Randomisation methods for adaptive and multi-arm clinical trials

Ruqayya Assad Azher

Thesis submitted for the degree of
Doctor of Philosophy

Faculty of Medical Sciences
Newcastle University
Newcastle upon Tyne
United Kingdom

May 2024

I declare that the work in this thesis has been done by myself and has not been submitted elsewhere for the award of any other degree.

Chapter 2 has been published as Azher RA, Wason JMS, Grayling MJ. A comparison of randomization methods for multi-arm clinical trials. *Stat Biopharm Res* 2023;doi:10.1080/19466315.2023.2238645.

Ruqayya Azher

Acknowledgements

First and foremost I am eternally grateful to God for the strength to reach this stage of my academic pursuits.

Secondly, I would like to express my deepest gratitude to my supervisors, Prof. James Wason and Dr. Michael Grayling, for their unwavering support and mentorship throughout my PhD journey. Their patience, especially as I am an international student, has been invaluable. I am incredibly grateful for their generosity with their time, providing countless hours of feedback and guidance that significantly improved the quality of my work and skills. I truly cannot envision a better PhD experience without Prof. Wason and Dr. Grayling's expertise.

I also would like to thank everyone at the Biostatistic Research Group (BRG) for creating such an enjoyable, stimulating environment for conducting this research.

Finally, I would like to give special thanks to my parents and all of the rest of my family for always encouraging and supporting me. Special thanks go to Hamzah, my son. He is more than just my son; he is my friend and partner on this journey. Completing this PhD often meant sacrificing precious time together, but his unwavering support, despite the impact on our time, made all the difference.

Abstract

Randomised control trials (RCTs) typically compare one experimental treatment to a control. However, over recent years, with the growing availability of many treatments for evaluation and the increasing complexity of determining which are promising, a growing need has emerged for more efficient trial designs. Accordingly, adaptive designs have gained popularity in clinical research, including multi-arm multi-stage designs and platform trial designs. Such designs aim to improve the clinical trial process by allowing the removal and/or addition of treatment arms during the course of the trial.

In almost all clinical trial designs, randomisation remains a fundamental principle to ensure unbiased treatment comparisons. This is no less true of adaptive designs, yet randomisation routines have received less attention for such studies compared to historical fixed sample designs. Therefore, throughout this thesis, the key considerations discussed include the importance of proper randomisation approaches and allocation ratios to achieve clinical trial objectives in adaptive trials.

First, I compare different randomisation approaches in the context of multi-arm trials to achieve various trial objectives, such as group size balance, covariate balance, effect precision, low allocation predictability, and high power. Next, two adaptive clinical trial designs are considered: multi-arm multi-stage designs build on multi-arm designs by incorporating interim analyses with stopping rules. In particular, if an experimental treatment shows poor performance, it can be dropped early. In this design, allocation ratios can be fixed throughout the trial, flexible (adjusted based on observed interim data), or pre-specified but variable across the study stages.

Finally, attention shifts to randomisation methods in a platform trial design, to which new treatments can be added or dropped over time, and allocation ratios might need to be modified to achieve the study objective(s). Here, the focus is on randomisation

approaches with different allocation ratios, aiming to achieve high covariate balance, especially for a newly added arm in the trial.

Contents

Acronyms

AD Adaptive Design

ARP BCM Allocation Ratio Preserving Biased Coin Minimisation

BCD Biased Coin Design

BCD-IT Biased Coin Design with Imbalance Tolerance

BCM Biased Coin Minimisation

BSD Big Stick Design

BUD Block Urn Design

ESS Expected Sample Size

FWER Family-Wise Error Rate

MAMS Multi-Arm Multi-Stage

Mini Minimisation

MP Maximal Procedure

MTI Maximal Tolerated Imbalance

PBR Permuted Block Randomisation

RAR Response Adaptive Randomisation

RCT Randomised Control Trial

SBR Stratified Block Randomisation

SBUD Stratified Block Urn Design

SR Simple Randomisation

UD Urn Design

List of Figures

List of Tables

Chapter 1

Background and motivation

1.1 Introduction

Before any new medical treatment becomes available to the public, clinical trials must be conducted to ensure that the treatment is safe and effective (Pocock, 2013). Clinical trials are research studies conducted in which human participants are involved to evaluate the efficacy and safety profile of a new treatment (or, more broadly, an intervention, procedure, or device) relative to the currently available treatments. Typically, this means ascertaining that the new treatment is superior to existing treatments or placebo (Matthews, 2006).

Such trials are usually divided into several phases. Phase I trials often involve a small number of healthy volunteers. However, in some circumstances, such as when testing treatments for a fatal disease, those involved may consist of patients who have previously exhausted all available treatment alternatives. Phase I trials focus on studying the pharmacokinetics (i.e., the movement of the treatment through the body), pharmacodynamics (i.e., the treatment's effect on the body), and toxicity of a treatment, with the primary objective being to establish a tolerable dose range. Phase II trials are then conducted on patients with the disease of interest. These are initial efficacy studies

conducted to determine the dose and frequency of dosing required to successfully treat patients (Peace and Chen, 2010). If a treatment is indicated as effective during phase II, then it proceeds to phase III. These are large-scale, costly confirmatory trials in which the experimental treatment is compared to a control (standard treatment or placebo), with the primary goal of establishing the treatment's effectiveness. After approval of the treatment, its long-term effects on the patient population are monitored during phase IV trials following release to the market.

A well-designed clinical trial must include: *a control group*, *blinded*, and *randomisation* (Byar et al., 1976; Matthews, 2006). These are essential components to design any successful clinical trial. In particular, designs that incorporate randomisation and a control group are referred to as a *randomised control trial* (RCT). More explicitly, such trial designs have two key features: i) the new treatment is administered to a group of patients (treatment arm), and another treatment, often the current standard-of-care or placebo, is administered to another group of patients at the same time (control arm). This is what makes the trial controlled; and ii) patients are allocated to the treatment and control groups with some element of chance (randomisation).

The idea of randomisation was first introduced by (Fisher, 1926), who randomly assigned treatments to blocks or plots of land in an agricultural study. Randomisation was then introduced into clinical studies in the early 1930s by Amberson Jr et al. (1931). However, the randomisation principle was not implemented in clinical trials until 1948 by the Medical Research Council. This landmark study, led by Sir Austin Bradford Hill, is recognised for establishing the use of random numbers to allocate patients to treatment and control groups; these random numbers helped eliminate potential biases in patient selection and achieve effective balancing of groups for comparison (Rosenberger et al., 2018).

Randomisation is the process of assigning patients to different treatment groups using random elements, ensuring that each patient has an equal chance of being allocated

to any treatment group. The use of randomisation in clinical trials has substantial advantages. First, randomisation (at least theoretically) ensures the treatment arms are more comparable in terms of patients baseline characteristics (covariates). In non-randomised observational studies, attempting to achieve such comparability typically involves adjusting for or matching for known covariates. However, this does not ensure control for other unknown covariates. Randomisation, on the other hand, also allows comparability for unknown covariates (Chow and Liu, 2008; Rosenberger and Lachin, 2015). This means that any observed difference between the treatment arms at the end of the trial is due to the treatment itself rather than any other factors (covariates).

Second, randomisation can also eliminate selection bias, which refers to the bias that can be introduced by selecting particular patients to receive specific treatments (Rosenberger and Lachin, 2015). In a RCT, this can occur if the clinician or investigator knows or can predict what treatment will be assigned next; the result can be biassed by selecting a patient expected to be more likely to benefit from that treatment. Selection bias has been discussed in the literature since the 1950s (e.g., Blackwell and Hodges Jr (1957)), modelled the impact of a clinician's attempt to guess the next allocation and then select a particular patient for a particular treatment (Rosenberger and Lachin, 2015).

Berger (2005) distinguishes between allocation concealment and masking. The author describes allocation concealment as the negation of the ability to observe upcoming allocations, so that investigators are unaware of the treatment assignment until after it has been assigned. Masking is more complex because it necessitates keeping the treatment assignment unrevealed until the end of the trial. It is not always possible to mask investigators, but it is always possible to conceal allocation. Allocation concealment plays a crucial role in preventing selection bias, as noted by Rosenberger and Lachin (2015). However, selection bias can occur even when allocation concealment is present, as investigators can still attempt to predict the next treatment assignment. Certain

randomisation procedures, as we shall see through the thesis, reduce the predictability of treatment assignments and mitigate selection bias further.

Based on the discussion above, a well-chosen randomisation routine in clinical trials should incorporate specific properties to ensure the success of the trial. The desirable properties of such a routine include:

- Balance. A well-designed randomisation process should ensure that the groups being compared are balanced in terms of known and unknown covariates. This minimises the risk of confounding variables affecting the results, as it ensures that these factors are equally distributed across all treatment groups, making the comparison of outcomes between groups fair and unbiased (See, e.g., Heritier et al. (2005) and Bruce et al. (2022)).

- Randomness. The allocation of patients to different treatment groups must be unpredictable to both the patients and the investigators. As I mentioned above, this helps prevent potential biases by ensuring that each participant has an equal and independent chance of being assigned to any group (Rosenberger and Lachin, 2015).

- Simple Implementation. The randomisation process should be easy to implement in practice. If the method is too complex, there may be errors in execution, which could lead to unintentional biases. A simple and clear procedure helps ensure that the randomisation is carried out consistently and correctly throughout the trial.

All the above randomisation properties are applicable for almost RCTs. However, it should be noted that several methods of randomisation are possible and that, some of these are more effective than others in achieving comparable groups. This issue will be explored in more details throughout this thesis.

Therefore, the scope of this thesis is to explore and evaluate several randomisation methods and allocation ratios within the context of multi-arm and adaptive clinical trial

designs. I begin by describing multi-arm and adaptive clinical trial designs in Sections 1.2 through 1.4. Following this, Sections 1.5 and 1.5.4 provide a comprehensive overview of randomisation methods and allocation ratios that have been discussed in the literature.

1.2 Two-arm fixed sample trials

The prototypical randomised control trial (RCT) is a two-arm parallel group trial that compares one experimental treatment to control treatment or placebo. This design is simple and is often used to evaluate distinct treatments. In this trial design, a pre-determined number of patients are randomly assigned to one of two treatment groups. The trial is then conducted, and the outcomes of the two treatment groups are compared to assess the effectiveness of the new treatment. However, this may not be always efficient for complex clinical objectives.

1.3 Multi-arm trials

While RCTs have become a mainstay of trial implementation in practice, investigators have recently sought to improve upon this design by comparing several treatment arms simultaneously. These multi-arm randomised control trials have become increasingly popular due to their myriad of benefits compared to two-arm trials. These benefits include: i) the efficient comparison of several treatments in a single trial reduces the required sample size, whilst also saving time and money, compared to conducting separate trials for each treatment (Juszczak et al., 2019); ii) evaluating different doses of the same treatment within a single trial design allows investigation of the optimal dose and minimises the need for additional trials (Mossop et al., 2022); and iii) assessing the efficacy of combining multiple experimental treatments can help identify the most successful combinations (Jones and Kenward, 2003).

Despite all the benefits of multi-arm trial designs, they are typically large and expensive, and they can take a long time to complete. In March 2004 the United States Food and Drug Administration (FDA) released a landmark report expressing concern about the slowdown in innovative treatments that were being submitted for approval (U.S. Food and Drug Administration, 2004). The report indicated "an urgent need to improve the efficiency and effectiveness of the clinical trial process, including improved trial design". Consequently, in March 2006, the FDA released a document indicating that biomarker development and streamlining clinical trials are the two most important areas for improving medical product development (U.S. Food and Drug Administration, 2006). Streamlining clinical trials includes the advancement of innovative trial designs, such as *adaptive designs* (ADs).

1.4 Adaptive clinical trial designs

An AD is defined as a clinical trial design that allows modification(s) to be made to one or more aspects of the trial design based on accumulating data from patients in the trial (Jaki, 2015; Pallmann et al., 2018; Burnett et al., 2020). These modifications are performed on the basis of interim analyses, without compromising the validity and integrity of the ongoing trial. The interim analyses are conducted at pre-specified times throughout the trial, to allow investigators to decide whether or not to, e.g., make modifications to the sample size, the number of treatment arms, and/or the treatment allocation ratios (that is the ratio of patients assigned to the experimental arms relative to the control arm) (Jaki, 2015; Pallmann et al., 2018; Burnett et al., 2020). Despite the recent growth in attention received by ADs, they have been around for more than 50 years, and ADs have now been developed to improve efficiency in all phases of clinical research (Bauer et al., 2016).

An example of an AD is the *group sequential design* (GSD), which was developed to

evaluate the efficacy/futility of a single treatment arm through a series of pre-planned interim analyses, allowing clinical benefit to be assessed under economic and ethical constraints (Pocock and Simon, 1975; O'Brien and Fleming, 1979; Gordon Lan and DeMets, 1983; Whitehead and Brunier, 1990; Jennison and Turnbull, 1999; Wassmer and Brannath, 2016). If multiple treatments are of interest, a Multi-Arm Multi-Stage (MAMS) design may be more beneficial (Stallard and Todd, 2003; Magirr et al., 2012; Urach and Posch, 2016).

1.4.1 Multi-arm multi-stage designs

The MAMS design allows investigators to evaluate multiple experimental treatments simultaneously over multiple study stages. MAMS designs are increasingly being implemented in practice due to their myriad of advantages over traditional clinical studies:

1. Multiple research questions can be answered simultaneously in a single trial rather than sequentially or through a series of separate trials. This helps investigators gain more information about treatments and make faster decisions about future development.

2. Instead of using a separate control group for each treatment, a shared control group can be used, which can reduce required sample sizes and trial costs.

3. A head-to-head comparison of treatments can be carried out directly, reducing the potential for bias arising from comparisons between treatments evaluated in different trials.

4. All patients are randomised upon enrollment to one of the treatment arms at the beginning of the trial. This ensures balance in covariates across treatment arms and mitigates selection bias.

5. At the end of each stage, an interim analysis is conducted to determine which, if any, of the treatment arms performs better or worse than the control arm to justify their removal from further study in subsequent stages. These interim analyses allow investigators to monitor the safety and efficacy of the treatments throughout the trial. This helps identify potential risks or lack of effectiveness early, minimising the exposure of patients to harm or ineffective treatments (Chow and Chang, 2006).

Some MAMS trials allow early stopping for efficacy performing several interim analyses on the same trial, increasing the chance of falsely rejecting a null hypothesis (type I error rate inflation). Consequently, several methods are available to control this error, such as: i) alpha spending functions, as proposed by Demets and Lan (1994), to specify how much of the overall type I error rate (typically 5% two-sided) can be spent at each interim analysis; ii) functional-form stopping boundaries (e.g., Pocock and Simon (1975); O'Brien and Fleming (1979)), which adjust the significance level for each interim analysis to compensate for multiple testing.

Figure 1.1 illustrates an example of a MAMS trial that includes two experimental arms compared to a shared control arm. At the interim analysis, experimental arm 2 dropped out of the trial, while experimental arm 1 continued to the end of the trial.

1.4.2 Platform trial designs

In addition to the implementation of ADs, a notable advancement in clinical trial study designs to improve efficiency is the use of a "master protocol" (Park et al., 2023). A master protocol is a trial design that allows testing of several treatments, individually or in combination, as well as testing multiple diseases simultaneously, all under a single protocol. The use of master protocols in clinical trials is increasing, as it provides a more streamlined and cost-effective approach to drug development (Berry et al., 2015). Investigators can effectively investigate multiple, potentially very distinct, hypotheses

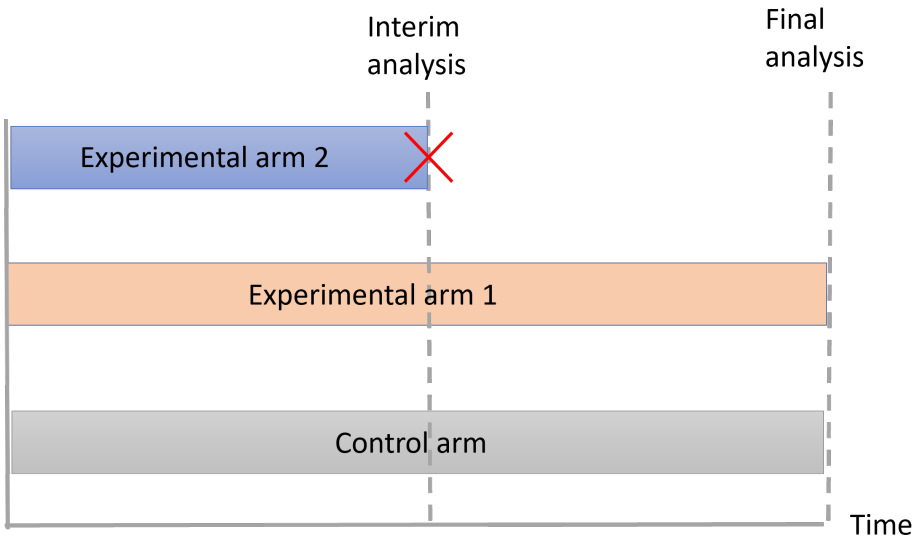

Figure 1.1: Illustration of a multi-arm multi-stage design.

within one study.

One of the master protocol designs is a MAMS *platform trial* which refer to ADs that allows new treatments to be introduced to the ongoing trial over time, even if they are not pre-specified in the planning stage. The platform trial designs have numerous advantages, encompassing all the benefits of the previously mentioned MAMS designs in Subsection 1.4.1, along with additional advantages such as:

1. Incorporating new treatment(s) for testing allows for an expedited evaluation and decision making regarding these treatment(s), saving time and resources.

2. Platform trials also provide the flexibility to update the control arm as the trial progresses. This is particularly useful when, for example, the platform trial demonstrates that a treatment is much more effective than a control treatment. In such cases, a successful treatment can be promptly introduced to benefit patients and become the new control treatment against which all subsequent treatments are tested in the trial, creating a potential cycle of improvement.

3. In platform trials, when new experimental arm(s) are added at a certain time point,

there will be a group of patients who were previously enrolled in the control arm. At analysis time, there are two different situations for comparing the newly added arm to the control arm: i) the newly added treatment arm is only compared to control arm patients recruited during the same time period as the enrolled patients in the newly added arm; ii) the newly added arm can be compared to all previously enrolled patients and current patients in the control arm. According to this, the control arm can be divided into a concurrent control and non-concurrent controls based on enrolment time relative to specific experimental arms. When non-concurrent controls can be leveraged, additional efficiency advantages may be possible (Berry et al., 2015).

Figure 1.2 illustrates an example of a platform trial that initially starts as two-arm, then later a new treatment arm is added. In this illustration, patients are initially randomised with fixed (often equal) allocation randomisation, then the new experimental arm is added to the trial, and patients are again randomised to all treatment arms using a fixed allocation ratio.

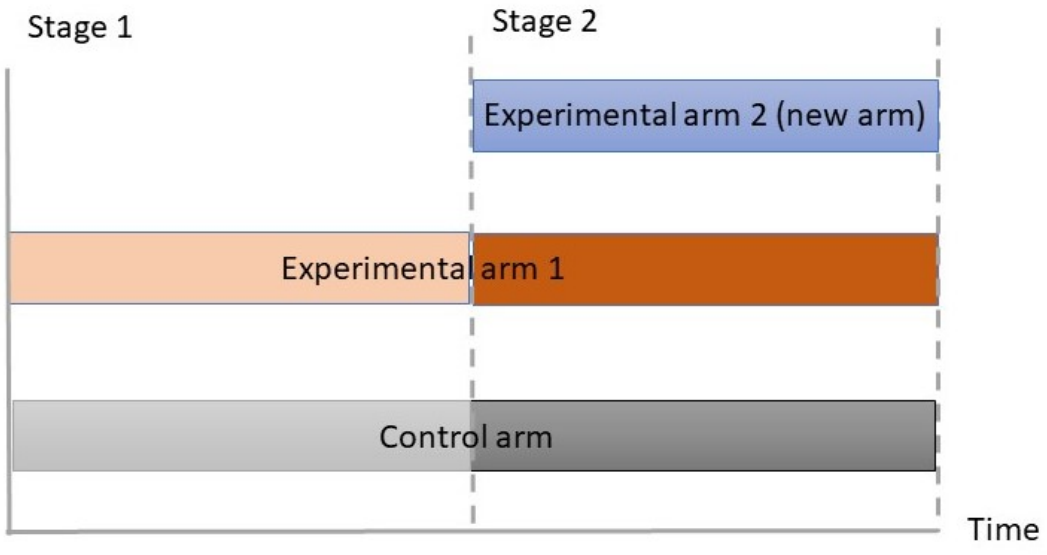

Figure 1.2: Illustration of a platform trial.

The most commonly cited example of a MAMS platform design is the STAMPEDE trial, which is a multi-centre randomised phase III trial that evaluated new treatments in men with hormone-naive prostate cancer (Parmar et al., 2017). This trial initially started with five experimental arms, then five additional experimental arms were added during the course of the trial. Another recent example of a platform trial is RECOVERY (Randomised Evaluation of COVID-19 Therapy). This is one of the largest international platform trials, aiming to identify effective treatments for patients hospitalised with COVID-19. The RECOVERY trial started with five treatment arms, and as it progressed, it added 10 additional treatment arms. The trial's objective is to determine whether any of the different treatments (as of June 2022) reduced mortality in COVID-19 patients. (Horby et al., 2020).

Despite the fact that clinical trial designs have undergone significant changes over the years, randomisation remains a crucial element in almost all controlled trial designs.

1.5 Randomisation in clinical trials

Randomisation approaches can be broadly categorised into two classes: *conventional (or fixed) randomisation* and *adaptive (or restricted) randomisation* (Chow and Liu, 2008). Conventional randomisation approaches use predetermined probabilities to independently allocate patients to various treatment arms. However, this method can accidentally result in imbalances between treatment arms.

Consequently, the concept of adaptive randomisation has been proposed to mitigate such imbalances. This can be achieved by dynamically adjusting the probability assignment to the treatment arms throughout the trial (Wei, 1978). The adaptive randomisation process takes place at each time of enrolling a new patient in the trial. In clinical trials, adaptive randomisation is often applied with respect to treatment, covariates, or clinical response (Chow and Liu, 2008). Therefore, adaptive randomisation

approaches can be referred to as *treatment adaptive randomisation*, *covariate adaptive randomisation*, or *response adaptive randomisation*.

The primary focus of this thesis will be on conventional randomisation, treatment adaptive randomisation, and covariate adaptive randomisation. For an overview of response adaptive randomisation, the reader is referred to Atkinson and Biswas (2013).

1.5.1 Conventional (or fixed) randomisation

Conventional randomisation, also known as fixed randomisation, involves treatment allocation probabilities that remain *constant* throughout the trial (Berger et al., 2021). In this method, investigators pre-specify allocation probabilities to each treatment arm before the trial begins. The patients are then allocated accordingly using a random number generator, which ensures that each patient has a fixed probability of being assigned to any treatment arm.

The simplest and most straightforward fixed randomisation method is *simple randomisation* (SR), also known as complete randomisation. SR can be thought of as randomly assigning patients to a particular treatment arm by the toss of a coin. The main feature of this method is its high level of randomness. However, SR can result in an imbalance between treatment arms; this may be particularly critical when there is an important covariate to balance (Matthews, 2006). This limitation was recognised earlier in agricultural studies by Fisher (1992). The author proposed dividing the experimental field into blocks (or groups) with similar conditions. Subsequently, different treatments were randomly assigned to these blocks, ensuring that each treatment received a fair representation within each block with similar conditions. This approach effectively minimised the impact of key factors that influence study outcomes and allowed a more accurate comparison of treatment effects (Fisher, 1992).

Since then, several authors (e.g., Cochran (1939)) have suggested the use of stratification within blocks in clinical trials to control for specific known covariates. This

is referred to as *stratified block randomisation* (SBR). The aim of SBR is to eliminate the potential bias that could occur when there are covariate imbalances between the treatment arms. Currently, SBR method is the most popular and widely used randomisation approach in practice due to its effective ability to balance on patients covariates (Zagoraiou, 2017; Berger et al., 2021; Bruce et al., 2022).

Although SBR method has been widely used in clinical trials, it is not without limitation. A particular limitation can occur when there are a large number of important covariates or a small number of covariates with many levels. In a clinical trial, a combination of these important covariates may be defined as a strata. In this case, these strata may not achieve balance because some of them may contain only a few patients, particularly in a small clinical trial. To account for this, several researchers, including Taves (1974) and Pocock and Simon (1975) proposed *covariate adaptive randomisation* to better balance a large number of strata; we discuss such approaches below. Another limitation of SBR is that it can introduce some level of allocation predictability compared to other randomisation methods, such as SR. For example, for a two-arm comparison with small block sizes (e.g., 2 or 4), the sequence of treatment assignments within a block can become predictable after the first few patients are enrolled. This predictability arises due to the limited number of possible combinations within the block. This predictability can be further increased if the allocation sequence within blocks is not randomised.

1.5.2 Treatment adaptive (or restricted) randomisation

Treatment adaptive randomisation, also known as restricted randomisation, aims to reduce the deviation from the target treatment allocation ratio. Treatment adaptive randomisation approaches effectively modify the assignment probabilities of new patients based on the current treatment imbalance (Chow and Chang, 2006; Chow and Liu, 2008). This helps to ensure a more balanced allocation of patients between treatment arms as

the trial progresses.

Examples of popular treatment adaptive randomisation methods include the *Biased Coin Design* (BCD), as introduced by Efron (1971). In this method, the randomisation process starts with an equal assignment probability for each treatment arm. Then, when a new patient is ready to be allocated, the assignment probability is biased in favour of the treatment group that currently has fewer patients. Specifically, Efron (1971) suggested using an assignment probability of 2/3 to the smaller arm. While easy to understand, the BCD may not be sufficiently effective in practice because this increased assignment probability is constant and fixed throughout the trial.

Consequently, several authors have modified the BCD to achieve better treatment balance. For example, Wei (1977, 1978) developed an adaptive BCD where the assignment probabilities update according to the degree of imbalance between treatment arms; this is known as an *urn design* UD. The UD has been extensively studied and developed in research due to its flexibility. Although, the UD design is advantageous in that it can be readily applied to clinical trials with two or more treatment arms, it is important to consider its limitations and the potential for treatment imbalance, especially in larger trials (Rosenberger and Lachin, 2015).

Furthermore, Soares and Wu (1983) extended the BCD in an attempt to reduce treatment imbalances, referring to their design as the *Big Stick Design* (BSD). In this method, a predetermined threshold specifies the maximum acceptable difference in the number of patients allocated to each treatment arm, this is known as *"maximum tolerated imbalance"* (MTI). The MTI intervenes when necessary to prevent treatment arms from deviating too far from each other. For illustration, in the BSD, patients are randomly assigned to treatment arms until a pre-specified MTI is reached. The following patient will then be allocated deterministically to the treatment group with fewer patients. This clearly effectively caps the imbalance between treatment arms. However, treatment allocation is completely predictable for certain participants, given knowledge of past

assignments.

In addition to the BSD, there are other proposed methods within the MTI family, such as Chen's procedure (Chen, 1999), the maximal procedure (MP) (Berger et al., 2003), and the block urn design (BUD) (Zhao and Weng, 2011). Chen's design combines elements of the BCD and BSD, referred to as *Biased Coin Design with Imbalance Tolerance* (BCD-IT). This method was proposed to achieve better treatment balance and reduce the assignment predictability that occurs with the BSD. In the BCD-IT method, patients are randomly assigned to treatment arms until a pre-specified MTI is reached. Then, a new enrolled patient will be randomised to treatment arms with a pre-defined biased assignment probability. Although the BCD-IT method can achieve a level of treatment balance while reducing the predictability of assignment, it is only suitable for a two-arm trial due to the difficulty in calculating the assignment probability based on both previous assignments and MTI thresholds (Zhao et al., 2018).

Furthermore, the MP design and BUD were proposed for two or more treatment arms. The MP design proposed to maximise the number of feasible allocation sequences under both the MTI threshold and the length of the allocation sequence (Berger et al., 2003). However, its implementation is difficult as there is no simple algorithm for calculating the conditional allocation probability based on previous assignments (Zhao et al., 2018). The BUD was proposed by combining aspects of both the UD and block randomisation, to achieve both of their objectives (Zhao and Weng, 2011). However, this approach may not be suitable for small clinical trials as it requires a reasonable sample to achieve balance.

In fact, adaptive treatment randomisation procedures have received considerable interest in recent decades as an effective approach to improve the balance between treatment arms (Chow and Liu, 2008). Therefore, in research studies, various methods have been developed in different formats, especially as extensions of the BCD and UD (e.g., Chen (2000); Antognini and Giovagnoli (2004); Bailey and Nelson (2003); Kuznetsova

and Tymofyeyev (2015)). However, not all are (often, if ever) considered for implementation in practice, for several reasons:

- Complexity. Some of the developed methods can enhance the level of complexity in implementation and interpretation (Zhao et al., 2018). In practice, simplicity and ease of implementation are often prioritised. This is particularly true for more complex clinical trial designs. More specifically, in adaptive randomised control trials, investigators look for a simpler and more effective implementation of randomisation approaches (Chow and Chang, 2006), since the trial design is already more complex compared to traditional two-arm trials.

- Some randomisation methods may not be applicable for multi-arm trials or when there is a need to account for covariates. For example, the BSD and BCD-IT were proposed and recommended only for two-arm trials (Berger et al., 2003; Zhao et al., 2018).

- Several randomisation methods are designed specifically for equal allocation ratios, such as the BSD. Therefore, they are not applicable for trials that wish to use unequal allocation ratios (Soares and Wu, 1983).

- Not all randomisation methods are applicable in all clinical trial settings. More specifically, in emergency settings or trials with significant time pressures, adaptive randomisation methods might be impractical. This is because these methods often require real-time data analysis and decision-making, which can be challenging under time constraints. In such cases, SR or SBR might be more practical options. These methods are easier to administer quickly and without the need for extensive computational support, ensuring that the trial can proceed without delays.

1.5.3 Covariate adaptive randomisation

In covariate adaptive randomisation, the assignment probability is adjusted based on previous assignments within strata. This concept was proposed in the early 1970s by several authors (Taves, 1974; Pocock and Simon, 1975), to balance a large number of covariates. Taves (1974) termed it *minimisation*, proposing it as an appropriate approach for a multi-arm clinical trial to minimise overall treatment imbalances for a large number of strata. His proposal was entirely deterministic. This means that a new enrolled patient is allocated with certainty to the treatment that minimises the imbalance. Subsequently, Pocock and Simon (1975) modified the minimisation approach by incorporating allocation probabilities (i.e., a random element) to reduce the deterministic assignment. Specifically, instead of strictly allocating the new patient to the treatment arm that minimises the imbalance, the patient will be allocated to that arm with a prespecified fixed probability. This allows the option that would minimise the imbalance to be weighted more heavily, while retaining a chance that the patient could be allocated to other treatment arms, to improve randomness (Taves, 1974; Pocock and Simon, 1975; Tymofyeyev et al., 2007; Lauzon et al., 2022).

Since then, minimisation with a random element has become widely used in clinical trials to achieve marginal balance between each covariate separately. It has been widely argued to be the best method to achieve covariate balance in practice (Kuznetsova and Tymofyeyev, 2012). However, this is only when equal allocation is desired for two or more treatment arms, expanding it to unequal allocation proved challenging (Han et al., 2009; Kuznetsova and Tymofyeyev, 2012). Han et al. (2009) showed that the minimisation process needs to be modified for trials involving unequal treatment allocations. The authors introduced two modifications: the first was referred to as *naive minimisation*. In this method, the unequal allocation ratio is accounted for in calculation of the required imbalance score. This can be achieved by dividing the treatment counts within each fac-

tor level by the treatment allocation ratios. Once the preferred treatment is determined, the assignment probability would be computed as for the minimisation process for equal treatment allocations. Han et al. (2009) showed, however, that this simple modification can lead to deviations from the target allocation ratio, especially when the allocation probability given to the preferred treatment is relatively low. To avoid this problem, the authors proposed a second modification, called *biased coin minimisation* (BCM). In BCM, the allocation probabilities assigned to each treatment are varied depending on their allocation ratio, when they are the preferred treatment.

Kuznetsova and Tymofyeyev (2012) showed that although BCM leads to an allocation ratio at the end of the trial that is close to the targeted one, it does not preserve the allocation ratio at every allocation step. The authors argued that if the allocation ratio is not preserved at every step, the probability of allocation to a particular treatment tends to fluctuate in periodic cycles, providing an opportunity for selection bias throughout enrolment. The authors then proposed a method to achieve unequal allocations while preserving the allocation ratio at every allocation, using a blocking approach. They showed that as long as the block size remained small, this method showed good balancing properties.

1.5.4 Allocation ratios

The choice of a randomisation method often also depends on the pre-determined allocation ratio that are needed to achieve the specific clinical trial objective. The allocation ratio is defined, as previously mentioned, as the ratio of patients assigned to the experimental arms relative to the control arm. At the planning stage of a RCT, it is paramount to pre-specify the allocation ratio for several reasons, such as: i) by pre-specifying the allocation ratio, investigators avoid the temptation to manipulate enrolment based on their own expectations or preliminary results. This helps maintain the trial's objectivity and the validity of its findings; ii) a carefully chosen allocation ratio can optimise the

trial power and/or minimise the expected sample size. The allocation ratio can indeed have a considerable impact on the clinical trial power. In two-arm trials, balanced (i.e., equal) allocation is often the preferred approach, partly due to the fact that it maximises power under quite general assumptions (Neuhäuser et al., 2021).

However, in multi-arm trials, balanced allocation is likely to be optimal only if the aim of the trial is to perform a pairwise comparison between all treatment arms. In particular, it may not be preferred when comparing a shared control arm with experimental arms (Neuhäuser et al., 2021). Most famously, Dunnett (1955) showed that, under certain conditions, the optimal allocation ratio to maximise the statistical power can be achieved by allocating the square root of the number of experimental arms to the control arm for every patient in an experimental arm. This approach enhances the precision of the control mean, which in turn improves the ability to detect significant differences between the treatments and the control. Furthermore, its appropriateness may be impacted when: i) there is heterogeneity in the variance of outcomes between treatment arms; ii) an investigator wants to change the allocation ratio to assign more patients to superior treatment(s); iii) there is unequal interest in certain treatment comparisons; and iv) there is a differential treatment cost with a limited budget.

One can also justify imbalanced allocation ratios for ethical reasons or recruitment purposes (Dumville et al., 2006). For example, in clinical trials where the experimental treatment is considered highly promising or has significant public health implications, such as a new AIDS treatment, an unequal allocation ratio might be used, such as 2:1 in favour of the experimental treatment over the control (Rosenberger and Lachin, 2015). This approach prioritises the experimental treatment and can enhance recruitment by allowing more patients to receive the potentially beneficial treatment. Furthermore, it is important to notice that the experimental treatment may be harmful. Therefore, investigators can reduce the number of patients allocated to the potentially harmful treatment. This approach allows for the assessment of the experimental treatment's

effects while reducing the risk of patients receiving harmful treatment.

1.6 Aims and objectives

The aim of this thesis is to determine efficient randomisation routines and allocation ratios that can be used in multi-arm trials, including adaptive MAMS and platform designs. The specific objectives are:

- To compare different randomisation methods in the context of multi-arm trials in the presence and/or absence of covariates.
- To explore and identify efficient stage-wise allocation ratios in the context of MAMS trials, in terms of maximising the marginal, disjunctive, or conjunctive power.
- To evaluate different methods of randomising patients in platform trial designs, focusing on covariate imbalance and assignment randomness as new arm(s) are added.

1.7 Outline

Chapter 2 provides a comprehensive simulation study, motivated by two real clinical trials, to empirically evaluate several different randomisation procedures in the context of multi-arm trials. Several randomisation procedures are considered, specifically simple randomisation, permuted block randomisation, stratified randomisation with permuted blocks, urn design, block urn design, stratified block urn design, and minimisation. Numerous properties of the randomisation procedures are evaluated, including group size balance, covariate balance, loss of precision (i.e., increase in the variance of treatment effect estimates), and predictability of assignment. Different definitions of statistical

power relevant in multi-arm trials, marginal, disjunctive, and conjunctive, are also compared.

Following this, Chapter 3 explores a range of practical combinations of stage-wise allocation ratios in Multi-Arm Multi-Stage designs to determine efficient stage-wise allocation ratios that can increase power. In this chapter, I begin by describing how to compute the operating characteristics of MAMS designs when allowing the stage-wise allocation ratios to vary. A real clinical trial is again used as motivation for the parameters assumed in the evaluation.

Chapter 4 provides a simulation study to assess different randomisation procedures and allocation ratios in the type of adaptive clinical trial design that recently gained expanded popularity during the COVID-19 pandemic: platform trials. The following randomisation approaches are considered for different allocation ratios: simple randomisation, stratified randomisation with permuted blocks, stratified block urn design, and minimisation. Covariate balance and the assignment predictability are once more computed and contrasted across the randomisation procedures.

To conclude, a general discussion about randomisation methods, allocation ratios, and directions for future research is presented in Chapter 5.

Chapter 2

A comparison of randomisation methods for multi-arm clinical trials

In this chapter, I compare and evaluate several commonly utilised randomisation methods, alongside other theoretically advantageous but less often employed methods, in order to provide general guidance on which randomisation methods should be considered for use in practice in a multi-arm trial. This work has been published in *Statistics in Biopharmaceutical Research* Azher et al. (2023).

2.1 Background

Multi-arm trials are a type of clinical trial in which at least two experimental treatment arms are evaluated and compared. These treatments can be compared to a common control arm, to different controls for each treatment arm, or directly to each other. This is in contrast to traditional two-arm trials, which only compare one intervention against a control. Multi-arm trials can be used to evaluate different doses of the same treatment, entirely different treatments, or different combinations of treatments. Most often, all treatment arms recruit patients at the same time, allowing for direct comparison from the start (an exception to this is multi-arm platform trials, which will be the focus of Chapter 4). Furthermore, the number of patients allocated to each arm is typically predetermined before the trial begins (exceptions to this include multi-arm multi-stage trials, which will be the focus of Chapter 3). As introduced in Section 1.1, a fundamental component of multi-arm trial design is the process of allocating patients to treatment arms (randomisation). This helps to control for bias and confounding factors that could otherwise influence the study results.

There are numerous methods that could be used for randomisation in a multi-arm trial, as described in Section 1.5, each with its own advantages and disadvantages. The choice of an appropriate randomisation routine in a multi-arm trial is also crucial and depends on several factors, including the trial's characteristics (e.g., sample size and the number of covariate) and the study objectives; for example, some trials might aim for equal allocation across all treatment arms, while others might aim to allocate more patients to promising treatments based on emerging data (adaptive randomisation).

In this chapter, I therefore provide an assessment of several more popular methods from across the categories of randomisation procedures, including: Simple Randomisation (SR), Permuted Block Randomisation (PBR) (Matts and Lachin, 1988), Stratified Block Randomisation (SBR) (Kernan et al., 1999), Urn Design (UD) (Wei, 1977), Block

Urn Design (BUD) (Zhao and Weng, 2011), Stratified Block Urn Design (SBUD) (Zhao, 2014), and Minimisation (Taves, 1974) (See table 2.1).

The simplest of these methods is SR, which is a procedure that assigns patients with a pre-specified probability to each of the treatment arms, without considering previous assignments, covariate data, or other information. In a very large trial, SR provides a high likelihood that covariates are well-balanced across all treatment arms, which means that randomisation is the first step towards ensuring an unbiased comparison (i.e., differences between groups are due to the treatment or random variation). However, for a smaller trial, the likelihood of a chance imbalance in important covariates can increase markedly for SR, as it does not take into account previous assignments. Furthermore, SR imparts a large possibility that the number of patients assigned to each treatment arm may differ notably, which may impact power. For these reasons, a variety of randomisation routines have now been developed that seek to increase the likelihood of balance on specified covariates believed to be of principal importance and/or improve balance in the number of patients assigned to each treatment.

Amongst these, Wei (1977) proposed the UD to force a small trial to be better-balanced in terms of group size, but to provide behaviour like SR as the size of the trial increases (Matts and Lachin, 1988). This is one example of a treatment adaptive randomisation approach, in which the treatment assignment probability is modified based on previous assignments. In this approach, an (hypothetical) urn contains balls with different labels based on the number of treatments. To assign a patient to a treatment, a ball is randomly selected from the urn, its label is observed, and the patient allocated to the arm associated with this label. The ball is then returned to the urn, along with α more balls of the same label, and β more balls of each of the other labels. This drawing procedure is repeated for each new assignment. The values of α and β can be any integer numbers. The UD can provide good allocation randomness whilst potentially increasing power due to the improved balance in the number of patients on each treatment,

particularly in a small trial. However, it may cause time related biases (the absolute imbalance can increase as the sample size increases) (Zhao and Weng, 2011).

Another popular randomisation method is block randomisation, which is one of the fixed randomisation methods that utilises a fixed probability assignment throughout the trial. In block randomisation, patients are assigned to treatment arms through 'blocks' that contain a specified number of allocations to each treatment arm in some given order. The size of each block is typically an integer multiple of the number of treatment arms, ensuring balance in the number of patients assigned to each arm within each block. PBR is a commonly used version of block randomisation, in which each block has its own specified randomly ordered treatment assignments. Then, a patient enrolling into a trial is allocated using the next assignment in the current block. In PBR, the block sizes may also be varied and randomly chosen. This helps to overcome a potential problem with block randomisation, at least for small block sizes, in which the next treatment assignment could become predictable.

Furthermore, PBR is a highly popular allocation method in clinical trials because it is easy to implement, and it has consistent imbalance control. However, PBR imparts deterministic assignments. In contrast, the UD has high randomness, but does not guarantee balance across treatment arms. To achieve the conflicting requirements of the balance of treatment allocation and the randomness of treatment assignments, the BUD was proposed by Zhao and Weng (2011) as a sub-type of MTI procedure (see Section 1.5.2 for more details on MTI procedures). The BUD is relatively simple and in theory combines the advantages of the PBR and UD procedures that I also consider. The BUD can be understood by considering two urns, termed active and inactive. Assignments are made based on random draws from the active urn, while the algorithmic returning of balls from the inactive urn to the active urn enables myopic control of assignment imbalances (see Subsection 2.2.1 for more details on implementing BUD).

Although the UD and PBR can help ensure balance in the number of patients as-

signed to each treatment arm, they do nothing to help balance between arms with respect to key covariates. A possible method to overcome this issue that I consider is SBR, in which patients are divided into strata according to important covariates. Then, PBR is used to generate the allocations within each stratum. Unfortunately, when the number of strata is large, SBR can perform poorly in ensuring balance between treatment arms (Matthews, 2006). In a similar method to the way SBR extends PBR, I also consider implementing the BUD within strata, to better balance covariates. I term this the SBUD.

Finally, I also consider a method from one of the more widely-used adaptive randomisation classes, covariate adaptive randomisation (Hu et al., 2014), which can provide balance over a large number of covariates. Specifically, I evaluate the use of minimisation. In minimisation, important covariates are identified before a trial starts and assignment of a new patient to a treatment arm is determined to minimise the differences between the arms in terms of these covariates. A number of metrics for evaluating such differences between arms have been proposed; each works in terms of a total imbalance across the specified covariates. In its purest form, minimisation does not involve any random allocation; knowledge of all patient covariates allows the treatment allocations to be entirely predicted. To avoid such predictability of assignment, in practice a random element is usually added to a minimisation procedure. This is typically in terms of both using SR for some initial period of a trial, and in terms of allocating patients to the arm recommended by the minimisation routine with only a certain non-guaranteed chance.

With numerous randomisation routines available, previous works have sought to both describe exactly how these procedures work and also to compare their performance in terms of key metrics such as group size imbalance and lack of randomness. See, for example, Stigsby and Taves (2010), Berger et al. (2021), and Bruce et al. (2024). Most of this work has been restricted to the context of two-arm RCTs. Multi-arm trials are now increasingly recommended due to their improved efficiency from the shared control arm,

as well as other operational advantages. It is critical that investigators make deliberate, informed choices about allocation methods in the context of multi-arm trials, with it possible that variation in performance of available routines in a two-arm setting may be exacerbated further in a multi-arm domain.

Previous works that have addressed treatment allocation procedures within the context of multi-arm trials include Atkinson (2002), which compared seven allocation methods for multi-arm studies, each being extensions to biased coin designs, in terms of two performance metrics: loss of precision, and selection bias. Ryeznik and Sverdlov (2018) also compared several adaptive randomisation methods in terms of allocation balance, randomness, and numerous other statistical operating characteristics. Furthermore, Baldi Antognini and Zagoraiou (2011) proposed the covariate-adaptive biased coin design and compared it to Atkinson's DA-biased coin design in terms of loss of precision and selection bias. Atkinson's DA-biased coin design is known as an optimal method in adaptive randomisation. This method uses a general linear regression model and a biased coin allocation strategy for balancing across covariates to minimise the variance of the estimated treatment difference (Begg and Iglewicz, 1980).

Note that, in particular, I am unaware of any previous work that has considered covariate imbalance under the more popular treatment allocation methods for multi-arm trials or evaluated whether different methods are more suited to studies with a certain number of treatment arms. Seeking to address these issues and more, my primary goal in this chapter is to guide researchers on how to best allocate patients to treatments in a fixed-sample multi-arm trial. To achieve this, as well as considering performance measures relevant to two-arm trials, I also consider the multi-arm specific issue of there being several types of statistical power.

2.2 Methods

I assume that the trial contains K treatment arms, indexed by $k = 0, \ldots, K-1$, where arm $k = 0$ is a shared control arm and arms $k = 1, \ldots, K-1$ are experimental treatment arms. In the trial, N patients are to be allocated to one of the K treatments; I index the patients by $i = 1, \ldots, N$. I assume that the ideal goal would be to achieve a $1 : 1 : 1 : \cdots : 1$ allocation across the treatment arms. Alternative allocation targets could be treated similarly: I return to comment on such a setting in Section 2.6. For each patient, J covariates will be measured, indexed by $j = 1, \ldots, J$. I use $T_i = 0, \ldots, K-1$ as an indicator variable for the treatment patient i was assigned to and $X_{i,j}$ as the value of covariate j for patient i. For brevity in what follows later, I denote by $N_{k,l}$ the number of patients allocated to arm k amongst patients $i = 1, \ldots, l$

$$N_{k,l} = \sum_{i=1}^{l} \mathbb{I}(T_i = k).$$

Here, $\mathbb{I}(A)$ is the indicator function on event A. For simplicity, I similarly set N_k as the number of patients in arm k on trial conclusion, i.e., $N_k = N_{k,N}$. Denoting the outcome variable from patient i by Y_i, the following model is then used to analyse the data

$$Y_i = \mu + \sum_{j=1}^{J} X_{i,j}\beta_j + \sum_{k=1}^{K-1} \mathbb{I}(T_i = k)\theta_k + \epsilon_i.$$

Here, μ is a fixed intercept term, β_j is a fixed effect for covariate j, and θ_k is a fixed effect for the difference between experimental arm $k = 1, \ldots, K-1$ and the control arm $k = 0$. I assume that $\epsilon_i \sim N(0, \sigma^2)$, such that the above is a multiple linear regression model for the continuous outcomes Y_i. Our interest is assumed to be in performing two-sided tests for the difference between the experimental arms and the control. That is, in testing

$$H_{0,k} : \theta_k = 0, \quad H_{1,k} : \theta_k \neq 0,$$

for $k = 1, \ldots, K-1$. To evaluate evidence regarding these hypotheses, I use t-tests, rejecting $H_{0,k}$ when

$$\left| \frac{\hat{\theta}_k}{\sqrt{Var(\hat{\theta}_k)}} \right| > \Psi^{-1}_{N-J-(K-1)}(1 - \alpha/2),$$

where $\hat{\theta}_k$ is the estimated effect of the experimental arm k relative to the control arm, and $Var(\hat{\theta}_k)$ is the variance of the estimated effect for k. In addition, $\Psi_\nu(x)$ is the cumulative distribution function of a central t-distribution on ν degrees of freedom evaluated at x. This provides a per-hypothesis type I error rate of α. For simplicity, and because many treatment allocation methods work in terms of strata defined through binary covariates, I assume $X_{i,j} = 0, 1$. I discuss extension to continuous covariates in Section 2.6. In what follows, I will consider the impact of varying the values for many of the parameters above. Next, I describe the six considered distinct methods for assigning the treatment allocations T_i in the above-described type of multi-arm trial.

2.2.1 Considered randomisation methods

Simple Randomisation (SR)

In SR, patients are allocated completely at random to one of the K treatment arms. That is, $\mathbb{P}(T_i = k) = 1/K$ for each patient $i = 1, \ldots, N$ and each arm $k = 0, \ldots, K-1$.

Permuted Block Randomisation (PBR)

For PBR, I assume that the blocks are of length B, and that each block contains B/K assignments to each treatment, based on a pre-specified (random) permuted block ordering. Furthermore, I assume that all possible blocks of length B are employed. For

example, when $B = 3$ and $K = 3$, the permuted block used in each block could be any of 012, 021, 102, 120, 201, or 210. I will refer to this design as PBR(B).

Denote by $b_i = 1, 2, \ldots$ the block to which patient i belongs. Then, the probability that each patient is assigned to a given treatment depends only on the assignment of the previous patients in their block. Specifically, the probability of assignment to treatment k for patient i is given by

$$\mathbb{P}(T_i = k) = \frac{B/K - \sum_{l=1}^{i-1} \mathbb{I}(b_l = b_i)\mathbb{I}(T_l = k)}{B - \sum_{l=1}^{i-1} \mathbb{I}(b_l = b_i)}.$$

Stratified Block Randomisation (SBR)

Often, PBR is used within pre-defined strata. This is SBR. I denote by $s_i = 1, \ldots, S$ the strata to which patient i belongs. As for PBR, I assume that all possible blocks of length B are used within each stratum, and refer to this design as SBR(B). Then the probability patient i is assigned to treatment k is given by

$$\mathbb{P}(T_i = k) = \frac{B/K - \sum_{l=1}^{i-1} \mathbb{I}(b_l = b_i)\mathbb{I}(s_l = s_i)\mathbb{I}(T_l = k)}{B - \sum_{l=1}^{i-1} \mathbb{I}(b_l = b_i)\mathbb{I}(s_l = s_i)}.$$

In what follows, I assume that the strata are defined based on all possible combinations of the presence/absence of covariates $j = 1, \ldots, J_{rand}$, for $J_{rand} \leq J$, such that there are $S = 2^{J_{rand}}$ strata, and J_{rand} is the number of covariates included in the randomisation scheme.

Urn Design (UD)

In the UD, the randomisation algorithm dynamically changes the treatment assignment probabilities based on the degree of assignment imbalance, with the goal of achieving balance between treatment arms over time. It can be understood by supposing an urn contains balls with K different labels (e.g., $0, \ldots, K-1$), with each label relating to

a specific treatment. Initially, there are w balls with each label. When a new patient enters the trial, a ball is drawn randomly; if its label is k then this patient is allocated to arm k. This ball is then replaced. In addition, α balls of label k are added to the urn, as well as β balls with labels $k' \neq k$. This procedure is repeated for each assignment. Although real urns may be used to demonstrate the procedure, in practice urns are simulated on a computer, especially when α or β are not integers. The probability mass function of the first assignment is given by $\mathbb{P}(T_1 = k) = 1/K$. Then, the probability of assignment to treatment k for patients $i = 2, \ldots, N$ is

$$\mathbb{P}(T_i = k) = \frac{w + \alpha N_{k,i-1} + \beta\{(i-1) - N_{k,i-1}\}}{Kw + \{\alpha + \beta(K-1)\}(i-1)}.$$

I will refer to an UD with particular parameter values as UD(w, α, β).

Block Urn Design (BUD)

The randomisation process for the BUD can be illustrated by a model with two urns, termed active and inactive. Suppose again that the block length is B, which is a multiple of K; explicitly, suppose that $B = \lambda K$, where λ is the number of 'minimal balanced sets' in each block. This λ is pre-defined by investigators to control the imbalance between treatment arms. Note that when the block contains only one minimal balanced set (i.e., $\lambda = 1$), the BUD is equivalent to PBR.

Then, the allocation procedure starts with an empty inactive urn and λ balls with K distinct labels (again $0, \ldots, K-1$) in the active urn. When a treatment assignment is requested, a ball is randomly selected; if this is of label k then this patient is assigned to arm k. This ball is then placed in the inactive urn. This process is repeated for each assignment until a minimal balanced set is present in the inactive urn. Then, one ball of each label is returned to the active urn from the inactive. Other balls, if any, remain in the inactive urn. Then, balls are again drawn from the active urn and placed into the

inactive urn until once more it contains at least one ball of each label, which are then again transferred to the active urn. This entire process is repeated until all patients have been assigned. In this case, it can be shown that

$$\mathbb{P}(T_i = k) = \frac{\lambda + \min_{l=0,\ldots,K-1} N_{l,i-1} - N_{k,i-1}}{B + K \min_{l=0,\ldots,K-1} N_{l,i-1} - (i-1)}.$$

I will refer to a BUD using a particular value of λ as BUD(λ).

Stratified Block Urn Design (SBUD)

In SBUD, I apply the BUD within pre-defined strata. As above, $s_i = 1, \ldots, S$ denotes the strata to which patient i belongs. Furthermore, as for the BUD, I assume that the block length within each stratum is $B = \lambda K$, with λ once more the number of 'minimal balanced sets' in each block. Then the probability patient i is assigned to treatment k is given by

$$\mathbb{P}(T_i = k) = \frac{\lambda + \min_{l=0,\ldots,K-1} N_{l,i-1,s_i} - N_{k,i-1,s_i}}{B + K \min_{l=0,\ldots,K-1} N_{l,i-1,s_i} - \{\sum_{l=1}^{i} \mathbb{I}(s_l = s_i) - 1\}}.$$

Here, $N_{k,i,s}$ is the number of patients that have been assigned to treatment k in stratum s, amongst the first i patients. That is

$$N_{k,l,s} = \sum_{l=1}^{i} \mathbb{I}(T_i = k)\mathbb{I}(s_i = s).$$

As for SBR I assume that the strata are defined based on all possible combinations of the presence/absence of covariates $j = 1, \ldots, J_{rand}$, for $J_{rand} \leq J$, such that there are $S = 2^{J_{rand}}$ strata. I will refer to a SBUD using a particular value of λ as SBUD(λ).

Minimisation

Minimisation can be illustrated as follows: set $N_{i,j,k,l}$ as the number of patients that have been assigned, amongst the first i patients, to treatment k, whose j^{th} covariate takes the value l. That is

$$N_{i,j,k,l} = \sum_{m=1}^{i} \mathbb{I}(T_m = k)\mathbb{I}(X_{m,j} = l).$$

Then, suppose that the next patient, patient $i+1$, has covariate information $X_{i+1,1}, \ldots, X_{i+1,J}$. The number of patients thus far at these levels, for arm k, is given by $N_{i,j,k,X_{i,j}}$ for $j = 1, \ldots, J$. The marginal imbalance on each level of covariate j across all treatment arms can be measured by the range as follows

$$RG(N_{i,j,0,X_{i,j}}, \ldots, N_{i,j,K-1,X_{i,j}}) = \max_{k=0,\ldots,K-1} N_{i,j,k,X_{i,j}} - \min_{k=0,\ldots,K-1} N_{i,j,k,X_{i,j}}.$$

For brevity, I focus here on the range as an intuitively simple measure of imbalance, but note that other measures are also possible. See Jin et al. (2019) for further details. The total hypothetical imbalance, if patient $i+1$ is assigned to arm k, is then defined as a weighted sum of the level-based imbalances for covariates that are including in the minimisation routine. I assume these are covariates $j = 1, \ldots, J_{rand}$, where $J_{rand} \leq J$ is the number of covariates included in the randomisation scheme

$$I_{i+1,k} = \sum_{j=1}^{J_{rand}} w_j RG(N_{i,j,0,X_{i,j}}, \ldots, N_{i,j,k,X_{i,j}} + 1, \ldots, N_{i,j,K-1,X_{i,j}}).$$

Here, the weights w_j are pre-specified and indicate the relative importance of covariates in measuring the imbalance (Jin et al., 2019). To incorporate a random component into the minimisation routine, patient i is allocated to a treatment arm as follows

$$\mathbb{P}(T_i = k) = \begin{cases} \frac{p_{min}}{|\underset{i=0,\ldots,K-1}{\operatorname{argmin}} \{I_{i-1,0},\ldots,I_{i-1,K-1}\}|} & : k \in \underset{i=0,\ldots,K-1}{\operatorname{argmin}} \{I_{i-1,0},\ldots,I_{i-1,K-1}\}, \\ \frac{1-p_{min}}{(K-1)-|\underset{i=0,\ldots,K-1}{\operatorname{argmin}} \{I_{i-1,0},\ldots,I_{i-1,K-1}\}|} & : \text{otherwise.} \end{cases}$$

That is, probability p_{min} is dedicated to assigning patient i to one of the arms that minimises the subsequent imbalance, and probability $1 - p_{min}$ to the remaining arms. Here, I also consider a modification to the algorithm above such that patients $i = 1, \ldots, \lceil qN \rceil$ are allocated using SR (i.e., the first $100q\%$ of patients are allocated via $\mathbb{P}(T_i = k) = 1/K$; this is sometimes termed a 'burn in' period). Only after this is the minimisation routine described above employed.

Later, I will fix $q = 0.1$, but will consider different values of p_{min} and thus will identify a particular minimisation routine as Mini(p_{min}).

2.2.2 Performance evaluation metrics

As I mentioned in the introduction Section 1.1, randomisation is the cornerstone of a clinical trial in achieving a balance between treatment arms while considering patient covariates that could influence outcomes and following the rules of random allocation. Therefore, in this section, I examine some important evaluation metrics that are used to compare the different randomisation methods. These include group size balance, covariate balance, loss of precision (i.e., an increase in the variance of treatment effect estimates), and allocation predictability. I also compare different definitions of statistical power relevant in multi-arm trials: marginal, disjunctive, and conjunctive powers.

Maximal imbalance according to group sizes

The maximum absolute difference between the number of patients in each experimental arm and the shared control arm, $\max(|N_1 - N_0|, \ldots, |N_{K-1} - N_0|)$, is used as our first

performance metric.

Degree of imbalance according to patient covariates

To evaluate the ability of the allocation techniques to balance covariate factors between arms, the proportion of patients in each arm with $X_{i,j} = 1$ was calculated as

$$c_{j,k} = \frac{N_{N,j,k,1}}{N_k}.$$

Then, the maximum absolute difference in these proportions for each experimental arm compared to the control, for each covariate j, was calculated as

$$C_j = \max\left(|c_{j,1} - c_{j,0}|, \ldots, |c_{j,K-1} - c_{j,0}|\right).$$

The value of $\max\left(C_1, C_2, \ldots, C_J\right)$ is used as our second performance metric. Note that the range of j here is $j = 1, 2, \ldots, J$, i.e., all covariates (not just those potentially used in the randomisation routine) are used in computing the maximal covariate imbalance.

Maximal treatment effect variance inflation

'Loss' is one of the evaluation methods that has previously been used to compare different multi-arm trial randomisation methods. It relates to the increase in the variance of treatment effect estimates due to the imbalance caused by randomisation and is defined based on a general formula for the variance of the treatment estimates when accounting for such imbalance (Atkinson, 2002; Baldi Antognini and Zagoraiou, 2011). For ease of interpretation, I here work directly in terms of the treatment effect variances, examining the percentage by which they increase relative to their theoretical value under perfect balance.

To this end, note that in the case of perfectly balanced assignment (in terms of

both arms and covariates), the variance of each treatment effect would be given by $2K\sigma^2/N$ (Atkinson, 2002). Using this, the maximal treatment effect variance inflation is computed as follows, and used as our third performance metric

$$\frac{\max\left\{Var(\hat{\theta}_1), \ldots, Var(\hat{\theta}_{K-1})\right\}}{\frac{2K\sigma^2}{N}}.$$

Allocation predictability

Predictability is a popular indicator that has been used to measure the lack of randomness of an allocation routine. It is defined as the proportion of times an investigator correctly guesses the next patient assignment when the investigator is assumed to know the allocation of all previous patients and they guess the next allocation as whichever arm currently has the fewest assignments (randomly allocating when multiple arms have equal lowest assignment). Formally, this is that the investigator guesses the assignment of patient i, G_i, as

$$\begin{aligned} G_i \sim\ & Multinomial(p_{i,0}, \ldots, p_{i,K-1}), \\ p_{i,k} = & \frac{\mathbb{I}\{N_{k,i-1} = \min\left(N_{0,i-1}, \ldots, N_{K-1,i-1}\right)\}}{\sum_{l=0}^{K-1} \mathbb{I}\{N_{l,i-1} = \min\left(N_{0,i-1}, \ldots, N_{K-1,i-1}\right)\}}. \end{aligned}$$

Then, the predictability, our fourth performance metric, is given by

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{I}(T_i = G_i).$$

Power

It is also important to consider the statistical power of the trial to detect present treatment effects (when they exist). However, the power of a multi-arm trial is more complex compared to a two-arm trial. In a multi-arm trial, power can be defined in various ways.

I consider

- The marginal power for a specific arm k, $P_k(\theta_k)$, which is the probability of rejecting the null hypothesis $H_{0,k}$ when the null hypothesis is not true.

$$P_k(\theta_k) = \mathbb{P}(\text{Reject } H_{0,k}|\theta_k).$$

- The disjunctive power $P_{dis}(\theta_1, \ldots, \theta_{K-1})$, which is the probability of rejecting at least one of the null hypotheses $H_{0,1}, \ldots, H_{0,K-1}$

$$P_{dis}(\theta_1, \ldots, \theta_{K-1}) = \mathbb{P}(\text{Reject at least one of } H_{0,1}, \ldots, H_{0,K-1}|\theta_1, \ldots, \theta_{K-1}).$$

- The conjunctive power $P_{con}(\theta_1, \ldots, \theta_{K-1})$, which is the probability that all of the null hypotheses $H_{0,1}, \ldots, H_{0,K-1}$ are rejected

$$P_{con}(\theta_1, \ldots, \theta_{K-1}) = \mathbb{P}(\text{Reject all of } H_{0,1}, \ldots, H_{0,K-1}|\theta_1, \ldots, \theta_{K-1})$$

Where all these powers are conditional on the treatment effects being non-zero ($\theta_k \neq 0$). The simulation results I present focus on the change in power. To contextualise this information across different settings, I also consider the change in sample size this represents under the assumption of equal allocation across the arms. For example, a 1% difference in power from 60% to 61% does not represent a large change in the sample size, but a change in the power from 98% to 99% represents a much larger change in the sample size.

2.3 Motivating examples

The two different settings that I considered based on real clinical trials were as follows.

The first setting (Burant et al., 2012) (Setting 1) is a randomised control trial of TAK-875 in outpatients with type 2 diabetes who had not responded to diet or metformin treatment. On the assumption of a standard deviation of 1.1% for the change from baseline in HbA1c, 420 patients were considered (for a 10% dropout rate) as sufficient to achieve at least 80% power to detect a relevant treatment effect of 0.6% between TAK-875 and placebo using a two-sample t-test at a 5% type I error rate (i.e., no multiplicity adjustment was made). Ultimately, 384 patients were randomly assigned to receive one of seven arms (i.e., $K = 7$): placebo, five doses of TAK-875, or glimepiride. In our simulation study, I considered 350 as the default total sample size distributed amongst the treatments, considering in all $N = 175, 350, 700$. Correspondingly, I set $\delta = 0.006$ and $\sigma = 0.011$.

The second setting (Diacon et al., 2012) (Setting 2) is a prospective RCT in which patients were recruited from outpatient clinics in Cape Town and randomised to receive one of the following treatments: bedaquiline, bedaquiline-pyrazinamide, PA-824-pyrazinamide, bedaquiline-PA-824, PA-824-moxifl oxacinpyrazinamide, or unmasked standard antituberculosis treatment as a positive control (i.e., $K = 6$). The primary outcome was the change in 14-day early bactericidal activity. Eighty-five eligible patients were planned for inclusion in the study, based on achieving 80% power for each comparison of an experimental arm against control at a standardized effect size of 1.2 with $\alpha = 0.05$ (i.e., without multiplicity adjustment), allowing for a 20% dropout rate. In our simulation study, I therefore considered $N = 43, 85, 170$, setting $\delta = 1.2$ and $\sigma = 1$.

2.4 Simulation Study

The goal of this simulation study is to evaluate the randomisation methods described in Section 2.2.1 in terms of group size imbalance, covariate imbalance, treatment effect variance, predictability, and power (marginal, disjunctive, and conjunctive). These evaluation metrics have been discussed in the literature as standard criteria for assessing randomisation methods in several studies (see, e.g., Bruce et al. (2024)). The choice of parameters is informed by two real clinical trial settings to make decisions and robustness assessment. Full details about the statistical models and parameter values underpinning the data simulations are presented below.

Parameter assumptions

I utilise simulation to evaluate the performance measures discussed in Section 2.2.2 for the considered allocation methods (see Table 2.1). I chose two quite different real trials (Burant et al., 2012; Diacon et al., 2012) that motivate the simulation studies' assumptions in terms of variation in sample size, the number of experimental arms, treatment effects, and the standard deviation.

The following summarises the assumptions common across the two considered settings. I consider three values for N in each instance: taken to be half of, equal to, and twice that in the motivating real trial. All simulations were performed under the 'global alternative hypothesis' in which $\theta_k = \delta > 0$ for $k = 1, \ldots, K-1$. I assumed the presence of four binary covariates ($J = 4$), with the value of $X_{i,j}$ drawn independently for all i and j as $X_{i,j} \sim Bern(0.25)$.

I assumed two different situations regarding these covariates use in the randomisation routines (where applicable). First, I assumed that the four binary covariates were all used in the randomisation (i.e., $J_{rand} = 4$). This assumption allows the evaluation of the impact of a large number of strata ($S = 16$) on small clinical trials, as demonstrated

in setting 2 (see, Table 2.2 for more details). Second, I assumed that only the first two covariates were used in the randomisation (i.e., $J_{rand} = 2$); this allows assessment of how well covariates not included in the randomisation routine are balanced.

I also assumed different effects of the four covariates, loosely classified as large, medium, small, and no effect on the mean of the outcome. This allowed us to determine how different randomisation procedures affect the power in the presence of different covariate effect sizes. Specific values for the effect of the covariates on the mean of the outcome data were nominated as $\beta_1 = \delta$, $\beta_2 = 0.5\delta$, $\beta_3 = 0.1\delta$, and $\beta_4 = 0$.

Each of the allocation methods discussed in Section 2.2 was examined. For the allocation methods utilising blocks (i.e., PBR and SBR) I considered $B = K$ and $B = 3K$. In the UD, I assumed $w = \alpha = 1$ and $\beta = 2$, i.e., UD$(1, 1, 2)$. In the BUD and SBUD, I assumed $\lambda = 3$. For minimisation, I considered $q = 0.1$ and $p_{min} = 0.7, 0.9$. Using different levels of p_{min} (moderate and stronger levels to achieve covariate balance) allows for a closer examination of the differences between moderate and strong probability assignments in achieving covariate balance while reducing predictability.

For each parameter combination, 10,000 simulation replicates were carried out to empirically estimate the performance metrics described in Subsection 2.2.2. Finally, I calculated the Monte Carlo error (MC) for each parameter combination by dividing the standard deviation by the square root of the number of simulation replicates.

Table 2.1: Randomisation Methods Evaluated.

Method	Acronym
Simple Randomisation	SR
Urn Design	UD
Permuted Block Randomisation	PBR
Block Urn Design	BUD
Stratified Block Randomisation	SBR
Stratified Block Urn Design	SBUD
Minimisation	Mini

Table 2.2: Settings and Scenarios parameters Considered in the Simulation Study.

Symbol	Definition	Setting 1	Setting 2
K	Number of experimental treatment arms.	$K = 7$	$K = 6$
N	Total sample size.	N=175, 350, 700	N=43, 85, 170
δ	Treatment effect	$\delta = 0.006$	$\delta = 1.2$
σ	Standard deviation	σ=0.011	σ=1

Table 2.3: Glossary of the assumptions common parameters across the two considered settings.

Symbol	Definition
$J = 4$	Total number of covariates.
$B = K$, $3K$	Number of blocks used in PBR, BUD, SBR, and SBUD.
$w = 1$	Number of balls with each label in the UD.
$\alpha = 1$	The number of additional balls added to the urn for the treatment arm that a newly enrolled patient is assigned to in the UD.
$\beta = 2$	The number of additional balls added to the urn for the other treatment arms in the UD.
$\lambda = 3$	'minimal balanced sets' in the BUD and SBUD.
$q = 0.1$	'burn-in' period in the minimisation.
$p_{min} = 0.7$, 0.9	probability assignment in the minimisation.

2.5 Results

2.5.1 Degree of imbalance according to group sizes

The simulation results comparing the performance of the randomisation methods with respect to balancing group sizes are illustrated in Figure 2.1. As expected, PBR(K), PBR($3K$), and BUD(3) have consistently low group size imbalance. In contrast, SR and UD$(1, 1, 2)$ performed less well in both settings, with the maximum mean imbalance slightly higher in SR. As would have been anticipated, the larger the block size, the greater mean maximal imbalance in PBR and SBR. Furthermore, SBUD(3) performed slightly less well in both settings than the SBR procedures. The results indicate SBR($3K$) or SBUD(3) should potentially not be considered if group size balance is critical. Overall though, there is little difference between the PBR, SBR, BUD, and SBUD designs across the considered scenarios.

When $J_{rand} = 2$, the mean maximum imbalance in the number of patients allocated to each arm for Mini(0.7) becomes much larger than its result with $J_{rand} = 4$. Our findings indicate that for the smaller considered value of J_{rand}, Mini(0.7) performs more like SR and the UD, while it is more comparable to PBR, SBR, BUD, and SBUD when $J_{rand} = 4$. In contrast, Mini(0.9) has low mean maximal imbalance for both $J_{rand} = 2$ and $J_{rand} = 4$. This highlights the importance of careful choice of p_{min}. To investigate this further, a small number of additional simulations were performed for the minimisation setting. They indicated, as would be expected, that the mean maximal imbalance increases as p_{min} decreases (as the procedure then functions more like SR).

In both settings and scenarios, the maximum MC error of 0.14 was observed for SR, with a mean maximum imbalance of 10.48 and a 95% confidence interval of [10.21, 10.75].

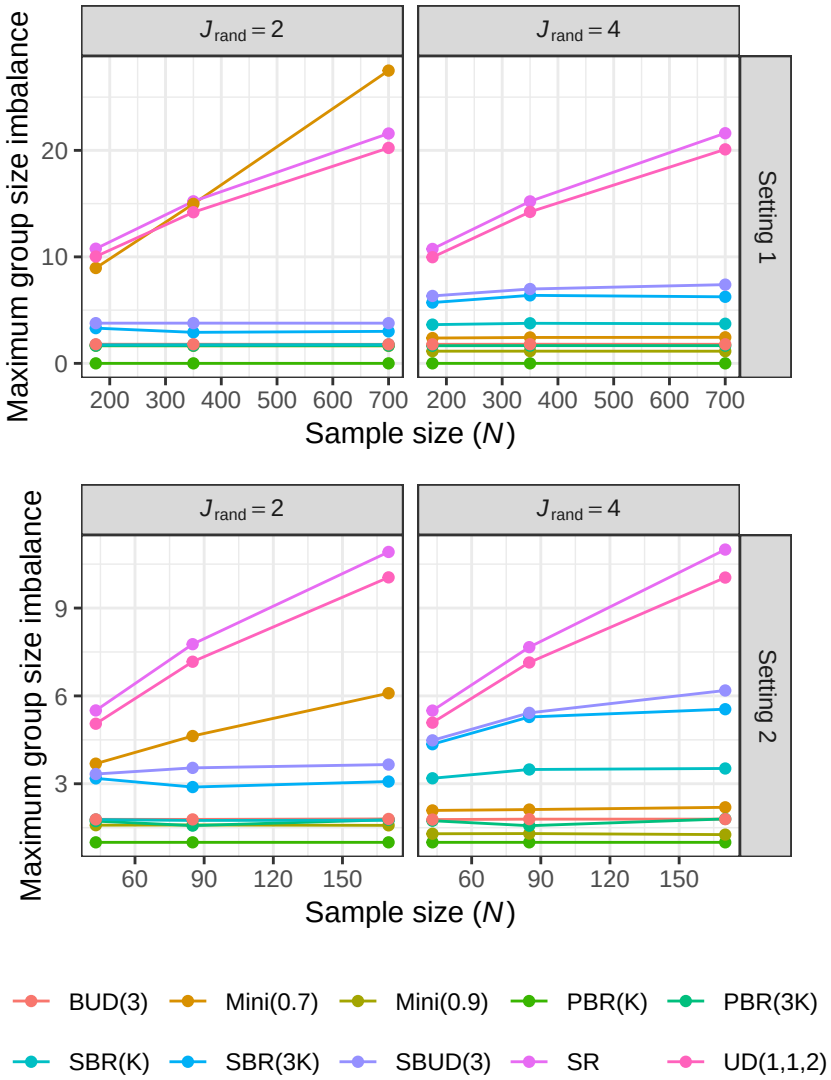

Figure 2.1: The empirical mean maximum group size imbalance between the control arm and each of the experimental arms is shown by setting and the value of J_{rand}.

2.5.2 Degree of imbalance according to patient covariates

The simulation results comparing covariate imbalance for different randomisation methods are displayed in Figure 2.2. As would be anticipated, the mean maximal covariate imbalance decreases for all methods as a function of the sample size. As many not have been anticipated, the degree of imbalance reduction as N is increased appears similar for each method.

In both settings, SR, UD$(1, 1, 2)$, the PBR designs, and BUD(3) have similar covariate imbalance based on our chosen metric. This is to be expected given none utilise covariate information in the patient allocation. For $J_{rand} = 4$, Mini(0.9) is the best performing method. Mini(0.7) and SBR(K) are the next best methods, performing similarly. SBR($3K$) and SBUD(3) have slightly weaker, but still strong, performance. This indicates a large p_{min} may be necessary to outperform SBR and SBUD in some scenarios.

By contrast, for $J_{rand} = 2$, I found that all considered allocation methods had similar mean maximal covariate imbalances, albeit with a small difference between randomisation methods for the lowest sample size in both settings. SBUD(3) had arguably the best performance for $J_{rand} = 2$. This highlights that methods which force balance on certain covariates provide no improvement in balance for other uncorrelated covariates of interest compared to simpler techniques such as SR and PBR.

To explore the influence of p_{min} on the performance of the minimisation method, Appendix A.1 presents results for $J_{rand} = 1, 2, 3, 4$. These indicate that, as would be anticipated, the imbalance for covariates not utilised in the allocation are unaffected by the value of p_{min}. However, for those covariates used in the allocation, a large reduction in imbalance can be achieved by increasing p_{min}. This is particularly true when $J_{rand} = 1$.

In both settings and scenarios, the maximum MC error of 0.002 was observed, with a mean maximum covariate imbalance of 0.20 and a 95% confidence interval of [0.196, 0.20].

2.5.3 Maximal treatment effect variance inflation

The simulation results comparing the treatment effect variance inflation are displayed in Figure 2.3, given as empirical mean percentages. Reflecting the results in Figures 2.1 and 2.2, they show that the highest treatment effect variance inflation results from SR and UD$(1, 1, 2)$, and that the inflation decreases as a function of the sample size. For Setting 1 with $J_{rand} = 2$, Mini(0.7) also performs poorly. SBR($3K$) and SBUD(3)

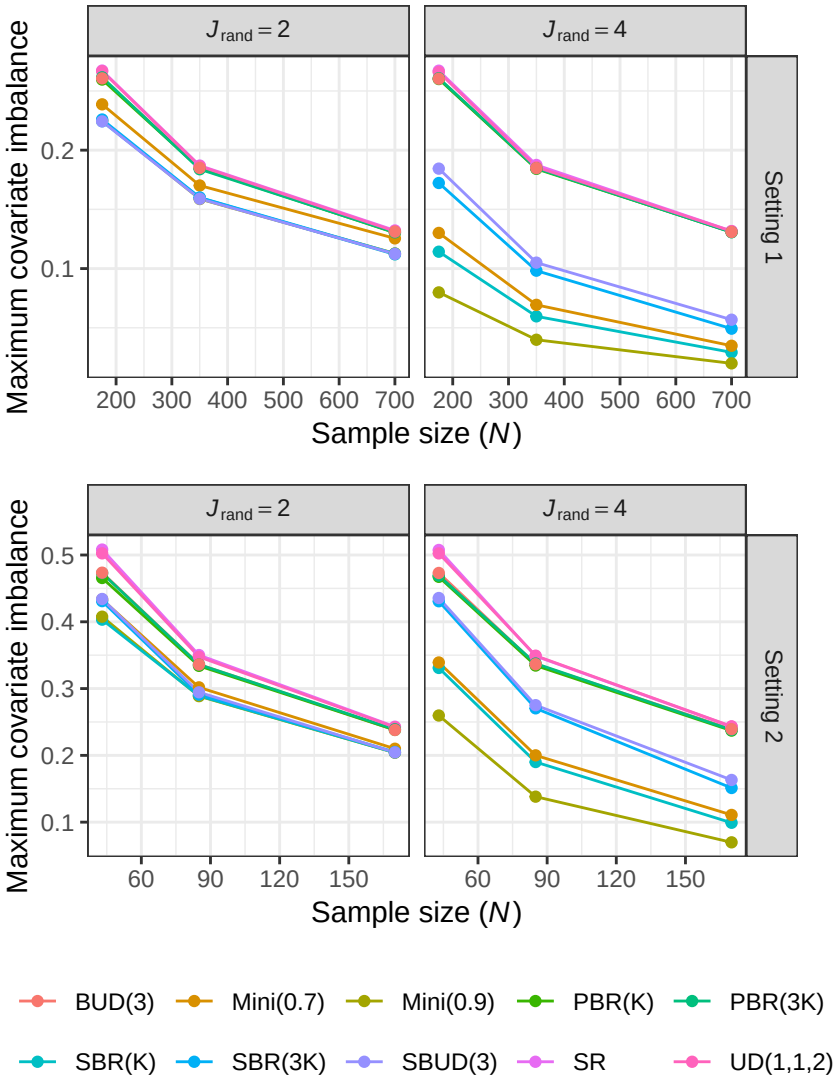

Figure 2.2: The empirical mean maximum covariate imbalance between the control arm and each of the experimental arms is shown by setting and the value of J_{rand}.

are generally the next worst performing procedures. Mini(0.9) has consistently low maximum treatment effect variance inflation.

In both settings and scenarios, the maximum MC error of 0.014 was observed for SR, with a maximum treatment effect variance inflation of 0.98 and a 95% confidence interval of [0.95, 1.00].

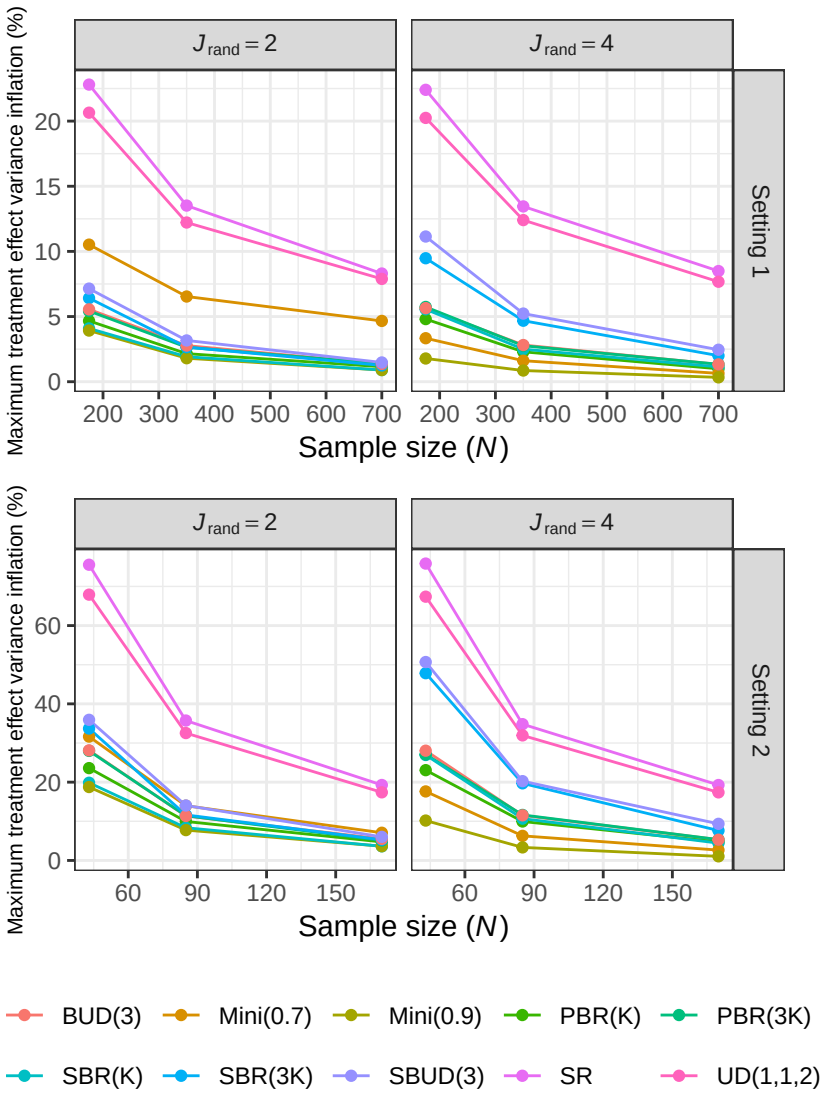

Figure 2.3: The empirical mean maximum treatment effect variance inflation (given as a percentage) is shown by setting and the value of J_{rand}.

2.5.4 Allocation predictability

Predictability of allocation is shown in Figure 2.4. It is clear that predictability is generally unaffected by the sample size. In Setting 1, the simulation results show that the lowest predictability results from using SR, UD$(1, 1, 2)$, SBUD(3), or SBR$(3K)$. For these methods, the mean proportion of correct guesses is similar to guessing at random. By contrast, the predictability of PBR(K) is much higher.

As expected, the predictability of Mini(0.7) is lower when compared to Mini(0.9). In Setting 2, notably, the predictability of Mini(0.7) is lower than PBR($3K$).

When $J_{rand} = 2$, as opposed to $J_{rand} = 4$, there was a small (3-6%) decrease in predictability for minimisation. For the SBR procedures and SBUD(3), predictability was instead lower for $J_{rand} = 4$.

In both settings and scenarios, the maximum MC error of 0.0003 was observed, with a mean predictability of 0.14 and a 95% confidence interval of [0.139, 0.14].

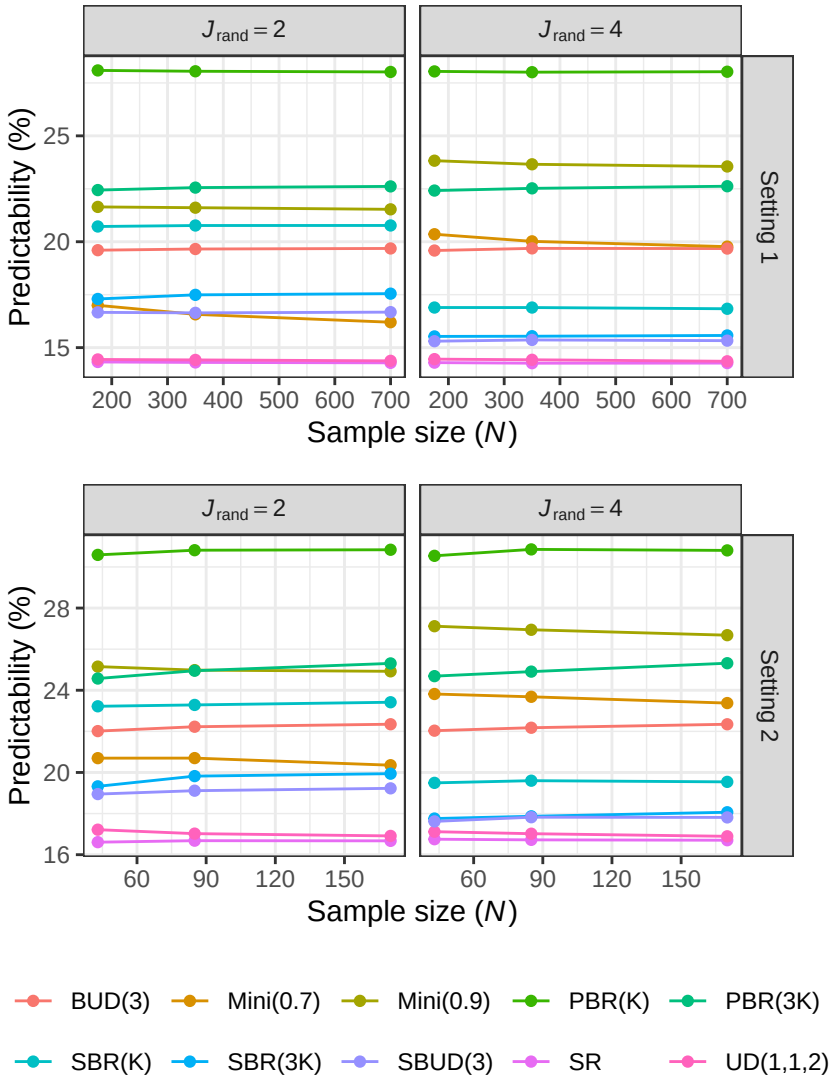

Figure 2.4: The empirical mean predictability is shown by setting and the value of J_{rand}.

2.5.5 Power

Figure 2.5 displays the results for the marginal (for $H_{0,1}$), disjunctive, and conjunctive power when $J_{rand} = 4$ (results for $J_{rand} = 2$ are not shown but are very similar to Figure 2.5). As expected, all powers increase with the sample size for all allocation methods. In most cases there is evidently little difference in the resultant powers across the considered methods. The only possible exception to this is the smallest considered sample size in Setting 2.

To examine whether the small observed variations in power seen in Figure 2.5 correspond to large changes in sample size, Figure 2.6 indicates how the powers vary under SR as a function of the sample size. Consider for example the marginal power in Setting 2 under Mini(0.9) when $N = 85$, estimated as approximately 88%. By contrast, UD$(1, 1, 2)$ has the smallest marginal power, estimated as approximately 84%. Figure 2.6 tells us that an increase in marginal power from 84% to 88% may be considered as corresponding to an increase in sample size from approximately 86 to 96 under SR. Thus, the variations in power in Figure 2.5 do not represent a large change in total sample size of the trial.

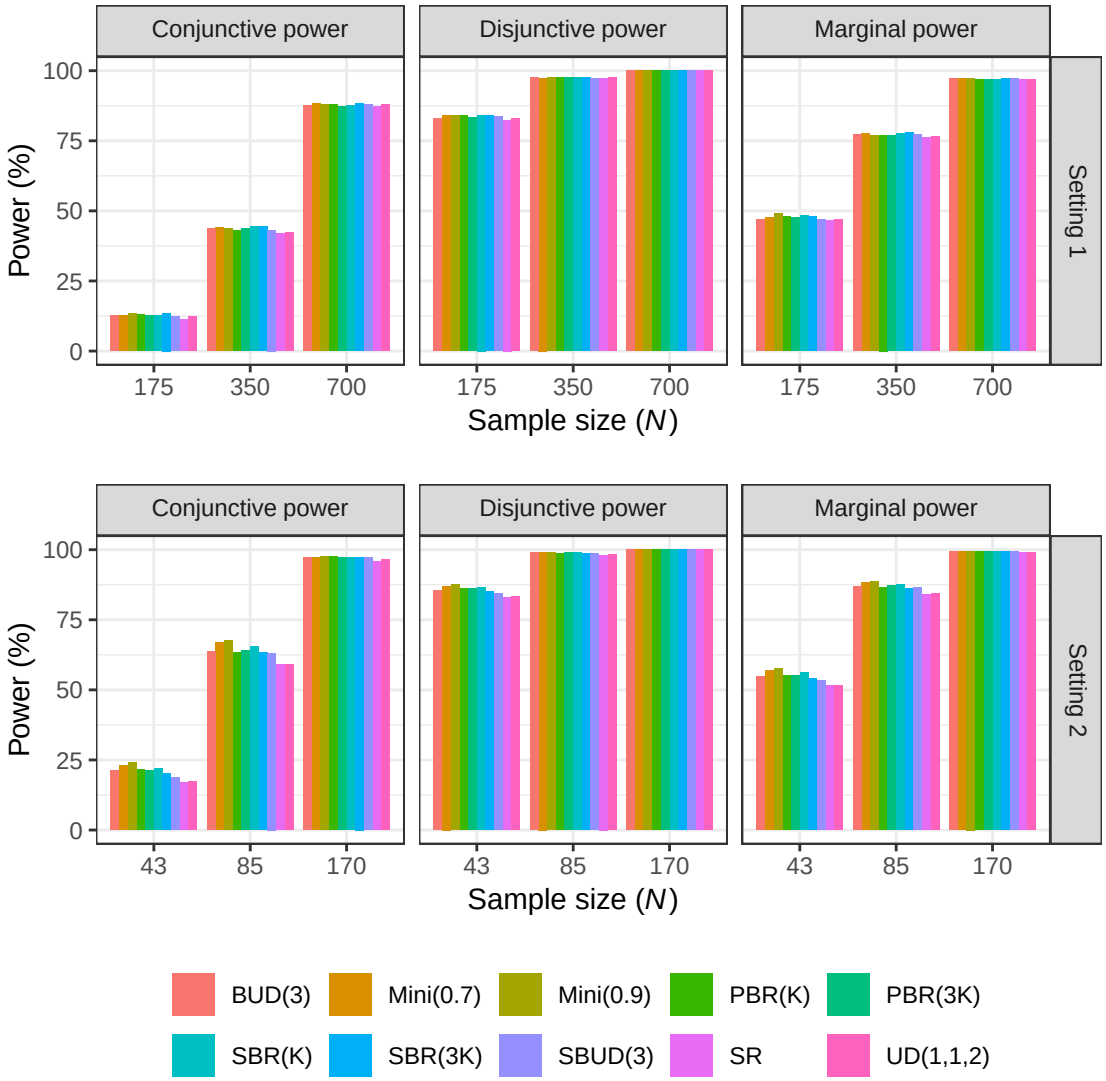

Figure 2.5: The empirical mean marginal, disjunctive, and conjunctive power are shown by setting, for $J_{rand} = 4$.

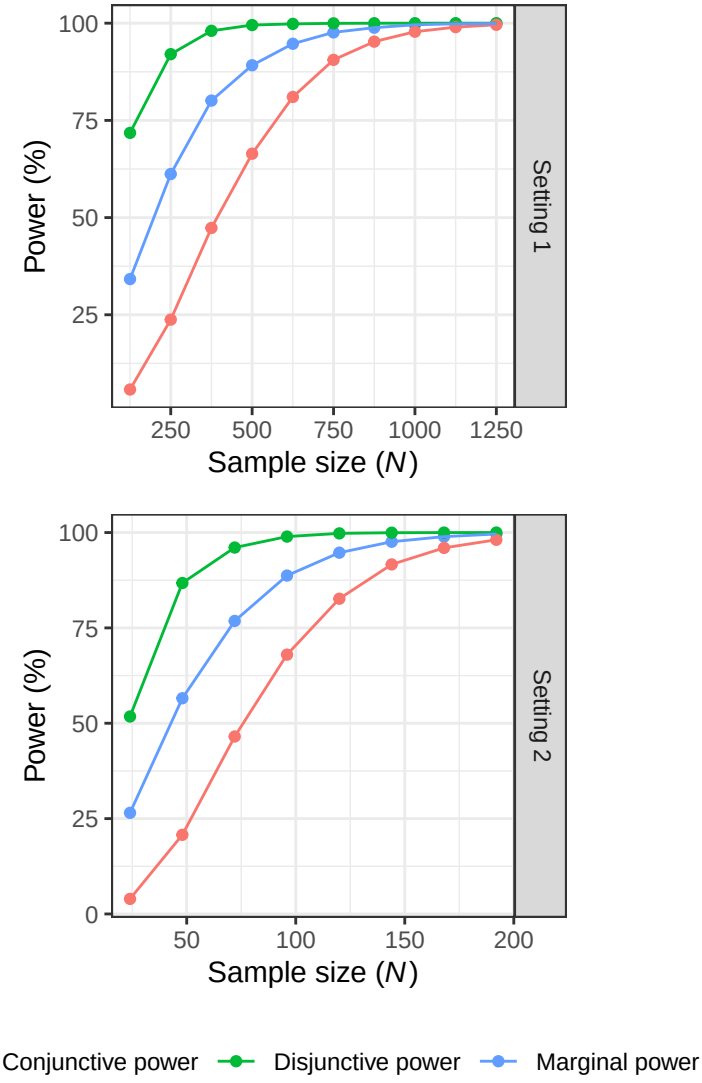

Figure 2.6: The estimated marginal, disjunctive, and conjunctive power under simple randomisation is shown by setting as a function of the trial's total sample size.

2.6 Discussion

Various allocation methods have been proposed to minimise the imbalance between treatment arms according to group sizes and important patient covariates. However, there has been little consideration of their relative performance for multi-arm trials. This study provides useful insights on this, with a simulation study motivated by two real multi-arm trials.

The selection of an appropriate allocation method for a trial may be influenced by many factors, including the number of arms, the sample size of the trial, and the number of covariates considered. However, the goals for patient randomisation in comparative clinical trials remain the same: to ensure the allocation is not predictable and to control imbalance in group sizes and according to key prognostic factors. Depending on the setting, these goals may be prioritised differently. To inform the selection of the most appropriate method, it is important to evaluate different options based on their performance.

In both settings, the best allocation methods for minimising imbalance in the group sizes were the PBR procedures and the considered BUD (Figure 2.1). However, arguably there was little overall difference in the mean maximal group size imbalance when considered relative to the total sample size.

With regard to covariate imbalance, when $J_{rand} = 2$, all allocation methods had similar maximal covariate imbalance in both settings, because of the two covariates not utilised in the allocation (Figure 2.2). This emphasises an important point, discussed many times previously in the two-arm context (see, e.g., Taves (1974) and Atkinson and Donev (1992)), that SBR, SBUD, and minimisation will not provide improved balance on covariates not included in the allocation routine that are uncorrelated with those included in the allocation routine. The effectiveness of SBR, SBUD, and minimisation may thus be highly dependent on pre-trial knowledge of which covariates may affect the

outcome data.

Using either PBR or BUD within strata (i.e., the SBR and SBUD designs) had similar, strong performance. The advantages of MTI procedures or PBR methods have been discussed extensively within the literature in relation to two-arm trials not attempting to balance for covariates (see, e.g., Berger et al. (2016)). Such advantages did not appear clear in the considered multi-arm setting with multiple strata.

I note again that Mini(0.9) performed better than SBR(K) when $J_{rand} = 4$ in terms of covariate balance. Thus, minimisation can be more effective at balancing larger numbers of covariates. However, this comes at a large cost in terms of the predictability (see below and Figures 2.3 and 2.4). Predictability of assignments provides the potential for selection bias, which increases with a higher frequency of predictable assignments. The simulation results demonstrate a differential effect on how the block size in the various methods has an impact on the mean predictability. Specifically, whilst the smaller block size increased predictability by a large amount for PBR, SBR was less affected. Significantly, the minimisation routines considered had a much larger predictability than the SBR and SBUD routines when $J_{rand} = 4$; evidencing a minimisation penalty noted above. We note, though, that this is for our particular definition of predictability and may not be the case if one assumes guesses are made with some knowledge of covariate data.

Regarding treatment effect variance inflation, a desirable allocation rule has low variance alongside low predictability. However, these factors generally work against each other. As a result, the randomisation rules with the highest predictability typically had the lowest variances.

Despite some variation in the mean group size imbalances, as well as in the covariate imbalance (Figures 2.1 and 2.2), the simulation results showed that the choice of allocation method did not have a major impact on the different types of power (Figure 2.5). The only notable differences appeared in Setting 2 due to the small assumed sample size.

Thus, there is little need to consider power when choosing an allocation methodology in the design scenario considered (Wason et al., 2014).

Providing a short precise summary of the findings is challenging given the volume of results and metrics considered. However, there are certain conclusions that we believe can be drawn. Evidently, the UD and SR had the highest imbalance in group sizes and covariates, causing them to have the largest treatment effect variance inflation. However, they had the lowest predictability. PBR with a small block size is perhaps the most commonly used randomisation method (McPherson et al., 2012; Parmar et al., 2014). Based solely on our considered metrics, a strong argument can be made that SBR is an overall better choice, providing the best trade off in performance overall and simplicity. The SBUD is theoretically superior, but such theoretical advantages may not outweigh the increased complexity in the multi-arm context unless the trial is very large in terms of both arms and strata. Despite controversy regarding the use of minimisation in randomised trials (Treasure and MacRae, 1998; Taves, 2010), minimisation was more successful in achieving a combination of group size balance and covariate balance compared to using either SBR or SBUD. Its predictability and treatment effect variance inflation were comparable to a considered PBR routine. However, its gains over SBR and SBUD may be considered insubstantial when weighed against a greatly increased predictability. A potential caveat to this statement is the dependence of the performance of minimisation on the choice of the parameter p_{min}, though Figures 2.2 and 2.4 indicate that it would likely not be possible to find a value for p_{min} resulting in similar predictability to SBR or SBUD that retains an advantage in covariate balance (Brown et al., 2005).

I conclude with some limitations of this work. In this chapter, I considered several popular randomisation methods. These represent only a small subset of available methods, and examination of other methods can be considered for future work. In addition, I only considered balancing up to four binary covariates. Such categorical covariates

are common in clinical trials. Even continuous covariates are typically discretised to be included in the randomisation routine (Taves, 2010). However, the breakdown of a continuous covariate into subcategories may lead to loss some of important information (Stigsby and Taves, 2010). Therefore, there are several allocation methods that have been proposed (Su, 2011; Ma and Hu, 2013) to balance continuous covariates as well as categorical. Investigation of such methods within the context of multi-arm trials for allocation methods is left for further research. Moreover, I assumed all covariates impacting the outcome data were included in the final analysis model. When there is one or more unobserved covariate correlated with treatment assignment, the estimate for the treatment effect can be biased (Liu and Hu, 2022). The effect of unobserved covariates on statistical inference under randomisation procedures in multi-arm trials is thus also an important issue for future study.

Furthermore, I considered statistical analysis based on a continuous outcome; results for other outcomes, such as binary or time to event, may differ. I also considered a statistical test based on population model assumptions. If randomisation-based inference was applied instead, this could affect our conclusions. Additionally, we only considered equal allocation ratios across all considered randomisation methods. The choice of randomisation procedure for a RCT with unequal allocation requires special additional considerations, such as the allocation ratio preserving property (Kuznetsova and Tymofyeyev, 2012). Finally adaptive changing of allocation ratios (i.e., (covariate) response adaptive randomisation) has been increasingly commonly considered within the context of multi-arm trials (Sverdlov, 2016; Park et al., 2020). Comparison of the methods discussed here to such adaptive randomisation methods is an additional potential topic for future work.

Chapter 3

Efficient stage-wise allocation ratios in multi-arm multi-stage designs

In this chapter, I determine stage-wise allocation ratio for MAMS designs that can help increase power, extending previous works that focus primarily on designs with a common allocation ratio across the trial. I start by first describing theory for MAMS designs that can incorporate stage-wise allocation ratios. I then explore combinations of common allocation ratios used in clinical practice.

3.1 Background

MAMS trials compare multiple experimental treatments simultaneously across multiple stages under a single trial. As was described earlier in the thesis in Subsection 1.4.1, these experimental treatments may comprise entirely different treatments, combinations of treatments, or different doses of the same treatment. MAMS designs can bring sizable gains in efficiency over conducting a series of single-stage two-armed trials (Grayling et al., 2018), and even compared to a single-stage multi-armed trial. Use of MAMS designs therefore may allow more treatments to be investigated when the potential number of patients to recruit is low (Wason et al., 2016) (see Subsection 1.4.1 for more details).

However, designing, implementing, and analysing MAMS trials presents unique challenges to traditional trials. MAMS trials are complex studies because of their involvement of multiple treatment arms, interim analyses, and even potential design adaptations based on accumulating data. Typically, MAMS studies are structured to have pre-specified interim analyses at specific time points (Jaki, 2015). For example, the first interim analysis may be when the response of interest has been observed from $n1$ patients, the second analysis planned when the response is available from $n2$ patients, and so on. At each of the interim analyses, a decision is made on which treatment arm(s) will continue and which will be stopped. This decision is based on pre-defined stopping rules based on test statistics for each treatment.

MAMS designs have many different sub-categories that differ mainly in the treatment selection mechanism used during the interim analyses (Wason et al., 2016). For example, a 'Pick-the-winner' design selects the most promising experimental treatment (at the first interim analysis) and compares it to the control arm in the subsequent stages (Stallard and Todd, 2003; Thall et al., 1988; Whitehead and Jaki, 2009). This may be extended to allow for more than one treatment to be continued after the first interim analysis, with instead a pre-determined number of treatment arms in each stage (Stallard and Friede,

2008). Kelly et al. (2005) recommended instead selecting all treatments that are 'similar' to the best performing one. By contrast, Royston et al. (2003) and Magirr et al. (2012) addressed study designs with several stages in which all treatments are continued at each stage provided that they are sufficiently promising. This class of design, often referred to as a group-sequential MAMS design, is the one that will be considered throughout this chapter.

Much work has now been done on MAMS designs. An aspect of designing MAMS trials that has not been exhaustively investigated to date is the choice of allocation ratio to the experimental arms, relative to the control arm, in each of the stages. As discussed previously in Subsection 1.5.4, allocation ratios have a considerable effect on the operating characteristics of a clinical trial, influencing, in particular, power, which in turn may impact the required sample size and overall trial cost (Dumville et al., 2006). The selection of an appropriate allocation ratio has thus received much attention in the literature within the context of fixed sample two-arm and multi-arm trials Sverdlov and Rosenberger (2013).

In two-arm trials where the outcome of interest has the same standard deviation in both arms, the optimal allocation ratio to maximise power is 1:1. However, this is not the case in single-stage multi-arm trials; the optimal allocation has been shown to be to approximately allocate the square root of the number of experimental arms to the control arm for each patient allocated to each experimental arm (Dunnett, 1955). This result is applicable under the assumptions of Dunnett's test of normally distributed data and homogeneity of variance to maximise marginal, disjunctive, and conjunctive powers (Neuhäuser et al., 2021). Beside maximising the power, this approach might be preferable in multi-arm trials because it reduces the chance that patients are allocated to potentially unsafe or toxic treatments compared to allocating equal numbers to all arms. The flip side is that this could negatively impact recruitment.

However, this square-root rule is not likely to be optimal in the case of a MAMS trial,

as experimental treatments can be stopped early in interim analyses. In the STAMPEDE trial (Sydes et al., 2009), it was decided to randomise two patients to the control arm for every one patient randomised to each experimental arm. The rationale for this was that the common control arm is used in every pairwise comparison, so higher precision provides greater power. Moreover, the effect of the allocation ratio can be seen in the correlation between pairs of test statistics that are used to test different treatments; Wason and Jaki (2012) showed that by increasing the allocation to the control arm, the correlation between the test statistics is reduced. This would increase statistical power because the control arm acts as a common reference group for all comparisons, and a larger control arm provides more data to estimate this reference, leading to potentially more similar estimates for each treatment compared to a smaller control group.

Furthermore, Wason and Jaki (2012) investigated the optimal allocation ratio in group-sequential MAMS trials, for various numbers of experimental arms and stages, as part of a wider search for optimal MAMS designs that minimise the expected sample size (ESS, the average number of patients that would be enrolled in the study). The authors found that in a trial with four experimental arms and three stages, the optimal allocation ratio to maximise the marginal power was approximately 1.33 patients to the control arm for each patient allocated in each experimental arm. They also found that the optimal allocation ratio increases when there are six experimental arms, but remains considerably below 2:1 (Wason et al., 2016).

The limitations of Wason and Jaki (2012) include the fact that the authors only considered one definition of power (marginal) and assumed that the same allocation ratio would be used at each stage. In addition, while they found that as the number of treatments increases the optimal allocation ratio also increased, a less clear pattern was reported on the effect of increasing the number of trial stages.

In practice, the allocation ratios could vary between stages, and this could be beneficial as it could reduce the pairwise correlation between test statistics at different stages.

Perhaps due to the complexity of assessing the operating characteristics of MAMS designs, no paper has yet explored identifying a design that maximises the marginal, disjunctive or conjunctive power.

The primary aim of this chapter is to address this question and to identify the most powerful stage-wise allocation in MAMS trials analytically. Furthermore, another study aim is to investigate how the stage-wise allocation ratio affects the ESS. I proceed by describing the type of MAMS design under consideration, including providing derivations of the expressions required to calculate their operating characteristics when allowing the stage-wise allocation ratios to vary. I then detail a motivating example, the TAILoR trial (Pushpakom et al., 2020), that is subsequently used in the investigation of the impact of the choice of a variety of stage-wise allocation ratios. I conclude with some recommendations based on my work.

3.2 Methods

3.2.1 Hypotheses and test statistics

I consider a MAMS trial that compares $K-1$ experimental treatment arms (indexed by $k = 1, \ldots, K-1$) against a common control arm ($k = 0$) through a maximum of $J \geq 2$ stages (such that there are K arms in the trial initially). I am interested in testing the efficacy of experimental treatments compared to the control, and thus I test $K-1$ one-sided hypotheses, given by $H_{0k} : \delta_k \leq 0$, $k = 1, \ldots, K-1$, where δ_k is a parameter that represents the difference in effect between experimental treatment k and the control. This difference may be defined in terms of, e.g., means if the outcome is normal, the log-odds ratio if the outcome is binary, or the log-hazard ratio if the outcome is time-to-event. For brevity, in what follows, I set $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_{K-1})^\top$ a vector of length $(K-1)$.

To test the hypothesis H_{0k}, after each stage $j = 1, \ldots, J$, I calculate standardised test statistics Z_{jk}, $k = 1, \ldots, K-1$, to compare the experimental arms k (still present in the trial) to the control arm, using all patients allocated up to analysis j. The test statistics, in general, depend on the type of outcome and the test statistic used. I restrict my attention to test statistics that follow a multivariate normal distribution. It should be noted that test statistics will be asymptotically normal when: i) the data is normally distributed, and the test statistic is either an unadjusted t-test or a t-test for the treatment effect after adjustment for covariates; ii) the data is binary, and the test statistic is either the unadjusted log odds ratio or the estimated log odds ratio from a logistic regression model with adjustments for covariates; iii) the data is time-to-event, and the test statistic is the log-rank test or the Cox model with adjustment for covariates (Jennison and Turnbull, 1999; Wason and Jaki, 2012; Wason and Mander, 2012).

For simplicity, I focus specifically on the case where $Y_{ijk} \sim N(\mu_k, \sigma^2)$ represents the primary outcome obtained from patient $i = 1, \ldots, n_{jk}$ in arm $k = 0, \ldots, K-1$

in stage $j = 1, \ldots, J$, with σ assumed to be known. Here, n_{jk} is the sample size of arm k in stage j. We now have $\delta_k = \mu_k - \mu_0$, with δ_k the mean difference between experimental treatment k and the control arm. I focus on the case where $n_{j0} = r_j n_{jk}$, $k = 1, \ldots, K-1$ (for arms k still present in the trial in stage j), such that there is a $r_j : 1$ stage-wise allocation ratio between the control arm and the present experimental arms in stage $j = 1, \ldots, J$. Note that with this, for simplicity and practical relevance, I have assumed that the sample size for all experimental arms still present in the trial is equal. Furthermore, this leaves $2J$ parameters to specify for all possible allocations in all stages to be known, without loss of generality $n_{10}, \ldots, n_{J0}$ and $r_1, \ldots, r_J$. I will write these as $\boldsymbol{n}_0 = (n_{10}, \ldots, n_{J0})^\top$ and $\boldsymbol{r} = (r_1, \ldots, r_J)^\top$ vectors of length J.

Now, we have

$$Z_{jk} = \frac{\bar{Y}_{jk} - \bar{Y}_{j0}}{\sigma\sqrt{\frac{1}{\tilde{n}_{j0}} + \frac{1}{\tilde{n}_{jk}}}} = \hat{\delta}_{jk} I_{jk}^{1/2},$$

with

$$\begin{aligned}
\bar{Y}_{jk} &= \frac{1}{\tilde{n}_{jk}} \sum_{l=1}^{j} \sum_{i=1}^{n_{lk}} Y_{ilk}, \\
\tilde{n}_{jk} &= n_{1k} + \cdots + n_{jk}, \\
\hat{\delta}_{jk} &= \bar{Y}_{jk} - \bar{Y}_{j0},
\end{aligned}$$

and I_{jk} the information for treatment k at stage j. I also define $\tilde{r}_j$ such that $\tilde{n}_{j0} = \tilde{r}_j \tilde{n}_{jk}$. Then, I can write

$$\begin{aligned}
I_{jk} &= \frac{1}{\sigma^2 \left(\frac{1}{\tilde{n}_{j0}} + \frac{1}{\tilde{n}_{jk}} \right)}, \\
&= \frac{1}{\sigma^2 \left(\frac{1}{\tilde{n}_{j0}} + \frac{\tilde{r}_j}{\tilde{n}_{j0}} \right)}, \\
&= \frac{\tilde{n}_{j0}}{\sigma^2 \left(1 + \tilde{r}_j\right)}.
\end{aligned}$$

It can be shown that the vector of test statistics $\boldsymbol{Z} = (Z_{11}, \ldots, Z_{J1}, \ldots, Z_{1(K-1)}, \ldots, Z_{J(K-1)})^\top$, a vector of length JK, is multivariate normal, with expectation and variance given by

$$\mathbb{E}(Z_{jk}) = \frac{\delta_k}{\sigma\sqrt{\frac{1}{\tilde{n}_{j0}} + \frac{1}{\tilde{n}_{jk}}}} = \delta_k I_{jk}^{1/2},$$
$$Cov(Z_{j_1k_1}, Z_{j_2k_2}) = \left(\frac{\tilde{n}_{j_10}}{\tilde{n}_{j_20}}\right)^{1/2} \frac{1 + \mathbb{I}(k_1 = k_2)\tilde{r}_{j_2}}{(1 + \tilde{r}_{j_1})^{1/2}(1 + \tilde{r}_{j_2})^{1/2}},\ 1 \leq j_1 \leq j_2 \leq J,$$

where $\mathbb{I}(X)$ is the indicator function on event X. A derivation of this covariance structure is given at the end of this chapter in Supplementary Materials 3.6.

3.2.2 Stopping rules

I consider a MAMS trial with both interim efficacy and futility monitoring, denoting the stopping boundaries for the various stages by $\boldsymbol{e} = (e_1, \ldots, e_J)^\top \in \mathbb{R}^J$ and $\boldsymbol{f} = (f_1, \ldots, f_J)^\top \in \mathbb{R}^J$, with $f_j < e_j$ for $j = 1, \ldots, J-1$ and $f_J = e_J$. Note that there are two broad sub-types of MAMS stopping rule: one terminates the entire trial as soon as any null hypothesis is rejected (simultaneous stopping) and the other continues until test statistics for all experimental arms have all crossed either an efficacy or a futility bound (separate stopping) (Urach and Posch, 2016). In this chapter, I consider only separate stopping. Thus, the decision rules are as follows. After stage $j = 1, \ldots, J$

- Denote the set of experimental arms still present in the trial at stage j by $\mathbb{K}_j \subseteq \{1, \ldots, K-1\}$. It is the case that $k \in \mathbb{K}_j$ if $Z_{1k} \in [f_1, e_1), \ldots, Z_{j-1k} \in [f_{j-1}, e_{j-1})$.
- For $k \in \mathbb{K}_j$
  - If $Z_{jk} < f_j$ drop arm k from the trial for futility.
  - Else if $Z_{jk} \geq e_j$ then reject H_{0k} and drop arm k from the trial for efficacy.
  - Else add k to $\mathbb{K}_{j+1}$.

- If $\mathbb{K}_{j+1} \neq \emptyset$ then proceed to stage $j+1$, retaining the control arm and those arms $k \in \mathbb{K}_{j+1}$.

3.2.3 Operating characteristics

I will be interested in the impact the choice of stage-wise allocation ratios has on several statistical quantities. The first of these will be the family-wise error rate (FWER), defined as the probability of incorrectly rejecting at least one true null hypothesis. It is common in MAMS designs, at least in a confirmatory setting, to control the FWER.

Next, as this is a multi-arm setting, there are several possible definitions of power that may be of interest depending on the research question that the trial is designed to address. Most typically these are the marginal, disjunctive, and conjunctive powers. As was described in the previous chapter, the marginal power, $P_k(\delta_k)$, is the probability of rejecting the null hypothesis H_{0k}. The disjunctive power, $P_{\text{dis}}(\delta_1, \ldots, \delta_{K-1})$, is the probability of rejecting at least one null hypothesis $H_{0,1}, \ldots, H_{0,K-1}$, whereas the conjunctive power, $P_{\text{con}}(\delta_1, \ldots, \delta_{K-1})$, is the probability of rejecting all of the null hypotheses $H_{0,1}, \ldots, H_{0,K-1}$. All these powers are conditional on the treatment effects being non-zero ($\delta_k \neq 0$). A comprehensive review of power definitions in multi-arm trials is given in Dmitrienko et al. (2009).

Finally, I will be interested in the ESSs of various designs, as this is the metric most often used to evaluate how efficient a MAMS design is. Next, I therefore describe how each of these statistical quantities can be computed.

I utilise the notation from Grayling et al. (2017). Specifically, I define two vectors of random variables that summarise the unknown study outcome. I use $\boldsymbol{\Omega} = (\Omega_1, \ldots, \Omega_{K-1})^\top$ to denote the final stage at which each experimental treatment was present, i.e., $\Omega_k = \omega_k \in \{1, \ldots, J\}$ if treatment k was present in the trial up to and including stage ω_k. Further, $\boldsymbol{\Psi} = (\Psi_1, \ldots, \Psi_{K-1})$ is used to define which null hypotheses are rejected. Explicitly, $\Psi_k = \psi_k = 1$ if H_{0k} is rejected and $\Psi_k = \psi_k = 0$ otherwise.

As an example, if $J = K = 2$ and the outcome is that treatment $k = 1$ is stopped for futility at the interim analysis, and treatment 2 has its null rejected at the final analysis, the random vectors would be observed to take the values $\boldsymbol{\Omega} = \boldsymbol{\omega} = (1, 2)^\top$ and $\boldsymbol{\Psi} = \boldsymbol{\psi} = (0, 1)^\top$. All possible values of $\{\boldsymbol{\Omega}, \boldsymbol{\Psi}\}$ can then be defined as

$$\Xi = \left\{ \{\boldsymbol{\omega}, \boldsymbol{\psi}\} : \boldsymbol{\omega} \in \{1, \ldots, J\}^{K-1}, \boldsymbol{\psi} \in \{0, 1\}^{K-1} \right\}.$$

Several subsets of Ξ will be of use to us in computing the operating characteristics of interest. These define when a specific null hypothesis (H_{0k}) is rejected (Ξ_k), when at least one null hypothesis is rejected (Ξ_{dis}), and when all of the null hypotheses are rejected (Ξ_{con}). They are given by

$$\begin{aligned}
\Xi_k &= \{\{\boldsymbol{\omega}, \boldsymbol{\psi}\} \in \Xi : \psi_k = 1\}, \\
\Xi_{\text{dis}} &= \left\{ \{\boldsymbol{\omega}, \boldsymbol{\psi}\} \in \Xi : \sum_{k=1}^{K-1} \psi_k > 0 \right\}, \\
\Xi_{\text{con}} &= \{\{\boldsymbol{\omega}, \boldsymbol{\psi}\} \in \Xi : \psi_1 = \cdots = \psi_{K-1} = 1\}.
\end{aligned}$$

Next, I write the probability of any outcome $\{\boldsymbol{\Omega}, \boldsymbol{\Psi}\} = \{\boldsymbol{\omega}, \boldsymbol{\psi}\}$, conditional on the choices of $\boldsymbol{n}_0$, $\boldsymbol{r}$, $\boldsymbol{e}$, $\boldsymbol{f}$ and the treatment effects $\boldsymbol{\delta}$, as $\mathbb{P}(\boldsymbol{\omega}, \boldsymbol{\psi} | \boldsymbol{n}_0, \boldsymbol{r}, \boldsymbol{e}, \boldsymbol{f}, \boldsymbol{\delta})$. Furthermore, I denote the sample size required by the trial when $\{\boldsymbol{\Omega}, \boldsymbol{\Psi}\} = \{\boldsymbol{\omega}, \boldsymbol{\psi}\}$ by $n(\boldsymbol{\omega})$ (as this is independent of $\boldsymbol{\psi}$). I have

$$\begin{aligned}
n(\boldsymbol{\omega}) &= \sum_{j=1}^{\max\{\omega_1, \ldots, \omega_{K-1}\}} n_{j0} + \sum_{k=1}^{K-1} \sum_{j=1}^{\omega_k} n_{jk}, \\
&= \tilde{n}_{\max\{\omega_1, \ldots, \omega_{K-1}\}0} + \sum_{k=1}^{K-1} \frac{\tilde{n}_{\omega_k 0}}{\tilde{r}_{\omega_k}}.
\end{aligned}$$

Using the above, I can determine the marginal power for H_{0k}, the disjunctive power,

the conjunctive power, and the ESS of a design as

$$\left.\begin{aligned} P_k(\boldsymbol{n}_0, \boldsymbol{r}, \boldsymbol{e}, \boldsymbol{f}, \boldsymbol{\delta}) &= \sum_{\{\boldsymbol{\omega},\boldsymbol{\psi}\}\in\Xi_k} \mathbb{P}(\boldsymbol{\omega}, \boldsymbol{\psi}|\boldsymbol{n}_0, \boldsymbol{r}, \boldsymbol{e}, \boldsymbol{f}, \boldsymbol{\delta}), \\ P_{\text{dis}}(\boldsymbol{n}_0, \boldsymbol{r}, \boldsymbol{e}, \boldsymbol{f}, \boldsymbol{\delta}) &= \sum_{\{\boldsymbol{\omega},\boldsymbol{\psi}\}\in\Xi_{\text{dis}}} \mathbb{P}(\boldsymbol{\omega}, \boldsymbol{\psi}|\boldsymbol{n}_0, \boldsymbol{r}, \boldsymbol{e}, \boldsymbol{f}, \boldsymbol{\delta}), \\ P_{\text{con}}(\boldsymbol{n}_0, \boldsymbol{r}, \boldsymbol{e}, \boldsymbol{f}, \boldsymbol{\delta}) &= \sum_{\{\boldsymbol{\omega},\boldsymbol{\psi}\}\in\Xi_{\text{con}}} \mathbb{P}(\boldsymbol{\omega}, \boldsymbol{\psi}|\boldsymbol{n}_0, \boldsymbol{r}, \boldsymbol{e}, \boldsymbol{f}, \boldsymbol{\delta}), \\ ESS(\boldsymbol{n}_0, \boldsymbol{r}, \boldsymbol{e}, \boldsymbol{f}, \boldsymbol{\delta}) &= \sum_{\{\boldsymbol{\omega},\boldsymbol{\psi}\}\in\Xi} \mathbb{P}(\boldsymbol{\omega}, \boldsymbol{\psi}|\boldsymbol{n}_0, \boldsymbol{r}, \boldsymbol{e}, \boldsymbol{f}, \boldsymbol{\delta}) n(\boldsymbol{\omega}). \end{aligned}\right\} \quad \delta > 0$$

Using the result from Magirr et al. (2012) on strong control of the FWER in MAMS trials, I can also compute the maximal FWER of a design as $\text{FWER}(\boldsymbol{n}_0, \boldsymbol{r}, \boldsymbol{e}, \boldsymbol{f}) = P_{\text{dis}}(\boldsymbol{n}_0, \boldsymbol{r}, \boldsymbol{e}, \boldsymbol{f}, \boldsymbol{0})$.

In practice, to power the trial, it is common to consider power under the least favourable configuration (LFC), as defined by Dunnett (1955). Without loss of generality, the LFC for treatment $k = 1$ may be assumed. To define the LFC, two thresholds are specified: a clinically interesting effect, $\theta_1 > 0$, and an uninteresting effect, $\theta_0 < \theta_1$. The LFC for $k = 1$ is then given by when $\delta_1 = \theta_1$ and $\delta_2 = \cdots = \delta_{K-1} = \theta_0$. The marginal power for H_{01} under its LFC is $P_1\{\boldsymbol{n}_0, \boldsymbol{r}, \boldsymbol{e}, \boldsymbol{f}, (\theta_1, \theta_0, \ldots, \theta_0)^\top\}$. The sample size (and allocation ratios) are then generally chosen such that $P_1\{\boldsymbol{n}_0, \boldsymbol{r}, \boldsymbol{e}, \boldsymbol{f}, (\theta_1, \theta_0, \ldots, \theta_0)^\top\} \geq 1 - \beta$ for some specified β.

All that remains is to describe how the probabilities $\mathbb{P}(\boldsymbol{\omega}, \boldsymbol{\psi}|\boldsymbol{n}_0, \boldsymbol{r}, \boldsymbol{e}, \boldsymbol{f}, \boldsymbol{\delta})$ can be determined. In brief, these may be computed easily and efficiently using the method of Genz and Bretz (2002) (see, e.g., Grayling et al. (2017) for details) or via simulation (see, e.g., Wason et al. (2016) for details). There is little practical difference between their relative merits (i.e., speed and precision) for small values of J and K.

3.3 Motivating Example: TAILoR

As motivation for many of the assumed parameter values, I use the TAILoR trial (Pushpakom et al., 2020). The TAILoR trial was originally designed to utilise a two-stage design ($J = 2$) to test four different doses of Telmisartan ($K = 5$), though in the actual trial only three different doses ($K = 4$) were included. The primary objective was to study the effects of Telmisartan, an angiotensin II receptor blocker, at 20, 40 and 80 mg dosages on the homeostasis model of insulin resistance (HOMA-IR) at 24 weeks compared to the common control arm. The maximum allowed sample size was set as 376 patients, with the interim analysis planned after 50% of these patients had observed data. A one-sided FWER of 5% was specified, while a 1:1:1:1 allocation was used at each stage to randomise patients between treatment arms (i.e., $r_1 = r_2 = 1$). The stopping boundaries were set as $\boldsymbol{f} = (0, 2.086)^\top$ and $\boldsymbol{e} = (2.782, 2.086)^\top$. Finally, the "interesting" treatment effect assumed in the power calculations was $\theta_1 = 0.545$. Using the above, I have examined the impact of the choice of stage-wise allocation ratios in two scenarios:

- Scenario 1: Sets $K = 4$, $J = 2$, $\max N = 376$, $\sigma = 1$, $\boldsymbol{f} = (0, 2.086)^\top$, $\boldsymbol{e} = (2.782, 2.086)^\top$, and $\theta_1 = 0.545$. I then consider multiple combinations of allocation ratios, specifically $(r_1, r_2) \in \{1, 1.\dot{3}, 1.5, \sqrt{3}, 2, 2.5, 3, 4\}^2$.

- Scenario 2: Sets $K = 5$, $J = 3$, and $\sigma = 1$. It then sets the treatment effects, maximum allowed sample size, and stopping boundaries based on those derived in Magirr et al. (2012) for these assumptions. That is, it sets $\max N = 465$, $\boldsymbol{f} = (0, 0, 2.182)^\top$, $\boldsymbol{e} = (3.779, 2.672, 2.182)^\top$, and $\theta_1 = 0.65$. I then consider $(r_1, r_2, r_3) \in \{1, 1.\dot{3}, 1.5, 2, 2.5\}^3$.

For each scenario, I divide the maximum allowed sample size equally between the stages, such that $n_j = \max N/J$. Then, the allocation to the control arm in stage j is set as $n_{0j} = r_j n_j / \{r_j + (K - 1)\}$.

Note that the choice of stage-wise allocation ratios has an impact on the FWER, for fixed values of $\boldsymbol{n}_0$, $\boldsymbol{e}$, and $\boldsymbol{f}$. For a fair comparison, I ensured all designs had the same FWER by modifying the values of $\boldsymbol{e}$ and $\boldsymbol{f}$ stated above. Specifically, for each considered $\boldsymbol{r}$ I searched for the value of c that would ensure that $\text{FWER}(\boldsymbol{n}_0, \boldsymbol{r}, c\boldsymbol{e}, c\boldsymbol{f}) = \alpha = 0.05$.

For each $\boldsymbol{r}$ considered in each scenario, I analytically compute the marginal, disjunctive, conjunctive power, and ESS when $\boldsymbol{\delta} = (\theta_1, \ldots, \theta_1)^\top$ (henceforth referred to as the global alternative hypothesis). I also calculated the ESS under the global null, given by $\boldsymbol{\delta} = \mathbf{0}$.

3.4 Results

Figure 3.1 displays the computed powers for Scenario 1, with Figure 3.2 giving the corresponding results for Scenario 2. Figures 3.3 and 3.4 give the ESSs under the global alternative hypothesis for Scenarios 1 and 2 respectively; findings were similar for the global null hypothesis.

3.4.1 Marginal power

In Scenario 1, the highest marginal power was approximately 76%, achieved for a range of r_1 across 1 to 2, with a range of r_2 over 1.33 to 2. In contrast, $r_1 = r_2 = 4$ generated the lowest observed marginal power of approximately 68%.

In Scenario 2, the highest marginal power was approximately 72%, achieved for $r_1 = 1$ with $r_2 \in \{1.\dot{3}, 2\}$ and $r_3 = 2.5$. Allocation ratios of $r_1 \in \{2, 2.5\}$ with $r_2 = r_3 = 1$ provided the lowest observed marginal power of 67%.

3.4.2 Disjunctive power

In Scenario 1, the highest disjunctive power of approximately 96% was achieved using $r_1 = 4$ with $r_2 = 2$. Interestingly, $r_1 = r_2 = 1$ resulted in the lowest disjunctive power of 92%.

In Scenario 2, the highest disjunctive power of 97% was achieved when $r_1 = r_2 = r_3 = 2.5$. Whilst, as for Scenario 1, the use of $r_1 = r_2 = r_3 = 1$ generated the lowest disjunctive power.

These findings indicate that, almost universally, increasing the number of patients in the control arm increased the disjunctive power across the two considered scenarios.

3.4.3 Conjunctive power

In Scenario 1, the highest conjunctive power of 54% was achieved using $(r_1, r_2) = (1, 1.\dot{3})$. Using $(r_1, r_2) = (4, 4)$ resulted in the lowest conjunctive power observed, which was approximately 37%.

In contrast, the highest conjunctive power in Scenario 2 was generated using $r_1 = r_2 = 1$, with $r_3 = 1.\dot{3}$. Whilst using an allocation of $r_1 = r_2 = 2.5$ with $r_3 = 1$ provided the lowest conjunctive power.

Overall, these results demonstrate that allocating more patients to the control arm caused a reduction in the conjunctive power.

3.4.4 Expected sample size

As expected, alteration of the stage-wise allocation ratios is associated with variation in the ESS. The results illustrate that lower ESSs were achieved by setting $r_1 = 2$ for various allocation ratios in the subsequent stage(s). Notably, the ESSs were higher for $r_1 = 1$ in both scenarios, which is significant given many trials use equal allocation in practice.

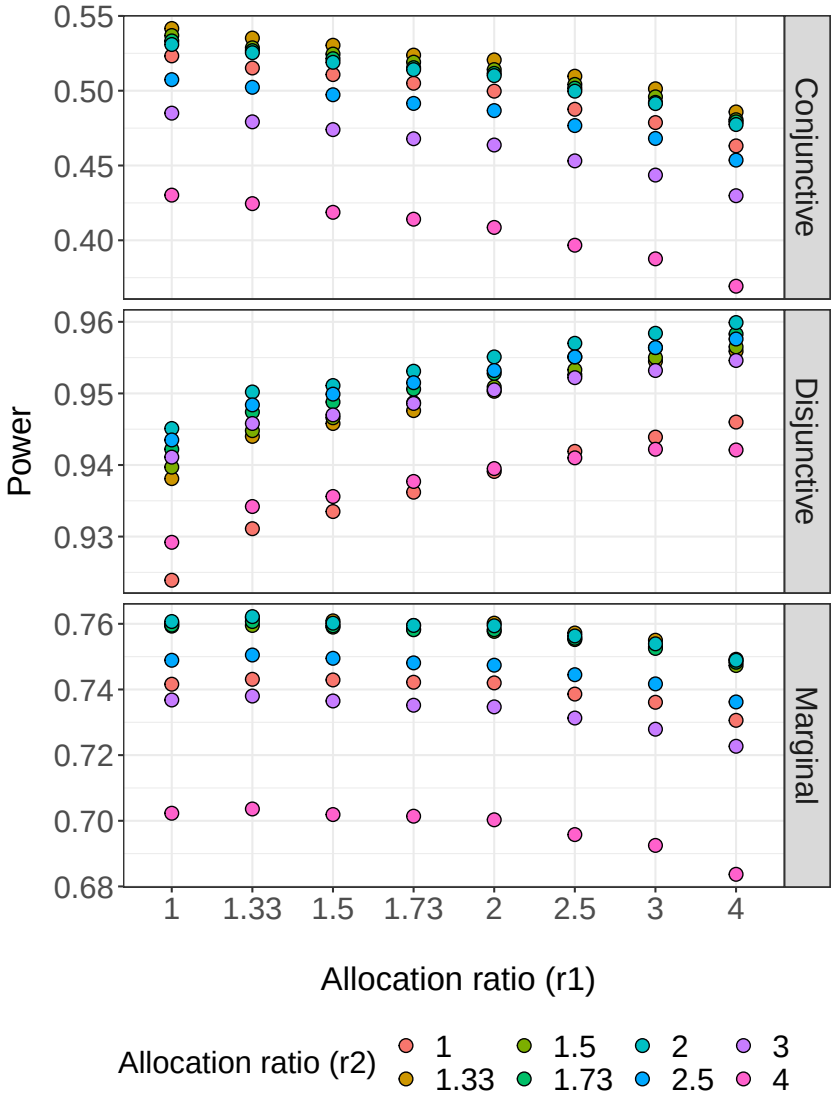

Figure 3.1: The empirical marginal, disjunctive, and conjunctive powers are shown for Scenario 1, as a function of the stage-wise allocation ratios r_1 and r_2.

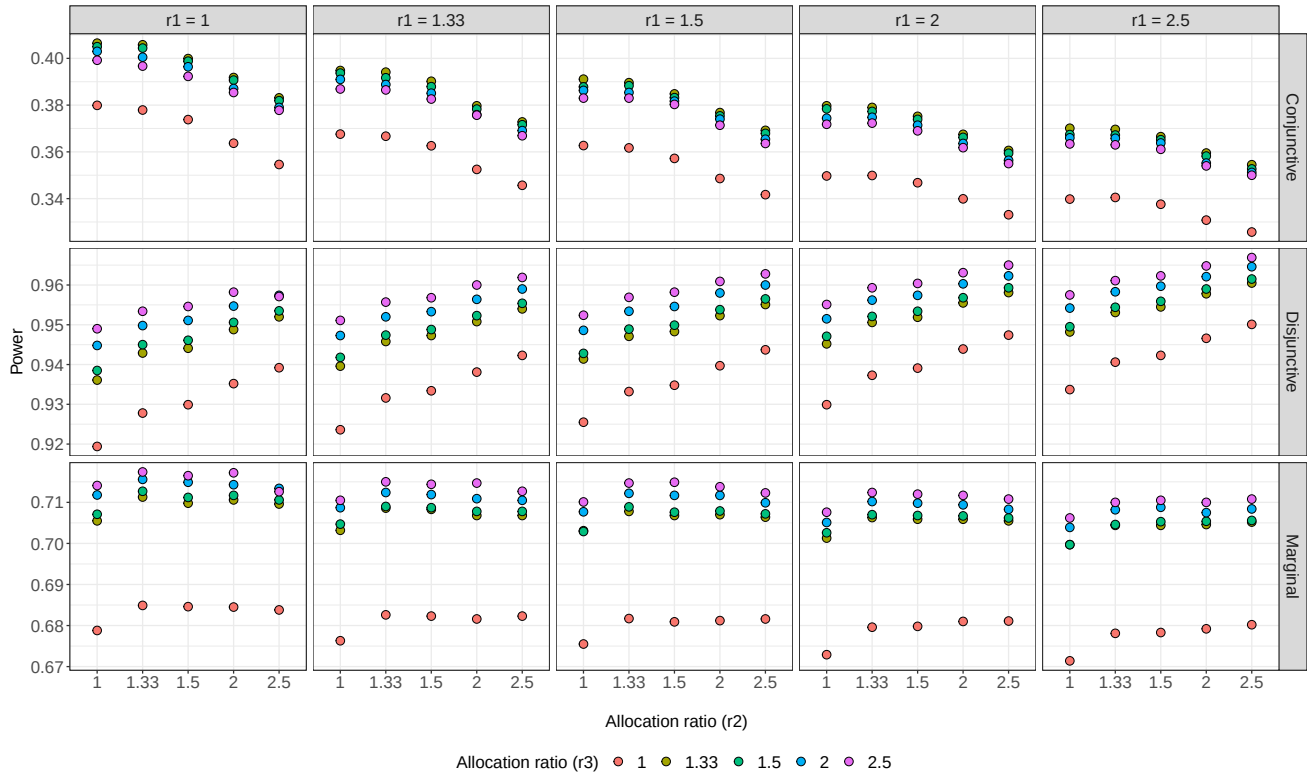

Figure 3.2: The empirical marginal, disjunctive, and conjunctive powers are shown for Scenario 2, as a function of the stage-wise allocation ratios r_1, r_2, and r_3.

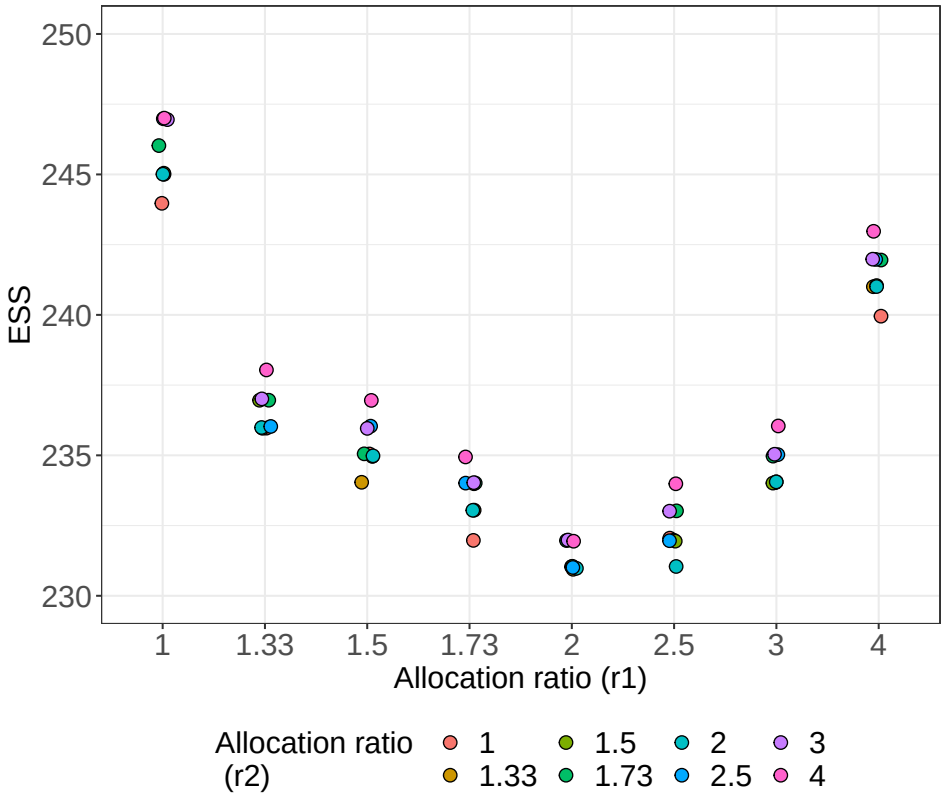

Figure 3.3: The empirical expected sample sizes under the global alternative hypothesis are shown for Scenario 1, as a function of the stage-wise allocation ratios r_1 and r_2.

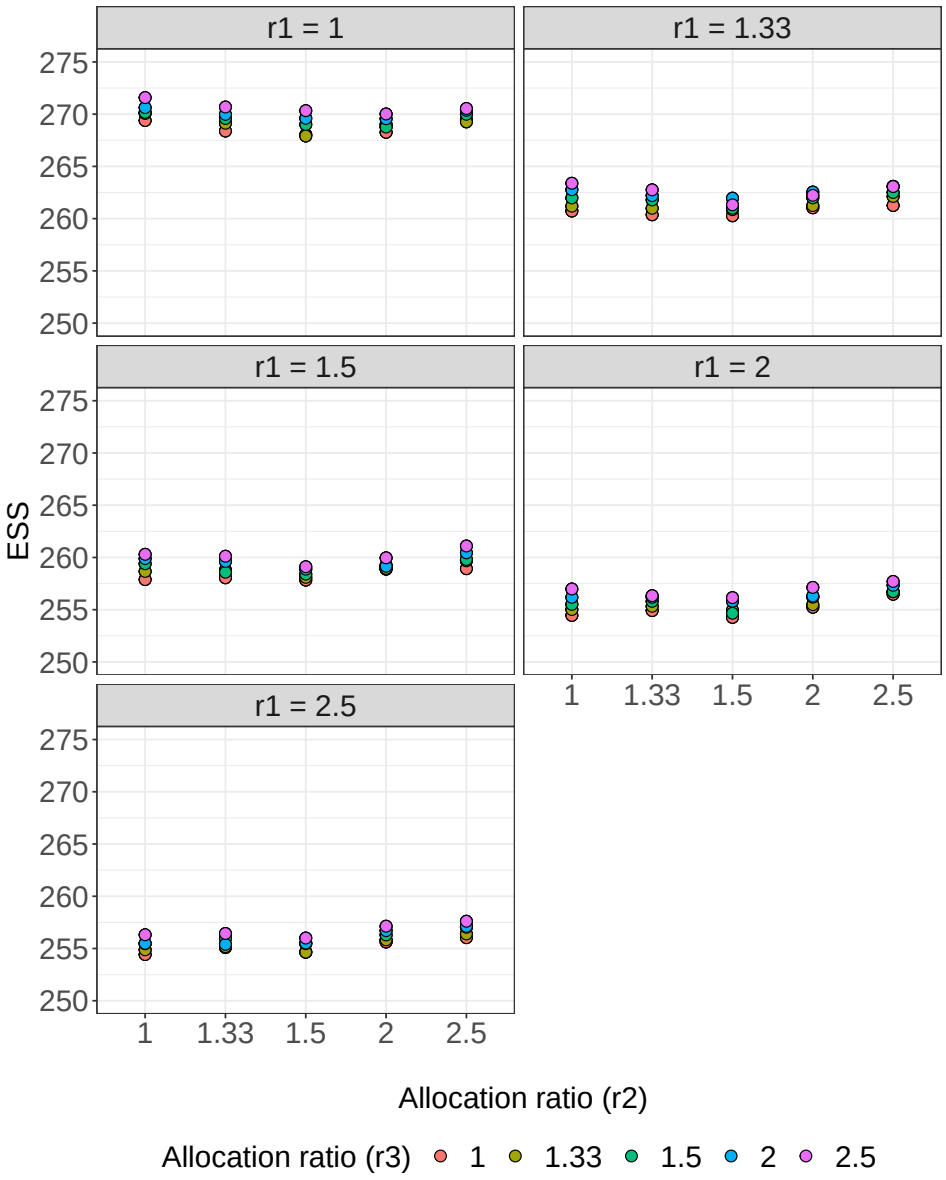

Figure 3.4: The empirical expected sample sizes under the global alternative hypothesis are shown for Scenario 2, as a function of the stage-wise allocation ratios r_1, r_2, and r_3.

3.5 Discussion

MAMS designs are a broad class of adaptive design that can improve the efficiency of clinical trials. In this study, I aimed to examine the effect on MAMS designs of modifying their stage-wise allocation ratios, examining in particular the impact on statistical power, to identify efficient stage-wise allocation ratios that can increase the marginal, disjunctive, or conjunctive power. Notably, modifying allocation ratios of MAMS at each stage increases the chance of FWER compared to use same allocation ratio in each stage. Therefore, I considered controlling the FWER at 0.05 for each allocation ratio (see Appendix A.2). Simultaneously, I determined the effect of various allocation ratios on the ESS, in all cases controlling the FWER to the desired level.

The results demonstrate differential effects of how the stage-wise allocation ratios impact the different types of power. In both scenarios, it was clear efficient allocation ratios for large conjunctive power were associated with a more balanced allocation between the control and experimental arms. In contrast, the disjunctive power was generally increased when more patients were allocated to the control arm vs. the experimental arms. However, exceedingly large allocation ratios did have a negative impact on the disjunctive power. These results reflect what has previously been demonstrated for how the disjunctive and conjunctive power in trials with multiple endpoints are impacted by the correlation between these outcomes (Senn and Bretz, 2007). Specifically, as the correlation between endpoints increases, the disjunctive power decreases, while the conjunctive power increases. As shown in Wason and Jaki (2012) increased control arm allocation reduces the correlation between test statistics.

Furthermore, the efficient stage-wise allocation ratios for marginal power seemingly increased in each stage. For example, in Scenario 2, the allocation ratios that resulted in maximal marginal power were similar in stages 1 and 2, but increased in stage 3. The results also indicate that there is a range of allocation ratios that provide similar power. For

example, in Scenario 1, the marginal power is similar for $(r_1, r_2) \in \{1.33, 1.5, 1.43, 2\}^2$, whereas the disjunctive power is similar for $(r_1, r_2) \in \{1.33, 1.5, 3\}^2$ and the conjunctive power for $(r_1, r_2) \in \{1.5, 1.73, 2, 2.5\}^2$.

Although the results presented indicate that the allocation ratios can have a notable impact on power, the ESS variation was relatively small. For example, Scenario 1 reported a notable possible increase in disjunctive power from 90% to 96% through the implementation of efficient stage-wise allocation ratios. However, the ESS for 90% power was 244, whereas for 96% power, it was 241, resulting in a reduction in the expected sample size of 1.23%. This reduction was calculated by dividing the difference in ESS between the two conditions (96% and 90% power) by the ESS for 90% power. This is consistent with previous papers about optimal multi-stage designs (Wason and Jaki, 2012). Thus, I suggest that modifying the allocation ratio so that there is an increased number of patients allocated to the control group can affect the power of a study without a meaningful increase to the required sample size.

Before concluding, it is important to recognise several limitations of this work. In this study, I only considered group-sequential stopping rules for selecting which treatments to retain for subsequent stages. The retained arms were then used in combination with pre-specified stage-wise allocation ratios. There is another, arguably more direct, approach to altering the allocation ratio in subsequent stages depending on the observed treatment outcomes; this method is known as RAR. RAR may be more effective than a MAMS design, even one with carefully chosen stage-wise allocation ratios, at increasing power under a range of potential treatment effects (Wason and Trippa, 2014). RAR is, however, more complex in practice than the MAMS designs considered here.

Furthermore, I do not consider dropping out of participants. These factors may influence the optimal allocation ratio and the required sample size. If there are variations in dropout rates between treatment arms, the intended allocation ratio might deviate from the desired one, affecting the desired statistical power and raising concerns about

potential biases. Several researchers, therefore, discuss this point (e.g., Möcks et al. (2002); Greene (2015)) to mitigate the impact of dropout and early termination on the study design. For example, investigators could consider adjusting the intended allocation ratio to account for the expected rate of dropout, ensuring that the desired allocation ratio between treatment arms is maintained throughout the trial. This adjustment may involve enrolling more patients in the expectation that some will drop out before reaching the analysis stage.

Another limitation is that I focused on optimising the power given a fixed sample size. This would be equivalent to minimising the sample size for a fixed power but may give different results to other optimisation criteria that are used in trials. Additionally, I focused on group-sequential MAMS designs which use pre-specified stopping boundaries. The optimal allocation ratio may be different for other MAMS designs, such as the drop-the-losers design.

Moreover, I only focused on the situation where the outcome is normally distributed, with known variance, and the outcome is the same at each analysis. These results should readily extend to alternative endpoint types, as asymptotically normally distributed test statistics are often used for binary and survival outcomes. However, many MAMS trials being carried out do use different outcomes at each stage (Jaki, 2015). For example, the outcome in the first stage could be binary, while the outcome in the second stage could be time-to-event. Further work exploring efficient stage-wise allocation ratios when different outcomes are used in each stage would be of interest. Additionally, I have assumed that the interim analyses are equally spaced in terms of the number of patient recruited in each stage. Further work is required to identify the most efficient stage-wise allocation ratios to maximise power for unequally spaced interim analyses.

3.6 Supplementary Materials: Test statistics covariance structure

Here, I describe how to drive the covariance formula for the test statistics between experimental arms at the same stage and across different stages. Any notation not defined here has been given in Subsection 3.2.

Set $\boldsymbol{Z}_j = (Z_{j1}, \ldots, Z_{j(K-1)})^\top$, $\boldsymbol{I}_j = (I_{j1}, \ldots, I_{j(K-1)})^\top$, and $\hat{\boldsymbol{\delta}}_j = (\hat{\delta}_{j1}, \ldots, \hat{\delta}_{j(K-1)})^\top$. I use the 'Hadamard product' of two vectors, $\circ$, defined as

$$(a_1, \ldots, a_n) \circ (b_1, \ldots, b_n) = (a_1 b_1, \ldots, a_n b_n),$$

and the 'Hadamard square-root' of a vector, $^{\circ 1/2}$, defined as

$$(a_1, \ldots, a_n)^{\circ 1/2} = (a_1^{1/2}, \ldots, a_n^{1/2}).$$

Finally, I set

$$Diag\{(a_1, \ldots, a_n)^\top\} = \begin{pmatrix} a_1 & 0 & 0 & \cdots & 0 \\ 0 & a_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & a_n \end{pmatrix}.$$

Then, for $1 \leq j_1 \leq j_2 \leq J$

$$\begin{aligned}
Cov(\boldsymbol{Z}_{j_1}, \boldsymbol{Z}_{j_2}) &= Cov\{(\hat{\delta}_{j_1 1} I_{j_1 1}^{1/2}, \ldots, \hat{\delta}_{j_1(K-1)} I_{j_1(K-1)}^{1/2})^\top, (\hat{\delta}_{j_2 1} I_{j_2 1}^{1/2}, \ldots, \hat{\delta}_{j_2(K-1)} I_{j_2(K-1)}^{1/2})^\top\}, \\
&= Cov\{(\hat{\delta}_{j_1 1}, \ldots, \hat{\delta}_{j_1(K-1)})^\top \circ (I_{j_1 1}^{1/2}, \ldots, I_{j_1(K-1)}^{1/2})^\top, \\
&\qquad\qquad (\hat{\delta}_{j_2 1}, \ldots, \hat{\delta}_{j_2(K-1)})^\top \circ (I_{j_2 1}^{1/2}, \ldots, I_{j_2(K-1)}^{1/2})^\top\}, \\
&= Cov(\hat{\boldsymbol{\delta}}_{j_1} \circ \boldsymbol{I}_{j_1}^{\circ 1/2}, \hat{\boldsymbol{\delta}}_{j_2} \circ \boldsymbol{I}_{j_2}^{\circ 1/2}), \\
&= Diag(\boldsymbol{I}_{j_1}^{\circ 1/2}) Cov(\hat{\boldsymbol{\delta}}_{j_1}, \hat{\boldsymbol{\delta}}_{j_2}) Diag(\boldsymbol{I}_{j_2}^{\circ 1/2}), \\
&= Diag(\boldsymbol{I}_{j_1}^{\circ 1/2}) Var(\hat{\boldsymbol{\delta}}_{j_2}) Diag(\boldsymbol{I}_{j_2}^{\circ 1/2}).
\end{aligned}$$

Note that $Cov(\hat{\boldsymbol{\delta}}_{j_1}, \hat{\boldsymbol{\delta}}_{j_2}) = Var(\hat{\boldsymbol{\delta}}_{j_2})$. This is a fundamental result in group sequential design theory. It can be found in the book (Jennison and Turnbull, 1999) as Equation (11.1) with a derivation for normal linear models given in Theorem 11.1.
This tells us that $Cov(Z_{j_1 k_1}, Z_{j_2 k_2}) = I_{j_1 k_1}^{1/2} Cov(\hat{\delta}_{j_2 k_1}, \hat{\delta}_{j_2 k_2}) I_{j_2 k_2}^{1/2}$. Note that

$$\begin{aligned}
Cov(\hat{\delta}_{jk_1}, \hat{\delta}_{jk_2}) &= \frac{\sigma^2}{\tilde{n}_{j0}} + \mathbb{I}(k_1 = k_2) \frac{\sigma^2}{\tilde{n}_{jk_1}}, \\
&= \frac{\sigma^2}{\tilde{n}_{j0}} \left\{1 + \mathbb{I}(k_1 = k_2) \tilde{r}_j\right\}.
\end{aligned}$$

Thus

$$\begin{aligned}
Cov(Z_{j_1 k_1}, Z_{j_2 k_2}) &= I_{j_1 k_1}^{1/2} Cov(\hat{\delta}_{j_2 k_1}, \hat{\delta}_{j_2 k_2}) I_{j_2 k_2}^{1/2}, \\
&= \left\{\frac{\tilde{n}_{j_1 0}}{\sigma^2 (\tilde{r}_{j_1} + 1)}\right\}^{1/2} \left(\frac{\sigma^2}{\tilde{n}_{j_2 0}}\right) \{1 + \mathbb{I}(k_1 = k_2) \tilde{r}_{j_2}\} \left\{\frac{\tilde{n}_{j_2 0}}{\sigma^2 (\tilde{r}_{j_2} + 1)}\right\}^{1/2}, \\
&= \left(\frac{\tilde{n}_{j_1 0}}{\tilde{n}_{j_2 0}}\right)^{1/2} \frac{1 + \mathbb{I}(k_1 = k_2) \tilde{r}_{j_2}}{(1 + \tilde{r}_{j_1})^{1/2} (1 + \tilde{r}_{j_2})^{1/2}},
\end{aligned}$$

as stated earlier.

Chapter 4

Randomisation methods in a multi-arm multi-stage platform trial design

In Chapter 2, it was clearly laid out how numerous randomisation methods were evaluated in the context of multi-arm clinical trials, where all treatment arms begin simultaneously. In this chapter, I evaluate the performance of several of the same randomisation methodologies (e.g., SR, SBR, SBUD, and Mini) in the context of a platform trial design in which a new experimental arm is introduced to the trial at a pre-specified time point. Particular focus is paid to performance under varying target allocation ratios after the addition of the new arm.

4.1 Background

Platform trials assess multiple experimental treatments for a particular indication continuously; these trials can persist until a treatment with the desired risk/benefit profile is identified (Woodcock and LaVange, 2017). That is, the key feature of platform trials is their ongoing nature in which new treatment arm(s) can be added to the trial as they become available, while ineffective or harmful ones can be removed. This continuous evaluation process allows for quicker identification of promising treatments. Such trial designs have substantial advantages over traditional trial designs (Saville and Berry, 2016), as detailed in Section 1.4.2. Recently, there has been an increasing interest in platform trials, leading to significant methodological advancements in trial designs within a relatively short period, particularly during the COVID-19 pandemic. During this time, Vanderbeek et al. (2022) identified 58 platform trials registered between January 2020 and May 2021 globally.

Platform trial designs can be categorised into two main types based on how the treatment arms are introduced (Park et al., 2019): *pre-planned platform trials* and *unplanned (adaptive) platform trials*. In pre-planned platform trials, the number of treatment arms and their characteristics are specified before the trial starts. These trials often involve specific treatment comparisons planned based on existing research or hypotheses. In such trials, the timing of introducing each arm might also be pre-defined. In adaptive platform trials, however, new arm(s) can be added or dropped at any time point. This allows for a more flexible and dynamic approach, allowing investigators to adapt the trial based on emerging findings (The Adaptive Platform Trial Coalition, 2019; Park et al., 2020).

Although platform trials offer considerable advantages, careful planning and execution are crucial for their success. Consequently, several authors have discussed some of the issues of introducing new treatment arm(s) to an ongoing trial using different

optimality criteria and statistical analysis procedures. For example, Elm et al. (2012) evaluated the operating characteristics of pairwise comparisons in trials that added a new arm. The authors focus on the family-wise type I error rate, power, sample size, and choice of analytical methods (linear model and pooled data methods). Cohen et al. (2015) summarised the literature regarding statistical methods and design considerations when adapting a trial by adding a new treatment midway through recruitment and investigated trials that have added an arm in practice and how they addressed the statistical considerations. Choodari-Oskooei et al. (2020) evaluated the type 1 error rate when introducing new arm(s). Furthermore, Bennett and Mander (2020) determined optimal allocation ratios for platform trials, in terms of minimising the total sample size and achieving a desirable marginal power, when not adjusting for potential time trends. Ren et al. (2021) described statistical considerations with respect to type 1 error and power in three-arm umbrella trials that add in an arm. The authors also discussed the optimal allocation ratio for the control arm in periods in which treatment arms overlap, in terms of minimising the sum of variances of the treatment effect estimators. Finally, Bofill Roig et al. (2023) investigated the optimal allocation ratios in terms of minimising the maximum standard error of the treatment effect estimators.

Although the current literature acknowledges some challenges of incorporating new treatment arm(s) into ongoing trials, there is a lack of research explicitly addressing randomisation methods. Perhaps unsurprisingly therefore, Pitre et al. (2023) identified that 17 (68%) out of the 25 published platform trials used SR, 7 (28%) applied RAR, and only one used minimisation methods. The randomisation process in platform trials is quite complex because each time a new arm is introduced, the randomisation algorithm must be adapted to account for the new arm, which of course affects the allocation to all arms. The complexity is particularly notable when:

- The trial has a fixed sample size, and therefore the remaining sample size is limited when a new treatment arm is introduced. The randomised allocation process to

assign patients to the initial and newly introduced treatment arms may become difficult.

- The study aims to achieve high covariate balance and compare multiple treatment arms simultaneously. This requires a sophisticated randomisation method that goes beyond simple random allocation and actively considers the distribution of covariates across existing arms and newly added arm(s), given one may anticipate the existing arms would be better balanced.

This raises the question of how patients should be allocated in the newly added arm, relative to the control arm and other treatment arms, to reduce the covariate imbalance. Furthermore, it is of interest to know if any randomisation approach can achieve high covariate balance whilst also maintaining strong assignment randomness.

The primary aim of this chapter, therefore, is to address this question and identify an efficient randomisation method, accounting for the selected allocation ratios, to reduce covariate imbalance and increase assignment randomness. I first describe several evaluated randomisation methods for the particular platform trial design under consideration. I consider both randomisation methods that are widely implemented in platform trials, such as SR, as well as other methods that performed well in the context of multi-arm trials as shown in Chapter 2, e.g., SBR and SBUD. I then detail a motivating example, the FLAIR trial (Howard et al., 2021), which is subsequently used in investigating the impact of the choice of a variety of allocation ratios with particular randomisation methods.

4.2 Methods

4.2.1 Two-stage platform trials

Consider a two-stage platform trial that initially has $K_1 = K$ treatment arms present, with a total planned sample size of $N = N_1 + N_2$, where N_s is the sample size of stage $s = 1, 2$. An additional treatment is added in stage 2, such that the number of treatment arms is then $K_2 = K + 1$. The randomisation process in this design is described in two steps:

- In stage 1, N_1 patients are to be allocated to one of the treatment arms, $k_1 = 0, 1, \ldots, K-1$, where $k_1 = 0$ is a shared control arm, present throughout the trial, and $k_1 = 1, \ldots, K-1$ are experimental arms. This is to be realised such that the allocation ratio is $r_{1,0} : r_{1,1} : \cdots : r_{1,K-1}$. In this stage, a popular allocation ratio choice is equal randomisation (Sverdlov et al., 2022). Therefore, I assume that patients are to be allocated to each initially-present treatment arm with $r_{1,0} = r_{1,1} = \cdots = r_{1,K-1} = 1$.

- Stage 2 starts with the introduction of a new treatment; the remaining N_2 patients are then to be allocated to the existing arms and the new arm, indexed $k_2 = 0, 1, \ldots, K$, with allocation $r_{2,0} : r_{2,1} : \cdots : r_{2,K}$. In this stage, I consider various possible allocation ratios, to allow flexibility in directing more patients to the new arm, compared to the initially-present arms.

I set R_s to be the sum of the allocation ratios in stage s. That is

$$R_s = \sum_{k=0}^{K_s} r_{s,k}.$$

Furthermore, I index the patients in stage s by $i = 1, \ldots, N_s$. I will then use $\mathbb{I}(T_{i,s} = k)$ as an indicator variable, taking the value 1 if patient i in stage s is assigned to treatment

k and 0 otherwise. Then I set

$$N_{s,k,l} = \sum_{i=1}^{l} \mathbb{I}(T_{i,s} = k).$$

That is, $N_{s,k,l}$ is the number of patients in stage s assigned to arm k, amongst the first l patients.

Finally, for simplicity, I assume that a set of J covariates of interest, indexed $j = 1, \ldots, J$, will be measured and be the same in both stages. I use $X_{i,j,s}$ for the value of covariate j for patient i in stage s. As many treatment allocation methods work in terms of strata defined through binary covariates, $X_{i,j,s}$ is assumed to take values 0 or 1.

4.2.2 Considered allocation procedures

I consider four different randomisation methods for the considered platform trial design: SR, SBR, SBUD, and Mini.

Simple Randomisation (SR)

SR can be implemented in the assumed platform trial design as follows. In stage 1, each patient has an equal chance of being assigned to one of $k_1 = 0, \ldots, K-1$. Then, in stage 2, when the new arm is introduced to the trial, SR can be easily implemented using unequal allocation ratios. Combined, the probability patient i in stage s is assigned to treatment k is given by

$$\mathbb{P}(T_{i,s} = k) = \frac{r_{s,k}}{R_s}.$$

Stratified Block Randomisation (SBR)

The SBR process in pre-planned platform trials is assumed to function as follows

1. 2^J strata are specified, defined from all possible combinations of covariates $j = 1, \ldots, J$. I use $c_{i,s}$ for the strata to which patient i in stage s belongs.

2. A chosen block size B, which is a multiple of the allocation ratios in both stages (see below for an example), is nominated.

3. In stage s, the N_s patients are allocated into fixed blocks of size B based on their stratum $c_{i,s}$. Each block contains a random permuted sequence of allocations containing a predetermined number of allocations to each treatment arm based on the desired allocation ratios.

The treatment assignments will therefore differ in each stage depending on the number of arms and the allocation ratio to each arm. For example, suppose that a platform trial was planned to begin with two arms using allocation ratio 1:1. Then, in stage 2, a new treatment is introduced, with the allocation ratio switched to 1:1:2. If $B = 8$ is chosen, 4 patients are randomly assigned to each treatment per block in stage 1, but in stage 2 the assignment changes to assign 2 patients each for the initially-present arms, and 4 for the new arm, within each block.

Define the block to which patient i in stage s is assigned to as $b_{i,s}$. Then, the probability patient i in stage s is assigned to treatment k is

$$\mathbb{P}(T_{i,s} = k) = \frac{B(r_{s,k}/R_s) - \sum_{l=1}^{i-1} \mathbb{I}(b_l = b_i)\mathbb{I}(c_l = c_i)\mathbb{I}(T_{l,s} = k)}{B - \sum_{l=1}^{i-1} \mathbb{I}(b_l = b_i)\mathbb{I}(c_l = c_i)}.$$

Stratified Block Urn Design (SBUD)

As in SBR, the randomisation process begins with predefined strata $c_{i,s}$ and allocation ratios $r_{s,k}$ for each arm k and stage s. Furthermore, a chosen MTI threshold should be predetermined that specifies the maximum acceptable difference within each stratum in the number of patients allocated to each treatment arm, indexed λ (see Subsection 2.2.1). Following this, the block size $B = \lambda R_s$ is predetermined, which is a multiple of the

allocation ratios in both stages as in SBR (see below for an example), is nominated.

The assignment process of SBUD can then be illustrated within each stratum in each stage by a model with two urns, termed active and inactive. The allocation procedure starts with an empty inactive urn and B balls in the active urn. In the active urn, the balls have k_s distinct labels, representing each of the arms present in stage s. When a treatment assignment is requested, a ball is randomly selected; if this is of label k then this patient is assigned to arm k. This ball is then placed in the inactive urn.

This process is repeated for each assignment until a minimal balanced set is collected in the inactive urn. These R_s balls are then returned to the active urn from the inactive one. Other balls, if any, remain in the inactive urn. Then, balls are again drawn from the active urn and placed into the inactive urn until it contains a minimal balanced set once more ,which are then again transferred to the active urn.

Returning to the previous example, if $\lambda = 2$ and $B = 6$ are chosen, 6 balls with distinct labels for each treatment are added to the active urn in stage 1, but in stage 2 the number of balls is changed to 3 balls for each of the initially-present treatments, and 6 balls for the new treatment.

With the above, the probability patient i in stage s is assigned to treatment k is given by

$$\mathbb{P}(T_{i,s} = k) = \frac{r_{s,k}\lambda + r_{s,k}N^*_{i-1,c_i} - N_{k,i-1,c_i}}{B + R_s N^*_{i-1,c_i} - \{\sum_{l=1}^{i} \mathbb{I}(c_l = c_i) - 1\}},$$

Where N^*_{i-1,c_i} is the number of minimal balanced sets in previous assignments in stratum c and $N_{k,i,c}$ is the number of patients who have been assigned in stage s and stratum c who have been assigned to treatment k, amongst the first i patients. That is,

$$N_{k,i,c} = \sum_{l=1}^{i} \mathbb{I}(T_{i,s} = k)\mathbb{I}(c_i = c).$$

Minimisation

The standard minimisation method proposed by Taves (1974); Pocock and Simon (1975), was later expanded to unequal allocation ratios using allocation ratio preserving biased coin minimisation (ARP BCM) to preserve the allocation ratio at every allocation step (Kuznetsova and Tymofyeyev, 2012; Jin et al., 2019) (see Subsection 1.5.3 for more details).

In platform trials, the minimisation algorithm can be described as follows: in each stage s, patients $i = 1, \ldots, qN_s$ are allocated to the k_s treatment arms using SR, where q is termed a 'burn-in; period, and takes range between 0 and 1. Following this, patients are allocated to a treatment k to minimise the marginal imbalance. The marginal imbalance can be measured by the range, variance, or standard deviation (Kuznetsova and Tymofyeyev, 2012; Jin et al., 2019). For brevity, I focus here on the range as an intuitively simple and commonly used measure of imbalance (Jin et al., 2019).

After a new arm is introduced (stage 2), minimisation could be implemented in two different cases:

- The minimisation algorithm takes into account the characteristics and allocation of patients already enrolled in the existing arms of the trial. This historical data is considered "non-concurrent" for the new arm since no patients have been assigned to it yet. This might be preferable in some settings, when a high covariate balance is highly desirable for the existing arms.

- The minimisation process can be implemented independently at each stage s. This approach ensures balance in treatment allocation across different stages of the trial. Each stage functions as a separate trial with its own minimisation process.

In this chapter, I consider both cases: where the minimisation algorithm in stage 2 relies on the allocations in stage 1, and where it is implemented independently in each stage s.

The ARP BCM can be illustrated as follows. Set $N_{i,j,s,k,l}$ as the number of patients who have been assigned, among the first i patients, to treatment k, whose j^{th} covariate takes the value l in stage s. That is

$$N_{i,j,s,k,l} = \sum_{m=1}^{i} \mathbb{I}(T_{m,s} = k)\mathbb{I}(X_{m,j} = l).$$

Then, suppose that the next patient, patient $i+1$, has covariate information $X_{i+1,1,s}, \ldots, X_{i,j,s}$. The number of patients up to this point at these levels in stage s, for arm k, is given by $N_{i,j,s,k,X_{i,j}}$ for $j = 1, \ldots, J$.

The marginal imbalance can then be defined by the range at each level of the covariate j across all treatment arms using ARP BCM to adjust the allocation ratio in each randomisation step. That is:

$$RG(N_{i,j,0,X_{i,j}}/r_0, \ldots, N_{i,j,K_s,X_{i,j}}/r_{s,k}) = \max_{s,k=0,\ldots} N_{i,j,s,k,X_{i,j}}/r_{s,k} - \min_{s,k=0\ldots} N_{i,j,s,k,X_{i,j}}/r_{s,k}.$$

The total hypothetical imbalance, if patient $i+1$ is assigned to arm k in stage s, is then defined as a weighted sum of the level-based imbalance for covariates included in the minimisation routine

$$I_{i+1,s,k} = \sum_{j=1}^{J} w_j RG(N_{i,j,0,X_{i,j}}/r_{s,0}, \ldots, N_{i,j,s,k,X_{i,j}}/r_{s,k}).$$

Here, the weights w_j are pre-specified and indicate the relative importance of covariates in measuring the imbalance (Jin et al., 2019). Then, probability p_{min} is dedicated to assigning patient i to one of the arms that minimises the subsequent imbalance, the probability $(1 - p_{min})$ to the remaining arms. Later, I will fix the burnin parameter $q = 0.1$, and identify a particular minimisation routine as Mini(p_{min}).

4.2.3 Performance evaluation metrics

In Chapter 2, I evaluated the randomisation methods in terms of covariate imbalance and allocation predictability at the end of the trial. However, in this chapter, I consider evaluating randomisation methods with a set of allocation ratios in terms of covariate balance and randomness as the trial progresses. I also examine the overall performance for each of the randomisation methods considered, which is the trade-off between covariate balance and randomness.

Degree of imbalance according to patient covariates

To evaluate the ability of the allocation techniques as the trial progresses to balance covariate factors between arms, the proportion of patients in each arm with $X_{i,j} = 1$ is calculated as:

$$\upsilon_{i,j,k} = \frac{N_{i,j,k,1}}{N_{i,k}}.$$

Then, the maximum absolute difference in these proportions is calculated for each experimental arm compared to the control arm, for each covariate j, as

$$\Upsilon_{i,j} = \max\left(|\upsilon_{i,j,1} - \upsilon_{i,j,0}|, \ldots, |\upsilon_{i,j,k} - \upsilon_{i,j,0}|\right).$$

These maximum of these differences $\max\left(\Upsilon_{i,1}, \Upsilon_{i,2}, \ldots, \Upsilon_{i,J}\right)$ is used as a performance metric. A randomisation procedure with low imbalance is preferable. Note that the maximal covariate imbalance is computed after each patient i allocated to treatment arm.

Allocation predictability

As mentioned in Subsection 2.2.2, the choice of randomisation method has a major impact on predictability, which is a popular indicator for evaluating randomisation rou-

tines. Predictability can be defined as the number of times the allocation could have been determined correctly by an investigator who knows the treatment allocation of all previous patients and tries to correctly identify the next allocation as whichever arm currently has the fewest assignments, factoring in the allocation ratio planned for each treatment arm.

Since the randomisation process for platform trial designs involves different stages, each with different treatment arms and allocation ratios, I calculate predictability independently for each stage s. I measured the predictability through:

$$\begin{aligned} G_i &\sim \ Multinomial(p_{i,s,0}, \ldots, p_{i,s,K}), \\ p_{i,s,k} &= \frac{\mathbb{I}\{N_{s,k,i-1} = \min\left(N_{s,0,i-1}/r_{s,0}, \ldots, N_{s,K,i-1}/r_{s,K}\right)\}}{\sum_{l=0}^{K} \mathbb{I}\{N_{s,l,i-1} = \min\left(N_{s,0,i-1}, \ldots, N_{s,K,i-1}\right)\}}. \end{aligned}$$

The predictability for each stage is given by

$$\Omega_s = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbb{I}\left(T_i = G_i\right).$$

Then, the predictability for both stages is defined by

$$\Omega_N = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}\left(T_i = G_i\right).$$

Covariate balance/randomness trade-off

In the simulation study, at the end of the trial, the maximum value of maximum covariate differences for each experimental treatment arm is defined as

$$\Upsilon_N = \max(\Upsilon_{1,N}, \Upsilon_{2,N}, \ldots, \Upsilon_{K,N})$$

The values of Υ_N and Ω_N are rescaled to (Υ_N^* and Ω_N^*) by the minimum and range for each allocation ratio to calculate an overall weighted score that can be used as an

evaluation metric. The weighted score is defined as

$$\xi(N) = \sqrt{\frac{(\Upsilon_N^*)^2 + (\Omega_N^*)^2}{2}}.$$

This assumes that covariate balance and unpredictability have equal importance. A randomisation procedure with a small $\xi(N)$ is preferable. This measure is useful as it takes into consideration the common trade-off between balance and predictability.

4.3 Motivating Example

As motivation for many of the assumed parameters values, I use the FLAIR trial (Howard et al., 2021). The FLAIR trial is a randomised controlled confirmatory trial in chronic lymphocyte leukaemia. It was originally planned as a two-arm trial ($K = 2$), with a planned total sample size of $N = 754$ to achieve 80% power. Then, a new treatment arm was planned to be added halfway through the trial ($N_1 = N_2 = 377$). Minimisation, with a random element, was implemented to achieve covariate balance with respect to important categorical covariates such as gender, age, disease severity, and medical centres. In this trial, patients were equally randomised to one of the open treatment arms throughout the trial (i.e., $r_{s,k} = 1$ for all k and s).

4.4 Simulation study

Using the above motivating trial, I have examined different randomisation methods with multiple allocation ratios in terms of the performance measures discussed in Subsection 4.2.3. The following summarises the common assumptions across two considered scenarios. I consider $N = 754$ to be the total sample size, with $N_1 = N_2 = N/2$. I assume that four binary covariates, $J = 4$, were used in the randomisation routine with a value of $X_{i,j}$ drawn independently for all i and j as $X_{i,j} \sim Bern(0.25)$. For minimisation, I considered $q = 0.1$ and $p_{min} = 0.7$. The above are applied in the following two scenarios:

- Scenario 1: Sets $K = 2$ initially. In stage 2, patients are to be allocated to one of the $K + 1$ treatments using one of several possible allocation ratios, specifically $(r_{2,0}, r_{2,1}, r_{2,2}) \in \{(1,1,1), (3,1,2), (2,1,3), (2,1,2), (1,1,2), (1,1,3), (2,1,1)\}$. For SBUD, I assume $\lambda = 4$. For SBR, I consider $B = 4R_2$; specifically, $B = 12$, for allocation ratios of 1:1:1; $B = 16$ for allocation ratios of 1:1:2, 2:1:1; $B = 24$ for

allocation ratios of 3:1:2 and 2:1:3; and $B = 20$ for allocation ratios of 2:1:2 and 1:1:3.

- Scenario 2: Sets $K = 3$ initially. In stage 2, patients are to be allocated to one of the $K + 1$ treatments using one of several possible allocation ratios, specifically $(r_{2,0}, r_{2,1}, r_{2,2}, r_{2,3}) \in \{(1,1,1,1), (3,1,1,2), (2,1,1,3), (2,1,1,2), (1,1,1,2), (1,1,1,3), (2,1,1,1)\}$.
  For SBUD method, I assume $\lambda = 3$. For SBR, I consider $B = 3R_2$; specifically, $B = 12$ for allocation ratios of 1:1:1:1; $B = 15$ for allocation ratios of 1:1:1:2 and 2:1:1:1; $B = 21$ for allocation ratios of 3:1:1:2 and 2:1:1:3; and $B = 18$ for allocation ratios of 1:1:1:3, and 2:1:1:2.

Furthermore, I also consider evaluating the implications of using Mini(0.7) with non-concurrent data in the stage 2 randomisation process. For each conidered set of parameter combinations, 10,000 simulation replicates were performed to empirically estimate the performance metrics.

4.5 Results

4.5.1 Degree of imbalance according to patient covariates

The simulation results comparing the covariate imbalance for the considered scenarios are presented in Figures 4.1 and 4.2. As expected, in both scenarios, the maximum covariate imbalance decreases for all randomisation methods as patient allocation progresses. In addition, as expected, I observe that the covariate imbalance for the treatment arms becomes closer to each other as the number of patients allocated increases. This trend is observed for all randomisation methods.

Furthermore, the simulation results illustrate that in both scenarios, SR consistently underperforms compared to other methods in terms of covariate balance in all treatment arms (existing and newly added arms). This is in line with the expected limitations of SR, as it does not attempt to control the potential imbalances in the covariates. Mini(0.7) on the other hand, appears to consistently perform slightly better in terms of covariate balance for all treatment arms and allocation ratios compared to the other randomisation methods considered.

Surprisingly, Figures 4.1 and 4.2 demonstrate that allocating more patients to only the newly added arm (e.g., 1:1:2, 1:1:3, 1:1:1:2, and 1:1:1:3) leads to a slightly higher covariate imbalance for this treatment arm as the trial progresses compared to other allocation ratios (e.g., 2:1:1, 2:1:2, etc); this applies to all randomisation methods.

Moreover, Figure 4.3 illustrates the simulation results exploring the influence of Mini(0.7) on covariate imbalance using past assignments in existing arms (non-concurrent data). Notable, the allocation ratios of 3:1:2 and 2:1:1 achieved a similar covariate balance at the end of the trial using non-concurrent assignments in the Mini(0.7). This imbalance is even lower compared to the existing arm. This suggests that using non-concurrent data in the Mini(0.7) process might actually improve covariate balance compared to using only concurrent data.

In addition, the allocation ratio of 2:1:2 achieves a similar covariate imbalance for the newly added arm and the existing arm. This indicates that considering non-concurrent data does not introduce bias towards the existing arm, which is crucial for fair comparisons at the end of the trial.

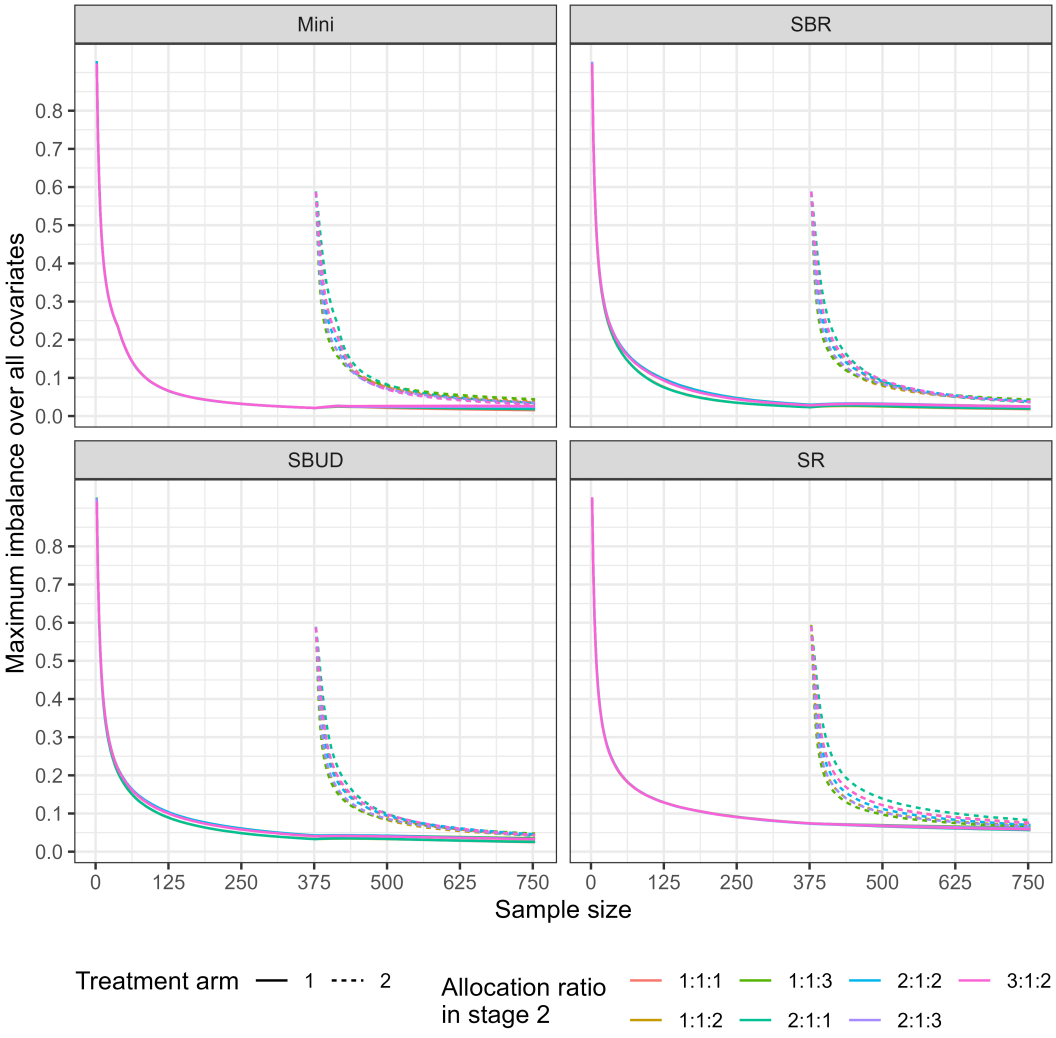

Figure 4.1: The empirical covariate imbalance between the control arm and each of the experimental arms, for each of the randomisation methods, under a variety of allocation ratios in stage 2 (Scenario 1).

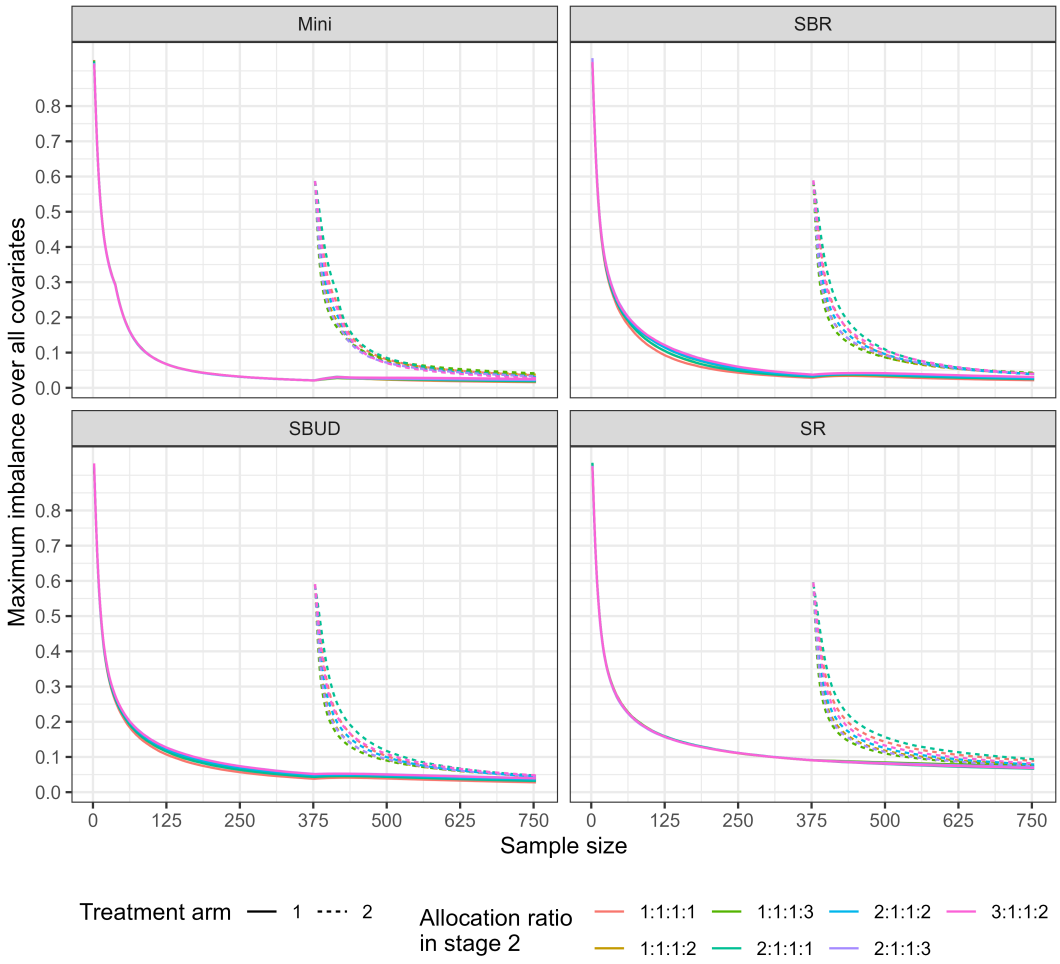

Figure 4.2: The maximum covariate imbalance between the control arm and existing arms and between the control arm and newly added arm for each of the randomisation methods and under a variety of allocation ratios in stage 2 (Scenario 2).

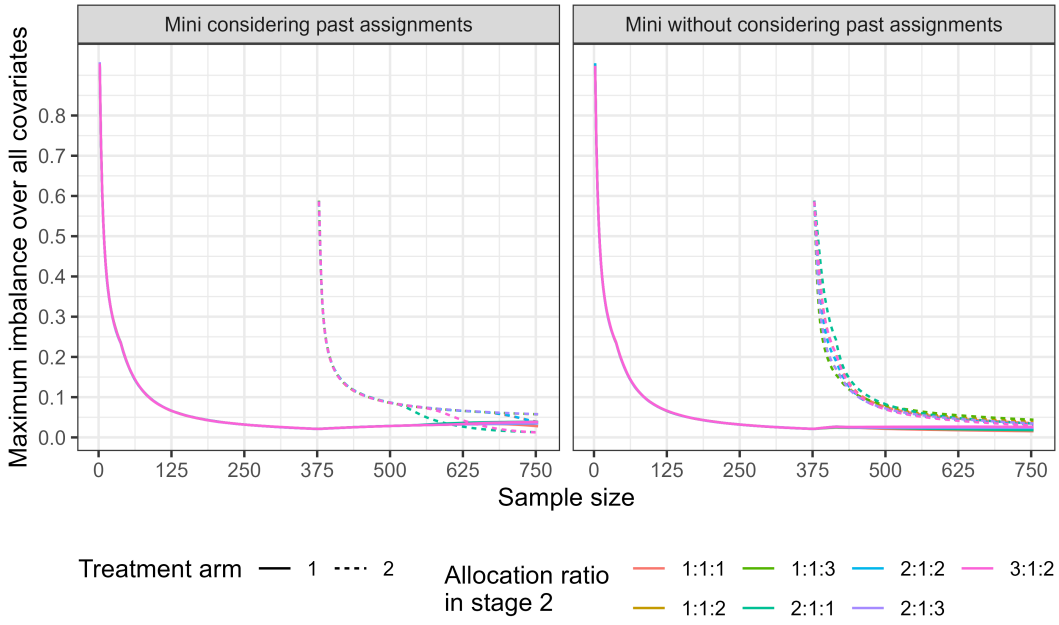

Figure 4.3: The empirical covariate imbalance between the control arm and each of the experimental arms, for minimisation method, under of allocation ratios in stage 2, when considering past assignments and without considering past assignments (Scenario 1).

4.5.2 Allocation predictability

The allocation predictability is shown in Figures 4.4 and 4.5 for scenarios 1 and 2 respectively. Mini(0.7) is more predictable relative to SBR, SBUD, and SR in both stages, according to the simulation results. In addition, in both scenarios, allocation predictability increases as patient allocation progresses except for SR.

As expected, use of a randomisation method with an unequal allocation ratio can have a considerable impact on predictability. For example, Mini(0.7) with unequal allocation ratios have higher predictability compared to Mini(0.7) with equal allocation ratio.

The predictability of Mini(0.7) is higher with allocation ratios of 1:1:3 (Scenario 1) and 1:1:1:3 (Scenario 2) compared to Mini(0.7) with all other allocation ratios considered.

By contrast, the predictability of SBR, SBUD, and SR with different unequal allocation ratios is lower compared to those with equal allocation ratio. This means that these methods rely more heavily on randomisation, especially with unequal allocation ratios. Furthermore, in both scenarios, the SBR and SBUD have similar predictability, although this is slightly lower in SBUD for allocation ratios of 2:1:1:1, 2:1:1:2, and 1:1:1:2.

4.5.3 Covariate balance/randomness trade-off

Tables 4.1 and 4.2 illustrate simulation results evaluating the overall performance of covariate balance and randomness at the end of the trial. The results indicate that SBUD performs slightly better than SBR for some but not all considered stage 2 allocation ratios; these two methods generally perform similarly. However, SBUD performs slightly better than SBR for some but not all considered stage 2 allocation ratios. On the other hand, SR and Mini have similar overall performance that is inferior to the other randomisation methods considered.

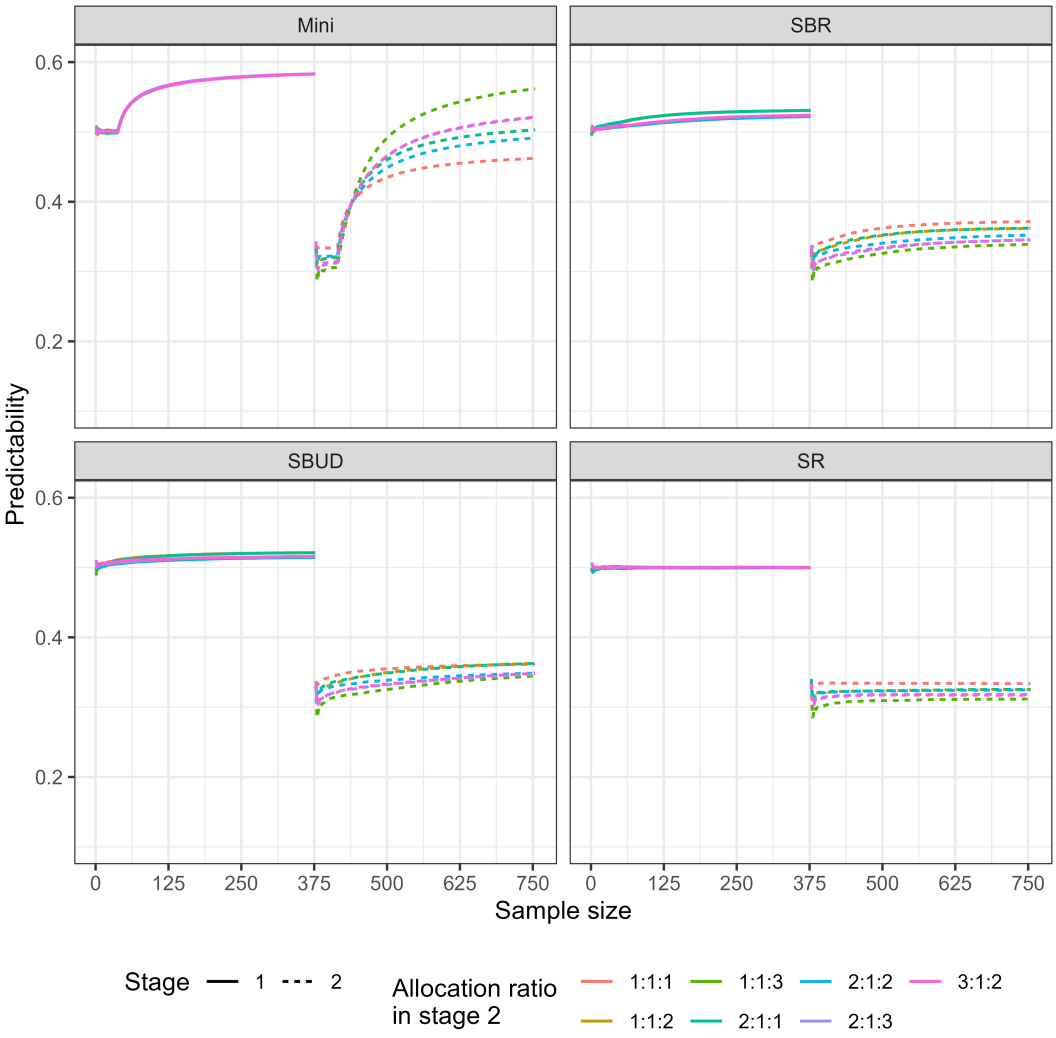

Figure 4.4: The empirical mean predictability for each stage, under a variety of stage 2 allocation ratios (Scenario 1).

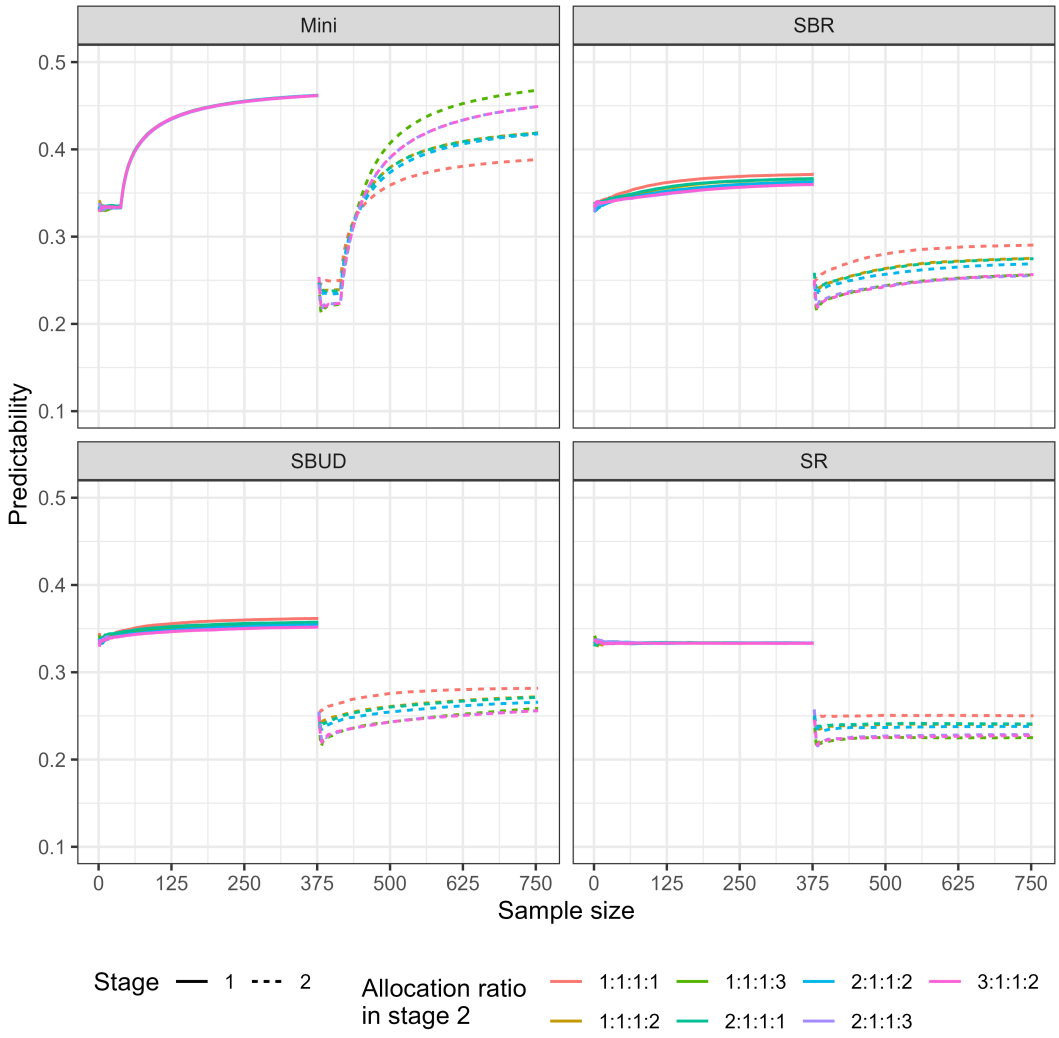

Figure 4.5: The empirical mean predictability for each, under a variety of stage 2 allocation ratios (Scenario 2).

Table 4.1: The trade-off between covariate balance and predictability at the end of the trial for different randomisation methods and allocation ratios (Scenario 1).

Method	Allocation ratio in stage 2	Maximum covariate imbalance	Predictability in both stages	Overall performance
SR	1:1:1	0.076	0.42	0.71
SBR	1:1:1	0.038	0.43	0.27
SBUD	1:1:1	0.043	0.43	0.22
Mini	1:1:1	0.035	0.46	0.71
SR	1:1:2	0.068	0.50	0.81
SBR	1:1:2	0.039	0.52	0.28
SBUD	1:1:2	0.042	0.51	0.20
Mini	1:1:2	0.039	0.54	0.71
SR	1:1:3	0.065	0.55	0.69
SBR	1:1:3	0.043	0.56	0.22
SBUD	1:1:3	0.047	0.56	0.18
Mini	1:1:3	0.044	0.59	0.71
SR	2:1:1	0.083	0.38	0.71
SBR	2:1:1	0.036	0.39	0.26
SBUD	2:1:1	0.041	0.39	0.21
Mini	2:1:1	0.034	0.42	0.71
SR	2:1:2	0.070	0.45	0.71
SBR	2:1:2	0.039	0.46	0.23
SBUD	2:1:2	0.046	0.46	0.28
Mini	2:1:2	0.032	0.49	0.71
SR	2:1:3	0.066	0.50	0.71
SBR	2:1:3	0.038	0.51	0.24
SBUD	2:1:3	0.043	0.51	0.23
Mini	2:1:3	0.035	0.54	0.71
SR	3:1:2	0.074	0.42	0.71
SBR	3:1:2	0.037	0.43	0.24
SBUD	3:1:2	0.044	0.42	0.26
Mini	3:1:2	0.030	0.46	0.71

Table 4.2: The trade-off between covariate balance and predictability at the end of the trial for different randomisation methods and allocation ratios (Scenario 2).

Method	Allocation ratio in stage 2	Maximum covariate imbalance	Predictability in both stages	Overall performance
SR	1:1:1:1	0.09	0.29	0.71
SBR	1:1:1:1	0.04	0.31	0.22
SBUD	1:1:1:1	0.04	0.31	0.20
Mini	1:1:1:1	0.04	0.36	0.71
SR	1:1:1:2	0.08	0.37	0.71
SBR	1:1:1:2	0.04	0.38	0.20
SBUD	1:1:1:2	0.05	0.38	0.19
Mini	1:1:1:2	0.04	0.43	0.71
SR	1:1:1:3	0.07	0.40	0.71
SBR	1:1:1:3	0.04	0.43	0.24
SBUD	1:1:1:3	0.05	0.42	0.22
Mini	1:1:1:3	0.04	0.48	0.71
SR	2:1:1:1	0.09	0.27	0.71
SBR	2:1:1:1	0.04	0.28	0.19
SBUD	2:1:1:1	0.05	0.28	0.21
Mini	2:1:1:1	0.03	0.33	0.71
SR	2:1:1:2	0.08	0.33	0.71
SBR	2:1:1:2	0.04	0.35	0.20
SBUD	2:1:1:2	0.05	0.34	0.24
Mini	2:1:1:2	0.03	0.40	0.71
SR	2:1:1:3	0.07	0.38	0.71
SBR	2:1:1:3	0.04	0.39	0.21
SBUD	2:1:1:3	0.05	0.39	0.26
Mini	2:1:1:3	0.03	0.44	0.71
SR	3:1:1:2	0.08	0.31	0.71
SBR	3:1:1:2	0.04	0.32	0.22
SBUD	3:1:1:2	0.05	0.32	0.28
Mini	3:1:1:2	0.03	0.37	0.71

4.6 Discussion

It can be practically advantageous to add new experimental arms to an existing trial, as it does not include the long process of starting a whole new trial (Sydes et al., 2012; Elm et al., 2012). This means that the evaluation of the newly emerging treatment arm(s) can be sped up, conducted at reduced costs, and with a lower number of patients required (Cohen et al., 2015). To date, though, no work has sought to assess what influence randomisation routines may have on the efficiency of platform trials. Therefore, in this chapter, I evaluated several randomisation methods under a variety of allocation ratios, focusing on their covariate imbalance and assignment randomness when a new experimental arm is added to an ongoing trial. I also considered the trade-off between covariate balance and randomness.

The results demonstrate that increasing the allocation ratio for only the newly added arm does not minimise the covariate imbalance if all treatment arms finish recruitment simultaneously. To minimise covariate imbalance, it would be preferable to decrease the allocation ratio in the existing arm and increase the allocation ratio in both the control and newly added arms. This is also recommended by Bennett and Mander (2020) to maximise the overall power when adding a new experimental arm, and all treatment arms finish recruitment simultaneously. In the paper, the authors stated that the optimal allocation ratio (in a trial concluding with a control and two treatment arms) would be 1.236:0.566:1 (Bennett and Mander, 2020). This ratio could be, at least approximately, implemented in practice by using minimisation or through using a large block size.

Furthermore, the results illustrate that Mini(0.7) performs better in terms of covariate balance compared to other randomisation methods for different allocation ratios, particularly when the platform trials initially start with more than two experimental arms, as in scenario 2. On the other hand, Mini(0.7) with unequal allocation ratios is

more predictable compared to the other randomisation methods considered. Note that in platform trial designs, it should be impossible for any investigator to have all of the knowledge needed for the above guessing strategy, even when the trial is open-label. In this study, predictability is rigorously defined to quantify its potential impact. However, in practice, it does not matter whether the next allocation is predictable, but rather whether those in a position to influence recruitment believe that they know what will happen next with higher chance, regardless of this being correct or not. Nevertheless, acting upon a belief that is not correct could introduce bias.

Although SBR is a common randomisation method in RCTs due to its advantages (Lin et al., 2015), it may be challenging to implement effectively in platform trials with varying numbers of treatment arms and unequal allocation ratios. SBR requires careful assessment of the number of covariates that can be stratified and careful choice of the size of the block, to avoid partially filled blocks and overall imbalances. Similarly, this may also be more challenging for SBUD due to the need to predetermine both the block size and MTI.

Furthermore, Tables 4.1 and 4.2 show that SR and Mini do about the same at the end of the trial when the goal is to get both high covariate balance and randomness. This is because SR is highly random, while Mini performs best in terms of covariate balance. The overall performance difference between SBR and SBUD depends on the allocation ratio and the number of initial arms.

Before concluding, it is important to recognise the limitations of this work. I have focused on the case where all treatment arms finished recruiting simultaneously. In practice, instead of recruiting to all treatment arms until the end of the study and reducing the allocation to the existing treatment arm, the existing treatment arm(s) could finish recruitment early. Reducing allocation to the existing experimental treatment arm(s) when the new treatment arm is added delays learning about the effect of the existing treatment arm. Further work is required to identify effective randomisation approaches

when recruitment does not continue for all arms until trial conclusion.

Moreover, in this chapter, I focused on covariate balance with a large sample size. However, evaluating different clinical trial power (marginal, disjunctive and conjunctive powers) and controlling FWER for different allocation ratios and randomisation methods would be valuable for future research. I also examined the trade-off between balance and randomness, assuming that both characteristics (as measured by the specified metrics) hold equal importance. However, this assumption may not universally apply in practice. Consequently, exploring different weights assigned to these characteristics could be an interesting for further research.

In addition, I explored the allocation ratios when a new treatment enters the ongoing trial halfway through the study. In many platform trials, a new treatment arm(s) would enter the trial at different time points. This could be a challenging point; particularly when a new treatment enters a trial after randomising half of the planned sample size. In this case, increasing allocation ratios further towards the control arm and newly added arm would be required to achieve high covariate balance. Further research would be useful to evaluate allocation ratios that achieve a high covariate balance. However, in the results, the considered allocation ratio in the first stage is always equal randomisation, and does not depend on the allocation ratios chosen for stage 2.

Furthermore, in this study, I assumed a fixed total sample size. Therefore, exploring multiple allocation ratios is arguably more limited, as it involves redistributing patients across existing and new arms. Possible future work could look at adjusting the allocation ratio to minimise the covariate imbalance with an adaptive sample size. Moreover, I assumed only binary covariates that are the same for both stages. Further research exploring different covariate scenarios would be of interest. Lastly, I have assumed that the new arm was added halfway through the trial. Future research examining performance when adding a new treatment arm at different time points could be helpful.

Chapter 5

Concluding remarks and recommendations

5.1 Summary

This thesis has focused on randomisation methods and the choice of allocation ratio in three different clinical trial designs, including multi-arm, MAMS, and platform trial designs. Each chapter sought to address a different issue that could potentially prevent these designs from being implemented in practice. In the following, I outline each chapter in turn and highlight the main contributions within.

5.1.1 Chapter 2: Randomisation in multi-arm trial designs

In Chapter 2, I evaluated and compared randomisation methods that investigators commonly use in practice for two-arm trials, within the context of multi-arm trials. I also included evaluation of several proposed methods that have less-often been used in practice, to examine if they are worth considering for future use. Several performance measures were used for comparison of the methods, including group size balance, covariate balance, loss of precision, and predictability of assignment. I also compared performance for

different definitions of statistical power: marginal, disjunctive, and conjunctive powers. Several conclusions were drawn from my work:

- Choosing an appropriate randomisation approach in RCTs is challenging because each approach has strong performance properties in one, but not all, of the evaluation metrics. Therefore, the choice of the randomisation approach depends on the clinical trial objectives. For example, if a clinical trial objective is to achieve balanced group sizes, PBR provides better performance.

- Investigators have argued extensively against minimisation because it is more predictable compared to other randomisation methods, but in fact including more covariates in the randomisation procedures leads to a reduction in the predictability. Therefore, I recommend using minimisation when there are a large number of covariates that need to be included in the randomisation routine.

- For smaller numbers of covariates, SBR arguably performs better compared to other randomisation methods, but investigators must be very careful about the choice of block size; a larger block size means larger group size and potential covariate imbalance.

- Larger block sizes increase allocation randomness but decrease the treatment balance, as at the end of recruitment a larger proportion of the block can remain unfilled.

5.1.2 Chapter 3: Multi-arm multi-stage trial designs

Before selecting a suitable randomisation method, it is necessary to identify a suitable allocation ratio, with consideration of how the choice of allocation ratio impacts power. In multi-arm trials, the optimal allocation ratio to maximise power has been identified as allocating the square root of the number of experimental arms in the control arm for

each patient in each experimental arm. Equivalent results were not available for MAMS trials. Therefore, in Chapter 3, I studied the impact of stage-wise allocation ratios on marginal, disjunctive, and conjunctive power in a MAMS trial. I also investigated the impact on the expected sample size of the choice of stage-wise allocation ratios. The results showed that:

- The stage-wise allocation ratio does not have a remarkable impact on expected sample size. Therefore, investigators should not hope that the use of alternative allocation ratio can reduce the required sample size.
- The most efficient stage-wise allocation ratios in MAMS trials can indeed vary significantly depending on several factors, including: i) the type of power (marginal, disjunctive, and conjunctive); ii) the number of arms; iii) the number of stages; and iv) other key design parameters (e.g., desired error rates).

5.1.3 Chapter 4: Platform trial designs

Motivated by the fact that platform trials have increased in popularity in drug development, particularly because of COVID-19, I sought to determine suitable randomisation methods (across possible allocation ratios) to reduce the covariate imbalance between a newly added arm and the control arm. The simulation results indicated:

- Minimisation is more amenable to the use of non-concurrent data or concurrent data only in the randomisation algorithm than other methods. Considering the use of non-concurrent data in the minimisation method can achieve a high covariate balance compared to using concurrent data only.
- Allocating more patients to only the newly added arm, did not minimise the covariate imbalance.

5.2 Areas for further work

5.2.1 Optimal allocation ratios for other types of MAMS design and platform trial design

With numerous types of MAMS design, each has different optimal/efficient allocation ratios. In Chapter 3, I only considered a group sequential MAMS design; numerically searching for the best performance amongst a set of allocation ratios considered. However, searching for the optimal performance allocation ratio would be of interest for further research.

In Chapter 4, I also considered a set of allocation ratios in platform trials to evaluate randomisation methods with different allocation ratios in terms of covariate balance and assignment randomness. Future research could investigate optimal allocation ratios and explore the initial allocation ratios at the start of trials for each randomisation method.

5.2.2 Multiple and different outcomes

In Chapters 2 and 3, my focus was principally on analysing continuous outcomes. However, it is important to recognise the importance of other outcome types, such as time-to-event and binary data. While we anticipate that our results would not materially change for other outcome types, verification of this expectation would be helpful. Additionally, researchers often encounter multiple outcomes of interest, which can present analytical challenges. Future research could explore methods to handle these complexities and incorporate various types of outcomes into RCT randomisation routines and analyses.

I also considered throughout statistical tests based on a model. If randomisation-based inference was applied, this could affect our conclusions. Thus, further aspects of interest include the evaluation of how randomisation-based inference would impact our conclusions.

5.2.3 Incorporation of continuous covariates

Throughout this thesis, I only assumed the presence of binary covariates. However, in practice, covariates could be categorical with different levels and also could be continuous. Often, an important continuous covariate is categorised to different levels to incorporate it into a randomisation routine. However, sometimes it may be preferable to keep it as it is; in this case investigators would benefit from the extension of our results to consider which available method can best balance continuous covariate data across treatment arms.

5.2.4 Response Adaptive Randomisation

Traditional clinical trials allocate patients to each treatment arm using a fixed allocation ratio. However, in some clinical trial cases, the allocation ratio is modified to assign more patients to the treatment arm that has demonstrated the best performance compared to other treatments (Hu and Rosenberger, 2006). RAR allows this allocation to change as the trial progresses. Therefore, RAR has been increasingly commonly considered within the context of adaptive randomisation (Saville and Berry, 2016).

Throughout this thesis, randomisation methods were evaluated for fixed allocation ratios to each treatment arm. However, evaluating (covariate) response adaptive randomisation to achieve, for example, a maximum overall power would be interesting in future research.

Appendix A

Appendix

A.1 Supplementary materials for Chapter 2

The simulation results for the minimisation method, in terms of the number of covariates used in the allocation method (J_{rand}) and the assumed value of p_{min}, are given below.

Table A.1: Simulation results for minimisation in Setting 1.

N	J_{rand}	p_{min}	**Mean maximal group-size imbalance (SD)**	**Mean minimal and maximal covariate imbalance**	**Mean predictability (SD)**	**Mean maximal treatment effect variance inflation (SD)**
175	1	0.7	8.40 (8.29)	[0.04,0.23]	[0.03,0.19]	0.16 (0.02)
	2		3.95 (4.59)	[0.02,0.13]	[0.03,0.19]	0.21 (0.03)
	3		1.84 (2.11)	[0.01,0.09]	[0.03,0.19]	0.25 (0.03)
	4		1.24 (1.21)	[0.01,0.09]	-	0.26 (0.03)
175	1	0.9	1.43 (1.40)	[0.01,0.06]	[0.03,0.18]	0.25 (0.03)
	2		0.89 (0.90)	[0.01,0.05]	[0.03,0.18]	0.28 (0.03)
	3		0.67 (0.76)	[0.01,0.05]	[0.03,0.19]	0.31 (0.03)
	4		0.58 (0.69)	[0.01,0.05]	-	0.31 (0.03)
350	1	0.7	14.48 (15.14)	[0.03,0.19]	[0.02,0.13]	0.16 (0.02)
	2		6.03 (7.59)	[0.01,0.09]	[0.02,0.13]	0.20 (0.02)
	3		2.06 (2.56)	[0.01,0.05]	[0.02,0.13]	0.24 (0.02)
	4		1.26 (1.21)	[0.01,0.05]	-	0.26 (0.02)
350	1	0.9	1.46 (1.44)	[0.00,0.03]	[0.02,0.13]	0.24 (0.02)
	2		0.85 (0.91)	[0.00,0.02]	[0.02,0.13]	0.28 (0.02)
	3		0.66 (0.74)	[0.00,0.02]	[0.02,0.13]	0.30 (0.02)
	4		0.59 (0.70)	[0.00,0.03]	-	0.31 (0.02)
700	1	0.7	26.41 (29.31)	[0.03,0.17]	[0.01,0.09]	0.16 (0.01)
	2		10.48 (14.11)	[0.01,0.06]	[0.01,0.09]	0.20 (0.01)
	3		2.24 (3.21)	[0.00,0.03]	[0.01,0.09]	0.23 (0.02)
	4		1.27 (1.21)	[0.00,0.02]	-	0.25 (0.01)
700	1	0.9	1.47 (1.43)	[0.00,0.01]	[0.01,0.09]	0.24 (0.01)
	2		0.86 (0.91)	[0.00,0.01]	[0.01,0.09]	0.28 (0.01)
	3		0.67 (0.75)	[0.00,0.01]	[0.01,0.09]	0.30 (0.01)
	4		0.58 (0.69)	[0.00,0.01]	-	0.31 (0.01)

Table A.2: Simulation results for minimisation in Setting 2.

N	J_{rand}	p_{min}	**Mean maximal group-size imbalance (SD)**	**Mean minimal and maximal covariate imbalance**	**Mean predictability (SD)**	**Mean maximal treatment effect variance inflation (SD)**
43	1	0.7	3.08 (2.81)	[0.05,0.30]	[0.06,0.35]	0.20 (0.06)
	2		1.89 (1.93)	[0.04,0.24]	[0.06,0.34]	0.26 (0.06)
	3		1.32 (1.37)	[0.03,0.23]	[0.05,0.33]	0.29 (0.06)
	4		1.07 (1.08)	[0.03,0.23]	-	0.30 (0.06)
43	1	0.9	1.13 (1.32)	[0.02,0.15]	[0.05,0.33]	0.29 (0.06)
	2		0.79 (0.84)	[0.02,0.15]	[0.05,0.33]	0.33 (0.06)
	3		0.67 (0.72)	[0.02,0.16]	[0.05,0.33]	0.34 (0.06)
	4		0.59 (0.67)	[0.02,0.17]	-	0.35 (0.06)
85	1	0.7	4.71 (4.28)	[0.04,0.23]	[0.04,0.24]	0.20 (0.04)
	2		2.31 (2.46)	[0.02,0.15]	[0.04,0.24]	0.26 (0.04)
	3		1.41 (1.45)	[0.02,0.13]	[0.04,0.24]	0.29 (0.04)
	4		1.14 (1.10)	[0.02,0.13]	-	0.30 (0.04)
85	1	0.9	1.17 (1.14)	[0.01,0.08]	[0.04,0.24]	0.29 (0.04)
	2		0.81 (0.85)	[0.01,0.08]	[0.04,0.24]	0.32 (0.04)
	3		0.67 (0.72)	[0.01,0.08]	[0.04,0.24]	0.34 (0.04)
	4		0.60 (0.68)	[0.01,0.09]	-	0.35 (0.04)
170	1	0.7	7.49 (7.08)	[0.03,0.17]	[0.03,0.17]	0.20 (0.03)
	2		2.96 (3.35)	[0.02,0.09]	[0.03,0.17]	0.25 (0.03)
	3		1.49 (1.53)	[0.01,0.07]	[0.03,0.17]	0.28 (0.03)
	4		1.17 (1.09)	[0.01,0.07]	-	0.30 (0.03)
170	1	0.9	1.22 (1.19)	[0.01,0.04]	[0.03,0.17]	0.29 (0.03)
	2		0.85 (0.84)	[0.01,0.04]	[0.03,0.17]	0.32 (0.03)
	3		0.70 (0.71)	[0.01,0.04]	[0.03,0.17]	0.34 (0.03)
	4		0.66 (0.65)	[0.01,0.04]	-	0.35 (0.03)

Table A.3: The simulation results on mean maximal group-size imbalance for minimisation with different values of p_{min}, when $J_{rand} = 2$.

Setting	N	p_{min}	Mean maximal group-size imbalance (SD)
1	350	0.65	8.98 (11.19)
		0.70	6.03 (7.59)
		0.75	3.64 (4.53)
		0.90	0.87 (0.90)
1	700	0.65	16.61 (21.87)
		0.70	10.48 (14.11)
		0.75	5.26 (7.38)
		0.90	0.86 (0.91)
2	43	0.65	2.25 (2.32)
		0.70	1.89 (1.93)
		0.75	1.53 (1.57)
		0.90	0.79 (0.84)
2	85	0.65	2.95 (3.16)
		0.70	2.32 (2.43)
		0.75	1.77 (1.81)
		0.90	0.81 (0.85)

A.2 Supplementary materials for Chapter 3

Simulation results for Chapter 3

The simulation results for the minimisation method, in terms of the number of covariates used in the allocation method (J_{rand}) and the assumed value of p_{min}, are given below.

Table A.4: Evaluating different allocation ratios in terms of resultant powers (Scenario 1).

Stopping boundaries		**Allocation ratios**	**Sample size**		**Power**			**ESS**	
(e_1, e_2)	(f_1, f_2)	(r_1, r_2)	$(n_{1(0)}, n_{2(0)})$	$(n_{1(k)}, n_{2(k)})$	P_1	P_{dis}	P_{con}	H_0	H_a
(2.793, 2.094)	(0, 2.094)	(1, 1.33)	(47, 58)	(47, 43)	0.7602	0.9381	0.5416	295	244
(2.793, 2.095)	(0, 2.095)	(1, 1.5)	(47, 62)	(47, 42)	0.7593	0.9397	0.5369	296	246
(2.799, 2.099)	(0, 2.099)	(1, 1.73)	(47, 68)	(47, 40)	0.7596	0.9422	0.5333	294	246
(2.803, 2.102)	(0, 2.102)	(1, 2)	(47, 75)	(47, 37)	0.7607	0.9451	0.5309	293	245
(2.810, 2.107)	(0, 2.107)	(1, 2.5)	(47, 85)	(47, 34)	0.7489	0.9435	0.5074	295	244
(2.812, 2.109)	(0, 2.109)	(1, 3)	(47, 94)	(47, 31)	0.7368	0.9411	0.4850	304	247
(2.819, 2.113)	(0, 2.113)	(1, 4)	(47, 107)	(47, 27)	0.7023	0.9292	0.4302	307	247
(2.793, 2.094)	(0, 2.094)	(1.33, 1)	(58, 47)	(43, 47)	0.7431	0.9311	0.5152	292	236
(2.799, 2.099)	(0, 2.099)	(1.33, 1.33)	(58, 58)	(43, 43)	0.7618	0.9440	0.5352	295	236
(2.805, 2.103)	(0, 2.103)	(1.33, 1.5)	(58, 62)	(43, 42)	0.7595	0.9448	0.5287	293	237
(2.806, 2.104)	(0, 2.104)	(1.33, 1.73)	(58, 68)	(43, 40)	0.7607	0.9474	0.5268	299	237
(2.808, 2.105)	(0, 2.105)	(1.33, 2)	(58, 75)	(43, 37)	0.7622	0.9502	0.5254	299	236
(2.807, 2.112)	(0, 2.112)	(1.33, 2.5)	(58, 85)	(43, 34)	0.7505	0.9484	0.5023	302	236
(2.818, 2.113)	(0, 2.113)	(1.33, 3)	(58, 94)	(43, 31)	0.7380	0.9458	0.4792	305	237
(2.824, 2.118)	(0, 2.118)	(1.33, 4)	(58, 107)	(43, 27)	0.7036	0.9342	0.4245	309	238
(2.793, 2.095)	(0, 2.095)	(1.5, 1)	(62, 47)	(42, 47)	0.7429	0.9335	0.5108	294	235

Table A.4 (continued): Evaluating different allocation ratios in terms of resultant powers (Scenario 1).

Stopping boundaries		**Allocation ratios**	**Sample size**		**Power**			**ESS**	
(e_1, e_2)	(f_1, f_2)	(r_1, r_2)	$(n_{1(0)}, n_{2(0)})$	$(n_{1(k)}, n_{2(k)})$	P_1	P_{dis}	P_{con}	H_0	H_a
(2.802, 2.101)	(0, 2.101)	(1.5, 1.33)	(62, 58)	(42, 43)	0.7609	0.9458	0.5304	296	234
(2.806, 2.104)	(0, 2.104)	(1.5, 1.5)	(62, 62)	(42, 42)	0.7590	0.9466	0.5245	298	235
(2.810, 2.107)	(0, 2.107)	(1.5, 1.73)	(62, 68)	(42, 40)	0.7594	0.9488	0.5214	298	235
(2.814, 2.110)	(0, 2.110)	(1.5, 2)	(62, 75)	(42, 37)	0.7602	0.9511	0.5190	300	235
(2.816, 2.111)	(0, 2.111)	(1.5, 2.5)	(62, 85)	(42, 34)	0.7495	0.9499	0.4973	304	236
(2.821, 2.115)	(0, 2.115)	(1.5, 3)	(62, 94)	(42, 31)	0.7365	0.9470	0.4740	306	236
(2.827, 2.119)	(0, 2.119)	(1.5, 4)	(62, 107)	(42, 27)	0.7019	0.9356	0.4187	310	237
(2.799, 2.099)	(0, 2.099)	(1.73, 1)	(68, 47)	(40, 47)	0.7422	0.9362	0.5050	294	232
(2.807, 2.107)	(0, 2.107)	(1.73, 1.33)	(68, 58)	(40, 43)	0.7595	0.9476	0.5238	295	233
(2.810, 2.107)	(0, 2.107)	(1.73, 1.5)	(68, 62)	(40, 42)	0.7583	0.9487	0.5191	297	234
(2.815, 2.111)	(0, 2.111)	(1.73, 1.73)	(68, 68)	(40, 40)	0.7582	0.9506	0.5154	300	234
(2.817, 2.112)	(0, 2.112)	(1.73, 2)	(68, 75)	(40, 37)	0.7595	0.9531	0.5141	301	233
(2.820, 2.115)	(0, 2.115)	(1.73, 2.5)	(68, 85)	(40, 34)	0.7481	0.9515	0.4915	305	234
(2.824, 2.118)	(0, 2.118)	(1.73, 3)	(68, 94)	(40, 31)	0.7352	0.9486	0.4679	307	234
(2.826, 2.119)	(0, 2.119)	(1.73, 4)	(68, 107)	(40, 27)	0.7014	0.9377	0.4141	311	235
(2.803, 2.102)	(0, 2.102)	(2, 1)	(75, 47)	(37, 47)	0.7420	0.9391	0.4997	292	231

Table A.4 (continued): Evaluating different allocation ratios in terms of resultant powers (Scenario 1).

Stopping boundaries		**Allocation ratios**	**Sample size**		**Power**			**ESS**	
(e_1, e_2)	(f_1, f_2)	(r_1, r_2)	$(n_{1(0)}, n_{2(0)})$	$(n_{1(k)}, n_{2(k)})$	P_1	P_{dis}	P_{con}	H_0	H_a
(2.809, 2.106)	(0, 2.106)	(2, 1.33)	(75, 58)	(37, 43)	0.7602	0.9503	0.5206	295	231
(2.814, 2.110)	(0, 2.110)	(2, 1.5)	(75, 62)	(37, 42)	0.7577	0.9509	0.5141	297	231
(2.817, 2.112)	(0, 2.112)	(2, 1.73)	(75, 68)	(37, 40)	0.7582	0.9528	0.5116	299	232
(2.818, 2.113)	(0, 2.113)	(2, 2)	(75, 75)	(37, 37)	0.7594	0.9551	0.5101	300	231
(2.823, 2.117)	(0, 2.117)	(2, 2.5)	(75, 85)	(37, 34)	0.7474	0.9532	0.4867	304	231
(2.826, 2.119)	(0, 2.119)	(2, 3)	(75, 94)	(37, 31)	0.7347	0.9505	0.4637	306	232
(2.829, 2.121)	(0, 2.121)	(2, 4)	(75, 107)	(37, 27)	0.7003	0.9395	0.4086	311	232
(2.812, 2.108)	(0, 2.108)	(2.5, 1)	(85, 47)	(34, 47)	0.7386	0.9419	0.4876	294	232
(2.815, 2.111)	(0, 2.111)	(2.5, 1.33)	(85, 58)	(34, 43)	0.7572	0.9526	0.5097	297	232
(2.818, 2.113)	(0, 2.113)	(2.5, 1.5)	(85, 62)	(34, 42)	0.7552	0.9533	0.5042	304	232
(2.820, 2.115)	(0, 2.115)	(2.5, 1.73)	(85, 68)	(34, 40)	0.7556	0.9551	0.5017	301	233
(2.823, 2.117)	(0, 2.117)	(2.5, 2)	(85, 75)	(34, 37)	0.7563	0.9570	0.4997	302	231
(2.826, 2.119)	(0, 2.119)	(2.5, 2.5)	(85, 85)	(34, 34)	0.7445	0.9551	0.4767	306	232
(2.829, 2.121)	(0, 2.121)	(2.5, 3)	(85, 94)	(34, 31)	0.7313	0.9522	0.4531	308	233
(2.832, 2.123)	(0, 2.123)	(2.5, 4)	(85, 107)	(34, 27)	0.6958	0.9410	0.3967	310	234
(2.815, 2.111)	(0, 2.111)	(3, 1)	(94, 47)	(31, 47)	0.7361	0.9439	0.4787	295	234

Table A.4 (continued): Evaluating different allocation ratios in terms of resultant powers (Scenario 1).

Stopping boundaries		**Allocation ratios**	**Sample size**		**Power**			**ESS**	
(e_1, e_2)	(f_1, f_2)	(r_1, r_2)	$(n_{1(0)}, n_{2(0)})$	$(n_{1(k)}, n_{2(k)})$	P_1	P_{dis}	P_{con}	H_0	H_a
(2.818, 2.113)	(0, 2.113)	(3, 1.33)	(94, 58)	(31, 43)	0.7550	0.9545	0.5013	297	234
(2.813, 2.117)	(0, 2.117)	(3, 1.5)	(94, 62)	(31, 42)	0.7530	0.9550	0.4959	299	234
(2.827, 2.119)	(0, 2.119)	(3, 1.73)	(94, 68)	(31, 40)	0.7525	0.9564	0.4922	301	235
(2.826, 2.119)	(0, 2.119)	(3, 2)	(94, 75)	(31, 37)	0.7539	0.9584	0.4914	302	234
(2.829, 2.121)	(0, 2.121)	(3, 2.5)	(94, 85)	(31, 34)	0.7417	0.9564	0.4681	306	235
(2.833, 2.124)	(0, 2.124)	(3, 3)	(94, 94)	(31, 31)	0.7279	0.9532	0.4436	309	235
(2.833, 2.123)	(0, 2.123)	(3, 4)	(94, 107)	(31, 27)	0.6925	0.9422	0.3876	314	236
(2.819, 2.113)	(0, 2.113)	(4, 1)	(107, 47)	(27, 47)	0.7306	0.9460	0.4631	296	240
(2.824, 2.118)	(0, 2.118)	(4, 1.33)	(107, 58)	(27, 43)	0.7492	0.9559	0.4858	299	241
(2.827, 2.119)	(0, 2.119)	(4, 1.5)	(107, 62)	(27, 42)	0.7473	0.9565	0.4806	301	241
(2.826, 2.119)	(0, 2.119)	(4, 1.73)	(107, 68)	(27, 40)	0.7483	0.9583	0.4792	303	242
(2.829, 2.121)	(0, 2.121)	(4, 2)	(107, 75)	(27, 37)	0.7489	0.9599	0.4774	304	241
(2.832, 2.123)	(0, 2.123)	(4, 2.5)	(107, 85)	(27, 34)	0.7362	0.9576	0.4536	308	242
(2.833, 2.123)	(0, 2.123)	(4, 3)	(107, 94)	(27, 31)	0.7227	0.9546	0.4298	311	242
(2.837, 2.127)	(0, 2.127)	(4, 4)	(107, 107)	(27, 27)	0.6837	0.9421	0.3693	316	243

Table A.5: Evaluating different allocation ratios in terms of resultant powers (Scenario 2).

Stopping boundaries		**Allocation ratios**	**Sample size**		**Power**			**ESS**	
(e_1, e_2, e_3)	(f_1, f_2, f_3)	(r_1, r_2, r_3)	$(n_{1(0)}, n_{2(0)}, n_{3(0)})$	$(n_{1(k)}, n_{2(k)}, n_{3(k)})$	P_1	P_{dis}	P_{con}	H_0	H_a
(3.794, 2.682, 2.190)	(0, 0, 2.190)	(1, 1, 1.33)	(31, 31, 39)	(31, 31, 29)	0.7055	0.9361	0.4065	310.58	270.07
(3.797, 2.685, 2.192)	(0, 0, 2.192)	(1, 1, 1.5)	(31, 31, 42)	(31, 31, 28)	0.7071	0.9385	0.4050	303.15	270.17
(3.801, 2.688, 2.195)	(0, 0, 2.195)	(1, 1, 2)	(31, 31, 51)	(31, 31, 26)	0.7118	0.9448	0.4030	313.81	270.63
(3.813, 2.696, 2.201)	(0, 0, 2.201)	(1, 1, 2.5)	(31, 31, 60)	(31, 31, 24)	0.7141	0.9490	0.3992	317.74	271.58
(3.786, 2.677, 2.186)	(0, 0, 2.186)	(1, 1.33, 1)	(31, 39, 31)	(31, 29, 31)	0.6849	0.9278	0.3779	309.79	268.39
(3.794, 2.682, 2.190)	(0, 0, 2.190)	(1, 1.33, 1.33)	(31, 39, 39)	(31, 29, 29)	0.7113	0.9429	0.4058	312.25	269.13
(3.797, 2.685, 2.192)	(0, 0, 2.192)	(1, 1.33, 1.5)	(31, 39, 42)	(31, 29, 28)	0.7127	0.9450	0.4044	313.10	269.58
(3.809, 2.693, 2.199)	(0, 0, 2.199)	(1, 1.33, 2)	(31, 39, 51)	(31, 29, 26)	0.7156	0.9498	0.4005	316.17	269.98
(3.820, 2.701, 2.206)	(0, 0, 2.206)	(1, 1.33, 2.5)	(31, 39, 60)	(31, 29, 24)	0.7174	0.9534	0.3967	319.68	270.68
(3.790, 2.680, 2.188)	(0, 0, 2.188)	(1, 1.5, 1)	(31, 42, 31)	(31, 28, 31)	0.6846	0.9299	0.3738	310.79	268.04
(3.801, 2.688, 2.195)	(0, 0, 2.195)	(1, 1.5, 1.33)	(31, 42, 39)	(31, 28, 29)	0.7098	0.9441	0.3999	313.03	267.91
(3.805, 2.690, 2.197)	(0, 0, 2.197)	(1, 1.5, 1.5)	(31, 42, 42)	(31, 28, 28)	0.7112	0.9461	0.3988	313.49	269.00
(3.813, 2.696, 2.201)	(0, 0, 2.201)	(1, 1.5, 2)	(31, 42, 51)	(31, 28, 26)	0.7149	0.9511	0.3964	316.81	269.60
(3.824, 2.704, 2.208)	(0, 0, 2.208)	(1, 1.5, 2.5)	(31, 42, 60)	(31, 28, 24)	0.7165	0.9546	0.3923	320.56	270.33
(3.797, 2.685, 2.192)	(0, 0, 2.192)	(1, 2, 1)	(31, 51, 31)	(31, 26, 31)	0.6845	0.9352	0.3637	313.14	268.27
(3.801, 2.688, 2.195)	(0, 0, 2.195)	(1, 2, 1.33)	(31, 51, 39)	(31, 26, 29)	0.7106	0.9488	0.3918	315.57	269.02

Table A.5 (continued): Evaluating different allocation ratios in terms of resultant powers (Scenario 2).

Stopping boundaries		**Allocation ratios**	**Sample size**		**Power**			**ESS**	
(e_1, e_2, e_3)	(f_1, f_2, f_3)	(r_1, r_2, r_3)	$(n_{1(0)}, n_{2(0)}, n_{3(0)})$	$(n_{1(k)}, n_{2(k)}, n_{3(k)})$	P_1	P_{dis}	P_{con}	H_0	H_a
(3.805, 2.690, 2.197)	(0, 0, 2.197)	(1, 2, 1.5)	(31, 51, 42)	(31, 26, 28)	0.7117	0.9506	0.3907	316.96	268.76
(3.816, 2.698, 2.203)	(0, 0, 2.203)	(1, 2, 2)	(31, 51, 51)	(31, 26, 26)	0.7143	0.9547	0.3871	320.15	269.56
(3.820, 2.701, 2.206)	(0, 0, 2.206)	(1, 2, 2.5)	(31, 51, 60)	(31, 26, 24)	0.7172	0.9582	0.3854	323.07	270.02
(3.805, 2.690, 2.197)	(0, 0, 2.197)	(1, 2.5, 1)	(31, 60, 31)	(31, 24, 31)	0.6838	0.9392	0.3546	316.68	269.46
(3.809, 2.693, 2.199)	(0, 0, 2.199)	(1, 2.5, 1.33)	(31, 60, 39)	(31, 24, 29)	0.7096	0.9520	0.3831	319.40	269.26
(3.813, 2.696, 2.201)	(0, 0, 2.201)	(1, 2.5, 1.5)	(31, 60, 42)	(31, 24, 28)	0.7106	0.9535	0.3819	319.77	269.99
(3.820, 2.701, 2.206)	(0, 0, 2.206)	(1, 2.5, 2)	(31, 60, 51)	(31, 24, 26)	0.7134	0.9574	0.3790	323.18	270.37
(3.824, 2.704, 2.208)	(0, 0, 2.208)	(1, 2.5, 2.5)	(31, 60, 60)	(31, 24, 24)	0.7126	0.9571	0.3778	326.56	270.54
(3.790, 2.680, 2.188)	(0, 0, 2.188)	(1.33, 1, 1)	(39, 31, 31)	(29, 31, 31)	0.6763	0.9236	0.3676	308.60	260.73
(3.801, 2.688, 2.195)	(0, 0, 2.195)	(1.33, 1, 1.33)	(39, 31, 39)	(29, 31, 29)	0.7032	0.9396	0.3948	311.66	261.19
(3.805, 2.690, 2.197)	(0, 0, 2.197)	(1.33, 1, 1.5)	(39, 31, 42)	(29, 31, 28)	0.7047	0.9418	0.3937	312.29	261.98
(3.813, 2.696, 2.201)	(0, 0, 2.201)	(1.33, 1, 2)	(39, 31, 51)	(29, 31, 26)	0.7087	0.9473	0.3910	316.00	262.77
(3.824, 2.704, 2.208)	(0, 0, 2.208)	(1.33, 1, 2.5)	(39, 31, 60)	(29, 31, 24)	0.7105	0.9511	0.3869	319.12	263.39
(3.797, 2.685, 2.192)	(0, 0, 2.192)	(1.33, 1.33, 1)	(39, 39, 31)	(29, 29, 31)	0.6826	0.9316	0.3667	311.91	260.37
(3.805, 2.690, 2.197)	(0, 0, 2.197)	(1.33, 1.33, 1.33)	(39, 39, 39)	(29, 29, 29)	0.7086	0.9458	0.3941	314.21	260.99
(3.813, 2.696, 2.201)	(0, 0, 2.201)	(1.33, 1.33, 1.5)	(39, 39, 42)	(29, 29, 28)	0.7090	0.9474	0.3917	315.23	261.79

Table A.5 (continued): Evaluating different allocation ratios in terms of resultant powers (Scenario 2).

Stopping boundaries		**Allocation ratios**	**Sample size**		**Power**			**ESS**	
(e_1, e_2, e_3)	(f_1, f_2, f_3)	(r_1, r_2, r_3)	$(n_{1(0)}, n_{2(0)}, n_{3(0)})$	$(n_{1(k)}, n_{2(k)}, n_{3(k)})$	P_1	P_{dis}	P_{con}	H_0	H_a
(3.820, 2.701, 2.206)	(0, 0, 2.206)	(1.33, 1.33, 2)	(39, 39, 51)	(29, 29, 26)	0.7124	0.9520	0.3888	318.74	262.20
(3.828, 2.706, 2.210)	(0, 0, 2.210)	(1.33, 1.33, 2.5)	(39, 39, 60)	(29, 29, 24)	0.7150	0.9557	0.3865	321.75	262.76
(3.801, 2.688, 2.195)	(0, 0, 2.195)	(1.33, 1.5, 1)	(39, 42, 31)	(29, 28, 31)	0.6823	0.9334	0.3626	312.15	260.28
(3.809, 2.693, 2.199)	(0, 0, 2.199)	(1.33, 1.5, 1.33)	(39, 42, 39)	(29, 28, 29)	0.7083	0.9473	0.3902	315.27	260.86
(3.816, 2.698, 2.203)	(0, 0, 2.203)	(1.33, 1.5, 1.5)	(39, 42, 42)	(29, 28, 28)	0.7087	0.9488	0.3879	315.50	260.99
(3.824, 2.704, 2.208)	(0, 0, 2.208)	(1.33, 1.5, 2)	(39, 42, 51)	(29, 28, 26)	0.7119	0.9533	0.3851	319.14	261.96
(3.831, 2.709, 2.212)	(0, 0, 2.212)	(1.33, 1.5, 2.5)	(39, 42, 60)	(29, 28, 24)	0.7144	0.9568	0.3826	323.04	261.32
(3.813, 2.696, 2.201)	(0, 0, 2.201)	(1.33, 2, 1)	(39, 51, 31)	(29, 26, 31)	0.6816	0.9381	0.3525	316.07	261.04
(3.820, 2.701, 2.206)	(0, 0, 2.206)	(1.33, 2, 1.33)	(39, 51, 39)	(29, 26, 29)	0.7068	0.9508	0.3797	318.09	261.25
(3.824, 2.704, 2.208)	(0, 0, 2.208)	(1.33, 2, 1.5)	(39, 51, 42)	(29, 26, 28)	0.7078	0.9523	0.3783	318.79	261.99
(3.831, 2.709, 2.212)	(0, 0, 2.212)	(1.33, 2, 2)	(39, 51, 51)	(29, 26, 26)	0.7109	0.9564	0.3759	322.30	262.54
(3.831, 2.709, 2.212)	(0, 0, 2.212)	(1.33, 2, 2.5)	(39, 51, 60)	(29, 26, 24)	0.7147	0.9600	0.3757	325.62	262.23
(3.816, 2.698, 2.203)	(0, 0, 2.203)	(1.33, 2.5, 1)	(39, 60, 31)	(29, 24, 31)	0.6823	0.9423	0.3457	319.41	261.26
(3.824, 2.704, 2.208)	(0, 0, 2.208)	(1.33, 2.5, 1.33)	(39, 60, 39)	(29, 24, 29)	0.7068	0.9540	0.3728	321.64	262.12
(3.828, 2.706, 2.210)	(0, 0, 2.210)	(1.33, 2.5, 1.5)	(39, 60, 42)	(29, 24, 28)	0.7078	0.9554	0.3716	322.97	262.51
(3.835, 2.712, 2.214)	(0, 0, 2.214)	(1.33, 2.5, 2)	(39, 60, 51)	(29, 24, 26)	0.7105	0.9590	0.3691	326.44	263.09

Table A.5 (continued): Evaluating different allocation ratios in terms of resultant powers (Scenario 2).

Stopping boundaries		**Allocation ratios**	**Sample size**		**Power**			**ESS**	
(e_1, e_2, e_3)	(f_1, f_2, f_3)	(r_1, r_2, r_3)	$(n_{1(0)}, n_{2(0)}, n_{3(0)})$	$(n_{1(k)}, n_{2(k)}, n_{3(k)})$	P_1	P_{dis}	P_{con}	H_0	H_a
(3.835, 2.712, 2.219)	(0, 0, 2.219)	(1.33, 2.5, 2.5)	(39, 60, 60)	(29, 24, 24)	0.7127	0.9619	0.3669	330.02	263.07
(3.794, 2.682, 2.190)	(0, 0, 2.190)	(1.5, 1, 1)	(42, 31, 31)	(28, 31, 31)	0.6755	0.9255	0.3627	308.43	257.88
(3.801, 2.688, 2.195)	(0, 0, 2.195)	(1.5, 1, 1.33)	(42, 31, 39)	(28, 31, 29)	0.7031	0.9414	0.3911	311.24	258.66
(3.813, 2.696, 2.201)	(0, 0, 2.201)	(1.5, 1, 1.5)	(42, 31, 42)	(28, 31, 28)	0.7029	0.9428	0.3878	312.19	259.40
(3.816, 2.698, 2.203)	(0, 0, 2.203)	(1.5, 1, 2)	(42, 31, 51)	(28, 31, 26)	0.7077	0.9486	0.3863	315.62	259.89
(3.824, 2.704, 2.208)	(0, 0, 2.208)	(1.5, 1, 2.5)	(42, 31, 60)	(28, 31, 24)	0.7101	0.9524	0.3830	318.97	260.30
(3.801, 2.688, 2.195)	(0, 0, 2.195)	(1.5, 1.33, 1)	(42, 39, 31)	(28, 29, 31)	0.6817	0.9332	0.3617	311.21	258.06
(3.809, 2.693, 2.199)	(0, 0, 2.199)	(1.5, 1.33, 1.33)	(42, 39, 39)	(28, 29, 29)	0.7078	0.9471	0.3896	314.45	258.80
(3.813, 2.696, 2.201)	(0, 0, 2.201)	(1.5, 1.33, 1.5)	(42, 39, 42)	(28, 29, 28)	0.7089	0.9489	0.3883	315.03	258.59
(3.820, 2.701, 2.206)	(0, 0, 2.206)	(1.5, 1.33, 2)	(42, 39, 51)	(28, 29, 26)	0.7122	0.9534	0.3854	317.94	259.62
(3.828, 2.706, 2.210)	(0, 0, 2.210)	(1.5, 1.33, 2.5)	(42, 39, 60)	(28, 29, 24)	0.7147	0.9569	0.3830	321.94	260.11
(3.809, 2.693, 2.199)	(0, 0, 2.199)	(1.5, 1.5, 1)	(42, 42, 31)	(28, 28, 31)	0.6809	0.9348	0.3572	311.72	257.83
(3.816, 2.698, 2.203)	(0, 0, 2.203)	(1.5, 1.5, 1.33)	(42, 42, 39)	(28, 28, 29)	0.7068	0.9483	0.3848	314.94	258.09
(3.820, 2.701, 2.206)	(0, 0, 2.206)	(1.5, 1.5, 1.5)	(42, 42, 42)	(28, 28, 28)	0.7076	0.9499	0.3832	315.48	258.40
(3.824, 2.704, 2.208)	(0, 0, 2.208)	(1.5, 1.5, 2)	(42, 42, 51)	(28, 28, 26)	0.7117	0.9546	0.3817	318.69	258.87
(3.828, 2.706, 2.210)	(0, 0, 2.210)	(1.5, 1.5, 2.5)	(42, 42, 60)	(28, 28, 24)	0.7149	0.9582	0.3803	322.38	259.11

Table A.5 (continued): Evaluating different allocation ratios in terms of resultant powers (Scenario 2).

Stopping boundaries		**Allocation ratios**	**Sample size**		**Power**			**ESS**	
(e_1, e_2, e_3)	(f_1, f_2, f_3)	(r_1, r_2, r_3)	$(n_{1(0)}, n_{2(0)}, n_{3(0)})$	$(n_{1(k)}, n_{2(k)}, n_{3(k)})$	P_1	P_{dis}	P_{con}	H_0	H_a
(3.816, 2.698, 2.203)	(0, 0, 2.203)	(1.5, 2, 1)	(42, 51, 31)	(28, 26, 31)	0.6812	0.9397	0.3486	315.09	258.88
(3.820, 2.701, 2.206)	(0, 0, 2.206)	(1.5, 2, 1.33)	(42, 51, 39)	(28, 26, 29)	0.7070	0.9523	0.3768	318.37	258.95
(3.824, 2.704, 2.208)	(0, 0, 2.208)	(1.5, 2, 1.5)	(42, 51, 42)	(28, 26, 28)	0.7079	0.9538	0.3754	318.60	259.25
(3.828, 2.706, 2.210)	(0, 0, 2.210)	(1.5, 2, 2)	(42, 51, 51)	(28, 26, 26)	0.7117	0.9580	0.3740	322.42	259.16
(3.835, 2.712, 2.214)	(0, 0, 2.214)	(1.5, 2, 2.5)	(42, 51, 60)	(28, 26, 24)	0.7138	0.9609	0.3714	325.79	259.96
(3.820, 2.701, 2.206)	(0, 0, 2.206)	(1.5, 2.5, 1)	(42, 60, 31)	(28, 24, 31)	0.6816	0.9437	0.3417	318.69	258.93
(3.828, 2.706, 2.210)	(0, 0, 2.210)	(1.5, 2.5, 1.33)	(42, 60, 39)	(28, 24, 29)	0.7064	0.9551	0.3692	321.93	259.68
(3.831, 2.709, 2.212)	(0, 0, 2.212)	(1.5, 2.5, 1.5)	(42, 60, 42)	(28, 24, 28)	0.7072	0.9565	0.3679	322.62	259.81
(3.839, 2.714, 2.216)	(0, 0, 2.216)	(1.5, 2.5, 2)	(42, 60, 51)	(28, 24, 26)	0.7099	0.9600	0.3654	325.75	260.44
(3.843, 2.717, 2.219)	(0, 0, 2.219)	(1.5, 2.5, 2.5)	(42, 60, 60)	(28, 24, 24)	0.7123	0.9628	0.3636	330.09	261.10
(3.809, 2.693, 2.199)	(0, 0, 2.199)	(2, 1, 1.33)	(51, 31, 39)	(26, 31, 29)	0.7013	0.9452	0.3797	313.62	255.01
(3.813, 2.696, 2.201)	(0, 0, 2.201)	(2, 1, 1.5)	(51, 31, 42)	(26, 31, 28)	0.7026	0.9471	0.3784	314.38	255.51
(3.824, 2.704, 2.208)	(0, 0, 2.208)	(2, 1, 2)	(51, 31, 51)	(26, 31, 26)	0.7051	0.9515	0.3744	317.48	256.18
(3.831, 2.709, 2.212)	(0, 0, 2.212)	(2, 1, 2.5)	(51, 31, 60)	(26, 31, 24)	0.7076	0.9551	0.3718	321.37	256.98
(3.813, 2.696, 2.201)	(0, 0, 2.201)	(2, 1.33, 1)	(51, 39, 31)	(26, 29, 31)	0.6796	0.9373	0.3499	313.38	254.92
(3.816, 2.698, 2.203)	(0, 0, 2.203)	(2, 1.33, 1.33)	(51, 39, 39)	(26, 29, 29)	0.7063	0.9506	0.3790	317.37	255.33

Table A.5 (continued): Evaluating different allocation ratios in terms of resultant powers (Scenario 2).

Stopping boundaries		**Allocation ratios**	**Sample size**		**Power**			**ESS**	
(e_1, e_2, e_3)	(f_1, f_2, f_3)	(r_1, r_2, r_3)	$(n_{1(0)}, n_{2(0)}, n_{3(0)})$	$(n_{1(k)}, n_{2(k)}, n_{3(k)})$	P_1	P_{dis}	P_{con}	H_0	H_a
(3.820, 2.701, 2.206)	(0, 0, 2.206)	(2, 1.33, 1.5)	(51, 39, 42)	(26, 29, 28)	0.7070	0.9521	0.3773	317.27	255.80
(3.828, 2.706, 2.210)	(0, 0, 2.210)	(2, 1.33, 2)	(51, 39, 51)	(26, 29, 26)	0.7102	0.9562	0.3748	320.49	256.18
(3.835, 2.712, 2.214)	(0, 0, 2.214)	(2, 1.33, 2.5)	(51, 39, 60)	(26, 29, 24)	0.7124	0.9593	0.3723	324.33	256.34
(3.816, 2.698, 2.203)	(0, 0, 2.203)	(2, 1.5, 1)	(51, 42, 31)	(26, 28, 31)	0.6798	0.9391	0.3468	314.47	254.26
(3.820, 2.701, 2.206)	(0, 0, 2.206)	(2, 1.5, 1.33)	(51, 42, 39)	(26, 28, 29)	0.7059	0.9519	0.3752	316.69	254.95
(3.824, 2.704, 2.208)	(0, 0, 2.208)	(2, 1.5, 1.5)	(51, 42, 42)	(26, 28, 28)	0.7068	0.9534	0.3739	318.02	254.66
(3.831, 2.709, 2.212)	(0, 0, 2.212)	(2, 1.5, 2)	(51, 42, 51)	(26, 28, 26)	0.7098	0.9574	0.3714	320.76	255.81
(3.839, 2.714, 2.216)	(0, 0, 2.216)	(2, 1.5, 2.5)	(51, 42, 60)	(26, 28, 24)	0.7120	0.9604	0.3690	324.53	256.15
(3.820, 2.701, 2.206)	(0, 0, 2.206)	(2, 2, 1)	(51, 51, 31)	(26, 26, 31)	0.6810	0.9439	0.3399	318.00	255.24
(3.828, 2.706, 2.210)	(0, 0, 2.210)	(2, 2, 1.33)	(51, 51, 39)	(26, 26, 29)	0.7059	0.9555	0.3675	320.87	255.48
(3.831, 2.709, 2.212)	(0, 0, 2.212)	(2, 2, 1.5)	(51, 51, 42)	(26, 26, 28)	0.7067	0.9568	0.3662	321.37	256.20
(3.839, 2.714, 2.216)	(0, 0, 2.216)	(2, 2, 2)	(51, 51, 51)	(26, 26, 26)	0.7094	0.9603	0.3636	325.30	256.31
(3.843, 2.717, 2.219)	(0, 0, 2.219)	(2, 2, 2.5)	(51, 51, 60)	(26, 26, 24)	0.7117	0.9631	0.3618	329.13	257.13
(3.828, 2.706, 2.210)	(0, 0, 2.210)	(2, 2.5, 1)	(51, 60, 31)	(26, 24, 31)	0.6811	0.9474	0.3331	321.90	256.48
(3.835, 2.712, 2.214)	(0, 0, 2.214)	(2, 2.5, 1.33)	(51, 60, 39)	(26, 24, 29)	0.7055	0.9581	0.3606	324.51	256.71
(3.839, 2.714, 2.216)	(0, 0, 2.216)	(2, 2.5, 1.5)	(51, 60, 42)	(26, 24, 28)	0.7062	0.9593	0.3594	325.39	256.73

Table A.5 (continued): Evaluating different allocation ratios in terms of resultant powers (Scenario 2).

Stopping boundaries		**Allocation ratios**	**Sample size**		**Power**			**ESS**	
(e_1, e_2, e_3)	(f_1, f_2, f_3)	(r_1, r_2, r_3)	$(n_{1(0)}, n_{2(0)}, n_{3(0)})$	$(n_{1(k)}, n_{2(k)}, n_{3(k)})$	P_1	P_{dis}	P_{con}	H_0	H_a
(3.847, 2.720, 2.221)	(0, 0, 2.221)	(2, 2.5, 2)	(51, 60, 51)	(26, 24, 26)	0.7083	0.9623	0.3564	329.05	257.33
(3.850, 2.722, 2.223)	(0, 0, 2.223)	(2, 2.5, 2.5)	(51, 60, 60)	(26, 24, 24)	0.7108	0.9650	0.3550	333.17	257.70
(3.813, 2.696, 2.201)	(0, 0, 2.201)	(2.5, 1, 1)	(60, 31, 31)	(24, 31, 31)	0.6714	0.9337	0.3398	311.95	254.42
(3.816, 2.698, 2.203)	(0, 0, 2.203)	(2.5, 1, 1.33)	(60, 31, 39)	(24, 31, 29)	0.6997	0.9482	0.3701	315.34	254.87
(3.824, 2.704, 2.208)	(0, 0, 2.208)	(2.5, 1, 1.5)	(60, 31, 42)	(24, 31, 28)	0.6997	0.9495	0.3674	315.55	255.45
(3.828, 2.706, 2.210)	(0, 0, 2.210)	(2.5, 1, 2)	(60, 31, 51)	(24, 31, 26)	0.7039	0.9542	0.3660	319.19	255.49
(3.835, 2.712, 2.214)	(0, 0, 2.214)	(2.5, 1, 2.5)	(60, 31, 60)	(24, 31, 24)	0.7062	0.9575	0.3634	323.10	256.31
(3.820, 2.701, 2.206)	(0, 0, 2.206)	(2.5, 1.33, 1)	(60, 39, 31)	(24, 29, 31)	0.6781	0.9406	0.3405	315.23	255.09
(3.824, 2.704, 2.208)	(0, 0, 2.208)	(2.5, 1.33, 1.33)	(60, 39, 39)	(24, 29, 29)	0.7044	0.9531	0.3696	318.09	255.18
(3.831, 2.709, 2.212)	(0, 0, 2.212)	(2.5, 1.33, 1.5)	(60, 39, 42)	(24, 29, 28)	0.7046	0.9544	0.3672	319.51	255.97
(3.835, 2.712, 2.214)	(0, 0, 2.214)	(2.5, 1.33, 2)	(60, 39, 51)	(24, 29, 26)	0.7082	0.9583	0.3658	322.55	255.40
(3.843, 2.717, 2.219)	(0, 0, 2.219)	(2.5, 1.33, 2.5)	(60, 39, 60)	(24, 29, 24)	0.7100	0.9611	0.3630	326.24	256.44
(3.824, 2.704, 2.208)	(0, 0, 2.208)	(2.5, 1.5, 1)	(60, 42, 31)	(24, 28, 31)	0.6783	0.9423	0.3376	316.99	254.66
(3.828, 2.706, 2.210)	(0, 0, 2.210)	(2.5, 1.5, 1.33)	(60, 42, 39)	(24, 28, 29)	0.7044	0.9545	0.3665	319.33	254.65
(3.831, 2.709, 2.212)	(0, 0, 2.212)	(2.5, 1.5, 1.5)	(60, 42, 42)	(24, 28, 28)	0.7053	0.9559	0.3652	319.81	255.46
(3.835, 2.712, 2.214)	(0, 0, 2.214)	(2.5, 1.5, 2)	(60, 42, 51)	(24, 28, 26)	0.7088	0.9597	0.3638	323.55	255.52

Table A.5 (continued): Evaluating different allocation ratios in terms of resultant powers (Scenario 2).

Stopping boundaries		**Allocation ratios**	**Sample size**		**Power**			**ESS**	
(e_1, e_2, e_3)	(f_1, f_2, f_3)	(r_1, r_2, r_3)	$(n_{1(0)}, n_{2(0)}, n_{3(0)})$	$(n_{1(k)}, n_{2(k)}, n_{3(k)})$	P_1	P_{dis}	P_{con}	H_0	H_a
(3.843, 2.717, 2.219)	(0, 0, 2.219)	(2.5, 1.5, 2.5)	(60, 42, 60)	(24, 28, 24)	0.7105	0.9623	0.3611	327.06	256.00
(3.831, 2.709, 2.212)	(0, 0, 2.212)	(2.5, 2, 1)	(60, 51, 31)	(24, 26, 31)	0.6792	0.9466	0.3308	320.40	255.62
(3.835, 2.712, 2.214)	(0, 0, 2.214)	(2.5, 2, 1.33)	(60, 51, 39)	(24, 26, 29)	0.7046	0.9578	0.3595	323.35	255.83
(3.839, 2.714, 2.216)	(0, 0, 2.216)	(2.5, 2, 1.5)	(60, 51, 42)	(24, 26, 28)	0.7054	0.9590	0.3582	323.83	256.29
(3.847, 2.720, 2.221)	(0, 0, 2.221)	(2.5, 2, 2)	(60, 51, 51)	(24, 26, 26)	0.7075	0.9621	0.3552	327.31	256.71
(3.850, 2.722, 2.223)	(0, 0, 2.223)	(2.5, 2, 2.5)	(60, 51, 60)	(24, 26, 24)	0.7100	0.9648	0.3540	331.65	257.13
(3.835, 2.712, 2.214)	(0, 0, 2.214)	(2.5, 2.5, 1)	(60, 60, 31)	(24, 24, 31)	0.6802	0.9501	0.3257	323.81	256.04
(3.839, 2.714, 2.216)	(0, 0, 2.216)	(2.5, 2.5, 1.33)	(60, 60, 39)	(24, 24, 29)	0.7052	0.9605	0.3545	326.85	256.39
(3.843, 2.717, 2.219)	(0, 0, 2.219)	(2.5, 2.5, 1.5)	(60, 60, 42)	(24, 24, 28)	0.7056	0.9615	0.3527	327.87	257.03
(3.847, 2.720, 2.221)	(0, 0, 2.221)	(2.5, 2.5, 2)	(60, 60, 51)	(24, 24, 26)	0.7084	0.9646	0.3512	331.57	257.12
(3.850, 2.722, 2.223)	(0, 0, 2.223)	(2.5, 2.5, 2.5)	(60, 60, 60)	(24, 24, 24)	0.7108	0.9669	0.3500	335.47	257.61

A.3 R code

- R code for Chapter 2 to reproduce the results is available from `https://github.com/Ruqayya20/allocation_methods-in-multi-arm-trial`
- R code for Chapter 3 to reproduce the results is available from `https://github.com/Ruqayya20/MAMS/blob/main/stage/wise-allocation-ratios`.
- R code for Chapter 4 to reproduce the results is available from `https://github.com/Ruqayya20/platform-trials.git`.

Bibliography

Amberson Jr, J. B., McMahon, B., and Pinner, M. (1931). A clinical trial of sanocrysin in pulmonary tuberculosis. *American Review of Tuberculosis*, 24(4):401–435.

Antognini, A. B. and Giovagnoli, A. (2004). A new 'biased coin design'for the sequential allocation of two treatments. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 53(4):651–664.

Atkinson, A. and Donev, A. (1992). *Optimum experimental designs*. Clarendon Press.

Atkinson, A. C. (2002). The comparison of designs for sequential clinical trials with covariate information. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 165(2):349–373.

Atkinson, A. C. and Biswas, A. (2013). *Randomised response-adaptive designs in clinical trials*. CRC Press.

Azher, R. A., Wason, J. M., and Grayling, M. J. (2023). A comparison of randomization methods for multi-arm clinical trials. *Statistics in Biopharmaceutical Research*, pages 1–13.

Bailey, R. and Nelson, P. (2003). Hadamard randomization: a valid restriction of random permuted blocks. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 45(5):554–560.

Baldi Antognini, A. and Zagoraiou, M. (2011). The covariate-adaptive biased coin design for balancing clinical trials in the presence of prognostic factors. *Biometrika*, 98(3):519–535.

Bauer, P., Bretz, F., Dragalin, V., König, F., and Wassmer, G. (2016). Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls. *Statistics in Medicine*, 35(3):325–347.

Begg, C. B. and Iglewicz, B. (1980). A treatment allocation procedure for sequential clinical trials. *Biometrics*, pages 81–90.

Bennett, M. and Mander, A. P. (2020). Designs for adding a treatment arm to an ongoing clinical trial. *Trials*, 21(1):1–12.

Berger, V. (2005). Selection bias and covariate imbalances. In *Randomized Clincal Trials*. John Wiley & Sons.

Berger, V. W., Bejleri, K., and Agnor, R. (2016). Comparing mti randomization procedures to blocked randomization. *Statistics in medicine*, 35(5):685–694.

Berger, V. W., Bour, L. J., Carter, K., Chipman, J. J., Everett, C. C., Heussen, N., Hewitt, C., Hilgers, R.-D., Luo, Y. A., Renteria, J., et al. (2021). A roadmap to using randomization in clinical trials. *BMC Medical Research Methodology*, 21:1–24.

Berger, V. W., Ivanova, A., and Deloria Knoll, M. (2003). Minimizing predictability while retaining balance through the use of less restrictive randomization procedures. *Statistics in medicine*, 22(19):3017–3028.

Berry, S. M., Connor, J. T., and Lewis, R. J. (2015). The platform trial: an efficient strategy for evaluating multiple treatments. *Jama*, 313(16):1619–1620.

Blackwell, D. and Hodges Jr, J. (1957). Design for the control of selection bias. *The Annals of Mathematical Statistics*, 28(2):449–460.

Bofill Roig, M., Glimm, E., Mielke, T., and Posch, M. (2023). Optimal allocation strategies in platform trials. *arXiv e-prints*, pages arXiv–2304.

Brown, S., Thorpe, H., Hawkins, K., and Brown, J. (2005). Minimization—reducing predictability for multi-centre trials whilst retaining balance within centre. *Statistics in medicine*, 24(24):3715–3727.

Bruce, C. L., Iflaifel, M., Montgomery, A., Ogollah, R., Sprange, K., and Partlett, C. (2024). Choosing and evaluating randomisation methods in clinical trials: a qualitative study. *Trials*, 25(1):199.

Bruce, C. L., Juszczak, E., Ogollah, R., Partlett, C., and Montgomery, A. (2022). A systematic review of randomisation method use in rcts and association of trial design characteristics with method selection. *BMC Medical Research Methodology*, 22(1):314.

Burant, C. F., Viswanathan, P., Marcinak, J., Cao, C., Vakilynejad, M., Xie, B., and Leifke, E. (2012). Tak-875 versus placebo or glimepiride in type 2 diabetes mellitus: a phase 2, randomised, double-blind, placebo-controlled trial. *The Lancet*, 379(9824):1403–1411.

Burnett, T., Mozgunov, P., Pallmann, P., Villar, S. S., Wheeler, G. M., and Jaki, T. (2020). Adding flexibility to clinical trial designs: an example-based guide to the practical use of adaptive designs. *BMC medicine*, 18(1):1–21.

Byar, D. P., Simon, R. M., Friedewald, W. T., Schlesselman, J. J., DeMets, D. L., Ellenberg, J. H., Gail, M. H., and Ware, J. H. (1976). Randomized clinical trials: perspectives on some recent ideas. *New England Journal of Medicine*, 295(2):74–80.

Chen, Y.-P. (1999). Biased coin design with imbalance tolerance. *Stochastic Models*, 15(5):953–975.

Chen, Y.-P. (2000). Which design is better? ehrenfest urn versus biased coin. *Advances in Applied Probability*, 32(3):738–749.

Choodari-Oskooei, B., Bratton, D. J., Gannon, M. R., Meade, A. M., Sydes, M. R., and Parmar, M. K. (2020). Adding new experimental arms to randomised clinical trials: impact on error rates. *Clinical Trials*, 17(3):273–284.

Chow, S. C. and Chang, M. (2006). *Adaptive design methods in clinical trials*. Chapman and Hall/CRC.

Chow, S. C. and Liu, J. p. (2008). *Design and analysis of clinical trials: concepts and methodologies*, volume 507. John Wiley & Sons.

Cochran, W. G. (1939). The use of the analysis of variance in enumeration by sampling. *Journal of the American Statistical Association*, 34(207):492–510.

Cohen, D. R., Todd, S., Gregory, W. M., and Brown, J. M. (2015). Adding a treatment arm to an ongoing clinical trial: a review of methodology and practice. *Trials*, 16(1):1–9.

Demets, D. L. and Lan, K. G. (1994). Interim analysis: the alpha spending function approach. *Statistics in medicine*, 13(13-14):1341–1352.

Diacon, A. H., Dawson, R., von Groote-Bidlingmaier, F., Symons, G., Venter, A., Donald, P. R., van Niekerk, C., Everitt, D., Winter, H., Becker, P., et al. (2012). 14-day bactericidal activity of pa-824, bedaquiline, pyrazinamide, and moxifloxacin combinations: a randomised trial. *The Lancet*, 380(9846):986–993.

Dmitrienko, A., Tamhane, A., and Bretz, F. (2009). *Multiple testing problems in pharmaceutical statistics*. CRC Press.

Dumville, J., Hahn, S., Miles, J., and Torgerson, D. (2006). The use of unequal randomisation ratios in clinical trials: A review. *Contemporary Clinical Trials*, 27(1):1–12.

Dunnett, C. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, 50(272):1096–121.

Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika*, 58(3):403–417.

Elm, J. J., Palesch, Y. Y., Koch, G. G., Hinson, V., Ravina, B., and Zhao, W. (2012). Flexible analytical methods for adding a treatment arm mid-study to an ongoing clinical trial. *Journal of biopharmaceutical statistics*, 22(4):758–772.

Fisher, R. (1926). Introduction to" the arrangement of field experiments. *J. Minist. Agric. GB*, 33:503–513.

Fisher, R. A. (1992). The arrangement of field experiments. In *Breakthroughs in statistics*, pages 82–91. Springer.

Genz, A. and Bretz, F. (2002). Comparison of methods for the computation of multivariate t probabilities. *Journal of Computational and Graphical Statistics*, 11(4):950–71.

Gordon Lan, K. and DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika*, 70(3):659–663.

Grayling, M., Wason, J., and AP, M. (2017). Efficient determination of optimised multi-arm multi-stage experimental designs with control of generalised error-rates. *arXiV*, page 1712.00229.

Grayling, M. J., Wason, J. M., and Mander, A. P. (2018). Group sequential clinical trial designs for normally distributed outcome variables. *The Stata Journal*, 18(2):416–431.

Greene, T. (2015). Randomized controlled trials 5: determining the sample size and power for clinical trials and cohort studies. *Clinical Epidemiology: Practice and Methods*, pages 225–247.

Han, B., Enas, N. H., and McEntegart, D. (2009). Randomization by minimization for unbalanced treatment allocation. *Statistics in medicine*, 28(27):3329–3346.

Heritier, S., Gebski, V., and Pillai, A. (2005). Dynamic balancing randomization in controlled clinical trials. *Statistics in medicine*, 24(24):3729–3741.

Horby, P. W., Mafham, M., Bell, J. L., Linsell, L., Staplin, N., Emberson, J., Palfreeman, A., Raw, J., Elmahi, E., Prudon, B., et al. (2020). Lopinavir–ritonavir in patients admitted to hospital with covid-19 (recovery): a randomised, controlled, open-label, platform trial. *The Lancet*, 396(10259):1345–1352.

Howard, D. R., Hockaday, A., Brown, J. M., Gregory, W. M., Todd, S., Munir, T., Oughton, J. B., Dimbleby, C., and Hillmen, P. (2021). A platform trial in practice: adding a new experimental research arm to the ongoing confirmatory flair trial in chronic lymphocytic leukaemia. *Trials*, 22:1–13.

Hu, F., Hu, Y., Ma, Z., and Rosenberger, W. F. (2014). Adaptive randomization for balancing over covariates. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(4):288–303.

Hu, F. and Rosenberger, W. F. (2006). *The theory of response-adaptive randomization in clinical trials*. John Wiley & Sons.

Jaki, T. (2015). Multi-arm clinical trials with treatment selection: What can be gained and at what price? *Clinical Investigation*, 5(4):393–9.

Jennison, C. and Turnbull, B. W. (1999). *Group sequential methods with applications to clinical trials*. CRC Press.

Jin, M., Polis, A., and Hartzel, J. (2019). Algorithms for minimization randomization and the implementation with an r package. *Communications in Statistics-Simulation and Computation*, 50(10):3077–3087.

Jones, B. and Kenward, M. G. (2003). *Design and analysis of cross-over trials*. Chapman and Hall/CRC.

Juszczak, E., Altman, D. G., Hopewell, S., and Schulz, K. (2019). Reporting of multi-arm parallel-group randomized trials: extension of the consort 2010 statement. *Jama*, 321(16):1610–1620.

Kelly, P. J., Stallard, N., and Todd, S. (2005). An adaptive group sequential design for phase ii/iii clinical trials that select a single treatment from several. *Journal of biopharmaceutical statistics*, 15(4):641–658.

Kernan, W. N., Viscoli, C. M., Makuch, R. W., Brass, L. M., and Horwitz, R. I. (1999). Stratified randomization for clinical trials. *Journal of clinical epidemiology*, 52(1):19–26.

Kuznetsova, O. M. and Tymofyeyev, Y. (2012). Preserving the allocation ratio at every allocation with biased coin randomization and minimization in studies with unequal allocation. *Statistics in Medicine*, 31(8):701–723.

Kuznetsova, O. M. and Tymofyeyev, Y. (2015). Brick tunnel and wide brick tunnel randomization for studies with unequal allocation. *Modern Adaptive Randomized Clinical Trials: Statistical, Operational, and Regulatory Aspects*, pages 83–114.

Lauzon, S. D., Zhao, W., Nietert, P. J., Ciolino, J. D., Hill, M. D., and Ramakrishnan, V. (2022). Impact of minimal sufficient balance, minimization, and stratified permuted blocks on bias and power in the estimation of treatment effect in sequential clinical trials with a binary endpoint. *Statistical Methods in Medical Research*, 31(1):184–204.

Lin, Y., Zhu, M., and Su, Z. (2015). The pursuit of balance: an overview of covariate-adaptive randomization techniques in clinical trials. *Contemporary clinical trials*, 45:21–25.

Liu, Y. and Hu, F. (2022). Balancing unobserved covariates with covariate-adaptive randomized experiments. *Journal of the American Statistical Association*, 117(538):875–886.

Ma, Z. and Hu, F. (2013). Balancing continuous covariates based on kernel densities. *Contemporary clinical trials*, 34(2):262–269.

Magirr, D., Jaki, T., and Whitehead, J. (2012). A generalized dunnett test for multi-arm multi-stage clinical studies with treatment selection. *Biometrika*, 99(2):494–501.

Matthews, J. N. (2006). *Introduction to randomized controlled clinical trials*. Chapman and Hall/CRC.

Matts, J. P. and Lachin, J. M. (1988). Properties of permuted-block randomization in clinical trials. *Controlled clinical trials*, 9(4):327–344.

McPherson, G. C., Campbell, M. K., and Elbourne, D. R. (2012). Use of randomisation in clinical trials: a survey of uk practice. *Trials*, 13(1):1–7.

Medical Research Council (1948). Streptomycin treatment of pulmonary tuberculosis. *Bmj*, 2(4582):769–82.

Möcks, J., Köhler, W., Scott, M., Maurer, J., Budde, M., and Givens, S. (2002). Dealing with selective dropout in clinical trials. *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry*, 1(2):119–130.

Mossop, H., Grayling, M. J., Gallagher, F. A., Welsh, S. J., Stewart, G. D., and Wason, J. M. (2022). Advantages of multi-arm non-randomised sequentially allocated cohort designs for phase ii oncology trials. *British Journal of Cancer*, 126(2):204–210.

Neuhäuser, M., Mackowiak, M., and Ruxton, G. (2021). Unequal sample sizes according to the square-root allocation rule are useful when comparing several treatments with a control. *Ethology*, 127(12):1094–100.

O'Brien, P. C. and Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics*, pages 549–556.

Pallmann, P., Bedding, A. W., Choodari-Oskooei, B., Dimairo, M., Flight, L., Hampson, L. V., Holmes, J., Mander, A. P., Odondi, L., Sydes, M. R., et al. (2018). Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC medicine*, 16:1–15.

Park, J. J., Harari, O., Dron, L., Lester, R. T., Thorlund, K., and Mills, E. J. (2020). An overview of platform trials with a checklist for clinical readers. *Journal of clinical epidemiology*, 125:1–8.

Park, J. J., Mills, E. J., and Wathen, J. K. (2023). *Introduction to Adaptive Trial Designs and Master Protocols*. Cambridge University Press.

Park, J. J., Siden, E., Zoratti, M. J., Dron, L., Harari, O., Singer, J., Lester, R. T., Thorlund, K., and Mills, E. J. (2019). Systematic review of basket trials, umbrella trials, and platform trials: a landscape analysis of master protocols. *Trials*, 20:1–10.

Parmar, M. K., Carpenter, J., and Sydes, M. R. (2014). More multiarm randomised trials of superiority are needed. *The Lancet*, 384(9940):283–284.

Parmar, M. K., Sydes, M. R., Cafferty, F. H., Choodari-Oskooei, B., Langley, R. E., Brown, L., Phillips, P. P., Spears, M. R., Rowley, S., Kaplan, R., et al. (2017). Testing many treatments within a single protocol over 10 years at mrc clinical trials unit at ucl: Multi-arm, multi-stage platform, umbrella and basket protocols. *Clinical trials*, 14(5):451–461.

Peace, K. E. and Chen, D. (2010). Clinical trial methodology (chapman & hall crc biostatistics series).

Pitre, T., Cheng, S., Cusano, E., Khan, N., Mikhail, D., Leung, G., Vernooij, R. W., Yarnell, C. J., Goligher, E., Murthy, S., et al. (2023). Methodology and design of platform trials: a meta-epidemiological study. *Journal of clinical epidemiology*, 157:1–12.

Pocock, S. J. (2013). *Clinical trials: a practical approach*. John Wiley & Sons.

Pocock, S. J. and Simon, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*, pages 103–115.

Pushpakom, S., Kolamunnage-Dona, R., Taylor, C., Foster, T., Spowart, C., García-Fiñana, M., Kemp, G., Jaki, T., Khoo, S., Williamson, P., Pirmohamed, M., and TAILoR Study Group (2020). TAILoR (TelmisArtan and InsuLin Resistance in Human Immunodeficiency Virus [HIV]): An adaptive-design, dose-ranging phase IIb randomized trial of telmisartan for the reduction of insulin resistance in HIV-positive individuals on combination antiretroviral therapy. *Clinical Infectious Diseases*, 70(10):2062–72.

Ren, Y., Li, X., and Chen, C. (2021). Statistical considerations of phase 3 umbrella trials allowing adding one treatment arm mid-trial. *Contemporary Clinical Trials*, 109:106538.

Rosenberger, W. F. and Lachin, J. M. (2015). *Randomization in clinical trials: theory and practice*. John Wiley & Sons.

Rosenberger, W. F., Uschner, D., and Wang, Y. (2018). Randomization: The forgotten component of the randomized clinical trial. *Statistics in medicine*, 38(1):1–12.

Royston, P., Parmar, M., and Qian, W. (2003). Novel designs for multi-arm clinical trials with survival outcomes with an application in ovarian cancer. *Statistics in Medicine*, 22(14):2239–56.

Ryeznik, Y. and Sverdlov, O. (2018). A comparative study of restricted randomization procedures for multiarm trials with equal or unequal treatment allocation ratios. *Statistics in Medicine*, 37(21):3056–3077.

Saville, B. R. and Berry, S. M. (2016). Efficiencies of platform clinical trials: a vision of the future. *Clinical Trials*, 13(3):358–366.

Senn, S. and Bretz, F. (2007). Power and sample size when multiple endpoints are considered. *Pharmaceutical Statistics*, 6(3):161–70.

Soares, J. F. and Wu, C. J. (1983). Some restricted randomization rules in sequential designs. *Communications in Statistics-Theory and Methods*, 12(17):2017–2034.

Stallard, N. and Friede, T. (2008). A group-sequential design for clinical trials with treatment selection. *Statistics in Medicine*, 27(29):6209–6227.

Stallard, N. and Todd, S. (2003). Sequential designs for phase III clinical trials incorporating treatment selection. *Statistics in Medicine*, 22(5):689–703.

Stigsby, B. and Taves, D. R. (2010). Rank-minimization for balanced assignment of subjects in clinical trials. *Contemporary clinical trials*, 31(2):147–150.

Su, Z. (2011). Balancing multiple baseline characteristics in randomized clinical trials. *Contemporary Clinical Trials*, 32(4):547–550.

Sverdlov, O. (2016). An overview of adaptive randomization designs in clinical trials. *Modern adaptive randomized clinical trials: Statistical and practical aspects*, pages 3–44.

Sverdlov, O., Glimm, E., and Mesenbrink, P. (2022). Platform trial designs. In *Principles and Practice of Clinical Trials*, pages 1455–1485. Springer.

Sverdlov, O. and Rosenberger, W. (2013). On recent advances in optimal allocation designs in clinical trials. *Journal of Statistical Theory and Practice*, 7(4):753–73.

Sydes, M., Parmar, M., James, N., Clarke, N., Dearnaley, D., Mason, M., Morgan, R., Sanders, K., and Royston, P. (2009). Issues in applying multi-arm multi-stage methodology to a clinical trial in prostate cancer: The MRC STAMPEDE trial. *Trials*, 10:39.

Sydes, M. R., Parmar, M. K., Mason, M. D., Clarke, N. W., Amos, C., Anderson, J., de Bono, J., Dearnaley, D. P., Dwyer, J., Green, C., et al. (2012). Flexible trial design in practice-stopping arms for lack-of-benefit and adding research arms mid-trial in stampede: a multi-arm multi-stage randomized controlled trial. *Trials*, 13:1–14.

Taves, D. R. (1974). Minimization: a new method of assigning patients to treatment and control groups. *Clinical Pharmacology & Therapeutics*, 15(5):443–453.

Taves, D. R. (2010). The use of minimization in clinical trials. *Contemporary clinical trials*, 31(2):180–184.

Thall, P. F., Simon, R., and Ellenberg, S. S. (1988). Two-stage selection and testing designs for comparative clinical trials. *Biometrika*, 75(2):303–310.

The Adaptive Platform Trial Coalition (2019). Adaptive platform trials: definition, design, conduct and reporting considerations. *Nature Reviews Drug Discovery*, 18(10):797–807.

Treasure, T. and MacRae, K. D. (1998). Minimisation: the platinum standard for trials?: Randomisation doesn't guarantee similarity of groups; minimisation does.

Tymofyeyev, Y., Rosenberger, W., and Hu, F. (2007). Implementing optimal allocation in sequential binary response experiments. *Journal of the American Statistical Association*, 102(477):224–34.

Urach, S. and Posch, M. (2016). Multi-arm group sequential designs with a simultaneous stopping rule. *Statistics in Medicine*, 35(30):5536–50.

U.S. Food and Drug Administration (2004). Innovation or stagnation: challenge and opportunity on the critical path to new medical products. *http://www. fda. gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/default. htm*.

U.S. Food and Drug Administration (2006). Approval summary: bortezomib for the treatment of progressive multiple myeloma after one prior therapy. *Clinical Cancer Research*, 12(10):2955–2960.

Vanderbeek, A. M., Bliss, J. M., Yin, Z., and Yap, C. (2022). Implementation of platform trials in the covid-19 pandemic: a rapid review. *Contemporary Clinical Trials*, 112:106625.

Wason, J. and Jaki, T. (2012). Optimal design of multi-arm multi-stage trials. *Statistics in Medicine*, 31(30):4269–79.

Wason, J., Magirr, D., Law, M., and Jaki, T. (2016). Some recommendations for multi-arm multi-stage trials. *Statistical Methods in Medical Research*, 25(2):716–27.

Wason, J. and Mander, A. (2012). Minimizing the maximum expected sample size in two-stage phase II clinical trials with continuous outcomes. *Journal of Biopharmaceutical Statistics*, 22(4):836–52.

Wason, J. and Trippa, L. (2014). A comparison of Bayesian adaptive randomization and multi-stage designs for multi-arm clinical trials. *Statistics in Medicine*, 33(13):2206–21.

Wason, J. M., Stecher, L., and Mander, A. P. (2014). Correcting for multiple-testing in multi-arm trials: is it necessary and is it done? *Trials*, 15:1–7.

Wassmer, G. and Brannath, W. (2016). *Group sequential and confirmatory adaptive designs in clinical trials*, volume 301. Springer.

Wei, L.-J. (1977). A class of designs for sequential clinical trials. *Journal of the American Statistical Association*, 72(358):382–386.

Wei, L.-J. (1978). The adaptive biased coin design for sequential experiments. *The Annals of Statistics*, 6(1):92–100.

Whitehead, J. and Brunier, H. (1990). The double triangular test: a sequential test for the two-sided alternative with early stopping under the null hypothesis. *Sequential Analysis*, 9(2):117–136.

Whitehead, J. and Jaki, T. (2009). One-and two-stage design proposals for a phase II trial comparing three active treatments with control using an ordered categorical endpoint. *Statistics in Medicine*, 28(5):828–47.

Woodcock, J. and LaVange, L. M. (2017). Master protocols to study multiple therapies, multiple diseases, or both. *New England Journal of Medicine*, 377(1):62–70.

Zagoraiou, M. (2017). Choosing a covariate-adaptive randomization procedure in practice. *Journal of biopharmaceutical statistics*, 27(5):845–857.

Zhao, W. (2014). A better alternative to stratified permuted block design for subject randomization in clinical trials. *Statistics in medicine*, 33(30):5239–5248.

Zhao, W., Berger, V. W., and Yu, Z. (2018). The asymptotic maximal procedure for subject randomization in clinical trials. *Statistical Methods in Medical Research*, 27(7):2142–2153.

Zhao, W. and Weng, Y. (2011). Block urn design—a new randomization algorithm for sequential trials with two or more treatments and balanced or unbalanced allocation. *Contemporary Clinical Trials*, 32(6):953–961.