Inference following biased coin designs in clinical trials

Wai Yin Yeung

A thesis submitted for the degree of Doctor of Philosophy

School of Mathematical Sciences,

Queen Mary, University of London

June 2013

Acknowledgements

I would like to thank Dr. D. S Coad, my supervisor, for his excellent academic guidance, invaluable suggestions and continuous encouragement throughout my research. I am extremely grateful for all his patience in answering my questions and queries during the course of my study, and for providing me with extremely constructive feedback on my writing. I must express my greatest gratitude for his understanding and support throughout my PhD.

Many thanks to the Engineering and Physical Sciences Research Council (EPSRC) for their generous funding during the past three and a half years. This research work would not have been able to proceed and completed without their financial support.

I am most thankful to Dr. V. V Anisimov as my supervisor in GlaxoSmithKline U.K. Limited (GSK) for my three-month Collaborative Research Venture (CRV). Results presented in Chapter 5 were completed during this time. Thanks for all his assistance during my time in GSK; his invaluable guidance, insightful ideas for the work and all his advice on programming and writing for the material included in Chapter 5. I am very grateful to the ImpactQM Scholarship Award Scheme, which provided me with the opportunity and financial support for my three-month (CRV) in the summer of 2011. Thanks to GSK for their hospitality. In particular, thanks to the Quantitative Sciences Division within GSK for selecting me as one of the four students to receive full financial support to present the work in Chapter 5 as a poster at its Internal Statistics Conference in October 2011.

Many thanks to the Society for Clinical Trials (SCT) for providing me with full financial support to travel to Miami in May 2012 as one of the three finalists of the SCT Thomas C.

Chalmers Student Scholarship and to present the work in Chapter 3 at the 33rd Annual Meeting of the SCT.

I would like to thank the International Biometric Society British and Irish Region for their bursary for me to attend the 26th International Biometric Conference in Kobe, Japan, in August 2012; and for awarding me the second prize for the Best Student Oral Presentation.

Thanks to all of my colleagues in the School of Mathematical Sciences at Queen Mary, University of London, particularly the statistics group, for providing me with a supportive working environment. My appreciation to Dr. B. Bogacka, my assessor and undergraduate advisor, and to Professor R. A Bailey, my assessor for their advice and feedback on my work at different stages of my research. Thanks also to Professor S. G Gilmour and Dr. R. A. Sugden for their help since my undergraduate studies.

I would like to express my sincerest gratitude to my family and friends, especially to my parents for their love and encouragement throughout my life. To my dearest father, Chak Lam, for his continuous support and belief in my ability. To my dearest mother, Yuk Ling, for her care and understanding during my study. To my brother, Chi Kit, for his patience in listening to all of my ups and downs during my research. I have lost my precious grandfather and grandmother. To my grandfather, Hung Shau, for teaching me never to give up and always to do my best. To my grandmother, Tak Hing, for her care and moral support.

Results presented in Chapters 2 and 4 will be combined to form a paper for submission to the *Journal of Statistical Planning and Inference*. Chapter 3 will form a paper to be sent to *Clinical Trials: Journal of the Society for Clinical Trials*. Finally, Chapter 5 is written as a paper to be submitted to *The Annals of Applied Statistics*.

Abstract

Randomization schemes for two-treatment clinical trials are studied. Theoretical expressions for the power are derived under both complete randomization and Efron's biased coin design for normal and binary responses. The better the scheme is at balancing the numbers of patients across treatments, the higher the power is. Efron's biased coin design is more powerful than complete randomization. Normal approximations to the powers are obtained. The power of the adjustable biased coin design is also investigated by simulation.

Covariate-adaptive randomization schemes are analysed when either global or marginal balance across cells is sought. By considering a fixed-effects linear model for normal treatment responses with several covariates, an analysis of covariance t test is carried out. Its power is simulated for global and marginal balance, both in the absence and in the presence of interactions between the covariates. Global balancing covariate-adaptive schemes are more efficient when there are interactions between the covariates.

Restricted randomization schemes for more than two treatments are then considered. Their asymptotic properties are provided. An adjustable biased coin design is introduced for which assignments are based on the imbalance across treatments. The finite-sample properties of the imbalance under these randomization schemes are studied by simulation. Assuming normal treatment responses, the power of the test for treatment differences is also obtained and is highest for the new design.

Imbalance properties of complete randomization and centre-stratified permuted block randomization for several treatments are investigated. It is assumed that the patient recruitment process follows a Poisson-gamma model. When the number of centres is large, the imbalance for both schemes is approximately multivariate normal. The power of a test for treatment differences is simulated for normal responses. The loss of power can be compensated for by a small increase in sample size.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Clinical trials

A clinical trial is a combination of studies, experiments and tests performed to test the efficacy and the safety of some existing or new treatments or drugs on human. A treatment refers to the method, the therapy or the remedy that is used on a patient for a disease or injury. A drug refers to the chemical substance intake by humans and it is used to cure, relieve or prevent diseases. During the trial, we collect data and gather information about the efficacy and the safety of the treatments and drugs. This is to identify any positive or adverse effects of the treatments or drugs on patients such that the most suitable treatments and drugs can be available to patients in the future.

In clinical trials, the patients can be healthy volunteers or patients with specific characteristics. For example, to test the effectiveness of a treatment for a particular disease, we may have to recruit patients suffering from this disease. In particular, if we need to recruit patients with several characteristics or we deal with trials for rare diseases, it may take a very long period to recruit enough patients. The number of patients to be recruited differs in the different phases of clinical trials.

There are different types of studies for clinical trials. These include testing the efficacy and safety of new treatments and drugs on patients, studies of the effectiveness and safety of existing treatments and drugs and assessing the effectiveness and safety of different doses of drugs other than the existing dose of a drug. In addition, two or more treatments or drugs are normally included in a trial.

The cost in completing all phases of a clinical trial is huge. Clinical trials are commonly sponsored and carried out by pharmaceutical companies or government-related organizations. These trials can be carried out in several centres or even in several countries. We have to ensure the same design and analysis are carried out in different centres and countries, and data are gathered from all centres and countries for the statistical analysis.

Pre-clinical studies are also needed to identify the drug for assessment in clinical trials. Before the trial, potential compounds for drugs have to be tested for several years and only very few of them can reach the stage to be tested in clinical trials. Clinical trials are normally classified into four phases. Each phases is usually considered as a separate clinical trial. It can take years and even decades to complete all four phases of a clinical trial before new treatments and drugs are available on the market for the general public. However, sometimes some of these phases are combined. The reasons for combined phases include making the assessment process quicker for a new treatment or drug to be available to patients, reducing the costs in carrying out the trials, requiring fewer patients and reducing the risk to the participants. For example, we often have Phases I and II or Phases II and III combined.

For Phase I of clinical trials, the main focus is on drug safety. In this stage, studies and experiments are carried out on around 20-80 patients. The main objective in this stage is to determine the maximum safe dose, that is, the maximum dose level without

causing toxicity. For Phase II, the studies will be carried out on a larger group of patients of size 100-200. At this stage, further investigation of the effectiveness and safety of drugs is carried out. This is also a stage to narrow down the number of drugs, by excluding those over-toxic or inactive drugs and choosing a few of the potential drugs to proceed to Phase III.

In Phase III, we apply the new drugs or treatments to a larger group of patients. Further tests will be carried out to confirm the effectiveness and safety of the treatments and drugs. In here, we may compare the new treatments or drugs to the standard or existing treatment. Details of different randomization schemes, the treatment assignment rules to patients, for this phase will be covered in this thesis. This phase is also referred to as the largest and most extensive study on the new treatments and drugs.

The last phase which is Phase IV is the postmarketing stage where remaining studies have to be carried out before the drug is approved for marketing. These include studies on adverse effects and benefits in using this drug.

1.2 Randomization schemes

Randomization refers to the process of randomly allocating patients to one of the treatments in clinical trials. These randomization schemes can be applied in all phases of a clinical trial. In this thesis, we will only study the randomization schemes used in Phase III, where we want to assess the effectiveness of our new treatments or drugs. We will test for differences in patients' responses on the treatments. One of the main aims of randomization schemes is to minimize selection bias.

Selection bias refers to the bias incurred when the experimenter or the investigator consciously or unconsciously decides which treatment is to be allocated to the next patient. The assignment process should be unpredictable so that both the investigator and

the patients will not know in advance the treatment to be allocated next in the trial. The selection bias is closely related to how likely the next treatment to be allocated to a subject can be guessed. The more likely that the treatment can be guessed the higher the selection bias will be. A good randomization scheme should have a low probability of correctly guessing which treatment is to be allocated to the next patient.

Another aim of randomization schemes is to balance the numbers of patients across treatments. Statistical power refers to the probability of detecting a genuine treatment effect in a test for treatment difference. The higher the power of the test under a randomization scheme, the more likely the scheme is to detect a genuine treatment effect. For example, for statistical analysis of a trial for comparing two treatments, assume the population variances of the patients' responses in the two treatment groups are the same. Then it is optimal when the numbers of patients in the two treatment groups are the same. In fact, the power of the test is maximized for a balanced trial when the population variances of the patients' responses in the two groups are the same (Lachin, 1981).

There are also different types of randomization schemes. The most straightforward randomization scheme is simple randomization. This is also called repeated simple random sampling, simple random sampling or complete randomization. Under this scheme, patients are randomly allocated to the treatments and it is the best at minimizing the selection bias.

We also have other types of randomization schemes such as restricted randomization. One of the main goals of restricted randomization schemes is to balance the numbers of patients across treatments. Some examples are Efron's (1971) biased coin design, the adjustable biased coin design (Baldi Antognini and Giovagnoli, 2004) and permuted block randomization. Except permuted-block randomization, most of these schemes have a greater probability of allocating the next patient to a treatment group that has

fewer patients and vice versa for a treatment group that is overrepresented. The current numbers of patients on the treatments are therefore an important basis for the treatment assignment of the next patient under these schemes. For permuted-block randomization, a fixed block size and an allocation proportion of treatments within blocks are given such that patients are allocated to treatments within a block randomly.

Covariate-adaptive randomization is also another type of randomization scheme that will be described in Chapter 3 of this thesis. One of the examples of covariate-adaptive randomization schemes is the minimization method of Pocock and Simon (1975). Examples of patients' covariates or prognostic factors are age, gender, stage of disease and so on. Under these randomization schemes, we aim to balance the numbers of patients across treatments for patients classified by their prognostic factors or covariates.

We also have response-adaptive randomization schemes. Under these schemes, patients' responses on treatments are recorded and are used for the treatment assignment of the forthcoming patient. The probability of assigning the next patients to a treatment that gives better responses will be higher. This is to increase the chance of allocating patients to treatments that are performing well. Details of response-adaptive randomization schemes will not be discussed.

1.3 Outline of thesis

The research work presented in this thesis mainly focuses on the comparison of biased coin designs with other randomization schemes. The biased coin design was developed by Efron (1971). This is a design for patient allocation to treatments in sequential clinical trials. Given two treatments, patients have to be allocated one of the two treatments upon their arrival. The biased coin design has a fixed probability greater than a half to allocate the next patient to a treatment that has been chosen less often. In particu-

lar, the powers of the test for treatment effect under all these randomization schemes are of our main concern throughout the thesis. Different assumptions under these randomization schemes will be investigated for their impact on the power.

The thesis consists of four main parts. Chapter 2 relates to randomization schemes with only two treatments. These two treatments are usually referred to as the standard treatment and the new treatment. A test is carried out to decide whether the new treatment is better than the standard one. The power of the test has been analysed theoretically and numerical results for the powers have been produced by Chen (2006). Here, both the theoretical expressions and the numerical results for the power will be given by assuming normal patient responses with different variances. The cases when these variances are known or unknown are studied under complete randomization and Efron's biased coin design. In Section 2.3, algorithms and numerical results for the power by simulation under Efron's biased coin design and the adjustable biased coin design will be given. The adjustable biased coin design is an extension of Efron's biased coin design where at each stage the probability of allocating the next patient to a treatment is a function instead of a fixed probability. This function depends on the current difference between the numbers of patients on the two treatments. We call the treatment with fewer of patients the under-represented treatment. Under the adjustable biased coin design, the fewer the number of patients on the under-represented treatment, the greater the probability of assigning the next patient to this treatment. In Section 2.4, the power functions are analysed under complete randomization and Efron's biased coin design when the patients' responses are binary. Details of the algorithm for the theoretical expression and the numerical values for the power are given for both designs. In Section 2.5, numerical values for the power using a normal approximation will be presented and hence compared with the exact powers under complete randomization and Efron's biased coin

design.

Chapter 3 covers material on covariate-adaptive randomization schemes. Patient prognostic factors or covariates such as age and gender will affect the responses of patients to the treatments. Therefore, it is sensible to study the treatment effect in a group of patients with the same or similar prognostic profiles. The covariate-adaptive randomization schemes refer to randomization schemes that apply to patients grouped by covariates. Such schemes are studied by Shao, Yu and Zhong (2010) when there is a single covariate and two treatments under comparison. One of the main aims of these covariate-adaptive randomization schemes is to balance the number of patients with the same prognostic profile. Here, three covariate-adaptive randomization schemes are studied when either global or marginal balance is sought. By considering a fixed-effects model for normal treatment responses when there are several covariates, an analysis of covariance t test is carried out. Numerical values for the power are simulated for the three randomization schemes for both global and marginal balance when there is an interaction between the covariates. It is shown that the covariate-adaptive adjustable biased coin design produces the highest power among the three. In addition, the power gain under global balance is higher than under marginal balance when there is an interaction between the covariates.

Chapter 4 gives results for different biased coin designs when there are more than two treatments. In this chapter, the assignment rules for more than two treatments under complete randomization and different biased coin designs will be demonstrated. A new class of assignment rules will be given such that for each treatment the probability of assigning the next patient to this treatment depends on the current value of the imbalance in this treatment. The asymptotic properties of complete randomization and these biased coin designs will be stated. This is followed by simulation results for the imbal-

ances across treatments, which confirms the theoretical asymptotic properties under all these schemes. Assuming patients' responses to be normally distributed and given the numbers of patients on the treatments, the analysis of variance F test is then used to test whether there are any treatment differences.

Chapter 5 is about the imbalance properties of complete randomization and permuted block randomization in clinical trials for several treatments. The number of patients to be recruited in different centres can be described by a patient recruitment process. The model is a Poisson process with arrival rate from a gamma distribution (Anisimov and Fedorov, 2007) such that the number of patients to be recruited in each centre has a beta-binomial distribution. Details about this model will be given in Section 5.3. We then have the imbalances defined for each treatment within centres and for all centres, and hence the column vector for the overall imbalance. In addition, for both of these randomization schemes, analytical results for the imbalances are investigated assuming that the number of patients to be recruited in different centres is fixed or when the patient recruitment process follows a Poisson-gamma model. When the number of centres involved in a trial is large, the overall imbalance for both schemes is approximately a multivariate normal. The accuracy of the approximations is assessed by simulation. The variances of the imbalance in a particular treatment within a centre are then compared under these two randomization schemes when the number of patients to be recruited in each centre is fixed or random. A test is then suggested for testing if there is at least one treatment difference when we compare each of the new treatments with the control treatment. By considering different simulated scenarios under centre-stratified permuted block randomization, the sample sizes are found in a balanced trial when a particular level of the power is achieved. We use the same sample sizes to study the impact on the power in the imbalanced case for each of the scenarios.

Chapter 6 draws conclusions and gives an indication of possible future work.

Supplementary information for Chapters 2-5 is given in three appendices.

Chapter 2

Restricted randomization schemes for two treatments

In this chapter, we will consider clinical trials where only two treatments are involved. Several randomization schemes will be covered and studied in detail with different assumptions. Consider a sequential clinical trial where patients arrive one by one and have to be assigned immediately to either of the two treatments. At each stage for each newly arrived patient, the assignments to be made under these randomization schemes are based on the current allocation status, that is, the current number of patients on each of the two treatments. One group of patients will receive the standard treatment and we call this the control group. The other group will receive the new treatment and is called the treatment group. For simplicity, treatment 1 will represent the standard treatment and treatment 2 will represent the new treatment.

2.1 Treatment assignment rules under different randomization schemes

2.1.1 Complete randomization/ Repeated simple random sampling

The assignment rules under complete randomization (CR) are very straightforward. Upon arrival, each patient is equally likely to be assigned to one of the two treatments. That is, both treatments 1 and 2 have the same probability of $1/2$ of being allocated the next patient. The probabilities of assignment for different patients are independent. This scheme achieves a high level of randomness in the assignment which means that the treatment to be assigned to the next patient is less likely to be predicted. In other words, the probability of correctly guessing the next treatment assignment for each newly arrived patient is low and hence the sequence of treatment allocations is less obvious. Furthermore, the selection bias discussed in Blackwell and Hodges (1957) and Efron (1971) which refers to the bias incurred when the experimenter's decision to allocate the subject to a treatment for which the experimenter thinks it will be the most suitable for the subject, is relatively low under complete randomization. In addition, the selection bias of the design defined in Baldi Antognini and Giovagnoli (2004) as the expected proportion of correct guesses equals $1/2$ under complete randomization for any sample size. Due to the randomness in the assignment of this design, the imbalances across the treatments produced under this scheme is very high compared to other designs.

2.1.2 Efron's biased coin design

In Efron (1971), the biased coin design BCD(p) was introduced and described as follows: in an experiment with two treatments, each patient arrives sequentially and has to be assigned immediately to one of the treatments. Based on the number of pa-

tients currently on each of the treatments, the biased coin design assigns the next patient to a treatment that has been chosen less often previously with probability p such that $p > 1/2$. When patients arrive sequentially, the assignment rule is to balance the number of patients on the treatments. It aims to achieve some randomness when assigning each patient to a treatment, which reduces any selection bias in the assignment of patients to treatments. In addition, this design also aims to maintain a balance in the number of patients on the two treatments. A balanced trial minimizes the variance of the test statistic and hence increases the power of the test when the variances for patients' responses are equal for the two groups.

2.1.3 The adjustable biased coin design

A new design called the adjustable biased coin design (ABCD) introduced by Baldi Antognini and Giovagnoli (2004) is an extension of Efron's biased coin design. This is also a design which aims to eliminate selection bias and maintain a balance in the number of patients in the treatment groups. Efron's biased coin design is similar to the adjustable biased coin design for which assignment of patients to treatment groups depends solely on the current number of patients in the two groups and biases the allocation towards the under-represented treatment. The only difference between them is their probability of selecting an under-represented treatment. The ABCD is a class of biased coin designs for which the probability of selecting a treatment at stage n is a decreasing function of the current difference between the numbers of patients on this treatment and the other treatment. Let n_1 and n_2 be the numbers of patients on treatment 1 and 2, respectively. Let $\delta_n = 1$ if the nth patient is allocated to treatment 1 and $\delta_n = -1$ otherwise. Let $p(n_1, n_2)$ be the conditional probability that the next patient is allocated to treatment 1

so that

$$p(n_1, n_2) = P(\delta_{n+1} = 1 | n, \tilde{D}_n),$$

where $\tilde{D}_n = \sum_{i=1}^{n} \delta_i = n_1 - n_2$.

Let $F(.)$ be a function $F : \mathbb{Z} \to [0, 1]$ for $\mathbb{Z}$ the set of integers such that

1. $F(x)$ is decreasing ;

2. $F(-x) = 1 - F(x)$.

Then $\mathcal{F}$ is the class of non-constant functions satisfying 1. and 2. above. Let

$$p(n_1, n_2) = F(n_1 - n_2),$$

so that $p(n_1, n_2) = 1/2$ if $n_1 = n_2$. The ABCD is generated by the function $F(.)$ in $\mathcal{F}$. Antognini and Giovagnoli (2004) suggested a class of functions which is

$$F_a(x) = \begin{cases} \frac{|x|^a}{|x|^a+1} & \text{if } x \leq -1, \\ \frac{1}{2} & \text{if } x = 0, \\ \frac{1}{|x|^a+1} & \text{if } x \geq 1, \end{cases} \tag{2.1}$$

where $a \in \mathbb{R}^+$ is a design parameter.

The ABCD generated by the above expression is abbreviated by ABCD(F_a). This class of functions $F_a(.)$ will be our main focus for studying the properties of the adjustable biased coin design. When $a = 0$, it becomes complete randomization and as $a \to \infty$, the design becomes deterministic. Note that $F(.)$ is the general form of the function that generates the ABCD and $F_a(.)$ is the particular class of functions that we are interested in. Therefore, by comparing $F(.)$ and $F_a(.)$ for (2.1), we see that the unknown variable x is $x = n_1 - n_2$. In general, $F_a(.)$ is the probability of allocating the next patient to treatment 1 and this function depends on a chosen value of a and x which is the difference between the numbers of patients on treatment 1 and treatment 2

at different stages. Hence, the probability of allocating the next patient to treatment 2 is $1 - F_a(x)$.

2.1.4 Wei's class of biased coin designs

Wei (1978) developed a new class of biased coin designs called the adaptive biased coin designs which are defined as follows.

Let $\tilde{D}_n = n_1 - n_2$ be the difference in the numbers of patients on treatments 1 and 2 after n assignments. Under this class of designs, the probability of assigning the $(n+1)$th patient to treatment 1 is a decreasing function of $\tilde{D}_n/n$, denoted by $p(\tilde{D}_n/n)$. Similarly, the probability of assigning the $(n+1)$th patient to treatment 2 is denoted by $q(\tilde{D}_n/n)$, where $p(\tilde{D}_n/n) + q(\tilde{D}_n/n) = 1$. The function $p(x)$ is chosen to be symmetric such that $p(x) = q(-x)$ for $x \in [-1, 1]$. In Wei (1978), the special case $p(x) = (1-x)/2$ is considered and studied. This special case can be written as

$$p(n_1, n_2) = \frac{n_2}{n_1 + n_2}$$

for the probability of assigning the $(n+1)$th patient to treatment 1.

2.1.5 Smith's class of designs

In Smith (1984), the decreasing function $p(\tilde{D}_n/n)$ from Wei's class of designs is studied. Assuming this function is differentiable at zero, a class of designs is suggested such that the probability of assigning treatment 1 to the $(n+1)$th patient is $p(x) = (1-x)^\rho/\{(1+x)^\rho + (1-x)^\rho\}$, which is the same as

$$p(n_1, n_2) = \frac{n_2^\rho}{n_1^\rho + n_2^\rho},$$

where $\rho = -2p'(0)$ and $p'(0)$ is the first derivative of the function p at 0. When $\rho = 0$, this class of designs reduces to complete randomization.

2.1.6 The D- and D_A-optimum biased coin designs

In Atkinson (1982), the theory of optimum design is used to obtain the probabilities of assigning the next patient to a treatment for the biased coin design. The D- and D_A-optimum biased coin designs are discussed. Consider a linear model for treatment responses $E(\mathbf{Y}) = x^T\boldsymbol{\beta}$ for which the responses on the two treatments are independent observations with the same variance σ^2. The covariance matrix of the least squares estimator of $\boldsymbol{\beta}$ is

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}.$$

The fitted value at x is denoted by $\hat{y}(x) = \hat{\boldsymbol{\beta}}^T x$ and the variance of the fitted value at x is $\mathrm{Var}\{\hat{y}(x)\} = \sigma^2 x^T(\mathbf{X}^T\mathbf{X})^{-1}x$. In the context of optimum experimental design, it is suggested that we can represent the n patient design by a measure ξ_n. After n assignments, the information matrix of the design ξ_n is denoted by $M(\xi_n) = n^{-1}(\mathbf{X}^T\mathbf{X})$, where $\mathbf{X}^T\mathbf{X}$ is a $p \times p$ matrix. In addition, the standardized variance at x is given by

$$d(x, \xi_n) = n\frac{\mathrm{Var}\{\hat{y}(x)\}}{\sigma^2} = x^T M^{-1}(\xi_n)x.$$

The D-optimum criterion is used when all the parameters in $\boldsymbol{\beta}$ are of interest in the study. The main purpose of the D-optimality criterion is to minimizes the generalized variance of these parameters' estimators. In other words, this maximizes the determinant of $M(\xi_n)$.

However, if the contrasts between the treatment effects are of interest, the D_A-optimality criterion is more appropriate to use for treatment allocations. The contrasts are the components of the vector $\mathbf{A}^T\boldsymbol{\beta}$ where $\mathbf{A}$ is a $p \times s$ matrix for the contrasts with rank $s < p$. The covariance matrix of the least squares estimator $\mathbf{A}^T\hat{\boldsymbol{\beta}}$ is proportional to $\mathbf{A}^T M^{-1}(\xi_n)\mathbf{A}$. The D_A-optimality criterion maximizes the determinant of $\{\mathbf{A}^T M^{-1}(\xi_n)\mathbf{A}\}^{-1}$. The standardized variance under the D_A-optimality criterion is de-

noted by

$$d_A(x, \xi_n) = x^T M^{-1}(\xi_n)\mathbf{A}\{\mathbf{A}^T M^{-1}(\xi_n)\mathbf{A}\}^{-1}\mathbf{A}^T M^{-1}(\xi_n)x.$$

For the biased coin design rules after n assignments, we have

$$p_j = \frac{d(j, \xi_n)}{\sum_{i=1}^{2} d(i, \xi_n)}$$

as the probability of allocating the next patient to treatment j for $j = 1, 2$.

The special case for which the model for treatment responses is $E(Y) = \beta_j$ for $j = 1, 2$ is used with two treatments and no prognostic factors involved. The contrast $\beta_1 - \beta_2$ is of interest and $\mathbf{A}^T = (1, -1)$ is the matrix of contrasts. The rank of this matrix is 1. After n assignments, there are n_1 patients on treatment 1 and n_2 patients on treatment 2. The covariance matrix of the least squares estimator in this case is $\sigma^2 \text{diag}(1/n_1, 1/n_2)$ and $M(\xi_n) = \text{diag}(n_1/n, n_2/n)$.

The standardized variance for the D-optimality criterion is $d(j, \xi_n) = n/n_j$ for treatment $j = 1, 2$. For the biased coin assignment rule, the probabilities of allocating the next patient to treatments 1 and 2 are $p_1 = n_2/n$ and $p_2 = n_1/n$, respectively.

For the D_A-optimality criterion, the standardized variance is $d_A(1, \xi_n) = n_2/n_1$ for treatment 1 and $d_A(2, \xi_n) = n_1/n_2$ for treatment 2. Here, the probabilities of allocating the next patient to treatments 1 and 2 are $p_1 = n_2^2/(n_1^2 + n_2^2)$ and $p_2 = n_1^2/(n_1^2 + n_2^2)$, respectively.

Under both the D- and the D_A-optimality criteria, the probabilities of treatment allocation are special cases of Smith's class of designs. When $\rho = 1$ in $p(x)$ or $p(n_1, n_2)$, this probability is the same as the p_1 in Atkinson's design for the D-optimality criterion. Similarly, $q(x)$ and p_2 are the same. When $\rho = 2$, $p(x)$ is the same as p_1 in Atkinson's design, and also $q(x)$ is the same as p_2, for the D_A-optimality criterion.

2.1.7 Asymptotic properties of the treatment assignment designs

Complete randomization

For complete randomization, the number of patients on treatment 1, n_1, has a binomial distribution with parameters n and $1/2$. Hence, the variance of n_1 is $n/4$. By the central limit theorem, we have

$$\sqrt{n}\left(\frac{n_1}{n}-\frac{1}{2}\right)\rightarrow N\left(0,\frac{1}{4}\right),$$

in distribution as $n\rightarrow\infty$. All of the above is also true for n_2 for the number of patients on treatment 2 under complete randomization.

Efron's biased coin design

In Hu, Zhang and He (2009), for a two-treatment trial, the asymptotic variance of $n_1/\sqrt{n}$ for the efficient randomized-adaptive designs (ERADE) attains the Cramér-Rao lower bound. Efron's biased coin design is a special case of ERADE with the target allocation $\xi=1/2$ on each treatment. As ξ is a constant under Efron's biased coin design, the Cramér-Rao lower bound is zero. Hence, under Efron's biased coin design, the asymptotic variance of $n_1/\sqrt{n}$ attains the Cramér-Rao lower bound of zero. Therefore, as $n\rightarrow\infty$,

$$\sqrt{n}\left(\frac{n_1}{n}-\frac{1}{2}\right)\rightarrow 0$$

in probability.

The adjustable biased coin design

The function $F:\mathbb{Z}\rightarrow[0,1]$, with $\mathbb{Z}$ the set of integers, is decreasing and symmetric. The probability of assigning the $(n+1)$th patient to treatment 1 is $F_a(x)$, where $x=\tilde{D}_n$. In Baldi Antognini and Giovagnoli (2004) and Baldi Antognini (2008), it is shown that

$n_1/n \to 1/2$ and $\tilde{D}_n/\sqrt{n} \to 0$ almost surely as $n \to \infty$. Therefore, as $n \to \infty$,

$$\sqrt{n}\left(\frac{n_1}{n} - \frac{1}{2}\right) \to 0$$

in probability.

Wei's, Smith's and Atkinson's classes of designs

It is shown in Wei (1978) that if $p(x)$ is differentiable at $x = 0$, $n^{-1/2}\tilde{D}_n$ converges to a normal distribution with mean 0 and variance $1/\{1 - 4p'(0)\}$. By rearranging terms as $n \to \infty$, $\sqrt{n}(n_1/n - 1/2)$ will converge to a normal distribution. Here, $E(n^{-1/2}\tilde{D}_n) = E\{\sqrt{n}(n_1/n - 1/2)\} = 0$ and $\mathrm{Var}(n^{-1/2}\tilde{D}_n) = 4\mathrm{Var}\{\sqrt{n}(n_1/n - 1/2)\}$. It follows that

$$\sqrt{n}\left(\frac{n_1}{n} - \frac{1}{2}\right) \to N\left(0, \frac{1}{4\{1 - 4p'(0)\}}\right)$$

in distribution as $n \to \infty$.

In Smith (1984), the particular class of designs with

$$p(x) = \frac{(1-x)^\rho}{(1+x)^\rho + (1-x)^\rho}$$

is considered with $\rho = -2p'(0)$. Since $p(x)$ is differentiable at $x = 0$, $n^{-1/2}\tilde{D}_n$ for this particular class of designs has the above asymptotic properties. Note that, we have $4p'(0) = -2\rho$. Therefore, under Smith's class of designs with this $p(x)$, we have

$$\sqrt{n}\left(\frac{n_1}{n} - \frac{1}{2}\right) \to N\left(0, \frac{1}{4(1+2\rho)}\right)$$

in distribution as $n \to \infty$.

In addition, Atkinson's class of D- and D_A-optimum biased coin designs are two special cases of Smith's class of designs when $\rho = 1$ and $\rho = 2$, respectively. So we have

$$\sqrt{n}\left(\frac{n_1}{n} - \frac{1}{2}\right) \to N\left(0, \frac{1}{12}\right)$$

in distribution as $n \to \infty$ under the D-optimum biased coin design and

$$\sqrt{n}\left(\frac{n_1}{n} - \frac{1}{2}\right) \to N\left(0, \frac{1}{20}\right)$$

in distribution as $n \to \infty$ under the D_A-optimum biased coin design.

From all of the above, we can see that Efron's biased coin design and the adjustable biased coin are less variable asymptotically in balancing the numbers of patients on two treatments. For both Efron's biased coin design and the adjustable biased coin design, $\sqrt{n}\,(n_1/n - 1/2)$ converges to zero instead of a normal distribution.

In Baldi Antognini and Giovagnoli (2004), the plots of the asymptotic values of the expected absolute differences $\tilde{D}_n$ and $\tilde{D}_n/n$ suggest that the adjustable biased coin design converges to balance faster than Wei's class of designs when $a = 2\rho$ for $a = 1, 2, 4$ and is preferable to Efron's biased coin design with $p = 2/3$ for $n > 10$. In addition, in terms of the asymptotic predictability, the adjustable biased coin design generated by (2.1) for any choice of a is preferable to Efron's biased coin design with $p = 2/3$ in balancing the numbers of patients on the two treatments.

2.2 Theoretical analysis of the power with normal responses for complete randomization and Efron's biased coin design

2.2.1 Background

In Chen (2006), the power of two designs for detecting treatment effects is investigated. The power is treated as the conditional probability of correctly detecting a treatment effect given a particular status of the treatment allocation. The powers of the complete randomization and the biased coin design with a deterministic value of p

are investigated and compared. Of course, if $p = 1/2$, the BCD(p) is just complete randomization. One of the treatment groups is defined to be the control group and the other the treatment group.

Firstly, it is assumed that there are n patients at the end of the trial with n_2 of them in the treatment group and n_1 of them in the control group, so that $n = n_2 + n_1$. Secondly, let the control responses $X_1, ..., X_{n_1}$ be independent and normally distributed with unknown mean μ_1, and similarly let the treatment responses $Y_1, .., Y_{n_2}$ be independent and normally distributed with unknown mean μ_2. The control responses are independent of the treatment responses. In addition, the control responses and treatment responses are assumed to have a common variance σ^2.

The null hypothesis $H_0 : \mu_1 = \mu_2$ is tested against $H_1 : \mu_2 > \mu_1$. With the variances assumed to be the same for the treatment and control responses, the situations of known and unknown variances are investigated. Expressions are derived for the conditional and unconditional powers of the two designs in both cases. The total number of patients n is assumed to be 20. Therefore, the absolute difference in the numbers of patients on the control group and treatment group are all even numbers from 0 to 20. In particular, four tables of numerical values for the powers are provided. In addition, the author has included the powers of the biased coin design with different values of p. The conclusion is that in both the known variance and the unknown variance cases, the biased coin design is uniformly more powerful than complete randomization and the power increases when the value of p in the biased coin design increases.

In what follows, extended work on the power based on Chen (2006) paper will be presented. The theoretical analysis and numerical values produced for the powers under complete randomization and the Efron's biased coin design will be given both when the variances of the responses are known and different and when the variances of the

responses are unknown and different.

2.2.2 Exact power with known and different variances

Let σ_1 and σ_2 be the standard deviations for the control and treatment groups, respectively. Further, let $\bar{X}_{n_1}$ and $\bar{Y}_{n_2}$ be the mean responses for the control and treatment groups, respectively. Then, for known variances, we carry out a z-test on the treatment responses. Let z_α be the number that satisfies $1 - \Phi(z_\alpha) = \alpha$ where Φ is the cumulative distribution function of the standard normal distribution.

We reject H_0 if and only if

$$\frac{\bar{Y}_{n_2} - \bar{X}_{n_1}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > z_\alpha.$$

Let

$$\delta = \frac{\mu_2 - \mu_1}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}. \tag{2.2}$$

Then the power of this test given a particular allocation status is the probability of rejecting H_0 for H_1 given a particular allocation status of n_1 and n_2. It is equal to

$$\beta_{Z,\alpha}(\mu_2 - \mu_1, \sigma_1, \sigma_2 | n_1, n_2) = \Phi(\delta - z_\alpha).$$

By, multiplying the conditional power by the probability of a particular allocation status and then summing over allocation status, the unconditional power of the test can be obtained.

We need to know the probability mass function of the allocation status. Under complete randomization, each of the patients is equally likely to be assigned to the treatment group or the control group. The probability of obtaining a particular allocation status n_1 and n_2 is

$$P_{CR}\{(n_1, n_2)\} = \binom{n}{n_1}(1/2)^n. \tag{2.3}$$

So the power of the α-level z-test under the complete randomization is

$$\begin{aligned}\beta_{Z,\alpha,CR}(\mu_2-\mu_1,\sigma_1,\sigma_2) &= \sum_{n_1+n_2=n} \beta_{Z,\alpha}(\mu_2-\mu_1,\sigma_1,\sigma_2|n_1,n_2)P_{CR}\{(n_1,n_2)\} \\ &= \sum_{n_1+n_2=n} \Phi(\delta-z_\alpha)\binom{n}{n_1}(1/2)^n.\end{aligned}$$

For the biased coin design, there is a probability $p > 1/2$ of allocating the next patient to the treatment that has been chosen less often. Under this design, the probability of obtaining a particular allocation status n_1 an n_2 depends on the absolute difference in the numbers of patients on the treatment and control groups. Full details about the allocation status under the biased coin design are given in Chen (1999). The key points are as follows. Let D_n be the absolute difference in the numbers of patients on the two treatments up to time n, so that $D_n = |n_1 - n_2|$. This D_n forms a Markov chain with period 2 on the state space $\{0, 1, 2, ...\}$ with $P^{(n)}_{l,m} = P(D_n = m|D_0 = l)$ as its n-step transition probabilities for $l, m, n \geq 0$. These probabilities yield the probability of obtaining a particular allocation status under the biased coin design. Note that

- $P^{(1)}_{0,1} = P(D_{n+1} = 1|D_n = 0) = 1$ for any $n \in \mathbb{Z}$ and $n \geq 0$.
- $P^{(1)}_{k,k+1} = P(D_{n+1} = k+1|D_n = k) = 1-p = q$ for $q < 1/2$ and $k \geq 1$.
- $P^{(1)}_{k,k-1} = P(D_{n+1} = k-1|D_n = k) = p$ for $k \geq 1$.
- for $i \neq j$, $i, j \geq 0$ and $|i-j| \neq 1$, $P^{(1)}_{i,j} = P(D_{n+1} = i|D_n = j) = 0$.

From Karlin and McGregor (1957), a formula for the n-step transition probabilities is

$$\begin{aligned}P^{(n)}_{l,m} &= P(D_n = m|D_0 = l) \\ &= \eta_m + (-1)^{l+m+n}\eta_m + 2pq^2\frac{\eta_m}{\eta_0}\int_0^1 \frac{\{2\sqrt{pq}\cos(\pi x)\}^n h_l(x)h_m(x)}{1-4pq\cos^2(\pi x)}\,dx,\end{aligned}$$

where

$$\eta_0 = \frac{p-q}{2p}$$

and

$$\eta_m = \frac{(p-q)q^{m-1}}{2p^{m+1}} \quad \text{for } m \geq 1$$

are the stationary distribution of D_n, and

$$h_m(x) = \left(\frac{p}{q}\right)^{\frac{m+1}{2}} \sin\{(m-1)\pi x\} - \left(\frac{p}{q}\right)^{\frac{m-1}{2}} \sin\{(m+1)\pi x\}.$$

The probability of obtaining a particular allocation status (n_1, n_2) is

$$P_{BCD(p)}\{(n_1,n_2)\} = \begin{cases} P_{0,0}^{(n)} = P(D_n = 0|D_0 = 0) & \text{if } n_1 = n_2, \\ \frac{1}{2}P_{0,|n_1-n_2|}^{(n)} = \frac{1}{2}P(D_n = |n_1 - n_2||D_0 = 0) & \text{if } n_1 \neq n_2. \end{cases}$$

So the unconditional power under Efron's biased coin design is

$$\begin{aligned} \beta_{Z,\alpha,BCD(p)}(\mu_2-\mu_1,\sigma_1,\sigma_2) &= \Phi(\frac{\mu_2-\mu_1}{\sqrt{(2\sigma_1^2+2\sigma_2^2)/n}} - z_\alpha)P_{0,0}^{(n)} \\ &+ \sum_{n_1+n_2=n, n_1\neq n_2} \Phi(\delta - z_\alpha)\frac{1}{2}P_{0,|n_1-n_2|}^{(n)}. \end{aligned}$$

One of the main aims of the above two designs is to achieve a balance in the numbers of patients on the two treatments. However, since the variances of the patients' responses in the two treatment groups are not the same, the power of the test is not maximized when the numbers of patients in the two treatment groups are the same. This means that the optimal allocation for maximum power is not $n_1 = n_2$. Neyman allocation gives the optimal allocation ratio for the numbers of patients on treatments 1 and 2. This ratio is determined by the values for the variances of the patients' responses in the two treatment groups and is given by

$$\rho_1 = \frac{\sigma_1}{\sigma_1 + \sigma_2}$$

for treatment 1 and $1 - \rho_1$ for treatment 2. We can use this ratio to determine the optimal numbers of patients in the two treatment groups. The probability of obtaining a particular allocation status under Neyman optimal allocation (NOA) will be determined by this ratio.

We consider both the deterministic and the random cases under Neyman allocation. In the deterministic case, the probability of obtaining a particular allocation status (n_1, n_2) under Neyman allocation is

$$P_{NOA(D)}\{(n_1, n_2)\} = \begin{cases} 1 & \text{if } n_1 = round(n\rho_1) \text{ and } n_2 = round\{n(1-\rho_1)\}, \\ 0 & \text{otherwise}, \end{cases}$$

where the notation $round(x)$ denotes rounding the value of x to the nearest integer. The above probability mass function is an indicator function and can also be written as $\mathbf{1}_{NOA(D)}\{(n_1, n_2)\}$. So the power of the α-level z-test under Neyman allocation in the deterministic case is

$$\begin{aligned} \beta_{Z,\alpha,NOA(D)}(\mu_2 - \mu_1, \sigma_1, \sigma_2) &= \sum_{n_1+n_2=n} \beta_{Z,\alpha}(\mu_2 - \mu_1, \sigma_1, \sigma_2 | n_1, n_2) P_{NOA(D)}\{(n_1, n_2)\} \\ &= \sum_{n_1+n_2=n} \Phi(\delta - z_\alpha) \mathbf{1}_{NOA(D)}\{(n_1, n_2)\}. \end{aligned}$$

Next, we study the random case under Neyman allocation. Here, each patient has a probability ρ_1 of being allocated to treatment 1 and a probability $1-\rho_1$ of being allocated to treatment 2. Therefore, $n_1 \sim \text{Bin}(n, \rho_1)$ and $n_2 \sim \text{Bin}(n, 1-\rho_1)$. The probability of obtaining a particular allocation status (n_1, n_2) in the random case is

$$P_{NOA(R)}\{(n_1, n_2)\} = \binom{n}{n_1} \rho_1^{n_1} (1-\rho_1)^{n_2}.$$

So the unconditional power in the random case for Neyman allocation is

$$\begin{aligned} \beta_{Z,\alpha,NOA(R)}(\mu_2 - \mu_1, \sigma_1, \sigma_2) &= \sum_{n_1+n_2=n} \beta_{Z,\alpha}(\mu_2 - \mu_1, \sigma_1, \sigma_2 | n_1, n_2) P_{NOA(R)}\{(n_1, n_2)\} \\ &= \sum_{n_1+n_2=n} \Phi(\delta - z_\alpha) \binom{n}{n_1} \rho_1^{n_1} (1-\rho_1)^{n_2}. \end{aligned}$$

The new numerical results for the powers under complete randomization and the biased coin design will be shown in the following three tables when we aim to balance the numbers of patients on the two treatments. In addition, results for the power for

Neyman allocation are given for comparison. We took $n = 20$, $\bar{d} = \mu_2 - \mu_1$ and $\alpha = 0.05$. Different values of p are considered. The greater the value of p in the BCD(p), the more deterministic the design. Different values for the treatment difference $\bar{d}$ have also been considered. Three sets of values for σ_1^2 and σ_2^2 are studied.

Table 2.1: Powers of CR, BCD(p), NOA(D) and NOA(R) with $\sigma_1^2 = 0.5$, $\sigma_2^2 = 1$, $n = 20$ and $\alpha = 0.05$

	$\bar{d} = 0$	$\bar{d} = 0.5$	$\bar{d} = 1$	$\bar{d} = 1.5$	$\bar{d} = 2$
CR(p=1/2)	0.05	0.3504	0.8074	0.9816	0.9993
BCD(p=7/12)	0.05	0.3567	0.8178	0.9849	0.9997
BCD(p=8/12)	0.05	0.3595	0.8223	0.9862	0.9997
BCD(P=9/12)	0.05	0.3607	0.8241	0.9867	0.9998
BCD(p=10/12)	0.05	0.3612	0.8249	0.9869	0.9998
BCD(p=11/12)	0.05	0.3615	0.8253	0.9870	0.9998
BCD(p=1)	0.05	0.3617	0.8257	0.9871	0.9998
NOA(D)	0.05	0.3686	0.8349	0.9888	0.9998
NOA(R)	0.05	0.3560	0.8163	0.9841	0.9995

Table 2.2: Powers of CR, BCD(p), NOA(D) and NOA(R) with $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $n = 20$ and $\alpha = 0.05$

	$\bar{d} = 0$	$\bar{d} = 0.5$	$\bar{d} = 1$	$\bar{d} = 1.5$	$\bar{d} = 2$
CR(p=1/2)	0.05	0.2257	0.5543	0.8458	0.9701
BCD(p=7/12)	0.05	0.2292	0.5641	0.8557	0.9746
BCD(p=8/12)	0.05	0.2308	0.5684	0.8598	0.9764
BCD(P=9/12)	0.05	0.2315	0.5702	0.8615	0.9771
BCD(p=10/12)	0.05	0.2318	0.5710	0.8623	0.9774
BCD(p=11/12)	0.05	0.2320	0.5715	0.8627	0.9775
BCD(p=1)	0.05	0.2321	0.5718	0.8630	0.9776
NOA(D)	0.05	0.2361	0.5819	0.8713	0.9802
NOA(R)	0.05	0.2289	0.5629	0.8541	0.9737

Table 2.3: Powers of CR, BCD(p), NOA(D) and NOA(R) with $\sigma_1^2 = 0.5$, $\sigma_2^2 = 2$, $n = 20$ and $\alpha = 0.05$

	$\bar{d} = 0$	$\bar{d} = 0.5$	$\bar{d} = 1$	$\bar{d} = 1.5$	$\bar{d} = 2$
CR(p=1/2)	0.05	0.2537	0.6231	0.8975	0.9851
BCD(p=7/12)	0.05	0.2570	0.6318	0.9060	0.9885
BCD(p=8/12)	0.05	0.2584	0.6357	0.9096	0.9899
BCD(P=9/12)	0.05	0.2590	0.6374	0.9111	0.9904
BCD(p=10/12)	0.05	0.2593	0.6381	0.9117	0.9906
BCD(p=11/12)	0.05	0.2594	0.6385	0.9121	0.9907
BCD(p=1)	0.05	0.2595	0.6388	0.9123	0.9907
NOA(D)	0.05	0.2771	0.6780	0.9352	0.9949
NOA(R)	0.05	0.2679	0.6568	0.9215	0.9915

The first column of each table gives the probabilities of rejecting the null hypothesis H_0 when H_0 is true. We can see that these values obtained from our equations match with the assumed significance level of the test $\alpha = 0.05$. From the tables, we can also conclude that the biased coin design is uniformly more powerful than complete randomization for the case where the control and treatment responses have different but known variances. The power function increases as the p in the biased coin design increases from $7/12$ to $11/12$ as well.

For Neyman allocation, the powers obtained are higher in the deterministic case than in the random case for the same value of $\bar{d}$. Furthermore, these powers in the deterministic case are higher than their corresponding powers under complete randomization and the biased coin design for all values of p. Consider the case when $\sigma_1^2 = 0.5$ and $\sigma_2^2 = 2$. The sum of the variances of the patients' responses on the two treatments is the largest in this case. The powers obtained here for Neyman allocation in the random case are higher than their corresponding powers under complete randomization and the biased coin design for all values of p. However, for the other two cases, the powers for Neyman allocation in the random case are higher than the corresponding powers under complete randomization but are lower than those for the biased coin design for all values of p.

2.2.3 Exact power with unknown and different variances

Welch's (1938) approximate t-test is often used when the variances of the treatment and control responses are unknown and different. The degrees of freedom are chosen so that the test statistic under the null hypothesis has approximately a t distribution. Let again $\bar{X}_{n_1}$ and $\bar{Y}_{n_2}$ be the mean responses for the control and treatment groups, respectively. Further, let s_2^2 and s_1^2 be the sample variances for the treatment and control

groups, respectively. Then the test statistic is

$$T = \frac{\bar{Y}_{n_2} - \bar{X}_{n_1}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

The degrees of freedom are

$$\nu = \frac{(\frac{s_2^2}{n_2} + \frac{s_1^2}{n_1})^2}{\frac{s_2^4}{n_2^2(n_2-1)} + \frac{s_1^4}{n_1^2(n_1-1)}}.$$

We reject H_0 if and only if $T > t_{\nu,\alpha}$, where $t_{\nu,\alpha}$ is the right-tailed α-level critical value of the t-test with degrees of freedom ν. Under $H_1 : \mu_2 > \mu_1$, the test statistic T has approximately a non-central t distribution with ν degrees of freedom and non-centrality parameter δ in (2.2). When the variances of the control and treatment responses are unknown and different, the power of this test given a particular allocation status is the probability of rejecting H_0 given a particular allocation status n_1 and n_2. It is equal to

$$\beta_{T,\alpha}(\mu_2 - \mu_1, \sigma_1, \sigma_2, s_1, s_2 | n_1, n_2) = 1 - \mathcal{T}_{\nu,\delta}(t_{\nu,\alpha}), \tag{2.4}$$

where $\mathcal{T}_{\nu,\delta}$ is the cumulative distribution function of the non-central t distribution with ν degrees of freedom and noncentrality parameter δ. By multiplying the conditional power by the probability of a particular allocation status and summing over allocation status the unconditional power of the test can be obtained.

Now, we know the probability mass function of the allocation status n_1 and n_2 under complete randomization from (2.3). Therefore, the power of the test under complete randomization is

$$\beta_{T,\alpha,CR}(\mu_2 - \mu_1, \sigma_1, \sigma_2, s_1, s_2) = \sum_{n_1+n_2=n} \{1 - \mathcal{T}_{\nu,\delta}(t_{\nu,\alpha})\} \binom{n}{n_1} (1/2)^n.$$

For the biased coin design, the probability of obtaining a particular allocation status n_1 and n_2 depends on the absolute difference D_n in the numbers of patients on the two treatments up to time n. This D_n forms a Markov chain on the state space $\{0, 1, 2, ...\}$

with $P_{l,m}^{(n)} = P(D_n = m|D_0 = l)$ as its n-step transition probabilities for $l, m, n \geq 0$ given in (2.4). So the power of the test is

$$\begin{aligned}\beta_{T,\alpha,BCD}(\mu_2 - \mu_1, \sigma_1, \sigma_2, s_1, s_2) &= \{1 - \mathcal{T}_{\nu,\delta}(t_{\nu,\alpha})\}P_{0,0}^{(n)} \\ &+ \sum_{n_1+n_2=n, n_1 \neq n_2} \{1 - \mathcal{T}_{\nu,\delta}(t_{\nu,\alpha})\}\frac{1}{2}P_{0,|n_1-n_2|}^{(n)}.\end{aligned}$$

The above two expressions give the powers of complete randomization and the biased coin design when the variances of the control and treatment responses are different and unknown. These are both expressed in terms of the means of the control and treatment responses, the standard deviations of the control and treatment responses, and the sample standard deviations of the control and treatment responses; which are the parameters μ_1, μ_2, σ_1 and σ_2, and the variables s_1 and s_2, respectively. The sample variances s_1^2 and s_2^2 are two random variables which take any non-negative values.

In order to produce numerical results for the powers of these two designs, the conditional power is needed where neither of the variables s_1 and s_2 are involved. This can be obtained by integrating the product of the conditional power and the joint probability density function of s_1 and s_2, with respect to s_1 and s_2. The powers of the two designs are then calculated by multiplying this new conditional power by the probability mass function of the allocation status and summing. The original conditional power of the test when the variances of the treatment and control responses are different and unknown is given in (2.4), where both $\mathcal{T}_{\nu,\delta}$ and ν involve the variables s_1 and s_2. Let $f(s_1^2, s_2^2|n_1, n_2)$ be the joint probability density function of s_1^2 and s_2^2 given the allocation status (n_1, n_2). Then the new conditional power is

$$\begin{aligned}& C_{T,\alpha}(\mu_2 - \mu_1, \sigma_1, \sigma_2|n_1, n_2) \\ = & \int_0^\infty \int_0^\infty \beta_{T,\alpha}(\mu_2 - \mu_1, \sigma_1, \sigma_2, s_1, s_2|n_1, n_2) f(s_2^2, s_1^2|n_1, n_2)\, \mathrm{d}s_1^2 \mathrm{d}s_2^2 \\ = & \int_0^\infty \int_0^\infty \{1 - \mathcal{T}_{\nu,\delta}(t_{\nu,\alpha})\} f(s_2^2, s_1^2|n_1, n_2)\, \mathrm{d}s_1^2 \mathrm{d}s_2^2.\end{aligned}$$

We know, $(n_1-1)s_1^2/\sigma_1^2 \sim \chi^2_{n_1-1}$ and that $(n_2-1)s_2^2/\sigma_2^2 \sim \chi^2_{n_2-1}$. Let $y_1 = s_1^2, y_2 = s_2^2, x_1 = (n_1-1)s_1^2/\sigma_1^2$ and $x_2 = (n_2-1)s_2^2/\sigma_2^2$. Then the probability density functions of x_1 and x_2 are

$$f(x_1|n_1) = \frac{x_1^{\frac{n_1-1}{2}-1}e^{-x_1/2}}{2^{\frac{n_1-1}{2}}\Gamma\left(\frac{n_1-1}{2}\right)}$$

for $x_1 > 0$ and

$$f(x_2|n_2) = \frac{x_2^{\frac{n_2-1}{2}-1}e^{-x_2/2}}{2^{\frac{n_2-1}{2}}\Gamma\left(\frac{n_2-1}{2}\right)},$$

for $x_2 > 0$, where Γ is the gamma function. By the method of change of variables, the probability density function of y_1 is

$$\begin{aligned} f(y_1|n_1) &= \left|\frac{\mathrm{d}x_1}{\mathrm{d}y_1}\right| f\left(x_1 = \frac{(n_1-1)y_1}{\sigma_1^2}\middle| n_1\right) \\ &= \left|\frac{n_1-1}{\sigma_1^2}\right| \frac{\left(\frac{(n_1-1)y_1}{\sigma_1^2}\right)^{\frac{n_1-1}{2}-1} e^{-\left(\frac{(n_1-1)y_1}{\sigma_1^2}\right)/2}}{2^{\frac{n_1-1}{2}}\Gamma\left(\frac{n_1-1}{2}\right)}, \end{aligned}$$

and similarly for the probability density function of y_2. So the probability density functions of s_1^2 and s_2^2 are

$$f(s_1^2|n_1) = \frac{n_1-1}{2^{\frac{n_1-1}{2}}\Gamma(\frac{n_1-1}{2})\sigma_1^2}(\frac{n_1-1}{\sigma_1^2}s_1^2)^{\frac{n_1-1}{2}-1}e^{-(\frac{n_1-1}{\sigma_1^2}s_1^2)/2} \tag{2.5}$$

and

$$f(s_2^2|n_2) = \frac{n_2-1}{2^{\frac{n_2-1}{2}}\Gamma(\frac{n_2-1}{2})\sigma_2^2}(\frac{n_2-1}{\sigma_2^2}s_2^2)^{\frac{n_2-1}{2}-1}e^{-(\frac{n_2-1}{\sigma_2^2}s_2^2)/2}.$$

The treatment responses are independent and normally distributed with mean μ_2 and variance σ_2^2, the control responses are independent and normally distributed with mean μ_1 and variance σ_1^2, and the control responses are independent of the treatment responses. Therefore, the random variables s_1^2 and s_2^2 are also independent. The joint probability density function of s_1^2 and s_2^2 is just the product of their individual probability density functions, so that

$$f(s_1^2, s_2^2|n_1, n_2) = f(s_1^2|n_1)f(s_2^2|n_2). \tag{2.6}$$

Finally, the power of complete randomization when the variances of the treatment and control responses are unknown and different is

$$\beta_{T,\alpha,CR}(\mu_2-\mu_1,\sigma_1,\sigma_2)=\sum_{n_1+n_2=n} C_{T,\alpha}(\mu_2-\mu_1,\sigma_1,\sigma_2|n_1,n_2)\binom{n}{n_1}(1/2)^n.$$

Similarly, the power of the biased coin design is

$$\begin{aligned}\beta_{T,\alpha,BCD(p)}(\mu_2-\mu_1,\sigma_1,\sigma_2) &= C_{T,\alpha}(\mu_2-\mu_1,\sigma_1,\sigma_2|n_1,n_2)P_{0,0}^{(n)} \\ &+ \sum_{n_1+n_2=n,n_1\neq n_2} C_{T,\alpha}(\mu_2-\mu_1,\sigma_1,\sigma_2|n_1,n_2)\frac{1}{2}P_{0,|n_1-n_2|}^{(n)}.\end{aligned}$$

We assume that the total number of patients in the trial is 20, so that $n=20$. There are in total 21 combinations of values for n_1 and n_2 which give $n=n_1+n_2=20$. There are two special cases for which a specific joint probability density function for s_1^2 and s_2^2 is used instead of (2.6).

- When $n_1=0$, then $n=n_2=20$. All the patients are allocated to the treatment group and no patient is in the control group, which means there will be no sample variance for the control responses and hence no probability density function for s_1^2. The joint probability density function of s_1^2 and s_2^2 is just the probability density function of s_2^2 alone, so that $f(s_1^2,s_2^2|n_1,n_2)=f(s_2^2|n_2)$. Similarly, when $n_2=0$, $n=n_1=20$ and there is no sample variance for the treatment group. The joint probability density function of s_1^2 and s_2^2 is now $f(s_1^2,s_2^2|n_1,n_2)=f(s_1^2|n_1)$.

- When $n_1=1$, then $n_2=n-n_1=19$. Only one patient is allocated to the control group. As there is only one patient in the control group, the sample variance of the control responses will be $s_1^2=0$. Then, from (2.5), $f(s_1^2|n_1)=0$ and the joint probability density function is $f(s_1^2,s_2^2|n_1,n_2)=0$. Similarly, when $n_2=1$, the sample variance of the treatment responses is $s_2^2=0$, and hence the joint probability density function is $f(s_1^2,s_2^2|n_1,n_2)=0$.

The new numerical values for the powers of complete randomization and the biased coin design are presented in the following three tables.

Table 2.4: Powers of CR and BCD(p) with $\sigma_1^2 = 0.5$, $\sigma_2^2 = 1$, $n = 20$ and $\alpha = 0.05$

	$\bar{d} = 0$	$\bar{d} = 0.5$	$\bar{d} = 1$	$\bar{d} = 1.5$	$\bar{d} = 2$
CR(p=1/2)	0.05	0.3270	0.7683	0.9686	0.9975
BCD(p=7/12)	0.05	0.3345	0.7830	0.9756	0.9990
BCD(p=8/12)	0.05	0.3379	0.7893	0.9781	0.9993
BCD(P=9/12)	0.05	0.3393	0.7918	0.9790	0.9994
BCD(p=10/12)	0.05	0.3400	0.7929	0.9794	0.9994
BCD(p=11/12)	0.05	0.3403	0.7935	0.9796	0.9995
BCD(p=1)	0.05	0.3406	0.7939	0.9797	0.9995

Table 2.5: Powers of CR and BCD(p) with $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $n = 20$ and $\alpha = 0.05$

	$\bar{d} = 0$	$\bar{d} = 0.5$	$\bar{d} = 1$	$\bar{d} = 1.5$	$\bar{d} = 2$
CR(p=1/2)	0.05	0.2124	0.5177	0.8090	0.9530
BCD(p=7/12)	0.05	0.2166	0.5299	0.8233	0.9615
BCD(p=8/12)	0.05	0.2185	0.5354	0.8293	0.9648
BCD(P=9/12)	0.05	0.2193	0.5376	0.8318	0.9660
BCD(p=10/12)	0.05	0.2197	0.5386	0.8328	0.9665
BCD(p=11/12)	0.05	0.2199	0.5392	0.8334	0.9668
BCD(p=1)	0.05	0.2200	0.5396	0.8338	0.9670

Table 2.6: Powers of CR and BCD(p) with $\sigma_1^2 = 0.5$, $\sigma_2^2 = 2$,$n = 20$ and $\alpha = 0.05$

	$\bar{d} = 0$	$\bar{d} = 0.5$	$\bar{d} = 1$	$\bar{d} = 1.5$	$\bar{d} = 2$
CR(p=1/2)	0.05	0.2367	0.5797	0.8618	0.9717
BCD(p=7/12)	0.05	0.2404	0.5906	0.8745	0.9791
BCD(p=8/12)	0.05	0.2420	0.5954	0.8800	0.9819
BCD(P=9/12)	0.05	0.2427	0.5974	0.8822	0.9829
BCD(p=10/12)	0.05	0.2430	0.5983	0.8832	0.9833
BCD(p=11/12)	0.05	0.2432	0.5988	0.8837	0.9836
BCD(p=1)	0.05	0.2433	0.5991	0.8841	0.9837

The first column of each table shows the significance level of the test, which gave the assumed significance level $\alpha = 0.05$. From the tables, we can see that the biased coin design is uniformly more powerful than complete randomization for the case where the control and treatment responses have different and unknown variances. Also, the power function increases as the p in the biased coin design increases. The power is smaller in the case when the assumed variances of the control and treatment responses are $\sigma_1^2 = 1$ and $\sigma_2^2 = 2$ than in the case when the assumed variances are $\sigma_1^2 = 0.5$ and $\sigma_2^2 = 1$. The sum of the variances is smaller in the latter case and leads to a higher power. Therefore, when the variances of the treatment and control responses are unknown and different, the power increases when the sum of the two variances is smaller.

By comparing the powers of complete randomization and Efron's biased coin design above to the corresponding powers when the variances of the treatment and control responses are known and different, the powers here are slightly less. This is due to the t-test being only approximate and the variances of the patients' responses on the two treatments being unknown. There is some variation in the values of the powers of the

two designs due to the approximation. The first column in the table has values which give the assumed significance level $\alpha = 0.05$.

2.3 Simulation of power under Efron's biased coin design and the adjustable biased coin design with normal responses

In this section, the powers under Efron's biased coin design and the adjustable biased coin design will be studied by simulation. Baldi Antognini (2008) considered the known variance case for the ABCD and investigated the power function of the z-test. He then describes the Markovian properties of the BCD. A theoretical analysis is given of the power by comparing the adjustable biased coin design with Efron's (1971) BCD and Wei's (1978) adaptive biased coin design. Theoretically, the adjustable biased coin design is shown to give a uniformly more powerful z-test than the other two designs.

Consider an experiment with $n = 20$ patients. Each of them arrive sequentially and have to be assigned to one of the treatments immediately. Both the biased coin design and the adjustable biased coin design have a probability which biases the allocation of a patient in favour of an under-represented treatment. The biased coin design has a fixed probability $p > 1/2$ of allocating a patient to an under-represented treatment. The adjustable biased coin design has a probability $F_a(x)$ of allocating a patient to a treatment. This probability varies with the numbers of patients currently in the two groups, unlike the p in Efron's biased coin design.

The powers of Efron's biased coin design and the adjustable biased coin design were produced by simulations in R. Patients arrive one by one and are assigned to one of the treatments immediately. The assignment rule is as follows. Firstly, we randomly

generate a number from a uniform distribution $U[0,1]$. When the first patient arrives, we have a probability of $1/2$ of assigning this patient to either of the treatments. The rules for the biased coin design and the adjustable biased coin design will be applied from the second patient onwards. For the biased coin design with fixed $p > 1/2$, if the random number is less than or equal to p, then we will assign the patient to the under-represented treatment group, and otherwise to the other group. If the random number generated is less than or equal to the value $F_a(x)$, we will allocate the patient to the control group, and otherwise to the treatment group. The difference in the numbers of patients on the two treatments is calculated after each patient is allocated to a treatment.

Twenty random numbers are generated in R to represent twenty patients and to allocate them to one of the treatments according to the biased coin rule. This is repeated for the adjustable biased coin design. These twenty assignments are considered to be one trial. The patients are assumed to have normal responses. At the end of a trial, the numbers of patients in the control and treatment groups, n_1 and n_2, will be known. With chosen values for the means and variances, we simulate values from a normal distribution to represent the patients' responses. Thus, we simulate n_1 values from a normal distribution with mean μ_1 and variance σ_1^2 and n_2 values from a normal distribution with mean μ_2 and variance σ_2^2. The averages of the responses for the control and treatment groups, $\bar{X}_{n_1}$ and $\bar{Y}_{n_2}$, respectively, can be calculated from the simulated responses. Then we test $H_0 : \mu_2 = \mu_1$ against $H_1 : \mu_2 > \mu_1$. The test statistics is

$$Z = \frac{\bar{Y}_{n_2} - \bar{X}_{n_1}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

when the variances for the control and treatment groups are the same, that is $\sigma_1^2 = \sigma_2^2 = \sigma^2$, and is

$$Z = \frac{\bar{Y}_{n_2} - \bar{X}_{n_1}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

when the variances for the control and treatment groups are different, that is $\sigma_1^2 \neq \sigma_2^2$. The significance level is assumed to be $\alpha = 0.05$. We reject H_0 if $Z > z_\alpha$.

The above is repeated 10,000 times with Efron's biased coin design and the adjustable biased coin design for different values for the means and variances. The number of tests with H_0 being rejected is counted for the two designs. The proportion of tests for which H_0 is rejected is the estimated power. Simulation results for the estimated powers of Efron's biased coin design and the adjustable biased coin design are shown in the following four tables. Again let $\bar{d} = \mu_2 - \mu_1$. Different values for $\bar{d}$ are used to study the power of the two designs.

Table 2.7: Powers of BCD(p) and ABCD with $\sigma^2 = 1$, $n = 20$ and $\alpha = 0.05$

	$\bar{d} = 0$	$\bar{d} = 0.5$	$\bar{d} = 1$	$\bar{d} = 1.5$	$\bar{d} = 2$
BCD(p=7/12)	0.050	0.300	0.714	0.952	0.998
BCD(p=8/12)	0.051	0.297	0.720	0.953	0.998
BCD(p=9/12)	0.051	0.297	0.728	0.954	0.997
BCD(p=10/12)	0.047	0.298	0.733	0.958	0.997
BCD(p=11/12)	0.052	0.302	0.721	0.957	0.998
ABCD(a=1)	0.047	0.292	0.723	0.954	0.998
ABCD(a=2)	0.050	0.295	0.729	0.955	0.998
ABCD(a=4)	0.048	0.304	0.720	0.959	0.997

Table 2.8: Powers of BCD(p) and ABCD with $\sigma_1^2 = 0.5$, $\sigma_2^2 = 1$, $n = 20$ and $\alpha = 0.05$

	$\bar{d} = 0$	$\bar{d} = 0.5$	$\bar{d} = 1$	$\bar{d} = 1.5$	$\bar{d} = 2$
BCD(p=7/12)	0.050	0.354	0.811	0.984	1
BCD(p=8/12)	0.050	0.358	0.821	0.986	1
BCD(p=9/12)	0.049	0.359	0.827	0.987	1
BCD(p=10/12)	0.052	0.356	0.830	0.986	1
BCD(p=11/12)	0.045	0.359	0.833	0.987	1
ABCD(a=1)	0.050	0.357	0.818	0.986	1
ABCD(a=2)	0.051	0.369	0.826	0.988	1
ABCD(a=4)	0.046	0.368	0.825	0.987	1

Table 2.9: Powers of BCD(p) and ABCD with $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $n = 20$ and $\alpha = 0.05$

	$\bar{d} = 0$	$\bar{d} = 0.5$	$\bar{d} = 1$	$\bar{d} = 1.5$	$\bar{d} = 2$
BCD(p=7/12)	0.052	0.225	0.568	0.849	0.975
BCD(p=8/12)	0.048	0.233	0.564	0.853	0.975
BCD(p=9/12)	0.050	0.240	0.567	0.859	0.979
BCD(p=10/12)	0.048	0.232	0.575	0.859	0.978
BCD(p=11/12)	0.049	0.234	0.571	0.865	0.977
ABCD(a=1)	0.050	0.238	0.581	0.867	0.980
ABCD(a=2)	0.048	0.232	0.579	0.866	0.980
ABCD(a=4)	0.051	0.228	0.573	0.869	0.980

Table 2.10: Powers of BCD(p) and ABCD with $\sigma_1^2 = 0.5$, $\sigma_2^2 = 2$, $n = 20$ and $\alpha = 0.05$

	$\bar{d} = 0$	$\bar{d} = 0.5$	$\bar{d} = 1$	$\bar{d} = 1.5$	$\bar{d} = 2$
BCD(p=7/12)	0.051	0.257	0.638	0.907	0.989
BCD(p=8/12)	0.047	0.249	0.646	0.910	0.991
BCD(p=9/12)	0.048	0.267	0.637	0.909	0.990
BCD(p=10/12)	0.049	0.259	0.639	0.914	0.991
BCD(p=11/12)	0.045	0.260	0.635	0.911	0.990
ABCD(a=1)	0.049	0.255	0.651	0.911	0.990
ABCD(a=2)	0.056	0.256	0.638	0.914	0.992
ABCD(a=4)	0.052	0.260	0.641	0.912	0.990

For both designs, the power increases when $\bar{d}$ increases. In addition, a larger value is obtained for the power when the sum of the variances is smaller. However, the trend in the power of the biased coin design when p increases is not obvious. Therefore, no conclusion can be drawn about the power of the biased coin design when p increases from the simulations. Similarly, no obvious pattern can be seen in the powers of the adjustable biased coin design when a increases. Hence, we cannot draw any conclusions about the power of the ABCD when a increases.

Theoretically, the adjustable biased coin design has been shown by Baldi Antognini (2008) to be uniformly more powerful than Efron's biased coin design. Here, we have tried to quantify the increase in power by simulation. The numerical values for the powers give no evidence that the adjustable biased coin design is uniformly more powerful than Efron's biased coin design. We have also simulated the trial 100,000 times and one million times and no significant increase in the power was shown by using the adjustable biased coin design. Therefore, from the values that have been obtained, we cannot con-

clude that the adjustable biased coin design is uniformly more powerful than Efron's biased coin design. Further theoretical work may be possible to study the degree of increase in power for the adjustable biased coin design over Efron's biased coin design.

2.4 Power under complete randomization and Efron's biased coin design with binary responses

Previously, we assumed that the treatment and control groups have normal responses. In this section, the situation when binary responses are available is considered. Therefore, each patient has a response which follows a Bernoulli distribution. Let X_i for $i = 1, ..., n_1$ represent the response of the ith patient in the control group and let Y_j for $j = 1, ..., n_2$ represent the response of the jth patient in the treatment group such that

$$X_i = \begin{cases} 1 & \text{if the } i\text{th patient in the control group survives,} \\ 0 & \text{otherwise,} \end{cases}$$

and

$$Y_j = \begin{cases} 1 & \text{if the } j\text{th patient in the treatment group survives,} \\ 0 & \text{otherwise.} \end{cases}$$

Let $P(X_i = 1) = p_1$ be the probability that a patient survives in the control group and let $P(Y_j = 1) = p_2$ be the probability that a patient survives and is in the treatment group.

By assumption, the random variables $X_1, X_2, ..., X_{n_1}$ are independent and identically distributed and the random variables of $Y_1, Y_2, ..., Y_{n_2}$ are independent and identically distributed. The control and treatment responses are also independent. Let $X = \sum_{i=1}^{n_1} X_i$ and $Y = \sum_{j=1}^{n_2} Y_j$. Then, given n_1 and n_2, we have $X|n_1 \sim \text{Bin}(n_1, p_1)$ and $Y|n_2 \sim \text{Bin}(n_2, p_2)$. The conditional probability mass functions of these two random

variables are

$$P(X = x|n_1) = \binom{n_1}{x} p_1^x (1-p_1)^{n_1-x}$$

and

$$P(Y = y|n_2) = \binom{n_2}{y} p_2^y (1-p_2)^{n_2-y}.$$

Since the control and treatment responses are independent, the joint conditional probability mass function of the numbers of patients who survive in the control and treatment groups is just the product of their individual conditional probability mass functions given by

$$\begin{aligned} P(X = x, Y = y|n_1, n_2) &= P(X = x|n_1)P(Y = y|n_2) \\ &= \binom{n_1}{x} p_1^x (1-p_1)^{n_1-x} \binom{n_2}{y} p_2^y (1-p_2)^{n_2-y}. \end{aligned}$$

We are interested in whether there is a treatment effect. Here, p_1 and p_2 are the two parameters of interest. A difference in these two parameters will indicate a treatment effect in the case of binary responses. We want to test $H_0 : p_1 = p_2$ against $H_1 : p_2 > p_1$. We reject H_0 if $Y - X > d$ for some positive integer d.

Numerical values for the powers of complete randomization and the biased coin design with binary responses will be produced. The conditional power, which is the probability of correctly detecting a treatment effect given a particular allocation status, is needed first. This is the same as the probability of rejecting H_0 given a particular allocation status, $P(Y - X = y - x > d|n_1, n_2)$, for a chosen critical value $d \in \mathbb{Z}^+$.

In order to obtain the probability mass function of $Y - X$, we use the joint conditional probability mass function of X and Y and a change of variables. Let $A = X + Y$ and $B = Y - X$. Then we have to obtain the joint conditional probability mass function of A and B denoted by $g(a, b|n_1, n_2)$. Then we can sum over all values of a in $g(a, b|n_1, n_2)$. The resulting expression is the conditional probability mass function $g(b|n_1, n_2)$ which

is also the conditional probability mass function of $Y - X$ given n_1 and n_2. Finally, by choosing a critical value d, the probability $P(Y - X = y - x > d|n_1, n_2)$ can be found.

By a change of variables, the joint conditional probability mass function $g(a, b|n_1, n_2)$ is

$$\begin{aligned} g(a, b|n_1, n_2) &= P\left(X = \frac{a-b}{2}, Y = \frac{a+b}{2}\middle| n_1, n_2\right) \\ &= \binom{n_1}{\frac{a-b}{2}} p_1^{\frac{a-b}{2}} (1-p_1)^{n_1-(\frac{a-b}{2})} \binom{n_2}{\frac{a+b}{2}} p_2^{\frac{a+b}{2}} (1-p_2)^{n_2-(\frac{a+b}{2})}. \end{aligned}$$

We have to sum over all values of a in $g(a, b|n_1, n_2)$ to obtain the conditional probability mass function of B. Therefore, the range for a has to be known. We know that x can take any integer value between 0 and n_1. Similarly, y can take any integer value between 0 and n_2. From $x = (a - b)/2$, we know that a is an integer that lies between b and $2n_1 + b$ inclusively. Similarly, from $y = (a + b)/2$, a is an integer that lies between $-b$ and $2n_2 - b$ inclusively. So we have four conditions for the range of a: $a \geq b$, $a \geq -b$, $a \leq 2n_1 + b$ and $a \leq 2n_2 - b$. By drawing the four lines $a = b$, $a = -b$, $a = 2n_1 + b$ and $a = 2n_2 - b$ on a graph, we can identify the values for a and b that satisfies these conditions. The values for a and b are the discrete points inside the shaded area. Three graphs are given to show the shaded area of interest, when $n_1 > n_2$, when $n_1 = n_2$ and when $n_1 < n_2$. The red, blue, yellow and green lines represent the lines with equations $a = b$, $a = -b$, $a = 2n_1 + b$ and $a = 2n_2 - b$, respectively.

Figure 2.1: Values of a and b when $n_1 > n_2$

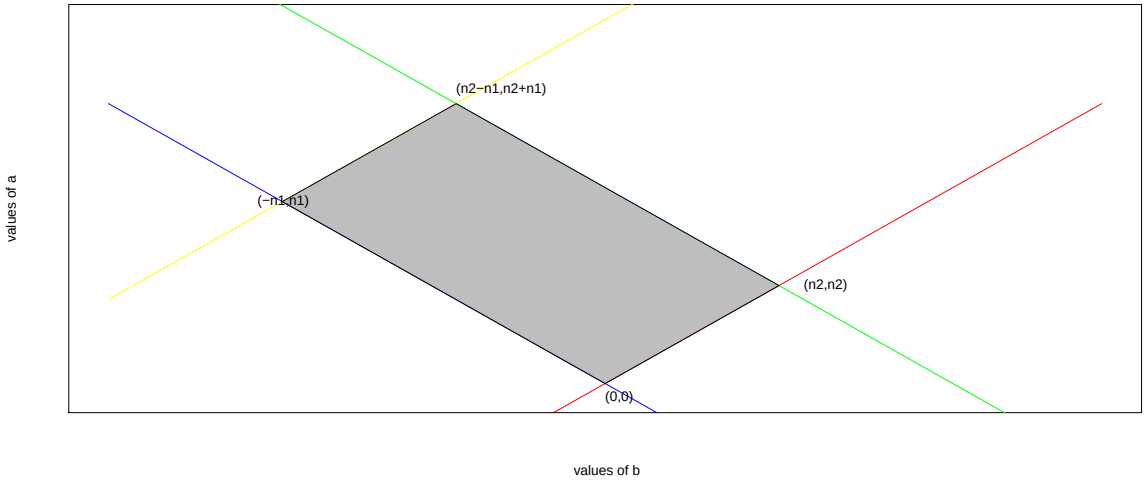


Figure 2.2: Values of a and b when $n_1 = n_2$

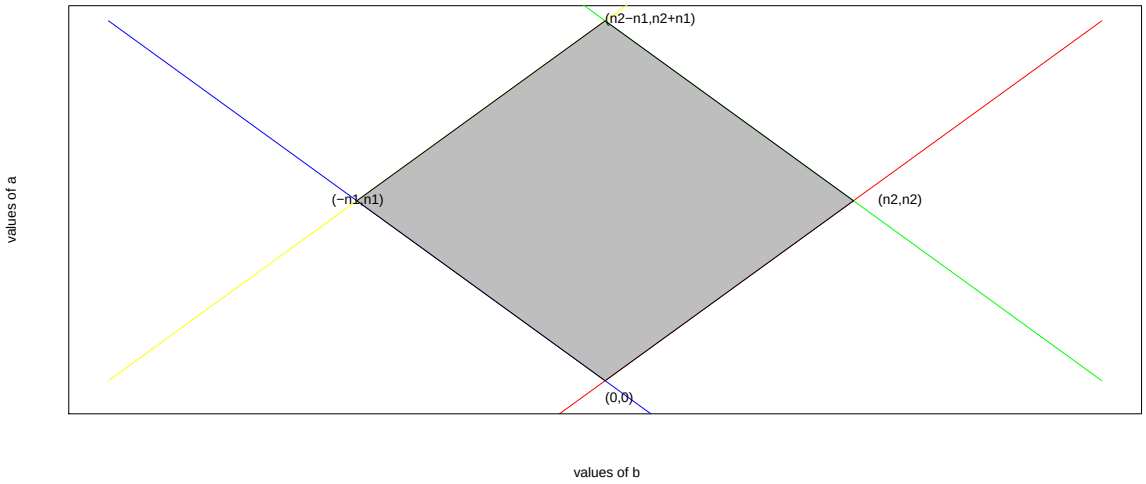
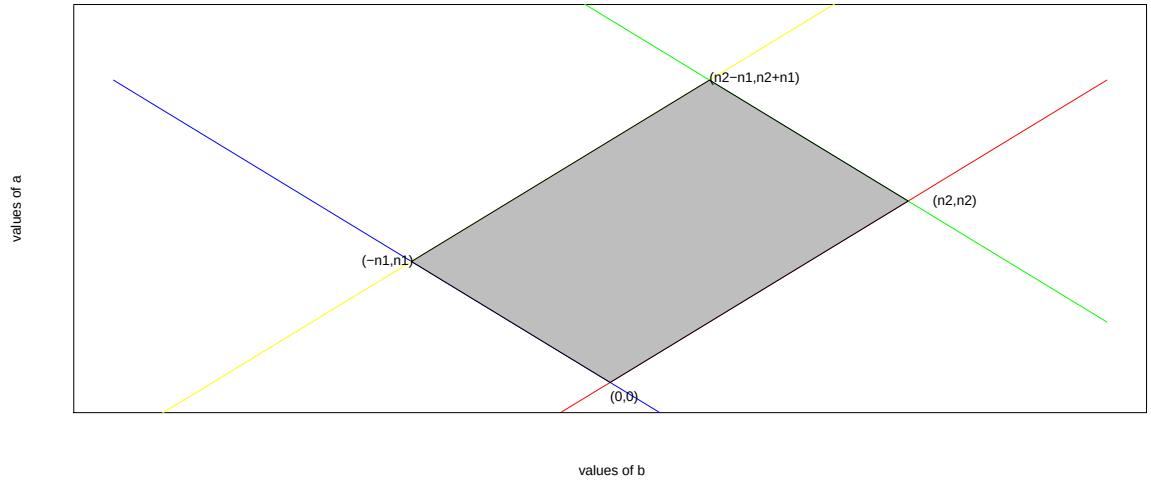


Figure 2.3: Values of a and b when $n_1 < n_2$



The first two conditions $a \geq b$ and $a \geq -b$ imply that a is greater than or equal to $|b|$ and the last two conditions $a \leq 2n_1 + b$ and $a \leq 2n_2 - b$ imply that a is less than or equal to $n - |b + n_1 - n_2|$. We know that all values for a depend on the current value for b. Consider all values for a when b is fixed. Then, we have $a = b + 2x$, so that a takes values in steps of two and $a = \{|b|, |b| + 2, ..., n - |b + n_1 - n_2|\}$. Therefore, we sum over a in steps of two for each value for b in the conditional power.

The conditional power $P(Y - X = b > d|n_1, n_2) = P(Y - X = y - x > d|n_1, n_2)$ for a chosen critical value d is

$$\sum_{b=d}^{n_2} \sum_{a=|b|}^{n-|b+n_1-n_2|} \binom{n_1}{\frac{a-b}{2}} p_1^{\frac{a-b}{2}} (1-p_1)^{n_1-(\frac{a-b}{2})} \binom{n_2}{\frac{a+b}{2}} p_2^{\frac{a+b}{2}} (1-p_2)^{n_2-(\frac{a+b}{2})}.$$

The above formula may be written as

$$(1-p_1)^{n_1}(1-p_2)^{n_2} \sum_{b=d}^{n_2} \left\{ \frac{p_2(1-p_1)}{p_1(1-p_2)} \right\}^{\frac{b}{2}} \sum_{a=|b|}^{n-|b+n_1-n_2|} \binom{n_1}{\frac{a-b}{2}} \binom{n_2}{\frac{a+b}{2}}$$
$$\times \left\{ \frac{p_1 p_2}{(1-p_1)(1-p_2)} \right\}^{\frac{a}{2}}.$$

This is the power given the allocation status n_1 and n_2.

The unconditional power can be obtained by multiplying the conditional power by the probability mass function of the allocation status and summing. The power of com-

plete randomization is

$$\sum_{n_1+n_2=n}(1-p_1)^{n_1}(1-p_2)^{n_2}\sum_{b=d}^{n_2}\left\{\frac{p_2(1-p_1)}{p_1(1-p_2)}\right\}^{\frac{b}{2}}\sum_{a=|b|}^{n-|b+n_1-n_2|}\binom{n_1}{\frac{a-b}{2}}\binom{n_2}{\frac{a+b}{2}}$$
$$\times\left\{\frac{p_1p_2}{(1-p_1)(1-p_2)}\right\}^{\frac{a}{2}}\binom{n}{n_1}\left(\frac{1}{2}\right)^n.$$

and the power of the biased coin design is

$$(1-p_1)^{\frac{n}{2}}(1-p_2)^{\frac{n}{2}}\sum_{b=d}^{\frac{n}{2}}\left\{\frac{p_2(1-p_1)}{p_1(1-p_2)}\right\}^{\frac{b}{2}}\sum_{a=|b|}^{n-|b|}\binom{\frac{n}{2}}{\frac{a-b}{2}}\binom{\frac{n}{2}}{\frac{a+b}{2}}\left\{\frac{p_1p_2}{(1-p_1)(1-p_2)}\right\}^{\frac{a}{2}}$$
$$\times P_{0,0}^{(n)}+\sum_{\substack{n_1+n_2=n,\\ n_1\neq n_2}}(1-p_1)^{n_1}(1-p_2)^{n_2}\sum_{b=d}^{n_2}\left\{\frac{p_2(1-p_1)}{p_1(1-p_2)}\right\}^{\frac{b}{2}}$$
$$\times\sum_{a=|b|}^{n-|b+n_1-n_2|}\binom{n_1}{\frac{a-b}{2}}\binom{n_2}{\frac{a+b}{2}}\left\{\frac{p_1p_2}{(1-p_1)(1-p_2)}\right\}^{\frac{a}{2}}\times\frac{1}{2}P_{0,|n_1-n_2|}^{(n)}.$$

We now consider the power when the significance level is fixed. Numerical values for the power are given for complete randomization and Efron's biased coin design in the following two tables. The first column of each table shows the significance level. By choosing the appropriate critical value d in each case, we want to fix the value for the significance level to be around 0.05. The critical value d is chosen so that this is less than or equal to 0.05. Here, d may not be the same for different designs as the probability distribution of the allocation status is different, since Efron's biased coin design is better than complete randomization in terms of balancing the numbers of patients on the two treatment groups.

Table 2.11: The power of CR and BCD with $\mu = p_2 - p_1$ and $n = 20$

		$\mu = 0$	$\mu = 0.2$	$\mu = 0.4$	$\mu = 0.6$	$\mu = 0.8$
CR(p=1/2)	d=6	0.0430	0.1313	0.3134	0.5745	0.8320
BCD(p=7/12)	d=5	0.0464	0.1708	0.4279	0.7492	0.9547
BCD(p=8/12)	d=5	0.0320	0.1478	0.4221	0.7762	0.9750
BCD(p=9/12)	d=5	0.0259	0.1366	0.4192	0.7900	0.9829
BCD(p=10/12)	d=5	0.0232	0.1311	0.4178	0.7970	0.9861
BCD(p=11/12)	d=5	0.0217	0.1279	0.4170	0.8012	0.9877

Table 2.12: The power of CR and BCD with $\mu = p_2 - p_1$ and $n = 50$

		$\mu = 0$	$\mu = 0.2$	$\mu = 0.4$	$\mu = 0.6$	$\mu = 0.8$
CR(p=1/2)	d=9	0.0443	0.2403	0.6249	0.9234	0.9970
BCD(p=7/12)	d=7	0.0513	0.3523	0.8281	0.9916	1
BCD(p=8/12)	d=7	0.0385	0.3417	0.8483	0.9960	1
BCD(p=9/12)	d=7	0.0489	0.3898	0.8820	0.9980	1
BCD(p=10/12)	d=7	0.0337	0.3370	0.8570	0.9972	1
BCD(p=11/12)	d=7	0.0330	0.3362	0.8584	0.9974	1

From the above tables, we can see that the biased coin design performs much better than complete randomization in terms of power for each value of $\mu = p_2 - p_1$. We took five values of p_1 from 0.5 to 0.1 in steps of 0.1 and p_2 takes five values from 0.5 to 0.9 in steps of 0.1. In Table 2.11, for $\mu = 0.6$ and $\mu = 0.8$, we can also see that the power of the biased coin design increases when p increases. Similar conclusions hold when the number of patients is $n = 50$ in Table 2.12. An increase in the number of patients in the trial gives a larger value for the power in each case.

2.5 Normal approximation for the power under complete randomization and Efron's biased coin design

In this section, we assume that patients' responses in the two treatment groups are normally distributed. We study the numerical values for the power obtained using a normal approximation for complete randomization and Efron's biased coin design, and compare these with the exact values. The exact values were obtained by Chen (2006) when the variances are the same, in Section 2.2.2 when the variances are known and different, and in Section 2.2.3 when the variances are unknown and different.

In Shao, Yu and Zhong (2010), normal approximations to the power are given under different randomization schemes. The responses of patients in the two treatment groups are assumed to be normally distributed. We wish to test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_2 > \mu_1$. The test statistic is

$$T = \frac{\bar{Y}_{n_2} - \bar{X}_{n_1}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}.$$

The sample variances for the standard and new treatments are denoted by s_1^2 and s_2^2, respectively. We will replace s_1^2 and s_2^2 by the population variances σ_1^2 and σ_2^2 when the variances are known. Given a significance level of α, we reject H_0 if $T > c_\alpha$, where c_α is the critical value of a Student's t-distribution or the standard normal distribution when the variances of the patients' responses are unknown or known, respectively.

When the variances of the responses in the two treatment groups are the same, we have $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Then we have,

$$\lim_{n\to\infty} P(T > c_\alpha) = \Phi(-c_\alpha).$$

The unconditional power under H_1 for the two sample t-test under complete randomiza-

tion and the biased coin design is given by

$$P(T > c_\alpha) \approx \Phi\left(\frac{\bar{d}\sqrt{n}}{2\sigma} - c_\alpha\right).$$

This represents the normal approximation to the power. The following three tables give the exact power and the normal approximation to the power for complete randomization and Efron's biased coin design. In the calculations, we took $\sigma^2 = 1$.

Table 2.13: Exact powers of CR and BCD(p) with normal approximation when $n = 20$, $\alpha = 0.05$, $\sigma = 1$ and known

	$\bar{d} = 0$	$\bar{d} = 0.5$	$\bar{d} = 1$	$\bar{d} = 1.5$	$\bar{d} = 2$
CR(p=1/2)	0.05	0.2893	0.7024	0.9461	0.9960
BCD(p=7/12)	0.05	0.2947	0.7139	0.9522	0.9971
BCD(p=8/12)	0.05	0.2972	0.7190	0.9546	0.9975
BCD(p=9/12)	0.05	0.2983	0.7211	0.9555	0.9976
BCD(p=10/12)	0.05	0.2987	0.7220	0.9560	0.9976
BCD(p=11/12)	0.05	0.2990	0.7225	0.9562	0.9976
N.Approx	0.05	0.2992	0.7228	0.9563	0.9977

We can see that the normal approximation to the power is slightly higher than the exact values obtained under complete randomization and Efron's biased coin design when the variances of the responses are the same and known. The differences in the exact and approximate powers under complete randomization are no more than about 2%, while those in the powers under Efron's biased coin design are no more than about 1%. As the exact power is higher for the biased coin design than for complete randomization, the over-approximation is less serious for the biased coin design than for complete randomization.

Table 2.14: Exact powers of CR and BCD(p) with normal approximation when $n = 20$, $\alpha = 0.05$, $\sigma = 0.5$ and unknown

	$\bar{d} = 0$	$\bar{d} = 0.5$	$\bar{d} = 1$	$\bar{d} = 1.5$	$\bar{d} = 2$
CR(p=1/2)	0.05	0.6732	0.9937	1	1
BCD(p=7/12)	0.05	0.6847	0.9952	1	1
BCD(p=8/12)	0.05	0.6897	0.9957	1	1
BCD(p=9/12)	0.05	0.6918	0.9959	1	1
BCD(p=10/12)	0.05	0.6927	0.9959	1	1
BCD(p=11/12)	0.05	0.6932	0.9960	1	1
N.Approx.	0.05	0.7228	0.9977	1	1

Table 2.15: Exact powers of CR and BCD(p) with normal approximation when $n = 20$, $\alpha = 0.05$, $\sigma = 2$ and unknown

	$\bar{d} = 0$	$\bar{d} = 0.5$	$\bar{d} = 1$	$\bar{d} = 1.5$	$\bar{d} = 2$
CR(p=1/2)	0.05	0.1313	0.2755	0.4709	0.6732
BCD(p=7/12)	0.05	0.1329	0.2806	0.4802	0.6847
BCD(p=8/12)	0.05	0.1337	0.2829	0.4843	0.6897
BCD(p=9/12)	0.05	0.1340	0.2839	0.4861	0.6918
BCD(p=10/12)	0.05	0.1341	0.2844	0.4868	0.6927
BCD(p=11/12)	0.05	0.1342	0.2846	0.4873	0.6932
N.Approx	0.05	0.1388	0.2992	0.5128	0.7228

In the case of unknown and equal variances, the normal approximation to the power again gives larger values for the power under both schemes than the exact calculations. The over-approximation is also more serious for complete randomization than for the biased coin design. However, the differences in the exact and approximate values are

greater than those when the variances are the same and known.

When the variances of the responses are different, the unconditional power under the two designs for the two sample t-test is

$$P(T > c_\alpha) \approx \Phi\left(\frac{\bar{d}\sqrt{n}}{\sqrt{2(\sigma_1^2+\sigma_2^2)}} - c_\alpha\right),$$

The next two tables present the exact power and the normal approximation to the power for the two designs.

Table 2.16: Exact powers of CR and BCD(p) with normal approximation when $n = 20$, $\alpha = 0.05$, $\sigma_1^2 = 0.5$, $\sigma_2^2 = 2$ and known

	$\bar{d}=0$	$\bar{d}=0.5$	$\bar{d}=1$	$\bar{d}=1.5$	$\bar{d}=2$
CR(p=1/2)	0.05	0.2537	0.6231	0.8975	0.9851
BCD(p=7/12)	0.05	0.2570	0.6318	0.9060	0.9885
BCD(p=8/12)	0.05	0.2584	0.6357	0.9096	0.9899
BCD(P=9/12)	0.05	0.2590	0.6374	0.9111	0.9904
BCD(p=10/12)	0.05	0.2593	0.6381	0.9117	0.9906
BCD(p=11/12)	0.05	0.2594	0.6385	0.9121	0.9907
BCD(p=1)	0.05	0.2595	0.6388	0.9123	0.9907
N.Approx	0.05	0.2595	0.6388	0.9123	0.9907

For known and different variances, the normal approximation to the power gives the most accurate approximation compared to the previous cases. The normal approximation to the power gives the same results as to the biased coin design when $p = 1$. There is still an over-approximation to the power under the two schemes. The differences in the exact and approximate values are no more than 1% for all $\bar{d}$ and $p \neq 1$ under the biased coin design. The normal approximation to the power also works better for the biased coin design than for complete randomization.

Table 2.17: Exact powers of CR and BCD(p) with normal approximation when $n = 20$, $\alpha = 0.05$, $\sigma_1^2 = 0.5$, $\sigma_2^2 = 1$ and unknown

	$\bar{d} = 0$	$\bar{d} = 0.5$	$\bar{d} = 1$	$\bar{d} = 1.5$	$\bar{d} = 2$
CR(p=1/2)	0.0500	0.3270	0.7683	0.9686	0.9975
BCD(p=7/12)	0.0500	0.3345	0.7830	0.9756	0.9990
BCD(p=8/12)	0.0500	0.3379	0.7893	0.9781	0.9993
BCD(p=9/12)	0.0500	0.3393	0.7918	0.9790	0.9994
BCD(p=10/12)	0.0500	0.3400	0.7929	0.9794	0.9994
BCD(p=11/12)	0.0500	0.3403	0.7935	0.9796	0.9995
BCD(p=1)	0.0500	0.3406	0.7939	0.9797	0.9995
N.Approx	0.05	0.3617	0.8257	0.9871	0.9998

When the variances of the responses are unknown and different, the exact power is obtained for Welch's approximate t-test. The normal approximation to the power again gives larger values for the power than the exact calculations under the two schemes. The differences in the exact and approximate values are no more than 4% for the biased coin design.

The asymptotic expression given for the normal approximation to the power are the same under complete randomization and the biased coin design. The values obtained for the power by the normal approximation are therefore the same under complete randomization and the biased coin design. For each case, the numerical values for the power given by the normal approximation are higher than the exact values. There is an over-estimation to the power. The problem of over approximation is more serious for complete randomization than for the biased coin design.

In the planning stage of a clinical trial, we have to estimate the number of patients

that are needed to achieve a certain power for a given treatment difference. The problem of over approximation to the power will lead the planning team to recruit too few patients. By knowing the degree of over-approximation can help the planners to adjust their sample size estimate accordingly.

2.6 Conclusions

It is clear that complete randomization gives the lowest power among the designs both when patients' responses are binary and normally distributed. Complete randomization schemes are less likely to detect a genuine treatment difference than the other designs. Efron's biased coin design is basically as good as the adjustable biased coin design in terms of balancing the numbers of patients on two treatments.

In this chapter, we have described the simplest case in clinical trials. Restricted randomization schemes are considered where patient assignment only depends on the current numbers of patients on the two treatments. In the next chapter, we will study schemes where covariate information will be used for treatment assignment. In other words, we study covariate-adaptive randomization schemes in two-treatment trials. In Chapter 4, we will study restricted randomization schemes for more than two treatments.

Chapter 3

Covariate-adaptive randomization schemes

3.1 Introduction

Consider a clinical trial in which there are two treatments under comparison. Patients arrive sequentially and have to be assigned to one of the treatments immediately. A randomization scheme can be used for treatment assignment to ensure that there are similar numbers of patients in the two treatment groups and to maintain some randomness in the assignment for valid statistical inference. These randomization schemes have their own assignment rules for which the assignments are made based on the current numbers of patients on the treatments.

In addition, some of the patients' prognostic factors like their age, gender, current health condition and so on will have an influential effect on their responses to different treatments. Therefore, it is sensible to include these prognostic factors as covariates both in the randomization stage and in the analysis stage. When the randomization schemes are applied to patients classified by their prognostic factors, we call these covariate-

adaptive randomization schemes. These schemes ensure that patients with the same prognostic profile are balanced across the treatment groups. One of the most well-known covariate adaptive randomization processes is the minimization procedure developed by Pocock and Simon (1975) and described as follows. We first identify patients' prognostic profiles by classifying them into different levels within each of the covariates. Let the total number of treatments involved be K. Then we calculate the imbalance between the treatments at different levels of the prognostic factors if the next patient is assigned to treatment j for $j = 1, .., K$. If some of the levels of the covariates are more important, more weight will be put on these. The overall imbalance is calculated from the imbalances of each level of the covariates. The next assignment will be made to the treatment which gives the minimum overall imbalance. Covariate-adaptive randomization schemes are also studied by Shao, Yu and Zhong (2010) when there is only one covariate and $K = 2$. Here, the covariate is considered to be a random variable.

In this chapter, we will investigate different covariate-adaptive randomization schemes when $K = 2$. In Section 3.2, the linear models for patients' responses is introduced when there are several covariates. We will only consider a fixed- effects model. Under any covariate-adaptive randomization scheme, we can achieve either global or marginal balance in the numbers of patients in the two treatment groups. Their properties are investigated for two covariates by Baldi Antognini and Zagoraiou (2011). In Section 3.3, global and marginal balance will be described for two or more covariates. The patient assignment rules will be introduced in Section 3.4 for covariate-adaptive simple random sampling, the covariate-adaptive biased coin design and the covariate-adaptive adjustable biased coin design for both global and marginal balance. In Section 3.5, an analysis of covariance t test on the treatment difference under the above three randomization schemes will be described. In Shao, Yu and Zhong (2010), an analytical approach is

given to obtain expressions for normal approximations to the power under different randomization schemes for a single covariate. In Section 3.6, we assess the accuracy of the normal approximation to the power for different scenarios. The corresponding expression for the normal approximation will also be given when there are $p \geq 2$ covariates. Numerical values for the power of the analysis of covariance t test will be simulated in Section 3.7. The values will be simulated for different scenarios under the three randomization schemes for both global and marginal balance. Finally, we draw conclusions in Section 3.8.

3.2 Models for covariate-adaptive randomization

Consider a trial with two treatments and n patients in total. Treatment 1 represents the standard treatment and treatment 2 represents the new treatment. We call the group of patients that receive treatment 1 the control group and the group of patients receiving treatment 2 the treatment group. Let n_1 and n_2 be the numbers of patients in the control and treatment groups, respectively. The response of the ith patient on treatment j is Y_{ij} for $i = 1, .., n$ and $j = 1$ if the ith patient is allocated to treatment 1 or $j = 2$ if the ith patient is allocated to treatment 2. Let I_i be the indicator variables such that

$$I_i = \begin{cases} 0 & \text{if the } i\text{th patient is allocated to treatment 1,} \\ 1 & \text{if the } i\text{th patient is allocated to treatment 2} \end{cases} \tag{3.1}$$

for $i = 1, ..., n$. The response of patient i can be represented as

$$Y_i = I_i Y_{i2} + (1 - I_i) Y_{i1}.$$

3.2.1 Linear model for patients' responses

We are interested in the model for the Y_{ij}. Let Z_{ik} be the kth covariate used in the

randomization process for the ith patient and let the total number of covariates used be p. Then our model for patients' responses may be written in matrix form as

$$\mathbf{Y} = \mathbf{W}\boldsymbol{\mu} + \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{3.2}$$

Here, $\mathbf{Y}$ is the $n \times 1$ column vector which contains all the responses Y_{ij} for $i = 1, ..., n$ and $j = 1$ or 2 such that

$$\mathbf{Y} = \left(Y_{11}, Y_{21}, ..., Y_{n_1 1}, Y_{(n_1+1)2}, Y_{(n_1+2)2}, ..., Y_{n2}\right)^T.$$

The column vectors $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$ represent the parameters in the model. Let μ_1 and μ_2 be the mean responses for treatment 1 and 2, respectively. Then $\boldsymbol{\mu}$ is 2×1 column vector with components μ_1 and $\mu_2 - \mu_1$. The ith patient will take μ_1 in the model when the subject is allocated to treatment 1 and μ_2 when the subject is allocated to treatment 2. Furthermore, W is an $n \times 2$ matrix with ones in the first column and the values of I_i in the second column.

The matrix $\mathbf{Z}$ contains the values of the covariates used in the randomization and their interactions. More specifically, it includes all values of the p covariates Z_{ik} for $i = 1, ..., n$ and $k = 1, ..., p$, together with any interaction terms between them. These interaction terms can be interactions between two or more covariates and will consist of at most

$$\binom{p}{2} + \binom{p}{3} + \ldots + \binom{p}{p}$$

terms. Therefore, $\mathbf{Z}$ is a $n \times q$ matrix, where $p \leq q$.

These covariates are either qualitative or quantitative variables. Quantitative variables refer to variables with numerical values. They can be discrete or continuous. Examples of quantitative covariates are the height and weight of a patient, the amount of alcohol intake per day, the number of packs of cigarettes smoked and so on. The first example is a continuous quantitative variable and the last is a discrete quantitative variable.

Qualitative variables refer to variables that cannot be measured in terms of numbers and are classified into different categories. These variables can be further divided into two categories which are categorical and ordinal. For ordinal variables, the categories are ordered. For example, the current health condition of a patient can be classified into categories in terms of status with levels such as 'very bad', 'normal','very good' and so on. Categorical covariates have no sense of ordering. Gender is an example of categorical covariates for which has the levels 'male' or female'.

We transform all the quantitative and qualitative covariates into factors variables as follows. Consider the covariate Z_{ik} which represents the prognostic profile for the kth covariate of the ith patient. This particular covariate is divided into l_k levels, where l_k represents the total number of levels for the kth covariate and $k = 1, ..., p$. These total numbers of levels for different covariates may be the same.

In the case where the covariates are quantitative, we have to define ranges for each of the levels within each covariate. For a particular covariate k, it will be classified as being in a particular level if its values lie within the range of that level. We allocate an integer value to each level of the covariate. For example, the covariate for age will take the value -1 if the age of a patient is between $0 - 20$, 0 for ages between 21 to 40 and so on. Another covariate, the height of a patient may take value -1 if their height is less than or equal to 100 cm, 0 for height above 100 cm and less than or equal to 120 cm and so on.

When the covariates are qualitative and ordinal, a value will be assigned to each of the levels of the covariates. For example, the health condition of a patient is represented by a value of -2 for 'very bad' condition and 2 for 'very good' condition. For qualitative and categorical covariates, we will also assign a value to each of the categories of a covariate. For example, we take -1 for a female patient and 1 for a male patient.

The values of all of the levels for each covariate add up to zero. This is to ensure that an effect of one level can be compensated for by inverse effects of all the other levels. Therefore, these values are used for the values of the covariates Z_{ik} for $i = 1, .., n$ and $k = 1, ..., p$ in the model.

The $q \times 1$ column vector $\boldsymbol{\beta}$ contains the regression parameters for the covariates and is

$$\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p, \beta_{p+1}, ..., \beta_q)^T .$$

Here, there are in total p covariates used in the randomization and β_1 to β_p are the regression parameters for each of the covariates. The rest of the components in $\boldsymbol{\beta}$ from β_{p+1} onwards are the regression parameters for the interactions between the covariates.

Finally, $\boldsymbol{\epsilon}$ is the $n \times 1$ column vector for the random errors ϵ_{ij} such that

$$\boldsymbol{\epsilon} = \left(\epsilon_{11}, ..., \epsilon_{n_1 1}, \epsilon_{(n_1+1)2}, ..., \epsilon_{n2}\right)^T ,$$

where the ϵ_{ij} for $i = 1, ..., n$ and $j = 1$ or 2 are independent and identically distributed with mean 0 and variance σ^2_ϵ. These random errors ϵ_{ij} are independent of all of the covariates Z_{ik}. The model can be considered as either a fixed-effects or a random-effects model.

3.2.2 Background

In Shao, Yu and Zhong (2010), the model for treatment responses is written in the form

$$Y_{ij} = \mu_j + bZ_i + \epsilon_{ij}.$$

The above equation is a special case of (3.2). Here, Y_{ij} and ϵ_{ij} for $i = 1, .., n$ and $j = 1$ or 2 are the components of the vectors $\mathbf{Y}$ and $\boldsymbol{\epsilon}$, respectively. Also,

$$\boldsymbol{\mu} = (\mu_1, \mu_2)^T \tag{3.3}$$

and

$$\mathbf{W} = \begin{pmatrix} 1 & 1 & \dots & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 \end{pmatrix}^T . \tag{3.4}$$

In this model, only one covariate is considered and hence there are no interaction terms. This covariate is assumed to be a univariate covariate with finite second order moment. The covariate Z_i here is the same as Z_{ik} in (3.2) with $k = p = 1$. So Z_{i1} can be written as Z_i in this case. This means that $\mathbf{Z} = (Z_1, Z_2, ..., Z_n)^T$. In addition, the column vector $\boldsymbol{\beta}$ in (3.2) which contains all the regression parameters for the covariates and their interactions is now just b.

In Shao, Yu and Zhong (2010), the covariates Z_i used in the randomization are random variables. In fact, (Y_{i1}, Y_{i2}, Z_i) for $i = 1, ..., n$ are assumed to be independent random variables from some distribution. The random variables Z_i are independent and identically distributed and can be from a discrete or continuous distribution. The random variables ϵ_{ij} are independent of the Z_i.

When Z_i is discrete, there will be a list of possible values for Z_i. These Z_i take any values in this list. If Z_i is continuous, then we have to define a discrete function of Z_i, denoted by $D(Z_i)$. Under any covariate-adaptive randomization scheme, when Z_i is discrete, we will apply its assignment rule for each value of Z_i. When Z_i is continuous, we will apply the assignment rules to patients in each of the categories defined by $D(Z_i)$. Our aim is to achieve balance between the two treatment groups for each of the prognostic factors.

3.2.3 Fixed-effects model

Consider now (3.2) as a fixed-effects model. The parameters in $\boldsymbol{\beta}$ for the covariates Z_{ik} and in $\boldsymbol{\mu}$ are some fixed quantities. These parameters are called the fixed-effect

coefficients in the model. Under the fixed-effects model, the covariates Z_{ik} that we use are also fixed quantities with integer values instead of random variables that takes value from some distribution. Under this model, all levels of each covariate are used. We will not randomly choose samples of levels for each covariate. All of the patients' responses have the same variance.

3.3 Global balance and marginal balance

3.3.1 Background

One of the main aims of randomization schemes is to balance the numbers of patients across treatment groups. When a patient arrives at a clinical centre, we have to assign this patient to one of the treatments immediately. Therefore, for a covariate-adaptive randomization scheme, one of its aims is to ensure a balance in the numbers of patients in the two treatment groups when patients are classified by their prognostic factors. When there are two or more covariates, either global or marginal balance can be sought under different covariate-adaptive randomization schemes. Global and marginal balancing are only used when the fixed-effects model is considered for patients' responses. There are in total p covariates to consider. Each of the covariates is classified into different levels. Let the total number of levels for the kth covariate be l_k for $k = 1, ..., p$. Let $L_k^{s_k}$ represents level s_k of the kth covariate for $k = 1, ..., p$ and $s_k = 1, ..., l_k$. For a particular covariate k, we will denote the levels of this covariate by $L_k^1, L_k^2, ..., L_k^{l_k}$. The values of l_k for $k = 1, ..., p$ can be the same or different.

3.3.2 Global balance

Global balance is where balance is achieved in the numbers of patients on the two

treatments for all combinations of the levels of the covariates. There are a total of $l_1 \times l_2 \times \ldots \times l_p$ possible combinations for all levels of the covariates. We will refer to each of the combinations as one of the $l_1 \times l_2 \times \ldots \times l_p$ cells for the covariates. When a patient arrives, the subject's prognostic profile will be recorded. We will know which cell the patient falls into. The assignment rule for a particular randomization scheme will be applied within that cell. We normally use the same randomization scheme for all cells. We may also use different randomization schemes in all or some of the cells. It will depend on how important it is to balance the numbers of patients between two treatments in a particular cell. Global balance therefore balances the numbers of patient in each of the cells $\{(L_1^1, L_2^1, ..., L_p^1), (L_1^2, L_2^1, ..., L_p^1), ..., (L_1^{l_1}, L_2^{l_2}, ..., L_p^{l_p})\}$.

We will let $D_m(L_1^{s_1}, L_2^{s_2}, ..., L_p^{s_p})$ be the difference between the number of patients on treatment 1 and the number of patients on treatment 2 up to the mth assignment based on all patients with the prognostic profile $\{(L_1^{s_1}, L_2^{s_2}, ..., L_p^{s_p})\}$. This $D_m(L_1^{s_1}, L_2^{s_2}, ..., L_p^{s_p})$ is used for the treatment assignment of the next patient in global balance for covariate-adaptive randomization schemes. Given the sign of this D_m, we know the under-represented treatment for this particular prognostic profile after m assignments have been made.

3.3.3 Marginal balance

Marginal balance is another way to balance the numbers of patients between two treatment groups. Here, we consider the difference in the numbers of patients on two treatments marginally for the levels of the covariates. When a patient arrives, its prognostic profile is noted and we know the levels of each of the p covariates for this patient. Let the prognostic profile of the next patient be $(L_1^{s_1}, L_2^{s_2}, ..., L_p^{s_p})$. Assume that m assignments have been made prior to this patient, let $D_m(L_1^{s_1})$ be the difference between the number of patients on treatment 1 and the number on treatment 2 based on all pa-

tients with level s_1 of the first covariate, and similarly define $D_m(L_2^{s_2}), ..., D_m(L_p^{s_p})$. These quantities measure the marginal imbalance at the individual levels of the p covariates for the next patient. Let $\bar{D}_m$ denote the overall imbalance used under covariate-adaptive randomization for the treatment assignment of the next patient based on the first m assignments in the marginal approach. This imbalance is a linear combination of the marginal imbalances at the individual levels of the p covariates and is defined as

$$\bar{D}_m = c_1 D_m(L_1^{s_1}) + c_2 D_m(L_2^{s_2}) + ... + c_p D_m(L_p^{s_p}),$$

where c_k for $k = 1, ..., p$ are any real numbers which represent the weights chosen to reflect the relative importance of the covariates. The overall imbalance $\bar{D}_m$ is calculated at each stage m and the sign of $\bar{D}_m$ will be noted. Hence, at each stage, the under-represented treatment will be known.

3.4 Assignment rules for covariate-adaptive randomization schemes

The treatment assignment rules are now described for covariate-adaptive simple random sampling, the covariate-adaptive biased coin design and the covariate-adaptive adjustable biased coin design which achieve either global or marginal balance when patients are classified by their prognostic factors.

Under covariate-adaptive simple random sampling for global balance, the next patient has a probability of $1/2$ of being allocated to either of the treatments in the appropriate cell for all values of $D_m(L_1^{s_1}, L_2^{s_2}, ..., L_p^{s_p})$. For marginal balance, the probability of allocating the next patient to either of the treatments is $1/2$ for all values of $\bar{D}_m$.

Next, the covariate-adaptive biased coin design is introduced. This is a design that applies Efron's (1971) biased coin design in the randomization process to patients grouped

by their prognostic profiles. Under the covariate-adaptive biased coin design, given the prognostic profile $(L_1^{s_1}, ..., L_p^{s_p})$ of the next patient, for global balance there is a fixed probability $p > 1/2$ of allocating the next patient to a treatment that has been chosen less often. Thus, we have

$$P(I_m = 1) = \begin{cases} p & \text{if } D_{m-1}(L_1^{s_1}, L_2^{s_2}, ..., L_p^{s_p}) > 0, \\ 1/2 & \text{if } D_{m-1}(L_1^{s_1}, L_2^{s_2}, ..., L_p^{s_p}) = 0, \\ 1-p & \text{if } D_{m-1}(L_1^{s_1}, L_2^{s_2}, ..., L_p^{s_p}) < 0 \end{cases}$$

as the probability of assigning the next patient to treatment 2. As $p \to 1$, the assignments tend to be more deterministic. For marginal balance, the allocation rule is similar but the difference $D_{m-1}(L_1^{s_1}, ..., L_p^{s_p})$ is replaced by $\bar{D}_{m-1}$ such that

$$P(I_m = 1) = \begin{cases} p & \text{if } \bar{D}_{m-1} > 0, \\ 1/2 & \text{if } \bar{D}_{m-1} = 0, \\ 1-p & \text{if } \bar{D}_{m-1} < 0. \end{cases}$$

Note that, for both global and marginal balance, when $p = 1/2$, the covariate-adaptive biased coin design reduces to covariate-adaptive simple random sampling.

The last covariate-adaptive randomization scheme to be introduced is the covariate-adaptive adjustable biased coin design. This design was developed by Baldi Antognini and Giovagnoli (2004). It is an extension of Efron's biased coin design in which the probability of assigning the next patient to a treatment depends on a decreasing and symmetric function $F(.)$ of the difference in the numbers of patients in the two treatment groups. For the covariate-adaptive adjustable biased coin design, given that the prognostic profile of the next patient is $(L_1^{s_1}, L_2^{s_2}, ..., L_p^{s_p})$, the probability of assigning this patient to treatment 1 is $F^a_{s_1,s_2,...,s_p}(.)$, where a is a design parameter. For global balance after m assignments, let $x = D_m(L_1^{s_1}, ..., L_p^{s_p})$. Then the probability or function

to assign the next patient to treatment 1 given the prognostic profile is $F^a_{s_1,\ldots,s_p}(x)$ and can be expressed as

$$F^a_{s_1,\ldots,s_p}(x) = \begin{cases} \frac{|x|^a}{|x|^a+1} & \text{if } x \leq -1, \\ 1/2 & \text{if } x = 0, \\ \frac{1}{|x|^a+1} & \text{if } x \geq 1, \end{cases} \tag{3.5}$$

where $a \geq 0$. As $a \to 0$, the design reduces to covariate-adaptive simple random sampling, whereas the design becomes more deterministic as $a \to \infty$. Similarly, for marginal balance under the covariate-adaptive adjustable biased coin design, the probability of assigning the next patient to treatment 1 given the same prognostic profile as above is $F^a_{s_1,\ldots,s_p}(x)$ in (3.5) with $x = \bar{D}_m$.

3.5 Analysis of covariance for fixed-effects model

After the treatment assignments have been made, the patients' responses in the two treatment groups will be studied under these covariate-adaptive randomization schemes. We are now interested in whether there is a genuine treatment difference between the two treatment groups.

Under the fixed-effects model, the covariates Z_{ik} used take integer values and all levels within any of the covariates are considered. In what follows, assume that the total number of levels of each of the covariates is the same, so that $l_1 = l_2 = \ldots = l_p$. We also assume that the values of all the levels of the covariates add up to zero. Finally, the patients' responses in the two treatment groups can be obtained from (3.2)

To test whether the population means are different, that is, there is a treatment effect, a two-sample t-test will be carried out. An analysis of covariance on the responses of the patients is considered. For each patient, the effect of the covariate will be removed from

the patients' responses to obtain the adjusted responses. The sample mean and variance of the adjusted responses will give estimated values for μ_j for $j = 1, 2$.

Let R_{ij} be the adjusted response of the ith patient allocated to treatment j for $i = 1, ..., n$ and $j = 1$ or 2. Let $\mathbf{R}$ represent the $n \times 1$ column vector of the adjusted responses R_{ij} given by

$$\mathbf{R} = (R_{11}, R_{21}, ..., R_{n_1 1}, R_{n_1+12}, ..., R_{n2})^T,$$

where

$$\mathbf{R} = \mathbf{Y} - \mathbf{Z}\boldsymbol{\beta}.$$

Each of the matrices $\mathbf{R}$, $\mathbf{Z}$ and $\mathbf{Y}$ can be split into two sub-matrices. We have $\mathbf{R} = (\mathbf{R_1}, \mathbf{R_2})^T$ as an $n \times 1$ column vector such that $\mathbf{R_1}$ is the $n_1 \times 1$ vector for the adjusted responses R_{i1} for the patients on treatment 1 and $\mathbf{R_2}$ is the $n_2 \times 1$ vector for the adjusted responses R_{i2} for the patients on treatment 2. Similarly, $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)^T$ is an $n \times 1$ column vector with $\mathbf{Y}_1$ as an $n_1 \times 1$ column vector for the responses Y_{i1} on treatment 1 and $\mathbf{Y}_2$ is an $n_2 \times 1$ column vector for the responses Y_{i2} on treatment 2. Finally, $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)^T$ is an $n \times q$ matrix such that $\mathbf{Z}_1$ represents the $n_1 \times q$ matrix with the values of the covariates and their interaction terms for those patients allocated to treatment 1 and $\mathbf{Z}_2$ is an $n_2 \times q$ matrix which contains the values of the covariates and their interaction terms for the patients allocated to treatment 2.

Let the sample means of the adjusted responses for treatments 1 and 2 be m_1 and m_2, respectively. Then we have

$$m_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} R_{i1}$$

and

$$m_2 = \frac{1}{n_2} \sum_{i=n_1+1}^{n} R_{i2}.$$

Consider a test of $H_0 : \mu_2 - \mu_1 = 0$ against $H_1 : \mu_2 - \mu_1 \neq 0$. The test statistic is

$$T_C = \frac{m_2 - m_1}{\sqrt{\widehat{\text{Var}(\hat{\mu}_2 - \hat{\mu}_1)}}}.$$

Given α as the significance level of the test, H_0 is rejected if $|T_C| > t_{n-2-q,\alpha/2}$. Under H_0, the test statistic T_C has a Student's t-distribution with $n-2-q$ degrees of freedom.

Next, we have to find an estimate for the variance of $\hat{\mu}_2 - \hat{\mu}_1$ for the test statistic. Using Hinkelmann and Kempthorne (2008) and (3.2), the estimation is as follows. Matrices $\mathbf{W}$ and $\mathbf{Z}$ are of full rank, 2 and q, respectively, so that $(\mathbf{W}^T\mathbf{W})^{-1}$ exists.

Model (3.2) reduces to $\mathbf{Y} = \mathbf{W}\boldsymbol{\mu}^* + \boldsymbol{\epsilon}^*$ if no covariates are included, where $\boldsymbol{\mu}^*$ represents the 2×1 column vector of population means and $\boldsymbol{\epsilon}^*$ is the $n \times 1$ column vector of error terms in this model. The least squares estimator of $\boldsymbol{\mu}^*$ in this model is $\widehat{\boldsymbol{\mu}}^* = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{Y}$. Now let P_W be the projection matrix or the orthogonal projection on $\mathbf{W}$ such that $P_W = \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T$. Also, let $R_W = \mathbf{I}_n - \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T = \mathbf{I}_n - P_W$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix, be the orthogonal projection on the orthogonal complement of $\mathbf{W}$. Then, $\mathbf{Y} = P_W\mathbf{Y} + R_W\mathbf{Y}$, where $P_W\mathbf{Y}$ and $R_W\mathbf{Y}$ are orthogonal, so that $(R_W\mathbf{Y})^T(P_W\mathbf{Y}) = 0$.

The least squares estimator of $\boldsymbol{\beta}$

$$\widehat{\boldsymbol{\beta}} = (\mathbf{Z}^T R_W \mathbf{Z})^{-1}\mathbf{Z}^T R_W \mathbf{Y}$$

and the estimator of $\boldsymbol{\mu}$ in (3.2) is

$$\widehat{\boldsymbol{\mu}} = \widehat{\boldsymbol{\mu}}^* - (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{Z}\widehat{\boldsymbol{\beta}}.$$

The error sum of squares is obtained as

$$\text{SSE} = \mathbf{Y}^T\mathbf{Y} - \widehat{\boldsymbol{\mu}}^{*T}\mathbf{W}^T\mathbf{Y} - \widehat{\boldsymbol{\beta}}^T\mathbf{Z}^T R_W \mathbf{Y}.$$

An unbiased estimator of the variance of the random error is then the mean square error given by

$$\hat{\sigma}^2_\epsilon = \text{SSE}/(n-2-q).$$

Since, $\widehat{\boldsymbol{\mu}}^*$ and $\widehat{\boldsymbol{\beta}}$ are uncorrelated, the variance of the estimator of $\boldsymbol{\mu}$ can be calculated as

$$\begin{aligned}\mathrm{Var}(\widehat{\boldsymbol{\mu}}) &= \mathrm{Var}(\widehat{\boldsymbol{\mu}}^*) + (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{Z}\mathrm{Var}(\widehat{\boldsymbol{\beta}})[(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{Z}]^T \\ &= [(\mathbf{W}^T\mathbf{W})^{-1} + (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{Z}(\mathbf{Z}^T R_W\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}]\sigma_\epsilon^2,\end{aligned}$$

which can be estimated by

$$\widehat{\mathrm{Var}(\widehat{\boldsymbol{\mu}})} = [(\mathbf{W}^T\mathbf{W})^{-1} + (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{Z}(\mathbf{Z}^T R_W\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}]\hat{\sigma}_\epsilon^2$$

An estimator of the variance of any linear combination of the $\hat{\mu}_j$ for $j = 1, 2$ can be calculated from the variance of $\widehat{\boldsymbol{\mu}}$. Let $\mathbf{a}$ be a 2×1 column vector such that it contains the coefficients of the linear combination of the $\hat{\mu}_j$, $j = 1, 2$. Then the linear combination of the μ_j can be written as $\mathbf{a}^T\widehat{\boldsymbol{\mu}}$. Therefore, an estimator of the variance of this linear combination of the μ_j is

$$\widehat{\mathrm{Var}(\mathbf{a}^T\widehat{\boldsymbol{\mu}})} = \hat{\sigma}_\epsilon^2\mathbf{a}^T[(\mathbf{W}^T\mathbf{W})^{-1} + (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{Z}(\mathbf{Z}^T R_W\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}]\mathbf{a}. \tag{3.6}$$

This can be used to obtain an estimator of the variance of the difference between the mean responses of the two treatment groups.

3.6 Normal approximation for power when the covariates are random variables

3.6.1 Background

Consider a clinical trial where patients are classified into different groups according to their prognostic profiles. The patients are allocated to one of the treatments by a covariate-adaptive randomization scheme. In Shao, Yu and Zhong (2010), the covariates

used in the randomization Z_i for $i = 1, ..., n$ are random variables. The responses of patients in the two treatment groups are compared by carrying out a two sample t-test to see whether there is a difference between the two treatment groups. The null hypothesis of the test is $H_0 : \mu_2 = \mu_1$ and the alternative is $H_1 : \mu_2 \neq \mu_1$. The test statistic is

$$T_s = \frac{\bar{Y}_2 - \bar{Y}_1}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}, \tag{3.7}$$

where $\bar{Y}_2$ and $\bar{Y}_1$ are the sample mean on treatments 2 and 1, respectively. With a significance level α, the null hypothesis is rejected if $|T_s| > c_\alpha$, where c_α is the critical value of a Student's t-distribution or the standard normal distribution.

Let $\mathcal{Y} = \{Y_{ij}, i = 1, .., n, j = 1 \text{ or } 2\}$ and $\mathcal{I} = \{I_i, i = 1, ..., n\}$, and let $\mathcal{Z} = \{D(Z_i), i = 1, .., n\}$. Further, let $\Delta_i = Z_i - E\{Z_i|D(Z_i)\}$ and $\Delta_i = 0$ if $Z_i = D(Z_i)$. Then, the difference in the sample means was calculated as

$$\bar{Y}_2 - \bar{Y}_1 = \mu_2 - \mu_1 + \frac{2}{n}\sum_{i=1}^{n}\{b(2I_i - 1)\Delta_i + I_i\epsilon_{i2} - (1 - I_i)\epsilon_{i1}\} + o_p(n^{-1/2})$$

conditionally on $\mathcal{Z} = \{D(Z_i), i = 1, .., n\}$, where $(\Delta_i, \epsilon_{i1}, \epsilon_{i2})$ are conditionally independent of $\mathcal{I}$ given $\mathcal{Z}$.

The asymptotic mean and variance of $\bar{Y}_2 - \bar{Y}_1$ were shown to be $\mu_2 - \mu_1$ and $4(b^2\sigma_\Delta^2 + \sigma_\epsilon^2)/n$, where $\sigma_\Delta^2 = \mathrm{Var}(\Delta_i)$. It was also shown that

$$\frac{(\bar{Y}_2 - \bar{Y}_1) - (\mu_2 - \mu_1)}{2\tau_\Delta/n^{1/2}} \to N(0, 1)$$

in distribution as $n \to \infty$, where $\tau_\Delta^2 = b^2\sigma_\Delta^2 + \sigma_\epsilon^2$, conditionally on $\mathcal{Z}$.

An estimate of the variance of $\bar{Y}_2 - \bar{Y}_1$ was shown to be

$$\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} = \frac{4\tau_z^2}{n} + o_p\left(\frac{1}{n}\right),$$

where $\tau_z^2 = b^2\sigma_z^2 + \sigma_\epsilon^2$ and $\sigma_z^2 = \mathrm{Var}(Z_i)$ conditionally on $\mathcal{Z}$. Under H_0, we have

$$\lim_{n\to\infty} P(|T_s| > c_\alpha) = 2\Phi(-c_\alpha\tau_z/\tau_\Delta).$$

Consider three randomization methods: simple randomization, biased coin randomization and covariate-adaptive biased coin randomization. Then, the unconditional power under H_1 for the two-sample t-test under simple randomization and the biased coin method is given by

$$P(|T_s| > c_\alpha) \approx \Phi\left(\frac{\bar{d}\sqrt{n}}{2\tau_z} - c_\alpha\right) + \Phi\left(-\frac{\bar{d}\sqrt{n}}{2\tau_z} - c_\alpha\right). \tag{3.8}$$

The power under covariate-adaptive biased coin randomization is given by

$$P(|T_s| > c_\alpha) \approx \Phi\left(\frac{\bar{d}\sqrt{n}}{2\tau_\Delta} - \frac{c_\alpha\tau_z}{\tau_\Delta}\right) + \Phi\left(-\frac{\bar{d}\sqrt{n}}{2\tau_\Delta} - \frac{c_\alpha\tau_z}{\tau_\Delta}\right). \tag{3.9}$$

3.6.2 One covariate

Consider a trial in which there is only one covariate. Assume that patients' responses follow a fixed-effects model of the form (3.2). Here, $\mathbf{Z}$ is an $n \times 1$ column vector containing the values for Z_i for $i = 1, ..., n$.

Two-sample t test

Consider a two sample t test and assume that the Z_i for all $i = 1, ..., n$ are discrete uniform random variables. We will study the power of this test under the covariate-adaptive biased coin design with null hypothesis $H_0 : \mu_2 - \mu_1 = 0$ and alternative hypothesis $H_1 : \mu_2 - \mu_1 > 0$. Since we assume that treatment 1 is the control treatment, we are interested in whether the new treatment which is treatment 2 is better than the standard or the control treatment. Therefore, the alternative hypothesis for this test is one-sided.

Simulation is carried out to study the numerical values for the power of the two sample t test under the covariate-adaptive biased coin design for different scenarios and to compare these with the numerical values obtained using a normal approximation.

When the Z_i are discrete uniformly distributed with possible values $1, ..., K$, each value has probability $1/K$. We will consider the values $K = 2, 4, 8$ such that they divide the real line into K categories with equal probabilities. The parameter b in the model is assumed to be 0.75 and the significance level is $\alpha = 0.05$. We have $n = 100$ and the random error term ϵ_{ij} is generated from a normal distribution with mean 0 and $\sigma_\epsilon = 0.5, 1, 2$. The probability p used in the covariate-adaptive biased coin design is $2/3$.

We first want to obtain the numerical values for the power using a normal approximation. For the above one-sided test, we have

$$P(T_s > c_\alpha) \approx \Phi\left(\frac{\bar{d}\sqrt{n}}{2\tau_\Delta} - \frac{c_\alpha \tau_z}{\tau_\Delta}\right) \tag{3.10}$$

as the normal approximation. When Z_i is discrete, Δ_i is by definition 0. It follows that $\mathrm{Var}(\Delta_i) = \sigma_\Delta^2 = 0$ and $\tau_\Delta^2 = \sigma_\epsilon^2$. The covariate Z_i is discrete uniformly distributed with variance $\sigma_z^2 = (K^2-1)/12$ for $K = 2, 4, 8$ and $\tau_z^2 = b^2(K^2-1)/12 + \sigma_\epsilon^2$. The following table gives numerical values for the power for various choices of $\bar{d} = \mu_2 - \mu_1$.

Table 3.1: Power by simulation and normal approximation when Z_i are discrete uniformly distributed

	$\bar{d}$	0	0.25	0.5	0.75	1	1.25	1.5
Simulation								
K	σ_ϵ							
2	0.5	0.023	0.661	0.998	1	1	1	1
	1	0.039	0.297	0.795	0.974	0.999	1	1
	2	0.043	0.144	0.336	0.568	0.787	0.921	0.979
4	0.5	0.004	0.269	0.924	1	1	1	1
	1	0.022	0.198	0.626	0.931	0.996	1	1
	2	0.038	0.126	0.296	0.537	0.760	0.906	0.975
8	0.5	0.003	0.048	0.317	0.773	0.974	0.999	1
	1	0.008	0.066	0.273	0.624	0.890	0.982	0.999
	2	0.024	0.083	0.196	0.384	0.614	0.796	0.922
Normal Approximation								
K	σ_ϵ							
2	0.5	0.020	0.672	0.998	1	1	1	1
	1	0.040	0.306	0.771	0.977	0.999	1	1
	2	0.047	0.147	0.336	0.580	0.796	0.927	0.981
4	0.5	0.001	0.238	0.963	1	1	1	1
	1	0.016	0.185	0.638	0.946	0.998	1	1
	2	0.037	0.123	0.297	0.536	0.763	0.910	0.975
8	0.5	0	0	0.187	0.947	1	1	1
	1	0.001	0.022	0.221	0.684	0.958	0.999	1
	2	0.015	0.061	0.179	0.385	0.630	0.831	0.943

The numerical values in the first column represent the significance level of the test for each scenario. Clearly, the values obtained by simulation and by the normal approximation are less than 0.05. In a two-sample t test, the significance level obtained is conservative under the covariate-adaptive biased coin design, as explained by Shao, Yu and Zhong (2010). As $\mathrm{Var}(\bar{Y}_2 - \bar{Y}_1)$ is less variable than $\mathrm{Var}(\bar{Y}_2) + \mathrm{Var}(\bar{Y}_1)$, there is a negative correlation between $\bar{Y}_1$ and $\bar{Y}_2$. The unbiased estimator of $\mathrm{Var}(\bar{Y}_2) + \mathrm{Var}(\bar{Y}_1)$, $s_1^2/n_1 + s_2^2/n_2$, is the denominator in the test statistic and gives a smaller value for the test statistic. Hence, the test is less likely to reject H_0 and produces a smaller value for the significance level. This explains why the numerical results obtained here are conservative.

For each K, the significance level increases when σ_ϵ increases. When K is larger, the power for the same values for $\bar{d}$ and σ_ϵ decreases. The power tends to 1 more quickly when σ_ϵ is smaller. For large values of $\bar{d}$, it is clear that the power obtained from the normal approximation is larger than the simulated value.

Analysis of covariance t test

An analysis of covariance t test can also be carried out on the treatment difference. This is a two-sample t test when the effect of the covariates is removed from the patients' responses. In Shao, Yu and Zhong (2010), the normal approximation to the power by analysis of covariance is given by

$$P(|T_C| > c_\alpha) \approx \Phi\left(\frac{\bar{d}\sqrt{n}}{2\sigma_\epsilon} - c_\alpha\right) + \Phi\left(-\frac{\bar{d}\sqrt{n}}{2\sigma_\epsilon} - c_\alpha\right),$$

where T_C is the test statistic for the analysis of covariance t test. Here, we consider a one-sided t test with null hypothesis $H_0 : \mu_2 - \mu_1 = 0$ and alternative hypothesis $H_1 : \mu_2 - \mu_1 > 0$. The normal approximation to the power is

$$P(T_C > c_\alpha) \approx \Phi\left(\frac{\bar{d}\sqrt{n}}{2\sigma_\epsilon} - c_\alpha\right).$$

Therefore, the power obtained from the normal approximation is not affected by the distribution of Z_i.

Now consider the general case where we have p covariates. Let Z_{ik} have mean 0 and variance $\sigma^2_{z_k}$ for $k = 1, ..., p$. Then using the same notation as in Section 3.5, $\boldsymbol{\mu}^*$ is the vector of population means in the model without covariates and $\widehat{\boldsymbol{\mu}}^* = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{Y}$. The variance of $\widehat{\boldsymbol{\mu}}^*$ then becomes

$$\begin{aligned} \mathrm{Var}(\widehat{\boldsymbol{\mu}}^*) &= (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathrm{Var}(\mathbf{Y})\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1} \\ &= (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\{\boldsymbol{\beta}^T\mathrm{Var}(\mathbf{Z})\boldsymbol{\beta}\mathbf{I}_n + \mathrm{Var}(\boldsymbol{\epsilon})\}\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}, \end{aligned}$$

where $\mathrm{Var}(\mathbf{Y})$ is the $n \times n$ covariance matrix for the column vector of patients' responses and $\mathbf{I}_n$ is the $n \times n$ identity matrix. As patients' responses are independent, the covariance of Y_{ij} and $Y_{i'j}$ for $i, i' = 1, ..., n$ and $i \neq i'$ is zero. The covariance matrix of $\mathbf{Y}$ will have the variance of Y_{ij} for $i = 1, ..., n$ as the ith diagonal element and zeros elsewhere.

Similarly, $\mathrm{Var}(\mathbf{Z})$ is the covariance matrix of $\mathbf{Z}$. As the Z_{ik} are independent and identically distributed for each $k = 1, ..., p$ with mean 0 and variance $\sigma^2_{z_k}$, $\mathrm{Var}(\mathbf{Z})$ is a $p \times p$ covariance matrix with $\sigma^2_{z_k}$ as the kth diagonal element and zeros elsewhere. In fact, for each patient $i = 1, ..., n$, $\boldsymbol{\beta}^T\mathrm{Var}(\mathbf{Z})\boldsymbol{\beta}$ is a scalar of the form $\beta_1^2\sigma^2_{z_1} + ... + \beta_p^2\sigma^2_{z_p}$. Finally, as ϵ_{ij} for $i = 1, ..., n$ are independent and identically distributed random errors with mean 0 and variance σ^2_ϵ, $\mathrm{Var}(\boldsymbol{\epsilon})$, the covariance matrix of $\boldsymbol{\epsilon}$ is an $n \times n$ matrix with diagonal entries σ^2_ϵ and zeros elsewhere.

So

$$\mathrm{Var}(\mathbf{Y}) = (\boldsymbol{\beta}^T\mathrm{Var}(\mathbf{Z})\boldsymbol{\beta} + \sigma^2_\epsilon)I_n.$$

Therefore,

$$\mathrm{Var}(\widehat{\boldsymbol{\mu}}^*) = (\boldsymbol{\beta}^T\mathrm{Var}(\mathbf{Z})\boldsymbol{\beta} + \sigma^2_\epsilon)(\mathbf{W}^T\mathbf{W})^{-1}.$$

Now, consider the variance of $\mathbf{Z}\widehat{\boldsymbol{\beta}}$. Then we have

$$\begin{aligned}\mathrm{Var}(\mathbf{Z}\widehat{\boldsymbol{\beta}}) &= E[(\mathbf{Z}\widehat{\boldsymbol{\beta}})(\mathbf{Z}\widehat{\boldsymbol{\beta}})^T] - E(\mathbf{Z}\widehat{\boldsymbol{\beta}})E(\mathbf{Z}\widehat{\boldsymbol{\beta}})^T \\ &= E[\mathbf{Z}\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}^T\mathbf{Z}^T] - E(\mathbf{Z}\widehat{\boldsymbol{\beta}})E(\mathbf{Z}\widehat{\boldsymbol{\beta}})^T.\end{aligned}$$

As $\mathbf{Z}$ and $\widehat{\boldsymbol{\beta}}$ are independent, as shown in the Appendix, we can now write

$$\begin{aligned}\mathrm{Var}(\mathbf{Z}\widehat{\boldsymbol{\beta}}) &= E(\mathbf{Z})E(\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}^T)E(\mathbf{Z}^T) - E(\mathbf{Z})E(\widehat{\boldsymbol{\beta}})E(\widehat{\boldsymbol{\beta}})^T E(\mathbf{Z})^T \\ &= E(\mathbf{Z})\{E(\widehat{\boldsymbol{\beta}}^T\widehat{\boldsymbol{\beta}}) - E(\widehat{\boldsymbol{\beta}})^T E(\widehat{\boldsymbol{\beta}})\}E(\mathbf{Z})^T \\ &= E(\mathbf{Z})\mathrm{Var}(\widehat{\boldsymbol{\beta}})E(\mathbf{Z})^T.\end{aligned}$$

We are interested in obtaining an estimate of the variance of $\widehat{\boldsymbol{\mu}}$. Since $\widehat{\boldsymbol{\mu}}^*$ and $\widehat{\boldsymbol{\beta}}$ are uncorrelated, the variance of $\widehat{\boldsymbol{\mu}}$ can be written as

$$\begin{aligned}\mathrm{Var}(\widehat{\boldsymbol{\mu}}) &= \mathrm{Var}(\widehat{\boldsymbol{\mu}}^*) + (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathrm{Var}(\mathbf{Z}\widehat{\boldsymbol{\beta}})\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1} \\ &= (\boldsymbol{\beta}^T\mathrm{Var}(\mathbf{Z})\boldsymbol{\beta} + \sigma_\epsilon^2)(\mathbf{W}^T\mathbf{W})^{-1} \\ &+ (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}\{E(\mathbf{Z})\mathrm{Var}(\widehat{\boldsymbol{\beta}})E(\mathbf{Z})^T\}\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}.\end{aligned}$$

The estimate of the variance of $\widehat{\boldsymbol{\mu}}$ becomes

$$\begin{aligned}\widehat{\mathrm{Var}(\widehat{\boldsymbol{\mu}})} &= (\widehat{\boldsymbol{\beta}}^T\widehat{\mathrm{Var}(\mathbf{Z})}\widehat{\boldsymbol{\beta}} + \hat{\sigma}_\epsilon^2)(\mathbf{W}^T\mathbf{W})^{-1} \\ &+ (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}\{E(\mathbf{Z})\widehat{\mathrm{Var}(\widehat{\boldsymbol{\beta}})}E(\mathbf{Z})^T\}\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1},\end{aligned}$$

where $\hat{\sigma_\epsilon}^2$ is defined earlier as the error mean square, $\widehat{\mathrm{Var}(\mathbf{Z})}$ is the estimator of the covariance matrix of $\mathbf{Z}$ and $\widehat{\mathrm{Var}(\widehat{\boldsymbol{\beta}})}$ is the estimator of the covariance matrix of the least square estimator $\boldsymbol{\beta}$. Let $\mathbf{a}$ be the column vector of contrasts of the population means. Then we have

$$\begin{aligned}\widehat{\mathrm{Var}(\mathbf{a}^T\widehat{\boldsymbol{\mu}})} &= \mathbf{a}^T[(\widehat{\boldsymbol{\beta}}^T\widehat{\mathrm{Var}(\mathbf{Z})}\widehat{\boldsymbol{\beta}} + \hat{\sigma}_\epsilon^2)(\mathbf{W}^T\mathbf{W})^{-1} \\ &+ (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}\{E(\mathbf{Z})\widehat{\mathrm{Var}(\widehat{\boldsymbol{\beta}})}E(\mathbf{Z})^T\}\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}]\mathbf{a}.\end{aligned}$$

In the simulations, we assume there is only one covariate. The matrix $\mathbf{Z}$ in (3.2) is just an $n \times 1$ column vector and all of the Z_i for $i = 1, ..., n$ take values from one distribution. Here, $E(\mathbf{Z}) = 0$ and $\mathrm{Var}(\mathbf{Z}) = \sigma_z^2$. So $\mathrm{Var}(\mathbf{Z}\widehat{\boldsymbol{\beta}}) = 0$. In addition, the vector $\boldsymbol{\beta}$ will only contain a single scalar and is denoted by b. Then we have

$$\mathrm{Var}(\widehat{\boldsymbol{\mu}}) = \mathrm{Var}(\widehat{\boldsymbol{\mu}^*}) = (b^2\sigma_z^2 + \sigma_\epsilon^2)(\mathbf{W}^T\mathbf{W})^{-1}.$$

The least squares estimator of b will also become a scalar and is denoted as $\hat{\beta}$. We have

$$\widehat{\mathrm{Var}(\widehat{\boldsymbol{\mu}})} = (b^2\hat{\sigma}_z^2 + \hat{\sigma}_\epsilon^2)(\mathbf{W}^T\mathbf{W})^{-1},$$

where $\hat{\sigma}_z^2$ is the estimator of the variance of $\mathbf{Z}$ and the sample variance of $\mathbf{Z}$ is used. Let $\mathbf{a}^T = (0, 1)$, as we want to study whether treatment 2 is better than treatment 1. We have to estimate the variance of $\hat{\mu}_2 - \hat{\mu}_1$, which can be expressed as

$$\hat{\sigma}_\epsilon^2 \, (0, 1) \, (b^2\hat{\sigma}_z^2 + \hat{\sigma}_\epsilon^2)(\mathbf{W}^T\mathbf{W})^{-1} \, (0, 1)^T .$$

The numerical values for the power obtained from analysis of covariance using simulation and the normal approximation are given below. The values for α, β, n, K and σ_ϵ are the same as for the two-sample t test.

Table 3.2: Power by simulation and normal approximation for analysis of covariance t test

	$\bar{d}$	0	0.25	0.5	0.75	1	1.25	1.5
Simulation								
K	σ_ϵ							
2	0.5	0.048	0.795	1	1	1	1	1
	1	0.054	0.345	0.801	0.981	0.999	1	1
	2	0.057	0.153	0.341	0.594	0.795	0.926	0.982
4	0.5	0.044	0.802	1	1	1	1	1
	1	0.050	0.341	0.800	0.983	0.999	1	1
	2	0.049	0.147	0.345	0.586	0.799	0.927	0.980
8	0.5	0.051	0.800	1	1	1	1	1
	1	0.054	0.346	0.798	0.980	0.999	1	1
	2	0.049	0.166	0.335	0.589	0.799	0.923	0.980
Normal Approximation								
All K	σ_ϵ							
	0.5	0.05	0.804	1	1	1	1	1
	1	0.05	0.347	0.804	0.982	1	1	1
	2	0.05	0.154	0.347	0.591	0.804	0.931	0.982

The numerical values for the power obtained using a normal approximation are usually larger than the simulated values. However, the significance level of the test is close to 0.05 for each scenario.

3.6.3 More than one covariate

The theoretical calculations for the normal approximations to the power in Shao, Yu, Zhong (2010) were considered when a single univariate covariate is involved in the randomization process. However, in real trials, more than one prognostic factor or a combination of several prognostic factors will affect the responses of patients to different treatments. We now consider the case where more than one univariate covariate is used in the randomization process.

Let p be the number of covariates used. Here, the covariates Z_{ik} for $i = 1, ..., n$ and $k = 1, ..., p$ are assumed to be random variables. In Shao, Yu and Zhong (2010), a first-order linear model for the patients' responses Y_{ij} was constructed for a univariate covariate. We extend the analysis to a second-order linear model for patients' responses where there is more than one univariate covariate.

The model for patients' responses is written as

$$Y_{ij} = \mu_j + \sum_{k=1}^{p} b_k Z_{ik} + \sum_{k=1}^{p} c_k Z_{ik}^2 + \sum_{\substack{k,m=1,\\ k<m}}^{p} d_{km} Z_{ik} Z_{im} + \epsilon_{ij},$$

where μ_j is the mean response on treatment j for $j = 1, 2$ and Z_{ik} is the kth covariate for the ith patient. There are p first-order terms. Next, we have two types of second-order terms in the model. The first type is the squared term in the covariate Z_{ik} and there are p of them. The second type is the interaction term between two different covariates and there are $p \times (p-1)$ of these. Finally, ϵ_{ij} is the random error term.

In (3.2), Y_{ij} and ϵ_{ij} are the components of the column vectors $\mathbf{Y}$ and $\boldsymbol{\epsilon}$, respectively. The column vector $\boldsymbol{\mu}$ will be the same as in (3.3) and the matrix $\mathbf{W}$ will be of the same

form as in (3.4). The matrix $\mathbf{Z}$ can be written as

$$\mathbf{Z} = \begin{pmatrix} Z_{11} & \dots & Z_{n_1 1} & Z_{(n_1+1)1} & \dots & Z_{n1} \\ \vdots & & \vdots & \vdots & & \vdots \\ Z_{1p} & \dots & Z_{n_1 p} & Z_{(n_1+1)p} & \dots & Z_{np} \\ Z_{11}^2 & \dots & Z_{n_1 1}^2 & Z_{(n_1+1)1}^2 & \dots & Z_{n1}^2 \\ \vdots & & \vdots & \vdots & & \vdots \\ Z_{1p}^2 & \dots & Z_{n_1 p}^2 & Z_{(n_1+1)p}^2 & \dots & Z_{np}^2 \\ Z_{11}Z_{12} & \dots & Z_{n_1 1}Z_{n_1 2} & Z_{(n_1+1)1}Z_{(n_1+1)2} & \dots & Z_{n1}Z_{n2} \\ Z_{11}Z_{13} & \dots & Z_{n_1 1}Z_{n_1 3} & Z_{(n_1+1)1}Z_{(n_1+1)3} & \dots & Z_{n1}Z_{n3} \\ \vdots & & \vdots & \vdots & & \vdots \\ Z_{1(p-1)}Z_{1p} & \dots & Z_{n_1(p-1)}Z_{n_1 p} & Z_{(n_1+1)(p-1)}Z_{(n_1+1)p} & \dots & Z_{n(p-1)}Z_{np} \end{pmatrix}^T.$$

Next, we have

$$\boldsymbol{\beta} = \left(b_1, b_2, ..., b_p, c_1, ..., c_p, d_{11}, d_{12}, ..., d_{(p-1)p}\right)^T.$$

Now let

$$\Delta_{ik} = Z_{ik} - E\{Z_{ik}|D(Z_{ik})\},$$

$$\Delta_{ik}^2 = Z_{ik}^2 - E\{Z_{ik}^2|D(Z_{ik})\}$$

and

$$\Delta_{ik}\Delta_{im} = Z_{ik}Z_{im} - E\{Z_{ik}Z_{im}|D(Z_{ik})\},$$

where $D(Z_{ik})$ is the discrete function of Z_{ik} if Z_{ik} is continuous. Then the covariate-adaptive biased coin design applies Efron's biased coin assignment rule within each category defined by $D(Z_{ik})$. In this case, $\mathcal{Z}$ is defined as $\mathcal{Z} = \{D(Z_{ik}), i = 1, ..., n, k = 1, ..., p\}$, so that when $Z_{ik} = D(Z_{ik})$, $\Delta_{ik} = 0$.

The same test will be carried out of $H_0 : \mu_2 - \mu_1 = 0$ against $H_1 : \mu_2 - \mu_1 > 0$. Here, we want to test if treatment 2 is better than treatment 1. We have the same test

statistic T_s as in (3.7) and H_0 will be rejected if $T_s > c_\alpha$, for c_α defined as earlier. The sample means of the responses on the two treatments $\bar{Y}_1$ and $\bar{Y}_2$ are needed for the test statistic. We may write

$$\begin{aligned}
\bar{Y}_2 - \bar{Y}_1 &= \mu_2 - \mu_1 + \frac{2}{n}\sum_{i=1}^{n}\sum_{k=1}^{p}(2I_i - 1)b_k\Delta_{ik} \\
&+ \frac{2}{n}\sum_{i=1}^{n}\sum_{k=1}^{p}(2I_i - 1)b_k E\{Z_{ik}|D(Z_{ik})\} \\
&+ \frac{2}{n}\sum_{i=1}^{n}\sum_{k=1}^{p}(2I_i - 1)c_k\Delta_{ik}^2 + \frac{2}{n}\sum_{i=1}^{n}\sum_{k=1}^{p}(2I_i - 1)c_k E\{Z_{ik}^2|D(Z_{ik})\} \\
&+ \frac{2}{n}\sum_{i=1}^{n}\sum_{\substack{m=1,\\ k<m}}^{p}\sum_{k=1}^{p}(2I_i - 1)d_{km}\Delta_{ik}\Delta_{im} \\
&+ \frac{2}{n}\sum_{i=1}^{n}\sum_{\substack{m=1,\\ k<m}}^{p}\sum_{k=1}^{p}(2I_i - 1)d_{km}E\{Z_{ik}Z_{im}|D(Z_{ik})\} \\
&+ \frac{2}{n}\sum_{i=1}^{n}\{I_i\epsilon_{i1} - (1 - I_i)\epsilon_{i0}\} + o_p(n^{-1/2}),
\end{aligned}$$

The asymptotic mean of $\bar{Y}_2 - \bar{Y}_1$ is calculated by taking the expectation of $\bar{Y}_2 - \bar{Y}_1$ conditionally on $\mathcal{Z}$ and is equal to $\mu_2 - \mu_1$.

The asymptotic variance of $\bar{Y}_2 - \bar{Y}_1$ can be written as

$$
\begin{aligned}
&\frac{4}{n}\Bigg(\sum_{k=1}^{p} b_k^2 E(\Delta_k^2) + \sum_{k=1}^{p} c_k^2 E(\Delta_k^4) + \sigma_\epsilon^2 + \sum_{\substack{k,m=1,\\ k<m}}^{p} d_{km}^2 E(\Delta_k^2 \Delta_m^2) \\
+ \;& 2\sum_{k=1}^{p} b_k c_k E(\Delta_k^3) + \sum_{\substack{k,m,s=1,\\ k<m<s}}^{p} d_{km} d_{ms} E(\Delta_k \Delta_m^2 \Delta_s) \\
+ \;& \sum_{\substack{k,m,s=1,\\ k<m<s}}^{p} d_{km} d_{ks} E(\Delta_k^2 \Delta_m \Delta_s) + 2\sum_{k=1}^{p} b_k c_k E(\Delta_k^2 Z_k) \\
+ \;& 2\sum_{\substack{k,m=1,\\ k<m}}^{p} b_k d_{km} E(\Delta_k^2 \Delta_m) + 2\sum_{k=1}^{p} c_k^2 E(\Delta_k^2 Z_k^2) + 2\sum_{\substack{k,m=1,\\ k<m}}^{p} c_k d_{km} E(\Delta_k^3 \Delta_m) \\
+ \;& 2\sum_{\substack{k,m=1,\\ k<m}}^{p} c_k d_{km} E(\Delta_k \Delta_m Z_k^2) + \sum_{\substack{k,l=1,\\ k\neq l}}^{p} b_k b_l E(\Delta_k \Delta_l) + \sum_{\substack{k,l=1,\\ k\neq l}}^{p} c_k c_l E(\Delta_k^2 \Delta_l^2) \\
+ \;& \sum_{\substack{k,m,s,r=1,\\ k<m<s<r}}^{p} d_{km} d_{sr} E(\Delta_k \Delta_m \Delta_s \Delta_r) + 2\sum_{\substack{k,l=1,\\ k<l}}^{p} b_k c_l E(\Delta_k \Delta_l^2) \\
+ \;& \sum_{\substack{k,m,s=1,\\ k<m<s}}^{p} d_{ks} d_{ms} E(\Delta_k \Delta_m \Delta_s^2) + 2\sum_{\substack{k,m,s=1,\\ k<m<s}}^{p} b_k d_{ms} E(\Delta_k \Delta_m \Delta_s) \\
+ \;& 2\sum_{\substack{k,l=1,\\ k\neq l}}^{p} c_k c_l E(\Delta_k^2 Z_l^2) + 2\sum_{\substack{k,m,s=1,\\ k<m<s}}^{p} c_k d_{ms} E(\Delta_k^2 \Delta_m \Delta_s) \\
+ \;& 2\sum_{\substack{k,m,s=1,\\ k<m<s}}^{p} c_k d_{ms} E(\Delta_m \Delta_s Z_k^2) + 2\sum_{\substack{k,l=1,\\ k<l}}^{p} b_k c_l E(\Delta_l^2 Z_k). \\
+ \;& 2\sum_{\substack{k,m=1,\\ k<m<}}^{p} b_k d_{km} E(\Delta_k \Delta_m Z_k) + 2\sum_{\substack{k,m,s=1,\\ k<m<s}}^{p} b_k d_{ms} E(\Delta_m \Delta_s Z_k)\Bigg)
\end{aligned}
$$

This is the form $4\tau_\Delta^{*2}/n$, where τ_Δ^{*2} is known from above. We also have

$$\frac{(\bar{Y}_2 - \bar{Y}_1) - (\mu_2 - \mu_1)}{2\tau_\Delta^*/n^{1/2}} \to N(0,1)$$

in distribution as $n \to \infty$. Clearly, τ_Δ^{*2} has a more complicated form than τ_Δ^2 for one covariate.

Similarly, we have

$$\begin{aligned}
\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2} &= \frac{4}{n}\Bigg(\sum_{k=1}^{p} b_k^2 E(Z_k^2)+\sum_{\substack{k,l=1,\\ k\neq l}}^{p} b_k b_l E(Z_k Z_l)+2\sum_{k=1}^{p} b_k c_k E(Z_k^3) \\
&+ \sum_{\substack{k,l=1,\\ k\neq l}}^{p} b_k c_l E(Z_k Z_l^2)+2\sum_{\substack{k,m=1,\\ k<m}}^{p} b_k d_{km} E(Z_k^2 Z_m) \\
&+ 2\sum_{\substack{k,m,s=1,\\ k<m<s}}^{p} b_k d_{ms} E(Z_k Z_m Z_s)+\sum_{\substack{k,l=1,\\ k\neq l}}^{p} b_l c_k E(Z_k^2 Z_l) \\
&+ \sum_{k=1}^{p} c_k^2 E(Z_k^4)+\sum_{\substack{k,l=1,\\ k\neq l}}^{p} c_k c_l E(Z_k^2 Z_l^2)+2\sum_{\substack{k,m=1,\\ k<m}}^{p} c_k d_{km} E(Z_k^3 Z_m) \\
&+ \sum_{\substack{k,m=1,\\ k<m}}^{p} d_{km}^2 E(Z_k^2 Z_m^2)+\sum_{\substack{k,m,s=1,\\ k<m<s}}^{p} c_k d_{ms} E(Z_k Z_m Z_s) \\
&+ \sum_{\substack{k,m,s=1,\\ k<m<s}}^{p} c_k d_{ms} E(Z_k^2 Z_m Z_s)+\sum_{\substack{k,m,s=1,\\ k<m<s}}^{p} d_{km} d_{ms} E(Z_k Z_m^2 Z_s) \\
&+ \sum_{\substack{k,m,s=1,\\ k<m<s}}^{p} d_{ks} d_{ms} E(Z_k Z_m Z_s^2)+\sum_{\substack{k,m,s=1,\\ k<m<s}}^{p} d_{km} d_{ks} E(Z_k^2 Z_m Z_s) \\
&+ \sum_{\substack{k,m,s,r,\\ k<m<s<r}}^{p} d_{km} d_{sr} E(Z_k Z_m Z_s Z_r)+\sigma_\epsilon^2\Bigg)+o_p\left(\frac{1}{n}\right).
\end{aligned}$$

This can be written in the form

$$\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}=\frac{4\tau_z^{*2}}{n}+o_p\left(\frac{1}{n}\right).$$

Again, as compared with τ_z^2, τ_z^{*2} is more complicated.

The normal approximation to the power when there are p covariates can be obtained as in (3.8) for simple random sampling and the biased coin design and as in (3.9) for the covariate-adaptive biased coin design by replacing τ_z and τ_Δ with τ_z^* and τ_Δ^*, respectively.

3.7 Simulation for global and marginal balance

3.7.1 Algorithms for global balance

In this section, we look at different covariate-adaptive randomization schemes when global and marginal balancing are sought in the numbers of patients on the two treatments by simulation.

The model for patients' responses to be considered here is the fixed-effects model. Assume that the covariates Z_{ik} are fixed and take integer values. For the kth covariate for $k = 1, .., p$, the values for all levels of this covariate will add up to zero. We assume that the number of levels for each covariate is the same.

Assume that two covariates or prognostic factors are considered to be of particular importance in affecting patients' responses to the treatments. Let Z_{i1} and Z_{i2} denote the first and second covariates. The numbers of levels for these are denoted by l_1 and l_2, respectively, where $l_1 = l_2$.

We consider two scenarios where both of the covariates either have two or three levels. For example, for three-level covariates, that is, $l_1 = l_2 = 3$, we denote the levels by a_1, a_2 and a_3 for the first covariate and b_1, b_2 and b_3 for the second covariate. Global balancing balances the numbers of patients at each of the combinations of the levels of the covariates. There will be in total nine possible combinations of the levels of the two covariates. Patients will be classified into nine cells or strata according to their prognostic profiles. The nine combinations for the two covariates are

$$\{(a_1, b_1), (a_1, b_2), (a_1, b_3), (a_2, b_1), (a_2, b_2), (a_2, b_3), (a_3, b_1), (a_3, b_2), (a_3, b_3)\}.$$

In other words, global balancing balances the numbers of patients in the two treatment groups in each of the nine cells.

In the simulations, we generate two random numbers from a uniform distribution

between 0 and 1 for each patient. The first number will represent the level of the first covariate and the second number the level of the second covariate for this patient. Suppose that we consider covariates with two levels. Then, if the first random number is greater than or equal to 0 and less than or equal to 0.5, then this patient belongs to level a_1 of the first covariate. If the first random number is greater than 0.5 and less than or equal to 1, then this patient belongs to level a_2 of the first covariate. For a three-level covariate, if the first random number has value greater than or equal to 0 and less than or equal to $1/3$, then this patient belongs to level a_1 of the first covariate. If the first random number has value greater than $1/3$ and less than or equal to $2/3$, then this patient belongs to level a_2 of the first covariate. Finally, if the first random number is greater than $2/3$ and less than or equal to 1, this patient belongs to level a_3 of the first covariate. Similarly, this applies to the second random number for the second covariate at two or three levels. By making use of the random numbers, each patient is equally likely to fall into one of two or three levels for both covariates. Therefore, we can obtain a prognostic profile for each patient by simulation.

We consider three covariate-adaptive randomization schemes, which are covariate-adaptive simple random sampling, covariate-adaptive biased coin randomization and covariate-adaptive adjustable biased coin randomization. The covariate-adaptive randomization schemes are applied to each of the nine cells. At each stage, when a patient arrives, the patient is classified into one of the nine cells according to the patient's prognostic characteristics. Within this cell, the current imbalance in the numbers of patients on the two treatments will be noted and this patient will be allocated to the under-represented treatment with a probability p. Under covariate simple random sampling p is fixed and equals $1/2$. For the covariate-adaptive biased coin design, p is also fixed and takes value greater than a half to allocate the patient to the treatment that has been

chosen less often. Here, we only consider the same p for all cells under the covariate-adaptive biased coin design. Under the covariate-adaptive adjustable biased coin design, this probability p varies with the current imbalance and equals $F^a(x)$ defined in Chapter 2 with some chosen design parameter a. Thus, we assume that a is the same for all cells and for all stages. At the same stage, $F^a(x)$ may be different in different cells as x the difference in the numbers of patients on treatments 1 and 2 may be different in different cells.

In the simulations for global balance, given the prognostic profile of a patient, we know which of the cells the patient belongs to. A third random number is generated for this patient. This random number is generated from a uniform distribution between 0 and 1 and will be used for treatment assignment. For this particular cell, we have to identify the current numbers of patients on the two treatments. Under covariate-adaptive simple random sampling in global balance, if the third random number has value greater than or equal to zero and less than 0.5, this patient will be allocated to treatment 1. If the third random number has value greater than or equal to 0.5 and less than or equal to 1, then this patient will be allocated to treatment 2.

Under the covariate-adaptive biased coin design for global balance, we take different values for p in the simulations. Given the prognostic profile of the next patient, we will just look at the current numbers of patients on the two treatments in this particular cell. If the third random number has value greater than or equal to 0 and less than p, this patient will be allocated to the treatment that has fewer patients. If the third random number has value greater than or equal to p and less than or equal to 1, then this patient will be allocated to the treatment that has more patients. If the numbers of patients on the two treatments are the same, then the covariate-adaptive biased coin assignment process will reduce to the covariate-adaptive simple random sampling assignment process with

$p = 1/2$.

Under the covariate-adaptive adjustable biased coin design for global balance, given the prognostic profile of the next patient, we have to calculate the difference in the numbers of patients on treatments 1 and 2 in this particular cell. This difference is denoted by x in the function $F^a_{s_1,s_2}(x)$ for the probability of allocating this patient to treatment 1 given the prognostic profile as (s_1, s_2). If the third random number for this patient has value greater than or equal to zero and less than $F^a_{s_1,s_2}(x)$, this patient will be allocated to treatment 1. If the third random number has value greater than or equal to $F^a_{s_1,s_2}(x)$ and less than or equal to 1, this patient will be allocated to treatment 2.

3.7.2 Algorithms for marginal balance

Marginal balancing balances the numbers of patients on the two treatments marginally for each level of the covariates. For example, suppose that the two covariates have three levels. Denote the three levels of the two covariates by a_1, a_2 and a_3 and b_1, b_2 and b_3. We aim to balance the numbers of patients on the two treatments at each level of the covariates. In other words, we balance across the levels $\{a_1, a_2, a_3\}$ of the first covariate and balance across the levels $\{b_1, b_2, b_3\}$ of the second covariate. The patient's prognostic profile is recorded upon arrival at a clinical centre and classifies the patient into one of the three levels of both covariates. Suppose that we have made m assignments so far and the prognostic profile of this next patient is level a_1 for the first covariate and level b_1 for the second covariate. First, we have to calculate the current imbalance in the numbers of patients on treatments 1 and 2 at level a_1 for the first covariate for all levels of the second covariate, that is, $D_m(a_1)$. We then calculate the imbalance at level b_1 for the second covariate for all levels of the first covariate, $D_m(b_1)$.

From Section 3.3, the overall imbalance after m assignments is

$$\bar{D}_m = c_1 D_m(a_1) + c_2 D_m(b_1),$$

where c_1 and c_2 are integers representing the weights given to $D_m(a_1)$ and $D_m(b_1)$, respectively. This overall imbalance is used for treatment assignments in marginal balance. More specifically, we either consider the overall imbalance as $\bar{D}_m = D_m(a_i)$, for $i = 1, .., 3$, so that $c_1 = 1$ and $c_2 = 0$, or $\bar{D}_m = D_m(b_i)$ for $i = 1, .., 3$, so that $c_1 = 0$ and $c_2 = 1$. In other words, we only consider the levels of one of the covariates.

In the simulations for marginal balance, the prognostic profile of a patient is created in exactly the same way as in the global balancing case. By using the two random numbers generated from a uniform distribution between 0 and 1, we can identify the levels of the two covariates for each patient. For the $(m + 1)$th patient, we can then calculate the overall imbalance $\bar{D}_m$. Similar to global balancing, based on the sign of the overall imbalance $\bar{D}_m = D_m(a_1)$ or $\bar{D}_m = D_m(b_1)$, we will know the current under-represented treatment. A third random number is generated for each patient from a uniform distribution between 0 and 1 for treatment assignment.

Under covariate-adaptive simple random sampling for marginal balance, the sign of $\bar{D}_m$ is ignored for treatment assignment. The assignment is made to treatment 1 if the third random number for this patient has value greater than or equal to 0 and less than 0.5. Similarly, assignment is made to treatment 2 if the third random number has value greater than or equal to 0.5 and less than or equal to 1.

Under the covariate-adaptive biased coin design, we will have a probability $p > 1/2$ of allocating the next patient to the treatment that has been chosen less often. As we know the sign of $\bar{D}_m$, we know which is the under-represented treatment. If the third random number for this patient has value greater than or equal to zero and less than p, we will allocate this patient to the treatment that has fewer patients based on $\bar{D}_m$. If the

third random number has values greater than or equal to p and less than or equal to 1, then this patient will be allocated to the treatment that has more patients.

Finally, under the covariate-adaptive adjustable biased coin design for marginal balance, the next patient is allocated to treatment 1 with probability $F^a(x)$ where $x = \bar{D}_m$. If the third random number for this patient has value greater than or equal to zero and less than $F^a(x)$, treatment 1 will be allocated to this patient. If the third random number has value greater than or equal to $F^a(x)$ and less than or equal to 1, then this patient is allocated to treatment 2.

3.7.3 Model and test for treatment difference

In the simulations, interactions can either be present or absent between the covariates. In both situations, we study the simulated power of the covariate-adaptive randomization schemes for global and marginal balance. Two different models for the patients' responses will be considered. Let Y_{ij} be the response of the ith patient on treatment j. Then the model used for two covariates with no interaction is

$$Y_{ij} = \mu_1 + (\mu_2 - \mu_1)I_i + \beta_1 Z_{i1} + \beta_2 Z_{i2} + \epsilon_{ij}$$

and

$$Y_{ij} = \mu_1 + (\mu_2 - \mu_1)I_i + \beta_1 Z_{i1} + \beta_2 Z_{i2} + \beta_3 (Z_{i1} * Z_{i2}) + \epsilon_{ij}$$

when there is an interaction between the covariates.

We have I_i as defined in (3.1) for $i = 1, ..., n$, and β_1 and β_2 are the unknown parameters for the first and second covariates Z_{i1} and Z_{i2}, respectively. As the levels for a covariate add up to zero, for two-level covariates we have either 1 or -1 for the values of Z_{i1} and Z_{i2}; and $-1, 0$ or 1 for the values of Z_{i1} and Z_{i2} when these two covariates have three levels. The term $Z_{i1} * Z_{i2}$ denotes the interaction term for the two covariates

and is the product of the two values from the first and the second covariate. Therefore, for two-level covariates, the interaction term will take values of 1 or -1, and similarly for three-level covariates, the interaction term will take values of -1, 0 or 1. The parameter for the interaction term in this model is β_3.

We can express the above two models in matrix form as in (3.2). The patients' responses Y_{ij} and the random error terms ϵ_{ij} are represented by the vectors $\mathbf{Y}$ and $\boldsymbol{\epsilon}$, respectively. The matrix $\mathbf{W}$ here is exactly the same as in (3.2). This is an $n \times 2$ matrix with all ones in the first column and the values for I_i in the second column. The column vector $\boldsymbol{\mu}$ is also the same as in (3.2).

Next we have $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$ when there is no interaction between the two covariates and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T$ when there is an interaction between the covariates. The matrix $\mathbf{Z}$ is

$$\mathbf{Z} = \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \\ \vdots & \vdots \\ Z_{n_1 1} & Z_{n_1 2} \\ Z_{(n_1+1)1} & Z_{(n_1+1)2} \\ \vdots & \vdots \\ Z_{n1} & Z_{n2} \end{pmatrix}$$

and

$$\mathbf{Z} = \begin{pmatrix} Z_{11} & Z_{12} & Z_{11} * Z_{12} \\ Z_{21} & Z_{22} & Z_{21} * Z_{22} \\ \vdots & \vdots & \vdots \\ Z_{n_1 1} & Z_{n_1 2} & Z_{n_1 1} * Z_{n_1 2} \\ Z_{(n_1+1)1} & Z_{(n_1+1)2} & Z_{(n_1+1)1} * Z_{(n_1+1)2} \\ \vdots & \vdots & \vdots \\ Z_{n1} & Z_{n2} & Z_{n1} * Z_{n2} \end{pmatrix},$$

respectively.

We construct a test of $H_0 : \mu_2 - \mu_1 = 0$ against $H_1 : \mu_2 - \mu_1 > 0$. The same procedure will be applied as in Section 3.5. An analysis of covariance t test is used and the power is simulated for the three covariate-adaptive randomization schemes under global and marginal balance, when there are two or three levels for the covariates and when interactions between the two covariates are either present or absent. This test has a Student's t distribution under H_0 with $n - 2 - q$ degrees of freedom.

3.7.4 Power under global and marginal balance

We now study the numerical values for the power obtained using the above test by simulation under the covariate-adaptive simple random sampling, the covariate-adaptive biased coin design and the covariate-adaptive adjustable biased coin design. Let the total number of patients in a trial be n. We will study two scenarios: $n = 100$ and $n = 200$. Let $\beta_1 = 1$, $\beta_2 = 0.75$ and $\beta_3 = 3$. For all $i = 1, ..., n$ and $j = 1, 2$, when $n = 100$, the random error term ϵ_{ij} is generated from a standard normal distribution, whereas, when $n = 200$, ϵ_{ij} is generated from a normal distribution with mean 0 and standard deviation 1.5. Let the significance level of the test be $\alpha = 0.05$.

We reject H_0 when the test statistic is greater than some critical value from a Stu-

dent's t distribution. When $n = 100$, we have 95 degrees of freedom when there are interactions and 96 degrees of freedom when there are no interactions. Similarly, when $n = 200$, we have 195 and 196 degrees of freedom, respectively.

The adjusted means m_1 and m_2 described in Section 3.5 for treatments 1 and 2, respectively, are calculated as the estimators of the population means μ_1 and μ_2. These two adjusted means are needed in the numerator of the test statistic T_C. To obtain the estimator of the variance of $\hat{\mu}_2 - \hat{\mu}_1$ in the denominator of the test statistic, recall from Section 3.5 that we need to use (3.6). Here, $\mathbf{a}$ is a column vector containing the coefficients in the linear combination of $\boldsymbol{\mu}$. In this case, we have $\mu_2 - \mu_1 = (0, 1)\,\boldsymbol{\mu}$. From the above, we know that $\mathbf{a}^T = (0, 1)$. Therefore, we have $\widehat{\mathrm{Var}(\hat{\mu}_2 - \hat{\mu}_1)}$, which is the estimator of the variance of $\hat{\mu}_2 - \hat{\mu}_1$ with $\mathbf{a}^T = (0, 1)$ in (3.6).

We simulate each such trial $10,000$ times and record the number of times that H_0 is rejected. The proportion of rejections is calculated and represents the simulated power of the test. The simulated power will be studied for different values of $\bar{d}$ and different values of the design parameters in the biased coin designs.

In what follows, we abbreviated covariate-adaptive simple random sampling, the covariate-adaptive biased coin design and the covariate-adaptive adjustable biased coin design by CSRS, CBCD and CABCD. The following four tables give values for the simulated power under global and marginal balance for three-level and two-level covariates with interactions.

Table 3.3: Powers of CSRS, CBCD and CABCD with interactions between the three-level covariates when $n = 100$ and $\alpha = 0.05$

Schemes	$\bar{d} = 0$	$\bar{d} = 0.25$	$\bar{d} = 0.5$	$\bar{d} = 0.75$	$\bar{d} = 1$	$\bar{d} = 1.25$
Global Balance						
CSRS(p=1/2)	0.053	0.329	0.782	0.977	0.999	1
CBCD(p=8/12)	0.053	0.349	0.795	0.981	1	1
CBCD(p=9/12)	0.051	0.337	0.795	0.979	0.999	1
CBCD(p=10/12)	0.049	0.339	0.802	0.981	1	1
CBCD(p=11/12)	0.050	0.338	0.801	0.983	1	1
CABCD(a=1)	0.049	0.335	0.795	0.981	0.999	1
CABCD(a=2)	0.050	0.344	0.795	0.979	1	1
CABCD(a=4)	0.050	0.342	0.797	0.980	1	1
Marginal Balance						
CSRS(p=1/2)	0.050	0.341	0.791	0.978	1	1
CBCD(p=8/12)	0.048	0.335	0.788	0.979	1	1
CBCD(p=9/12)	0.047	0.337	0.784	0.979	0.999	1
CBCD(p=10/12)	0.049	0.346	0.790	0.978	0.999	1
CBCD(p=11/12)	0.054	0.345	0.797	0.978	1	1
CABCD(a=1)	0.052	0.334	0.785	0.976	1	1
CABCD(a=2)	0.050	0.350	0.792	0.979	0.999	1
CABCD(a=4)	0.050	0.332	0.790	0.980	0.999	1

Table 3.4: Powers of CSRS, CBCD and CABCD with interactions between the three-level covariates when $n = 200$ and $\alpha = 0.05$

Schemes	$\bar{d} = 0$	$\bar{d} = 0.25$	$\bar{d} = 0.5$	$\bar{d} = 0.75$	$\bar{d} = 1$	$\bar{d} = 1.25$
Global Balance						
CSRS(p=1/2)	0.048	0.317	0.753	0.966	0.999	1
CBCD(p=8/12)	0.050	0.315	0.760	0.971	0.999	1
CBCD(p=9/12)	0.047	0.319	0.754	0.970	0.999	1
CBCD(p=10/12)	0.050	0.316	0.754	0.969	0.999	1
CBCD(p=11/12)	0.050	0.317	0.767	0.970	0.999	1
CABCD(a=1)	0.053	0.322	0.757	0.965	0.999	1
CABCD(a=2)	0.048	0.315	0.766	0.968	0.999	1
CABCD(a=4)	0.054	0.320	0.757	0.971	0.998	1
Marginal Balance						
CSRS(p=1/2)	0.052	0.328	0.754	0.969	0.998	1
CBCD(p=8/12)	0.051	0.316	0.755	0.968	0.999	1
CBCD(p=9/12)	0.048	0.312	0.761	0.971	0.998	1
CBCD(p=10/12)	0.049	0.317	0.760	0.969	0.999	1
CBCD(p=11/12)	0.048	0.321	0.753	0.969	0.998	1
CABCD(a=1)	0.052	0.320	0.758	0.969	0.999	1
CABCD(a=2)	0.049	0.323	0.754	0.967	0.999	1
CABCD(a=4)	0.052	0.311	0.749	0.969	0.999	1

Table 3.5: Powers of CSRS, CBCD and CABCD with interactions between the two-level covariates when $n = 100$ and $\alpha = 0.05$

Schemes	$\bar{d} = 0$	$\bar{d} = 0.25$	$\bar{d} = 0.5$	$\bar{d} = 0.75$	$\bar{d} = 1$	$\bar{d} = 1.25$
Global Balance						
CSRS(p=1/2)	0.048	0.336	0.785	0.979	1	1
CBCD(p=8/12)	0.050	0.342	0.795	0.979	1	1
CBCD(p=9/12)	0.049	0.343	0.802	0.982	1	1
CBCD(p=10/12)	0.051	0.347	0.794	0.983	1	1
CBCD(p=11/12)	0.052	0.343	0.800	0.979	0.999	1
CABCD(a=1)	0.052	0.345	0.800	0.981	1	1
CABCD(a=2)	0.048	0.333	0.796	0.978	1	1
CABCD(a=4)	0.053	0.344	0.789	0.983	0.999	1
Marginal Balance						
CSRS(p=1/2)	0.048	0.336	0.790	0.982	0.999	1
CBCD(p=8/12)	0.049	0.340	0.791	0.977	0.999	1
CBCD(p=9/12)	0.053	0.347	0.790	0.979	1	1
CBCD(p=10/12)	0.048	0.336	0.791	0.981	0.999	1
CBCD(p=11/12)	0.045	0.349	0.787	0.979	1	1
CABCD(a=1)	0.047	0.350	0.794	0.982	1	1
CABCD(a=2)	0.049	0.341	0.796	0.981	1	1
CABCD(a=4)	0.052	0.345	0.793	0.978	1	1

Table 3.6: Powers of CSRS, CBCD and CABCD with interactions between the two-level covariates when $n = 200$ and $\alpha = 0.05$

Schemes	$\bar{d} = 0$	$\bar{d} = 0.25$	$\bar{d} = 0.5$	$\bar{d} = 0.75$	$\bar{d} = 1$	$\bar{d} = 1.25$
Global Balance						
CSRS(p=1/2)	0.050	0.317	0.756	0.971	0.999	1
CBCD(p=8/12)	0.047	0.318	0.760	0.971	0.998	1
CBCD(p=9/12)	0.050	0.316	0.759	0.969	0.999	1
CBCD(p=10/12)	0.051	0.325	0.762	0.967	0.999	1
CBCD(p=11/12)	0.051	0.308	0.762	0.969	0.999	1
CABCD(a=1)	0.054	0.320	0.764	0.972	0.999	1
CABCD(a=2)	0.051	0.321	0.758	0.971	0.999	1
CABCD(a=4)	0.051	0.321	0.759	0.973	0.999	1
Marginal Balance						
CSRS(p=1/2)	0.054	0.316	0.753	0.968	0.999	1
CBCD(p=8/12)	0.049	0.324	0.756	0.969	0.999	1
CBCD(p=9/12)	0.050	0.317	0.763	0.966	0.999	1
CBCD(p=10/12)	0.050	0.317	0.754	0.969	0.999	1
CBCD(p=11/12)	0.049	0.314	0.751	0.967	0.999	1
CABCD(a=1)	0.051	0.323	0.755	0.970	0.998	1
CABCD(a=2)	0.049	0.315	0.768	0.969	0.999	1
CABCD(a=4)	0.050	0.312	0.756	0.969	0.999	1

The first column in the tables shows the actual significance level of the test in each scenario. Due to the variation in the simulations, we obtain values close to 0.05 but not exactly this value. The power increases when $\bar{d}$ increases. For the CBCD and CABCD,

the power also increases when the design becomes more deterministic, that is, when p in the CBCD and a in the CABCD increase. For a fixed sample size, we can see from the tables that the power is highest for the CABCD and lowest for CSRS. This means that the former design will require fewer patients than the other two to achieve the same level of power.

Obviously, the simulated power under covariate-adaptive simple random sampling gives the lowest power among all schemes. The standard error of the values in the second column is $\sqrt{0.3 \times 0.7/10,000} \approx 0.004$. For example, for 3-level covariates and $n = 100$, the value of the power for CSRS under global balance is 0.329 and is 0.349 for the CBCD with $p = 8/12$. The difference between the two powers is 0.02, which is more than three standard errors. In addition, under marginal balance, the increase in the power is around one to two standard errors. We can see that the increase in the power is greater under global balance than marginal balance. These conclusions are also generally true for other values of $\bar{d}$ and when $n = 200$. Furthermore, the same conclusions apply when there are only two levels for the covariates.

The results show a genuine increase in the power when using the covariate-adaptive biased coin design instead of covariate-adaptive simple random sampling. In other words, the covariate-adaptive biased coin design achieves a more balanced trial than covariate-adaptive simple random sampling when patients are classified according to their prognostic profiles. In addition, both covariate-adaptive designs gain more power when we consider global balance instead of marginal balance when there are interactions between the covariates.

In the situation without covariates, the adjustable biased coin design has been proved theoretically to give a more balanced trial than Efron's biased coin design. We will first discuss the difference in the simulated powers for the covariate-adaptive adjustable

biased coin design and the covariate-adaptive biased coin design when there is an interaction between the covariates. As p in the covariate-adaptive biased coin design and a in the covariate-adaptive adjustable biased coin design increase, theoretically both of the designs will become more deterministic in treatment allocation and hence increase the power. From the values of the simulated power obtained, for both two-level and three-level covariates and when $n = 100$ or $n = 200$, under the CABCD, the increase in the power when a increases is not very obvious. It shows around one to two standard errors of variation with different values of a in the above four tables for global and marginal balance. The tables also suggested that the power under the CBCD shows no obvious increase when p increases. For both global and marginal balance, the variation in the power is less than two standard errors.

We can see that, when there is an interaction between the covariates, there is no difference in the powers between global and marginal balance for covariate-adaptive simple random sampling. However, global balance gives around 1% more power than marginal balance under the covariate-adaptive biased coin design and the covariate-adaptive adjustable biased coin design. This means that, under these two covariate-adaptive randomization schemes, global balance is more efficient at detecting a genuine treatment difference than marginal balance.

However, by comparing the powers obtained under global balance for two-level and three-level covariates, there is no obvious difference in their values for the same value of $\bar{d}$ under the three covariate-adaptive randomization schemes. Therefore, we cannot conclude in the case where there are interactions between covariates whether two-level or three-level covariates give a higher power under global balance. This conclusion holds for marginal balance as well.

Next, the simulated power will be shown in the following four tables for the covariate-

adaptive randomization schemes when there are no interactions between the two covariates under global and marginal balance.

Table 3.7: Powers of CSRS, CBCD and CABCD with no interactions between the three-level covariates when $n = 100$ and $\alpha = 0.05$

Schemes	$\bar{d} = 0$	$\bar{d} = 0.25$	$\bar{d} = 0.5$	$\bar{d} = 0.75$	$\bar{d} = 1$	$\bar{d} = 1.25$
Global Balance						
CSRS(p=1/2)	0.049	0.331	0.789	0.977	1	1
CBCD(p=8/12)	0.050	0.336	0.795	0.978	1	1
CBCD(p=9/12)	0.047	0.343	0.799	0.981	1	1
CBCD(p=10/12)	0.049	0.346	0.797	0.982	0.999	1
CBCD(p=11/12)	0.046	0.342	0.792	0.981	1	1
CABCD(a=1)	0.049	0.348	0.794	0.980	0.999	1
CABCD(a=2)	0.049	0.343	0.797	0.978	0.999	1
CABCD(a=4)	0.053	0.338	0.799	0.979	0.999	1
Marginal Balance						
CSRS(p=1/2)	0.047	0.333	0.789	0.978	0.999	1
CBCD(p=8/12)	0.048	0.334	0.789	0.979	1	1
CBCD(p=9/12)	0.048	0.344	0.798	0.980	1	1
CBCD(p=10/12)	0.052	0.339	0.794	0.977	1	1
CBCD(p=11/12)	0.049	0.337	0.798	0.979	0.999	1
CABCD(a=1)	0.051	0.338	0.792	0.980	0.999	1
CABCD(a=2)	0.051	0.353	0.795	0.977	1	1
CABCD(a=4)	0.051	0.335	0.795	0.979	0.999	1

Table 3.8: Powers of CSRS, CBCD and CABCD with no interactions between the three-level covariates when $n = 200$ and $\alpha = 0.05$

Schemes	$\bar{d} = 0$	$\bar{d} = 0.25$	$\bar{d} = 0.5$	$\bar{d} = 0.75$	$\bar{d} = 1$	$\bar{d} = 1.25$
Global Balance						
CSRS(p=1/2)	0.048	0.320	0.752	0.970	0.999	1
CBCD(p=8/12)	0.049	0.314	0.764	0.967	0.999	1
CBCD(p=9/12)	0.046	0.315	0.763	0.971	0.998	1
CBCD(p=10/12)	0.056	0.327	0.763	0.971	0.999	1
CBCD(p=11/12)	0.051	0.311	0.767	0.972	0.999	1
CABCD(a=1)	0.049	0.323	0.758	0.968	0.999	1
CABCD(a=2)	0.053	0.326	0.757	0.972	0.999	1
CABCD(a=4)	0.051	0.318	0.760	0.970	0.999	1
Marginal Balance						
CSRS(p=1/2)	0.050	0.314	0.748	0.965	0.999	1
CBCD(p=8/12)	0.056	0.316	0.751	0.967	0.999	1
CBCD(p=9/12)	0.049	0.312	0.755	0.968	0.998	1
CBCD(p=10/12)	0.049	0.317	0.760	0.969	0.999	1
CBCD(p=11/12)	0.043	0.318	0.766	0.970	0.999	1
CABCD(a=1)	0.047	0.314	0.761	0.971	0.999	1
CABCD(a=2)	0.053	0.323	0.752	0.971	0.998	1
CABCD(a=4)	0.049	0.322	0.750	0.968	0.998	1

Table 3.9: Powers of CSRS, CBCD and CABCD with no interaction between the two-level covariates when $n = 100$ and $\alpha = 0.05$

Schemes	$\bar{d} = 0$	$\bar{d} = 0.25$	$\bar{d} = 0.5$	$\bar{d} = 0.75$	$\bar{d} = 1$	$\bar{d} = 1.25$
Global Balance						
CSRS(p=1/2)	0.051	0.335	0.794	0.975	0.999	1
CBCD(p=8/12)	0.051	0.333	0.803	0.983	1	1
CBCD(p=9/12)	0.050	0.352	0.798	0.981	1	1
CBCD(p=10/12)	0.052	0.343	0.796	0.981	0.999	1
CBCD(p=11/12)	0.049	0.343	0.801	0.982	1	1
CABCD(a=1)	0.049	0.339	0.790	0.979	0.999	1
CABCD(a=2)	0.049	0.345	0.797	0.981	1	1
CABCD(a=4)	0.053	0.345	0.794	0.981	1	1
Marginal Balance						
CSRS(p=1/2)	0.049	0.333	0.789	0.979	1	1
CBCD(p=8/12)	0.048	0.348	0.789	0.982	0.999	1
CBCD(p=9/12)	0.051	0.338	0.795	0.982	0.999	1
CBCD(p=10/12)	0.047	0.343	0.792	0.981	1	1
CBCD(p=11/12)	0.050	0.346	0.800	0.980	0.999	1
CABCD(a=1)	0.047	0.346	0.796	0.980	0.999	1
CABCD(a=2)	0.056	0.333	0.802	0.980	0.999	1
CABCD(a=4)	0.055	0.346	0.798	0.982	0.999	1

Table 3.10: Powers of CSRS, CBCD and CABCD with no interaction between the two-level covariates when $n = 200$ and $\alpha = 0.05$

Schemes	$\bar{d} = 0$	$\bar{d} = 0.25$	$\bar{d} = 0.5$	$\bar{d} = 0.75$	$\bar{d} = 1$	$\bar{d} = 1.25$
Global Balance						
CSRS(p=1/2)	0.052	0.325	0.752	0.967	0.999	1
CBCD(p=8/12)	0.051	0.316	0.763	0.969	0.999	1
CBCD(p=9/12)	0.051	0.321	0.755	0.965	0.999	1
CBCD(p=10/12)	0.049	0.325	0.761	0.969	0.999	1
CBCD(p=11/12)	0.049	0.327	0.753	0.970	0.999	1
CABCD(a=1)	0.050	0.315	0.759	0.970	0.999	1
CABCD(a=2)	0.050	0.309	0.757	0.967	0.999	1
CABCD(a=4)	0.050	0.321	0.756	0.971	0.999	1
Marginal Balance						
CSRS(p=1/2)	0.047	0.319	0.753	0.967	0.998	1
CBCD(p=8/12)	0.051	0.314	0.756	0.967	0.999	1
CBCD(p=9/12)	0.046	0.310	0.760	0.967	0.999	1
CBCD(p=10/12)	0.055	0.313	0.761	0.968	0.999	1
CBCD(p=11/12)	0.051	0.324	0.754	0.970	0.999	1
CABCD(a=1)	0.054	0.321	0.757	0.968	0.999	1
CABCD(a=2)	0.052	0.317	0.749	0.967	0.999	1
CABCD(a=4)	0.049	0.323	0.759	0.971	0.999	1

When there is no interaction between the two covariates, the powers are similar for global and marginal balance under all three covariate-adaptive randomization schemes. Covariate-adaptive simple random sampling again gives the lowest power for each value

of $\bar{d}$. We cannot conclude whether global balance or marginal balance is better when there is no interaction between the covariates. In this case, we can say that global balance is as good as marginal balance when there is no interaction between the covariates. In addition, under each of the covariate-adaptive randomization schemes, it is not clear whether two-level or three-level covariates give a higher power under global or marginal balance.

3.8 Conclusions

Covariate-adaptive randomization schemes provide methods for patient allocation when we want to study treatment effects in patients classified by prognostic factors. This is to ensure that we have the same numbers of patients in the two treatment groups for each combination of the covariates' profiles. One of the main aims of covariate-adaptive randomization schemes is to balance the numbers of patients on the two treatments with patients classified by their prognostic factors. Under any such scheme, we can further achieve global or marginal balance. When there are two covariates, the simulated power obtained under global balance is higher than that under marginal balance when there are interactions between the covariates. The powers obtained are similar for global and marginal balance when there are no interactions between the covariates.

Numerical values for the power using a normal approximation are also given under covariate-adaptive simple random sampling and the biased coin design when a single covariate is uniformly distributed. It is shown that these values for the power over-approximate the actual value for the power in both cases. This will cause the planner for the trial to prepare less resources than are needed to achieve a particular level of power for the study. For example, we can estimate the sample size needed to achieve a certain level of power by the normal approximation method. However, in reality, under

both randomization schemes, this sample size will give a power too low. The theoretical properties and the expression for the normal approximation to the power when we consider more than one covariate are also provided.

Chapter 4

Randomization schemes with more than two treatments

In the previous two chapters, we studied different randomization schemes and their power properties when two treatments are compared in a clinical trial. It is sometimes more efficient to compare several treatments in the same trial for ease of comparison and to identify their effects. In this chapter, we consider different randomization schemes when there are more than two treatments.

4.1 Introduction

Most of the randomization schemes that we have studied and which are considered in the literature are based on trials with two treatments. Here, assume that patients enter a clinical trial sequentially and have to be assigned immediately to one of $K > 2$ treatments. Assume also that the variances of the patients' responses on different treatments are the same. It will be most efficient to have a balanced trial. These randomization schemes will have to maintain a balance in the numbers of patients across the K treatments and preserve randomness in the allocation. The assignment rules for these ran-

domization schemes can be based on different criteria. One of the most commonly used criteria is the number of patients currently on each of the treatments. For each new patient, the probabilities of assigning this patient to each of the K treatments are obtained by incorporating this information into the calculations. The probabilities obtained may be all different or some of them may be the same. The larger the probability for a given treatment, the more likely the next patient will be allocated to that particular treatment. In most cases, the goal is to obtain a larger probability of assigning the next patient to an under-represented treatment. Another criterion used for treatment assignment is the imbalance across treatments. The imbalance is defined in the two-treatment case as simply the difference in the numbers of patients on the two treatments. In the case of more than two treatments, the imbalance at each stage for each treatment has to be defined.

In Section 4.2, different randomization schemes for more than two treatments will be introduced. Their probabilities of assigning the next patient to the different treatments will also be given. In addition, a new class of designs called the adjustable biased coin design will be proposed such that each assignment is made based on all of the current imbalances at each stage. The values for these imbalances are incorporated into the probabilities of assignment to all K treatments at each stage. Section 4.3 will be concerned with comparing the covariances of the numbers of patients on different treatments under complete randomization and the biased coin design. In Section 4.4, the asymptotic properties of the randomization schemes will be described. This is followed by simulations of the imbalances under different randomization schemes in Section 4.5, where the finite-sample behaviour of the imbalances under different randomization schemes will be demonstrated. In Section 4.6, numerical values for the power of an F test are given by simulation for different randomization schemes. Finally, conclusions will be drawn in Section 4.7. Note that, similar to previous chapters, treatment 1 is always the standard

treatment. The patients that are allocated to this treatment group form the control group.

4.2 Classes of treatment assignment rules

4.2.1 Complete randomization

Under complete randomization, each patients is equally likely to be assigned to one of the K treatments. In other words, the probability of allocating treatment j to a patient for each $j = 1, ..., K$ is $1/K$. These probabilities remain the same for all stages. This design ignores the current numbers of patients on the treatments at each stage. Complete randomization gives the highest level of randomness in assignment among all schemes and selection bias is a minimum. Selection bias refers to the bias where the experimenter's decision depends on the suitability of a subject, and is discussed by Blackwell and Hodges (1957) and Efron (1971). However, complete randomization is more likely to produce severe imbalances in the numbers of patients across the treatments.

4.2.2 Permuted-block randomization

Permuted-block randomization is another example of a randomization scheme for several treatments. The block length and the proportion of patients allocated to a particular treatment within a block are pre-specified. The sequence of treatments is randomly permuted within a block. By combining the sequences for all blocks, a complete sequence for the treatment assignments is formed. Each patient arrives in a trial and is allocated to a treatment following this sequence. Under permuted-block randomization, an equal proportion of patients is usually allocated to each treatment in order to balance the numbers of patients across the treatment groups. In addition, the block size or the length of the block K_1 can be chosen to be a multiple of the number of treatments in

the trial so as to obtain a balance across the treatment groups for every K_1 patients in the trial. Permuted-block randomization is very efficient at maintaining a balance in the numbers of patients across the treatment groups. However, the assignments are very deterministic and the selection bias for this design is very high.

In this chapter, we will not consider permuted-block randomization. In practice, a clinical trial can be carried out in different centres. The assignment rules for different centres may be the same or different. When a patient arrives at a particular centre, the assignment rules for this particular centre will be applied. This is called a centre-stratified randomization scheme and centre-stratified permuted-block randomization will be discussed in the next chapter.

4.2.3 Efron's biased coin design

The biased coin design introduced by Efron (1971) gives a compromise between randomness and maintaining balance in the treatment allocation. Under the biased coin design for two-treatment trials, we have a fixed probability $p > 1/2$ of allocating the next patient to a treatment that has been chosen less often and $1 - p$ otherwise. This idea can be extended to the case of more than two treatments. The aim is to balance the proportion of patients on each of the treatments. We have a fixed probability $p > 1/K$ of allocating the next patient to the treatment that has the least number of patients. Furthermore, the probability of allocating the next patient to one of the other treatments is $(1-p)/(K-1)$. If we have more than one treatment that has the least number of patients, the following can be applied. Let n^* be the number of treatments that have the least number of patients. Then, for a chosen probability $p > 1/K$, we have p/n^* as the probability of assigning the next patient to one of the n^* treatments and $(1 - p)/(K - n^*)$ as the probability of allocating the next patient to one of the other treatments.

4.2.4 Optimum biased coin design

In Atkinson (1982, 2004), the ideas of D- and D_A-optimality are used to obtain the probabilities in the biased coin design for allocating the next patient to one of the K treatments. The linear model $E(\mathbf{y}) = x^T\boldsymbol{\beta}$ is for the responses of the patients, which are independent and have variance σ^2. The variance of the least square estimates of $\boldsymbol{\beta}$ is $\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$, where the $p \times p$ information matrix $\mathbf{X}^T\mathbf{X}$ is assumed to be of full rank from n observations. In the construction of optimum experimental designs, let the measure be ξ_n for an n-point design over the design region $\mathbb{X}$. The information matrix for this design is $M(\xi_n) = n^{-1}(\mathbf{X}^T\mathbf{X})$. The fitted value is $\hat{y}(x) = \hat{\beta}^T x$ at x with variance $\mathrm{Var}\{\hat{y}(x)\} = \sigma^2 x^T(X^TX)^{-1}x$. The scaled variance of the predicted responses is $\mathrm{Var}\{\hat{y}(x)\}/\sigma^2$. The standardized variance is obtained by scaling the variance by σ^2 and the number of observations, so that it is given by

$$d(x, \xi_n) = n\frac{\mathrm{Var}(\hat{y}(x))}{\sigma^2} = x^T M^{-1}(\xi_n)x.$$

To achieve D-optimality, the determinant of $M(\xi_n)$ is maximized and this minimizes the variances of the estimates of the parameters. For an n-point design, the $(n+1)$st point is added where $d(x, \xi_n)$ is a maximum.

If all of the parameters in the model are of interest, the D-optimal criterion will be appropriate. Alternatively, if the contrasts between the treatment effects are of interest, the D_A-optimal criterion can be used. For most of this chapter, the D_A-optimal criterion will be appropriate to use instead of the D-optimal criterion. The contrasts are s linear combinations. These are the components of the vector $\mathbf{A}^T\boldsymbol{\beta}$, where $\mathbf{A}$ is an $p \times s$ matrix with rank $s < p$. The least squares estimate of the contrast vector will be denoted by $\mathbf{A}^T\hat{\boldsymbol{\beta}}$. The covariance matrix of $\mathbf{A}^T\hat{\boldsymbol{\beta}}$ is proportional to $\mathbf{A}^T M^{-1}(\xi_n)\mathbf{A}$ and the

standardized variance is given by

$$d_A(x, \xi_n) = x^T M^{-1}(\xi_n)\mathbf{A}\{\mathbf{A}^T M^{-1}(\xi_n)\mathbf{A}\}^{-1}\mathbf{A}^T M^{-1}(\xi_n)x.$$

Sequential D_A -optimum designs are generated one point at a time by adding an observation at x where $d_A(x, \xi_n)$ is a maximum.

In a clinical trial with $K > 2$ treatments, we will study the design region $\mathbb{X}$ which consists of K points. Here, we replace $d(x, \xi_n)$ and $d_A(x, \xi_n)$ with $d(j, \xi_n)$ and $d_A(j, \xi_n)$, respectively. The sequential construction of the D- or the D_A- optimum designs allocates the $(n+1)$st patient to the treatment such that its respective standardized variance $d(j, \xi_n)$ or $d_A(j, \xi_n)$ is maximized.

The probability of allocating the next patient to treatment j for the D-optimal biased coin design is

$$p_j = \frac{d(j, \xi_n)}{\sum_{i=1}^{K} d(i, \xi_n)}$$

and

$$p_j = \frac{d_A(j, \xi_n)}{\sum_{i=1}^{K} d_A(i, \xi_n)},$$

for the D_A-optimal biased coin design. Assume that the $s = (K-1)$-dimensional space of the contrasts is orthogonal to the overall mean for which the matrix $\mathbf{A}^T$ is arbitrary. Then the standardized variance can be written as

$$d_A(j, \xi_n) = (n - n_j)/n_j = r_j - 1,$$

where $r_j = n/n_j$ is the reciprocal of the proportion of patients on treatment j. Under the D_A-optimality criterion, the probability of allocating treatment j for $j = 1, ..., K$ to the next patient is

$$p_j = \frac{r_j - 1}{\sum_{i=1}^{K} r_i - K}.$$

4.2.5 Wei's class of biased coin designs

A class of treatment assignment rules was suggested by Wei (1978) for more than two treatments. Suppose that, in a clinical trial after n assignments, there are $n_{j,n}$ patients on treatment j for $j = 1, ..., K$ such that $\sum_{j=1}^{K} n_{j,n} = n$. Let $\mathbf{p}$ be the $K \times 1$ vector $\mathbf{p} = (p_1, p_2, ..., p_K)^T$, where p_j is the probability of allocating treatment j to the $(n+1)$st patient. Then the vector $\mathbf{p}$ is a function of the vector $(n_{1,n}/n, n_{2,n}/n, ..., n_{K,n}/n)^T$ and has the following properties.

1. The vector $\mathbf{p}$ is a function from Ω to Ω where

$$\Omega = \left\{ \mathbf{y} = (y_1, y_2, ..., y_K)^T : \sum_{j=1}^{K} y_j = 1 \right\}.$$

2. If $y_i < y_j$, then $p_i \geq p_j$ and, if $y_i = y_j$, then $p_i = p_j$ for $i \neq j$ and $i, j = 1, ..., K$.

3. We have $p_j\left(y_1, ..., y_{j-1}, 1/K, y_{j+1}, ..., y_K\right) = 1/K$ for all y_i such that $i \neq j$ and $i = 1, ..., K$. In addition, $\sum_{i \neq j} y_i = 1 - 1/K$.

4. Finally, $p_j(\mathbf{y})$ is continuous at the point $(1/K, ..., 1/K)^T$ for $j = 1, ..., K$.

The numbers of patients on the treatments after n assignments are represented by the vector $\mathbf{D}_n = (n_{1,n}, n_{2,n}, ..., n_{K,n})^T$. This vector forms a Markov chain with transition probabilities

$$p_j\left(\mathbf{D}_n/n\right) = P\left\{\mathbf{D}_{n+1} = (n_{1,n}, ..., n_{K,n})^T + \mathbf{e}_j | \mathbf{D}_n = (n_{1,n}, ..., n_{K,n})^T\right\},$$

where $\mathbf{e}_j$ is a $K \times 1$ vector which contains 1 in its jth component and zeros elsewhere for $j = 1, ..., K$. Wei (1978) proves that $E(n_{j,n}) = n/K$ and $E[p_j(\mathbf{D}_n/n)] = 1/K$, and also that $\mathbf{D}_n/n \to (1/K, ..., 1/K)^T$ in probability as $n \to \infty$. It is also shown by induction that the variance of the number of patients on treatment j after n assignments under this class of treatment assignment rules is smaller than the corresponding variance

under complete randomization. In the next section, these results will be extended to the covariance of the numbers of patients on any two treatments.

4.2.6 Smith's extension of Wei's class of biased coin designs

Wei's class of treatment assignment rules was extended by Smith (1984). A new procedure was proposed by generalizing Wei's procedure to achieve a limiting design measure $(\xi_1, ..., \xi_K)$ for $\xi_j \geq 0$, $j = 1, ..., K$ and $\sum_{j=1}^{K} \xi_j = 1$. Here, we want the limiting proportion of patients allocated to treatment j to be ξ_j. In this case, all ξ_j for $j = 1, .., K$ are known at the start of the trial. The vector $\mathbf{p}$ has to satisfy two properties.

1. The vector $\mathbf{p}$ is twice continuously differentiable on Ω.

2. If $y_j \geq \xi_j$, then $p_j(\mathbf{y}) \leq \xi_j$.

The main result is given as the following theorem. Let ρ be the parameter which is the sum of all the limiting proportions of patients allocated to the treatments such that $\rho = \sum_{j=1}^{K} \xi_j$. Further, let $\delta_{j,n}$ be 1 if the nth patient is allocated to treatment j and 0 otherwise. Then we have the approximation

$$\mathrm{Cov}(n_{j,l}, n_{m,n}) \approx \begin{cases} (1+2\rho)^{-1} l^{1+\rho} n^{-\rho} (\xi_j - \xi_j^2), & j = m, \\ -(1+2\rho)^{-1} l^{1+\rho} n^{-\rho} \xi_j \xi_m, & j \neq m, \end{cases}$$

for $l \leq n$ as $l \to \infty$ with the joint distributions asymptotically normal and

$$\mathrm{Cov}(\delta_{j,l}, \delta_{m,n}) \approx \begin{cases} \xi_j - \xi_j^2, & l = n, j = m, \\ -\xi_j \xi_m, & l = n, j \neq m, \\ -\rho(1+\rho)(1+2\rho)^{-1} l^{\rho} n^{-1-\rho} (\xi_j - \xi_j^2), & l \neq n, j = m, \\ \rho(1+\rho)(1+2\rho)^{-1} l^{\rho} n^{-1-\rho} \xi_j \xi_m, & l \neq n, j \neq m. \end{cases}$$

The above two approximations for the covariances are very useful in obtaining the asymptotic form of the covariance matrix under Wei's class of biased coin designs. These will be given explicitly in Section 4.4.

4.2.7 The adjustable biased coin design

The idea of the biased coin assignment rule can be extended to situations where more than two treatments are used in a trial. The simplest case was given for Efron's biased coin assignment rule in Section 4.2.3. This assignment rule only depends on the number of patients on each treatment at each stage. The probabilities are fixed and do not take into account the numbers of patients on all treatments at different stages and the degree of imbalance at different stages. A new class of treatment assignment rules is now suggested. Here, the probability of assigning the next patient to treatment j is calculated by taking into account the imbalances on all treatments.

The imbalance on treatment j at stage n is defined by

$$\Delta_{j,n} = n_{j,n} - \frac{n}{K}.$$

At each stage n, the imbalance shows the difference between the number of patients on treatment j and the average number of patients per treatment. When the imbalance is zero, this implies that we have an average number of patients on treatment j at stage n and that there is no imbalance. A negative imbalance shows that treatment j is under-represented with fewer patients than the average at stage n and vice versa for a positive imbalance. The sum of these imbalances at any stage n is 0, that is, $\sum_{j=1}^{K} \Delta_{j,n} = 0$.

Let $F_j(z)$ be a function of the current imbalance on treatment j. This function will be different according to the sign of the current imbalance. Define this function as $F_j : \mathbb{R} \rightarrow [0, 1]$, where $\mathbb{R}$ is the set of real numbers. Then the function depends on the value $z \in \Delta_{j,n}$ for any treatment j after n assignments. For a particular stage n and treatment j, the

imbalance is zero when the number of patients on treatment j equals the average number of patients per treatment. So, at stage $n+1$, we define $F_j(z) = 1/K$. If the value of the imbalance is negative, which means that treatment j is under-represented, we will need more patients on this treatment to achieve a balanced trial. Therefore, for the $(n+1)$st assignment, we should have $F_j(z) > 1/K$. Similarly, for a positive imbalance, there are more patients on treatment j than on average, and we should have $F_j(z) < 1/K$. The values for $F_j(z)$ take into account the degree of imbalance on treatment j at stage n. The function $F_j(z)$ is not symmetric and is decreasing. More specifically, after n assignments, the function $F_j(z)$ is defined as

$$F_j(z) = \begin{cases} \frac{1}{|z|+K}, & z > 0, \\ \frac{1}{K}, & z = 0, \\ \frac{|z|+1}{|z|+K}, & z < 0, \end{cases}$$

where $z \in \Delta_{j,n}$.

For $j = 1, ..., K$, the probability of assigning the next patient to treatment j is defined as

$$p_j = \frac{F_j(z)}{\sum_{i=1}^{K} F_i(z)}, \tag{4.1}$$

where $\sum_{j=1}^{K} p_j = 1$. As $F_j(z)$ depends on the imbalance on treatment j, the probability p_j of the next patient being allocated to treatment j depends on $F_i(z)$ for all treatments $i = 1, ..., K$ and is therefore obtained from the imbalances on all treatments from the previous stage. We call this class of designs the adjustable biased coin design.

Remark For the above design, assume that $K = 2$. Then the probability p_j for $j = 1, 2$ of assigning the next patient to treatment j becomes $p_j = F_j(z)$, as the denominator in (4.1) is $\sum_{i=1}^{2} F_i(z) = 1$. It is clear that this does not give the same results as the adjustable biased coin design for two treatments proposed by Baldi Antognini and Giovagnoli (2004). Recall from Chapter 2 that $p_j = F_a(x)$ for this design. For $K = 2$,

the value z in $F_j(z)$ is the difference between the number of patients on treatment j for $j = 1, 2$ and the average number of patients on each treatment up to n assignments. This z can be any real number. However, the value x in $F_a(x)$ is the difference between the number of patients on treatment 1 and the number of patients on treatment 2. The values for x can only be integers and this is the imbalance defined by Baldi Antognini and Giovagnoli (2004) for their adjustable biased coin design. This means that the imbalances for the two designs are different.

4.3 Comparison of the covariances under complete randomization and Wei's class of biased coin designs

In this section, we will compare the covariances of the numbers of patients on any two treatments for complete randomization and Wei's class of biased coin designs. We want to show by induction that the covariances under complete randomization are at least as large as those under Wei's class of biased coin designs.

Let $\mathbf{S}_n = (L_{1,n}, ..., L_{K,n})^T$ be the $K \times 1$ vector, where $L_{j,n}$ is the number of patients on treatment j after n assignments under complete randomization. It is clear that $\mathbf{S}_n$ has a multinomial distribution with parameters n and $\mathbf{p}$. Under complete randomization, each patient is equally likely to be assigned to one of the K treatments. The probability p_j of assigning treatment j to the next patient will be $1/K$ for $j = 1, ..., K$. So the covariance of the numbers of patients on treatment j and m for $j \neq m$ and $j, m = 1, ..., K$ is $\mathrm{Cov}(L_{j,n}, L_{m,n}) = -np_jp_m = -n/K^2$ after n assignments. We want to show that $\mathrm{Cov}(L_{j,n}, L_{m,n}) \geq \mathrm{Cov}(n_{j,n}, n_{m,n})$ for all $n \geq 1$ and $j \neq m$. Now, if $n = 1$, the covariance under complete randomization is $\mathrm{Cov}(L_{j,1}, L_{m,1}) = -1/K^2$. Under the biased coin design, the assignment rule for the first assignment is just complete

randomization. Therefore, if $n = 1$, we have $\mathrm{Cov}(L_{j,1}, L_{m,1}) \geq \mathrm{Cov}(n_{j,1}, n_{m,1})$. We assume that the result holds at stage n, so that

$$\mathrm{Cov}(L_{j,n}, L_{m,n}) \geq \mathrm{Cov}(n_{j,n}, n_{m,n})$$

for $j \neq m$ and $j, m = 1, ..., K$.

We can write the covariance after $n + 1$ assignments for the biased coin design as

$$\begin{aligned}
\mathrm{Cov}(n_{j,n+1}, n_{m,n+1}) &= E[\{n_{j,n+1} - E(n_{j,n+1})\}\{n_{m,n+1} - E(n_{m,n+1})\}] \\
&= E\left[\left\{n_{j,n+1} - \frac{n+1}{K}\right\}\left\{n_{m,n+1} - \frac{n+1}{K}\right\}\right] \\
&= E\left(E\left[\left\{n_{j,n+1} - \frac{n+1}{K}\right\}\left\{n_{m,n+1} - \frac{n+1}{K}\right\}\right]\Big|\mathbf{D}_n\right).
\end{aligned}$$

After $n + 1$ assignments, for any treatment $j = 1, ..., K$, the number of patients on treatment j will be $n_{j,n+1} = n_{j,n}$ or $n_{j,n+1} = n_{j,n} + 1$. Now consider treatment m where $m \neq j$ and $m = 1, ..., K$. If $n_{j,n+1} = n_{j,n}$, we will have either $n_{m,n+1} = n_{m,n} + 1$ or $n_{m,n+1} = n_{m,n}$. However, if $n_{j,n+1} = n_{j,n} + 1$, then $n_{m,n+1} = n_{m,n}$. Let $q_j(\mathbf{D}_n/n) = 1 - p_j(\mathbf{D}_n/n)$ for any treatment $j = 1, ..., K$, so that $p_j(\mathbf{D}_n/n) + q_j(\mathbf{D}_n/n) = 1$.

Using the above, the covariance for the biased coin design after $n + 1$ assignments can be written as

$$\begin{aligned}
& E\left[\left\{n_{j,n} + 1 - \frac{n+1}{K}\right\} p_j\left(\frac{\mathbf{D}_n}{n}\right)\left(n_{m,n} - \frac{n+1}{K}\right) q_m\left(\frac{\mathbf{D}_n}{n}\right)\right] \\
+ \; & E\left\{\left(n_{j,n} - \frac{n+1}{K}\right) q_j\left(\frac{\mathbf{D}_n}{n}\right)\left(n_{m,n} + 1 - \frac{n+1}{K}\right) p_m\left(\frac{\mathbf{D}_n}{n}\right)\right\} \\
+ \; & E\left\{\left(n_{j,n} - \frac{n+1}{K}\right) q_j\left(\frac{\mathbf{D}_n}{n}\right)\left(n_{m,n} - \frac{n+1}{K}\right) q_m\left(\frac{\mathbf{D}_n}{n}\right)\right\}.
\end{aligned}$$

We can expand the first expectation as

$$\begin{aligned}
& E\left[\left\{\left(n_{j,n} - \frac{n}{K}\right) p_j\left(\frac{\mathbf{D}_n}{n}\right) + \left(1 - \frac{1}{K}\right) p_j\left(\frac{\mathbf{D}_n}{n}\right)\right\}\right. \\
& \left.\times \left\{\left(n_{m,n} - \frac{n}{K}\right) q_m\left(\frac{\mathbf{D}_n}{n}\right) - \frac{1}{K} q_m\left(\frac{\mathbf{D}_n}{n}\right)\right\}\right].
\end{aligned}$$

Similarly, the second expectation can be expanded as

$$E\left[\left\{\left(n_{j,n}-\frac{n}{K}\right)q_j\left(\frac{\mathbf{D}_n}{n}\right)-\frac{1}{K}q_j\left(\frac{\mathbf{D}_n}{n}\right)\right\}\right.$$
$$\left.\times\left\{\left(n_{m,n}-\frac{n}{K}\right)p_m\left(\frac{\mathbf{D}_n}{n}\right)+\left(1-\frac{1}{K}\right)p_m\left(\frac{\mathbf{D}_n}{n}\right)\right\}\right].$$

Finally, the third expectation can be written as

$$E\left[\left\{\left(n_{j,n}-\frac{n}{K}\right)q_j\left(\frac{\mathbf{D}_n}{n}\right)-\frac{1}{K}q_j\left(\frac{\mathbf{D}_n}{n}\right)\right\}\right.$$
$$\left.\times\left\{\left(n_{m,n}-\frac{n}{K}\right)q_m\left(\frac{\mathbf{D}_n}{n}\right)-\frac{1}{K}q_m\left(\frac{\mathbf{D}_n}{n}\right)\right\}\right].$$

By combining the three expectations, the covariance of $n_{j,n+1}$ and $n_{m,n+1}$ for the biased coin design is

$$\begin{aligned}
& E\left[\left(n_{j,n}-\frac{n}{K}\right)\left(n_{m,n}-\frac{n}{K}\right)\left\{p_j\left(\frac{\mathbf{D}_n}{n}\right)q_m\left(\frac{\mathbf{D}_n}{n}\right)\right.\right.\\
& \left.\left.+\,p_m\left(\frac{\mathbf{D}_n}{n}\right)q_j\left(\frac{\mathbf{D}_n}{n}\right)+q_j\left(\frac{\mathbf{D}_n}{n}\right)q_m\left(\frac{\mathbf{D}_n}{n}\right)\right\}\right]\\
+\;& E\left[\left(n_{j,n}-\frac{n}{K}\right)\left\{-\frac{1}{K}p_j\left(\frac{\mathbf{D}_n}{n}\right)q_m\left(\frac{\mathbf{D}_n}{n}\right)\right.\right.\\
& \left.\left.+\left(1-\frac{1}{K}\right)p_m\left(\frac{\mathbf{D}_n}{n}\right)q_j\left(\frac{\mathbf{D}_n}{n}\right)-\frac{1}{K}q_j\left(\frac{\mathbf{D}_n}{n}\right)q_m\left(\frac{\mathbf{D}_n}{n}\right)\right\}\right]\\
+\;& E\left[\left(n_{m,n}-\frac{n}{K}\right)\left\{\left(1-\frac{1}{K}\right)p_j\left(\frac{\mathbf{D}_n}{n}\right)q_m\left(\frac{\mathbf{D}_n}{n}\right)\right.\right.\\
& \left.\left.-\,\frac{1}{K}p_m\left(\frac{\mathbf{D}_n}{n}\right)q_j\left(\frac{\mathbf{D}_n}{n}\right)-\frac{1}{K}q_j\left(\frac{\mathbf{D}_n}{n}\right)q_m\left(\frac{\mathbf{D}_n}{n}\right)\right\}\right]\\
+\;& E\left[-\frac{1}{K}\left(1-\frac{1}{K}\right)p_j\left(\frac{\mathbf{D}_n}{n}\right)q_m\left(\frac{\mathbf{D}_n}{n}\right)\right.\\
& \left.-\,\frac{1}{K}\left(1-\frac{1}{K}\right)p_m\left(\frac{\mathbf{D}_n}{n}\right)q_j\left(\frac{\mathbf{D}_n}{n}\right)+\frac{1}{K^2}q_j\left(\frac{\mathbf{D}_n}{n}\right)q_m\left(\frac{\mathbf{D}_n}{n}\right)\right].
\end{aligned}$$

Since we know that, for any treatment j, $E\left\{p_j\left(\mathbf{D}_n/n\right)\right\}=1/K$ and $E(n_{j,n})=n/K$,

the covariance can be written as

$$
\begin{aligned}
& E\left\{\left(n_{j,n}-\frac{n}{K}\right)\left(n_{m,n}-\frac{n}{K}\right)\right\}+E\left\{\left(n_{j,n}-\frac{n}{K}\right)p_m\left(\frac{\mathbf{D}_n}{n}\right)\right\} \\
- \quad & E\left\{\left(n_{j,n}-\frac{n}{K}\right)\left(n_{m,n}-\frac{n}{K}\right)p_j\left(\frac{\mathbf{D}_n}{n}\right)p_m\left(\frac{\mathbf{D}_n}{n}\right)\right\} \\
- \quad & \left(1-\frac{1}{K}\right)E\left\{\left(n_{j,n}-\frac{n}{K}\right)p_j\left(\frac{\mathbf{D}_n}{n}\right)p_m\left(\frac{\mathbf{D}_n}{n}\right)\right\} \\
+ \quad & E\left\{\left(n_{m,n}-\frac{n}{K}\right)p_j\left(\frac{\mathbf{D}_n}{n}\right)\right\} \\
- \quad & \left(1-\frac{1}{K}\right)E\left\{\left(n_{m,n}-\frac{n}{K}\right)p_j\left(\frac{\mathbf{D}_n}{n}\right)p_m\left(\frac{\mathbf{D}_n}{n}\right)\right\} \\
+ \quad & \left(\frac{2}{K}-\frac{1}{K^2}\right)E\left\{p_j\left(\frac{\mathbf{D}_n}{n}\right)p_m\left(\frac{\mathbf{D}_n}{n}\right)\right\}-\frac{1}{K^2}.
\end{aligned}
$$

Further, we know that $\mathrm{Cov}(n_{j,n}, n_{m,n}) = E\left\{\left(n_{j,n}-n/K\right)\left(n_{m,n}-n/K\right)\right\}$, and so $\mathrm{Cov}(n_{j,n+1}, n_{m,n+1})$ has the form

$$
\begin{aligned}
& \mathrm{Cov}(n_{j,n}, n_{m,n})-\frac{1}{K^2}-E\left\{\left(n_{j,n}-\frac{n}{K}\right)\left(n_{m,n}-\frac{n}{K}\right)p_j\left(\frac{\mathbf{D}_n}{n}\right)p_m\left(\frac{\mathbf{D}_n}{n}\right)\right\} \\
+ \quad & E\left\{\left(n_{j,n}-\frac{n}{K}\right)p_m\left(\frac{\mathbf{D}_n}{n}\right)\right\}+\left(\frac{2}{K}-\frac{1}{K^2}\right)E\left\{p_j\left(\frac{\mathbf{D}_n}{n}\right)p_m\left(\frac{\mathbf{D}_n}{n}\right)\right\} \\
- \quad & \left(1-\frac{1}{K}\right)E\left\{\left(n_{j,n}-\frac{n}{K}\right)p_j\left(\frac{\mathbf{D}_n}{n}\right)p_m\left(\frac{\mathbf{D}_n}{n}\right)\right\} \\
+ \quad & E\left\{\left(n_{m,n}-\frac{n}{K}\right)p_j\left(\frac{\mathbf{D}_n}{n}\right)\right\} \\
- \quad & \left(1-\frac{1}{K}\right)E\left\{\left(n_{m,n}-\frac{n}{K}\right)p_j\left(\frac{\mathbf{D}_n}{n}\right)p_m\left(\frac{\mathbf{D}_n}{n}\right)\right\}.
\end{aligned}
$$

We know that, for any treatment j, $0 \leq p_j\left(\mathbf{D}_n/n\right) \leq 1$. The covariance of $L_{j,n+1}$ and $L_{m,n+1}$ is $\mathrm{Cov}(L_{j,n+1}, L_{m,n+1}) = -(n+1)/K^2 = \mathrm{Cov}(L_{j,n}, L_{m,n}) - 1/K^2$. It is clear that, for Wei's class of biased coin designs, if $n_{j,n}-n/K = 0$, then $p_j(\mathbf{D}_n/n) = 1/K$. If $n_{j,n}-n/K > 0$, then $p_j(\mathbf{D}_n/n) < 1/K$, and, if $n_{j,n}-n/K < 0$, then $p_j(\mathbf{D}_n/n) > 1/K$. The two random variables $p_j(\mathbf{D}_n/n)$ and $n_{j,n}-n/K$ are negatively correlated. We know that

$$
\mathrm{Cov}\left\{\left(n_{j,n}-\frac{n}{K}\right), p_j\left(\frac{\mathbf{D}_n}{n}\right)\right\} = E\left\{\left(n_{j,n}-\frac{n}{K}\right)p_j\left(\frac{\mathbf{D}_n}{n}\right)\right\} \leq 0.
$$

We conjecture that

$$
E\left\{\left(n_{j,n}-\frac{n}{K}\right)\left(n_{m,n}-\frac{n}{K}\right)p_j\left(\frac{\mathbf{D}_n}{n}\right)p_m\left(\frac{\mathbf{D}_n}{n}\right)\right\} \geq 0.
$$

We conjecture that

$$\mathrm{Cov}(n_{j,n+1}, n_{m,n+1}) \leq \mathrm{Cov}(L_{j,n+1}, n_{m,n+1}).$$

4.4 Asymptotic properties of randomization schemes

One of the main aims of randomization schemes in clinical trials is to achieve balance in the numbers of patients across treatment groups. It is clear that, for any treatment $j = 1, ..., K$, $L_{j,n}/n \to 1/K$ almost surely under complete randomization and $n_{j,n}/n \to 1/K$ almost surely under the biased coin design as $n \to \infty$.

4.4.1 Complete randomization

By the central limit theorem

$$\sqrt{n}\left[\left(\frac{L_{1,n}}{n}, ..., \frac{L_{K,n}}{n}\right)^T - \left(\frac{1}{K}, ..., \frac{1}{K}\right)^T\right] \to \mathcal{N}_K(\mathbf{0}, \mathbf{\Sigma}^c)$$

in distribution as $n \to \infty$. Here, $\mathcal{N}_K(\mathbf{0}, \mathbf{\Sigma}^c)$ represents a multivariate normal distribution with mean vector a $K \times 1$ vector of zeros and the $K \times K$ covariance matrix Σ^c. For any treatment $j = 1, ..., K$, the variance of $\sqrt{n}(L_{j,n}/n - 1/K)$ is

$$\Sigma^c_{j,j} = \mathrm{Var}\left[\sqrt{n}\left(\frac{L_{j,n}}{n} - \frac{1}{K}\right)\right] = \frac{1}{K}\left(1 - \frac{1}{K}\right),$$

since $L_{j,n}$ has a binomial distribution with parameters n and $1/K$. For any two different treatments j and m, the covariance of $\sqrt{n}(L_{j,n}/n - 1/K)$ and $\sqrt{n}(L_{m,n}/n - 1/K)$ is

$$\Sigma^c_{j,m} = \mathrm{Cov}\left[\sqrt{n}\left(\frac{L_{j,n}}{n} - \frac{1}{K}\right), \sqrt{n}\left(\frac{L_{m,n}}{n} - \frac{1}{K}\right)\right] = -\frac{1}{K^2}.$$

The covariance matrix Σ^c has $\Sigma^c_{j,j}$ as its jth diagonal element and $\Sigma^c_{j,m}$ as element (j, m).

4.4.2 Wei's class of biased coin designs with Smith's approximation

The approximation for the covariance of the numbers of patients on two different treatments was obtained by Smith (1984) and is given in Section 4.2.6. We will only look at a particular stage n. For this class of biased coin designs, we aim to achieve the same limiting proportion of patients on each treatment. Therefore, the limiting proportion is $\xi_j = 1/K$ for $j = 1, ..., K$ and $\rho = \sum_{j=1}^{K} \xi_j = 1$. By the central limit theorem,

$$\sqrt{n}\left[\left(\frac{n_{1,n}}{n}, ..., \frac{n_{K,n}}{n}\right)^T - \left(\frac{1}{K}, ..., \frac{1}{K}\right)^T\right] \to \mathcal{N}_K(\mathbf{0}, \mathbf{\Sigma}^w)$$

in distribution as $n \to \infty$. For any stage n, the covariance matrix Σ^w can be obtained using the approximation obtained by Smith (1984). We have

$$\begin{aligned}
\Sigma^w_{j,j} &= \text{Var}\left[\sqrt{n}\left(\frac{n_{j,n}}{n} - \frac{1}{K}\right)\right] \\
&= \frac{1}{n}(1+2\rho)^{-1} n^{1+\rho} n^{-\rho}(\xi_j - \xi_j^2) \\
&= \frac{1}{3K}\left(1 - \frac{1}{K}\right).
\end{aligned}$$

Similarly, for $j \neq m$,

$$\begin{aligned}
\Sigma^w_{j,m} &= \text{Cov}\left[\sqrt{n}\left(\frac{n_{j,n}}{n} - \frac{1}{K}\right), \sqrt{n}\left(\frac{n_{m,n}}{n} - \frac{1}{K}\right)\right] \\
&= -\frac{1}{n}(1+2\rho)^{-1} n^{1+\rho} n^{-\rho} \xi_j \xi_m \\
&= -\frac{1}{3K^2}.
\end{aligned}$$

The $K \times K$ covariance matrix has $\Sigma^w_{j,j}$ as its jth diagonal element and $\Sigma^w_{j,m}$ as element (j, m). It is clear that, with Smith's approximation, both the variance and the covariance are one-third of their respective values under complete randomization.

4.4.3 D_A-optimum biased coin design

In the case where there are only two treatments, Smith (1984) studied Wei's (1978) class of designs when $f : [-1, 1] \to [0, 1]$ is the conditional probability that the next

patient is allocated to treatment 1 given the numbers of patients n_1 and n_2 on treatments 1 and 2, respectively. This function is differentiable at 0 with special case

$$p(n_1, n_2) = f_t(x) = \frac{(1-x)^t}{(1-x)^t + (1+x)^t},$$

where

$$x = \frac{n_1 - n_2}{n_1 + n_2}.$$

When $t = 1$ or 2, this is the same as Atkinson's (1982) designs. The asymptotic properties of Smith's class of designs apply to Atkinson's designs when there are two treatments.

Smith (1984) showed that, for $K \geq 2$, the probabilities p_j in Atkinson's D_A-optimum biased coin design satisfy the two conditions in Section 4.2.6 with $\xi_j = 1/K$ for all treatments $j = 1, ..., K$. So, by the central limit theorem,

$$\sqrt{n}\left[\left(\frac{n_{1,n}}{n}, ..., \frac{n_{K,n}}{n}\right)^T - \left(\frac{1}{K}, ..., \frac{1}{K}\right)^T\right] \rightarrow \mathcal{N}_K(\mathbf{0}, \mathbf{\Sigma}^w)$$

in distribution as $n \rightarrow \infty$.

4.4.4 Efron's biased coin design

For Efron's biased coin design, the assignment is made based on the numbers of patients on the treatments from the previous stage. There is a fixed probability $p > 1/K$ of allocating the next patient to the treatment that has been chosen least often. This probability is a fixed constant instead of a continuous function used in Wei's class of biased coin designs.

In Hu, Zhang and He (2009), Efron's biased coin design is a special case of the efficient randomized-adaptive design which is a family of response-adaptive designs and asymptotically attains a lower bound for the variance of the allocation proportion. Hu, Rosenberger and Zhang (2006) state that any procedure that attains the lower bound is

asymptotically best. For $K > 2$ treatments, the asymptotic properties of the response-adaptive design are given. The regularity conditions are as follows.

- The parameter space $\boldsymbol{\Theta}_j$ is an open subset of $\mathbb{R}^d$, $d \geq 1$, for $j = 1, ..., K$.

- The limiting proportion of patients on treatment j for $j = 1, ..., K$ is $\rho_j(\boldsymbol{\theta})$ and $\rho(\boldsymbol{\theta}) = (\rho_1(\boldsymbol{\theta}), ..., \rho_K(\boldsymbol{\theta}))^T \in (0,1)^K$, where

$$\frac{n_{j,n}}{n} \to \rho_j(\boldsymbol{\theta})$$

  almost surely as $n \to \infty$.

- For some positive definite matrix $V(\boldsymbol{\theta})$,

$$\sqrt{n}\left[\left(\frac{n_{1,n}}{n}, ..., \frac{n_{K,n}}{n}\right)^T - (\rho_1(\boldsymbol{\theta}), ..., \rho_K(\boldsymbol{\theta}))^T\right] \to \mathcal{N}_K(\mathbf{0}, V(\boldsymbol{\theta}))$$

  in distribution as $n \to \infty$.

A procedure is said to be asymptotically best if $V(\boldsymbol{\theta})$ attains the lower bound

$$B(\boldsymbol{\theta}) = \nabla\rho(\boldsymbol{\theta})\mathbf{I}^{-1}(\boldsymbol{\theta})\nabla\rho(\boldsymbol{\theta})^T,$$

where $\mathbf{I}(\boldsymbol{\theta}) = \text{diag}\{\rho_1(\boldsymbol{\theta})I_1(\theta_1), ..., \rho_K(\boldsymbol{\theta})I_K(\theta_K)\}$ and $I_j(\theta_j)$ is the information matrix for treatment j. The form of $V(\boldsymbol{\theta})$ is

$$V(\boldsymbol{\theta}) = \frac{1}{1+2\gamma}\Sigma_1(\boldsymbol{\theta}) + \frac{2(1+\gamma)}{1+2\gamma}B(\boldsymbol{\theta}),$$

when the above conditions hold, $\Sigma_1(\boldsymbol{\theta}) = \text{diag}\{\rho(\boldsymbol{\theta})\} - \rho(\boldsymbol{\theta})\rho(\boldsymbol{\theta})^T$ and γ is some non-negative integer. When $\gamma = 0$, the procedure becomes deterministic. The larger the value of γ, the greater the randomness in the design. As $\gamma \to \infty$, $1/(1+2\gamma) \to 0$ and $2(1+\gamma)/(1+2\gamma) \to 1$, so that $V(\boldsymbol{\theta})$ attains the lower bound $B(\boldsymbol{\theta})$. In other words, as $\gamma \to \infty$, the response-adaptive design is asymptotically the best.

Under Efron's biased coin design for $K > 2$ treatments, the limiting proportion of patients on treatment j is $\rho_j = 1/K$ for $j = 1, ..., K$, so that $\rho(\boldsymbol{\theta}) = (1/K, ..., 1/K)^T$.

As $\rho(\boldsymbol{\theta})$ is a vector of constants, the lower bound $B(\boldsymbol{\theta})$ is 0 and $V(\boldsymbol{\theta})$ attains this lower bound. Therefore,

$$\sqrt{n}\left[\left(\frac{n_{1,n}}{n},...,\frac{n_{K,n}}{n}\right)^T-\left(\frac{1}{K},...,\frac{1}{K}\right)^T\right]\rightarrow(0,...,0)^T$$

in probability as $n\rightarrow\infty$.

4.5 Imbalance for different randomization schemes

In this section, we will study the imbalance in the numbers of patients on the treatments for four different randomization schemes, complete randomization, Efron's biased coin design, D_A-optimal randomization and our new class of randomization schemes, the adjustable biased coin design. It is always most efficient to test for treatment differences using balanced treatment groups when the variances of patients' responses on different treatments are the same. Imbalances reduce the power of the test.

We define the overall imbalance across treatments to be a vector which contains the imbalances Δ_j for $j=1,...,K$, which are the imbalances obtained at stage n. In other words, we have $\Delta_j=\Delta_{j,n}$ for $j=1,...,K$ in the vector. Denote this vector by $\boldsymbol{\Delta}=(\Delta_1,\Delta_2,...,\Delta_K)^T$. From Lemma 2 of Wei (1978), we have

$$\frac{\boldsymbol{\Delta}}{n}\rightarrow 0$$

in probability as $n\rightarrow\infty$. This result holds for each of the four randomization schemes.

We are interested in the asymptotic distribution and the properties of the imbalances under these designs. For any particular treatment j, $\sqrt{n}(n_{j,n}/n-1/K)$ can be written as $\sqrt{n}(\Delta_j/n)$. Hence, from Section 4.4,

$$\sqrt{n}\left(\frac{\boldsymbol{\Delta}}{n}\right)\rightarrow\mathcal{N}_K(\mathbf{0},\boldsymbol{\Sigma}^c)$$

in distribution as $n \to \infty$ under complete randomization. As the D_A-optimum biased coin design is a special case of Wei's class of designs, from Smith's approximation,

$$\sqrt{n}\left(\frac{\boldsymbol{\Delta}}{n}\right) \to \mathcal{N}_K(\mathbf{0}, \boldsymbol{\Sigma}^w)$$

in distribution as $n \to \infty$ under D_A-optimum biased coin design. Finally,

$$\sqrt{n}\left(\frac{\boldsymbol{\Delta}}{n}\right) \to (0, ..., 0)^T$$

in probability as $n \to \infty$ for Efron's biased coin design.

4.5.1 Simulations for the imbalances

We will now study the imbalance properties of the different randomization schemes by simulation. First, the values for the imbalances under these randomization schemes are produced. Then the properties of the imbalances will be shown using their quartiles and the spike plot for one of the K treatments.

In the simulations, assume that there are 60 patients in the trial. They are allocated to one of the treatments upon arrival according to the assignment rules under complete randomization, Efron's biased coin design, the D_A- optimum biased coin design and the new class of designs, the adjustable biased coin design. We took $K = 3$ and $K = 4$. The trial is simulated $10,000$ times under each scheme. When each patient arrives, we generate a random number for this patient from a uniform distribution between 0 and 1.

Complete randomization

Under complete randomization, each patient is equally likely to be allocated to one of the K treatments. We divide the real line between 0 and 1 into K equal intervals and identify the interval in which the simulated values lies. When the simulated value is greater than or equal to 0 and less than $1/K$, we allocate this patient to treatment 1.

If the value is greater than or equal to $1/K$ and less than $2/K$, the patient is allocated to treatment 2, and similarly for all other treatments, until an assignment is made. The imbalances across all treatments will be calculated after each assignment. At the end of a trial, the imbalance $\mathbf{\Delta}$ is obtained by computing the imbalances after the last assignment for all treatments $j = 1, ..., K$.

Efron's biased coin design

For Efron's biased coin design, we fix the probability $p > 1/K$ of allocating the treatment that has been chosen least often. We considered the values for p which are $8/12, 9/12, 10/12$ and $11/12$. As p increases, the assignment becomes more deterministic. The trial starts off with complete randomization.

The process of treatment assignment of the first patient is the same as that described for complete randomization. The first patient is assigned to treatment 1 if the simulated value is greater than or equal to 0 and less than $1/K$, to treatment 2 if the simulated value is greater than or equal to $1/K$ and less than $2/K$, and so on for all other treatments until an assignment is made. From the second patient onwards, the current numbers of patients on the treatments are noted. The treatments that have the least number of patients are identified. Let the number of treatments that have the least number of patients at the current stage be n^* for $n^* = 1, ..., K$. At any particular stage, if there is only one treatment that has the least number of patients, we will allocate the next patient to this treatment if the simulated value is less than p. Otherwise, for all other $K - 1$ treatments, the assignment is to the first of these treatments if the simulated value is greater than or equal to p and less than $p+(1-p)/(K-1)$, to the second of these treatments if the value is greater than or equal to $p + (1 - p)/(K - 1)$ and less than $p + 2(1 - p)/(K - 1)$, and so on. When there are more than two treatments that have the least number of patients,

the next patient will be allocated to the first of these treatments if the simulated value is greater than or equal to zero and less than p/n^*, to the second of these if the value is greater than or equal to p/n^* and less than $2p/n^*$, and so on. If the simulated value is greater than p, the patient will be allocated to the group of treatments that do not have the least number of patients at the current stage. The patient will be allocated to the first of these if the simulated value is greater than or equal to $p + (1-p)/(K-n^*)$ and less than $p + 2(1-p)/(K-n^*)$, to the second of these if the value is greater than or equal to $p + 2(1-p)/(K-n^*)$ and less than $p + 3(1-p)/(K-n^*)$, and similarly for the rest of the intervals until an assignment is made. After all assignments are made at the end of the trial, the vector $\mathbf{\Delta}$ is obtained.

D_A-optimum biased coin design

For the D_A-optimum biased coin design, the probability p_j of assigning the next patient to treatment j is calculated at each stage based on the current number of patients on the jth treatment for $j = 1, ..., K$. For the initial K patients, we will allocate the first patient to treatment 1, the second patient to treatment 2 and so on. We then have one patient on each of the K treatments before the assignment rules are applied. For the $(K+1)$st patient onwards, a simulated value is generated for each patient. For the $(K+1)$st patient, we have $p_j = 1/K$ for all $j = 1, ..., K$. The $(K+1)$st patient will be allocated to treatment 1 if its simulated value is greater than or equal to 0 and less than $1/K$, to treatment 2 if the simulated value is greater than or equal to $1/K$ and less than $2/K$, and so on for all other treatments until a treatment assignment is made. For the $(K+2)$nd patient onwards, the probabilities p_j obtained for each patient are sorted into ascending order. Some of these probabilities may be the same. The next patient will be allocated to the treatment with the smallest p_j if the simulated value is greater than or

equal to zero and less than p_j. For the next largest p_l, treatment $l \neq j$ will be assigned to the next patient if the simulated value is greater than or equal to p_j and less than $p_j + p_l$, and similarly for the other probabilities. At the end of the trial, we will obtain the vector of imbalances under this scheme.

The adjustable biased coin design

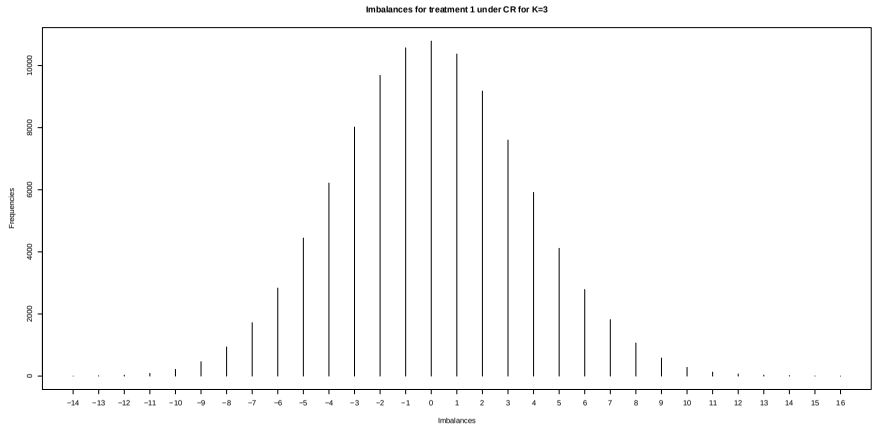
Finally, we will describe the new class of designs proposed, the adjustable biased coin design. The probability p_j of assigning the next patient to treatment j for $j = 1, ..., K$ is calculated at each stage based on the current imbalance across the treatments. For each trial, the values for the imbalances on the treatments at each stage are needed to obtain p_j for $j = 1, ..., K$. The first patient under this scheme is equally likely to be assigned to one of the K treatments with $p_j = 1/K$ for all $j = 1, ..., K$. For the second patient, we need to obtain $\Delta_{j,1}$, the imbalances on treatment j for $j = 1, .., K$. Then values for $F_j(z)$ for all j can be obtained and hence the p_j for the treatment assignment of the second patient. Once the values for p_j are known, they are put into ascending order. The treatment assignment process is then the same as that described for the D_A-optimum biased coin design. The next patient will be allocated to the treatment with the smallest p_j if the simulated value for this patient is greater than or equal to 0 and less than p_j, and to the treatment with the next smallest p_l if the simulated value is greater than or equal to p_j and less than $p_j + p_l$. The process continues for the other probabilities until a treatment is assigned. This process of assignment applies to all other patients until the assignment is made for the last patient. Although the vector of imbalances $\mathbf{\Delta}$ is obtained here at each stage, we will only study the vector of imbalances obtained at the end of the trial.

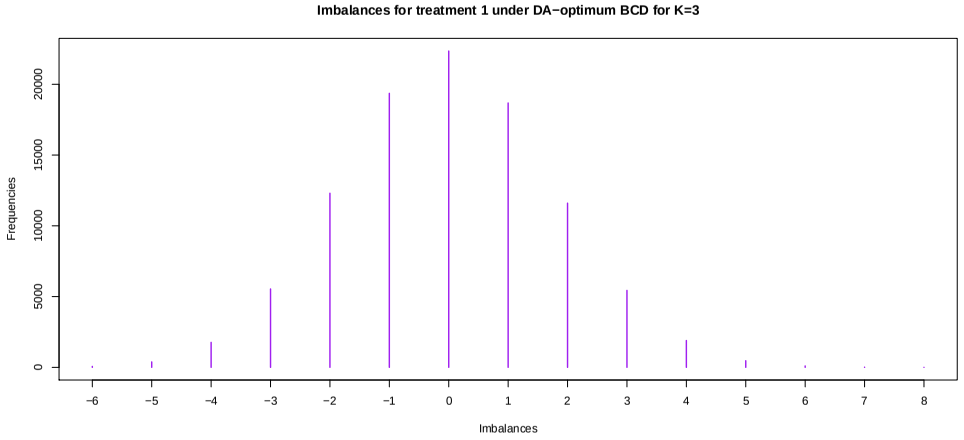
After $10,000$ trials under each randomization scheme are simulated, the values for

the $10,000$ vectors of imbalances $\mathbf{\Delta}$ are recorded and the imbalances obtained for each of the treatments are plotted. Here, we show the plots for the first treatment under each randomization scheme when $K = 3$ and $K = 4$. The rest of the plots for the other treatments will be given in Appendix B.

When $K = 3$, the plots under these four randomization schemes for treatment 1 are given. Different values of p which are $8/12, 9/12, 10/12$ and $11/12$ are considered under Efron's biased coin design. Then, we show the plots of the imbalances for treatment 1 when $K = 4$.

Figure 4.1: Plots of imbalances in treatment 1 under complete randomization, the D_A-optimum biased coin design and the adjustable biased coin design when $K = 3$





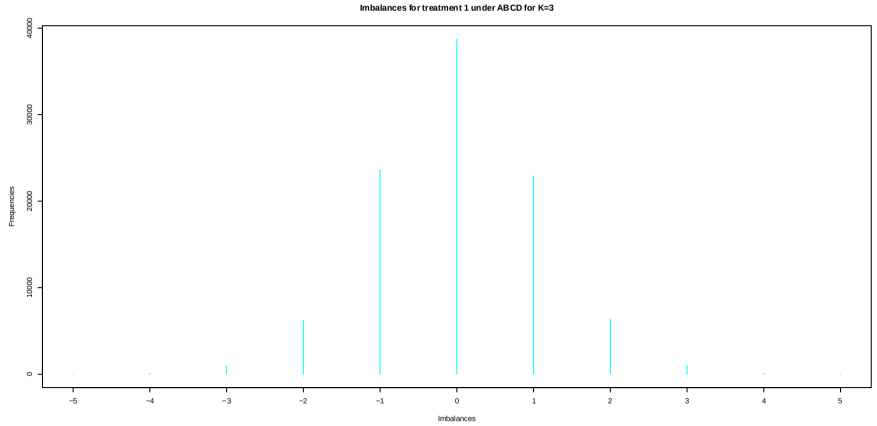


Figure 4.2: Plots of imbalances in treatment 1 under Efron's biased coin design with $p = 8/12, 9/12, 10/12$ and $11/12$ when $K = 3$

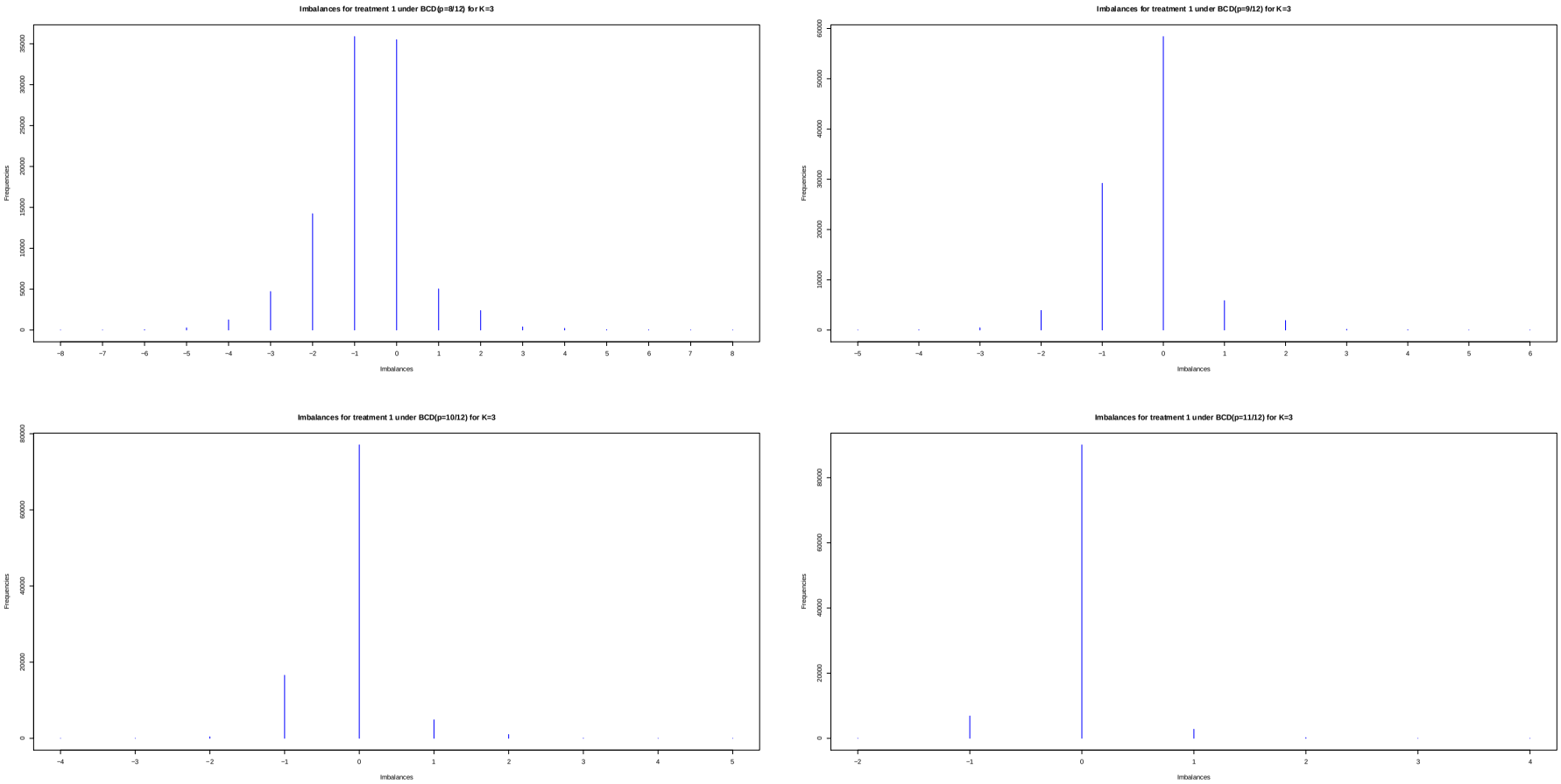
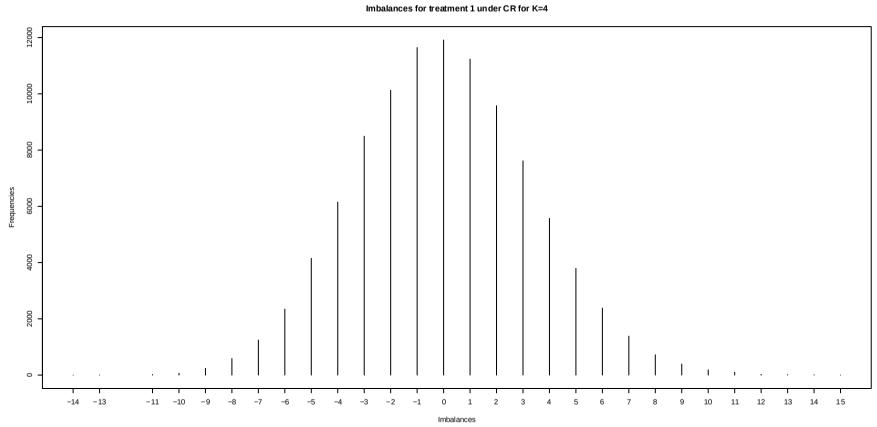
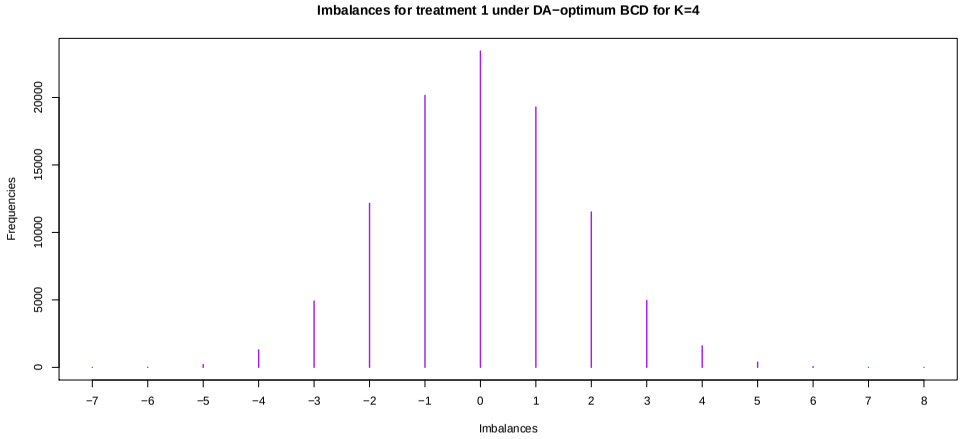


Figure 4.3: Plots of imbalances in treatment 1 under complete randomization, the D_A-optimum biased coin design and the adjustable biased coin design when $K = 4$





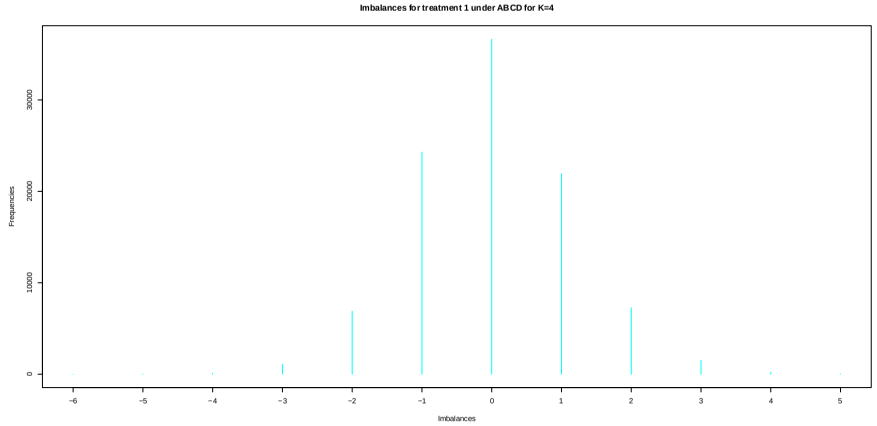
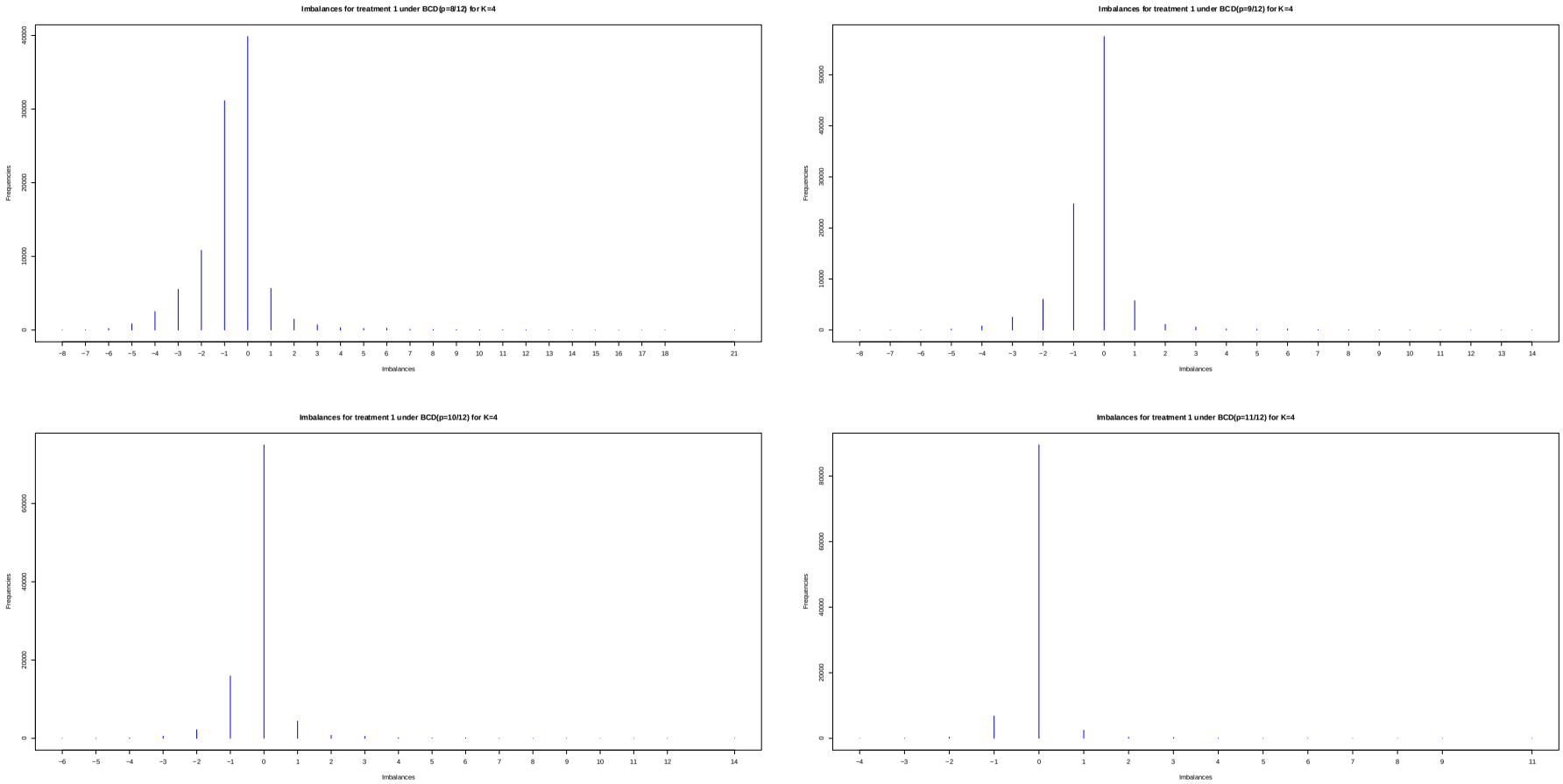


Figure 4.4: Plots of imbalances in treatment 1 under Efron's biased coin design with $p = 8/12, 9/12, 10/12$ and $11/12$ when $K = 4$



We also give the numerical values for the quartiles of the imbalances on treatment 1 under the four randomization schemes for the above two scenarios. Here, CR represents complete randomization, $\mathrm{D_A}$ represents the D_A-optimum biased coin design, ABCD represents the adjustable biased coin design and BCD for the Efron's biased coin design. Also $\min$ denotes the minimum value and $\max$ the maximum value. The numerical values of the quartiles for treatments 2, 3 and 4 for both scenarios are given in Appendix B.

Table 4.1: Numerical values of the quartiles of imbalance under all schemes for $K = 3$ for treatment 1

	CR	$\mathrm{D_A}$	ABCD	BCD			
				$p = 8/12$	$p = 9/12$	$p = 10/12$	$p = 11/12$
min	-14	-6	-5	-8	-5	-4	-2
Q1	-2	-1	-1	-1	-1	0	0
Q2	0	0	0	-1	0	0	0
mean	-0.0067	-0.0112	-0.0027	-0.7298	-0.2821	-0.1041	-0.03632
Q3	2	1	1	0	0	0	0
max	16	8	5	8	6	5	4

Table 4.2: Numerical values of the quartiles of imbalance under all schemes for $K = 4$ for treatment 1

	CR	D_A	ABCD	BCD			
				$p = 8/12$	$p = 9/12$	$p = 10/12$	$p = 11/12$
min	-14	-7	-6	-8	-8	-6	-4
Q1	-2	-1	-1	-1	-1	0	0
Q2	0	0	0	-1	0	0	0
mean	-0.0188	0.0034	0.0018	-0.6709	-0.3430	-0.1295	-0.0316
Q3	2	1	1	0	0	0	0
max	15	8	5	21	14	14	11

From the values of the quartiles and the shapes of the plots above, we can see that these confirm the theoretical results that we stated in Section 4.4. The range of the imbalances is largest for complete randomization and smallest for the D_A-optimum biased coin design and the adjustable biased coin design. This means that complete randomization is the most variable randomization scheme. For complete randomization, the D_A-optimum biased coin design and the adjustable biased coin design, the imbalances are approximately normally distributed. The values for the imbalances under these three randomization schemes for each of the treatments have peaks at 0. These results suggest that each of the vectors of imbalances $\mathbf{\Delta}$ under these two schemes are multivariate normal with mean vector $(0, ..., 0)^T$, which confirms the theoretical results for complete randomization and Wei's class of designs.

For Efron's biased coin design, the increase in p from $8/12$ to $11/12$ means that the design becomes more deterministic. The plots for this design show that the imbalances on each of the treatments in both scenarios have sharp peaks at 0 and do not follow a nor-

mal distribution. These results suggest that $\mathbf{\Delta}$ for Efron's biased coin design converges to $(0, ..., 0)^T$ in probability and confirm the theoretical results.

The idea of this design is to give a greater probability of assignment to the treatment that has been chosen less often. Therefore, we have to choose the appropriate p in order to balance the numbers of patients across treatments. For example, $p = 8/12$ may not be a good choice. For $K = 3$, when there are two treatments that have the least number of patients, then we have a probability $1/2 \times 8/12 = 1/3$ of allocating the next patient to either of these treatments and $1/3$ to the treatment that has the largest number of patients. The probabilities are the same and the design at this stage becomes complete randomization. For $K = 4$, when three of the four treatments have the least number of patients, the probability $p = 8/12$ will have to be divided by three to give $2/9$ as the probability of assigning each of these three treatments. The treatment with the largest number of patients has a probability $1/3$ of being allocated to the next patient. Since $1/3$ is greater than $2/9$, this invalidates the idea of Efron's biased coin design in balancing the numbers of patients across the treatments. As a result, there are a few extreme values for the imbalances. These values are either too large or too small, particularly when p is small. Although the plots have peaks at zero, the median and mean will be affected by these few extreme values. We can see that the range of the values is becoming smaller as p increases. Therefore, the value for p has to be chosen carefully according to the number of treatments involved in the trial.

For the adjustable biased coin design, the numerical values of the quartiles for different treatments show that the range of the imbalances is the smallest compared to the other designs for both scenarios. In other words, the variability under this design is the smallest. The shapes of all of the plots suggest that the imbalances on each of the treatments have a normal distribution with mean 0. This may indicate that, under the ad-

justable biased coin design, the vector of imbalances has asymptotically a multivariate normal distribution with mean vector $(0, ..., 0)^T$.

4.6 Power under different randomization schemes

Let μ_j be the mean response on treatment j for $j = 1, ..., K$. We have the null hypothesis $H_0 : \mu_1 = \mu_2 = ... = \mu_K$ of no treatment differences and the alternative hypothesis H_1 : at least two of the mean responses are different. We test these hypotheses by constructing the analysis of variance table. It is assumed that the responses on treatment j are normally distributed with variance σ^2, all independent.

With the assignment rules at each stage for different randomization schemes, we will know the number of patients on each of the treatments at the end of the trial. Let n be the total number of patients at the end of the trial with n_j the number of patients on treatment j for $j = 1, ..., K$. Also, let y_{ij} be the response of the ith patient on treatment j, so that we have the treatment total $T_j = \sum_{i=1}^{n_j} y_{ij}$ for the jth treatment. After obtaining K treatment totals, we calculate the grand total $G = \sum_{j=1}^{K} T_j$ and the correction factor G^2/n. Next, the treatment sum of squares is

$$S_T = \sum_{j=1}^{K} \frac{T_j^2}{n_j} - \frac{G^2}{n}.$$

Then we have the total sum of squares given by

$$S_G = \sum_{j=1}^{K} \sum_{i=1}^{n_j} y_{ij}^2 - \frac{G^2}{n}.$$

Finally, the residual sum of squares is

$$S_R = S_G - S_T.$$

The analysis of variance table can now be constructed. The degrees of freedom for the treatment, residual and total sums of squares are $K-1$, $n-K$ and $n-1$, respectively.

Each of their mean squares are then defined as their sums of squares divided by their corresponding degrees of freedom. The mean treatment sum of squares is

$$M_T = \frac{S_T}{K-1}$$

and the mean residual sum of squares is

$$M_R = \frac{S_R}{n-K}.$$

An F test is carried out to test whether there is a difference between the μ_j. The F statistic is

$$F = \frac{M_T}{M_R}.$$

Under H_0 given the number of patient on treatment j for $j = 1, ..., K$, F has an F distribution with $K-1$ and $n-K$ degrees of freedom. We reject H_0 at the $100\alpha\%$ level if $F > F_{K-1,n-K,\alpha}$, where $F_{K-1,n-K,\alpha}$ is the upper $100\alpha\%$ value of the $F_{K-1,n-K}$ distribution. The power given the n_j, or, in other words, the conditional power of the test, is the probability of rejecting H_0 given that there is a genuine treatment difference, that is, when H_1 is true.

We will study the power of the test under the four randomization schemes by simulation. We will again consider the two scenarios $K = 3$ and $K = 4$. Consider a trial where patients are allocated to treatments according to the assignment rules for the randomization schemes. The numbers of patients on the treatments are then recorded at the end of the trial. Next we generate the patients responses for the test. In the simulations, n_j responses will be generated from a normal distribution with mean μ_j and variance σ^2 for $j = 1, ...K$. We took $\sigma^2 = 1$. Various values for the mean responses on different treatments will be considered. The trial is simulated $10,000$ times and the number of rejections of H_0 is counted. The proportion of rejections of H_0 will be the

estimated power for the test for different randomization schemes. This estimated power is the unconditional power.

For simplicity, we assume that the value for μ_1 is always 0. For μ_2, we take nine values from 0 to 2 with increments of 0.25. We take 0 together with eight values from 0.5 to 2.25 with increments of 0.25 for μ_3. Finally, we have 0 and eight values from 0.75 to 2.5 with increments of 0.25 for μ_4. The values for μ_1, μ_2 and μ_3 are the same in both scenarios. Let $d_j = \mu_j - \mu_1$ for $j = 2, ..., K$. Then the numerical values for the power will be given under the four randomization schemes for different values of d_2 and d_3 when $K = 3$ and d_2, d_3 and d_4 when $K = 4$. Let $\alpha = 0.05$ the significance level of the test.

Table 4.3: Powers under four randomization schemes for $K = 3$

	$d_2 = 0$	$d_2 = 0.25$	$d_2 = 0.5$	$d_2 = 0.75$	$d_2 = 1$	$d_2 = 1.25$	$d_2 = 1.5$
	$d_3 = 0$	$d_3 = 0.5$	$d_3 = 0.75$	$d_3 = 1$	$d_3 = 1.25$	$d_3 = 1.5$	$d_3 = 1.75$
CR	0.050	0.258	0.540	0.815	0.957	0.994	1
BCD(p=8/12)	0.050	0.261	0.544	0.824	0.961	0.995	1
BCD(p=9/12)	0.051	0.262	0.548	0.825	0.962	0.996	1
BCD(p=10/12)	0.050	0.262	0.547	0.824	0.963	0.996	1
BCD(p=11/12)	0.050	0.263	0.547	0.827	0.962	0.996	1
$\mathrm{D_A}$	0.050	0.262	0.542	0.824	0.962	0.995	1
ABCD	0.050	0.260	0.545	0.825	0.963	0.996	1

Table 4.4: Powers under four randomization schemes for $K = 4$

	$d_2 = 0$	$d_2 = 0.25$	$d_2 = 0.5$	$d_2 = 0.75$	$d_2 = 1$	$d_2 = 1.25$	$d_2 = 1.5$
	$d_3 = 0$	$d_3 = 0.5$	$d_3 = 0.75$	$d_3 = 1$	$d_3 = 1.25$	$d_3 = 1.5$	$d_3 = 1.75$
	$d_4 = 0$	$d_4 = 0.75$	$d_4 = 1$	$d_4 = 1.25$	$d_4 = 1.5$	$d_4 = 1.75$	$d_4 = 2$
CR	0.050	0.384	0.621	0.833	0.950	0.989	0.998
BCD(p=8/12)	0.050	0.393	0.633	0.845	0.957	0.993	0.999
BCD(p=9/12)	0.050	0.393	0.630	0.846	0.958	0.993	0.999
BCD(p=10/12)	0.052	0.391	0.634	0.846	0.960	0.993	1
BCD(p=11/12)	0.051	0.392	0.632	0.849	0.960	0.994	0.999
D_A	0.050	0.387	0.629	0.847	0.958	0.993	0.999
ABCD	0.051	0.391	0.629	0.845	0.960	0.993	0.999

The first column in the two tables represents the significance level of the simulated test. The values are very close to $\alpha = 0.05$, which is the assumed significance level of the test. We can see that the values for the powers increase when the differences in the mean treatment responses increase. The powers obtained under complete randomization are the lowest among the randomization schemes in both scenarios. The power under Efron's biased coin design usually increases as p increases. In addition, for both scenarios, when Efron's biased coin design is very deterministic with $p = 10/12$ or $p = 11/12$, the power obtained is higher than that under the adjustable biased coin design. For $K = 3$, the D_A-optimum biased coin design sometimes achieves a higher power than Efron's biased coin design with $p = 8/12$, and $p = 9/12$.

For each randomization scheme, the powers achieved when $K = 3$ are lower than those obtained when $K = 4$. When more treatments are involved in the trial, the probability of detecting a treatment effect under different randomization schemes is higher. When $K = 4$, the powers under the D_A-optimum biased coin design is lower than that under Efron's biased coin design for most values of p. In general, we can say that the

adjustable biased coin design is very efficient regardless of the number of treatments involved and the values for the treatment differences.

4.7 Conclusions

In the literature, randomization schemes for more than two treatments have not gained great attention over the years and there are only limited theoretical results available about these randomization schemes. In this chapter, a new class of designs called the adjustable biased coin design is proposed and the results given show that this design is generally more efficient at balancing the numbers of patients across the treatments than complete randomization, Efron's biased coin design when $p = 8/12$ and $p = 9/12$, and the D_A-optimum biased coin design.

The adjustable biased coin design is the only design which uses the imbalances instead of the numbers of patients on the treatments at each stage. The advantage of this design is that the imbalances are taken into account in the calculation of the assignment probabilities. Furthermore, under this design, the probability of assigning treatment j to the next patient not only depends on the current imbalance on treatment j, but also on all current imbalances on the other treatments.

The results from simulations indicate that the vector of imbalances under the adjustable biased coin design has a multivariate normal distribution asymptotically with mean vector $(0, ..., 0)^T$. The form of the covariance matrix is not known. Further work can be carried out on the theoretical properties of the imbalance and power under this design. The rate of convergence of the imbalance and the structure of the covariance matrix are of particular interest.

Chapter 5

Imbalance properties of centre-stratified permuted-block randomization and complete randomization for several treatments in clinical trials

5.1 Introduction

Randomization schemes used in clinical trials are considered as essential and of great importance to maintain a balance in the numbers of patients across treatment groups and to gain some randomness in assigning a treatment to a patient to avoid any selection or accidental bias. A trial having similar numbers of patients across treatment groups is better for comparison purposes. Statistical inference based on an equal number of patients in each treatment group is the most efficient method to detect a genuine treatment effect

when the variances of the patients' responses across treatments are the same. However, imbalance in assignments will still occur, even in well-defined randomization schemes. The problem of imbalance in the numbers of patients across treatment groups is more serious in clinical trials where fewer patients are recruited and involved in the study. An imbalance in the numbers of patients in the treatment groups will decrease the power, the probability of correctly detecting a genuine treatment effect (McPherson, Campbell and Elbourne, 2012).

Under complete randomization for two-treatment assignments, each patient is equally likely to be assigned to either of the treatments. The number of patients in any of the two treatment groups will follow a binomial distribution with parameters the sample size and the probability of $1/2$ to assign either treatment to a patient. Complete randomization is very likely to produce very serious imbalance and have most of the patients assigned to one treatment group. It is known that imbalance caused by complete randomization is unlikely to occur in a large trial when the number of patients involved is greater than 200.

Another randomization scheme that is used in clinical trials by pharmaceutical companies is randomly permuted-block randomization. There are two basic schemes: unstratified and centre-stratified permuted-block randomization. Given an assigned proportion of patients for each treatment within a block, a sequence of treatment assignments is generated by randomly listing all of the possibilities for different permuted blocks. Patients are then assigned to the treatment according to this sequence in order upon their arrival. Here, unequal allocations are allowed within a block. However, in order to achieve a balance in the numbers of patients across treatment groups, the treatment allocation ratio within a block is usually assumed to be the same. Under this assumption, imbalance will not occur in any of the complete blocks, but it may occur in the incom-

plete block. In other words, under the permuted-block design, it is sufficient to study the imbalance in the numbers of patients across treatments by investigating the properties of the incomplete block.

The treatment imbalance properties for different randomization schemes have been investigated by Hallstrom and Davis (1988) for stratified block randomization, Lachin (1988a, 1988b) for complete randomization and urn randomization, and Anisimov (2007, 2010, 2011) for centre-stratified permuted-block randomization. All of the above authors have considered the imbalance in the number of patients for two treatments. This chapter extends work on the analysis of the imbalance properties to more than two treatment groups. In the previous papers, the expectation and the variance of the imbalance are obtained. The distribution of the overall imbalance is approximately normal with the same expectation and variance. In this chapter, the main focus is to investigate the imbalance properties for two randomization schemes, centre-stratified permuted-block randomization and complete randomization when more than two treatments are studied in a clinical trial. The overall imbalance is represented as a vector of imbalances on different treatments rather than a scalar in the two-treatment case, and asymptotically it has a multivariate normal distribution with a vector of means and a covariance matrix.

In Section 5.2, we consider centre-stratified permuted-block randomization and complete randomization for cases when there are more than two treatments. In Section 5.3, the imbalance for both randomization schemes will be defined for a particular treatment in a particular centre and for all centres. The means, variances and covariances of the imbalances within a centre or for all centres will be evaluated when the numbers of patients to recruit in different centres are known.

In Section 5.4, we will consider the case where the numbers of patients recruited in different centres are random variables. This is based on the Poisson-gamma patient

recruitment model developed by Anisimov and Fedorov (2007), where the recruitment process follows a Poisson process with recruitment rates from a gamma distribution. Therefore, the numbers of patients recruited in different centres have beta-binomial distributions. The expectations, variances and covariances of the imbalances within a centre or for all centres will be evaluated. In Section 5.5, the variances of the imbalance for the two randomization schemes are compared. In Section 5.6, numerical values are simulated for the imbalances on treatments for two particular scenarios. The values for the expectations and the covariance matrices are given together with histograms of the imbalances, which confirm the theoretical results produced in Section 5.4. In Section 5.7, a test will be described for all pairs of treatment differences with the control group and hence how the power can be obtained from this test. Numerical results for the power and the sample size are also given by simulation for different scenarios. By fixing a particular level of power to be achieved for the balanced case, the sample size can be found. The same sample size will be used to study the power in the imbalanced case. In addition, the number of patients that need to be added in each scenario to achieve the same level of power in the balanced case is given. Finally, conclusions will be drawn in Section 5.8.

5.2 Randomization schemes for more than two treatments

Consider a multi-centre trial study, where in total n patients have to be recruited by N clinical centres. Patients have to be assigned to one of the $K \geq 3$ treatments. Let n_i be the number of patients who have to be recruited in centre i. Then $\sum_{i=1}^{N} n_i = n$.

There are two types of permuted-block randomization schemes that are commonly used in clinical trials: unstratified permuted-block randomization and centre-stratified permuted-block randomization. Unstratified randomization means that patients are ran-

domized to treatment according to independent randomly permuted blocks of fixed size without regard to centres. Centre-stratified permuted-block randomization means that each of the centres has a separate permuted-block randomization scheme. Patients for the study are randomized to treatments according to the independent randomly permuted blocks of fixed size within each centre.

For example, suppose that there are three treatments a, b and c, the size of the block is 3 and the ratio within blocks is $1 : 1 : 1$. Then there will be six possibilities for the different permuted blocks: $\{(a, b, c), (a, c, b), (b, a, c), (b, c, a), (c, a, b), (c, b, a)\}$. A randomly chosen sequence of blocks forms a sequence of treatments. The patients are assigned to treatments according to this sequence in the order of their registration.

It is obvious that unstratified permuted-block randomization will minimise the imbalance in the number of patients on different treatments for the whole study and increase the imbalance in each centre compared to centre-stratified permuted-block randomization.

Assume that there are in total $K \geq 3$ treatments with the allocation in the block $(k_1, k_2, ..., k_K)$, where k_j is the number of patients within a block that are allocated to treatment j for $j = 1, ..., K$. Let $K_1 = \sum_{j=1}^{K} k_j$ be the block size.

Now consider centre-stratified permuted-block randomization. The total number of patients recruited n_i in centre i may not be a multiple of the block size K_1. This may lead to incomplete blocks in some of the centres. These incomplete blocks will, however, have the chance to contain an unequal number of patients on each treatment and cause imbalance. In a multi-centre clinical trial, if there exist many incomplete blocks, this will cause serious imbalance in the total numbers of patients on the treatments. Therefore, we can study the imbalance properties of a randomization scheme by studying the properties of these incomplete blocks. The last block can be incomplete. If the size of the last block

is K_1, n_i is a multiple of K_1 and contains no incomplete block for this study. For some centre i, let the size of the incomplete block be r such that $r = 1, ..., K_1 - 1$.

Let $\xi_j(r)$ be the number of patients on treatment j in an incomplete block of size r. We can see that $\xi_j(r)$ has a hypergeometric distribution, that is,

$$\mathbf{P}[\xi_j(r) = l] = \frac{\binom{k_j}{l}\binom{K_1 - k_j}{r-l}}{\binom{K_1}{r}}$$

for $l = 0, 1, ..., \min(k_j, r)$ with

$$E[\xi_j(r)] = \frac{k_j r}{K_1}$$

and

$$\mathrm{Var}[\xi_j(r)] = \frac{k_j r(K_1 - k_j)(K_1 - r)}{K_1^2(K_1 - 1)}.$$

If $r = 0$, we set $\xi_j(r) = 0$.

Let the total number of patients in centre i on treatment j be n_{ij}. Denote by $\lfloor z \rfloor$ the integer part of z and $mod(z, K_1) = z - \lfloor z/K_1 \rfloor K_1$. Then we have

$$n_{ij} = \left\lfloor \frac{n_i}{K_1} \right\rfloor k_j + \xi_j \left[mod(n_i, K_1) \right].$$

The centre-stratified permuted-block design will be studied and referred to as permuted-block randomization in this chapter.

Consider now complete randomization. For ease of comparison, centre-stratified complete randomization is studied and referred to as complete randomization. Each patient is assigned to a treatment j with probability p_j. The total number of patients in centre i on treatment j, n_{ij}, has a binomial distribution with parameters n_i and p_j. In general, p_j is the proportion of patients that are allocated to treatment j in a complete block such that $p_j = k_j/K_1$. For equal treatment proportions within each complete block, we have $k_1 = k_2 = ... = k_j$ with block size $K_1 = Kk_1$. In this case, we have $p_j = 1/K$.

For investigation of the properties of imbalance, we will consider two cases: when the number of patients recruited in a centre is given or is random. Let $\{n_i\}$ be the numbers of patients to be recruited in centre i. Define for centre i the imbalance on treatment j under the permuted block design as

$$\Delta_{ij} = n_{ij} - \frac{n_i}{K_1} k_j \tag{5.1}$$

for $j = 1, ..., K$ and $i = 1, ..., N$. Similarly, the imbalance on treatment j under complete randomization is defined as

$$\Delta^*_{ij} = n_{ij} - \frac{n_i}{K_1} k_j. \tag{5.2}$$

5.3 The numbers of patients recruited in a centre are given

5.3.1 Permuted-block design

First, assume that the $\{n_i\}$ are given. By (5.1), the expectation of Δ_{ij} given n_i under the permuted-block design is

$$E(\Delta_{ij}) = \frac{n_i}{K_1} - \frac{n_i}{K_1} = 0.$$

The variance of Δ_{ij} is

$$\mathrm{Var}(\Delta_{ij}) = E(\Delta^2_{ij}),$$

where

$$E(\Delta^2_{ij}) \quad = \quad E\left(\left\{ \left\lfloor \frac{n_i}{K_1} \right\rfloor k_j + \xi_j[mod(n_i, K_1)] - \frac{n_i}{K_1} k_j \right\}^2 \right).$$

Let $r = mod(n_i, K_1)$ be the size of the incomplete block in this particular centre i. Then

we have

$$\begin{aligned}
\mathrm{Var}(\Delta_{ij}) &= E\left\{\left\lfloor\frac{n_i}{K_1}\right\rfloor k_j\xi_j(r) + [\xi_j(r)]^2 - \frac{n_i}{K_1}k_j\xi_j(r)\right\} \\
&= E\left(-\frac{r^2}{K_1^2}k_j^2\right) + E\left\{[\xi_j(r)]^2\right\} \\
&= -\frac{r^2}{K_1^2}k_j^2 + \frac{rk_j(K_1-k_j)(K_1-r)}{K_1^2(K_1-1)} + \frac{k_j^2r^2}{K_1^2} \\
&= \frac{rk_j(K_1-k_j)(K_1-r)}{K_1^2(K_1-1)}.
\end{aligned}$$

We now consider the covariance of the imbalances on two different treatments j and m in a trial. Let i and l be two different centres. Then, due to the independence of the numbers of patients on a particular treatment j in different centres, the imbalances on a particular treatment j in different centres are also independent, so that $\mathrm{Cov}(\Delta_{ij}, \Delta_{lm}) = 0$. Therefore, we have to calculate the covariance for two different treatments in the same centre i, that is,

$$\mathrm{Cov}(\Delta_{ij}, \Delta_{im}) = E(\Delta_{ij}\Delta_{im}) - E(\Delta_{ij})E(\Delta_{im}) = E(\Delta_{ij}\Delta_{im}).$$

It follows that

$$\begin{aligned}
\mathrm{Cov}(\Delta_{ij}, \Delta_{im}) &= E\left(\left\{\left\lfloor\frac{n_i}{K_1}\right\rfloor k_j + \xi_j[mod(n_i, K_1)] - \frac{n_i}{K_1}k_j\right\}\right. \\
&\quad \left.\times\left\{\left\lfloor\frac{n_i}{K_1}\right\rfloor k_m + \xi_m[mod(n_i, K_1)] - \frac{n_i}{K_1}k_m\right\}\right) \\
&= E\left\{\left\lfloor\frac{n_i}{K_1}\right\rfloor k_j\xi_m[mod(n_i, K_1)]\right. \\
&\quad +\xi_j[mod(n_i, K_1)]\xi_m[mod(n_i, K_1)] \\
&\quad \left.- \frac{n_i}{K_1}k_j\xi_m[mod(n_i, K_1)]\right\}.
\end{aligned}$$

Putting $r = mod(n_i, K_1)$, we have

$$\mathrm{Cov}(\Delta_{ij}, \Delta_{im}) = -\frac{r^2k_jk_m}{K_1^2} + \mathrm{Cov}[\xi_j(r), \xi_m(r)] + E[\xi_j(r)E(\xi_m(r)],$$

where $\mathrm{Cov}[\xi_j(r), \xi_m(r)]$ is the covariance for a multivariate hypergeometric distribution.

Lemma 5.3.1 *The covariance of $\xi_j(r)$ and $\xi_m(r)$ is*

$$\begin{aligned}\mathrm{Cov}[\xi_j(r), \xi_m(r)] &= -\frac{k_j k_m r(K_1 - r)}{K_1^2(K_1 - 1)},\end{aligned}$$

and the covariance of Δ_{ij} and Δ_{im} is equal to the covariance of $\xi_j(r)$ and $\xi_m(r)$.

Proof. See Appendix C. ■

Now consider the imbalance on any particular treatment for all centres. The overall imbalance on a particular treatment j for all centres is defined as

$$\Delta_j = \sum_{i=1}^{N} n_{ij} - \frac{n}{K_1} k_j = \sum_{i=1}^{N} \left(n_{ij} - \frac{n_i}{K_1} k_j \right) = \sum_{i=1}^{N} \Delta_{ij}. \tag{5.3}$$

The expectation of Δ_j is

$$E(\Delta_j) = \sum_{i=1}^{N} E\left(n_{ij} - \frac{n_i}{K_1} k_j \right) = \sum_{i=1}^{N} \left[E(n_{ij}) - \frac{n_i}{K_1} k_j \right] = 0.$$

The variance of Δ_j is

$$\mathrm{Var}(\Delta_j) = \mathrm{Var}\left(\sum_{i=1}^{N} \Delta_{ij} \right) = E\left(\sum_{i=1}^{N} \Delta_{ij}^2 \right),$$

so that

$$\begin{aligned}\mathrm{Var}(\Delta_j) &= \sum_{i=1}^{N} E\left\{ \left\lfloor \frac{n_i}{K_1} \right\rfloor k_j \xi_j[mod(n_i, K_1)] + [\xi_j(mod(n_i, K_1))]^2 \right. \\ &- \left. \frac{n_i}{K_1} k_j \xi_j[mod(n_i, K_1)] \right\}.\end{aligned}$$

Let $r_i = mod(n_i, K_1)$ be the size of the incomplete block in centre i. Then we have

$$\begin{aligned}\mathrm{Var}(\Delta_j) &= \sum_{i=1}^{N} \left\lfloor \frac{n_i}{K_1} \right\rfloor k_j E[\xi_j(r_i)] + \{E[\xi_j(r_i)]\}^2 - \frac{n_i}{K_1} k_j E[\xi_j(r_i)] \\ &= E\left\{ \sum_{i=1}^{N} \frac{r_i k_j (K_1 - k_j)(K_1 - r_i)}{K_1^2(K_1 - 1)} \right\} \\ &= \frac{k_j(K_1 - k_j)}{K_1(K_1 - 1)} \sum_{i=1}^{N} r_i - \frac{k_j(K_1 - k_j)}{K_1^2(K_1 - 1)} \sum_{i=1}^{N} r_i^2. \end{aligned} \tag{5.4}$$

Similarly, the covariance of Δ_j and Δ_m can be written as

$$\begin{aligned}\mathrm{Cov}(\Delta_j, \Delta_m) &= \sum_{i=1}^{N} \left\lfloor \frac{n_i}{K_1} \right\rfloor k_j E[\xi_m(r_i)] + E[\xi_j(r_i)\xi_m(r_i)] - \frac{n_i}{K_1} k_j E[\xi_m(r_i)] \\ &= -\frac{k_j k_m}{K_1^2(K_1-1)} \sum_{i=1}^{N} r_i(K_1 - r_i) \\ &= -\frac{k_j k_m}{K_1(K_1-1)} \sum_{i=1}^{N} r_i + \frac{k_j k_m}{K_1^2(K_1-1)} \sum_{i=1}^{N} r_i^2. \end{aligned} \tag{5.5}$$

5.3.2 Complete randomization

General Case

Now consider another randomization scheme, complete randomization, where the $\{n_i\}$ for different centres are given. Let n_{ij} be the number of patients in centre i assigned to treatment j. Then n_{ij} has a binomial distribution with parameters n_i and k_j/K_1.

The expectation of the imbalance defined in (5.2) is

$$E(\Delta_{ij}^*) = E\left(n_{ij} - \frac{n_i}{K_1} k_j\right) = \frac{n_i}{K_1} k_j - \frac{n_i}{K_1} k_j = 0.$$

The variance is then

$$\mathrm{Var}(\Delta_{ij}^*) = E(\Delta_{ij}^{*2}) = E\left[\left(n_{ij} - \frac{n_i}{K_1} k_j\right)^2\right],$$

and so

$$\begin{aligned}\mathrm{Var}(\Delta_{ij}^*) &= E(n_{ij}^2) - \frac{2n_i}{K_1} k_j E(n_{ij}) + \frac{n_i^2}{K_1^2} k_j^2 \\ &= \mathrm{Var}(n_{ij}) + [E(n_{ij})]^2 - \frac{n_i^2}{K_1^2} k_j^2 \\ &= \frac{n_i k_j (K_1 - k_j)}{K_1^2}. \end{aligned} \tag{5.6}$$

The variables Δ_{ij}^* for a particular treatment j in different centres are independent. If we have two different treatments j and m and two different centres i and l, then Δ_{ij}^* and Δ_{lm}^* are independent. Therefore, $\mathrm{Cov}(\Delta_{ij}^*, \Delta_{lm}^*) = 0$. The covariance is then calculated

for the same centre i as

$$\mathrm{Cov}(\Delta_{ij}^*, \Delta_{im}^*) = E(\Delta_{ij}^* \Delta_{im}^*) - E(\Delta_{ij}^*)E(\Delta_{im}^*) = E(\Delta_{ij}^* \Delta_{im}^*).$$

Thus,

$$\begin{aligned} \mathrm{Cov}(\Delta_{ij}^*, \Delta_{im}^*) &= E\left[\left(n_{ij} - \frac{n_i}{K_1}k_j\right)\left(n_{im} - \frac{n_i}{K_1}k_m\right)\right] \\ &= \mathrm{Cov}(n_{ij}, n_{im}) + E(n_{ij})E(n_{im}) - \frac{n_i^2}{K_1^2}k_j k_m. \end{aligned}$$

At centre i, each of the $\{n_{ij}\}$ for $j = 1, \ldots, K$ has a binomial distribution. The covariance of n_{ij} and n_{im} is deduced from the multinomial distribution of n_{ij} and n_{im} as $\mathrm{Cov}(n_{ij}, n_{im}) = -n_i k_j k_m / K_1^2$. Finally, the covariance of Δ_{ij}^* and Δ_{im}^* under complete randomization is

$$\begin{aligned} \mathrm{Cov}(\Delta_{ij}^*, \Delta_{im}^*) &= -\frac{n_i}{K_1^2}k_j k_m + \frac{n_i^2}{K_1^2}k_j k_m - \frac{n_i^2}{K_1^2}k_j k_m \\ &= -\frac{n_i k_j k_m}{K_1^2}. \end{aligned}$$

We then define the imbalance on a particular treatment j over all centres under complete randomization by

$$\Delta_j^* = \sum_{i=1}^{N} n_{ij} - \frac{n}{K_1}k_j = \sum_{i=1}^{N}\left(n_{ij} - \frac{n_i}{K_1}k_j\right) = \sum_{i=1}^{N}\Delta_{ij}^*. \tag{5.7}$$

The expectation of Δ_j^* is

$$\begin{aligned} E(\Delta_j^*) &= E\left(\sum_{i=1}^{N}\Delta_{ij}^*\right) \\ &= E\left(\sum_{i=1}^{N} n_{ij}\right) - \sum_{i=1}^{N}\frac{n_i}{K_1}k_j \\ &= \sum_{i=1}^{N}\frac{n_i}{K_1}k_j - \frac{n}{K_1}k_j = 0. \end{aligned}$$

The variance of Δ_j^* is then

$$\mathrm{Var}(\Delta_j^*) = E\left(\sum_{i=1}^{N}\Delta_{ij}^{*2}\right).$$

The variables Δ_{ij}^* for a particular treatment j in different centres are independent. It follows that

$$\begin{aligned}
\mathrm{Var}(\Delta_j^*) &= \sum_{i=1}^{N} \mathrm{Var}(\Delta_{ij}^*) \\
&= \sum_{i=1}^{N} \frac{n_i k_j (K_1 - k_j)}{K_1^2} \\
&= \frac{n k_j (K_1 - k_j)}{K_1^2}. \qquad (5.8)
\end{aligned}$$

Similarly, the covariance of Δ_j^* and Δ_m^* is

$$\begin{aligned}
\mathrm{Cov}(\Delta_j^*, \Delta_m^*) &= E\left(\sum_{i=1}^{N} \Delta_{ij}^* \Delta_{im}^*\right) \\
&= E\left[\sum_{i=1}^{N} \left(n_{ij} - \frac{n_i}{K_1} k_j\right)\left(n_{im} - \frac{n_i}{K_1} k_m\right)\right] \\
&= \sum_{i=1}^{N} \left[\mathrm{Cov}(n_{ij}, n_{im}) + E(n_{ij})E(n_{im}) - \frac{n_i^2}{K_1^2} k_j k_m\right] \\
&= -\sum_{i=1}^{N} \frac{n_i}{K_1^2} k_j k_m \\
&= -\frac{n}{K_1^2} k_j k_m. \qquad (5.9)
\end{aligned}$$

Equal Treatment Allocation

Assume that the $\{n_i\}$ are given. Then n_{ij} has a binomial distribution with parameters n_i and $1/K$ for equal treatment proportions within each block.

Let Δ_{ij}^* be the imbalance on treatment j in centre i. Then we have

$$\Delta_{ij}^* = n_{ij} - \frac{n_i}{K}. \qquad (5.10)$$

The expectation of this imbalance is

$$E(\Delta_{ij}^*) = E\left(n_{ij} - \frac{n_i}{K}\right) = 0.$$

Its variance is

$$\mathrm{Var}(\Delta_{ij}^*) = E(n_{ij}^2) - \frac{n_i}{K} E(n_{ij}) = \frac{n_i (K-1)}{K^2}.$$

Since the covariance of n_{ij} and n_{im} from the multinomial distribution is $\mathrm{Cov}(n_{ij}, n_{im}) = -n_i/K^2$, the covariance of Δ^*_{ij} and Δ^*_{im} is

$$\begin{aligned}\mathrm{Cov}(\Delta^*_{ij}, \Delta^*_{im}) &= -\frac{n_i}{K^2} - \frac{n_i^2}{K^2} + \frac{n_i^2}{K^2} \\ &= -\frac{n_i}{K^2}.\end{aligned}$$

The imbalance on a particular treatment j over all centres is defined as

$$\Delta^*_j = \sum_{i=1}^{N} n_{ij} - \frac{n}{K} = \sum_{i=1}^{N} \Delta^*_{ij}, \tag{5.11}$$

with expectation

$$E(\Delta^*_j) = E\left(\sum_{i=1}^{N} n_{ij}\right) - \sum_{i=1}^{N} \frac{n_i}{K} = 0.$$

The variance of Δ^*_j is

$$\mathrm{Var}(\Delta^*_j) = \sum_{i=1}^{N} E(\Delta^{*2}_{ij}) = \frac{n(K-1)}{K^2}. \tag{5.12}$$

Similarly, the covariance of Δ^*_j and Δ^*_m is

$$\mathrm{Cov}(\Delta^*_j, \Delta^*_m) = E\left[\sum_{i=1}^{N} \left(n_{ij} - \frac{n_i}{K}\right)\left(n_{im} - \frac{n_i}{K}\right)\right] = -\frac{n}{K^2}. \tag{5.13}$$

5.3.3 Overall imbalance

When the number of centres N involved in a clinical trial is large, the imbalance defined in (5.3) and (5.7) is approximated by a normal distribution with mean 0, and variance (5.4) and (5.8), respectively, for the permuted-block design and complete randomization. Let the overall imbalance for the permuted-block design and complete randomization be $\mathbf{\Delta}$ and $\mathbf{\Delta}^*$, respectively. Then each of these is a vector of imbalances on different treatments for all centres. In other words, $\mathbf{\Delta} = (\Delta_1, \Delta_2, ..., \Delta_K)^T$ and $\mathbf{\Delta}^* = (\Delta^*_1, \Delta^*_2, ..., \Delta^*_K)^T$.

As each of the Δ_j and Δ_j^* are asymptotically normally distributed, both $\boldsymbol{\Delta}$ and $\boldsymbol{\Delta}^*$ are asymptotically multivariate normal. Under the permuted-block design, the asymptotic multivariate normal distribution of $\boldsymbol{\Delta}$ has zero mean vector and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1K} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{K1} & \sigma_{K2} & \dots & \sigma_K^2 \end{pmatrix},$$

where $\sigma_j^2 = \mathrm{Var}(\Delta_j)$ in (5.4) and $\sigma_{jm} = \mathrm{Cov}(\Delta_j, \Delta_m)$ in (5.5). Similarly, for complete randomization, the overall imbalance $\boldsymbol{\Delta}^*$ has zero mean vector and covariance matrix

$$\boldsymbol{\Sigma}^* = \begin{pmatrix} \sigma_1^{*2} & \sigma_{12}^* & \dots & \sigma_{1K}^* \\ \sigma_{21}^* & \sigma_2^{*2} & \dots & \sigma_{2K}^* \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{K1}^* & \sigma_{K2}^* & \dots & \sigma_K^{*2} \end{pmatrix},$$

where $\sigma_j^{*2} = \mathrm{Var}(\Delta_j^*)$ in (5.8) or (5.12) and $\sigma_{jm}^* = \mathrm{Cov}(\Delta_j^*, \Delta_m^*)$ in (5.9) or (5.13).

5.4 The numbers of patients recruited in a centre are random variables

5.4.1 Patient recruitment model

We assume that the $\{n_i\}$ are random and use the Poisson-gamma model of Anisimov and Fedorov (2007) for the patient recruitment process. Assume that the number of patients recruited in centre i follows a Poisson process with rate λ_i. The rates $\{\lambda_i\}$ are gamma distributed with known parameters (α, β), where these are the shape and the rate parameters, respectively. For given rates, the number of patients recruited in centre i, n_i, has a binomial distribution with parameters n and p_i, where n is the total number of

patients recruited in all centres and p_i is the probability of recruiting a patient in centre i given by

$$p_i = \frac{\lambda_i}{\sum_{k=1}^{N} \lambda_k}.$$

Since the $\{\lambda_i\}$ are gamma distributed, p_i has a beta distribution with parameters $(\alpha, \alpha(N-1))$. Therefore, n_i has a beta-binomial distribution with

$$P(n_i = l) = P(n, N, \alpha, l) = \binom{n}{l} \frac{B(\alpha + l, \alpha(N-1) + n - l)}{B(\alpha, \alpha(N-1))}, \tag{5.14}$$

where $B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$ denotes the beta function, and expectation

$$E(n_i) = \frac{n}{N} \tag{5.15}$$

and variance

$$\mathrm{Var}(n_i) = \frac{n(N-1)(\alpha N + n)}{N^2(\alpha N + 1)}.$$

5.4.2 Permuted block design

Under the permuted block design with the $\{n_i\}$ random, consider the properties of the imbalance defined in (5.1). The expectation is

$$E(\Delta_{ij}) = E[E(\Delta_{ij}|n_i)] = 0.$$

The variance of Δ_{ij} is then

$$\begin{aligned}
\mathrm{Var}(\Delta_{ij}) &= E[E(\Delta_{ij}^2|n_i)] \\
&= E\left(E\left\{\left\lfloor\frac{n_i}{K_1}\right\rfloor k_j\xi_j[mod(n_i, K_1)] + \{\xi_j[mod(n_i, K_1)]\}^2 \right.\right. \\
&\quad \left.\left. - \frac{n_i}{K_1}k_j\xi_j[mod(n_i, K_1)]\right| n_i\right\}\Bigg) \\
&= E\left(\left\lfloor\frac{n_i}{K_1}\right\rfloor k_j E\{\xi_j[mod(n_i, K_1)]|n_i\} + E(\{\xi_j[mod(n_i, K_1)]\}^2|n_i)\right. \\
&\quad \left. - \frac{n_i}{K_1}k_j E\{\xi_m[mod(n_i, K_1)]|n_i\}\right).
\end{aligned}$$

Let $R_i = mod(n_i, K_1)$ be a random variable which represents the size of the incomplete block in a particular centre i. The variance can be written as

$$\begin{aligned}
\mathrm{Var}(\Delta_{ij}) &= E\left(\left\lfloor \frac{n_i}{K_1} \right\rfloor k_j \frac{R_i}{K_1} k_j + \mathrm{Var}[\xi_j(R_i)|n_i] + \{E[(\xi_j(R_i)|n_i]\}^2 - \frac{n_i}{K_1}\frac{k_j^2 R_i}{K_1}\right) \\
&= E\left[\frac{R_i}{K_1}k_j^2\left(-\frac{R_i}{K_1}\right) + \frac{R_i k_j (K_1 - k_j)(K_1 - R_i)}{K_1^2(K_1 - 1)} + \frac{k_j^2 R_i^2}{K_1^2}\right] \\
&= E\left[\frac{R_i k_j (K_1 - k_j)(K_1 - R_i)}{K_1^2(K_1 - 1)}\right] \\
&= \frac{k_j(K_1 - k_j)}{K_1(K_1 - 1)}E(R_i) - \frac{k_j(K_1 - k_j)}{K_1^2(K_1 - 1)}E(R_i^2). \qquad (5.16)
\end{aligned}$$

The covariance for two different treatments j and m is calculated in a similar way using

$$\mathrm{Cov}(\Delta_{ij}, \Delta_{im}) = E(\Delta_{ij}\Delta_{im}) - E(\Delta_{ij})E(\Delta_{im}) = E(\Delta_{ij}\Delta_{im}),$$

so that

$$\begin{aligned}
\mathrm{Cov}(\Delta_{ij}, \Delta_{im}) &= E[E(\Delta_{ij}\Delta_{im}|n_i)] \\
&= E\left(\left\lfloor \frac{n_i}{K_1} \right\rfloor k_j E\{\xi_m[mod(n_i, K_1)|n_i]\}\right) \\
&\quad + E\left(E\{\xi_m[mod(n_i, K_1)]\xi_j[mod(n_i, K_1)]|n_i\}\right) \\
&\quad - E\left(\frac{n_i}{k_1}k_j E\left\{\xi_m[mod(n_i, K_1)]|n_i\right\}\right).
\end{aligned}$$

Let $R_i = mod(n_i, K_1)$. Then

$$\begin{aligned}
\mathrm{Cov}(\Delta_{ij}, \Delta_{im}) &= E\left(\left\lfloor \frac{n_i}{K_1} \right\rfloor k_j \frac{k_m R_i}{K_1} + \mathrm{Cov}(\xi_j(R_i), \xi_m(R_i)|n_i)\right. \\
&\quad \left. + E[\xi_j(R_i)|n_i]E[\xi_m(R_i)|n_i] - \frac{n_i}{K_1}k_j\frac{R_i k_m}{K_1}\right) \\
&= E\left[-\frac{k_j k_m R_i(K_1 - R_i)}{K_1^2(K_1 - 1)}\right] \\
&= -\frac{k_j k_m}{K_1(K_1 - 1)}E(R_i) + \frac{k_j k_m}{K_1^2(K_1 - 1)}E(R_i^2). \qquad (5.17)
\end{aligned}$$

In Anisimov (2007, 2011), the probability $P(mod(n_i, K_1) = r) = P(R_i = r)$ is given as

$$q_r(n, N, \alpha, K_1) = \sum_{s=0}^{n/K_1 - 1} P(n, N, \alpha, r + sK_1) \qquad (5.18)$$

for $r = 0, 1, ..., K_1 - 1$, where $P(n, N, \alpha, r + sK_1)$ is defined in (5.14). So

$$\begin{aligned} E(R_i) &= \sum_{r=0}^{K_1-1} rP(R_i = r) \\ &= \sum_{r=0}^{K_1-1} \sum_{s=0}^{n/K_1-1} r\binom{n}{r+sK_1} \frac{B(\alpha + r + sK_1, \alpha(N-1) + n - r - sK_1)}{B(\alpha, \alpha(N-1))}. \end{aligned}$$

Similarly,

$$E(R_i^2) = \sum_{r=0}^{K_1-1} \sum_{s=0}^{n/K_1-1} r^2\binom{n}{r+sK_1} \frac{B(\alpha + r + sK_1, \alpha(N-1) + n - r - sK_1)}{B(\alpha, \alpha(N-1))}.$$

Note that, if $n/N > 2K_1$, we can assume that the random variable R_i has an approximate discrete uniform distribution on the interval between 0 and $K_1 - 1$. The probability of obtaining each of the values between 0 and $K_1 - 1$ is $1/K_1$. Therefore, the expectation of R_i is

$$E(R_i) = \frac{K_1 - 1}{2}$$

and its variance is

$$\mathrm{Var}(R_i) = \frac{K_1^2 - 1}{12}.$$

Then we can obtain

$$E(R_i^2) = \mathrm{Var}(R_i) + [E(R_i)]^2 = \frac{(K_1 - 1)(2K_1 - 1)}{6}.$$

Therefore, (5.16) and (5.17) become

$$\begin{aligned} \mathrm{Var}(\Delta_{ij}) &= \frac{k_j(K_1 - k_j)}{2K_1} - \frac{k_j(K_1 - k_j)(2K_1 - 1)}{6K_1^2} \\ &= \frac{k_j(K_1 - k_j)(K_1 + 1)}{6K_1^2} \end{aligned} \tag{5.19}$$

and

$$\begin{aligned} \mathrm{Cov}(\Delta_{ij}, \Delta_{im}) &= -\frac{k_j k_m}{2K_1} + \frac{k_j k_m (2K_1 - 1)}{6K_1^2} \\ &= -\frac{k_j k_m (K_1 + 1)}{6K_1^2}. \end{aligned}$$

The imbalance Δ_j on treatment j for all centres is defined in (5.3) and has expectation

$$\begin{aligned} E(\Delta_j) &= E\left[E\left(\sum_{i=1}^{N}\Delta_{ij}\Big|n_i\right)\right] \\ &= E\left\{\sum_{i=1}^{N}E\left(n_{ij}-\frac{n_i}{K_1}k_j\Big|n_i\right)\right\} \\ &= E\left[\sum_{i=1}^{N}E\left(n_{ij}|n_i\right)\right]-E\left(\sum_{i=1}^{N}\frac{n_i}{K_1}k_j\right) \\ &= \frac{nk_j}{K_1}-\frac{nk_j}{K_1}=0. \end{aligned}$$

Given the $\{n_i\}$, the variables $\{\Delta_{ij}\}$ for a particular treatment j in different centres are independent, and so

$$\begin{aligned} \mathrm{Var}(\Delta_j) &= E[E(\Delta_j^2|n_i)] \\ &= E\left[\sum_{i=1}^{N}\mathrm{Var}(\Delta_{ij}|n_i)\right] \\ &= \sum_{i=1}^{N}\mathrm{Var}(\Delta_{ij}). \end{aligned}$$

From (5.16), we obtain

$$\mathrm{Var}(\Delta_j) = \sum_{i=1}^{N}\left[\frac{k_j(K_1-k_j)}{K_1(K_1-1)}E(R_i)-\frac{k_j(K_1-k_j)}{K_1^2(K_1-1)}E(R_i^2)\right]. \qquad (5.20)$$

For $n/N > 2K_1$, the random variable R_i has an approximate discrete uniform distribution. Using the results for $E(R_i)$ and $E(R_i^2)$, and (5.19), the variance of Δ_j is

$$\mathrm{Var}(\Delta_j) = \frac{Nk_j(K_1-k_j)(K_1+1)}{6K_1^2}. \qquad (5.21)$$

Similarly, the covariance of Δ_j and Δ_m is

$$\mathrm{Cov}(\Delta_j,\Delta_m) = \sum_{i=1}^{N}E\left[\mathrm{Cov}(\Delta_{ij},\Delta_{im}|n_i)\right] = \sum_{i=1}^{N}\mathrm{Cov}(\Delta_{ij},\Delta_{im}).$$

Using (5.17), we obtain

$$\mathrm{Cov}(\Delta_j,\Delta_m) = \sum_{i=1}^{N}\left\{-\frac{k_jk_m}{K_1(K_1-1)}E(R_i)+\frac{k_jk_m}{K_1^2(K_1-1)}E(R_i^2)\right\}. \qquad (5.22)$$

Finally, when R_i has an approximate discrete uniform distribution, the covariance of Δ_j and Δ_m is

$$\mathrm{Cov}(\Delta_j, \Delta_m) = -\frac{Nk_jk_m(K_1+1)}{6K_1^2}. \tag{5.23}$$

5.4.3 Complete randomization

General Case

Now we study the imbalance of complete randomization when the $\{n_i\}$ are random.

Under complete randomization, the imbalance on treatment j in centre i is defined in (5.2). So its expectation is

$$\begin{aligned} E(\Delta_{ij}^*) &= E[E(\Delta_{ij}^*|n_i)] \\ &= E[E(n_{ij}|n_i)] - E\left(\frac{n_i}{K_1}k_j\right) \\ &= E\left(\frac{n_i}{K_1}k_j\right) - E\left(\frac{n_i}{K_1}k_j\right) = 0. \end{aligned}$$

The variance of Δ_{ij}^* is

$$\begin{aligned} \mathrm{Var}(\Delta_{ij}^*) &= E[E(\Delta_{ij}^{*2})|n_i] \\ &= E\left\{E\left[\left(n_{ij} - \frac{n_i}{K_1}k_j\right)\left(n_{ij} - \frac{n_i}{K_1}k_j\right)\middle| n_i\right]\right\} \\ &= E\left[E(n_{ij}^2|n_i) - \frac{n_i}{K_1}k_jE(n_{ij}|n_i)\right] \\ &= E\left\{\mathrm{Var}(n_{ij}|n_i) + [E(n_{ij}|n_i)]^2 - \frac{n_i^2}{K_1^2}k_j^2\right\} \\ &= E\left[\frac{n_ik_j(K_1-k_j)}{K_1^2} + \frac{n_i^2}{K_1^2}k_j^2 - \frac{n_i^2}{K_1^2}k_j^2\right] \\ &= \frac{k_j(K_1-k_j)}{K_1^2}E(n_i) \\ &= \frac{nk_j(K_1-k_j)}{NK_1^2}, \end{aligned} \tag{5.24}$$

using (5.15).

Since Δ_{ij}^* and Δ_{lm}^* are independent variables for two different centres i and l, their

covariance is zero. The covariance of Δ_{ij}^* and Δ_{im}^* for a particular centre i is

$$\begin{aligned}
\mathrm{Cov}(\Delta_{ij}^*, \Delta_{im}^*) &= E[E(\Delta_{ij}^*\Delta_{im}^*|n_i)] \\
&= E\left\{E\left[\left(n_{ij} - \frac{n_i}{K_1}k_j\right)\left(n_{im} - \frac{n_i}{K_1}k_m\right)\middle|n_i\right]\right\} \\
&= E\left[E(n_{ij}n_{im}|n_i) - E\left(\frac{n_i}{K_1}k_j n_{im}\middle|n_i\right)\right] \\
&= E\left[\mathrm{Cov}(n_{ij}, n_{im}|n_i) + E(n_{ij}|n_i)E(n_{im}|n_i) - \frac{n_i}{K_1}k_j E(n_{im}|n_i)\right].
\end{aligned}$$

For centre i, the conditional covariance of n_{ij} and n_{im} can be obtained from the multinomial distribution and is given by $-n_i k_j k_m/K_1^2$. So the covariance of Δ_{ij}^* and Δ_{im}^* is

$$\begin{aligned}
\mathrm{Cov}(\Delta_{ij}^*, \Delta_{im}^*) &= E\left(-\frac{n_i}{K_1^2}k_j k_m + \frac{n_i^2}{K_1^2}k_j k_m - \frac{n_i^2}{K_1^2}k_j k_m\right) \\
&= -\frac{k_j k_m}{K_1^2}E(n_i) \\
&= -\frac{n k_j k_m}{N K_1^2}.
\end{aligned}$$

The imbalance over all centres on treatment j defined in (5.7) has expectation

$$\begin{aligned}
E(\Delta_j^*) &= E\left\{E\left[\sum_{i=1}^{N}\left(n_{ij} - \frac{n_i}{K_1}k_j\right)\middle|n_i\right]\right\} \\
&= E\left\{\sum_{i=1}^{N}\left[E(n_{ij}|n_i) - \frac{n_i}{K_1}k_j\right]\right\} \\
&= \sum_{i=1}^{N}\left[E\left(\frac{n_i}{K_1}k_j\right) - E\left(\frac{n_i}{K_1}k_j\right)\right] = 0.
\end{aligned}$$

Given the $\{n_i\}$, the variables $\{\Delta_{ij}^*\}$ for a particular treatment j are independent for different centres, so that

$$\begin{aligned}
\mathrm{Var}(\Delta_j^*) &= \sum_{i=1}^{N}\mathrm{Var}(\Delta_{ij}^*) \\
&= \sum_{i=1}^{N}\frac{n k_j(K_1 - k_j)}{N K_1^2} \\
&= \frac{n k_j(K_1 - k_j)}{K_1^2}. \qquad (5.25)
\end{aligned}$$

The covariance of Δ_j^* and Δ_m^* is

$$\begin{aligned}\mathrm{Cov}(\Delta_j^*, \Delta_m^*) &= E(\Delta_j^* \Delta_m^*) \\ &= \sum_{i=1}^{N} E(\Delta_{ij}^* \Delta_{im}^*) \\ &= -\sum_{i=1}^{N} \frac{n k_j k_m}{N K_1^2} \\ &= -\frac{n k_j k_m}{K_1^2}. \end{aligned} \tag{5.26}$$

Here, it is clearly seen that, under complete randomization, the variance and covariance of the imbalance on treatments for all centres are the same as those when n_i is given.

Equal Treatment Allocation

From (5.10), the imbalance on treatment j in centre i has expectation

$$E(\Delta_{ij}^*) = E[E(\Delta_{ij}^* | n_i)] = 0.$$

The variance of Δ_{ij}^* is

$$\begin{aligned}\mathrm{Var}(\Delta_{ij}^*) &= E[E(\Delta_{ij}^{*2}) | n_i] \\ &= E\left\{ \mathrm{Var}(n_{ij}|n_i) + [E(n_{ij}|n_i)]^2 - \frac{n_i^2}{K^2} \right\} \\ &= \frac{n(K-1)}{NK^2}. \end{aligned}$$

We only consider the covariance of the imbalance on two different treatments in the same centre. Given the $\{n_i\}$, Δ_{ij}^* and Δ_{lm}^*, the imbalances on two treatments j and m in two different centres i and l are independent with covariance equal to zero. The covariance of Δ_{ij}^* and Δ_{im}^* is

$$\begin{aligned}\mathrm{Cov}(\Delta_{ij}^*, \Delta_{im}^*) &= E[E(\Delta_{ij}^* \Delta_{im}^* | n_i)] \\ &= E\left[\mathrm{Cov}(n_{ij}, n_{im}|n_i) + E(n_{ij}|n_i) E(n_{im}|n_i) - \frac{n_i}{K} E(n_{im}|n_i) \right]. \end{aligned}$$

Since the conditional covariance of n_{ij} and n_{im} is $-n_i/K^2$, we have

$$\begin{aligned}\mathrm{Cov}(\Delta^*_{ij}, \Delta^*_{im}) &= E\left(-\frac{n_i}{K^2} + \frac{n_i^2}{K^2} - \frac{n_i^2}{K^2}\right) \\ &= -\frac{n}{NK^2}. \end{aligned} \tag{5.27}$$

The imbalance over all centres on treatment j from (5.11) has expectation

$$E(\Delta^*_j) = E\left\{\sum_{i=1}^{N}\left[E(n_{ij}|n_i) - \frac{n_i}{K}\right]\right\} = 0.$$

The variance of Δ^*_j is

$$\mathrm{Var}(\Delta^*_j) = \sum_{i=1}^{N}\frac{n(K-1)}{NK^2} = \frac{n(K-1)}{K^2}. \tag{5.28}$$

Using (5.27), the covariance of Δ^*_j and Δ^*_m is

$$\begin{aligned}\mathrm{Cov}(\Delta^*_j, \Delta^*_m) &= \sum_{i=1}^{N}\mathrm{Cov}(\Delta^*_{ij}, \Delta^*_{im}) \\ &= -\sum_{i=1}^{N}\frac{n}{NK^2} \\ &= -\frac{n}{K^2}. \end{aligned} \tag{5.29}$$

5.4.4 Overall imbalance

When N the number of centres involved in a trial is large, the imbalances Δ_j and Δ^*_j will be asymptotically normally distributed with mean 0 and variance given by (5.20) and (5.25), respectively. Therefore, the overall imbalances $\boldsymbol{\Delta}$ and $\boldsymbol{\Delta}^*$ are asymptotically multivariate normal.

Under the permuted block design, the imbalance $\boldsymbol{\Delta}$ has zero mean vector and covariance matrix Σ with entries σ^2_j in (5.21) and σ_{jm} in (5.23). We use the approximation for the variables $\{R_i\}$ that they have an approximate discrete uniform distribution. For large N, the variance of Δ_j is just N times the variance of Δ_{ij} and similarly for the covariance. The covariance matrix of the overall imbalance is N times the covariance matrix for a

particular centre. Similarly, for complete randomization, the overall imbalance $\mathbf{\Delta}^*$ has an approximate multivariate normal distribution with zero mean vector. The covariance matrix Σ^* has entries σ_j^{*2} in (5.25) or (5.28) and σ_{jm}^* in (5.26) or (5.29), or it can be obtained by multiplying the covariance matrix for a particular centre by N.

Remark 1 For $K = 2$ when there are only two treatments, a and b, the overall imbalance was defined by Anisimov (2007) as the difference between the numbers of patients on treatments a and b. We denote this overall imbalance by $\widetilde{\mathbf{\Delta}} = n_{.a} - n_{.b}$, where $n_{.a} = \sum_{i=1}^{N} n_{ia}$ and $n_{.b} = \sum_{i=1}^{N} n_{ib}$. In our case when $K = 2$, treatment 1 refers to treatment a and treatment 2 refers to treatment b. The overall imbalance $\widetilde{\mathbf{\Delta}}$ defined by Anisimov (2007) can be written in terms of our notation as

$$\begin{aligned} \widetilde{\mathbf{\Delta}} &= n_{.a} - n_{.b} \\ &= \Delta_a - \Delta_b + \left(\frac{n}{K_1} k_a - \frac{n}{K_1} k_b \right). \end{aligned}$$

When there is an equal proportion of patients for the two treatments within a complete block, we have $\widetilde{\mathbf{\Delta}} = \Delta_a - \Delta_b$. Now, the overall imbalance in our notation is $\mathbf{\Delta} = (\Delta_a, \Delta_b)^T$ with expectation $E(\mathbf{\Delta}) = (0, 0)^T$ and covariance matrix Σ with entries σ_j^2 in (5.21) and σ_{jm} in (5.23).

Let $\mathbf{u} = (1, -1)^T$. Then we can write the overall imbalance $\widetilde{\mathbf{\Delta}}$ in terms of $\mathbf{\Delta}$ as $\widetilde{\mathbf{\Delta}} = \mathbf{u}^T \mathbf{\Delta}$. The expectation of $\widetilde{\mathbf{\Delta}}$ is zero and its variance can be written as

$$\begin{aligned} \mathrm{Var}(\widetilde{\mathbf{\Delta}}) &= \mathbf{u}^T \Sigma \mathbf{u} \\ &= (1, -1) \begin{pmatrix} \frac{N k_a (K_1 - k_a)(K_1 + 1)}{6K_1^2} & -\frac{N k_a k_b (K_1 + 1)}{6K_1^2} \\ -\frac{N k_a k_b (K_1 + 1)}{6K_1^2} & \frac{N k_b (K_1 - k_b)(K_1 + 1)}{6K_1^2} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\ &= \frac{2N k_a k_b (K_1 + 1)}{3K_1^2}. \end{aligned}$$

Remark 2 For equal proportions $k_a = k_b = K_1/2$ within a complete block, we have $\mathrm{Var}(\widetilde{\mathbf{\Delta}}) = N(K_1 + 1)/6$.

For large N, Δ_a and Δ_b are approximately normal. Therefore, the imbalance $\tilde{\mathbf{\Delta}}$ as a linear combination of Δ_a and Δ_b is also approximately normal. The result calculated here in terms of our notation for the expectation and variance of $\tilde{\mathbf{\Delta}}$ matches with the expectation and variance of the overall imbalance calculated by Anisimov (2007, 2010).

5.5 Comparison of imbalance of complete randomization and the permuted block design

5.5.1 The numbers of patients recruited in different centres are given

One of the main aims of a randomization scheme is to balance the numbers of patients across treatment groups. This ensures that an adequate level of power can be achieved. The higher the power, the more likely a test will detect a genuine treatment difference. The variance of the imbalance is always considered to be a good indicator of which randomization schemes provide better balance. The greater the variance of the imbalance of a randomization scheme, the less efficient is the design for balancing the numbers of patients across groups.

By using the calculated variance of the imbalance in (5.3), (5.6), (5.19) and (5.24), comparisons can be made between the two randomization schemes on their effectiveness in balancing the numbers of patients across treatment groups. First suppose that n_i, the number of patients in a particular centre i, is given. The ratio of the variances of the imbalances for the two randomization schemes is

$$\begin{aligned}\frac{\mathrm{Var}(\Delta_{ij})}{\mathrm{Var}(\Delta_{ij}^*)} &= \frac{rk_j(K_1-k_j)(K_1-r)}{K_1^2(K_1-1)} \Big/ \frac{n_ik_j(K_1-k_j)}{K_1^2} \\ &= \frac{r(K_1-r)}{n_i(K_1-1)}.\end{aligned}$$

Let $s \geq 0$ such that $n_i = sK_1 + r$. Then the ratio becomes

$$\frac{\text{Var}(\Delta_{ij})}{\text{Var}(\Delta^*_{ij})} = \frac{r(K_1 - r)}{(sK_1 + r)(K_1 - 1)}.$$

If $s = 0$ and $r = 0$, there will be no patients in the trial and we cannot calculate the ratio. When $s = 0$, there is only one incomplete block, so that the ratio becomes

$$\frac{K_1 - r}{K_1 - 1} = \begin{cases} 1, & r = 1, \\ < 1, & r > 1. \end{cases}$$

When $r = 1$, there will be only one patient in centre i and the variance of the imbalance is the same for complete randomization and the permuted-block design. When $r > 1$, there is only one incomplete block in centre i. The variance of that imbalance under complete randomization is greater than that of the permuted block design for a non-empty incomplete block, which means that the permuted-block design is more efficient than complete randomization.

Now consider the case $s \geq 1$. When $r = 0$, centre i contains only complete blocks and the ratio becomes zero. It is obvious that there is no imbalance in a complete block. Therefore, there is no imbalance in this centre and hence no variance for the imbalance can be calculated for both randomization schemes. Now suppose that $r \geq 1$, which implies that $K_1 - r \leq K_1 - 1$, and, for $s \geq 1$ and $K_1 \geq 1$, implies that $sK_1 + r \geq r$. Therefore, we have $r(K_1 - r) \leq (sK_1 + r)(K_1 - 1)$. If in centre i, there are more than two blocks which include at least one complete block and one incomplete block, the variance of the imbalance for complete randomization is greater than that of the permuted-block design.

To conclude, the variance of the imbalance in a particular centre i on a particular treatment j is greater under complete randomization than the permuted-block design in all cases. In other words, the permuted-block design is more efficient than complete randomization, except when there is only one patient in each centre.

We will now study the overall imbalance on treatments under these two randomization schemes. We can look at the covariance matrices for the randomization schemes to see which randomization scheme is better. Therefore, we will study the entries in the covariance matrices, which are $\mathrm{Var}(\Delta_j)$, $\mathrm{Var}(\Delta_j^*)$, $\mathrm{Cov}(\Delta_j, \Delta_m)$ and $\mathrm{Cov}(\Delta_j^*, \Delta_m^*)$.

The imbalances on a particular treatment j are independent for different centres. The variance of the imbalance on a particular treatment j for all centres is just the sum of all the imbalances on treatment j for each centre i for $i = 1, ..., N$. Therefore, the variance of the imbalance on treatment j for all centres under permuted-block randomization will be less than the variance of the imbalance on treatment j for all centres under complete randomization. Similarly, consider the covariances for two different treatments j and m in centre i under these two randomization schemes. The ratio of the two covariances is

$$\begin{aligned}\frac{\mathrm{Cov}(\Delta_{ij}, \Delta_{im})}{\mathrm{Cov}(\Delta_{ij}^*, \Delta_{im}^*)} &= -\frac{k_j k_m r(K_1 - r)}{K_1^2(K_1 - 1)} \Big/ -\frac{n_i k_j k_m}{K_1^2} \\ &= \frac{r(K_1 - r)}{(sK_1 + r)(K_1 - 1)},\end{aligned}$$

for $n_i = sK_1 + r$.

The covariance ratio of the imbalance is exactly what we obtained for the variance ratio of the imbalance. Therefore, we can draw the same conclusions here for the covariance ratio of the imbalance. The covariance of the imbalance in a particular centre i on two different treatments j and m is greater under complete randomization than permuted block randomization in all cases. Therefore, all the entries in the covariance matrix Σ^* for complete randomization have values greater than all the entries in the covariance matrix Σ under permuted-block randomization. For the overall imbalance across treatments, the covariance under permuted-block randomization is less than that under complete randomization, which implies that, under permuted-block randomization, a more balanced trial can be achieved across treatments for all centres than complete randomization.

5.5.2 The numbers of patients recruited in different centres are random

Consider now the case where the numbers of patients $\{n_i\}$ to be recruited are random variables and the patient recruitment process follows a Poisson-gamma model.

As proved above,

$$\mathrm{Var}(\Delta_{ij}|n_i) = \mathrm{Var}(\Delta^*_{ij}|n_i) \text{ if } n_i = 1$$

and

$$\mathrm{Var}(\Delta_{ij}|n_i) < \mathrm{Var}(\Delta^*_{ij}|n_i) \text{ if } n_i > 1.$$

Now, when the n_i are random, if $n > 1$, then $P(n_i > 1) > 0$. Thus,

$$\begin{aligned}
\mathrm{Var}(\Delta_{ij}) &= E[\mathrm{Var}(\Delta_{ij} \mid n_i)] \\
&= \sum_{s=0}^{n} \mathrm{Var}(\Delta_{ij}|n_i = s)P(n_i = s) \\
&< \sum_{s=0}^{n} \mathrm{Var}(\Delta^*_{ij}|n_i = s)P(n_i = s) \\
&= \mathrm{Var}(\Delta^*_{ij}).
\end{aligned}$$

Therefore, if we have more than one patient in a trial, $\mathrm{Var}\ (\Delta_{ij}) < \mathrm{Var}(\Delta^*_{ij})$ and permuted-block randomization always has lower variability in the imbalance than complete randomization.

Similarly, for the covariance of the imbalance for two different treatments j and m in centre i,

$$\mathrm{Cov}(\Delta_{ij}, \Delta_{im}|n_i) = \mathrm{Cov}(\Delta^*_{ij}, \Delta^*_{im}|n_i) \text{ if } n_i = 1$$

and

$$\mathrm{Cov}(\Delta_{ij}, \Delta_{im}|n_i) < \mathrm{Cov}(\Delta^*_{ij}, \Delta^*_{im}|n_i) \text{ if } n_i > 1.$$

Therefore, for n_i random, if there is more than one patient in the trial, we have $\mathrm{Cov}\ (\Delta_{ij}, \Delta_{im}) < \mathrm{Cov}(\Delta^*_{ij}, \Delta^*_{im})$.

For the overall imbalance, we look at the covariance matrices under the two randomization schemes. The conclusion is the same as when n_i is fixed. Permuted-block randomization is better than complete randomization and thus permuted-block randomization provides less imbalance in the number of patients on a particular treatment for all centres than complete randomization. The entries in the covariance matrix for permuted-block randomization are all less than those for complete randomization. Therefore, permuted-block randomization gives less overall imbalance than complete randomization.

5.6 Simulation for the expectation and covariance matrix of the overall imbalance

Results of simulation support what we have found theoretically in Section 5.4. Consider the centre-stratified randomization process in a study. Assume that the patient recruitment process is modelled by the Poisson-gamma model, where the number of patients to be recruited in each centre is simulated from a beta-binomial distribution. Within each centre, patients are allocated to a treatment according to some randomly permuted blocks. Assume that an equal proportion of patients is to be allocated to each treatment within each complete block. Some randomly permuted blocks of size K_1 will be generated in the simulation and the patients will be allocated to treatments according to the sequence formed by these randomly permuted blocks. As the imbalance on treatments is found in the incomplete block for each centre, the treatment allocation to patients in these incomplete blocks will be simulated from a multivariate hypergeometric distribution. The imbalance on treatments can be calculated by subtracting the simulated number of patients allocated to each of the treatments from the expected number of pa-

tients on each treatment. The imbalance on treatments can be calculated within each centre and for all centres. Finally, the vector for the overall imbalance on treatments will be obtained.

The above procedure will be simulated s times. The sample mean vector and the sample covariance matrix of the overall imbalance can be obtained after s runs. For K treatments, let $\widehat{\boldsymbol{\Delta}}_p$ be the vector for the overall imbalance obtained in the pth simulation run. The sample mean vector is calculated by

$$\bar{\widehat{\boldsymbol{\Delta}}} = \frac{1}{s}\sum_{p=1}^{s}\widehat{\boldsymbol{\Delta}}_p.$$

The sample covariance matrix is obtained by

$$\hat{\Sigma} = \frac{1}{s-1}\sum_{p=1}^{s}(\widehat{\boldsymbol{\Delta}}_p - \bar{\widehat{\boldsymbol{\Delta}}})(\widehat{\boldsymbol{\Delta}}_p - \bar{\widehat{\boldsymbol{\Delta}}})^T.$$

The values for the sample mean vector and covariance matrix will then be compared with the theoretical mean and covariance matrix. We consider two particular scenarios. The first scenario has $n = 168$, $N = 100$ $K = 4$ and $K_1 = 8$, with $\alpha = 1.2$ and $\beta = 2$ for the patient recruitment process. In the second scenario, we have $n = 232$, $N = 100$ and the same K, K_1, α and β as in the first. The theoretical mean vector and covariance matrix of the overall imbalance can be calculated by using results in Section 5.4. In both scenarios, the theoretical expectation of the overall imbalance is the vector of zeros. The theoretical covariance matrix of the overall imbalance has diagonal entries in (5.20) and off-diagonal entries in (5.22). If the size of the incomplete block R_i in centre i has an approximate discrete uniform distribution, the theoretical covariance matrix has entries (5.21) and (5.23).

For both scenarios, we took $s = 100,000$. For the first scenario, the sample mean vector is

$$\bar{\widehat{\boldsymbol{\Delta}}} = (-0.003, -0.015, 0.025, -0.007)^T,$$

which is close to the theoretical mean vector. The theoretical covariance matrix for this scenario is

$$\Sigma = \begin{pmatrix} 18.961 & -6.320 & -6.320 & -6.320 \\ -6.320 & 18.961 & -6.320 & -6.320 \\ -6.320 & -6.320 & 18.961 & -6.320 \\ -6.320 & -6.320 & -6.320 & 18.961 \end{pmatrix}. \tag{5.30}$$

If the size of the incomplete block R_i in centre i has an approximate discrete uniform distribution, the theoretical covariance matrix is

$$\Sigma = \begin{pmatrix} 28.125 & -9.375 & -9.375 & -9.375 \\ -9.375 & 28.125 & -9.375 & -9.375 \\ -9.375 & -9.375 & 28.125 & -9.375 \\ -9.375 & -9.375 & -9.375 & 28.125 \end{pmatrix}. \tag{5.31}$$

The sample covariance matrix is

$$\hat{\Sigma} = \begin{pmatrix} 18.863 & -6.284 & -6.258 & -6.320 \\ -6.284 & 19.095 & -6.454 & -6.356 \\ -6.258 & -6.454 & 18.962 & -6.250 \\ -6.320 & -6.356 & -6.250 & 18.926 \end{pmatrix},$$

We can see that $\hat{\Sigma}$ has values close to the theoretical values in (5.30).

Consider the second scenario with $n = 232, N = 100, K = 4, K_1 = 8, \alpha = 1.2$ and $\beta = 2$. The sample mean vector is

$$\bar{\hat{\boldsymbol{\Delta}}} = (0.002, -0.019, 0.025, -0.008)^T,$$

which is again close to the theoretical mean vector. The theoretical covariance matrix is

$$\Sigma = \begin{pmatrix} 21.668 & -7.223 & -7.223 & -7.223 \\ -7.223 & 21.668 & -7.223 & -7.223 \\ -7.223 & -7.223 & 21.668 & -7.223 \\ -7.223 & -7.223 & -7.223 & 21.668 \end{pmatrix}. \tag{5.32}$$

When R_i, the size of the incomplete block in centre i, has an approximate discrete uniform distribution, we have

$$\Sigma = \begin{pmatrix} 28.125 & -9.375 & -9.375 & -9.375 \\ -9.375 & 28.125 & -9.375 & -9.375 \\ -9.375 & -9.375 & 28.125 & -9.375 \\ -9.375 & -9.375 & -9.375 & 28.125 \end{pmatrix}. \tag{5.33}$$

The sample covariance matrix of the overall imbalance is

$$\hat{\Sigma} = \begin{pmatrix} 21.723 & -7.080 & -7.404 & -7.239 \\ -7.080 & 21.444 & -7.136 & -7.228 \\ -7.404 & -7.136 & 21.667 & -7.126 \\ -7.239 & -7.228 & -7.126 & 21.593 \end{pmatrix},$$

which also has values close to the theoretical values in (5.32).

For both scenarios, the results from simulation are consistent with the numerical values for the theoretical mean vector and covariance matrices in (5.30) and (5.32). When the size of the incomplete block R_i in centre i has an approximate discrete uniform distribution, the theoretical covariance matrix is less accurate. In the second scenario, the value assumed for n, the number of patients recruited, is higher than the corresponding value assumed in the first scenario. The difference between the two theoretical covariance matrices is larger when we have fewer patients.

The theoretical covariance matrices in (5.30) and (5.32) hold when N the number of centres involved in the trial is large. Furthermore, those in (5.31) and (5.33) hold when $n/N > 2K_1$ such that R_i has an approximate discrete uniform distribution. This is a good approximation, since for any size of the incomplete block r, $P(R_i = r) \approx 1/K_1$ when $n/N > 2K_1$. For example, take $n = 720$, $N = 80$, $K_1 = 4$ and $\alpha = 1.2$ such that $n/N = 9$ and $2K_1 = 8$. Using (5.18), $P(R_i = r)$ for $r = 0, 1, 2, 3$ is 0.2761,

0.2616, 0.2416 and 0.2207, respectively. We can see that all of these values are close to $1/K_1 = 0.25$, and so it will be appropriate for us to use the approximate discrete uniform for R_i when $n/N > 2K_1$. We have n/N equal to 1.68 for the first scenario and 2.32 for the second scenario, which are both far less than $2K_1 = 16$. The size of the incomplete block in any centre i for both scenarios cannot be approximated by a discrete uniform distribution. Therefore, the values for the sample covariance matrices in the two scenarios are not close to the theoretical covariance matrices in (5.31) and (5.33). In general, the discrete uniform approximation is quite good when $n/N > K_1$.

As defined in (5.3) , Δ_j the imbalance on a particular treatment j for $j = 1, ..., K$ for all centres is approximated by a normal distribution with mean 0 and variance (5.20), or (5.21) if the size of the incomplete block has an approximate discrete uniform distribution. The results of the simulations support this. For both scenarios, we consider $K = 4$ treatments with block size $K_1 = 8$. The values of Δ_j for $j = 1, .., 4$ are calculated for $100,000$ simulations. Below are the histograms of the values of Δ_j for each treatment $j = 1, ..., 4$ for both scenarios.

Figure 5.1: Histograms for simulated values of Δ_j for $j = 1, ..., 4$ for the first scenario

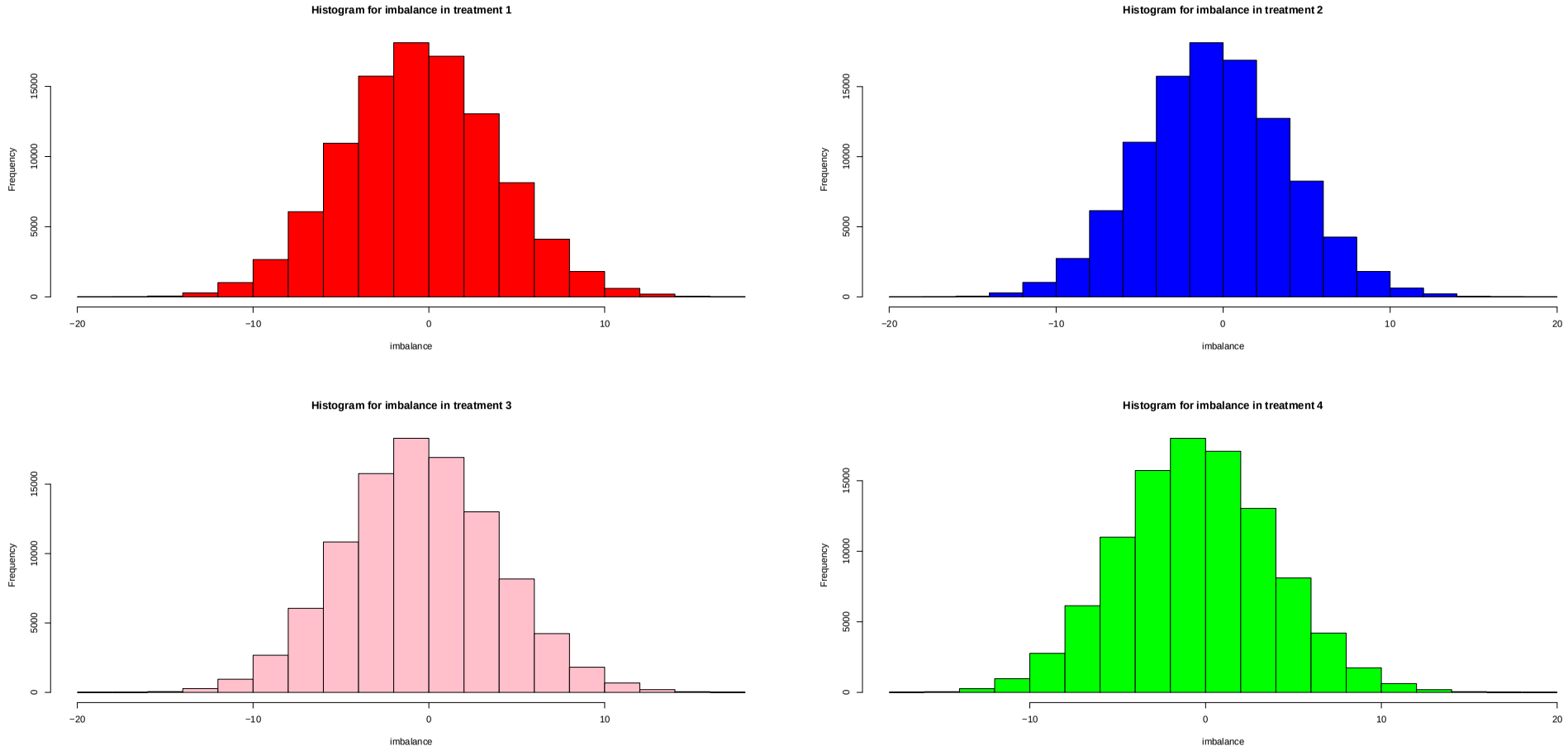
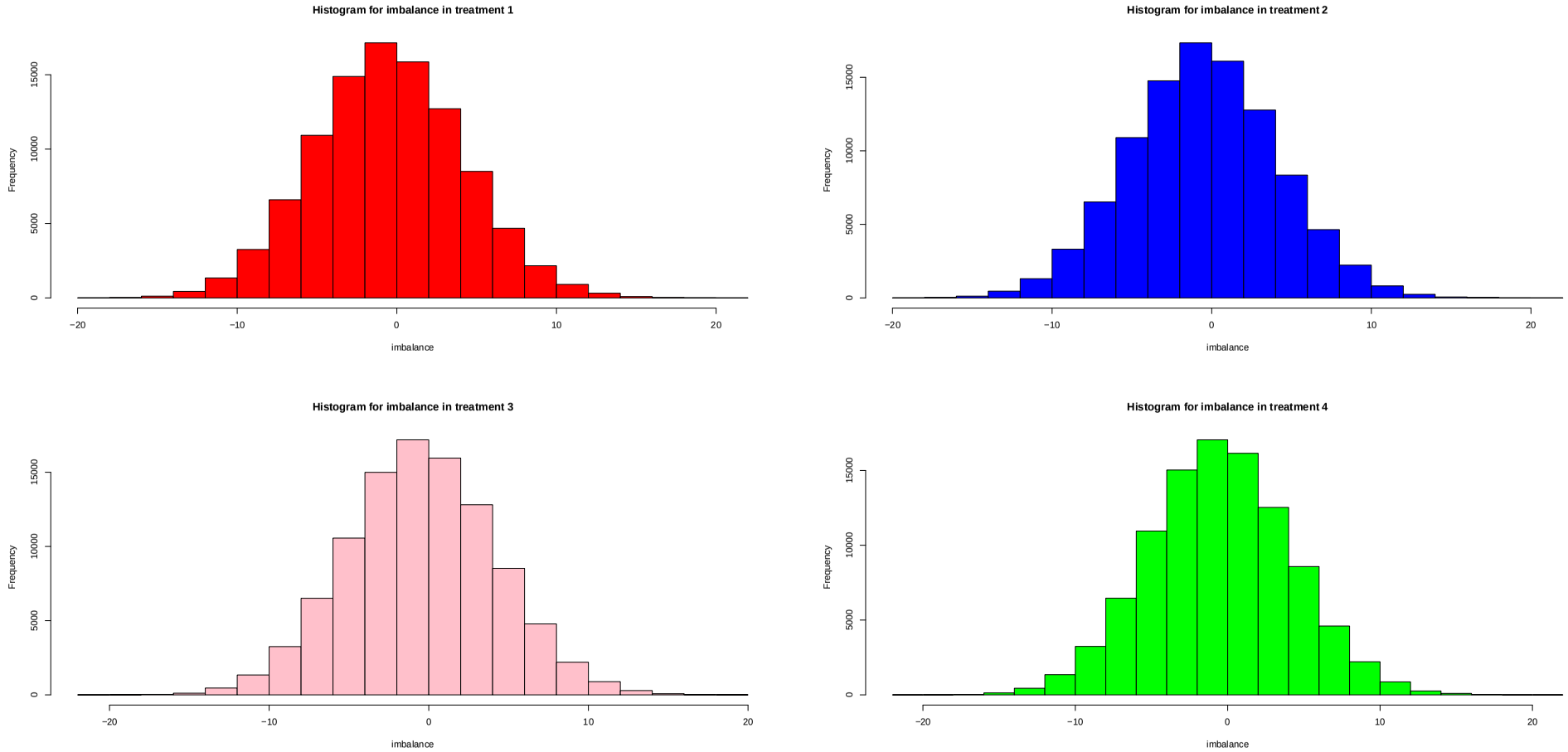
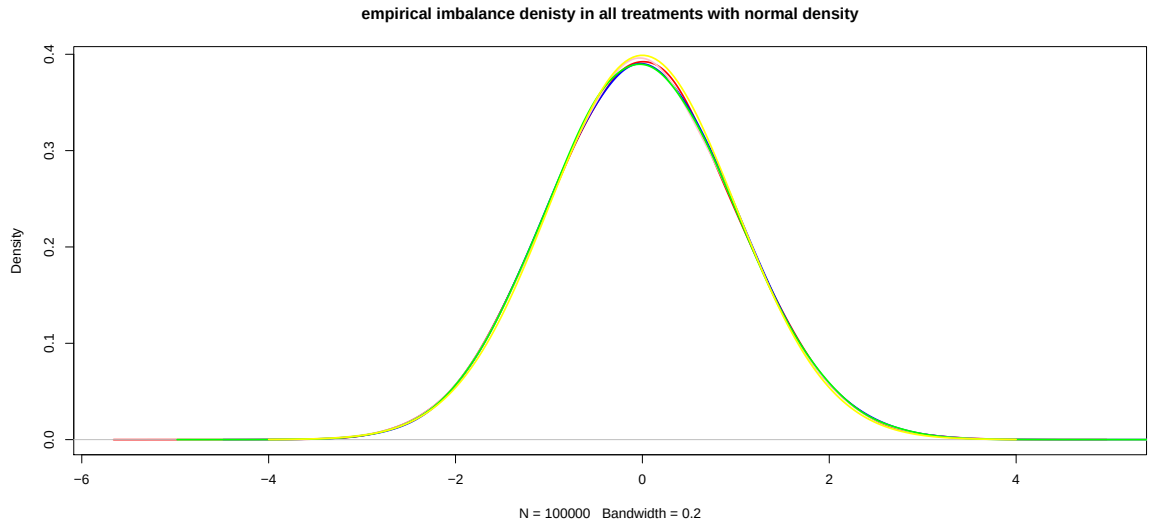


Figure 5.2: Histograms for simulated values of Δ_j for $j = 1, ..., 4$ for the second scenario



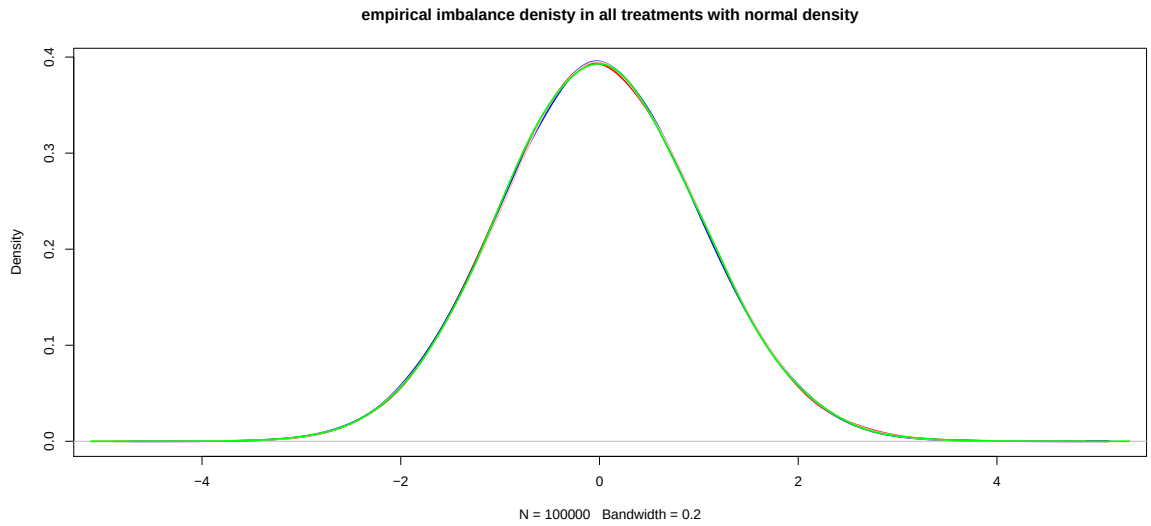
The curves for the empirical density functions of $\Delta_j/\sqrt{\text{Var}(\Delta_j)}$ for $j = 1, ..., 4$ are also shown in the two figures below. It is clear that, in both scenarios, the curves for the empirical density functions coincide with that for the density function of the standard normal distribution.

Figure 5.3: Empirical density functions of $\Delta_j/\sqrt{\text{Var}(\Delta_j)}$ for $j = 1, \ldots, 4$ for the first scenario



The yellow line represents the density function of the standard normal distribution. The red, blue, pink and green lines represent the values of $\Delta_j/\sqrt{\text{Var}(\Delta_j)}$ for $j = 1, 2, 3, 4$, respectively.

Figure 5.4: Empirical density functions of $\Delta_j/\sqrt{\text{Var}(\Delta_j)}$ for $j = 1, \ldots, 4$ for the second scenario



The yellow line represents the density function of the standard normal distribution. The red, blue, pink and green lines represent the values of $\Delta_j/\sqrt{\text{Var}(\Delta_j)}$ for $j = 1, 2, 3, 4$, respectively.

5.7 Power and sample size

Here, we will study the impact of imbalance on treatments on the power for centre-stratified randomization. We will also see in the imbalanced case how an increase in the sample size can compensate for the loss in power. This will be shown by numerical results from simulation.

Let n be the total number of patients to be randomized to $K > 2$ treatments at N clinical centres. Centre-stratified randomization has blocks of size K_1. Let n_j be the number of patients randomized to treatment j for $j = 1, ..., K$ and $\bar{X}_j$ be the mean of the patient responses on treatment j. Assume that the observations are independent normal with unknown means m_j for $j = 1, ..., K$ and known variance σ^2. Suppose that the first group of patients receive the standard treatment and let this group be the control group.

We will be carrying out $K-1$ tests assuming no centre effect.

Consider testing the null hypothesis $H_0 : m_1 = m_2 = ... = m_K = m$ for any non-negative constant m against H_1 : at least one $m_j - m_1 = h_j > 0$ for any $j = 2, ..., K$. The test statistics are

$$S_j = \frac{\bar{X}_j - \bar{X}_1}{\sigma\sqrt{1/n_j + 1/n_1}}$$

for $j = 2, ..., K$. Under H_0, the S_j are dependent standard normal random variables. Given γ as the significance level of one test, let Φ be the standard normal distribution function and let z_γ satisfy $1 - \Phi(z_\gamma) = \gamma$. We will reject H_0 if, for at least one j with $j = 2, ..., K$, $S_j > z_\gamma$. Let the significance level of the overall test be γ^*. We can represent the significance level as

$$P\left\{\cup_{j=2}^{K}\left(S_j > z_\gamma\right)|H_0\right\} = \gamma^*.$$

Under H_1, we will have

$$\frac{\bar{X}_j - \bar{X}_1}{\sigma\sqrt{1/n_j + 1/n_1}} - \frac{h_j}{\sigma\sqrt{1/n_j + 1/n_1}} = S_j - \frac{h_j}{\sigma\sqrt{1/n_j + 1/n_1}}$$

as a standard normal random variable for each $j = 2, ..., K$. Let β^* be the probability of a type II error. The power is $1 - \beta^*$. For a given level of significance γ, the power can be written as

$$P\left\{\cup_{j=2}^{K}\left(S_j - \frac{h_j}{\sigma\sqrt{1/n_j + 1/n_1}} > z_\gamma - \frac{h_j}{\sigma\sqrt{1/n_j + 1/n_1}}\right)\middle| H_1\right\} = 1 - \beta^*.$$

Now let

$$\begin{aligned}\eta_j &= \frac{\bar{X}_j - \bar{X}_1}{\sigma\sqrt{1/n_j + 1/n_1}} - \frac{h_j}{\sigma\sqrt{1/n_j + 1/n_1}} \\ &= \frac{\sqrt{n_j}(\bar{X}_j - m_j)}{\sigma}\frac{1}{\sqrt{n_j}\sqrt{1/n_j + 1/n_1}} - \frac{\sqrt{n_1}(\bar{X}_1 - m_1)}{\sigma}\frac{1}{\sqrt{n_1}\sqrt{1/n_j + 1/n_1}}\end{aligned}$$

for $j = 2, ..., K$. The power of the test can be written as

$$\begin{aligned} & P\left\{\cup_{j=2}^{K}\left(\eta_j > z_\gamma - \frac{h_j}{\sigma\sqrt{1/n_j + 1/n_1}}\right)\middle| H_1\right\} \\ = \; & 1 - P\left\{\cap_{j=2}^{K}\left(\eta_j \leq z_\gamma - \frac{h_j}{\sigma\sqrt{1/n_j + 1/n_1}}\right)\middle| H_1\right\} \\ = \; & 1 - \beta^*. \end{aligned}$$

Under H_1, for each $j = 2, ..., K$,

$$\zeta_j = \frac{\sqrt{n_j}(\bar{X}_j - m_j)}{\sigma}$$

and

$$\zeta_1 = \frac{\sqrt{n_1}(\bar{X}_1 - m_1)}{\sigma}$$

are two independent standard normal random variables. So each η_j can be written as a linear combination of two independent normal random variables, since

$$\eta_j = \zeta_j c_{j1} - \zeta_1 c_{j2}, \tag{5.34}$$

where c_{j1} and c_{j2} are scalars such that

$$c_{j1} = \frac{1}{\sqrt{1 + \frac{n_j}{n_1}}} \tag{5.35}$$

and

$$c_{j2} = \frac{1}{\sqrt{1 + \frac{n_1}{n_j}}}. \tag{5.36}$$

The form in (5.34) for η_j can be used to simulate values for η_j and hence simulate numerical values for the power of the test.

5.7.1 Impact of imbalance on power and sample size

Here, we will study the power of the test by simulation for four scenarios when $\gamma = 0.05$ and $\gamma = 0.05/(K-1)$. First, consider the balanced case where we have the

same number of patients on each of the treatments. We will simulate the test $100,000$ times to find n, the total number of patients to be recruited for each of the scenarios such that the power is at least 0.95. In each simulation, we will generate n_j patient responses on treatment j from a normal distribution with mean m_j and variance σ^2 for $j = 2, ..., K$. We take the values $m_1 = 0$, $m_2 = 0.5$, $m_3 = 0.55$, $m_4 = 0.6$, $m_5 = 0.65$, $m_6 = 0.7$, $m_7 = 0.75$ and $m_8 = 0.8$ in each of the scenarios, and $\sigma^2 = 1$. Then we can obtain the sample mean of the responses $\bar{X}_j$ and hence values of ζ_j for $j = 1, ..., K$. Finally, the values for η_j are obtained from the linear combination of ζ_j and ζ_1 for $j = 2, ..., K$. The values for c_{j1} and c_{j2} can be calculated in the balanced case, since $n_1 = n_2 = ... = n_K$. By (5.35) and (5.36), $c_{j1} = c_{j2} = 1/\sqrt{2}$. If at least one of the η_j for $j = 2, ..., K$ is greater than z_γ, we will reject H_0 for this simulation. After $100,000$ simulations, the proportion of rejections of H_0 will be the estimated power of the test.

Let K be the total number of treatments involved in the study and let K_1 be the size of a complete block. In the balanced case, the number of patients on each treatment will be n/K. We will find n for each scenario such that there are at least $95,000$ rejections of H_0 in a total of $100,000$ simulations.

Once n is known for each scenario, we can study the estimated power in the imbalanced case to see how much less it is compared to the balanced case and how many extra patients we need to compensate for the loss. Under centre-stratified permuted-block randomization when the patient recruitment process is modelled by the Poisson-gamma model, the number of patients to be recruited in centre i, n_i, may not be a multiple of the block size K_1. Incomplete blocks will be formed in centres and the numbers of patients across treatments will not be equal. For simulation in the imbalanced case, for each scenario, we will use the same n as in the balanced case. We will first model the patient recruitment process by the Poisson-gamma model. Once the numbers of patients to be

recruited in different centres are known, the numbers of patients on different treatments for all centres are known. Then we will simulate again n_j responses from a normal distribution with mean m_j and variance σ^2. The values for c_{j1}, c_{j2}, ζ_j and ζ_1, and hence η_j can be obtained. As in the balanced case, in each simulation for all $j = 2, ..., K$, it will be checked if at least one of the η_j is greater than z_γ. The whole process will be simulated $100,000$ times and the proportion of rejections of H_0 in these $100,000$ runs will be the estimated power for the imbalanced case. We will also simulate imbalanced cases when extra patients are added to the trial and study the numerical values of the estimated power in these cases.

We will consider four scenarios, each with a different number of treatments. We have $N = 5$ for the total number of centres. All the numerical results for the estimated overall significance level and power are summarised in the two tables below.

Table 5.1: Simulated overall significance levels and powers for four scenarios when $\gamma = 0.05$

	$K = 4$ $K_1 = 8$	$K = 5$ $K_1 = 10$	$K = 6$ $K_1 = 12$	$K = 8$ $K_1 = 8$
Balanced case				
n	168	160	168	152
γ^*	0.121	0.147	0.168	0.205
Power	0.957	0.952	0.961	0.956
Imbalanced case				
n	168	160	168	152
γ^*	0.118	0.143	0.166	0.207
Power	0.956	0.950	0.959	0.953
n	169	161	169	153
γ^*	0.118	0.144	0.166	0.203
Power	0.956	0.951	0.959	0.956
n	170	162	170	154
γ^*	0.118	0.144	0.166	0.202
Power	0.958	0.951	0.960	0.956
n	171	163	171	155
γ^*	0.117	0.143	0.168	0.203
Power	0.958	0.952	0.961	0.957

Table 5.2: Simulated overall significance levels and powers for four scenarios when $\gamma = 0.05/(K-1)$

	$K=4$ $K_1=8$	$K=5$ $K_1=10$	$K=6$ $K_1=12$	$K=8$ $K_1=8$
Balanced case				
n	232	250	264	280
γ^*	0.045	0.042	0.041	0.039
Power	0.954	0.952	0.952	0.952
Imbalanced case				
n	232	250	264	280
γ^*	0.043	0.041	0.040	0.037
Power	0.953	0.950	0.951	0.952
n	233	251	265	281
γ^*	0.044	0.042	0.039	0.038
Power	0.954	0.950	0.950	0.951
n	234	252	266	282
γ^*	0.042	0.041	0.040	0.037
Power	0.954	0.950	0.952	0.953
n	235	253	267	283
γ^*	0.043	0.042	0.040	0.037
Power	0.954	0.952	0.951	0.953

We can see from the tables that the test is anticonservative when $\gamma = 0.05$ and slightly conservative when $\gamma = 0.05/(K-1)$, becoming more so as K increases, and that there is a slight loss in power in the imbalanced case compared to the balanced case

when n is the same. No more than three patients are needed in the imbalanced case in order to achieve the same level of power as in the balanced case.

5.8 Conclusions

In this chapter, imbalances are defined for complete randomization and the permuted block design for clinical trials with more than two treatments. Most of the literature studied previously dealt with imbalance properties for different randomization schemes for two treatment groups only. The overall imbalance for more than two treatments is no longer the difference in the numbers of patients on two treatments, but a vector that contains the imbalance for each treatment group from the expected number of patients in the group.

The imbalance for each of the treatments is defined here within centres and for all centres. The calculations of the expectations, variances and covariances of the imbalances in a centre or for all centres are shown for the two randomization schemes. Two cases are considered: the number of patients recruited in a centre is given and is known; and the number of patients recruited in a centre follows the Poisson-gamma model of Anisimov and Fedorov (2007). The overall imbalance is defined and asymptotically it has a multivariate normal distribution. Furthermore, the variances of the two randomization schemes are compared. For trials with several treatments, in general, centre-stratified permuted-block randomization performs better than complete randomization in balancing the numbers of patients across treatment groups. In other words, complete randomization provides more uncertainty in the numbers of patients in different treatment groups. A test is also developed by comparing each treatment to the control group. The numerical values for the power of the test are given in the balanced and imbalanced cases. These suggest that we can compensate for the loss in power in the imbalanced

case by adding no more than three patients.

In real clinical trials, it is common for pharmaceutical companies to have more than two treatments under investigation at the same time in order to obtain more results for comparison, and to reduce costs and time to recruit patients in a trial to expedite earlier availability of an effective treatment to the general public. The imbalance properties for the several treatments case are of particular importance as they will affect the power of the test for treatment differences. The more serious the imbalance is in a trial, the less power it will have to detect a genuine treatment difference. The loss in power due to the imbalance can be compensated for by an increase in the sample size.

Chapter 6

Discussion

6.1 Conclusions

In this thesis, we have studied three types of randomization schemes, restricted randomization schemes, covariate-adaptive randomization schemes and the permuted-block randomization scheme.

In Chapter 2, the treatment assignment rules under different restricted randomization schemes are given. In addition, the asymptotic properties are stated under these randomization schemes. The theoretical power of a test for a treatment effect is obtained for both complete randomization and Efron's biased coin design for normal and binary responses. The parameters of interest are the population mean difference for normal responses and the population difference in probabilities of survival for binary responses. The test is for a treatment effect represented by a larger value of the population mean of patients' responses in the treatment group than in the control group for normal responses. For binary responses, the test is for a treatment effect represented by a larger value of the probability of surviving in the treatment group than in the control group. In particular, for normal responses, we studied the cases where the variances of the pa-

tients' responses in the two treatment groups are different when they are either known or unknown. Efron's biased coin design is more efficient at balancing the numbers of patient on the two treatments than complete randomization. We see that the numerical values obtained for the power are higher for Efron's biased coin design than for complete randomization in all cases. Simulation results for the power under the adjustable biased coin design and Efron's biased coin design for normal responses are also given, which show that the adjustable biased coin design is as good as Efron's biased coin design in balancing the numbers of patients on the two treatments. Finally, numerical values for the power obtained by a normal approximation are compared with the exact values under complete randomization and Efron's biased coin design. For all cases, the power obtained by the normal approximation is higher than the actual power.

In Chapter 3, covariate-adaptive randomization schemes are introduced. They are covariate-adaptive simple random sampling, the covariate-adaptive biased coin design and the covariate-adaptive adjustable biased coin design. A normal linear model for patients' responses involving covariates is considered. This model takes into account the values of the covariates and their interactions. Then the methodologies for global and marginal balancing the numbers of patients grouped by their prognostic factors are given for different covariate-adaptive randomization schemes. Next, we detail the theoretical calculations for the analysis of covariance t test under the fixed-effects linear model. In addition, normal approximations to the power are given for the two-sample t test and the analysis of covariance t test when the covariates are considered to be random variables. Numerical values for the power obtained by normal approximations with one covariate in the model are compared with the corresponding values obtained by simulation for the two-sample t test and the analysis of covariance t test. In both tests, the powers obtained by normal approximations are higher than the simulated values. Furthermore, the nor-

mal approximation to the power for more than two covariates is given theoretically. We also simulated the power under the fixed-effects model for either global or marginal balancing when interactions between the two covariates are both present and absent. They show that global balancing is better than marginal balancing when there is an interaction between the covariates under the three covariate-adaptive randomization schemes. When there is no interaction between the covariates, global balancing is as good as marginal balancing.

In Chapter 4, restricted randomization schemes for more than two treatments are considered. The assignment rules under complete randomization, Efron's biased coin design, the D_A-optimum biased coin design, Wei's class of biased coin design and a new class of designs, the adjustable biased coin design, are given. In addition, their asymptotic properties are provided. In this chapter, the overall imbalance is defined as a vector which contains the imbalances in the numbers of patients on all treatments. The properties of the overall imbalance have been studied under different randomization schemes by simulation. The distributional properties of the imbalances represented by plots and the values for the quartiles of the imbalances under different designs are also given for two scenarios. The simulated results show that, except for Efron's biased coin design, the overall imbalance has an asymptotically multivariate normal distribution. For Efron's bised coin design, the overall imbalance tends to the zero vector in probability asymptotically. Furthermore, we have studied the simulated power of these randomization schemes for the analysis of variance F test. Numerical values for the power are given for two scenarios. The results showed that, for all schemes, the power increases when we consider more treatments in the trial. Efron's biased coin design gives the highest power when the design is very deterministic with $p = 11/12$ and $10/12$. Other than these two situations, the adjustable biased coin design achieves the highest power in all

cases in the two scenarios.

In Chapter 5, we consider situations for multi-centre trials for more than two treatments. The model for patient recruitment introduced by Anisimov and Fedorov (2007) is used. The recruitment process is described by a Poisson process with rate following a gamma distribution. By using this patient recruitment model, the number of patients to be recruited in each centre follows a beta-binomial distribution. The imbalance on each treatment is defined both within centres and for all centres under complete randomization and centre-stratified permuted-block randomization. We studied the imbalance of each treatment within centres and for all centres when the number of patients recruited in each centre is fixed or has a beta-binomial distribution. The expectations and the variances of the imbalances are given theoretically under the two schemes for the two situations. The overall imbalance which is defined as a vector which contains the imbalance on each treatment for all centres has a multivariate normal distribution when the number of centres involved is large. Hence, we have the vectors of expectations and the covariance matrices of these overall imbalances for the two situations. The variances of the imbalance on a treatment within centres are compared for complete randomization and centre-stratified permuted-block randomization. We showed that, under centre-stratified permuted-block randomization, the imbalance is less variable. Furthermore, we studied the power under centre-stratified permuted-block randomization. We have a test for at least one treatment difference when we compare the treatment groups with the control group. We found the sample size for a balanced trial in each scenario for which a certain level of power is achieved. Then we studied the change in the power in the imbalanced case when the patient recruitment model is used. The results showed that the loss of power in the imbalanced case can be compensated for by an increase of no more than three patients for each scenario.

6.2 Future work

The results presented in this thesis provide a basis for further study and possible extensions. Some of the ideas for further work are mentioned throughout the thesis.

In Chapter 2, the theoretical power function is given for binary responses for complete randomization and Efron's biased coin design. The critical value used to reject the null hypothesis is chosen so that the significance level is less than or equal to the assumed level $\alpha = 0.05$. This is due to the discrete distribution of the test statistics. Further research can be carried out so that we obtain the critical value corresponding to the assumed significance level of the test. One possibility is a randomized test.

In Chapter 3, covariate-adaptive randomization schemes are studied. When the covariates are considered to be random variables, an expression is given for the normal approximation to the power when there is more than one covariate. Simulated values for the power can be compared with numerical values obtained using the normal approximation in this case. These results can also be compared with those obtained in Section 3.6.2, when there is only one covariate. Furthermore, under the fixed-effects model, numerical values for the power are shown by simulation for global and marginal balance with two covariates when interactions between the covariates are both present and absent. Further work may be possible for more than two covariates.

In Chapter 4, a new class of designs called the adjustable biased coin design for more than two treatments has been introduced. The treatment assignment probabilities under this design are obtained based on all of the current imbalances on the treatments. Using simulation, the imbalance properties and the power are obtained under this design. These suggest that this design is as good as Efron's biased coin design and the D_A-optimum biased coin design in balancing the numbers of patients across treatments. Further research can be pursued on the asymptotic properties of the adjustable biased

coin design. In particular, the theoretical power of the test for treatment differences for Efron's biased coin design and the adjustable biased coin design is of interest.

In Chapter 5, we studied the imbalance properties of centre-stratified permuted-block randomization in the case of more than two treatments. We assumed that, within a block, the number of patients to be allocated to each of the treatments is the same. Thus, we aim for a balance in the numbers of patients across treatments to maximize the power of the test for treatment differences. However, it may be more powerful in some cases to have unequal numbers of patients on the treatments. Therefore, further research in this direction is possible. Moreover, the test for comparing each of the treatments with the control does not have the same significance level for all situations. Further work can be carried out on different tests for treatment differences with a chosen significance level, and hence the power of these tests. We have only considered the test when the variances of the patients' responses are known and are the same. This can be extended to cases where the variances are different and unknown.

Appendix A

Supplementary information for Chapter 3

A.1 Uncorrelated $\widehat{\boldsymbol{\mu}}^*$ and $\widehat{\boldsymbol{\beta}}$

We know that

$$\begin{pmatrix} \widehat{\boldsymbol{\mu}}^* \\ \widehat{\boldsymbol{\beta}} \end{pmatrix} = \begin{pmatrix} \mathbf{W}^T\mathbf{W} & \mathbf{W}^T R_W \mathbf{Z} \\ \mathbf{Z}^T R_W \mathbf{W} & \mathbf{Z}^T R_W \mathbf{Z} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{W}^T\mathbf{Y} \\ \mathbf{Z}^T R_W \mathbf{Y} \end{pmatrix}.$$

It follows that

$$\begin{aligned} \mathrm{Cov}(\widehat{\boldsymbol{\mu}}^*, \widehat{\boldsymbol{\beta}}) &= (\mathbf{W}^T\mathbf{W})^{-1}\mathrm{Cov}\left(\mathbf{W}^T\mathbf{Y}, \mathbf{Z}^T R_W \mathbf{Y}\right)(\mathbf{Z}^T R_W \mathbf{Z})^{-1} \\ &= (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathrm{Var}(\mathbf{Y})R_W\mathbf{Z}(\mathbf{Z}^T R_W \mathbf{Z})^{-1}, \end{aligned}$$

where $\mathrm{Var}(\mathbf{Y})$ is the $n \times n$ covariance matrix of the patients' responses with σ_ϵ^2 as the diagonal elements and zeros elsewhere, denoted by $\sigma_\epsilon^2\mathbf{I}_n$. Thus, we have

$$\mathrm{Cov}(\widehat{\boldsymbol{\mu}}^*, \widehat{\boldsymbol{\beta}}) = \sigma_\epsilon^2(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T R_W \mathbf{Z}(\mathbf{Z}^T R_W \mathbf{Z})^{-1}.$$

Since $\mathbf{W}^T R_W \mathbf{Z}$ gives a null matrix, the covariance of $\widehat{\boldsymbol{\mu}}^*$ and $\widehat{\boldsymbol{\beta}}$ is the zero matrix. Hence, $\widehat{\boldsymbol{\mu}}^*$ and $\widehat{\boldsymbol{\beta}}$ are uncorrelated.

A.2 The independence of Z and $\widehat{\beta}$ when the covariates are random variables

Since $\widehat{\boldsymbol{\beta}}$ is the least squares estimator of $\boldsymbol{\beta}$, we have

$$\widehat{\boldsymbol{\beta}} = (\mathbf{Z}^T R_W \mathbf{Z})^{-1} \mathbf{Z}^T R_W \mathbf{Y},$$

where $R_W = \mathbf{I}_n - \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T$. If $\mathbf{Z}$ and $\widehat{\boldsymbol{\beta}}$ are independent, then $E(\mathbf{Z}\widehat{\boldsymbol{\beta}}) = E(\mathbf{Z})E(\widehat{\boldsymbol{\beta}})$. Now, we have

$$\begin{aligned} E(\mathbf{Z}\widehat{\boldsymbol{\beta}}) &= E\Big(\mathbf{Z}\left[\mathbf{Z}^T\left\{\mathbf{I}_n - \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\right\}\mathbf{Z}\right]^{-1} \\ &\quad \times \mathbf{Z}^T\left\{\mathbf{I}_n - \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\right\}(\mathbf{W}\boldsymbol{\mu} + \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon})\Big) \end{aligned}$$

Expanding gives

$$\begin{aligned} E(\mathbf{Z}\widehat{\boldsymbol{\beta}}) &= E\Big[\mathbf{Z}\Big\{\mathbf{Z}^T\mathbf{Z} - \mathbf{Z}^T\mathbf{W}\left(\mathbf{W}^T\mathbf{W}\right)^{-1}\mathbf{W}^T\mathbf{Z}\Big\}^{-1}\left\{\mathbf{Z}^T\mathbf{W}\boldsymbol{\mu} + \mathbf{Z}^T\mathbf{Z}\boldsymbol{\beta} + \mathbf{Z}^T\boldsymbol{\epsilon}\right. \\ &\quad \left. - \mathbf{Z}^T\mathbf{W}\boldsymbol{\mu} - \mathbf{Z}^T\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{Z}\boldsymbol{\beta} - \mathbf{Z}^T\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\boldsymbol{\epsilon}\right\}\Big], \end{aligned}$$

which reduces to

$$\begin{aligned} E(\mathbf{Z}\widehat{\boldsymbol{\beta}}) &= E\Big[\mathbf{Z}\left\{\mathbf{Z}^T\mathbf{Z} - \mathbf{Z}^T\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{Z}\right\}^{-1} \\ &\quad \times \left\{\mathbf{Z}^T\mathbf{Z} - \mathbf{Z}^T\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{Z}\right\}\left\{\boldsymbol{\beta} + \mathbf{Z}^T(\mathbf{Z}\mathbf{Z}^T)^{-1}\boldsymbol{\epsilon}\right\}\Big]. \end{aligned}$$

Finally, we obtain

$$E(\mathbf{Z}\widehat{\boldsymbol{\beta}}) = E\left\{\mathbf{Z}\left(\boldsymbol{\beta} + \mathbf{Z}^T(\mathbf{Z}\mathbf{Z}^T)^{-1}\boldsymbol{\epsilon}\right)\right\} = E(\mathbf{Z}\boldsymbol{\beta}) = E(\mathbf{Z})\boldsymbol{\beta}.$$

Hence, we have $E(\mathbf{Z}\widehat{\boldsymbol{\beta}}) = E(\mathbf{Z})\boldsymbol{\beta} = E(\mathbf{Z})E(\widehat{\boldsymbol{\beta}})$.

A.3 Normal approximation for the power when there are more than two randomly distributed covariates

From the Appendix of Shao, Yu and Zhong (2010), we know that

$$\frac{n_j}{n} - \frac{1}{2} = o_p(n^{-1/2})$$

for $j = 1, 2$ conditionally on $\mathcal{Z}$. Further, $(\Delta_{ik}, \epsilon_{i1}, \epsilon_{i2})$ are conditionally independent of $\mathcal{I}$ given $\mathcal{Z}$, $E(\epsilon_{ij}|\mathcal{Z}) = 0$, $E(2I_i - 1|\mathcal{Z}) = 0$ and $E(\Delta_{ik}|\mathcal{Z}) = E(\Delta_{ik}|D(Z_{ik})) = 0$. The asymptotic mean of $\bar{Y}_2 - \bar{Y}_1$ is calculated by taking the expectation of $\bar{Y}_2 - \bar{Y}_1$ given $\mathcal{Z}$. It is also known that $E(Z_{ik}|D(Z_{ik}))$, $E(Z_{ik}^2|D(Z_{ik}))$ and $E(Z_{ik}Z_{im}|D(Z_{ik}))$ are discrete, and that $\sum_{i=1}^{n}(2I_i - 1) = o_p(n^{-1/2})$. When $n \to \infty$, $E(\bar{Y}_2 - \bar{Y}_1|\mathcal{Z}) = \mu_2 - \mu_1$.

Now, we can write

$$\bar{Y}_2 - \bar{Y}_1 = \mu_2 - \mu_1 + \frac{2}{n}(A + B + C + D + E + F + G) + o_p(n^{-1/2}),$$

where A, B, C, D, E,F and G are the terms in order in (3.11). We need the variance of $\bar{Y}_2 - \bar{Y}_1$ given $\mathcal{Z}$. From above, this will be a sum of the variances of all the terms involved and all the covariances of any two different terms. The term G is expressed in terms of the random errors ϵ_{ij}. Since $(\Delta_{ik}, \epsilon_{i1}, \epsilon_{i2})$ are independent of $\mathcal{I}$ given $\mathcal{Z}$ and the expectation of ϵ_{ij} given $\mathcal{Z}$ is zero, the covariance of any term with G will be equal to

zero. Also, $\mu_2 - \mu_1$ is a constant, so that $\mathrm{Var}(\mu_2 - \mu_1|\mathcal{Z}) = 0$. Thus, we have

$$\begin{aligned}
\mathrm{Var}(\bar{Y}_2 - \bar{Y}_1|\mathcal{Z}) &= \frac{4}{n^2}\{\mathrm{Var}(A|\mathcal{Z}) + \mathrm{Var}(B|\mathcal{Z}) + \mathrm{Var}(C|\mathcal{Z}) + \mathrm{Var}(D|\mathcal{Z}) \\
&+ \mathrm{Var}(E|\mathcal{Z}) + \mathrm{Var}(F|\mathcal{Z}) + \mathrm{Var}(G|\mathcal{Z}) + 2\mathrm{Cov}(A, B|\mathcal{Z}) \\
&+ 2\mathrm{Cov}(A, C|\mathcal{Z}) + 2\mathrm{Cov}(A, D|\mathcal{Z}) + 2\mathrm{Cov}(A, E|\mathcal{Z}) \\
&+ 2\mathrm{Cov}(A, F|\mathcal{Z}) + 2\mathrm{Cov}(B, C|\mathcal{Z}) + 2\mathrm{Cov}(B, D|\mathcal{Z}) \\
&+ 2\mathrm{Cov}(B, E|\mathcal{Z}) + 2\mathrm{Cov}(B, F|\mathcal{Z}) + 2\mathrm{Cov}(C, D|\mathcal{Z}) \\
&+ 2\mathrm{Cov}(C, E|\mathcal{Z}) + 2\mathrm{Cov}(C, F|\mathcal{Z}) + 2\mathrm{Cov}(D, E|\mathcal{Z}) \\
&+ 2\mathrm{Cov}(D, F|\mathcal{Z}) + 2\mathrm{Cov}(E, F|\mathcal{Z})\} + o_p(n^{-1}).
\end{aligned}$$

Note that $(2I_i - 1)^2 = 4I_i^2 - 4I_i + 1 = 1$, because $I_i^2 = I_i$

We can split the above expression into two parts. We have one part that contains all the variance terms and one that contains all covariance terms. Consider the first part. As an example, we take $\mathrm{Var}(A|\mathcal{Z})$, which is equal to

$$\begin{aligned}
\mathrm{Var}\left\{\sum_{i=1}^{n}\sum_{k=1}^{p}(2I_i - 1)b_k\Delta_{ik}|D(Z_{ik})\right\} &= \sum_{k=1}^{p}\sum_{i=1}^{n} b_k^2\mathrm{Var}\left\{\Delta_{ik}|D(Z_{ik})\right\} \\
&+ \sum_{\substack{k,l=1,\\ l\neq k}}^{p}\sum_{i=1}^{n} b_k b_l E\{\Delta_{ik}\Delta_{il}|D(Z_{ik})\}.
\end{aligned}$$

We have similar results for $\mathrm{Var}(C|\mathcal{Z})$ and $\mathrm{Var}(D|\mathcal{Z})$. For $\mathrm{Var}(G|\mathcal{Z})$, we have $n\sigma_\epsilon^2$. As $E\{Z_{ik}|D(Z_{ik})\}$, $E\{Z_{ik}^2|D(Z_{ik})\}$ and $E\{Z_{ik}Z_{im}|D(Z_{ik})\}$ are constants given $\mathcal{Z}$, the

variance of the above three terms equals 0. So first part of the variance of $\bar{Y}_2 - \bar{Y}_1$ is

$$
\begin{aligned}
& \frac{4}{n}\sum_{k=1}^{p} b_k^2 \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}\{\Delta_{ik}|D(Z_{ik})\} + \frac{4}{n}\sum_{\substack{k,l=1,\\ l\neq k}}^{p} b_k b_l \frac{1}{n}\sum_{i=1}^{n} E\{\Delta_{ik}\Delta_{il}|D(Z_{ik})\} \\
+ \quad & \frac{4}{n}\sum_{k=1}^{p} c_k^2 \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}\{\Delta_{ik}^2|D(Z_{ik})\} + \frac{4}{n}\sum_{\substack{k,l=1,\\ l\neq k}}^{p} c_k c_l \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}\{\Delta_{ik}\Delta_{il}|D(Z_{ik})\} \\
+ \quad & \frac{4}{n}\sum_{\substack{k,m=1,\\ k<m}}^{p} d_{km}^2 \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}\{\Delta_{ik}\Delta_{im}|D(Z_{ik})\} \\
+ \quad & \frac{4}{n}\sum_{\substack{k,m,s,r=1,\\ k<m<s<r}}^{p} d_{km}d_{sr} \frac{1}{n}\sum_{i=1}^{n} E\{\Delta_{ik}\Delta_{im}\Delta_{is}\Delta_{ir}|D(Z_{ik})\} \\
+ \quad & \frac{4}{n}\sum_{\substack{k,m,s=1,\\ k<m<s}}^{p} d_{km}d_{ms} \frac{1}{n}\sum_{i=1}^{n} E\{\Delta_{ik}\Delta_{im}^2\Delta_{is}|D(Z_{ik})\} \\
+ \quad & \frac{4}{n}\sum_{\substack{k,m,s=1,\\ k<m<s}}^{p} d_{km}d_{ks} \frac{1}{n}\sum_{i=1}^{n} E\{\Delta_{ik}^2\Delta_{im}\Delta_{is}|D(Z_{ik})\} \\
+ \quad & \frac{4}{n}\sum_{\substack{k,m,s=1,\\ k<m<s}}^{p} d_{ks}d_{ms} \frac{1}{n}\sum_{i=1}^{n} E\{\Delta_{ik}\Delta_{im}\Delta_{is}^2|D(Z_{ik})\} + \frac{4\sigma_\epsilon^2}{n} \qquad \text{(A.1)}
\end{aligned}
$$

For the second part involving the covariances, we take $\mathrm{Cov}(A, B|\mathcal{Z})$ and $\mathrm{Cov}(A, C|\mathcal{Z})$ as examples. Now, $E(AB|\mathcal{Z}) = BE(A|\mathcal{Z})$ and $E\{A|D(Z_{ik})\} = 0$, so we have $\mathrm{Cov}(A, B|\mathcal{Z}) = 0$.

Since $E\{A|D(Z_{ik})\} = 0$, we have $\mathrm{Cov}(A, C|\mathcal{Z}) = E\{AC|D(Z_{ik})\}$. It follows that

$$
\begin{aligned}
2\mathrm{Cov}(A, C|\mathcal{Z}) \quad = \quad & \frac{8}{n}\sum_{k=1}^{p} b_k c_k \frac{1}{n}\sum_{i=1}^{n} E\{\Delta_{ik}^3|D(Z_{ik})\} \\
+ \quad & \frac{8}{n}\sum_{\substack{k,l=1,\\ l\neq k}}^{p} b_k c_l \frac{1}{n}\sum_{i=1}^{n} E\{\Delta_{ik}\Delta_{il}^2|D(Z_{ik})\}. \qquad \text{(A.2)}
\end{aligned}
$$

The rest of the second part of the variance of $\bar{Y}_2 - \bar{Y}_1$ can be written as

$$
\begin{aligned}
& \frac{8}{n} \sum_{\substack{k,m=1,\\ k<m}}^{p} b_k d_{km} \frac{1}{n} \sum_{i=1}^{n} E\{\Delta_{ik}^2 \Delta_{im} | D(Z_{ik})\} \\
+ \; & \frac{8}{n} \sum_{k=1}^{p} b_k c_k \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}\{\Delta_{ik} | D(Z_{ik})\} E\{Z_{ik} | D(Z_{ik})\} \\
+ \; & \frac{8}{n} \sum_{\substack{k,m=1,\\ k<m}}^{p} b_k d_{km} \frac{1}{n} \sum_{i=1}^{n} E\{Z_{ik} | D(Z_{ik})\} E\{\Delta_{ik} \Delta_{im} | D(Z_{ik})\} \\
+ \; & \frac{8}{n} \sum_{k=1}^{p} c_k^2 \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}\{Z_{ik} | D(Z_{ik})\} \mathrm{Var}\{\Delta_{ik} | D(Z_{ik}\} \\
+ \; & \frac{8}{n} \sum_{\substack{k,m=1,\\ k<m}}^{p} c_k d_{km} \frac{1}{n} \sum_{i=1}^{n} E\{\Delta_{ik}^3 \Delta_{im} | D(Z_{ik})\} \\
+ \; & \frac{8}{n} \sum_{\substack{k,m=1,\\ k<m}}^{p} c_k d_{km} \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}\{Z_{ik} | D(Z_{ik})\} E\{\Delta_{ik} \Delta_{im} | D(Z_{ik})\} \\
+ \; & \frac{8}{n} \sum_{\substack{k,m,s=1,\\ k<m<s}}^{p} b_k d_{ms} \frac{1}{n} \sum_{i=1}^{n} E\{\Delta_{ik} \Delta_{im} \Delta_{is} | D(Z_{ik})\} \\
+ \; & \frac{8}{n} \sum_{\substack{k,l=1,\\ k\neq l}}^{p} b_k c_l \frac{1}{n} \sum_{i=1}^{n} E\{Z_{ik} | D(Z_{ik})\} \mathrm{Var}\{\Delta_{il} | D(Z_{ik})\} \\
+ \; & \frac{8}{n} \sum_{\substack{k,m,s=1,\\ k<m<s}}^{p} b_k d_{ms} \frac{1}{n} \sum_{i=1}^{n} E\{Z_{ik} | D(Z_{ik})\} E\{\Delta_{im} \Delta_{is} | D(Z_{ik})\} \\
+ \; & \frac{8}{n} \sum_{\substack{k,l=1,\\ k\neq l}}^{p} c_k c_l \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}\{Z_{il} | D(Z_{ik})\} \mathrm{Var}\{\Delta_{ik} | D(Z_{ik})\} \\
+ \; & \frac{8}{n} \sum_{\substack{k,m,s=1,\\ k<m<s}}^{p} c_k d_{ms} \frac{1}{n} \sum_{i=1}^{n} E\{\Delta_{ik}^2 \Delta_{im} \Delta_{is} | D(Z_{ik})\} \\
+ \; & \frac{8}{n} \sum_{\substack{k,m,s=1\\ k<m<s}}^{p} c_k d_{ms} \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}\{Z_{ik} | D(Z_{ik})\} E\{\Delta_{im} \Delta_{is} | D(Z_{ik})\}. \qquad \text{(A.3)}
\end{aligned}
$$

The variance of $\bar{Y}_2 - \bar{Y}_1$ can be written as the sum of (A.1), (A.2), (A.3) and $o_p(n^{-1})$.

We now look at its asymptotic properties as $n \to \infty$. We know that

$$\mathrm{Var}\{\Delta_{ik} | D(Z_{ik})\} = E\{\Delta_{ik}^2 | D(Z_{ik})\} - [E\{\Delta_{ik} | D(Z_{ik})\}]^2 = E\{\Delta_{ik}^2 | D(Z_{ik})\}.$$

As $n \to \infty$, by the law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} E\{Z_{ik}|D(Z_{ik})\} \to E[E\{Z_k|D(Z_k)\}] = E(Z_k)$$

for $k = 1, ..., p$. Similarly, as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} E(\Delta_{ik}^2|D(Z_{ik})) \to E(E(\Delta_k^2|D(Z_k))) = E(\Delta_k^2).$$

for $k = 1, .., p$.

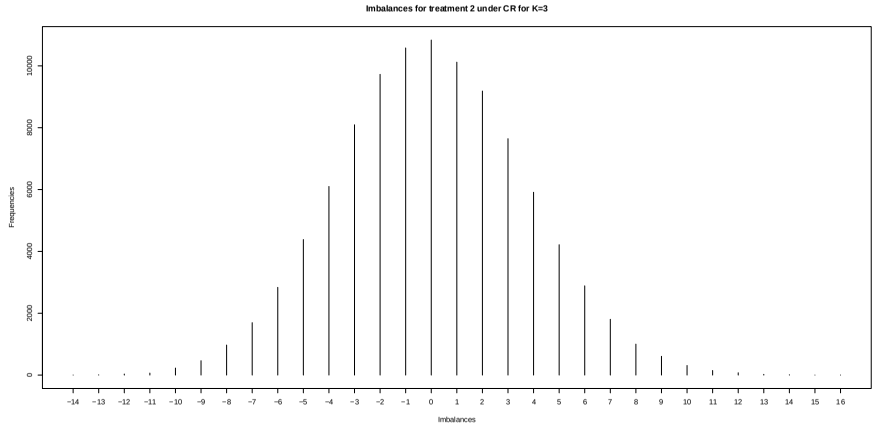
Appendix B

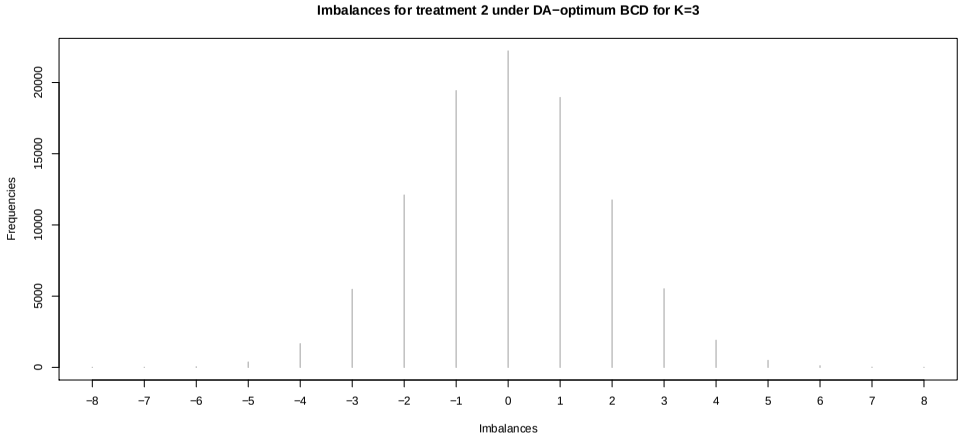
Supplementary information for Chapter 4

B.1 Plots of imbalances under different randomization schemes for both scenario

First, the plots of the imbalances on treatments 2 and 3 under each of the randomization schemes will be given when $K = 3$. Then we have the plots of the imbalances in treatments $2, 3$ and 4 when $K = 4$.

Figure B.1: Plots of imbalances on treatment 2 under complete randomization, the D_A-optimum biased coin design and the adjustable biased coin design when $K = 3$





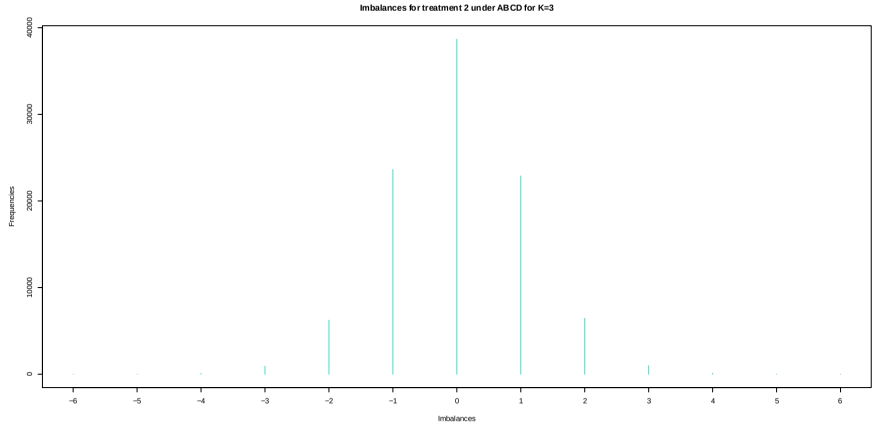


Figure B.2: Plots of imbalances on treatment 2 under Efron's biased coin design with $p = 8/12, 9/12, 10/12$ and $11/12$ when $K = 3$

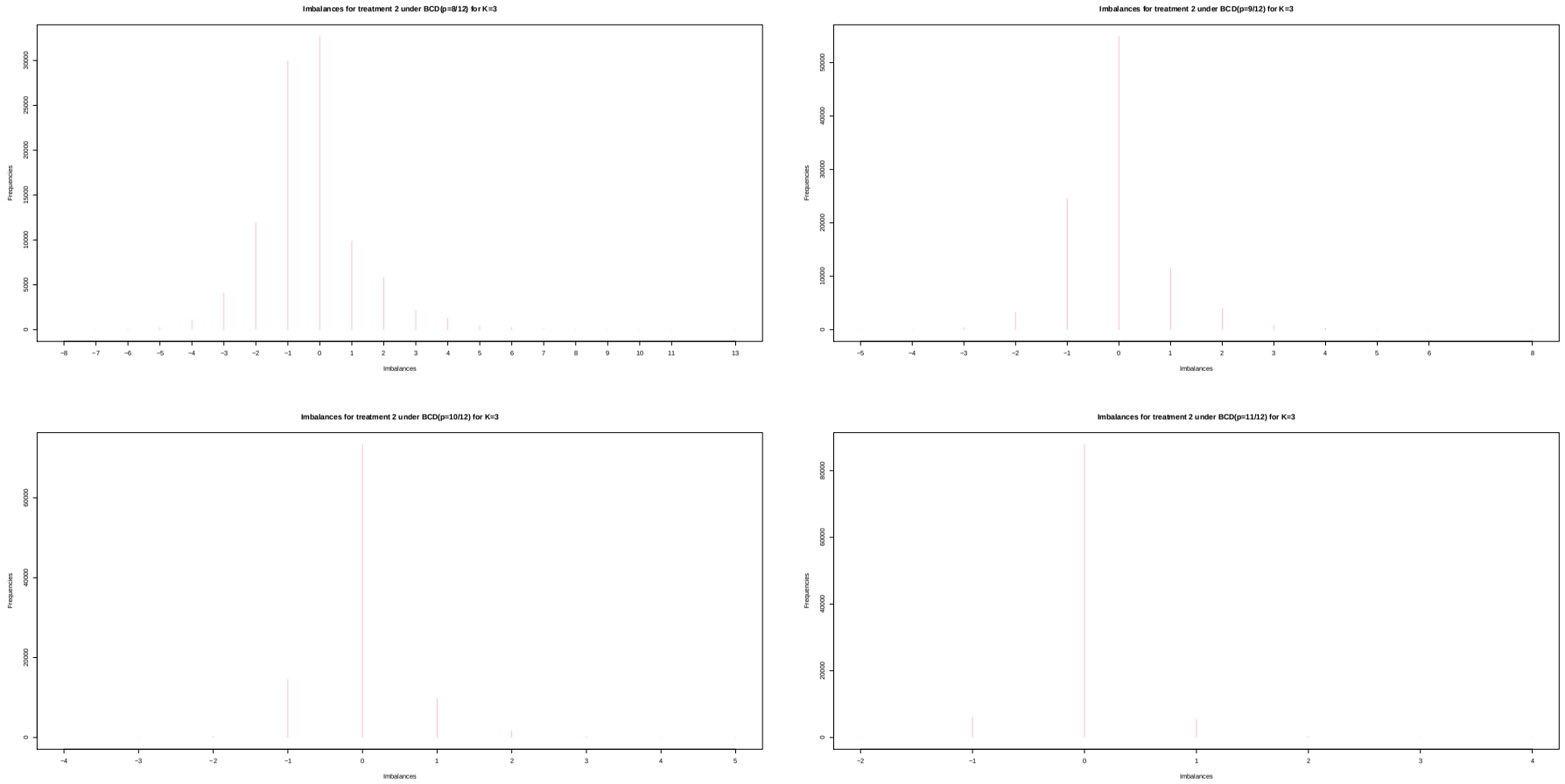
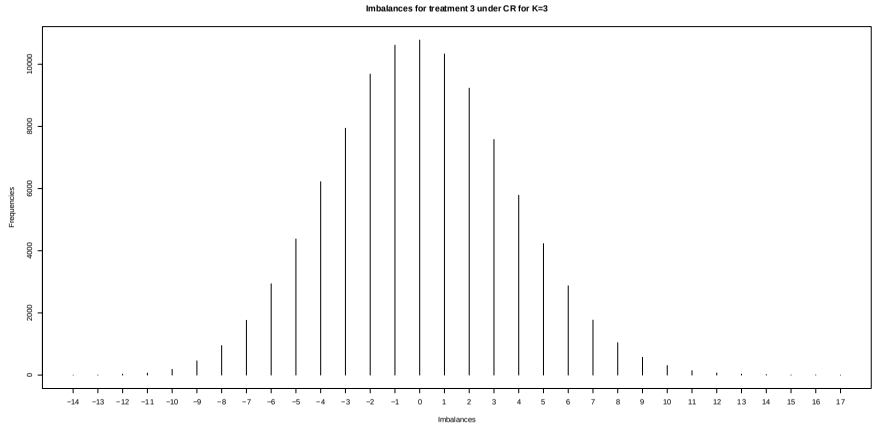
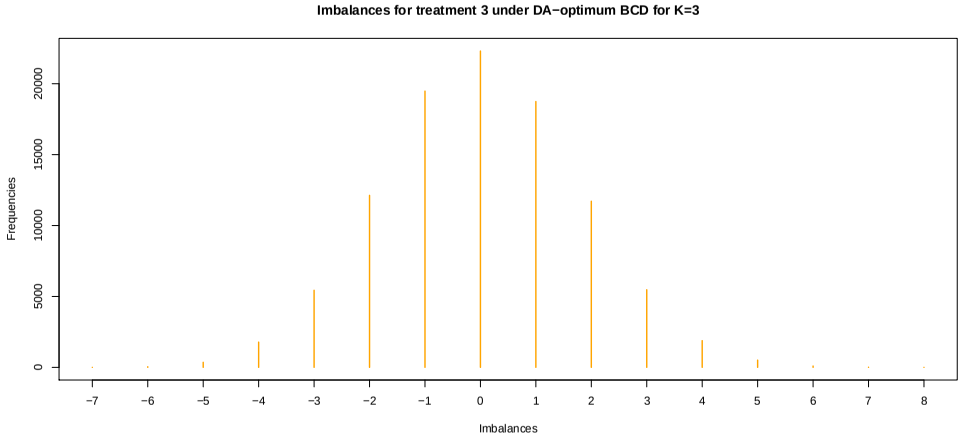


Figure B.3: Plots of imbalances on treatment 3 under complete randomization, the D_A-optimum biased coin design and the adjustable biased coin design when $K = 3$





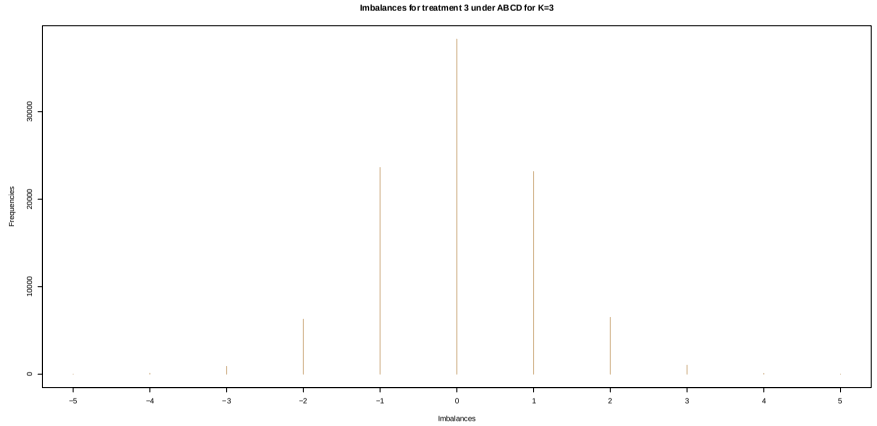


Figure B.4: Plots of imbalances on treatment 3 under Efron's biased coin design with $p = 8/12, 9/12, 10/12$ and $11/12$ when $K = 3$

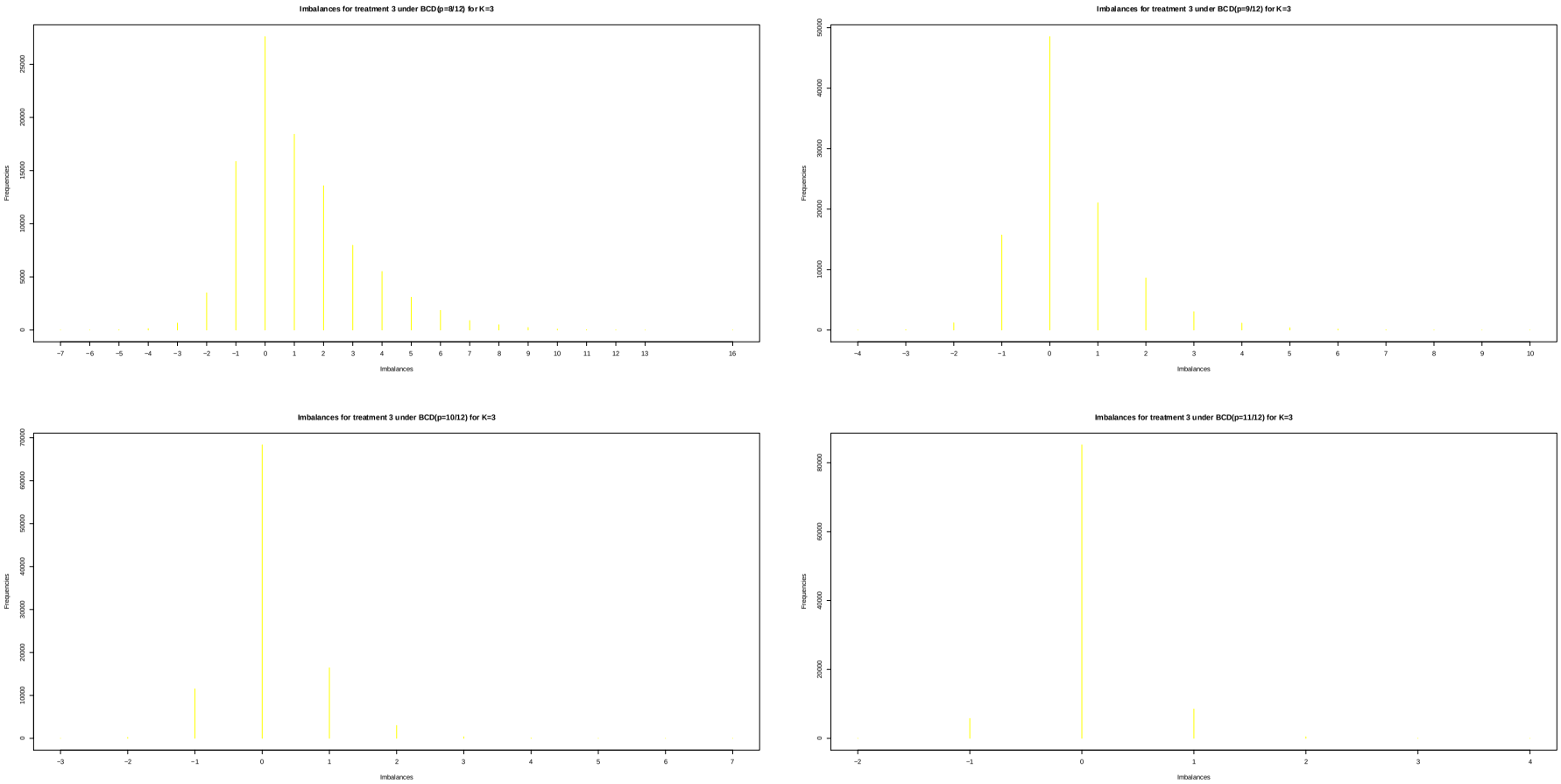
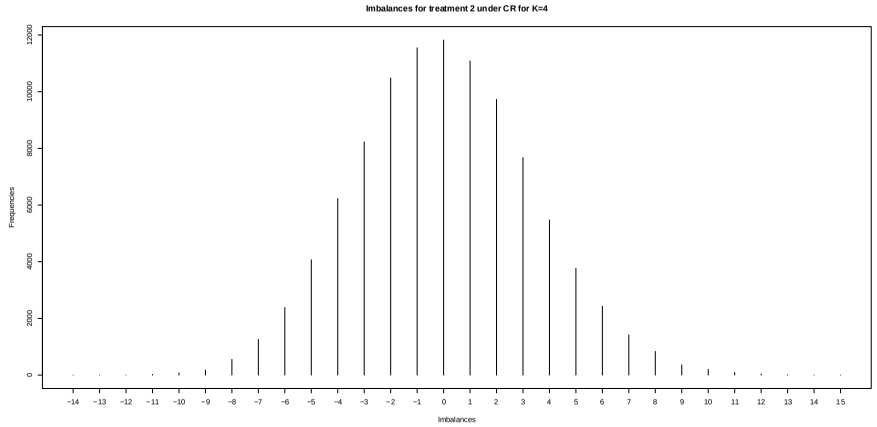
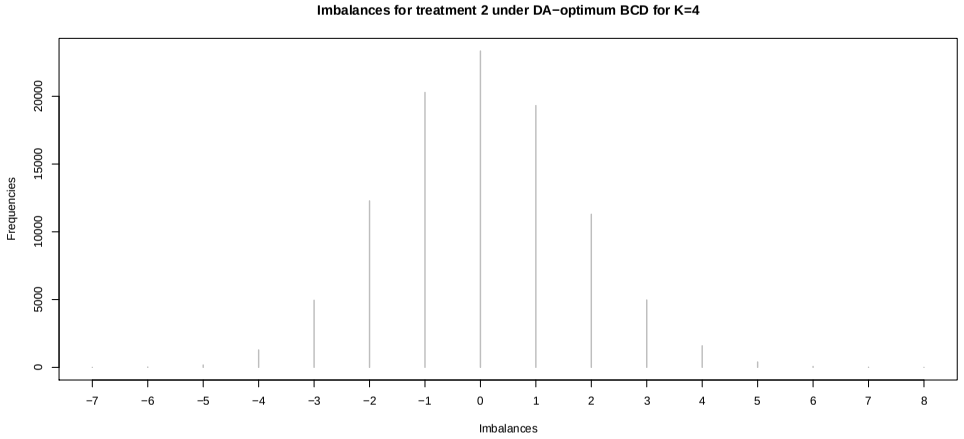


Figure B.5: Plots of imbalances on treatment 2 under complete randomization, the D_A-optimum biased coin design and the adjustable biased coin design when $K = 4$





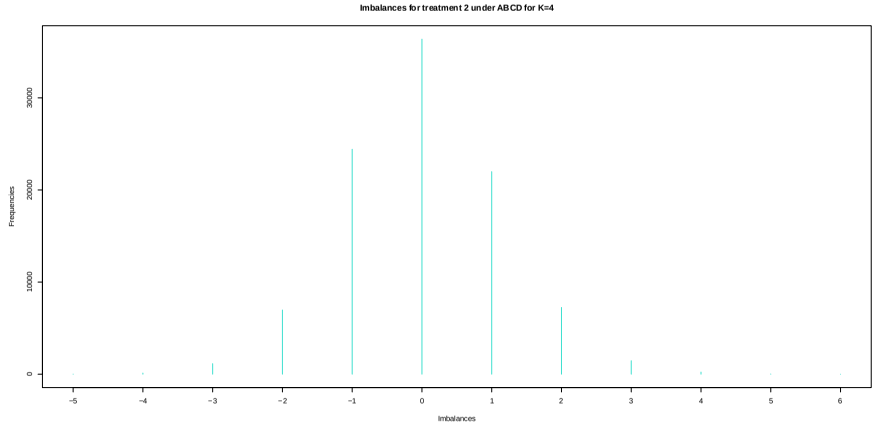


Figure B.6: Plots of imbalances on treatment 2 under Efron's biased coin design with $p = 8/12, 9/12, 10/12$ and $11/12$ when $K = 4$

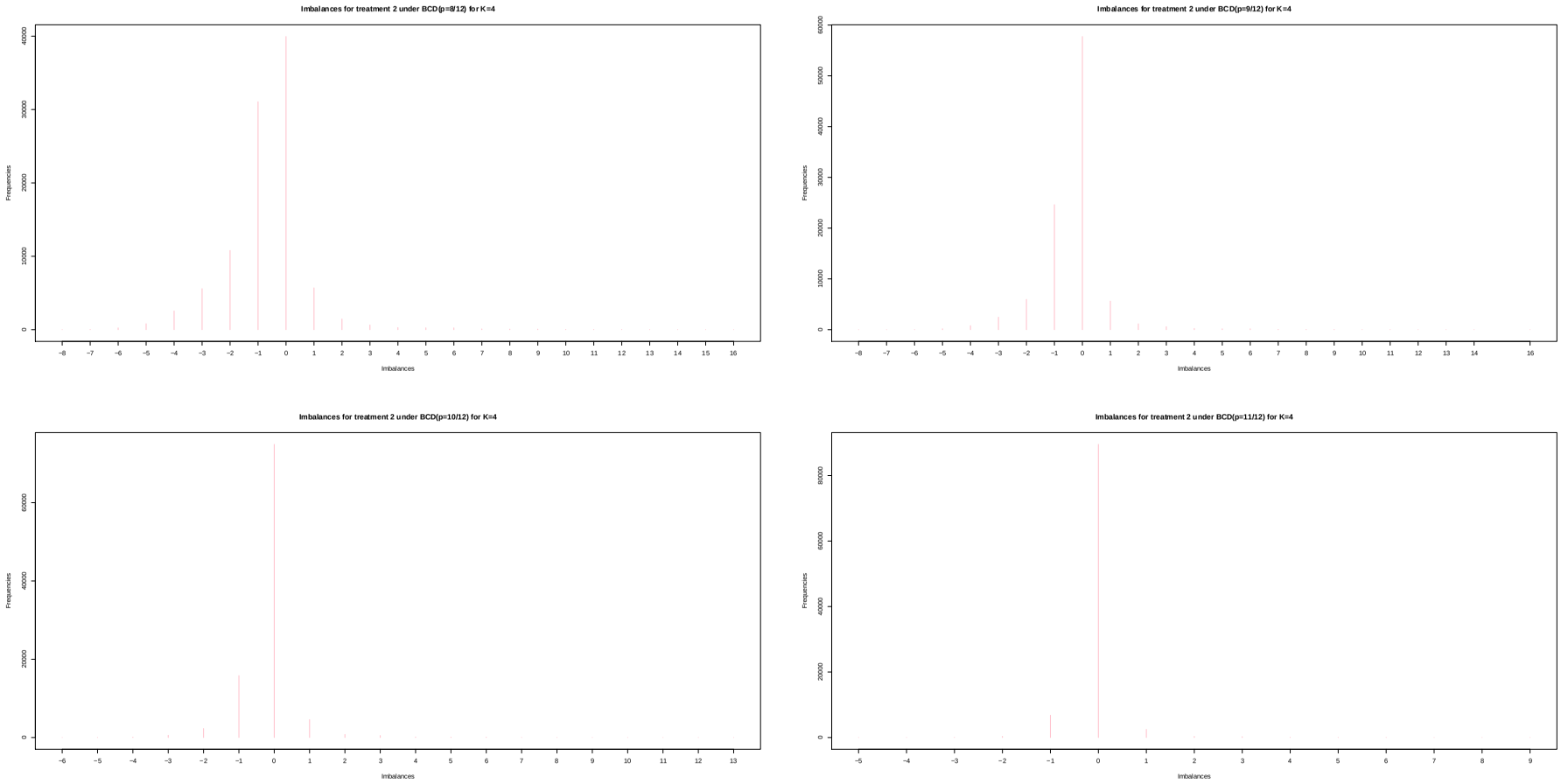
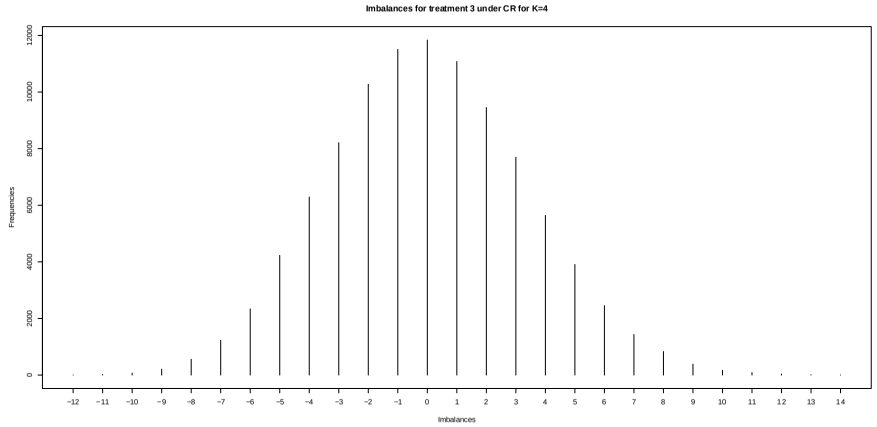
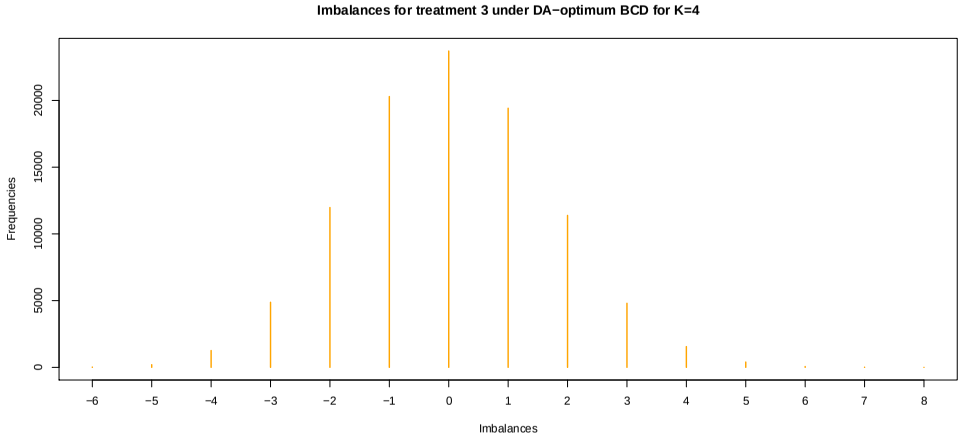


Figure B.7: Plots of imbalances on treatment 3 under complete randomization, the D_A-optimum biased coin design and the adjustable biased coin design when $K = 4$





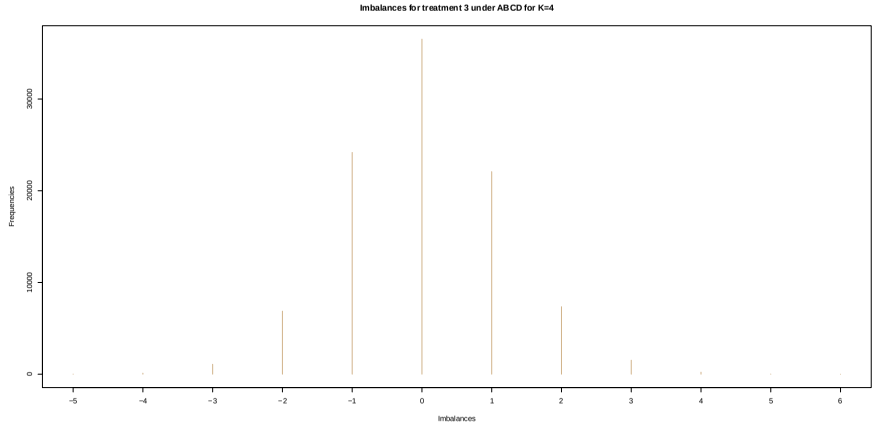


Figure B.8: Plots of imbalances on treatment 3 under Efron's biased coin design with $p = 8/12, 9/12, 10/12$ and $11/12$ when $K = 4$

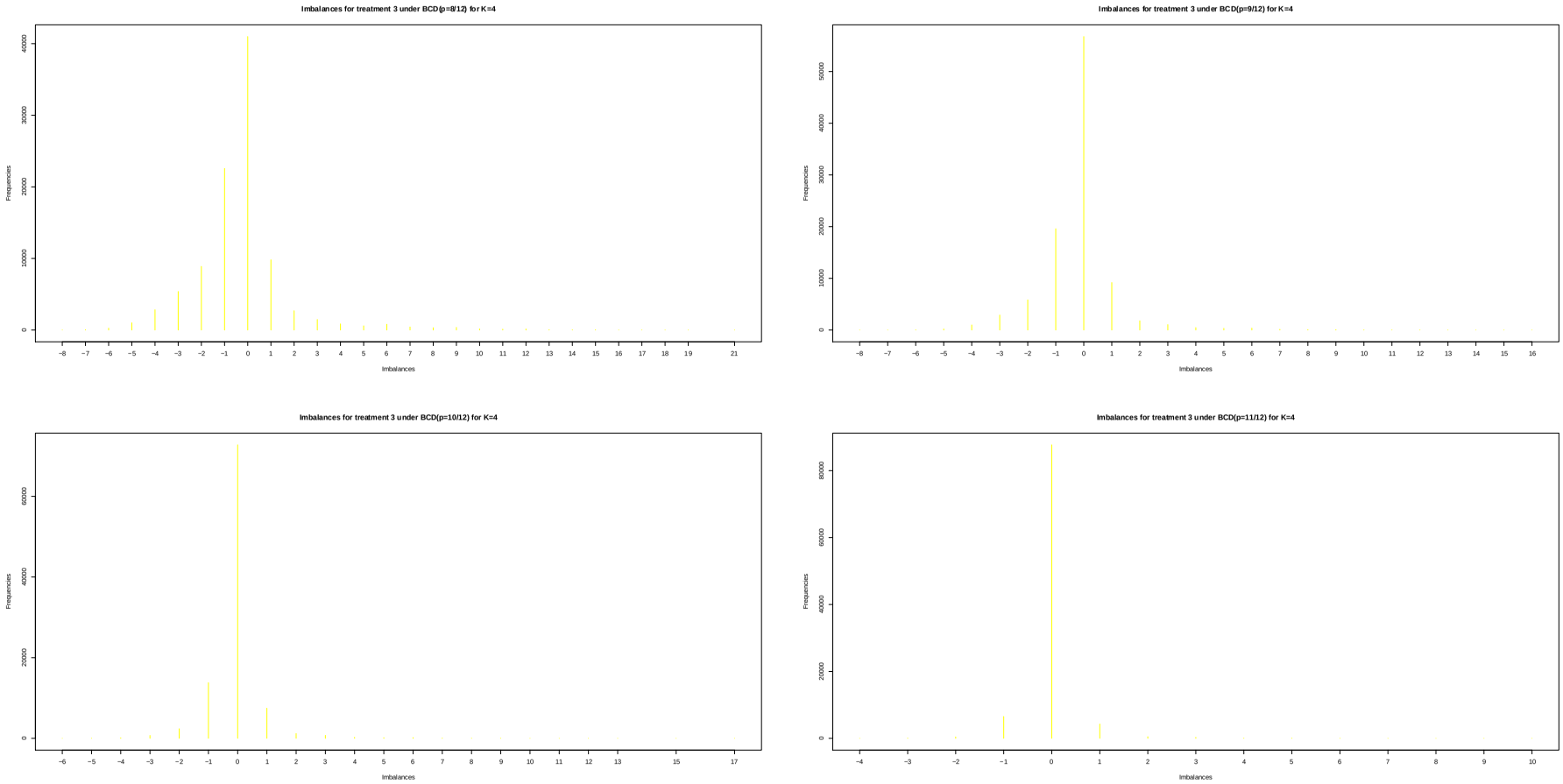
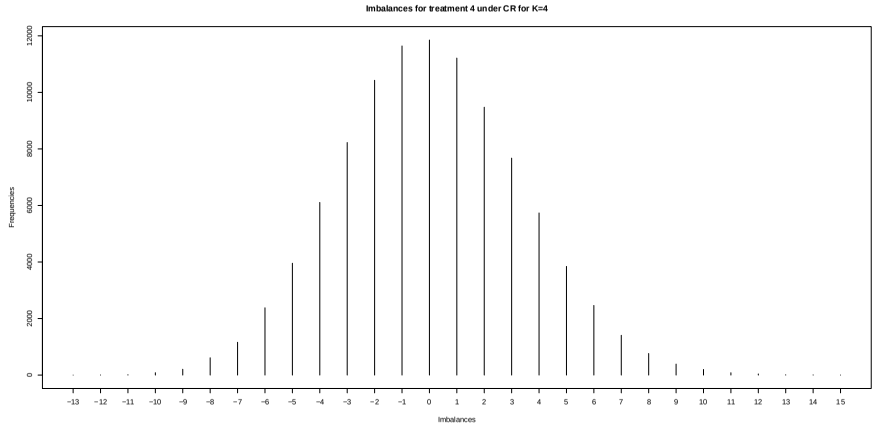
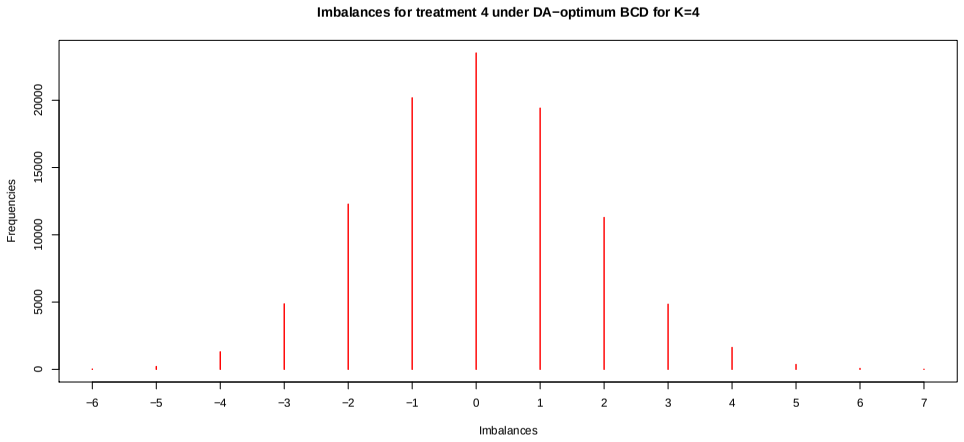


Figure B.9: Plots of imbalances on treatment 4 under complete randomization, the D_A-optimum biased coin design and the adjustable biased coin design when $K = 4$





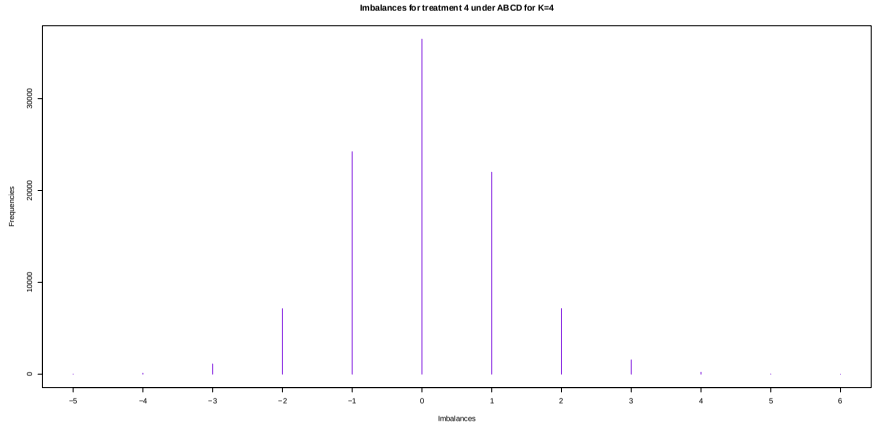
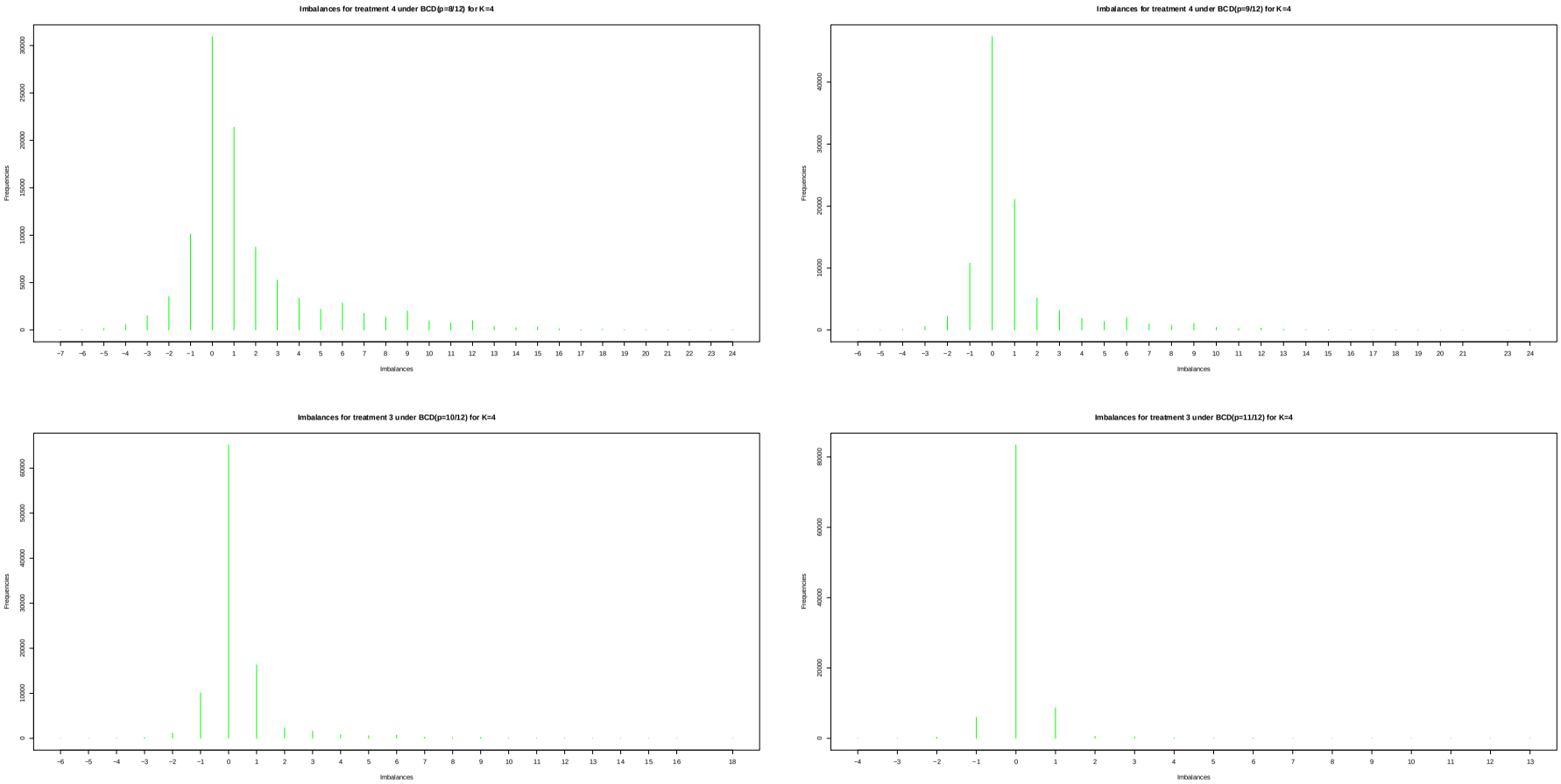


Figure B.10: Plots of imbalances on treatment 4 under Efron's biased coin design with $p = 8/12, 9/12, 10/12$ and $11/12$ when $K = 4$



B.2 Tables of quartiles of imbalance

We have the numerical values of the quartiles of the imbalances for all schemes by simulation when $K = 3$ for treatments 2 and 3 and when $K = 4$ for treatments 2, 3 and 4. T2, T3 and T4 represent treatments 2, 3 and 4, respectively.

Table B.1: Numerical values of the quartiles of the imbalances under all schemes when $K = 3$

	CR		D_A		ABCD		BCD($p = 8/12$)		BCD($p = 9/12$)		BCD($p = 10/12$)		BCD($p = 11/12$)	
	T2	T3	T2	T3	T2	T3	T2	T3	T2	T3	T2	T3	T2	T3
min	-14	-14	-8	-7	-6	-5	-8	-8	-5	-5	-4	-4	-2	-2
Q1	-2	-2	-1	-1	-1	-1	-1	0	-1	0	0	0	0	0
Q2	0	0	0	0	0	0	0	1	0	0	0	0	0	0
mean	0.0073	-0.0001	0.0100	0.0013	-0.0006	0.0033	-0.3361	1.0660	-0.0883	0.3704	-0.0154	0.1194	-0.0009	0.0373
Q3	2	2	1	1	1	1	0	2	0	1	0	0	0	0
max	16	17	8	8	6	5	13	16	8	10	5	7	4	4

Table B.2: Numerical values of the quartiles of the imbalances under complete randomization, the D_A-optimum biased coin design and the adjustable biased coin design when $K = 4$

	CR			D_A			ABCD		
	T2	T3	T4	T2	T3	T4	T2	T3	T4
min	-14	-12	-13	-7	-6	-6	-5	-5	-5
Q1	-2	-2	-2	-1	-1	-1	-1	-1	-1
Q2	0	0	0	0	0	0	0	0	0
mean	-0.0026	0.0047	0.0167	-0.0019	0.0025	-0.0041	-0.0042	0.0066	-0.0042
Q3	2	2	2	1	1	1	1	1	1
max	15	14	15	8	8	7	6	6	6

Table B.3: Numerical values of the quartiles of the imbalances under Efron's biased coin design when $K = 4$

	$p = 8/12$			$p = 9/12$			$p = 10/12$			$p = 11/12$		
	T2	T3	T4	T2	T3	T4	T2	T3	T4	T2	T3	T4
min	-8	-8	-7	-8	-8	-6	-6	-6	-6	-5	-5	-4
Q1	-1	-1	0	-1	-1	0	0	0	0	0	0	0
Q2	-1	0	1	0	0	0	0	0	0	0	0	0
mean	-0.6781	-0.2773	1.6260	-0.3385	-0.1991	0.8805	-0.1285	-0.0512	0.3092	-0.0304	-0.0064	0.0683
Q3	0	0	2	0	0	1	0	0	0	0	0	0
max	16	21	24	16	16	24	13	17	18	9	10	13

Appendix C

Supplementary information for Chapter 5

C.1 Proof of Lemma 5.3.1

For a multivariate hypergeometric distribution, the joint probability distribution is

$$\mathbf{P}[\xi_1(r) = s_1, \ldots, \xi_K(r) = s_K] = \frac{\prod_{i=1}^{K} \binom{k_i}{s_i}}{\binom{K_1}{r}}$$

for $s_i = 0, 1, \ldots, \min(k_i, r)$ and $i = 1, \ldots, K$. As all marginal distributions of the multivariate hypergeometric distribution are hypergeometric, it is sufficient to consider the bivariate hypergeometric distribution to obtain the covariance of $\xi_j(r)$ and $\xi_m(r)$. The probability distribution of a bivariate hypergeometric distribution can be written as

$$\mathbf{P}[\xi_j(r) = s_1, \xi_m(r) = s_2] = \frac{\binom{k_j}{s_1}\binom{k_m}{s_2}}{\binom{K_1}{r}}$$

for $s_1 = 0, 1, \ldots, \min(k_j, r)$ and $s_2 = 0, 1, \ldots, \min(k_m, r)$.

We have

$$E[\xi_j(r)\xi_m(r)] = \sum_{s_1=0}^{\min(r,k_j)} \sum_{s_2=0}^{\min(r,k_m)} s_1 s_2 \frac{\binom{k_j}{s_1}\binom{k_m}{s_2}}{\binom{K_1}{r}}.$$

Using the equalities

$$x\binom{z}{x} = z\binom{z-1}{x-1}$$

and

$$\binom{z}{x} = \frac{z}{x}\binom{z-1}{x-1},$$

we obtain

$$\begin{aligned}
E[\xi_j(r)\xi_m(r)] &= \sum_{s_1=1}^{\min(r,k_j)} \sum_{s_2=1}^{\min(r,k_m)} \frac{k_j k_m \binom{k_j-1}{s_1-1}\binom{k_m-1}{s_2-1}}{\binom{K_1}{r}} \\
&= k_j k_m \sum_{s_1=2}^{\min(r,k_j)} \sum_{s_2=2}^{\min(r,k_m)} \frac{\binom{k_j-1}{s_1-1}\binom{k_m-1}{s_2-1}}{\frac{K_1(K_1-1)}{r(r-1)}\binom{K_1-2}{r-2}} \\
&= \frac{k_j k_m r(r-1)}{K_1(K_1-1)} \sum_{s_1=2}^{\min(r,k_j)} \sum_{s_2=2}^{\min(r,k_m)} \frac{\binom{k_j-1}{s_1-1}\binom{k_m-1}{s_2-1}}{\binom{K_1-2}{r-2}} \\
&= \frac{k_j k_m r(r-1)}{K_1(K_1-1)}.
\end{aligned}$$

Here,

$$\sum_{s_1=2}^{\min(r,k_j)} \sum_{s_2=2}^{\min(r,k_m)} \frac{\binom{k_j-1}{s_1-1}\binom{k_m-1}{s_2-1}}{\binom{K_1-2}{r-2}} = 1,$$

since this is the sum of all the probabilities for a bivariate hypergeometric distribution.

The covariance of $\xi_j(r)$ and $\xi_m(r)$ is then

$$\begin{aligned}
\mathrm{Cov}[\xi_j(r), \xi_m(r)] &= \frac{k_j k_m r(r-1)}{K_1(K_1-1)} - \frac{k_j k_m r^2}{K_1^2} \\
&= -\frac{k_j k_m r(K_1-r)}{K_1^2(K_1-1)}.
\end{aligned}$$

Finally, the covariance of Δ_{ij} and Δ_{im} is

$$\begin{aligned}
\mathrm{Cov}(\Delta_{ij}, \Delta_{im}) &= -\frac{r^2 k_j k_m}{K_1^2} - \frac{k_j k_m r(K_1-r)}{K_1^2(K_1-1)} + \frac{r^2 k_j k_m}{K_1^2} \\
&= -\frac{r k_j k_m (K_1-r)}{K_1^2(K_1-1)},
\end{aligned}$$

so that

$$\mathrm{Cov}[\xi_j(r), \xi_m(r)] = \mathrm{Cov}(\Delta_{ij}, \Delta_{im}).$$

Bibliography

[1] Anisimov, V. (2007). Effect of imbalance in using stratified block randomization in clinical trials. *Proceedings of the 56th Session of the International Statistical Institute*, 1-4.

[2] Anisimov, V. (2010). Impact of stratified randomization in clinical trials. In *mODa 9 - Advances in Model-Oriented Design and Analysis*, eds. A. Giovagnoli, A. C. Atkinson, B. Torsney and C. May, pp.1-8. Berlin : Physica-Verlag HD.

[3] Anisimov, V. (2011). Effect of unstratified and centre-stratified randomization in multi-centre clinical trials. *Pharmaceutical Statistics* **10**, 50-59.

[4] Anisimov, V. and Fedorov, V. (2007). Modelling, prediction and adaptive adjustment of recruitment in multicentre trials. *Statistics in Medicine* **26**, 4958-4975.

[5] Atkinson, A. C. (1982). Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* **69**, 61-67.

[6] Atkinson, A. C. (1999). Optimum biased-coin designs for sequential treatment allocation with covariate information. *Statistics in Medicine* **18**, 1741-1752.

[7] Atkinson, A. C. (2002). The comparison of designs for sequential clinical trials with covariate information. *Journal of the Royal Statistical Society Series A* **165**, 349-373.

[8] Atkinson, A. C. (2004). Adaptive biased coin designs for clinical trials with several treatments. *Discussiones Mathematicae Probability and Statistics* **24**, 85-108.

[9] Bailey, R. A. and Nelson, P. R. (2003). Hadamard randomization: a valid restriction of random permuted blocks. *Biometrical Journal* **45**, 554-560.

[10] Baldi Antognini, A. (2008). A theoretical analysis of the power of biased coin designs. *Journal of Statistical Planning and Inference* **138**, 1792-1798.

[11] Baldi Antognini, A. and Giovagnoli, A. (2004). A new 'biased coin design' for the sequential allocation of two treatments. *Journal of the Royal Statistical Society Series C* **53**, 651-664.

[12] 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**, 519-535.

[13] Begg, C. B. (1990). On inferences from Wei's biased coin design for clinical trials. *Biometrika* **77**, 467-478.

[14] Blackwell, D. and Hodges, J. L. (1957). Design for the control of selection bias. *Annals of Mathematical Statistics* **28**, 449-460.

[15] Chen, Y. P. (1999). Biased coin design with imbalance tolerance. *Communications in Statistics - Stochastic Models* **15**, 953-975.

[16] Chen, Y. P. (2000). Which design is better? Ehrenfest urn versus biased coin. *Advances in Applied Probability* **32**, 738-749.

[17] Chen, Y. P. (2006). The power of Efron's biased coin design. *Journal of Statistical Planning and Inference* **136**, 1824-1835.

[18] Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association* **50**, 1096-1121.

[19] Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58**, 403-417.

[20] Efron, B. (1980). Randomizing and balancing a complicated sequential experiment. In *Biostatistics Casebook*, eds. R. G. Miller, B. Efron, B. Brown and L. E. Moss, pp.19-30. New York : Wiley.

[21] Eisele, J. R. (1995). Biased coin designs: some properties and applications. In *Adaptive Designs*, eds. N. Flournoy and W. F. Rosenberger, M. Hayward, pp.48-64. California : Institute of Mathematical Statistics.

[22] Grizzle, J. E. (1982). A note on stratifying versus complete random assignment in clinical trials. *Controlled Clinical Trials* **3**, 365-368.

[23] Hallstrom, A. and Davis, K. (1988). Imbalance in treatment assignments in stratified blocked randomization. *Controlled Clinical Trials* **9**, 375-382.

[24] Hinkelmann, K. and Kempthorne, O. (2008). *Design and Analysis of Experiments Volume 1: Introduction to Experimental Design*, 2nd edition. New York : Wiley.

[25] Hu, F., Rosenberger, W. F. and Zhang, L. X. (2006). Asymptotically best response-adaptive randomization procedures. *Journal of Statistical Planning and Inference* **136**, 1911-1922.

[26] Hu, F., Zhang, L. X. and He, X. (2009). Efficient randomized-adaptive designs. *The Annals of Statistics* **37**, 2543-2560.

[27] Karlin, S. and McGregor, J. (1957). Random walks. *Illinois Journal of Mathematics* **3**, 66-81.

[28] Karlin, S. and Taylor, M. H. (1975). *A First Course in Stochastic Processes*, 2nd edition. New York : Academic Press.

[29] Lachin, J. M. (1981). Introduction to sample size determination and power analysis for clinical trials. *Controlled Clinical Trials* **2**, 93-113.

[30] Lachin, J. M. (1988a). Statistical properties of randomization in clinical trials. *Controlled Clinical Trials* **9**, 289-311.

[31] Lachin, J. M. (1988b). Properties of simple randomization in clinical trials. *Controlled Clinical Trials* **9**, 312-326.

[32] Lachin, J. M., Matts, J. P. and Wei, L. J. (1988). Randomization in clinical trials: conclusion and recommendations. *Controlled Clinical Trials* **9**, 365-374.

[33] Markaryan, T. and Rosenberger, W. F. (2010). Exact properties of Efron's biased coin randomization procedure. *The Annals of Statistics* **38**, 1546-1567.

[34] Matts, J. P. and Lachin, J. M. (1988). Properties of permuted-block randomization in clinical trials. *Controlled Clinical Trials* **9**, 327-345.

[35] McPherson, G. C., Campbell, M. K. and Elbourne, D. R. (2012). Use of randomization in clinical trials: a survey of UK practice. *Trials* **13** in press.

[36] Meier, P. (1981). Stratification in the design of a clinical trial. *Controlled Clinical Trials* **1**, 355-361.

[37] Morrissey, M., McEntegart, D. and Lang, M. (2010). Randomization in double-blind multicentre trials with many treatments. *Contemporary Clinical Trials* **31**, 381-391.

[38] Pocock, S. J. (1983). *Clinical Trials: A Practical Approach.* New York : Wiley.

[39] Pocock, S. J. and Simon, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* **31**, 103-115.

[40] Pocock, S. J. (1979). Allocation of patients to treatment in clinical trials. *Biometrics* **35**, 183-197.

[41] Rosenberger, W. F. and Lachin, J. M. (2002). *Randomization in Clinical Trials: Theory and Practice*. New York : Wiley.

[42] Senn, S. (1997). *Statistical Issues in Drug Development*. Chichester : Wiley.

[43] Senn, S. (1998). Some controversies in planning and analysing multi-centre trials. *Statistics in Medicine* **17**, 1753-1756.

[44] Shao, J., Yu, X. and Zhong, B. (2010). A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika* **97**, 347-360.

[45] Simon, R. (1979). Restricted randomization designs in clinical trials. *Biometrics* **35**, 503-512.

[46] Smith, R. L. (1984). Sequential treatment allocation using biased coin designs. *Journal of the Royal Statistical Society Series B* **46**, 519-543.

[47] Smith, R. L. (1984). Properties of biased coin designs in sequential clinical trials. *The Annals of Statistics* **12**, 1018-1034.

[48] Taves, D. R. (1947). Minimization: a new method of assigning patients to treatment and control groups. *Clinical Pharmacology and Therapeutics* **15**, 443-453.

[49] Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika* **29**, 350-362.

[50] Welch, B. L. (1947). The generalization of "Student's" problem when several different population variances are involved. *Biometrika* **34**, 28-35.

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

[52] Wei, L. J. (1978). A class of treatment assignment rules for sequential experiments. *Communications in Statistics - Theory and Methods* **7**, 285-295.

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

[54] Wei, L. J., Smythe, R. T. and Smith, R. L. (1986). K-treatment comparisons with restricted randomization rules in clinical trials. *The Annals of Statistics* **14**, 265-274.