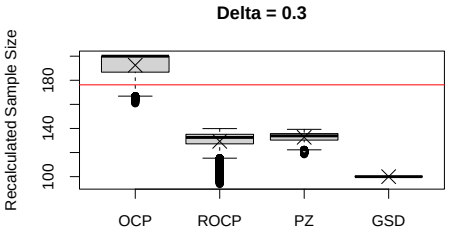


(f) Plot 6

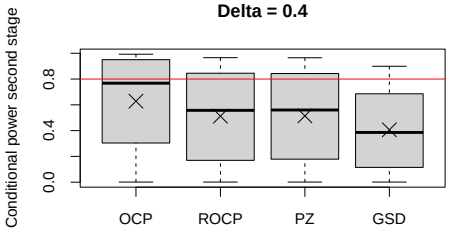
Figure 3.11: R2 Approach when both treatment and control generated from MDM with p=0.1



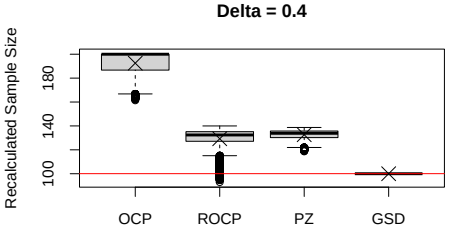
(a) Plot 1



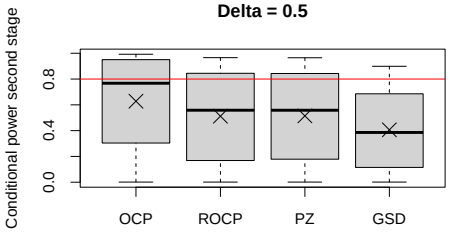
(b) Plot 2



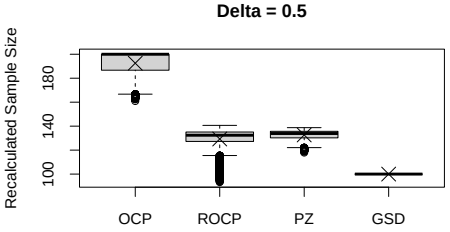
(c) Plot 3



(d) Plot 4



(e) Plot 5



(f) Plot 6

Figure 3.12: R2 Approach when both treatment and control generated from MDM with p=0.1

3.7 Results from measurement error model (MEM)

Table 3.9: Conditional power score when both treatment and control generated from MEM

Delta	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

From Table (3.9) one can see that the GSD is the performance winner in standard simulation without resampling. This is due to fact recalculated sample sizes have no variation. Moreover, it is interesting to note that CPS in standard simulation remained same for Δ= 0,0.1,0.2,0.3. In R1 approach, the ROCP and PZ have approximately the same CPS, while the OCP has less value of the CPS. Furthermore, the performance of R1 approach, this approach consider mean as summary location measure, is better against standard simulation. This is due to decreased variation by 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 generation of treatment and control groups from MEM. The detailed simulation results are presented in (3.10), (3.11), (3.12). The box plots for second stage conditional power and recalculated sample size for standard simulation, R1 approach and R2 approach are illustrated in (3.13, 3.14), (3.15), (3.16,3.17)

respectively.

Table 3.10: Standard simulation results when both treatment and control generated from MEM

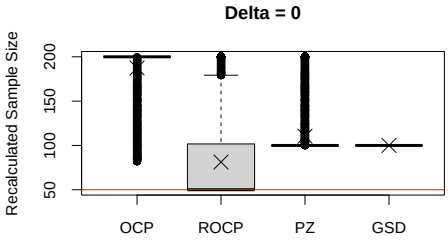
Delta	Metric	mean_cp	e_cp	var_cp	v_cp	score_cp	mean_n	e_n	var_n	v_n	score_n	score_cond
0	OCP	0.3319	0.6852	0.1003	0.3668	0.5260	187.4409	0.0837	872.6132	0.6062	0.3449	0.4355
0	ROCP	0.2240	0.7959	0.1226	0.2997	0.5478	81.2458	0.7917	2857.3641	0.2873	0.5395	0.5436
0	PZ	0.2453	0.7740	0.0987	0.3717	0.5729	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 3.11: R1 approach when both treatment and control generated from MEM

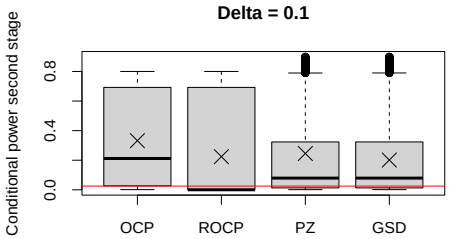
Delta	Metric	mean_cp	e_cp	var_cp	v_cp	score_cp	mean_n	e_n	var_n	v_n	score_n	score_cond
0	OCP	0.2674	0.7514	0.0835	0.4220	0.5867	134.2499	0.4383	91.6999	0.8723	0.6553	0.6210
0	ROCP	0.1566	0.8650	0.0426	0.5874	0.7262	75.0924	0.8327	57.6779	0.8987	0.8657	0.7960
0	PZ	0.2043	0.8161	0.0587	0.5155	0.6658	103.4696	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 3.12: R2 Approach when both treatment and control generated from MEM

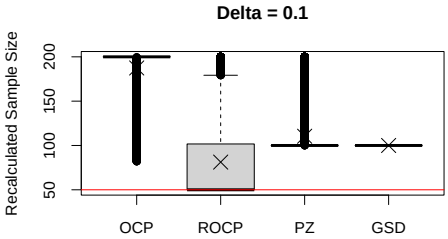
Delta	Metric	mean_cp	e_cp	var_cp	v_cp	score_cp	mean_n	e_n	var_n	v_n	score_n	score_cond
0	OCP	0.2674	0.7514	0.0835	0.4220	0.5867	134.2499	0.4383	91.6999	0.8723	0.6553	0.6210
0	ROCP	0.1566	0.8650	0.0426	0.5874	0.7262	75.0924	0.8327	57.6779	0.8987	0.8657	0.7960
0	PZ	0.2042	0.8161	0.0587	0.5155	0.6658	103.4696	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



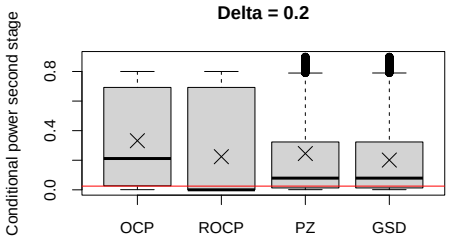
(a) Plot 2



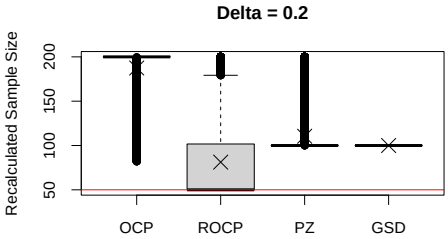
(b) Plot 3



(c) Plot 4

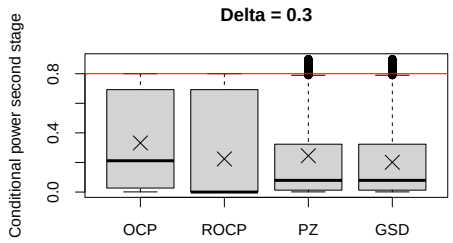


(d) Plot 5

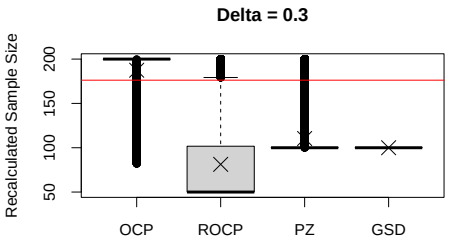


(e) Plot 6

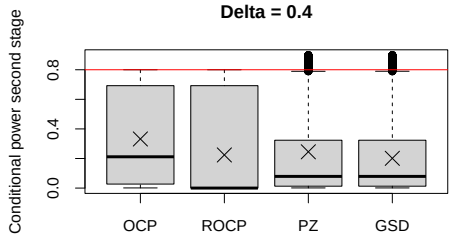
Figure 3.13: Standard simulation when both treatment and control generated from MEM



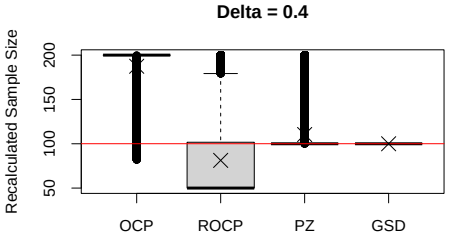
(a) Plot 1



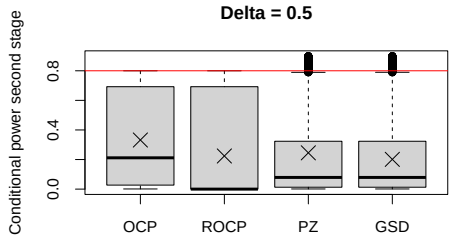
(b) Plot 2



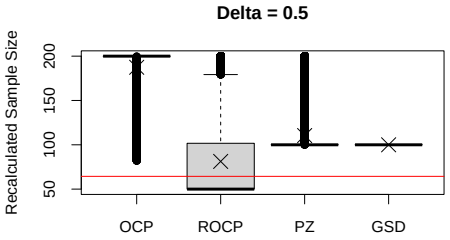
(c) Plot 3



(d) Plot 4

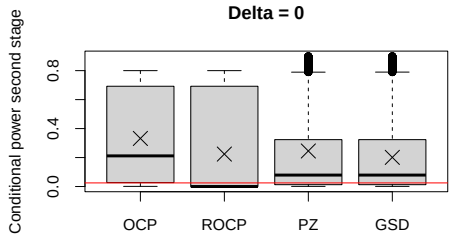


(e) Plot 5

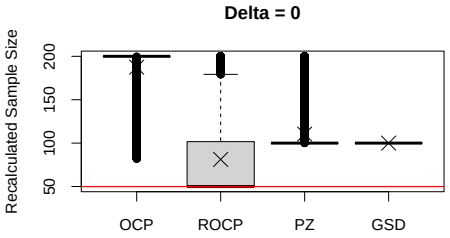


(f) Plot 6

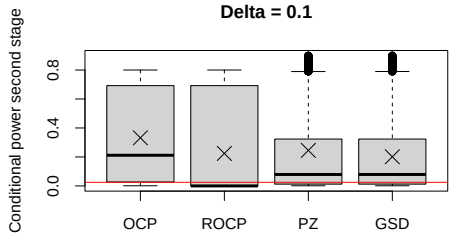
Figure 3.14: Standard simulation when both treatment and control generated from MEM



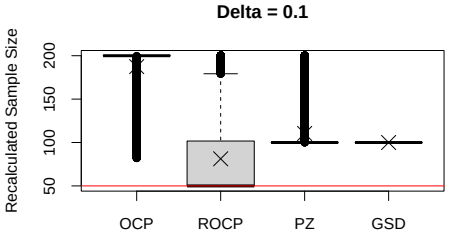
(a) Plot 1



(b) Plot 2

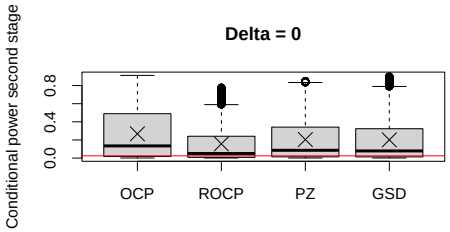


(c) Plot 3

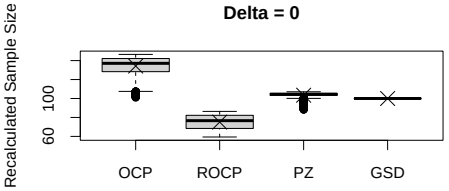


(d) Plot 4

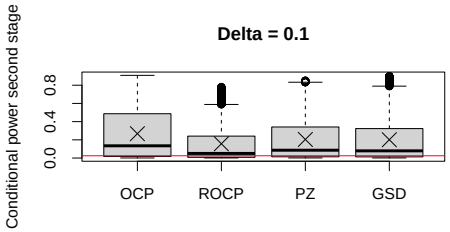
Figure 3.15: R1 Approach when both treatment and control generated from MEM



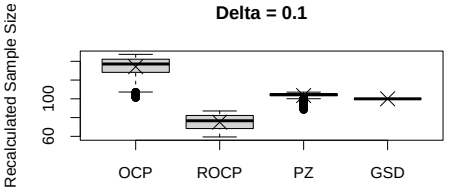
(a) Plot 1



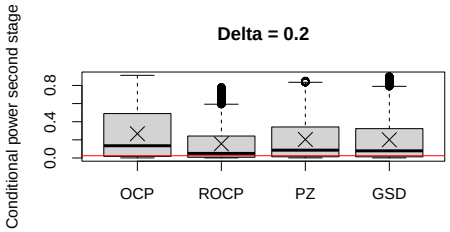
(b) Plot 2



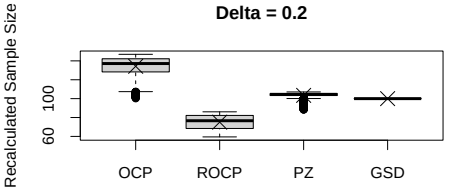
(c) Plot 3



(d) Plot 4

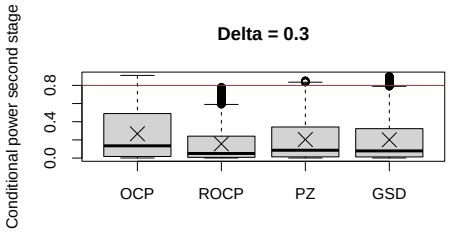


(e) Plot 5

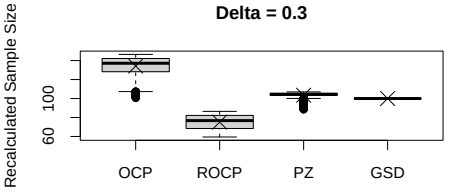


(f) Plot 6

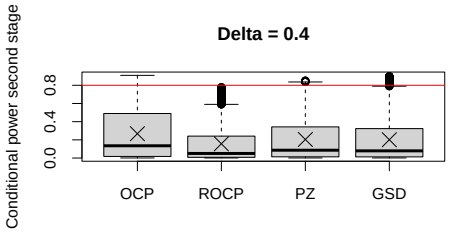
Figure 3.16: R2 Approach when both treatment and control generated from MEM



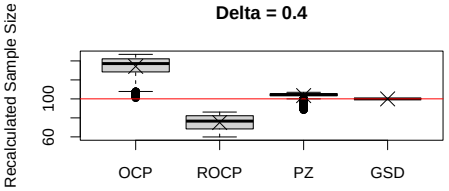
(a) Plot 1



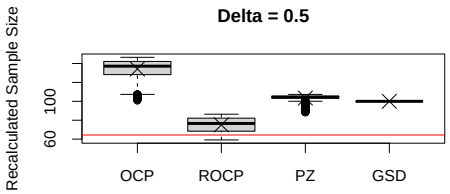
(b) Plot 2



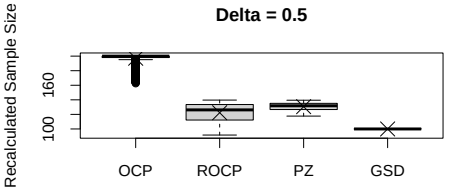
(c) Plot 3



(d) Plot 4



(e) Plot 5



(f) Plot 6

Figure 3.17: R2 Approach when both treatment and control generated from MEM

Table 3.13: Conditional power score where only treatment is generated with MEM and control from a normal distribution

Delta	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	0.7650	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	0.4790	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	0.6650	0.6650	0.6650

In standard sample size recalculation, GSD performed well against all other designs for all delta(3.13) This is due to no variation in recalculated sample size. If we compare the designs in R1 approach with GSD we can see that ROCP and PZ have approximately similar performance as GSD, while OCP have a little less score. This is due to better conditional power of ROCP and PZ against OCP. The R1 approach demonstrates competitive performance across different delta 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(3.16. This behavior can be seen in the similar CPSs for certain scenarios. The CPS of R1 approach is higher than R2 approach. This is due to lack of R2 to target the values of sample size effectively. The detailed simulation results are presented in (3.14), (3.15), (3.16). The box plots for second stage conditional power and recalculated sample size for standard simulation, R1 approach and R2 approach are illustrated in (3.18), (3.19) respectively.

Table 3.14: Standard simulation where only treatment is generated with MEM and control from a normal distribution

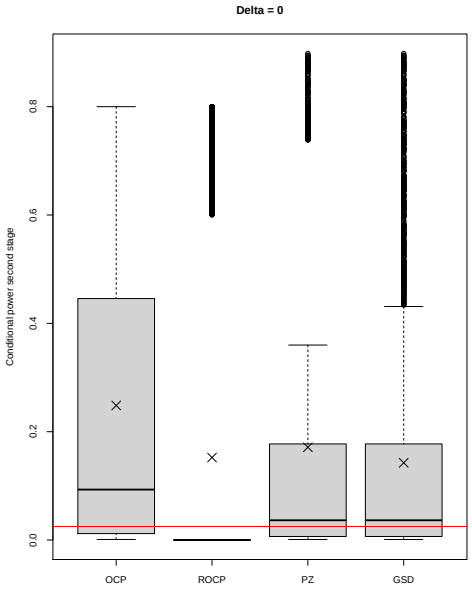
Delta	Metric	mean_cp	e_cp	var_cp	v_cp	score_cp	mean_n	e_n	var_n	v_n	score_n	score_cond
0	OCP	0.2484	0.7708	0.0858	0.4141	0.5925	192.1952	0.0520	569.4070	0.6819	0.3669	0.4797
0	ROCP	0.1523	0.8694	0.0933	0.3890	0.6292	72.2506	0.8517	2302.4635	0.3602	0.6060	0.6176
0	PZ	0.1713	0.8500	0.0737	0.4570	0.6535	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 3.15: R1 Approach where only treatment is generated with MEM and control from a normal distribution

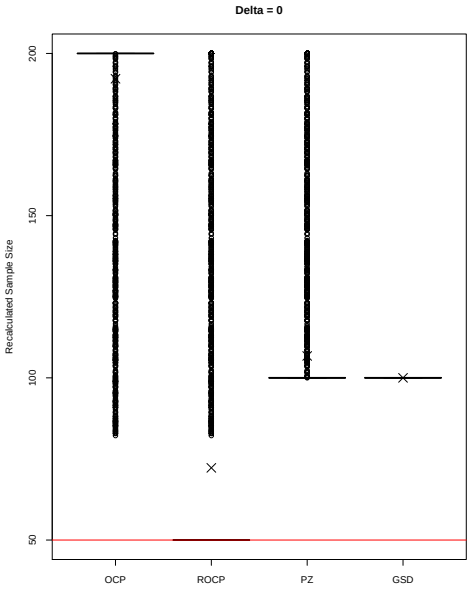
Delta	Metric	mean_cp	e_cp	var_cp	v_cp	score_cp	mean_n	e_n	var_n	v_n	score_n	score_cond
0	OCP	0.1944	0.8262	0.0664	0.4848	0.6555	134.2621	0.4383	79.1251	0.8814	0.6598	0.6577
0	ROCP	0.1095	0.9134	0.0310	0.6481	0.7807	73.0042	0.8466	59.4450	0.8972	0.8719	0.8263
0	PZ	0.1459	0.8760	0.0445	0.5782	0.7271	103.8817	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	0.7794
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	79.7480	0.8809	0.8007	0.6162
0.3	ROCP	0.1095	0.2918	0.0310	0.6477	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	0.4319	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	0.1459	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 3.16: R2 approach where only treatment is generated with MEM and control from a normal distribution

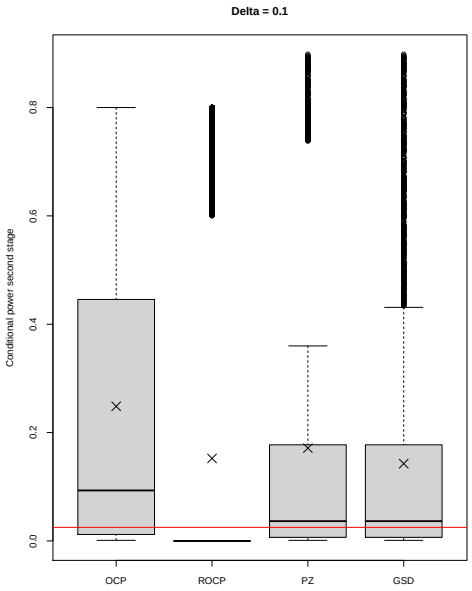
Delta	Metric	mean_cp	e_cp	var_cp	v_cp	score_cp	mean_n	e_n	var_n	v_n	score_n	score_cond
0	OCP	0.2631	0.7558	0.1039	0.3552	0.5555	197.8316	0.0145	28.8411	0.9284	0.4714	0.5135
0	ROCP	0.1901	0.8306	0.0724	0.4618	0.6462	118.9021	0.5407	170.0101	0.8262	0.6834	0.6648
0	PZ	0.1934	0.8273	0.0721	0.4630	0.6451	129.2970	0.4714	28.8894	0.9283	0.6998	0.6725
0	GSD	0.1426	0.8794	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	0.0144	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



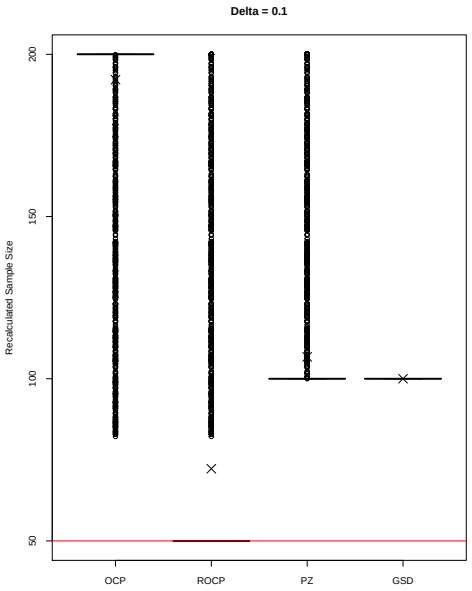
(a) Plot 1



(b) Plot 2

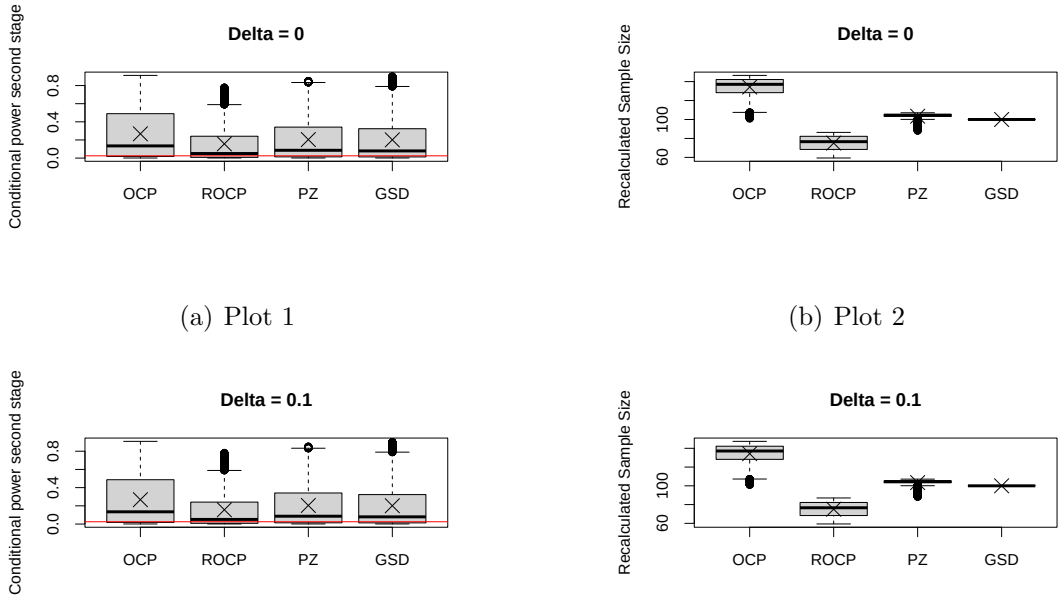


(c) Plot 3



(d) Plot 4

Figure 3.18: Standard simulation where only treatment is generated with MEM and control from a normal distribution



(a) Plot 1

(b) Plot 2

(c) Plot 3

(d) Plot 4

Figure 3.19: R2 approach where only treatment is generated with MEM and control from a normal distribution

3.8 Illustrative Clinical trial example

In the context of a clinical trial conducted by Bowden and Mander (2014), 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 a 2-week period compared to the baseline. Pain relief levels are measured using the McGill pain scale Melzack and Torgerson (1971), which ranges from 0 (indicating no pain) to 50 (indicating the highest pain level). In this trial, we assume that data is generated through MDM having both distributions normal and mixing proportion is 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 (2021). Initially, a pilot study indicated the superiority of the new treatment. However, further evidence is needed to quantify its actual effect, considering both the MEM and the MDM. As a result, the following hypotheses are formulated:

$$H_0 : \mu^T_{\text{baseline}} - \mu^P_{\text{baseline}} \leq 0 \quad \text{vs} \quad H_1 : \mu^T_{\text{baseline}} - \mu^P_{\text{2 weeks}} > 0. \tag{3.14}$$

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 potential measurement inaccuracies and then utilized the results from these tables (3.2, 3.3 and 3.4). Also, utilized results from these tables (3.10, 3.11 and 3.12). 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, with 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 performance at $\Delta = 0.2$, neighboring effect sizes ($\Delta = 0.1$ and 0.3) are also considered, taking into account the mixed distribution performance metrics are presented in Tables (3.1), 3.2, 3.3 and 3.4). Moreover, for measurement errors performance metrics are presented in Table (3.9), along with Tables (3.10, 3.11 and 3.12) 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 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, 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. Upon 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 conditional sample size and power due to the resampling approach, which is particularly relevant when dealing with complex data distributions (as illustrated in Tables (3.2) and (3.3)). 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, showcasing 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, courtesy of 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.

Chapter 4

Discussion and Conclusion

Integrating resampling into established sample size recalculation rules enhances the robustness of recalculation approaches, leading to significantly improved performance across various individual characteristics and a conditional performance score. This improvement is primarily due to the decreased variance in conditional sample size and conditional power. It is important to note that reference values and the weighting scheme for the conditional performance score could also be adjusted. Additionally, it is observed that 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 these 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 original sample size recalculation rules concerning the conditional performance score, it does not imply that 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 balance 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.

Particularly, 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.

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.

4.1 Recommendations and future work

Recommendations:

1. **Applicability to Different Endpoints:** Consider that the resampling techniques outlined in formulas (3.12) and (3.13) can be extended to studies involving various types of endpoints, as long as the test statistics exhibit approximate normal distribution characteristics.

2. **Binary Endpoints:** For studies with binary endpoints, you can readily apply the resampling approach by utilizing the normal approximation to binomial.

3. **Time-to-Event Endpoints:** When dealing with time-to-event endpoints, explore the possibility of employing the resampling approach through the utilization of the logrank test within an adaptive design framework, as this test is also suitable in such scenarios.

Future work:

1. **Direct Performance Optimization:** 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.

2. **Numerical Constrained Optimization:** Explore the idea of implementing the aforementioned 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.

Appendix A

Conditional performance score formulas

In the ensuing discussion, we will delve into the formulas governing the conditional performance score, along with its constituent sub-scores. For a more comprehensive understanding of the rationale behind this scoring system, readers are encouraged to consult Herrmann et al. (2020). Let's commence by elucidating the four key components, denoted as e_n, v_n, e_cp, and v_cp, as well as the two sub-scores, SCN and SCP, which collectively constitute the overarching total score, CS.

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

$$N_{\text{target}} = \begin{cases} n^{fix}, & \text{if } n^{fix} \leq n_{max} \text{ } and \text{ } \Delta \neq 0, \\ n_1, & \text{if } n^{fix} \geq n_{max} \text{ } or \text{ } \Delta = 0, \end{cases} \tag{A.1}$$

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

$$CP_{\text{target}} = \begin{cases} 1-\beta, & \text{if } n^{fix} \leq n_{max} \text{ } and \text{ } \Delta \neq 0, \\ \alpha, & \text{if } n^{fix} \geq n_{max} \text{ } or \text{ } \Delta = 0, \end{cases} \tag{A.2}$$

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

$$SCP(\Delta) = \gamma_{\text{loc}} \left(1 - \frac{|E[CN_{\Delta}^{RA}(T_1)] - N_{\text{target}}|}{n_{\max} - n_1}\right) + \gamma_{\text{var}} \left(1 - \sqrt{\frac{\text{Var}\left(N_{\Delta}^{RA}(T_1)\right)}{\text{Var}_{\max}\left(N_{\Delta}^{RA}(T_1)\right)}}\right) \tag{A.3}$$

where as $\left(1 - \frac{|E[CN_{\Delta}^{RA}(T_1)] - N_{\text{target}}|}{n_{\max} - n_1}\right)$ = 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 $\gamma_{loc} + \gamma_{var}$ =1. Both sub-scores, SCN and SCP, have a range of [0, 1]. They 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, CS, as follows:

$$CS(\Delta) = \frac{1}{2} \cdot [SCP(\Delta) + SCN(\Delta)] \tag{A.4}$$

References

Boos, D. D. and Brownie, C. (1992). A rank-based mixed model approach to multisite clinical trials. *Biometrics*, 48:61–72.

Bowden, J. and Mander, A. (2014). A review and re-interpretation of a group-sequential approach to sample size re-estimation in two-stage trials. *Pharmaceutical Statistics*, 13(3):163–172.

Brakenhoff, T. B., Mitroiu, M., Keogh, R. H., Moons, K. G., Groenwold, R. H., and van Smeden, M. (2018). Measurement error is often neglected in medical literature: a systematic review. *Journal of Clinical Epidemiology*, 98:89–97.

Chakraborty, H. and Gu, H. (2009). A mixed model approach for intent-to-treat analysis in longitudinal clinical trials with missing values. *Psychiatry Res*, 129.

Charles, P., Giraudeau, B., Dechartres, A., Baron, G., and Ravaud, P. (2009). Reporting of sample size calculation in randomised controlled trials. *BMJ*, 338.

Chow, S.-C. and Chang, M. (2008). Adaptive design methods in clinical trials–a review. *Orphanet Journal of Rare Diseases*, 3(1):1–13.

Das, S., Mitra, K., and Mandal, M. (2016). Sample size calculation: Basic principles. *Indian Journal of Anaesthesia*, 60(9):652.

Deng, Y., Tu, D., O'Callaghan, C. J., Liu, G., and Xu, W. (2022). Two-stage multivariate mendelian randomization on multiple outcomes with mixed distributions. *Statistical Methods in Medical Research*, page 09622802231181220.

Denne, J. S. (2001). Sample size recalculation using conditional power. *Statistics in Medicine*, 20(17-18):2645–2660.

Friede, T. and Kieser, M. (2001). A comparison of methods for adaptive sample size adjustment. *Statistics in Medicine*, 20(24):3861–3873.

Harden, M. and Friede, T. (2018). Sample size calculation in multi-centre clinical trials. *BMC Medical Research Methodology*, 18(1):1–10.

Herrmann, C., Kluge, C., Pilz, M., Kieser, M., and Rauch, G. (2021). Improving sample size recalculation in adaptive clinical trials by resampling. *Pharmaceutical Statistics*, 20(6):1035–1050.

Herrmann, C., Pilz, M., Kieser, M., and Rauch, G. (2020). A new conditional performance score for the evaluation of adaptive group sequential designs with sample size recalculation. *Statistics in Medicine*, 39(15):2067–2100.

Herrmann, C. and Rauch, G. (2021). Smoothing corrections for improving sample size recalculation rules in adaptive group sequential study designs. *Methods of Information in Medicine*, 60(01/02):001–008.

Jennison, C. and Turnbull, B. W. (2003). Mid-course sample size modification in clinical trials based on the observed treatment effect. *Statistics in Medicine*, 22(6):971–993.

Kieser, M. and Friede, T. (2003). Simple procedures for blinded sample size adjustment that do not affect the type i error rate. *Statistics in Medicine*, 22(23):3571–3581.

Liang, H., Wu, H., and Carroll, R. J. (2003). The relationship between virologic and immunologic responses in aids clinical research using mixed-effects varying-coefficient models with measurement error. *Biostatistics*, 4(2):297–312.

Mehta, C. R. and Pocock, S. J. (2011). Adaptive increase in sample size when interim results are promising: a practical guide with examples. *Statistics in medicine*, 30(28):3267–3284.

Melzack, R. and Torgerson, W. S. (1971). On the language of pain. *The Journal of the American Society of Anesthesiologists*, 34(1):50–59.

Morgan, T. M. and Elashoff, R. M. (1987). Effect of covariate measurement error in randomized clinical trials. *Statistics in Medicine*, 6(1):31–41.

Nagin, D. S. and Odgers, C. L. (2010). Group-based trajectory modeling in clinical research. *Annual Review of Clinical Psychology*, 6:109–138.

Pritchett, Y. L., Menon, S., Marchenko, O., Antonijevic, Z., Miller, E., Sanchez-Kam, M., Morgan-Bouniol, C. C., Nguyen, H., and Prucka, W. R. (2015). Sample size re-estimation designs in confirmatory clinical trials—current state, statistical considerations, and practical guidance. *Statistics in Biopharmaceutical Research*, 7(4):309–321.

Spanbauer, C. and Sparapani, R. (2021). Nonparametric machine learning for precision medicine with longitudinal clinical trials and Bayesian additive regression trees with mixed models. *Statistics in Medicine*, 40(11):2665–2691.

Wang, N., Lin, X., Gutierrez, R. G., and Carroll, R. J. (1998). Bias analysis and simex approach in generalized linear mixed measurement error models. *Journal of the American Statistical Association*, 93(441):249–261.

Yang, Y., Li, G., and Tong, T. (2015). Corrected empirical likelihood for a class of generalized linear measurement error models. *Science China Mathematics*, 58:1523–1536.