EXACT PROPERTIES OF RESTRICTED RANDOMIZATION PROCEDURES

by

Hui Shao

A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
In Partial fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Statistical Science

Committee:

Dr. William F. Rosenberger, Dissertation Director

Dr. Daniel B. Carr, Committee Member

Dr. Guoqing Diao, Committee Member

Dr. Liansheng Larry Tang, Committee Member

Dr. Tigran Markaryan, Committee Member

Dr. William F. Rosenberger, Department Chair

Dr. Kenneth S. Ball, Dean,
Volgenau School of Engineering

Date: 8 December 2015

Fall Semester 2015
George Mason University
Fairfax, VA

Exact Properties of Restricted Randomization Procedures

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Hui Shao
Master of Science
University of Toledo, 2012
Master of Science
University of Chinese Academy of Sciences, 2010
Bachelor of Science
China Agricultural University, 2006

Director: Dr. William F. Rosenberger, University Professor and Chairman
Department of Statistics

Fall Semester 2015
George Mason University
Fairfax, VA

Dedication

I dedicate this dissertation to my parents.

Acknowledgments

I would like to express my deepest gratitude to my dissertation advisor Dr. William F. Rosenberger, for his guidance, encouragement and patience throughout the time of my dissertation research. He is an expert in randomization and I am so grateful that he was willing to share his knowledge and insight with me. I also thank his trust in me, providing me the opportunity to serve as his research assistant for more than two years. He put much effort to help me to improve my English writing, as well as providing suggestions and strong support on my way of achieving my career goal. Every time when I was sick, he always told me to go home and take a rest. He is such a great advisor and I am tremendously fortunate to be his student.

I feel fortunate to have Dr. Carr, Dr. Diao, Dr. Markaryan and Dr. Tang serve on my dissertation committee, and am deeply appreciative to each of them for their insightful comments, encouragement and support. In particular, I am grateful that Dr. Markaryan took the time out of his busy schedule to discuss the mathematical problems in my dissertation research and share his experience with me. I thank Dr. Carr for his excellent guidance and invaluable suggestions on better visualizing my research results. I want to express my gratitude to Dr. Vidyashankar, from whom I took half of my doctoral courses. He has great knowledge in the field of both mathematics and statistics.

Special thanks to Mrs. Liz Quigley and Ms. Leslie J. Carl for helping me schedule numerous meetings and booking conference rooms, as well as their kindness and patience.

I would never have been able to finish my dissertation without the love and support from my family. I owe an irredeemable debt to my father and mother. They provide me unconditional love and constant support. They are always proud of me and forgive all my mistakes. I thank my husband for encouraging me to pursue this doctoral degree. He always sees my potential and stimulates me from ordinary to extraordinary. My parents and my husband are also my best friends. They always stay with me and try their best to help me when I have a difficult time. I am so filled with gratitude and joy for being surrounded by love and understanding.

Last but not the least, I would like to thank all my colleagues for their company in the past three and half years.

Table of Contents

List of Tables

List of Figures

Abstract

EXACT PROPERTIES OF RESTRICTED RANDOMIZATION PROCEDURES

Hui Shao, PhD

George Mason University, 2015

Dissertation Director: Dr. William F. Rosenberger

Restricted randomization is employed in clinical trials to achieve balanced or near balanced treatment assignments, but is associated with a loss of randomness. For example, the permuted block design, probably the most commonly used restricted randomization procedure, randomizes patients block by block and achieves perfect balance within each block. However, there is at least one predictable treatment assignment within each block. To avoid deterministic treatment allocation in the permuted block design, many clinical trialists prefer randomizing the block sizes as well. While such a procedure is rarely formalized, it is generally assumed that the design will be less predictable.

In this research, we first formalize the random block design by assuming a discrete uniform distribution for block size. We obtain its distributional properties such as the joint distribution of the block size and position within the block. With this distribution theory, we can analyze properties of prediction and balance. We then explore properties of restricted randomization procedures in clinical trials, including the permuted block design, random block design, Efron's biased coin design and the big stick design. We investigate the degree of predictability and balancing properties theoretically, and discuss the performance of randomization tests by simulation. Additionally, we conduct graphical comparisons between these randomization procedures with respect to bivariate objectives.

The goal of this study is first to provide a statistical understanding of the random block design, and second to provide properties of some commonly used randomization procedures for guidance in designing clinical trials. Our analysis allows clinical trialists to quantitatively assess predictability and balance of a restricted randomization procedure without doing simulations.

Chapter 1: Introduction and Literature Review

1.1 Randomization in Clinical Trials

Participants in a clinical trial are randomly assigned to one of two or more treatment groups. Randomization offers at least three major advantages. First, it prevents both the investigators and participants from guessing further treatment assignments. In unmasked studies, investigators may be able to guess the treatment assignment for further patients if they know the past treatment assignments. The investigators may wait to enter a patient into a treatment which they consider to be better suited to that specific patient if they know the next treatment assignment. In this case, a potential bias will be introduced into the trial. Randomization mitigates this type of bias, which is known as *selection bias*. Selection bias will be described more in Section 1.3.

Second, randomization generates comparable treatment groups in both known and unknown covariates with high probability. Any baseline differences in participant characteristics such as gender, age, race and some medical baseline measurements between treatment groups may cause bias. These covariates are known in advance and their distribution within each treatment group can be equalized by certain randomization techniques, for example, stratified randomization and covariate-adaptive randomization. Randomization also helps increase the probability of comparability with respect to unknown important covariates as well (Cornfield, 2012).

Third, randomization provides an assumption-free statistical test of the treatment effect. These tests are known as randomization tests, originally proposed by Fisher (Fisher, 1935). Randomization tests are performed simply on the basis that participants were randomly assigned to treatment groups. Proper randomization ensures that statistically significant differences between treatment groups are indeed due to an actual treatment effect.

Hu and Rosenberger (2006) described five classes of randomization procedures: complete randomization, restricted randomization, covariate-adaptive randomization, response-adaptive randomization, and covariate-adjusted response-adaptive randomization. Let T_j be a random variable that is 1 if the jth patient is assigned to treatment A and 0 if the jth patient is assigned to treatment B. Then $T_1, \ldots, T_n$ be a sequence of random treatment assignments. In complete randomization, a balanced coin is used to determine the assignment. The assignment T_j is Bernoulli distributed with $P(T_j = 1) = 1/2$ and $T_1, \ldots, T_n$ are independent. Complete randomization has the largest degree of randomness but also a large probability of producing imbalanced treatment groups and a small probability of serious imbalance.

Restricted randomization is used to achieve equal or approximately equal assignments among treatment groups. In a restricted randomization procedure, the jth patient is assigned to either treatment A or B with an allocation probability based on the prior treatment assignment. The allocation probability is defined as $\phi_j = P(T_j = 1|T_1, \ldots, T_{j-1})$.

Covariate-adaptive randomization is used to assign patients to minimize the imbalances of certain known important covariates within treatment groups. Response-adaptive randomization is employed when it is desirable to allow more patients to be assigned to the superior treatment. For example, 50-50 allocation may be unethical if the outcome of the control group is serious disease or high probability of death. In response-adaptive randomization, the treatment assignments depend on the responses of the previous patients. The aim of covariate-adjusted response-adaptive randomization is to allocate greater numbers of patients to the superior treatment according to the patients' covariate values, achieving high statistical efficiency and power in comparing treatment effects. In a covariate-adjusted response-adaptive randomization procedure, the allocation probabilities depend on the previous patients' responses and covariates as well as the incoming patient's covariates.

A statistician always faces the question of selecting an appropriate randomization procedure when designing a randomized clinical trial. The goal of this research is to provide properties of some commonly used restricted randomization procedures that are useful in

selecting the appropriate procedure. We only focus on some of the restricted randomization procedures that could be easily implemented in clinical trials. We establish theoretical results for the degree of predictability and treatment assignments balancing properties and investigate the performance of randomization tests by simulation.

1.2 Restricted Randomization

Many restricted randomization procedures in two-armed trials are described in Chapter 3 of Rosenberger and Lachin (2002). Let $N_A(j)$ and $N_B(j)$ be the number of patients assigned to treatment A and treatment B after j patients have been assigned, respectively. Let n be the total number of patients. Both the random allocation rule and the truncated binomial design are ways to assign exactly $n/2$ patients to each treatment. The *random allocation rule* (RAR) is defined by the following allocation probabilities

$$\phi_j = \begin{cases} 1/2, & j = 1, \\ \frac{n/2 - N_A(j-1)}{n-(j-1)}, & j = 2, \ldots, n. \end{cases} \tag{1.1}$$

The *truncated binomial design* (TBD) is defined by the following allocation probabilities

$$\phi_j = \begin{cases} 1/2, & \text{if } \max\{N_A(j-1), N_B(j-1)\} < n/2, \\ 0, & \text{if } N_A(j-1) = n/2, \\ 1, & \text{if } N_B(j-1) = n/2. \end{cases} \tag{1.2}$$

In a *permuted block design* (PBD) procedure, first a number of blocks with even block size are established, then treatment assignments are randomized within each block. For the purpose of treatment balance, one can use either the random allocation rule (1.1) or the truncated binomial design (1.2) within each block. We refer to the PBD with block size

$2B$ using the RAR to fill each block as PBD(B; R), and the PBD using the TBD within blocks as PBD(B; T). Because there is at least one deterministic assignment within each block under the PBD, sometimes the *random block design* (RBD) is used by varying the block sizes. We have not seen a formal description of the RBD. We will devote a chapter to formalize the RBD and study its distributional properties.

The family of biased coin designs includes Efron's biased coin design (Efron, 1971), Wei's urn design (Wei, 1977), Wei's adaptive biased coin design (Wei, 1978), Atkinson's optimum design (1982), Smith's design (1984b), accelerated biased coin design (Baldi Antognini and Giovagnoli, 2004) and dominant biased coin design (Baldi Antognini and Zagoraiou, 2014). Among these biased coin designs, Efron's biased coin design is the easiest one to implement. *Efron's biased coin design* is defined by

$$\phi_j = \begin{cases} 1-p, & D_{j-1} > 0, \\ 1/2, & D_{j-1} = 0, \\ p, & D_{j-1} < 0, \end{cases} \tag{1.3}$$

where D_j is the difference between the number of treatment A and number of treatment B, $D_j = N_A(j) - N_B(j) = 2N_A(j) - j$, and p is a constant between $(0.5, 1]$. We refer to this design as BCD(p). Unlike the BCD(p) that p_j is a constant, under other biased coin designs, p_j is a function of D_j.

Another class of restricted randomization procedure includes the big stick design (Soares and Wu, 1983) and the biased coin design with imbalance intolerance (Chen, 1999). Under these two designs, a maximum tolerated imbalance bound is pre-specified. The big stick design (BSD) is defined by

$$\phi_j = \begin{cases} 1, & D_{j-1} = -a, \\ 1/2, & -a < D_{j-1} < a, \\ 0, & D_{j-1} = a, \end{cases} \tag{1.4}$$

where a is the pre-specified maximum tolerated imbalance. We refer to this design as the BSD(a). The biased coin design with imbalance intolerance (BCDII) is defined by

$$\phi_j = \begin{cases} 1, & D_{j-1} = -a, \\ 1-p, & -a < D_{j-1} < 0, \\ 1/2, & D_{j-1} = 0, \\ p, & 0 < D_{j-1} < a, \\ 0, & D_{j-1} = a, \end{cases} \tag{1.5}$$

Berger, Ivanova and Deloria-Knoll (2003) introduce the maximal procedure. The maximal procedure is constructed by randomly selecting an allocation sequence with equal probability from the whole set of sequences that satisfy final balance and the maximum tolerated imbalance. The allocation rule is similar as the big stick design, but the transition probabilities are different. This design requires the exact number of patient n is known, and making the randomization-based inference from this design is more difficult.

One can think of the big stick design as complete randomization with an imbalance intolerance; BCDII is Efron's BCD with imbalance intolerance; and the maximal procedure is the random allocation rule with imbalance intolerance.

1.3 Criteria for Restricted Randomization

The degree of randomness and variability of treatment balance are the two main criteria for qualifying a restricted randomization procedure. A higher degree of randomness means a lower chance of predictability of further assignments. We want the randomization sequence to be less predictable because predictability can introduce selection bias. If the investigators are unblinded to the previous treatment assignments, they may be able to guess the next assignment hence may try to alter the entry sequence of patients to achieve the success of the new treatment. This may happen in masked studies too; for example, if the results of different treatments are adverse, or the masked treatment has distinguishing features. This type of prediction of future allocations could lead to the imbalance of baseline covariates across treatment groups which in turn could bias the study result when the unbalanced covariates happen to be high related with the outcome. Green and Byar (1984) stated that "even small imbalances in important prognostic factors could overwhelm treatment differences, either producing apparent treatment effects when none in fact are present or masking true treatment differences when they do exist". Berger (2005) quantified the covariate imbalance resulting from selection bias when a permuted block randomization procedure is used. The covariate imbalance is a function of the block size and the level of certainty about upcoming allocations. He found that, on average, a binary covariate can be up to 50% unbalanced by selection bias. The induced covariate imbalance from selection bias may result in biased parameter estimation, inflated type I error rates, or overly narrow confidence intervals (Proschan, 1994; Berger and Exner, 1999). Therefore, "the elimination of selection bias is the most essential requirement for a good clinical trial" (Chalmers, 1990).

Regarding the quantification of the prediction of future allocations, there are several models. The most commonly used is the expected selection bias factor proposed by Blackwell and Hodges (Blackwell and Hodges, 1957). They proposed a *convergence strategy* for guessing the upcoming assignment, which was to guess the treatment that has fewer prior allocations, or to guess one of the treatments if both treatments have equal numbers of prior allocations. Their model calculates the *expected selection bias factor*, $E(F)$, which is the

expected excess of correct guesses of treatment assignments beyond that expected by chance when the investigator uses the convergence strategy. It is also equivalent to the difference between the expected number of correct and incorrect guesses among all the guesses made when the two treatment groups have different prior assignments. The proof can be found in Chapter 6 of Rosenberger and Lachin (2002). If the treatment assignments remain masked then there is no potential selection bias regardless of the methods of treatment assignment. For any unmasked clinical trial with complete randomization, $E(F) = 0$, since the number of correct guesses guessing with probability $1/2$ for each treatment is $n/2$, which is the same as that using the optimal strategy. For the PBD(B, R), let M be the number of blocks, then the expected selection bias factor is given by (Matt and Lachin, 1988):

$$E(F) = M\left[\frac{2^{2B-1}}{\binom{2B}{B}} - \frac{1}{2}\right]. \tag{1.6}$$

While $E(F)$ measures the difference between the expected number of correct and incorrect guesses, Smith (1984a) suggested measuring the difference between expected percentage of correct guesses and that of incorrect ones, which is equivalent to $\sum_{j=1}^{n} E(F)/n$. This metric has the advantage of ranging from 0 to 1. Many researchers have employed this measure of selection bias in their studies (Baldi Antognini and Giovagnoli, 2004; Zhao et al, 2011; Baldi Antognini and Zagoraiou, 2014; Atkinson, 2014). Smith also proposed a measure to quantify the effect of selection bias based on the Blackwell and Hodges model, which is $E(F)x_d/(2n)$, where x_d is the mean difference in response between a patient selected in the belief that the next treatment assignment will be treatment A and one selected in the belief of treatment B (Smith, 1984b). Another measure of the selection bias based on $E(F)$ is Chen's (1999, 2000) method. He introduced the average excess selection bias up to epoch n for assessing randomness:

$$\sum_{j=1}^{n} \frac{E[\max(\phi_j, 1-\phi_j)]}{n} - \frac{1}{2}. \tag{1.7}$$

He applied this quantity to measure the randomness of the BCDII, BCD, and BSD. We will show that this is mathematically equivalent to $\sum_{j=1}^{n} E(F)/n$ later in Chapter 3.

Matts and Lachin (1988) proposed a type of predictability: *prediction with certainty*. They assume the investigators only try to influence the patients' entry when they know what the next assignment is with certainty. For a PBD(B, R), the upcoming assignments are deterministic after B allocations have occurred to one of the treatments within the block. Therefore, the number of predictions with certainty can be from 1 to B within a block. Matts and Lachin modified the Blackwell-Hodges model to measure $E(F')$, the expected number of assignments that are predictable with certainty. For an unmasked two-armed PBD(B, R), $E(F')$ is given by

$$E(F') = \frac{2BM}{B+1}. \tag{1.8}$$

Deterministic allocation does not occur in all the restricted randomization procedures, for example, the BCD and Wei's urn design. Dupin-Spriet, Fermanian and Spriet (2004) derived a more general formula to measure the predictability which can also be used in a PBD with three or more arms, and trials with unbalanced treatment allocation. They define *the predictability within a block* as the proportion of the expected number of treatments which are predictable within a block of known length. In the case of trials with two treatment arms and balanced PBD, the predictability within a block is $1/(B+1)$, which is equivalent to the result (1.8) of Matts and Lachin. From this formula, we can see that the predictability within a block decreases as in the block size increases.

Berger (2007) pointed out that the predictable allocations defined by Dupin-Spriet *et al.* are actually deterministic allocations. He redefined *predictable allocations* as those

whose conditional distribution differs from the unconditional distribution specified by the allocation proportions. He also defines *deterministic allocations* as those for which the conditional distribution is degenerate, having a positive probability of only one outcome. For illustration, consider the sequence ABAB for a two-armed balanced PBD(2, R). The fourth allocation in the sequence is deterministic because we know it must be B, given that each block has only two allocations to A and two allocations to A have been already made in this block. The second allocation is predictable because, given the first assignment to A, there remain two assignments to B and one assignment to A. The conditional probability of allocation to B is 2/3 while the unconditional probability is 1/2. The first and third allocations are unpredictable because their conditional probabilities are 1/2 which is the same as the unconditional probability. The results appear in Table 4.4 of the book are reproduced here in Table 1.1 (Berger, 2007).

The correct guesses showed in the table are calculated based on the use of convergence strategy. The results show that the probability of deterministic allocations decreases greatly from 50% to 25% when the block size increases from 2 to 6. However, the chance of predictable allocation increases and the probability of correct guesses only decreases slightly from 75% to 68%.

Since restricted randomization is employed for achieving exact or approximate treatment assignment balance, the treatment assignment balancing property, is another important criteria for qualifying a restricted randomization procedure. The main reason for equal allocation is statistical precision. The test and estimation procedure to be used in analyzing the data are usually more efficient if the allocation is balanced (Soares and Wu, 1983). Large imbalances in treatment assignments may create imbalances in important covariates; hence they negatively impact the study results as we discussed before.

There is no uniform measure for balance in randomization procedures. Efron (1971) looked at the probability of achieving exact balance. Soares and Wu (1983) proposed to look at both the probability that the design produces an unacceptable imbalance, $P(|D_n| \geq n/3)$, and the maximum absolute imbalance throughout the trial. Based on the fact that the

variance of the estimated treatment difference is proportional to $1/n_A+1/n_B$, Smith (1984a) used $E(1/n_A + 1/n_B - 4/n)$ to measure the imbalance for biased coin design. Chen (1999) introduced the *average imbalance between the treatment allocations up to epoch* n to measure the imbalance of the BCD, BSD and BCDII, which is given by

$$\sum_{j=1}^{n} \frac{E(|D_j|)}{n}. \tag{1.9}$$

Burman (1996) considered the lack of balance at step n in the form

$$L_n = E\left(\frac{D_n}{\sqrt{n}}\right)^2,$$

which in fact equals $Var(D_n)/n$ since $E(D_n) = 0$. This measure is also called loss of precision and has been widely used in research (Atkinson 1999, 2002, 2003 and 2014, Baldi Antognini and Zagoraiou 2014). Markaryan and Rosenberger measured the exact variance of terminal imbalance of the BCD, $Var(D_n)$.

Ideally a restricted randomization procedure would be random and would produce balanced allocation. In fact, randomness and balance are competing goals. The more restrictive the allocation procedure, the less randomness, and treatment imbalance, the greater the predictability and potential for selection bias. The random allocation rule, the truncated binomial design and the PBD with a completion of the last block produce exact balance in treatment assignment at the cost of prediction. The BCD and Wei's urn design are less predictable but have potentially large imbalances when the sample size is large. Much of the research on randomization procedures has been to find a trade-off between the degree of randomness and balance in the treatment assignments. Researchers have attempted to find a less restricted randomization procedure that minimizes the predictability while retaining balance; for example, the big stick design, the biased coin design with imbalance intolerance and the maximal procedure.

Berger, Ivanova and Deloria-Knoll (2003) define three conditions for describing balance. *Condition T* specifies the final treatment balance. *Condition B* specifies the maximum tolerated imbalance. *Condition F* specifies perfect balance within each block. The big stick design and biased coin design with imbalance intolerance are defined by Condition B while the maximal procedure is defined by Conditions T and B. Because all the allocation sequences for a PBD have to satisfy all three conditions, the big stick design, the biased coin design with imbalance intolerance and the maximal procedure have less predictability than the PBD. However, for the maximal procedure, this reduction of prediction does not require a corresponding increase of imbalance, because Condition B specifies the largest imbalance.

1.4 Permuted Block Design

The PBD is the most frequently used restricted randomization procedure in clinical trials. If the enrollment of the last block is complete, this randomization procedure can achieve a perfect balance between treatment groups, otherwise, it generates treatment groups with unequal sample sizes. The imbalance at point n, which is $|D_n|$, is always less than or equal to B.

The PBD can prevent the problem of severe imbalance during the course of the trial that may occur in complete randomization, the random allocation rule and truncated binomial design. If the baseline characteristics of patients change with time, such severe imbalance could lead to differences between treatment groups in these important covariates, hence introducing potential bias especially when the trial size is small. Choosing a small block size could fix this problem. Block sizes greater than 2 are recommended, as a procedure with a block size of 2 requires identical pairs and incurs a high risk of selection bias (Rosenberger and Lachin, 2002).

The main disadvantage of the PBD is that selection bias arises frequently since the allocation sequence will become known or predictable when the block size is fixed and the previous treatment assignments are unmasked. For example, within a block, if the

current assignment sequence is BAA, given a block size of 4, the last assignment must be B in order to ensure treatment balance. Such a deterministic treatment assignment allows the investigators to guess future treatment assignment correctly and they may alter the patients' entry. Using a large block size will help protect against the investigator predicting the treatment assignment. However, if one treatment occurs frequently at the beginning of a block, there will be a large imbalance if the trial is terminated midway through a block. Therefore, the block size should be short enough to prevent a large imbalance and long enough to decrease the predictability of further assignment.

The random block design (RBD), in which the block size is randomly selected, is proposed to reduce the predictability of further assignments when the trial is not double-blinded. Each of the M blocks has $2b_i$ patients, $i = 1, 2, \ldots, M$. Each b_i is an integer from $1, ..., K$. The block size of the ith block is randomly selected with a probability of $1/K$. Since the block sizes are varied and concealed, the procedure helps preserve unpredictability.

Matts and Lachin (1988) calculated the expected selection bias factor $E(F)$ and the modified selection bias factor $E(F')$ for both PBD and RBD when the block sizes are unmasked. For a PBD with equal block size $2B$, $E(F)$ is given by (1.6) and $E(F')$ is given by (1.8). For a RBD with possibly unequal block sizes $(2b_i, i = 1, 2, \ldots, M)$, if every block is filled, the overall selection bias factor $E(F)$ is given by

$$E(F) = \sum_{i=1}^{M} \left[\frac{2^{2b_i - 1}}{\binom{2b_i}{b_i}} - \frac{1}{2} \right]. \tag{1.10}$$

The overall $E(F')$ is given by

$$E(F') = \sum_{i=1}^{M} \frac{2b_i}{b_i + 1}. \tag{1.11}$$

Note here that b_i is, itself, a random variable, and there is no assurance that the final block will be filled. Based on the calculation, they argued that the selection bias factor of a RBD

is approximately equal to the average of the bias factors over all blocks. For example, the selection bias factor of a design with block sizes of 6 and 10 will be approximately the same as that of a block size of 8. In this case, the RBD does not reduce the chance of selection bias. Matts and Lachin further illustrated that the prediction with certainty is considerably reduced by varying the block sizes instead of having a fixed block size only if the block sizes are unknown to the investigators, by comparing the $E(F')$ of an RBD with two block sizes 4 and 6 to a PBD having a block size of 6. When the sequence of block sizes is masked and the sequence of treatment assignments is unmasked, treatment assignments are known with certainty only when the imbalance reaches one-half of the block size. Hence, decreasing the proportion of blocks with the largest block size will reduce the expected selection bias.

Many researchers pointed out that, even when the sequence of block sizes is unknown, the RBD does not reduce the chance of selection bias and may introduce more prediction than the PBD (Rosengerger and Lachin, 2002; Berger, Ivanova and Deloria-Knoll, 2003; Berger, 2007; Salama, Ivanova and Qaqish, 2008). Berger, Ivanova and Deloria-Knoll (2003) compared the type I error rate inflation of the PBD, RBD and the maximal procedure with the same maximum imbalance to access the presence of selection bias based on a simulation study. They showed that the RBD generates the most predictable allocations and the least number of deterministic allocations; the PBD yields the most deterministic allocations. Salama, Ivanova and Qaqish (2008) stated that some blocks are smaller in an RBD than they could have been in a PBD, which makes it easier to predict the upcoming assignments if the investigator is using the convergence strategy.

As stated earlier, when the sequence of block size is randomized, (1.10) and (1.11) are inappropriate to measure the potential selection bias of an RBD due to unknown b_i and M. We will propose a metric to measure the degree of randomness of the RBD.

1.5 Efron's Biased Coin Design

1.5.1 Efron's Original Article

Efron (1971) proposed his famous biased coin design (BCD) as a method to force a sequential experiment to be balanced. The allocation rule is defined in (1.3). Efron pointed out in his article that $|D_j|$ forms a Markov chain with states 0,1,2,.... The transition probabilities are given by

$$P(|D_{j+1}| = k+1||D_j| = k) = 1 - p \ (k \geq 1),$$

$$P(|D_{j+1}| = k-1||D_j| = k) = p \ (k \geq 1),$$

$$P(|D_{j+1}| = 1||D_j| = 0) = 1.$$

This is a random walk which has a reflecting barrier at the origin, and the stationary probabilities are the following:

$$\pi_0 = \frac{r-1}{2r}, \quad \pi_j = \frac{r-1}{2r}\frac{r+1}{r^j} \ (j \geq 1),$$

where $r = p/1 - p$. The period of this Markov chain is 2 because $|D_j|$ can only be odd or even values as j is odd or even. Therefore, the limiting probabilities of exact balance and imbalance of 1 are given by

$$\lim_{m\to\infty} P(|D_{2m}| = 0) = 2\pi_0 = \frac{r-1}{r},$$

$$\lim_{m\to\infty} P(|D_{2m+1}| = 1) = 2\pi_1 = \frac{r^2-1}{r^2}.$$

Efron suggested that investigators use $p = 2/3$ (p. 405). The asymptotic probability of the allocation being exactly balanced is 1/2 for even j, and the asymptotic probability of being almost balanced (imbalance of 1) is 3/4 for odd j. The distribution of $|D_j|$ gets closer to 0

when p is increasing for any j, but Efron failed to provide the exact distribution of $|D_j|$ to support his claim.

Efron calculated the selection bias of BCD based on the definition proposed by Blackwell and Hodges (1957). A measure of selection bias is the expectation of correct guesses of the assignment if the investigator guesses optimally. The optimal guess against a BCD is the treatment which has occurred least often so far. The probability of a correct guess for the jth allocation is

$$\frac{1}{2}P(|D_{j-1}| = 0) + pP(|D_{j-1}| > 0).$$

As $j \to \infty$, the asymptotic probability is

$$\frac{1}{2}\pi_0 + p(1 - \pi_0) = \frac{1}{2} + \frac{r-1}{4r}.$$

Then the asymptotic selection bias in j assignments can be computed as

$$\frac{r-1}{4r}j. \tag{1.12}$$

1.5.2 Markaryan and Rosenberger's Work

In 2010, Markaryan and Rosenberger derived the exact distribution of the imbalance of the treatment assignment and the variance-covariance matrix of the treatment assignments. These results provide the means to derive an explicit form of the selection bias and the variance of the imbalance, which reflect the degree of randomness and the variability of the BCD procedure, respectively. Let n be the total number of assignments, $n = 1, 2, 3, \ldots$. Let $0 \leq k \leq n$, k and n have the same parity. Then the distribution of D_n is given by

$$P(D_n = \pm k) = \frac{1}{2} p^{(n-1)/2} \sum_{l=0}^{(n-k)/2} \frac{n+k-2l}{n+k+2l} \binom{\frac{n+k}{2}+l}{l} (1-p)^{k+l-1}, \quad \text{when } k > 0,$$
$$P(D_n = 0) = p^{n/2} \sum_{l=0}^{n/2-1} \frac{n-2l}{n+2l} \binom{\frac{n}{2}+l}{l} (1-p)^{l}, \quad \text{when } k = 0. \tag{1.13}$$

Hence the variance of D_n is given by

$$Var(D_n) = \sum_{k=1,\, n-k\ even}^{n} k^2 p^{(n-k)/2} \sum_{l=0}^{(n-k)/2} \frac{n+k-2l}{n+k+2l} \binom{\frac{n+k}{2}+l}{l} (1-p)^{k+l-1}. \tag{1.14}$$

Note that this variance is increasing in n when p if fixed and is decreasing in p for given n. Using the steady state distribution property of the induced Markov chain, the limiting variance of the imbalance is given by

$$\frac{4r(r^2+1)}{(r^2-1)^2}, \quad \text{when the number of trials is even,}$$

$$\frac{8r^2}{(r^2-1)^2} + 1, \quad \text{when the number of trials is odd.}$$

The variances become stable and approach their limit when the size of the trials is between 75 to 100, which indicates the balancing properties of the BCD procedure stabilize for clinical trials with sizes of 75 to 100.

The total selection bias in n trials is given by

$$\frac{1}{2} + (n-1)p - (p - \frac{1}{2}) \sum_{m=1}^{[(n-1)/2]} p^m \frac{m-l}{m+l} \binom{m+l}{l} (1-p)^{l}, \tag{1.15}$$

where $[a]$ denotes the integer part of a, and the sum is treated as zero if the upper limit of

summation is smaller than the lower limit. Compared with Efron's asymptotic result (1.12), Markaryan and Rosenberger claim that the asymptotic limit is reached when n is as small as 50.

1.6 Big Stick Design

1.6.1 Soares and Wu's Original Article

The BSD was originally proposed by Soares and Wu in 1983. The allocation rule is defined in (1.4). In the original paper, the final imbalance distribution was provided. Under the BSD, $|D_j|$ forms a finite Markov chain with states $0, 1, 2, \ldots, a$ and transition probabilities given by

$$\begin{aligned} P(|D_{j+1}| = k+1 || D_j| = k) &= \frac{1}{2}, \text{ for } \quad 0 < k < a, \\ P(|D_{j+1}| = a-1 || D_j| = a) &= 1, \\ P(|D_{j+1}| = 1 || D_j| = 0) &= 1. \end{aligned}$$

This is a symmetric random walk with two reflecting barriers at the origin and at a, then its stationary distribution can be found in Cox and Miller (1965):

$$\pi_0 = \pi_a = \frac{1}{2a}, \quad \pi_m = \frac{1}{a}, \quad 0 < m < a.$$

The initial distribution $P(|D_0| = 0)$ is 0 and $|D_j|$ has the same parity as j; in other words, $|D_j|$ takes on only odd (or even) values when j is odd (or even). Thus the period of the Markov chain is 2 and the limiting probabilities are twice the stationary probabilities. When a is even, the limiting probabilities are given by

$$\lim_{j\to\infty} P(|D_{2j}| = 0) = \lim_{j\to\infty} P(|D_{2j}| = a) = \frac{1}{a},$$

$$\lim_{j\to\infty} P(|D_{2j}| = m) = \frac{2}{a}, \quad 0 < m < a.$$

And when a is odd, the limiting distribution is

$$\lim_{j\to\infty} P(|D_{2j+1}| = 0) = \lim_{j\to\infty} P(|D_{2j+1}| = a) = 0,$$

$$\lim_{j\to\infty} P(|D_{2j+1}| = m) = \frac{2}{a+1}, \quad 0 < m < a.$$

Soares and Wu also studied the excess selection bias $E(F)$ defined by Blackwell and Hodges, as well as the accidental bias for the BSD. For the calculation, they applied the result of high-order transition probabilities of a random walk with two reflecting barriers in Karlin's book (1968). However, Karlin's solution was proved to be incorrect later and omitted in his second edition. Therefore, Soares and Wu's calculation is incorrect.

1.6.2 Chen's Work on the BSD

Chen (1999) made a correction of Karlin's formula for the BSD(a), $a > 1$, in Corollary 2.1:

> Corollary 2.1: Under the BSD(a), the absolute difference process D_n forms a Markov chain and its j-step transition matrix, $j \geq 0$, has entries
>
> $$P_{l,m}^{(j)} = \eta_m + (-1)^{j+l+m}\eta_m + 2\eta_m \sum_{k=1}^{a-1} \left(cos\frac{k\pi}{a}\right)^j cos\left(\frac{lk\pi}{a}\right) cos\left(\frac{mk\pi}{a}\right), \qquad (1.16)$$
>
> where $0 \leq l, m \leq a$, $\eta_0 = 1/(2a)$, $\eta_1 = \cdots = \eta_{a-1} = 1/a$, and $\eta_a = 1/(2a)$ are the components of the stationary distribution of D_n.

With this formula, he found that, under the BSD (a), the average imbalance between the treatment allocations up to epoch n (1.9) converges to $a/2$. The average excess selection

bias up to epoch n (1.7) is given by

$$\frac{1}{4a} + \frac{(-1)^a - (-1)^{a+n}}{8an} + \frac{1}{2an}\sum_{k=1}^{a-1}\frac{1 - cos^n(k\pi/a)}{1 - cos(k\pi/a)}. \tag{1.17}$$

The asymptotic average excess selection bias is $1/(4a)$. Thus, the asymptotic average excess selection bias is monotonically increasing as a decreases.

1.7 Outline of the Thesis

This thesis is structured as follows. In Chapter 2, we formalize the RBD and derive some important distributional properties. We investigate the degree of predictability of six restricted randomization procedures in Chapter 3 and the variability of treatment assignments imbalance in Chapter 4. In Chapter 5, randomization-base inference from the six designs is explored. Comparisons among the six designs are made in Chapter 6. Chapter 7 is devoted to general conclusions and remarks.

1.8 Contributions of the Thesis

The contributions of this thesis are listed here:

- Rigorously define the RBD.
- Find the exact distributional properties of the RBD such as the joint distribution of the block size and position within the block.
- Compare the RBD to other restricted randomization procedures and conclude that randomizing the block sizes is not better with respect to selection bias.
- Draw conclusions about appropriate procedures in practice.
- Investigate the properties of randomization tests.

Table 1.1: Prediction of future allocations with balanced (1:1) blocks (from Berger 2007, p. 90)

Size	Ratio	Deterministic	Predictable	Correct guesses
2	1:01	50%	50%	75%
4	2:02	33%	58%	71%
6	3:03	25%	63%	68%

Chapter 2: Distributional Properties of the Random Block Design

To avoid deterministic treatment allocations in the PBD, many clinical trialists prefer the random block design (RBD). However, this procedure is rarely carefully defined when it is employed. In this chapter, we formalize the RBD by assuming a discrete uniform distribution for block size. We also provide the distributional properties of the RBD.

Consider a RBD with integer block sizes from $b = 1, ..., B_{max}$, where $2b$ subjects are randomized to either treatment A or B within each block using the random allocation rule. Each block is randomly selected uniformly with a probability $1/B_{max}$. If $T_1, ..., T_n$ are the treatment assignments, where $T_j = 1$ if A and 0 if B, $j = 1, ..., n$, then the randomization procedure is defined as $\phi_j = E(T_j|T_1, ..., T_{j-1})$.

Let B_j be the *block size* of the jth subject, $j = 1, ..., n$, which can take the values 1,...,B_{max}. Note that for n subjects, each block will be filled with the possible exception of the last. We do not assume that n is known in advance. The *position number*, given B_j is denoted R_j, $j = 1, ..., n$, and takes the values $1, ..., 2B_j$. If $R_n = 2B_n$, then every block is filled; otherwise, the last block is unfilled. Let $N_A(n)$ be the total number of patients assigned to treatment A after assigning the nth patient, then $N_A(n) = \sum_{j=1}^{n} T_j$ and is a random variable. Let $N_A^S(j)$ denote the number of patients assigned to treatment A only within the block of the jth patient, before assigning the jth patient to any treatment, where $N_A^S(j) = \sum_{l=j+1-R_j}^{j-1} T_l$. Note that if $R_j = 1$, $N_A^S(j)$ is 0. Given the position number, $N_A^S(j)$ is still a random variable since T_l is random. Using the random allocation rule and knowing the block size B_j and the position number R_j, the probability of assigning the jth patient

to treatment A is given by

$$\phi_j = E(T_j|T_1, ..., T_{j-1}, B_j, R_j) = \frac{B_j - N_A^S(j)}{2B_j - R_j + 1}. \tag{2.1}$$

When assigning patients to each treatment using the truncated binomial design, the allocation probability with the knowledge of block size and position number is given by

$$\phi_j = \begin{cases} 1/2, & \text{if } \max\{N_A^S(j), N_B^S(j)\} < B_j, \\ 0, & \text{if } N_A^S(j) = B_j, \\ 1, & \text{if } N_B^S(j) = B_j. \end{cases}$$

In the future, we refer to each design as RBD(B_{max}; R) for those filled using the RAR and RBD(B_{max}; T) for those filled using the TBD.

This chapter lists 12 lemmas, which cover the following distributional results:

- joint distribution of the block size and the position number,
- distribution of the block size,
- conditional distribution of the position number within a block, given the block size,
- conditional distribution of the number of patients assigned to treatment A within a block so far, given the block size and position number,
- conditional distribution of the imbalance, given the block size and position number,
- distribution of the imbalance.

All these results are obtained for the first time. With the help of these results, we are able to derive the exact selection bias of the RBD, which will be given in a later chapter.

2.1 Distributional Properties of the RBD

In this section, we present some distributional properties of the RBD. In all the proofs of this chapter, we treat summations as 0 if the upper limit of the summation is smaller than the lower limit.

Lemma 1 presents the joint distribution of the position number and the block size for $j \leq 2B_{max} + 2$.

Lemma 1. *Let b be an integer from 1 to B_{max}. Let r be an integer from 1 to $2b$. Let r and j have the same parity. For $j \leq 2B_{max} + 2$, the joint distribution of B_j and R_j is given by*

$$P(R_j = r, B_j = b) = \begin{cases} \dfrac{1}{B_{max}}, & 1 \leq r = j \leq 2b, \\ \dfrac{1}{B_{max}(B_{max}+1)} \left(\dfrac{B_{max}+1}{B_{max}}\right)^{\frac{j-r}{2}}, & 1 \leq r \leq \min(j-2, 2b), \\ 0, & r > \min(j, 2b). \end{cases} \quad (2.2)$$

Proof. This proof proceeds by induction.

Note that since r and j have the same parity, the following equation holds when j is odd:

$$P(R_j = r, B_j = b) = P(R_{j+1} = r + 1, B_{j+1} = b). \quad (2.3)$$

However, equation (2.3) does not hold when j is even.

We prove equation (2.2) is true when j is odd first. Equation (2.2) is trivially true for the case $j = 1$ because of the following:

$$P(R_1 = 1, B_1 = b) = P(B_1 = b) = \frac{1}{B_{max}}.$$

We assume equation (2.2) is true for j when j is any positive odd integer up to and including $2B_{max} - 1$. Then we prove that (2.2) works for $j + 2$. We separate the proof into two parts:

$r = 1$ and $r \geq 3$.

Case 1: $r = 1$ and $j \leq 2B_{max} - 1$ and j is odd:

$$\begin{aligned} P(R_{j+2} = r, B_{j+2} = b) &= P(R_{j+2} = 1, B_{j+2} = b) \\ &= P(R_{j+1} = 2B_{j+1}, B_{j+2} = b) \\ &= P(B_{j+2} = b | R_{j+1} = 2B_{j+1}) P(R_{j+1} = 2B_{j+1}) \\ &= \frac{1}{B_{max}} P(R_{j+1} = 2B_{j+1}) \\ &= \frac{1}{B_{max}} P(R_j = 2B_j - 1). \end{aligned} \tag{2.4}$$

To complete the proof, we need to find $P(R_j = 2B_j - 1)$.

$$\begin{aligned} P(R_j = 2B_j - 1) &= \sum_{b=1}^{(j+1)/2} P(R_j = 2b - 1, B_j = b) \qquad \text{since } j + 1 \leq 2B_{max} \\ &= \sum_{b=1}^{(j-1)/2} \frac{1}{B_{max}(B_{max}+1)} \left(\frac{B_{max}+1}{B_{max}} \right)^{\frac{j+1-2b}{2}} + \frac{1}{B_{max}} \\ &= \frac{1}{B_{max}(B_{max}+1)} \left(\frac{B_{max}+1}{B_{max}} \right)^{\frac{j+1}{2}} \sum_{b=1}^{(j-1)/2} \left(\frac{B_{max}}{B_{max}+1} \right)^{b} + \frac{1}{B_{max}} \\ &= \frac{1}{B_{max}(B_{max}+1)} \left(\frac{B_{max}+1}{B_{max}} \right)^{\frac{j+1}{2}} \left[\frac{\left(\frac{B_{max}}{B_{max}+1} \right)^{(j+1)/2} - \left(\frac{B_{max}}{B_{max}+1} \right)}{\frac{B_{max}}{B_{max}+1} - 1} \right] + \frac{1}{B_{max}} \\ &= \frac{1}{B_{max}(B_{max}+1)} \left(\frac{B_{max}+1}{B_{max}} \right)^{\frac{j+1}{2}} \left[B_{max} - \left(\frac{B_{max}}{B_{max}+1} \right)^{\frac{j+1}{2}} (B_{max}+1) \right] + \frac{1}{B_{max}} \\ &= \frac{1}{B_{max}+1} \left(\frac{B_{max}+1}{B_{max}} \right)^{\frac{j+1}{2}} - \frac{1}{B_{max}} + \frac{1}{B_{max}} \\ &= \frac{1}{B_{max}+1} \left(\frac{B_{max}+1}{B_{max}} \right)^{\frac{j+1}{2}}. \end{aligned} \tag{2.5}$$

Now we plug equation (2.5) back into equation (2.4), and we get

$$P(R_{j+2}=1, B_{j+2}=b) = \frac{1}{B_{max}(B_{max}+1)} \left(\frac{B_{max}+1}{B_{max}}\right)^{\frac{j+1}{2}},$$

which is exactly the right hand side (RHS) of equation (2.2) with j replaced by $j+2$.

Case 2: $r \geq 3$ and $j \leq 2B_{max}-1$ and j is odd:

$$P(R_{j+2}=r, B_{j+2}=b) = P(R_j = r-2, B_j = b)$$

$$= \begin{cases} \frac{1}{B_{max}}, & j+2=r\geq 3, \\ \frac{1}{B_{max}(B_{max}+1)} \left(\frac{B_{max}+1}{B_{max}}\right)^{\frac{j+2-r}{2}}, & j+2>r\geq 3, \\ 0, & j+2<r. \end{cases}$$

This proves equation (2.2) holds for all $1 \leq j \leq 2B_{max}+2$ when j is odd.

Next, we prove equation (2.2) holds for all $1 \leq j \leq 2B_{max}+2$ when j is even. It is true since $P(R_j = r, B_j = b) = P(R_{j-1} = r-1, B_{j-1} = b)$ when j is even. □

There are two corollaries following Lemma 1.

Corollary 1. *Let b be an integer from 1 to B_{max}. Let r be an integer from 1 to $2b$. Let r and j have the same parity. The joint distribution of the block size and the position number for any integer j, $P(R_j = r, B_j = b)$, does not depend on the value of b.*

Proof. For $j \leq 2B_{max}+2$, we see the joint distribution is only a function of the position number r and does not depend on the block size b from Lemma 1. Therefore, we need to prove it also does not depend on b for $j \geq 2B_{max}+3$. We have separated the proof into two cases: j is odd and j is even.

Case 1: $j \geq 2B_{max}+3$ and j is odd, for any $1 < r \leq 2b$. We have

$$P(R_j = r, B_j = b) = P(R_{j-(r-1)} = 1, B_{j-(r-1)} = b) = P(R_{j-r+1} = 1, B_{j-r+1} = b).$$

Hence, we only need to prove that $P(R_j = 1, B_j = b)$ is independent of b in this case.

$$\begin{aligned}
P(R_j = 1, B_j = b) &= P(R_{j-1} = 2B_{j-1}, B_j = b) \\
&= P(B_j = b | R_{j-1} = 2B_{j-1}) P(R_{j-1} = 2B_{j-1}) \\
&= \frac{1}{B_{max}} P(R_{j-1} = 2B_{j-1}) \\
&= \frac{1}{B_{max}} \sum_{b=1}^{B_{max}} P(R_{j-1} = 2b, B_{j-1} = b) \\
&= \frac{1}{B_{max}} \sum_{b=1}^{B_{max}} P(R_{j-1-(2b-2)} = 2, B_{j-1-(2b-2)} = b) \\
&= \frac{1}{B_{max}} \sum_{b=1}^{B_{max}} P(R_{j-2b+1} = 2, B_{j-2b+1} = b) \\
&= \frac{1}{B_{max}} \sum_{b=1}^{B_{max}} P(R_{j-2b} = 1, B_{j-2b} = b).
\end{aligned}$$

This is a recursive formula, hence it can be rewritten as a function of the initial B_{max} values: $P(R_1 = 1, B_1 = b), P(R_3 = 1, B_3 = b), \ldots, P(R_{2B_{max}-1} = 1, B_{2B_{max}-1} = b)$. As we know from Lemma 1, these initial values are independent of b. Therefore, $P(R_j = 1, B_j = b)$ is not a function of b when $j \geq 2B_{max} + 3$ and j is odd.

Case 2: $j \geq 2B_{max} + 3$ and j is even, for any $1 < r \leq 2b$. We have

$$P(R_j = r, B_j = b) = P(R_{j-(r-2)} = 2, B_{j-(r-2)} = b) = P(R_{j-r+2} = 2, B_{j-r+2} = b).$$

Hence, we only need to prove $P(R_j = 2, B_j = b)$ is independent of b when $j \geq 2B_{max}+3$ and j is even. This is trivially true because $P(R_j = 2, B_j = b) = P(R_{j-1} = 1, B_{j-1} = b)$. $\square$

By Corollary 1, the following equation is true:

$$P(R_j = r, B_j = b) = P(R_j = r, B_j = B_{max}), \quad \forall\, r \leq 2b.$$

Corollary 2. *For any odd j and $j > 1$,*

$$P(R_j = 1, B_j = b) = \frac{1}{B_{max}} P(B_{j-1} = B_{max}).$$

Proof. We separate the proof into two parts: $3 \leq j \leq 2B_{max}$ and $j > 2B_{max}$. When $3 \leq j \leq 2B_{max}$,

$$P(R_j = 1, B_j = b) = \frac{1}{B_{max}} P(R_{j-1} = 2B_{j-1})$$

$$= \frac{1}{B_{max}} \sum_{b=1}^{(j-1)/2} P(R_{j-1} = 2b, B_{j-1} = b)$$

$$= \tfrac{1}{B_{max}} \left[P(R_{j-1} = 2, B_{j-1} = 1) + P(R_{j-1} = 4, B_{j-1} = 2) + \ldots + P(R_{j-1} = j-1, B_{j-1} = \tfrac{j-1}{2})\right]$$

$$= \tfrac{1}{B_{max}} \left[P(R_{j-1} = 2, B_{j-1} = B_{max}) + P(R_{j-1} = 4, B_{j-1} = B_{max}) + \ldots + P(R_{j-1} = j-1, B_{j-1} = B_{max}\right]$$

$$= \frac{1}{B_{max}} P(B_{j-1} = B_{max}).$$

When $j > 2B_{max}$,

$$P(R_j = 1, B_j = b) = \frac{1}{B_{max}} P(R_{j-1} = 2B_{j-1})$$

$$= \frac{1}{B_{max}} \sum_{b=1}^{B_{max}} P(R_{j-1} = 2b, B_{j-1} = b)$$

$$= \frac{1}{B_{max}} \sum_{r=2}^{2B_{max}} P(R_{j-1} = r, B_{j-1} = B_{max})$$

$$= \frac{1}{B_{max}} P(B_{j-1} = B_{max}).$$

Corollary 2 is proven. □

Corollary 1 and Corollary 2 will be significant later in this chapter in the proof of Lemma 4.

The following Lemma provides the distribution of the block size.

Lemma 2. *For $j \leq 2B_{max}+2$ and j is odd, the distribution of B_j is given by*

$$P(B_j = b) = \begin{cases} \frac{1}{B_{max}} \left(\frac{B_{max}+1}{B_{max}}\right)^{(j-1)/2} \left[1 - \left(\frac{B_{max}}{B_{max}+1}\right)^b\right], & 1 \leq b \leq \frac{j-1}{2}, \\ \frac{1}{B_{max}} \left(\frac{B_{max}+1}{B_{max}}\right)^{(j-1)/2}, & \frac{j+1}{2} \leq b \leq B_{max}. \end{cases} \tag{2.6}$$

For $j \leq 2B_{max}+2$ and j is even, the distribution of B_j is given by

$$P(B_j = b) = \begin{cases} \frac{1}{B_{max}} \left(\frac{B_{max}+1}{B_{max}}\right)^{(j-2)/2} \left[1 - \left(\frac{B_{max}}{B_{max}+1}\right)^b\right], & 1 \leq b \leq \frac{j}{2} - 1, \\ \frac{1}{B_{max}} \left(\frac{B_{max}+1}{B_{max}}\right)^{(j-2)/2}, & \frac{j}{2} \leq b \leq B_{max}. \end{cases} \tag{2.7}$$

Proof. By Lemma 1, for $1 \leq j \leq 2B_{max}+2$, we have

$$P(B_j = b) = \begin{cases} \sum_{r=1}^{2b} P(B_j = b, R_j = r), & 2 \leq 2b \leq j-1, \\ \sum_{r=1}^{j} P(B_j = b, R_j = r), & 2b \geq j. \end{cases} \tag{2.8}$$

Note that r and j have the same parity.

We first prove (2.6) is true. Using (2.2), the first part of (2.8) is given by

$$P(B_j = b) = \sum_{r=1}^{2b} P(B_j = b, R_j = r)$$

$$= \sum_{r=1}^{2b-1} \frac{1}{B_{max}(B_{max}+1)} \left(\frac{B_{max}+1}{B_{max}}\right)^{\frac{j-r}{2}}$$

$$= \frac{1}{B_{max}(B_{max}+1)} \left(\frac{B_{max}+1}{B_{max}}\right)^{\frac{j}{2}} \sum_{r=1}^{2b-1} \left(\frac{B_{max}}{B_{max}+1}\right)^{\frac{r}{2}}$$

$$= \tfrac{1}{B_{max}(B_{max}+1)} \left(\tfrac{B_{max}+1}{B_{max}}\right)^{\frac{j}{2}} \left[\left(\tfrac{B_{max}}{B_{max}+1}\right)^{\frac{1}{2}} + \left(\tfrac{B_{max}}{B_{max}+1}\right)^{\frac{3}{2}} + \cdots + \left(\tfrac{B_{max}}{B_{max}+1}\right)^{\frac{2b-1}{2}}\right]$$

$$= \frac{1}{B_{max}(B_{max}+1)} \left(\frac{B_{max}+1}{B_{max}}\right)^{\frac{j}{2}} \frac{\left(\frac{B_{max}}{B_{max}+1}\right)^{n+\frac{1}{2}} - \left(\frac{B_{max}}{B_{max}+1}\right)^{\frac{1}{2}}}{\left(\frac{B_{max}}{B_{max}+1}\right) - 1}$$

$$= \frac{1}{B_{max}} \left(\frac{B_{max}+1}{B_{max}}\right)^{\frac{j}{2}} \left[\left(\frac{B_{max}}{B_{max}+1}\right)^{\frac{1}{2}} - \left(\frac{B_{max}}{B_{max}+1}\right)^{b+\frac{1}{2}}\right]$$

$$= \frac{1}{B_{max}} \left(\frac{B_{max}+1}{B_{max}}\right)^{\frac{j-1}{2}} \left[1 - \left(\frac{B_{max}}{B_{max}+1}\right)^{b}\right]. \tag{2.9}$$

The second part of (2.8) is given by

$$P(B_j = b) = \sum_{r=1}^{j} P(B_j = b, R_j = r)$$

$$= \tfrac{1}{B_{max}(B_{max}+1)} \left(\tfrac{B_{max}+1}{B_{max}}\right)^{\frac{j}{2}} \left[\left(\tfrac{B_{max}}{B_{max}+1}\right)^{\frac{1}{2}} + \left(\tfrac{B_{max}}{B_{max}+1}\right)^{\frac{3}{2}} + \cdots + \left(\tfrac{B_{max}}{B_{max}+1}\right)^{\frac{j-2}{2}}\right] + \tfrac{1}{B_{max}}$$

$$= \frac{1}{B_{max}(B_{max}+1)} \left(\frac{B_{max}+1}{B_{max}}\right)^{\frac{j}{2}} \frac{\left(\frac{B_{max}}{B_{max}+1}\right)^{\frac{j}{2}} - \left(\frac{B_{max}}{B_{max}+1}\right)^{\frac{1}{2}}}{\left(\frac{B_{max}}{B_{max}+1}\right) - 1} + \frac{1}{B_{max}}$$

$$= \frac{1}{B_{max}} \left(\frac{B_{max}+1}{B_{max}}\right)^{\frac{j-1}{2}} - \frac{1}{B_{max}} + \frac{1}{B_{max}}$$

$$= \frac{1}{B_{max}} \left(\frac{B_{max}+1}{B_{max}}\right)^{\frac{j-1}{2}}. \tag{2.10}$$

Equations (2.9) and (2.10) complete the proof of (2.6) in Lemma 2.

The proof of (2.7) is trivial because when j is even, $P(B_j = b) = P(B_{j-1} = b)$. This

completes the proof of Lemma 2. □

Lemma 3 describes the conditional distribution of the position number given the block size.

Lemma 3. *For $j \leq 2B_{max}+2$ and j is odd, the distribution of R_j, given the block size B_j, is given by*

$$P(R_j = r|B_j) = \begin{cases} \left(\frac{B_{max}}{B_{max}+1}\right)^{\frac{r-1}{2}}, & r = j \leq 2B_j - 1, \\ \frac{1}{B_{max}+1}\left(\frac{B_{max}}{B_{max}+1}\right)^{\frac{r-1}{2}}, & r < j \leq 2B_j - 1, \\ \frac{\frac{1}{B_{max}+1}\left(\frac{B_{max}}{B_{max}+1}\right)^{\frac{r-1}{2}}}{1-\left(\frac{B_{max}}{B_{max}+1}\right)^{B_j}}, & r < 2B_j + 1 \leq j \leq 2B_{max}+1, \\ 0, & r > \min\{2B_j - 1, j\}. \end{cases}$$

For $j \leq 2B_{max}+2$ and j is even, the distribution of R_j, given the block size B_j, is given by

$$P(R_j = r|B_j) = \begin{cases} \left(\frac{B_{max}}{B_{max}+1}\right)^{\frac{r-1}{2}}, & r = j \leq 2B_j - 1, \\ \frac{1}{B_{max}+1}\left(\frac{B_{max}}{B_{max}+1}\right)^{\frac{r}{2}-1}, & r < j \leq 2B_j - 1, \\ \frac{\frac{1}{B_{max}+1}\left(\frac{B_{max}}{B_{max}+1}\right)^{\frac{r}{2}-1}}{1-\left(\frac{B_{max}}{B_{max}+1}\right)^{B_j}}, & r < 2B_j + 1 \leq j \leq 2B_{max}+1, \\ 0, & r > \min\{2B_j - 1, j\}. \end{cases}$$

Proof. The Lemma 3 can be proved by deriving the conditional probability using Lemma 1 and Lemma 2:

$$P(R_j = r|B_j) = \frac{P(R_j = r, B_j = b)}{P(B_j = b)}.$$ □

In Lemma 4, we will show that the probability of the block size is a recursive sequence.

Lemma 1 and Lemma 4 together can be used to derive the exact distribution of the block size for a RBD. Lemma 1, Lemma 4 and Lemma 5 can also be used together to obtain the exact joint distribution of the position number and the block size.

Lemma 4. *For $j \geq 2B_{max}+1$, the distribution of the block size B_j is given by*

$$P(B_j = b) = \frac{1}{B_{max}} \sum_{i=1}^{B_{max}} P(B_{j-2i} = b). \tag{2.11}$$

Proof. We will prove Lemma 4 by induction. The proof is broken into two parts: j is odd and j is even.

Case 1: j is odd. When $j = 2B_{max}+1$, the LHS of (2.11) is

$$\begin{aligned} P(B_{2B_{max}+1} = b) =& \frac{1}{B_{max}} \left(\frac{B_{max}+1}{B_{max}}\right)^{(2B_{max}+1-1)/2} \left[1 - \left(\frac{B_{max}}{B_{max}+1}\right)^b\right] \\ =& \frac{1}{B_{max}} \left(\frac{B_{max}+1}{B_{max}}\right)^{B}_{max} \left[1 - \left(\frac{B_{max}}{B_{max}+1}\right)^b\right]. \end{aligned}$$

The RHS of (2.11) is

$$\begin{aligned} & \frac{1}{B_{max}} \sum_{i=1}^{B_{max}} P(B_{2B_{max}+1-2i} = b) \\ &= \frac{1}{B_{max}} [P(B_1 = b) + P(B_3 = b) + \ldots + P(B_{2B_{max}-1} = b)] \\ &= \frac{1}{B_{max}} \left\{ \sum_{i=1}^{b} \frac{1}{B_{max}} \left(\frac{B_{max}+1}{B_{max}}\right)^{i-1} + \frac{1}{B_{max}} \left[1 - \left(\frac{B_{max}}{B_{max}+1}\right)^b\right] \sum_{i=b}^{B_{max}-1} \left(\frac{B_{max}+1}{B_{max}}\right)^i \right\} \\ &= \frac{1}{B_{max}} \left\{ \frac{1}{B_{max}+1} \sum_{i=1}^{b} \left(\frac{B_{max}+1}{B_{max}}\right)^i + \frac{1}{B_{max}} \left[1 - \left(\frac{B_{max}}{B_{max}+1}\right)^b\right] \sum_{i=b}^{B_{max}-1} \left(\frac{B_{max}+1}{B_{max}}\right)^i \right\} \\ &= \frac{1}{B_{max}} \left\{ \frac{1}{B_{max}+1} \frac{\left(\frac{B_{max}+1}{B_{max}}\right)^{b+1} - \frac{B_{max}+1}{B_{max}}}{\frac{B_{max}+1}{B_{max}} - 1} + \frac{1}{B_{max}} \left[1 - \left(\frac{B_{max}}{B_{max}+1}\right)^b\right] \frac{\left(\frac{B_{max}+1}{B_{max}}\right)^{B_{max}} - \left(\frac{B_{max}+1}{B_{max}}\right)^b}{\frac{B_{max}+1}{B_{max}} - 1} \right\} \end{aligned}$$

$$= \frac{1}{B_{max}} \left\{ \left[\left(\frac{B_{max}}{B_{max}+1} \right)^b - 1 \right] + \left[1 - \left(\frac{B_{max}}{B_{max}+1} \right)^b \right] \left[\left(\frac{B_{max}+1}{B_{max}} \right)^{B_{max}} - \left(\frac{B_{max}+1}{B_{max}} \right)^b \right] \right\}$$

$$= \frac{1}{B_{max}} \left\{ \left(\frac{B_{max}+1}{B_{max}} \right)^b - 1 + \left(\frac{B_{max}+1}{B_{max}} \right)^{B_{max}} - \left(\frac{B_{max}+1}{B_{max}} \right)^b - \left(\frac{B_{max}+1}{B_{max}} \right)^{B_{max}-b} + 1 \right\}$$

$$= \frac{1}{B_{max}} \left(\frac{B_{max}+1}{B_{max}} \right)^{B_{max}} \left[1 - \left(\frac{B_{max}}{B_{max}+1} \right)^b \right].$$

So both sides are equal and (2.11) is true for $2B_{max}+1$.

Now we suppose (2.11) is true for any odd $j \geq 2B_{max}+1$:

$$P(B_j = b) = \frac{1}{B_{max}} \sum_{i=1}^{B_{max}} P(B_{j-2i} = b)$$

$$= \frac{1}{B_{max}} \left[P(B_{j-2} = b) + P(B_{j-4} = b) + \ldots + P(B_{j-2B_{max}} = b) \right].$$

We will show that (2.11) holds for $j+2$. We consider two cases: $b < B_{max}$ and $b = B_{max}$. In the case of $b < B_{max}$, the LHS of (2.11) is

$$P(B_{j+2} = b) = \sum_{r=1}^{2b-1} P(B_{j+2} = b, R_{j+2} = r)$$

$$= P(R_{j+2} = 1, B_{j+2} = b) + P(R_{j+2} = 3, B_{j+2} = b) + \ldots + P(R_{j+2} = 2b-1, B_{j+2} = b)$$

$$= P(R_{j+2} = 1, B_{j+2} = b) + P(R_j = 1, B_j = b) + P(R_j = 3, B_j = b) + \ldots + P(R_j = 2b-3, B_j = b)$$

$$= P(R_{j+2} = 1, B_{j+2} = b) + \sum_{r=1}^{2b-1} P(R_j = r, B_j = b) - P(R_j = 2b-1, B_j = b)$$

$$= \frac{1}{B_{max}} P(R_{j+1} = 2B_{j+1}) + P(B_j = b) - P(R_j = 2b-1, B_j = b). \tag{2.12}$$

We will find $P(R_{j+1=}2B_{j+1)}$ first.

$$P(R_{j+1} = 2B_{j+1}) = \sum_{b=1}^{B_{max}} P(R_{j+1} = 2b, B_{j+1} = b) = \sum_{b=1}^{B_{max}} P(R_j = 2b-1, B_j = b)$$

$$= P(R_j = 1, B_j = 1) + P(R_j = 3, B_j = 2) + \ldots + P(R_j = 2B_{max} - 1, B_j = B_{max})$$

$$= P(R_j = 1, B_j = b) + \ldots + P(R_j = 2b-1, B_j = b)$$

$$+ P(R_j = 2b+1, B_j = B_{max}) + \ldots + P(R_j = 2B_{max} - 1, B_j = B_{max}) \quad \text{by Corollary 1}$$

$$= \sum_{r=1}^{2b-1} P(R_j = r, B_j = b) + \sum_{r=2b+1}^{2B_{max}-1} P(R_j = r, B_j = B_{max})$$

$$= P(B_j = b) + \sum_{r=2b+1}^{2B_{max}-1} P(R_j = r, B_j = B_{max})$$

$$= P(B_j = b) + \sum_{r=1}^{2B_{max}-2b-1} P(R_{j-2b} = r, B_{j-2b} = B_{max}). \tag{2.13}$$

Next we rewrite $P(R_j = 2b-1, B_j = b)$ using Corollary 2.

$$P(R_j = 2b-1, B_j = b) = P(R_{j-2b+2} = 1, B_{j-2b+2} = b)$$

$$= \frac{1}{B_{max}} P(B_{j-2b+1} = B_{max}) \quad \text{by Corollary 2}$$

$$= \frac{1}{B_{max}} P(B_{j-2b} = B_{max}) \quad \text{since } j \text{ is odd}$$

$$= \frac{1}{B_{max}} \sum_{r=1}^{\min(2B_{max}-1, j-2b)} P(R_{j-2b} = r, B_{j-2b} = B_{max}). \tag{2.14}$$

Now we plug (2.13) and (2.14) back into (2.12), we get

$$\frac{1}{B_{max}} \left[P(B_j = b) + P(B_j = b) + \sum_{r=1}^{2B_{max}-2b-1} P(R_{j-2b} = r, B_{j-2b} = B_{max}) - \sum_{r=1}^{\min(2B_{max}-1, j-2b)} P(R_{j-2b} = r, B_{j-2b} = B_{max}) \right]$$

$$= \frac{1}{B_{max}} \left[P(B_j = b) + P(B_j = b) - \sum_{r=2B_{max}-2b+1}^{\min(2B_{max}-1, j-2b)} P(R_{j-2b} = r, B_{j-2b} = B_{max}) \right]$$

$$= \frac{1}{B_{max}} \left[P(B_j = b) + P(B_j = b) - \sum_{r=1}^{\min(2b-1, j-2B_{max})} P(R_{j-2B_{max}} = r, B_{j-2B_{max}} = B_{max}) \right]$$

$$= \frac{1}{B_{max}} \left[P(B_j = b) + P(B_j = b) - \sum_{r=1}^{\min(2b-1, j-2B_{max})} P(R_{j-2B_{max}} = r, B_{j-2B_{max}} = b) \right]$$

by Corollary 1

$$= \frac{1}{B_{max}} \left[P(B_j = b) + P(B_j = b) - P(B_{j-2B_{max}} = b) \right].$$

The RHS of (2.11) is

$$\frac{1}{B_{max}} \sum_{i=1}^{B_{max}} P(B_{j+2-2i} = b)$$

$$= \frac{1}{B_{max}} \left[P(B_j = b) + P(B_{j-2} = b) + P(B_{j-4} = b) + \ldots + P(B_{j+2-2B_{max}} = b) \right]$$

$$= \frac{1}{B_{max}} P(B_j = b) + \frac{1}{B_{max}} \left[P(B_{j-2} = b) + \ldots \right.$$

$$\left. + P(B_{j+2-2B_{max}} = b) + P(B_{j-2B_{max}} = b) \right] - \frac{1}{B_{max}} P(B_{j-2B_{max}} = b)$$

$$= \frac{1}{B_{max}} P(B_j = b) + P(B_j = b) - \frac{1}{B_{max}} P(B_{j-2B_{max}} = b).$$

The LHS and the RHS are equal, hence (2.11) holds for the case j is odd and $b < B_{max}$.

When $b = B_{max}$,

$$P(B_{j+2} = B_{max}) = \sum_{r=1}^{2B_{max}-1} P(R_{j+2} = r, B_{j+2} = B_{max})$$

$$= P(R_{j+2} = 1, B_{j+2} = B_{max}) + \sum_{r=3}^{2B_{max}-1} P(R_{j+2} = r, B_{j+2} = B_{max})$$

$$= P(R_{j+2} = 1, B_{j+2} = B_{max}) + \sum_{r=1}^{2B_{max}-1} P(R_j = r, B_j = B_{max}) - P(R_j = 2B_{max} - 1, B_j = B_{max})$$

$$= \frac{1}{B_{max}} P(B_{j+1} = B_{max}) + P(B_j = B_{max}) - P(R_{j-2B_{max}+2} = 1, B_{j-2B_{max}+2} = B_{max})$$

by Corollary 2

$$= \frac{1}{B_{max}} P(B_{j+1} = B_{max}) + P(B_j = B_{max}) - \frac{1}{B_{max}} P(B_{j-2B_{max}+1} = B_{max}) \quad \text{by Corollary 2 again}$$

$$= \frac{1}{B_{max}} P(B_j = B_{max}) + P(B_j = B_{max}) - \frac{1}{B_{max}} P(B_{j-2B_{max}} = B_{max}) \quad \text{since } j \text{ is odd}$$

$$= \frac{1}{B_{max}} P(B_j = B_{max}) + \frac{1}{B_{max}} [P(B_{j-2} = B_{max}) + P(B_{j-4} = B_{max}) + \dots$$

$$+P(B_{j-2B_{max}} = B_{max})] - \frac{1}{B_{max}} P(B_{j-2B_{max}} = B_{max})$$

$$= \frac{1}{B_{max}} P(B_j = B_{max}) + \frac{1}{B_{max}} P(B_{j-2} = B_{max}) + \dots + \frac{1}{B_{max}} P(B_{j+2-2B_{max}} = B_{max})$$

$$= \frac{1}{B_{max}} \sum_{i=1}^{B_{max}} P(B_{j+2-2i} = B_{max}).$$

The proof of Lemma 4 when j is odd is complete.

Case 2: j is even.

$$P(B_j = b) = P(B_{j-1} = b)$$

$$= \frac{1}{B_{max}} \sum_{i=1}^{B_{max}} P(B_{j-1-2i} = b) \quad \text{since we have proved that Lemma 4 holds for odd numbers,}$$

$$= \frac{1}{B_{max}} \sum_{i=1}^{B_{max}} P(B_{j-2i} = B_{max}) \quad \text{since } j \text{ is even.}$$

The proof of Lemma 4 is complete. □

We treat $P(B_j = 0) = 0$ in the following lemma.

Lemma 5. *Let b be an integer from 1 to B_{max}. Let r be an integer from 1 to $2b$. Let r and j have the same parity. When j is odd, the joint distribution of R_j and B_j is given by*

$$P(R_j = r, B_j = b) = \begin{cases} P\left(B_j = \frac{r+1}{2}\right) - P\left(B_j = \frac{r-1}{2}\right), & r \leq \min\{2b-1, j\}, \\ 0, & \textit{otherwise.} \end{cases} \tag{2.15}$$

When j is even, the joint distribution of R_j and B_j is given by

$$P(R_j = r, B_j = b) = \begin{cases} P\left(B_j = \frac{r}{2}\right) - P\left(B_j = \frac{r}{2} - 1\right), & r \leq \min\{2b, j\}, \\ 0, & \textit{otherwise.} \end{cases} \tag{2.16}$$

Proof. We first prove Lemma 5 is true when j is odd, and separate the proof into two cases: $j = 1$ and $j > 1$. For $j = 1$,

$$P(R_1 = 1, B_1 = b) = P(R_1 = 1, B_1 = b) = \frac{1}{B_{max}} = P(B_1 = 1).$$

For $j > 1$, the RHS of (2.15) is

$$\begin{aligned} & P\left(B_j = \frac{r+1}{2}\right) - P\left(B_j = \frac{r-1}{2}\right) \\ &= \sum_{i=1}^{r} P\left(R_j = i, B_j = \frac{r+1}{2}\right) - \sum_{i=1}^{r-2} P\left(R_j = i, B_j = \frac{r-1}{2}\right) \\ &= \sum_{i=1}^{r} P(R_j = i, B_j = b) - \sum_{i=1}^{r-2} P(R_j = i, B_j = b) \quad \text{by Corollary 1 and } r \leq 2b \\ &= P(R_j = r, B_j = b). \end{aligned}$$

Both sides are equal and the lemma holds when j is odd.

Next we prove equation (2.16) is true when j is even. For $j = 2$,

$$P(R_2 = 2, B_2 = b) = P(R_1 = 1, B_1 = b) = \frac{1}{B_{max}} = P(B_2 = 1).$$

For $j > 1$, the RHS of (2.16) is

$$\begin{aligned} P\left(B_j = \frac{r}{2}\right) - P\left(B_j = \frac{r}{2} - 1\right) &= \sum_{i=2}^{r} P\left(R_j = i, B_j = \frac{r}{2}\right) - \sum_{i=2}^{r-2} P\left(R_j = i, B_j = \frac{r}{2} - 1\right) \\ &= \sum_{i=2}^{r} P(R_j = i, B_j = b) - \sum_{i=2}^{r-2} P(R_j = i, B_j = b) \quad \text{by Corollary 1 and } r \leq 2b \\ &= P(R_j = r, B_j = b). \end{aligned}$$

This completes the proof. □

Lemma 6. *Let b be an integer from 1 to B_{max}. The distribution of B_j when j is odd and greater than $2B_{max}$ is given by*

$$P(B_j = b) = \frac{1}{B_{max}(B_{max}+1)} \sum_{h=1}^{B_{max}} \left(2h + \sum_{i=1}^{B_{max}-1} \sum_{m=1}^{h} \lambda_i^{\frac{j+1}{2} - B_{max} - m}\right) P(B_{2h-1} = b),$$

and for even $j > 2B_{max}$, the distribution of B_j is given by

$$P(B_j = b) = \frac{1}{B_{max}(B_{max}+1)} \sum_{h=1}^{B_{max}} \left(2h + \sum_{i=1}^{B_{max}-1} \sum_{m=1}^{h} \lambda_i^{\frac{j}{2} - B_{max} - m}\right) P(B_{2h-1} = b),$$

where $\lambda_1, \ldots, \lambda_{B_{max}-1}$ are all roots of the equation

$$x^{B_{max}-1} + \frac{B_{max}-1}{B_{max}} x^{B_{max}-2} + \frac{B_{max}-2}{B_{max}} x^{B_{max}-3} + \cdots + \frac{2}{B_{max}} x + \frac{1}{B_{max}} = 0. \quad (2.17)$$

Proof. We only prove 2.11 holds for odd j. We use the following simplified notation throughout the proof.

- $a_1, a_2, \ldots, a_{B_{max}}$ denote $P(B_1 = b), P(B_3 = b), \ldots, P(B_{2B_{max}-1} = b)$, which is known from Lemma 2;
- $a_{n+B_{max}}$ denotes $P(B_j = b)$ for odd $j > 2B_{max}$, where $n = \frac{j+1}{2} - B_{max}$.

Lemma 4 shows that $a_{n+B_{max}}$ is the nth order of a linear recursive sequence and

$$a_{n+B_{max}} = \frac{1}{B_{max}} \left(a_{n+B_{max}-1} + a_{n-B_{max}-2} + \cdots + a_n\right)$$

This linear recursive sequence system satisfies

$$\begin{bmatrix} a_{n+B_{max}} \\ a_{n+B_{max}-1} \\ a_{n+B_{max}-2} \\ \vdots \\ a_{n+1} \end{bmatrix} = \begin{bmatrix} \frac{1}{B_{max}} & \frac{1}{B_{max}} & \cdots & \cdots & \frac{1}{B_{max}} \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} a_{n+B_{max}-1} \\ a_{n+B_{max}-2} \\ a_{n+B_{max}-3} \\ \vdots \\ a_n \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{B_{max}} & \frac{1}{B_{max}} & \cdots & \cdots & \frac{1}{B_{max}} \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}^2 \begin{bmatrix} a_{n+B_{max}-2} \\ a_{n+B_{max}-3} \\ a_{n+B_{max}-4} \\ \vdots \\ a_{n-1} \end{bmatrix}$$

$$= \cdots = \begin{bmatrix} \frac{1}{B_{max}} & \frac{1}{B_{max}} & \cdots & \cdots & \frac{1}{B_{max}} \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}^n \begin{bmatrix} a_{B_{max}} \\ a_{B_{max}-1} \\ a_{B_{max}-2} \\ \vdots \\ a_1 \end{bmatrix},$$

then $a_{a+B_{max}}$ is the inner product of the first row of the matrix P^n and the vector $(a_{B_{max}}, a_{B_{max}-1}, \cdots, a_1)$. Let

$$P = \begin{bmatrix} \frac{1}{B_{max}} & \frac{1}{B_{max}} & \cdots & \cdots & \frac{1}{B_{max}} \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix},$$

the characteristic polynomial of matrix P is :

$$f(x) = x^{B_{max}} - \frac{1}{B_{max}} x^{B_{max}-1} - \frac{1}{B_{max}} x^{B_{max}-2} - \cdots - \frac{1}{B_{max}} x - \frac{1}{B_{max}}$$

$$= (x-1)\left(x^{B_{max}-1} + \frac{B_{max}-1}{B_{max}} x^{B_{max}-2} + \frac{B_{max}-2}{B_{max}} x^{B_{max}-3} + \cdots + \frac{2}{B_{max}} x + \frac{1}{B_{max}} \right),$$

so P has a eigenvalue 1, and the rest $B_{max} - 1$ eigenvalues are all roots of the equation 2.17, which are denoted as $\lambda_1, \ldots, \lambda_{B_{max}-1}$. It is easy to check the corresponding right and left eigenvectors for the eigenvalue 1 are $\vec{\upsilon}_0 = (1, 1, \ldots, 1)^t$ and $\vec{\nu}_0 = (B_{max}, B_{max} - 1, \ldots, 2, 1)$, respectively. The right and left eigenvectors corresponding to eigenvalue λ_i, $i = 1, 2, \ldots, B_{max} - 1$, are $\vec{\upsilon}_i$ and $\vec{\nu}_i$, respectively, where $\upsilon_{i,h} = \lambda_i^{B_{max}-h}$, and $\nu_{i,h} = B_{max}\lambda_i^h - \sum_{m=1}^{h-1} \lambda_i^m$,$h = 1, 2, \ldots, B_{max}$.

In order to calculate P^n, we need to construct a pair of biorthogonal families to diagonalize P. Note that if $\lambda_i \neq \lambda_j$, then $\lambda_i \vec{\nu_i}\vec{\upsilon_j} = \vec{\nu_i} P \vec{\upsilon_j} = \vec{\nu_i}\lambda_j\vec{\upsilon_j} = \lambda_j \vec{\nu_i}\vec{\upsilon_j}$, which implies that $\vec{\nu_i}\vec{\upsilon_j} = 0$. Let $\vec{\phi_i} = \vec{\upsilon_i}$ and $\vec{\varphi_i} = c_i\vec{\nu_i}$, where c_i is a constant such that $\vec{\varphi_i}\vec{\phi_i} = 1$ for $i = 0, 1, \ldots, B_{max} - 1$. We solve c_i for $i = 0$ and $i = 1, 2, \ldots, B_{max} - 1$ separately. It is easy to get $c_0 = \frac{2}{B_{max}(B_{max}+1)}$ providing $c_0\vec{\nu_0}\vec{\upsilon_0} = c_0(B_{max}, B_{max} - 1, \ldots, 2, 1)(1, 1, \ldots, 1)^t = c_0 \sum_{h=1}^{B_{max}} h = 1$. For $i = 1, 2, \ldots, B_{max} - 1$,

$$c_i\vec{\nu_i}\vec{\upsilon_i} = c_i \sum_{h=1}^{B_{max}} \nu_{i,h}\upsilon_{i,h} = c_i \sum_{h=1}^{B_{max}} \left[\lambda_i^{B_{max}-h}\left(B_{max}\lambda_i^h - \sum_{m=1}^{h-1}\lambda_i^m\right)\right]$$

$$= c_i \sum_{h=1}^{B_{max}} \left[B_{max}\lambda_i^{B_{max}} - \lambda_i^{B_{max}-h}\sum_{m=1}^{h-1}\lambda_i^m\right]$$

$$= c_i \sum_{h=1}^{B_{max}} \left[B_{max}\lambda_i^{B_{max}} - \frac{\lambda_i^{B_{max}-h+1} - \lambda_i^{B_{max}}}{1-\lambda_i}\right]$$

$$= c_i \left[B_{max}^2\lambda_i^{B_{max}} + \frac{B_{max}\lambda_i^{B_{max}}}{1-\lambda_i} - \sum_{h=1}^{B_{max}}\frac{\lambda_i^{B_{max}-h+1}}{1-\lambda_i}\right]$$

$$= c_i \left[B_{max}^2\lambda_i^{B_{max}} + \frac{B_{max}\lambda_i^{B_{max}}}{1-\lambda_i} - \frac{\lambda_i(1-\lambda_i^{B_{max}})}{(1-\lambda_i)^2}\right]$$

$$= c_i \left[B_{max}^2\lambda_i^{B_{max}} + \frac{B_{max}\lambda_i^{B_{max}}}{1-\lambda_i} - \frac{\lambda_i(1-\lambda_i)(\lambda_i^{B_{max}-1}+\lambda_i^{B_{max}-2}+\cdots+\lambda_i+1)}{(1-\lambda_i)^2}\right]$$

$$= c_i \left[B_{max}^2\lambda_i^{B_{max}} + \frac{B_{max}\lambda_i^{B_{max}}}{1-\lambda_i} - \frac{B_{max}\lambda_i^{B_{max}+1}}{1-\lambda_i}\right]$$

$$= c_i B_{max}(B_{max}+1)\lambda_i^{B_{max}} = 1,$$

Hence, $c_i = \frac{\lambda_i^{-B_{max}}}{B_{max}(B_{max}+1)}$, and

$$\varphi_{i,h} = c_i\nu_{i,h} = \frac{\lambda_i^{-B_{max}}}{B_{max}(B_{max}+1)}\left(B_{max}\lambda_i^h - \sum_{m=1}^{h-1}\lambda_i^m\right), i = 1,\ldots,B_{max}-1, h = 1,\ldots,B_{max}. \tag{2.18}$$

If $B = [\vec{v}_0, \vec{v}_1, \ldots, \vec{v}_{B_{max}-1}]$ and $C^t = [\vec{\varphi}_0, \vec{\varphi}_1, \ldots, \vec{\varphi}_{B_{max}-1}]$, then $CB = I_{B_{max}\times B_{max}}$, the identity matrix, and $CPB = D$, the diagonal matrix with the eigenvalues of P on the main diagonal. B, C, D are listed below.

$$B = \begin{bmatrix} 1 & \lambda_1^{B_{max}-1} & \cdots & \lambda_{B_{max}-1}^{B_{max}-1} \\ 1 & \lambda_1^{B_{max}-2} & \cdots & \lambda_{B_{max}-1}^{B_{max}-2} \\ \vdots & \ddots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

$$D = \begin{bmatrix} 1 & & & \\ & \lambda_1^n & & \\ & & \ddots & \\ & & & \lambda_{B_{max}-1}^n \end{bmatrix}$$

$C =$

$$\begin{bmatrix} \frac{2B_{max}}{B_{max}(B_{max}+1)} & \frac{2(B_{max}-1)}{B_{max}(B_{max}+1)} & \cdots & \frac{2}{B_{max}(B_{max}+1)} \\ \frac{B_{max}\lambda_1^{1-B_{max}}}{B_{max}(B_{max}+1)} & \frac{\lambda_1^{-B_{max}}}{B_{max}(B_{max}+1)}\left(B_{max}\lambda_1^2 - \lambda_1\right) & \cdots & \frac{\lambda_1^{-B_{max}}}{B_{max}(B_{max}+1)}\left(B_{max}\lambda_1^{B_{max}} - \sum_{m=1}^{B_{max}-1}\lambda_1^m\right) \\ \vdots & \ddots & \ddots & \vdots \\ \frac{B_{max}\lambda_{B_{max}-1}^{1-B_{max}}}{B_{max}(B_{max}+1)} & \frac{\lambda_{B_{max}-1}^{-B_{max}}}{B_{max}(B_{max}+1)}\left(B_{max}\lambda_{B_{max}-1}^2 - \lambda_{B_{max}-1}\right) & \cdots & \frac{\lambda_{B_{max}-1}^{-B_{max}}}{B_{max}(B_{max}+1)}\left(B_{max}\lambda_{B_{max}-1}^{B_{max}} - \sum_{m=1}^{B_{max}-1}\lambda_{B_{max}-1}^m\right) \end{bmatrix}$$

This diagonalization provides us $P^n = BD^nC$. Hence,

$$a_{n+B_{max}} = P^n_{1\cdot}(a_{B_{max}}, a_{B_{max}-1}, \cdots, a_1)^t$$

$$= \frac{1}{B_{max}(B_{max}+1)} \sum_{h=1}^{B_{max}} a_{B_{max}-h+1}\left\{2(B_{max}-h+1) + \sum_{i=1}^{B_{max}-1}\left[\lambda_i^{B_{max}-1+n}\lambda_i^{-B_{max}}\left(B_{max}\lambda_i^h - \sum_{m=1}^{h-1}\lambda_i^m\right)\right]\right\}$$

$$= \frac{1}{B_{max}(B_{max}+1)} \sum_{h=1}^{B_{max}} a_{B_{max}-h+1}\left\{2(B_{max}-h+1) + \sum_{i=1}^{B_{max}-1}\left(B_{max}\lambda_i^{n-1+h} - \sum_{m=1}^{h-1}\lambda_i^{n-1+m}\right)\right\}$$

$$= \frac{1}{B_{max}(B_{max}+1)} \sum_{h=1}^{B_{max}} a_{B_{max}-h+1}\left\{2(B_{max}-h+1) + \sum_{i=1}^{B_{max}-1}\lambda_i^{n-1+h-B_{max}}\left(B_{max}\lambda_i^{B_{max}} - \sum_{m=1}^{h-1}\lambda_i^{B_{max}-h+m}\right)\right\}$$

$$= \frac{1}{B_{max}(B_{max}+1)} \sum_{h=1}^{B_{max}} a_{B_{max}-h+1}\left\{2(B_{max}-h+1) + \sum_{i=1}^{B_{max}-1}\lambda_i^{n-1+h-B_{max}}\left(\sum_{m=0}^{B_{max}-1}\lambda_i^m - \sum_{m=B_{max}-h+1}^{B_{max}-1}\lambda_i^m\right)\right\}$$

$$= \frac{1}{B_{max}(B_{max}+1)} \sum_{h=1}^{B_{max}} a_{B_{max}-h+1} \left[2(B_{max}-h+1) + \sum_{i=1}^{B_{max}-1} \lambda_i^{n-1+h-B_{max}} \sum_{m=0}^{B_{max}-h} \lambda_i^m \right]$$

$$= \frac{1}{B_{max}(B_{max}+1)} \sum_{h=1}^{B_{max}} a_{B_{max}-h+1} \left[2(B_{max}-h+1) + \sum_{i=1}^{B_{max}-1} \sum_{m=0}^{B_{max}-h} \lambda_i^{n-1+h-B_{max}+m} \right]$$

$$= \frac{1}{B_{max}(B_{max}+1)} \sum_{h=1}^{B_{max}} a_h \left[2h + \sum_{i=1}^{B_{max}-1} \sum_{m=0}^{h-1} \lambda_i^{n-h+m} \right]$$

$$= \frac{1}{B_{max}(B_{max}+1)} \sum_{h=1}^{B_{max}} a_h \left[2h + \sum_{i=1}^{B_{max}-1} \sum_{m=1}^{h} \lambda_i^{n-m} \right].$$

This completes the proof. □

With the help of Lemma 1, Lemma 4 and Lemma 6, we can obtain the limiting properties of the distribution of the block size, the joint distribution of B_j and R_j, and the conditional distribution of R_j given B_j.

Lemma 7. *Let b be an integer from 1 to B_{max}. The limiting distribution of B_j is given by*

$$\lim_{j\to\infty} P(B_j = b) = \frac{2b}{B_{max}(B_{max}+1)}.$$

The limit of the joint distribution of B_j and R_j is given by

$$\lim_{j\to\infty} P(R_j = r, B_j = b) = \frac{2}{B_{max}(B_{max}+1)}.$$

The limit of the conditional distribution of R_j given B_j is given by

$$\lim_{j\to\infty} P(R_j = r|B_j) = \frac{1}{B_j}, \quad 1 \le r \le 2B_j.$$

Proof. We prove the limiting distribution of B_j first.

Since $\lambda_i < 1, i = 1, 2, \ldots, B_{max} - 1$, from Lemma 6, we have

$$\lim_{j\to\infty} P(B_j = b) = \lim_{j\to\infty} \frac{1}{B_{max}(B_{max}+1)} \sum_{h=1}^{B_{max}} \left(2h + \sum_{i=1}^{B_{max}-1} \sum_{m=1}^{h} \lambda_i^{\frac{j+1}{2}-B_{max}-m}\right) P(B_{2h-1} = b)$$

$$= \frac{2}{B_{max}(B_{max}+1)} \sum_{h=1}^{B_{max}} hP(B_{2h-1} = b)$$

$$= \frac{2}{B_{max}(B_{max}+1)} \left\{ \sum_{i=1}^{b} \frac{i(B_{max}+1)^{i-1}}{B_{max}^i} + \sum_{i=b+1}^{B_{max}} \frac{i(B_{max}+1)^{i-1}}{B_{max}^i} \left[1 - (\frac{B_{max}}{B_{max}+1})^b\right]\right\}$$

By $\sum_{i=1}^{n} ia^i = \frac{a - a^{n+1}(1+n-na)}{(1-a)^2}$ and $\sum_{i=1}^{n} a^i = \frac{a-a^{n+1}}{1-a}$,

$$\lim_{j\to\infty} P(B_j = b) = \frac{2}{B_{max}(B_{max}+1)} \left[\sum_{i=1}^{B_{max}} i\frac{(B_{max}+1)^{i-1}}{B_{max}^i} - \sum_{i=b+1}^{B_{max}} \frac{i(B_{max}+1)^{i-1}}{B_{max}^i} (\frac{B_{max}}{B_{max}+1})^b\right]$$

$$= \frac{2}{B_{max}(B_{max}+1)} \left[B_{max} - \sum_{i=1}^{B_{max}-b} \frac{(b+i)(B_{max}+1)^{i-1}}{B_{max}^i}\right]$$

$$= \frac{2}{B_{max}(B_{max}+1)} \left\{ B_{max} - \frac{1}{B_{max}} \left[\sum_{i=1}^{B_{max}-b} b(\frac{B_{max}+1}{B_{max}})^i + \sum_{i=1}^{B_{max}-b} i(\frac{B_{max}+1}{B_{max}})^i\right]\right\}$$

$$= \frac{2}{B_{max}(B_{max}+1)} \left[B_{max} - (B_{max} - b)\right]$$

$$= \frac{2b}{B_{max}(B_{max}+1)}.$$

By Lemma 5, the limit of the joint distribution of B_j and R_j can be derived by following.

$$\lim_{j\to\infty} P(R_j = r, B_j = b) = \lim_{j\to\infty} \left[P\left(B_j = \frac{r+1}{2}\right) - P\left(B_j = \frac{r-1}{2}\right)\right]$$

$$= \lim_{j\to\infty} P\left(B_j = \frac{r+1}{2}\right) - \lim_{j\to\infty} P\left(B_j = \frac{r-1}{2}\right)$$

$$= \frac{r+1}{B_{max}(B_{max}+1)} - \frac{r-1}{B_{max}(B_{max}+1)}$$

$$= \frac{2}{B_{max}(B_{max}+1)}.$$

The proof of the limit of the conditional distribution is trivial. □

Lemma 8, Lemma 9 and Lemma 10 give some distributional properties of the RBD $(B_{max};\, R)$ to assign patients within a block.

Lemma 8. *The conditional distribution of $N_A^S(j)$ given B_j and R_j of the RBD $(B_{max};\, R)$ is given by*

$$P(N_A^S(j) = t|R_j, B_j,) = \frac{\binom{R_j-1}{t}\binom{2B_j-R_j+1}{B_j-t}}{\binom{2B_j}{B_j}}, \quad t \in \begin{cases} [0, R_j-1], & \text{if } 1 \le R_j \le B_j, \\ [R_j-B_j-1, B_j], & \text{if } B_j+1 \le R_j \le 2B_j. \end{cases}$$

Proof. Think of an urn of $2B_j$ balls, B_j type A balls and B_j type B balls. We randomly select R_j-1 balls without replacement. The number of ways to draw R_j-1 balls from the urn is $\binom{2B_j}{R_j-1}$. The number of ways to select t type A balls from B_j type A balls in the urn is $\binom{B_j}{t}$. Similarly, the number of ways to select the remaining R_j-1-t type B balls from the B_j type B balls is $\binom{B_j}{R_j-1-t}$. Therefore we have a hypergeometric distribution with parameters $2B_j$, B_j and R_j-1. The probability mass function of $N_A^B(j-1)$ is given by

$$P(N_A^S(j) = t|R_j, B_j,) = \frac{\binom{B_j}{t}\binom{B_j}{R_j-1-t}}{\binom{2B_j}{R_j-1}}$$

$$= \frac{\binom{R_j-1}{t}\binom{2B_j-R_j+1}{B_j-t}}{\binom{2B_j}{B_j}}, t \in \begin{cases} [0, R_j-1], & \text{if } 1 \le R_j \le B_j, \\ [R_j-B_j-1, B_j], & \text{if } B_j+1 \le R_j \le 2B_j. \end{cases}$$ □

The following lemma provides the conditional distribution of the imbalance after the jth allocation and Lemma 8 displays the distribution of the imbalance.

Lemma 9. *Let d be a nonnegative integer, d, R_j and j have the same parity. The conditional distribution of D_j given R_j and B_j of the RBD $(B_{max}; R)$ is given by*

$$P(D_j = \pm d | R_j, B_j) = \frac{\binom{R_j}{\frac{d+R_j}{2}}\binom{2B_j - R_j}{B_j - \frac{d+R_j}{2}}}{\binom{2B_j}{B_j}}, \quad d \in [0, \min\{R_j, 2B_j - R_j\}].$$

Proof. The proof is trivial. This is another hypergeometric distribution with parameters $2B_j$, B_j and R_j. □

Lemma 10. *Let d be a nonnegative integer, r, d and j have the same parity. For $j \leq 2B_{max} + 2$, the distribution of D_j of the RBD $(B_{max}; R)$ is given by*

$$P(D_j = \pm d) = \sum_{b=1}^{B_{max}} \sum_{r=1}^{\min(2b,j)} \frac{\binom{r}{\frac{d+r}{2}}\binom{2b-r}{b-\frac{d+r}{2}}}{\binom{2b}{b}} \frac{1}{B_{max}} \left[\frac{1}{B_{max}+1}\left(\frac{1}{B_{max}}+1\right)^{\frac{j-r}{2}}\right]^{\mathbb{I}_{(r<j)}},$$

$$d \leq \min(r, 2b - r).$$

For $j \geq 2B_{max} + 3$, the distribution of D_j is given by

$$P(D_j = \pm d) \approx \sum_{b=1}^{B_{max}} \sum_{r=1}^{2b} \frac{\binom{r}{\frac{d+r}{2}}\binom{2b-r}{b-\frac{d+r}{2}}}{\binom{2b}{b}} \frac{2}{B_{max}(B_{max}+1)}, \quad d \leq \min(r, 2b - r).$$

Proof. This lemma can be proven by deriving the joint distribution of D_j, B_j and R_j using Lemma 9 and Lemma 1, then calculating the marginal distribution of D_j. □

Lemma 11 and Lemma 12 present the distribution of imbalance of the RBD $(B_{max}; T)$.

Lemma 11. *Let d be a nonnegative integer, r, d and j have the same parity. The conditional distribution of D_j given R_j and B_j of the RBD (B_{max}; T) is given by*

$$P(D_j = \pm d|R_j, B_j) = \begin{cases} \binom{R_j}{\frac{R_j \pm d}{2}} \frac{1}{2^{R_j}}, & d \in [0, R_j], 1 \le R_j \le B_j, \\ \binom{R_j}{\frac{R_j \pm d}{2}} \frac{1}{2^{R_j}} \sum_{x=1}^{2B_j - R_j} \binom{2B_j - x - 1}{B_j - 1} \frac{1}{2^{2B_j - x - 1}}, & d \in [0, 2B_j - R_j), B_j + 1 \le R_j \le 2B_j, \\ \sum_{x=2B_j - R_j + 1}^{B_j} \binom{2B_j - x - 1}{B_j - 1} \frac{1}{2^{2B_j - x}}, & d = 2B_j - R_j, B_j + 1 \le R_j \le 2B_j. \end{cases}$$

Proof. The distribution of D_j can be obtained by finding the distribution of the number of patients allocated to treatment A after the allocation of the jth patient, which is denoted by $N_A^*(j)$. Hence

$$P(D_j = \pm d|R_j, B_j) = P\left(N_A^*(j) = \frac{R_j \pm d}{2} \middle| R_j, B_j\right),$$

Given $1 \le R_j \le B_j$, it is easy to see that $N_A^*(j)$ follows a binomial distribution $(R_j, \frac{1}{2})$.

When $B_j + 1 \le R_j \le 2B_j$, then randomization sequence will be entirely deterministic after B_j patients have been assigned to either treatment A or B. Let X be the random number of the trials in the tail. Then $2B_j - X$ follows a negative binomial distribution. Therefore,

$$P(X = x) = \binom{2B_j - x - 1}{B_j - 1} \frac{1}{2^{2B_j - x - 1}}, x = 1, 2, \ldots, B_j.$$

If $X \le 2B_j - R_j$, $N_A^*(j)$ still follows a binomial distribution $Bin(n = 2B_j, p = 1/2)$ and $R_j - B_j < N_A^*(j) < B_j$, which is equivalent to $0 \le d < 2B_j - R_j$. Therefore

$$P\left(N_A^*(j) = \frac{R_j \pm d}{2}, X \le 2B_j - R_j | R_j, B_j, B_j + 1 \le R_j \le 2B_j\right)$$

$$= P\left(N_A^*(j) = \frac{R_j \pm d}{2}, X = x | R_j, B_j, B_j + 1 \leq R_j \leq 2B_j\right) \sum_{x=1}^{2B_j - R_j} P(X = x)$$

$$= \binom{R_j}{\frac{R_j \pm d}{2}} \frac{1}{2^{R_j}} \sum_{x=1}^{2B_j - R_j} \binom{2B_j - x - 1}{B_j - 1} \frac{1}{2^{2B_j - x - 1}}.$$

If $X > 2B_j - R_j$, which indicates R_j is in the tail and either $N_A^*(j) = B_j$ or $N_B^*(j) = B_j$ each with a probability of 1/2. In this case, $d = 2B_j - R_j$.

$$P\left(D_j = \pm(2B_j - R_j), X > 2B_j - R_j | R_j, B_j, B_j + 1 \leq R_j \leq 2B_j\right)$$

$$= P\left(D_j = \pm(2B_j - R_j), X = x | R_j, B_j, B_j + 1 \leq R_j \leq 2B_j\right) \sum_{x=2B_j - R_j + 1}^{B_j} P(X = x)$$

$$= \frac{1}{2} \sum_{x=2B_j - R_j + 1}^{B_j} \binom{2B_j - x - 1}{B_j - 1} \frac{1}{2^{2B_j - x - 1}}. \qquad \square$$

Lemma 12. *Let d be a nonnegative integer from 0 up to B_{max}. Let r, d and j have the same parity. For $j \leq 2B_{max} + 2$, the distribution of D_j of the RBD $(B_{max}; T)$ is given by*

$$P(D_j = \pm d) = \sum_{b=1}^{B_{max}} \sum_{r=max(d,1)}^{\min(b,j)} \binom{r}{\frac{r \pm d}{2}} \frac{1}{2^r} \frac{1}{B_{max}} \left[\frac{1}{B_{max} + 1} \left(\frac{B_{max} + 1}{B_{max}}\right)^{\frac{j-r}{2}}\right]^{\mathbb{I}_{(r<j)}}$$

$$+ \sum_{b=d+2}^{B_{max}} \sum_{r=b+1}^{\min(j, 2b-d-2)} \sum_{x=1}^{2b-r} \binom{r}{\frac{r \pm d}{2}} \binom{2b - x - 1}{b - 1} \frac{1}{2^{2b+r-x-1}} \frac{1}{B_{max}} \left[\frac{1}{B_{max} + 1} \left(\frac{B_{max} + 1}{B_{max}}\right)^{\frac{j-r}{2}}\right]^{\mathbb{I}_{(r<j)}}$$

$$+ \sum_{b=d+1}^{B_{max}} \sum_{x=d+1}^{b} \binom{2b - x - 1}{b - 1} \frac{1}{2^{2b-x}} \frac{1}{B_{max}} \left[\frac{1}{B_{max} + 1} \left(\frac{B_{max} + 1}{B_{max}}\right)^{\frac{j-2b+d}{2}}\right]^{\mathbb{I}_{(2b-d<j)}}$$

For $j \geq 2B_{max} + 3$, the distribution of D_j is given by

$$P(D_j = \pm d) \approx \sum_{b=1}^{B_{max}} \sum_{r=max(d,1)}^{b} \binom{r}{\frac{r \pm d}{2}} \frac{1}{2^r} \frac{2}{B_{max}(B_{max} + 1)}$$

$$+\sum_{b=d+2}^{B_{max}}\sum_{r=b+1}^{2b-d-2}\sum_{x=1}^{2b-r}\binom{r}{\frac{r\pm d}{2}}\binom{2b-x-1}{b-1}\frac{1}{2^{2b+r-x-1}}\frac{2}{B_{max}(B_{max}+1)}$$

$$+\sum_{b=d+1}^{B_{max}}\sum_{x=d+1}^{b}\binom{2b-x-1}{b-1}\frac{1}{2^{2b-x}}\frac{2}{B_{max}(B_{max}+1)}.$$

Proof. This lemma can be proven by deriving the joint distribution of D_j, B_j and R_j by

$$P(D_j = \pm d, R_j = r, B_j = b) = P(D_j = \pm d | R_j = r, B_j = b)P(R_j = r, B_j = b)$$

using Lemma 1 and Lemma 11. Then calculate the marginal distribution. □

2.2 Conclusion

In this section, we have formalized the RBD and obtained some important distributional properties of the RBD. With the help of these distributional results, we will be able to assess the predictability and balancing properties of the RBD, which will be given in the following two chapters.

One may note that we are unable to provide an exact form for the $\lambda_1, \ldots, \lambda_{B_{max}-1}$ in Lemma 6 when calculating the exact distribution of $P(B_j = b)$ for $j > 2B_{max}+2$. However, one can solve the equation (2.17) using statistical or mathematical software, or solve the exact distribution by the recursive equation in Lemma 5. In next two chapters, we are going to use the asymptotic result from Lemma 7, which simulation shows is quite accurate.

Chapter 3: Degree of Predictability of Restricted Randomizations

As we discussed in Chapter 1, the degree of randomness is a very important criteria for qualifying a randomization procedure. In this chapter, we study the degree of randomness of restricted randomization procedures. In Section 3.1, we propose a metric to measure the degree of randomness of restricted randomization procedures and call it the degree of predictability. We also provide exact formulas for calculating the degree of predictability of the RBD and PBD. Section 3.2 is the computation and comparison of the degree of predictability of six randomization procedures. Section 3.3 gives a short summary and discussion.

3.1 Degree of Predictability

Blackwell and Hodges (1957) developed a simple model to measure the potential for selection bias. They proposed a convergence strategy for guessing the upcoming assignment, which was to guess the treatment that has fewer prior allocations, or to guess one of the treatments if both treatments have an equal number of prior allocations. Their model calculates the expected selection bias factor, $E(F)$, which is the expected excess number of correct guesses of treatment assignments beyond that expected by chance when the investigator uses the convergence strategy. It is also equivalent to the difference between the expected number of correct and incorrect guesses among all the guesses made when the two treatment groups have different prior assignments.

Based on the fact that a larger degree of randomness corresponds to less predictability, we propose a metric for measuring the degree of randomness of a restricted randomization

procedure, which is the degree of predictability

$$\rho_{PRED} = \sum_{j=1}^{n} E\left|\phi_j - \frac{1}{2}\right|. \tag{3.1}$$

This is similar to the metric proposed by Chen (1999). This degree of predictability describes the expected deviation of a randomization procedure from complete randomization. For complete randomization, (3.1) is 0 since the treatment allocation probability is 1/2; the allocation has the largest degree of randomness and is unpredictable. When the allocation probability differs from 1/2, the allocation is restricted and predictable. A large value of (3.1) indicates a low degree of randomness, hence a high chance of correct prediction.

It turns out that the degree of predictability is mathematically equivalent to Blackwell-Hodges expected selection bias factor.

Theorem 1. *For all restricted randomization procedures that assign the next patient to any treatment using a probability based on previous assignments, the degree of predictability is mathematically equivalent to the Blackwell-Hodges selection bias factor with the convergence strategy for guessing the upcoming assignment*

$$\rho_{PRED} = E(F).$$

Proof. Call a correct guess a "hit" and an incorrect guess a "miss" when the two treatment arms have different prior allocations. Letting H and M denote the number of hits and misses, respectively, then the expected number of hits and misses, given $\phi_1, \ldots, \phi_n$ is calculated as

$$E(H|\phi_1, \cdots, \phi_n) = \sum_{j=1}^{n} \left[\frac{1}{2}\mathbb{I}\left(\phi_j = \frac{1}{2}\right) + \phi_j \mathbb{I}\left(\phi_j > \frac{1}{2}\right) + (1-\phi_j)\mathbb{I}\left(\phi_j < \frac{1}{2}\right)\right],$$

and

$$E(M|\phi_1,\cdots,\phi_n)=\sum_{j=1}^{n}\left[\frac{1}{2}\mathbb{I}\left(\phi_j=\frac{1}{2}\right)+(1-\phi_j)\mathbb{I}\left(\phi_j>\frac{1}{2}\right)+\phi_j\mathbb{I}\left(\phi_j<\frac{1}{2}\right)\right].$$

The expected selection bias factor is then given by

$$\begin{aligned}E(F)=&EE(F|\phi_1,\cdots,\phi_n)=E\left(\frac{E(H|\phi_1,\cdots,\phi_n)-E(M|\phi_1,\cdots,\phi_n)}{2}\right)\\ =&E\sum_{j=1}^{n}\left[\left(\phi_j-\frac{1}{2}\right)\mathbb{I}\left(\phi_j>\frac{1}{2}\right)+\left(\frac{1}{2}-\phi_j\right)\mathbb{I}\left(\phi_j<\frac{1}{2}\right)\right]\\ =&\sum_{j=1}^{n}E\left|\phi_j-\frac{1}{2}\right|\\ =&\rho_{PRED}.\end{aligned}$$

□

Hence the metrics are equivalent. Next, with the help of (3.1), we show the degree of predictability of the RBD(B_{max}; R), RBD(B_{max}; T), PBD(B_{max}; R) and PBD(B_{max}; T). Note that we treat summations as 0 if the upper limit of the summation is smaller than the lower limit.

Theorem 2. *Let r and j have the same parity, the ρ_{PRED} in n, $n\geq 1$, trials for the RBD (B_{max}; R) is given by*

$$\begin{aligned}\rho_{PRED}\approx&\sum_{j=1}^{\min(n,2B_{max}+2)}\sum_{b=1}^{B_{max}}\left[\sum_{r=1}^{\min(b,j)}\sum_{t=0}^{r-1}\left|\frac{b-t}{2b-r+1}-\frac{1}{2}\right|\frac{\binom{r-1}{t}\binom{2b-r+1}{b-t}}{\binom{2b}{b}}\right.\\ &\left.+\sum_{r=b+1}^{\min(2b,j)}\sum_{t=r-b-1}^{b}\left|\frac{b-t}{2b-r+1}-\frac{1}{2}\right|\frac{\binom{r-1}{t}\binom{2b-r+1}{b-t}}{\binom{2b}{b}}\right]\frac{1}{B_{max}}\left(\frac{1}{B_{max}+1}\left(1+\frac{1}{B_{max}}\right)^{\frac{j-r}{2}}\right)^{\mathbb{I}(r<j)}\end{aligned}$$

$$+\sum_{j=2B_{max}+3}^{n}\sum_{b=1}^{B_{max}}\left[\sum_{r=1}^{b}\sum_{t=0}^{r-1}\left|\frac{b-t}{2b-r+1}-\frac{1}{2}\right|\frac{\binom{r-1}{t}\binom{2b-r+1}{b-t}}{\binom{2b}{b}}\right.$$

$$\left.+\sum_{r=b+1}^{2b}\sum_{t=r-b-1}^{b}\left|\frac{b-t}{2b-r+1}-\frac{1}{2}\right|\frac{\binom{r-1}{t}\binom{2b-r+1}{b-t}}{\binom{2b}{b}}\right]\frac{2}{B_{max}(B_{max}+1)}.$$

Proof. It is straightforward to verify Theorem 2 using equation (3.1), Lemma 1 and Lemma 8 by

$$\sum_{j=1}^{n}E\left|\phi_j-\frac{1}{2}\right|=\sum_{j=1}^{n}\sum_{over\ t,b,r}\sum\sum\left|\frac{b-t}{2b-r+1}-\frac{1}{2}\right|P(N_A^B(j)=t|B_j=b,R_j=r)P(B_j=b,R_j=r).$$

□

Theorem 3. *Let r and j have the same parity, ρ_{PRED}, $n\geq 1$ trials for the RBD(B_{max}; T) is given by*

$$\rho_{PRED}\approx\sum_{j=1}^{\min(n,2B_{max}+2)}\sum_{b=1}^{B_{max}}\sum_{r=b+1}^{\min(j,2b)}\sum_{x=2b-r+1}^{b}\binom{2b-x-1}{b-1}\frac{1}{2^{2b-x}}\frac{1}{B_{max}}\left(\frac{1}{B_{max}+1}\left(\frac{1}{B_{max}}+1\right)^{\frac{j-r}{2}}\right)^{\mathbb{I}(r<j)}$$

$$+\sum_{j=2B_{max}+3}^{n}\sum_{b=1}^{B_{max}}\sum_{r=b+1}^{2b}\sum_{x=2b-r+1}^{b}\binom{2b-x-1}{b-1}\frac{1}{2^{2b-x}}\frac{2}{B_{max}(B_{max}+1)}. \tag{3.2}$$

Proof.

$$\sum_{j=1}^{n}E\left|\phi_j-\frac{1}{2}\right|=\sum_{j=1}^{n}\left[\left|\frac{1}{2}-\frac{1}{2}\right|P\left(\phi_j=\frac{1}{2}\right)+\left|1-\frac{1}{2}\right|P(\phi_j=1)+\left|0-\frac{1}{2}\right|P(\phi_j=0)\right]$$

$$=\frac{1}{2}\sum_{j=1}^{n}\left[P(\phi_j=0)+P(\phi_j=1)\right]$$

$$=\frac{1}{2}\sum_{j=1}^{n}\left[P(R_j\text{ is in the tail}|R_j,B_j)P(B_j+1\leq R_j\leq 2B_j)\right]$$

$$= \frac{1}{2}\sum_{j=1}^{n}[P(2B_j - R_j + 1 \le X \le B_j)P(B_j + 1 \le R_j \le 2B_j)]$$

$$= \sum_{j=1}^{n}\sum_{b=1}^{B_{max}}\sum_{r=b+1}^{2b}\sum_{x=2B_j-R_j+1}^{b}\binom{2b-x-1}{b-1}\frac{1}{2^{2b-x}}P(B_j = b, R_j = r)$$

The proof is completed by plugging in Lemma 1 in Chapter 2. □

Theorem 4. *With $n \ge 1$ trials for the PBD(B; R), ρ_{PRED} is given by*

$$\rho_{PRED} = \left\lfloor \frac{n}{2B} \right\rfloor \left[\frac{2^{2B-1}}{\binom{2B}{B}} - \frac{1}{2} \right]$$

$$+ \sum_{r=1}^{m} \left[\mathbb{I}_{(r \le B)} \sum_{t=0}^{r-1} \left| \frac{B-t}{2B-r+1} - \frac{1}{2} \right| \frac{\binom{r-1}{t}\binom{2B-r+1}{B-t}}{\binom{2B}{B}} \right.$$

$$\left. + \mathbb{I}_{(r \ge B+1)} \sum_{t=r-B-1}^{B} \left| \frac{B-t}{2B-r+1} - \frac{1}{2} \right| \frac{\binom{r-1}{t}\binom{2B-r+1}{B-t}}{\binom{2B}{B}} \right],$$

where $\lfloor a \rfloor$ denotes the integer part of a and $m = n - 2B\left\lfloor \frac{n}{2B} \right\rfloor$.

Proof. Note that $\lfloor a \rfloor$ denotes the number of blocks that are fully filled. We can apply Matts and Lachin's result for the PBD(B; R) since $\rho_{PRED} = E(F)$. For the last block, if it is not fully filled, we can apply Lemma 8 in Chapter 2 with non-random B_j and R_j. □

Theorem 5. *With $n \ge 1$ trials for the PBD(B; T), ρ_{PRED} is given by*

$$\rho_{PRED} = \left\lfloor \frac{n}{2B} \right\rfloor \left[\frac{2B}{2^{2B+1}} \binom{2B}{B} \right] + \sum_{r=B+1}^{m} \sum_{x=2B-r+1}^{B} \binom{2B-x-1}{B-1} \frac{1}{2^{2B-x}}.$$

where $\lfloor a \rfloor$ denotes the integer part of a and $m = n - 2B\left\lfloor \frac{n}{2B} \right\rfloor$.

Proof. For the blocks that all fully filled, we apply the results of equation (6.8) in Rosenberger and Lachin (2002). For the last block, if it is incomplete, the proof is similar to the proof of Theorem 3, but with non-random B_j and R_j. □

3.2 Computation of the Degree of Predictability

We compare ρ_{PRED} for six restricted randomization procedures, including RBD(B_{max}; R), RBD(B_{max}; T), PBD(B; R), PBD(B; T), BCD(p) and BSD(a), as a function of both n and their parameters B_{max}, B, p and a. Because we have proved the equivalence of ρ_{PRED} and $E(F)$ in Theorem 1, we use equation (1.15) to calculate the degree of predictability for the BCD(p), and equation (1.17) multiplying by n for the BSD(a).

Figure 3.1 presents an overall view of ρ_{PRED} for the six procedures. All the graphs have been generated for $n = 1, 2, \ldots, 100$, $B_{max} = 1, 2, \ldots, 20$ for the RBD(B_{max}, R) and RBD(B_{max}, T), $B = 1, 2, \ldots, 20$ for the PBD(B, R) and PBD(B, T), $p = 0.5, 0.525, \ldots, 1$ for the BCD and $a = 1, 2, \ldots, 20$ for the BSD. Color diverging from red to blue indicates the increase of ρ_{PRED}.

Figure 3.2 shows the behavior of ρ_{PRED} for the six restricted randomization procedures with some examples. All the graphs are generated for $n = 1, 2, \ldots, 50$. For the RBD, PBD and BSD, we choose $B_{max}/B/a = 3, 4, 6$ and 10. For the BCD, we choose $p = 3/5, 2/3, 3/4$ and $5/6$.

All the procedures RBD(1, R), RBD(1, T), PBD(1, R), PBD(1, T), BCD(1) and BSD(1) generate the same deterministic sequence $ABABAB\cdots$ or $BABABA\cdots$. Therefore these designs have the same ρ_{PRED} of $n/4$, as shown in Figure 3.1. For fixed $B_{max}/p/a$, ρ_{PRED} of the RBD(B_{max}, R), RBD(B_{max}, T), BCD(p) and BSD(a) are strictly monotone increasing functions of n. For a fixed B, ρ_{PRED} of the PBD(B) increases periodically with a period of half of the block size B.

The degree of predictably of the RBD and PBD for different values of n, B_{max} and B is provided in Tables 3.1 and 3.2. As expected, ρ_{PRED} is a monotone decreasing function

in B_{max} for the RBD. However, this does not hold for the PBD for some small size trials. Observe that ρ_{PRED} is monotone increasing in p for the BCD, with the smallest $\rho_{PRED} = 0$ when $p = 1/2$ and largest $\rho_{PRED} = n/4$ when $p = 1$. For the BSD, as a increases, ρ_{PRED} decreases, and ρ_{PRED} for $a > 1$ is greatly less than that when $a = 1$.

As can be seen in the graphs and tables, for the same value of B_{max}/B, the block design using the TBD to allocate patient to treatment group within blocks generates less predictable treatment assignment sequences than the design using the RAR within blocks. This agrees with Blackwell and Hodges (1957). It also shows that when $B_{max} = B$, ρ_{PRED} of the RBD(B_{max}; R) is higher than the PBD(B; R). We see that the selection bias factor of the RBD(3; R) is approximately the same as that of the PBD(2; R), and the selection bias factor of the RBD(5; R) is approximately the same as that of the PBD(3; R). In order to generate a randomization sequence with a comparable degree of predictability of PBD(B; R), a larger value of B_{max} has to be selected for the RBD(B_{max}; R). It is also true between the RBD(B_{max}; T) and PBD(B; T). However, the RBD(B_{max}; T) and PBD(B, R) have almost the same degree of predictability when $B_{max} = B$.

The BSD has the smallest ρ_{PRED} comparing with the other five restricted randomization procedures since it has the largest red area in Figure 3.1.

3.3 Conclusion

In this chapter, we investigated the degree of randomness of restricted randomization procedures. We provide a metric equivalent to the Blackwell-Hodges selection bias factor, which is the degree of predictability ρ_{PRED}, to find the exact selection bias of restricted randomization procedures. The Blackwell-Hodges model works well for the case that the sample size n is known in advance, but it would be difficult for the case n is unknown, like a permuted block design with an incomplete last block or a random block design with unknown block sizes. Therefore, in light of this new approach, for the first time we find the selection bias of the RBD when both the trial size and block size of all blocks are unknown,

and the exact selection bias of the PBD when the trial size n is unknown.

It is generally assumed that the RBD will be less predictable. However, in this chapter, we find that the RBD does not reduce predictability as one might expect compared to the PBD. This agrees with the simulation results of Zhao, et al. (2011). One has to choose a relatively larger value of B_{max} for the RBD(B_{max}) to produce a comparable or less predictable allocation sequence than that under the PBD(B).

Matts and Lachin (1988) concluded that the selection bias factor of a design employing multiple block sizes is approximately the same as that of a design with a block size equal to the average of all the block sizes of the former design. Our results confirm their conclusion and show that the degree of predictability of RBD(3; R) is slightly smaller than that of the PBD(2; R). The RBD does not reduce the degree of predictability greatly, and there is an explicit explanation in Berger's book (2005).

When $a = B_{max} = B$, the BSD(a) has a significantly smaller degree of predictability compared with the RBD(B_{max}; R), RBD(B_{max}; T), PBD(B; R) and PBD(B; T).

Table 3.1: Degree of predictability of the RBD(B_{max}; R) and PBD(B; R) for different values of n, B_{max} and B

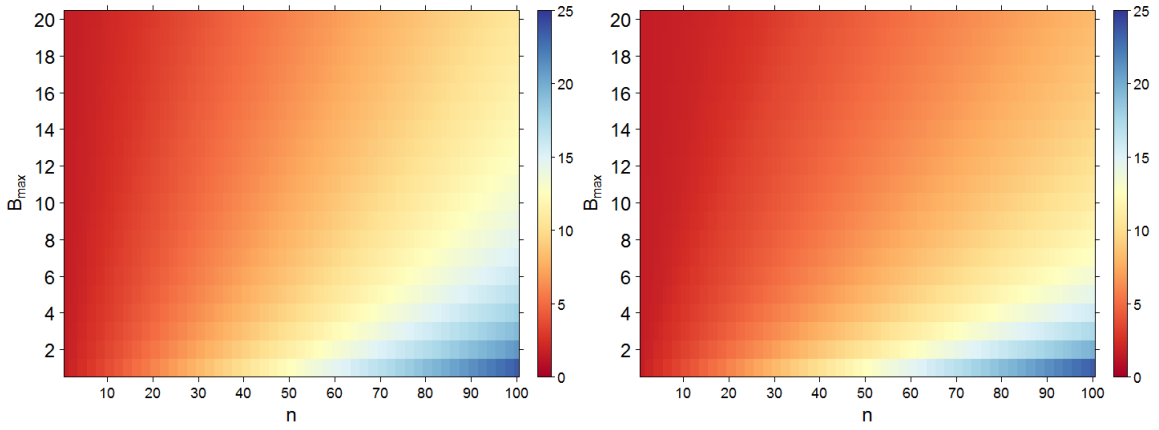
.

	RBD(B_{max}; R)						PBD(B; R)					
n	3	4	5	6	7	8	2	3	4	5	6	7
5	0.76	0.66	0.58	0.52	0.47	0.43	0.83	0.60	0.40	0.30	0.24	0.20
10	1.90	1.71	1.60	1.45	1.32	1.21	1.83	1.50	1.40	1.53	0.99	0.77
15	2.78	2.55	2.36	2.21	2.09	1.97	2.83	2.40	2.16	1.83	1.81	1.89
20	3.92	3.59	3.33	3.11	2.93	2.77	4.17	3.40	2.93	3.06	2.33	2.18
25	4.81	4.43	4.13	3.88	3.66	3.47	5.00	4.40	3.99	3.37	3.43	2.81
50	10.01	9.24	8.62	8.12	7.69	7.33	10.17	8.90	8.04	7.66	6.91	6.16
75	14.95	13.84	12.95	12.22	11.60	11.07	15.33	13.40	12.10	11.02	10.39	9.64
100	20.15	18.64	17.45	16.46	15.64	14.93	20.83	18.00	16.21	15.32	13.90	13.25

Table 3.2: Degree of predictability of the RBD(B_{max}; T) and PBD(B; T) for different values of n, B_{max} and B

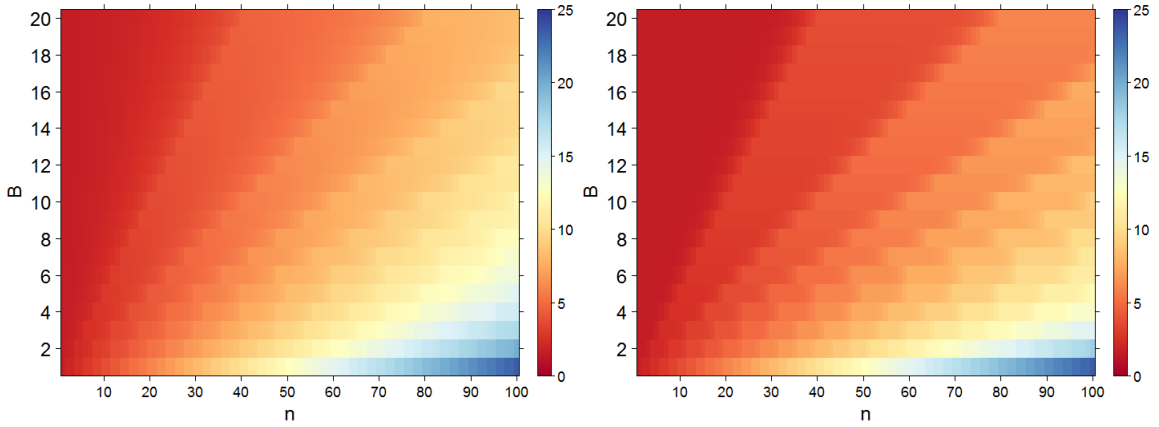
.

	RBD(B_{max}; T)						PBD(B; T)					
n	3	4	5	6	7	8	2	3	4	5	6	7
5	0.65	0.48	0.38	0.31	0.27	0.23	0.75	0.44	0.06	0.00	0.00	0.00
10	1.64	1.41	1.28	1.09	0.92	0.78	1.50	1.06	1.09	1.23	0.48	0.13
15	2.46	2.16	1.93	1.75	1.61	1.48	2.50	1.88	1.69	1.23	1.35	1.47
20	3.46	3.05	2.74	2.49	2.28	2.11	3.75	2.81	2.19	2.46	1.43	1.47
25	4.29	3.80	3.44	3.14	2.90	2.69	4.50	3.75	3.28	2.46	2.71	1.77
50	8.93	7.97	7.25	6.68	6.21	5.82	9.00	7.50	6.56	6.15	5.41	4.41
75	13.40	12.01	10.96	10.13	9.45	8.88	13.75	11.25	9.84	8.61	8.12	7.33
100	18.05	16.18	14.77	13.66	12.76	12.00	18.75	15.12	13.12	12.30	10.83	10.26



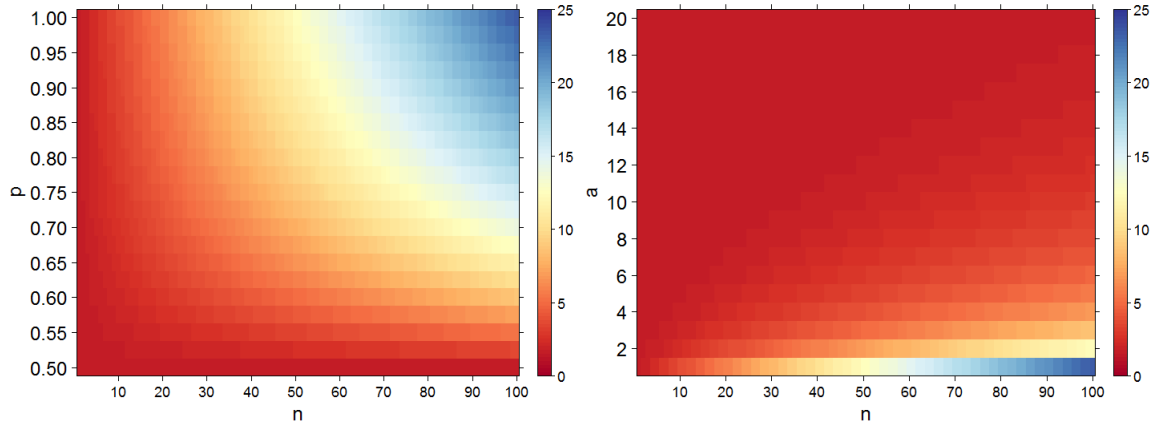
(a) RBD(B_{max}; R)

(b) RBD(B_{max}; T)



(c) PBD(B; R)

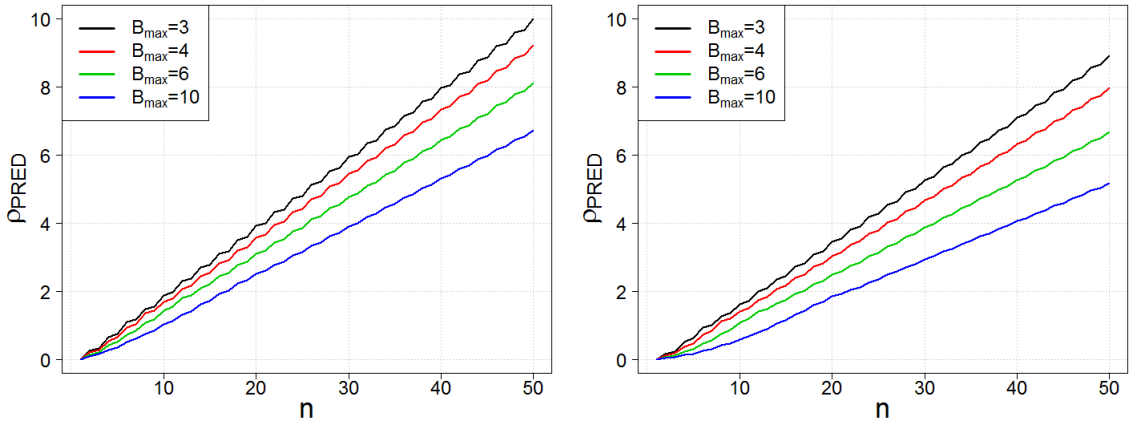
(d) PBD(B; T)



(e) BCD(p)

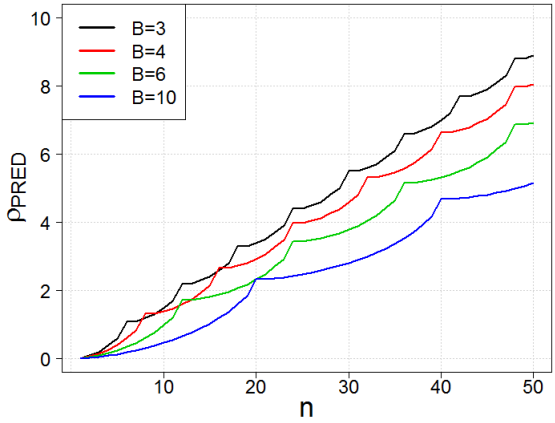
(f) BSD(a)

Figure 3.1: Degree of predictability of the six restricted randomization procedures. Note that each panel has different color scales.

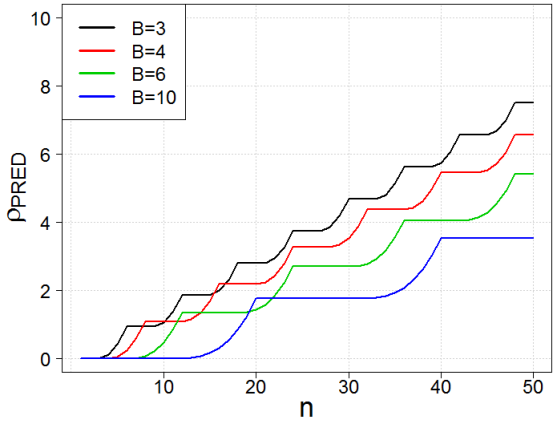


(a) RBD(B_{max}; R)

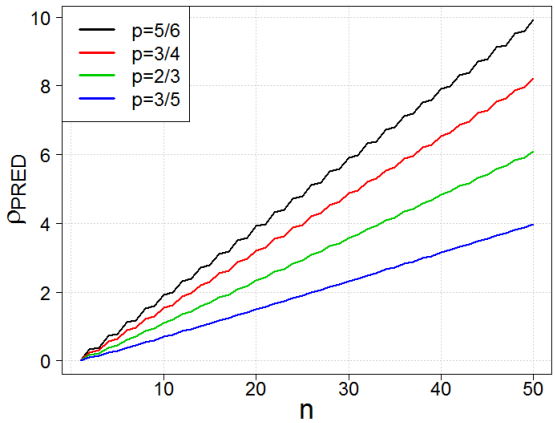
(b) RBD(B_{max}; T)



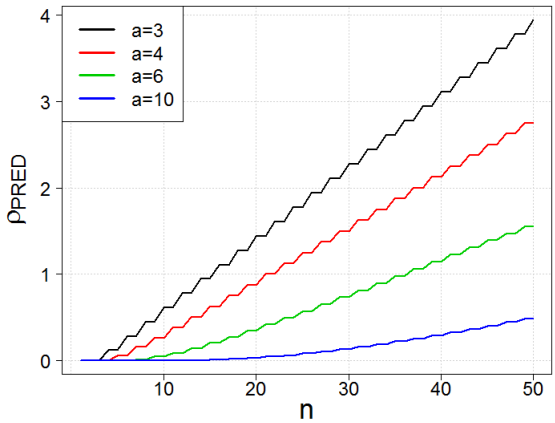
(c) PBD(B; R)



(d) PBD(B; T)



(e) BCD(p)



(f) BSD(a)

Figure 3.2: Degree of predictability of the six restricted randomization procedures. Note that each panel has different y-axis scales.

Chapter 4: Variability of the Imbalance of Restricted Randomization Procedures

In this chapter, we study the balancing properties of restricted randomization procedures. Large imbalances in treatment assignments may result in large or significant differences of baseline characteristics between treatment groups, hence lessen the creditability of the trial results.

In complete randomization, patients are assigned to treatment groups without any constraint; however, there is a large probability of imbalanced treatment groups, which is a special concern for a small clinical trials. Rosenberger and Lachin (2002) provide the asymptotic distribution of $|D_n|$ for complete randomization based on the normal approximation of binomial distribution. For $r > 0$, $P(|D_n| \leq r) \approx 2\Phi(r/\sqrt{n}) - 2$. Table 4.1 shows the probabilities of $|D_n|$ falling in three intervals for different values of n. When $n = 50$, the chance of an imbalance of more than 20% of 50 is 0.16. And for $n = 100$, that probability is 0.05, corresponding to an imbalance of 60/40 in favor of treatment A (or B). When n is large, the probability of large imbalances converges to 0.

For Efron's biased coin design, with $p = 2/3$, the asymptotic probability of perfectly balanced treatment groups is 0.5 for even n, and 0.75 for odd n. By equation (1.13), $P(0 \leq |D_n| \leq 0.10n) = 0.97$ for $n = 50$ and almost 1 for $n = 100$. Compared with complete randomization, the BCD(p) has a much smaller probability of producing extreme imbalances. However, Soares and Wu (1983) believe that no experimenter would like to take the risk to incur a serve imbalance in a trial, especially when the trial cannot be repeated. In their BSD (a), a maximum imbalance bound a is prescribed to constrain the maximum imbalance throughout the trial.

Block randomization controls the imbalance by randomizing the treatment assignment

block by block. It ensures periodical treatment balances, and that the largest imbalance of the whole course of the trial is not greater than half of the largest block size. If the number of patients in the trial is known in advance, the PBD(B) produces balanced treatment group by choosing B appropriately.

We have summarized the metrics for balance given in the literature in Section 1.3. Soares and Wu (1983) argue that the expected final imbalance $E(D_n)$ is a misleading way of summarizing the final imbalance distribution, because it is more desirable to have an imbalance distribution that does not produce extreme final imbalances. The simulation results of Rosenberger and Lachin (2002) are consistent with Soares and Wu's opinion. When $n = 50$, $E(D_n)$ is close to 0 for all the designs; for example, $E(D_n) = -0.017$ for complete randomization and 0.002 for the BCD (2/3). However, $Var(D_n)$ for complete randomization is 49.92, while $Var(D_n)$ for the other designs are less than 5 except the urn design, which is 16.58. This example indicates that a small variability of the imbalance is more important than controlling its expectation around 0. More recently, researchers have proposed that measuring large imbalances throughout the trial is as important as terminal imbalance, particularly for the trials with time-varying baseline covariates. One measure of interest is $E(\max_{1\leq j\leq n}|D_j|)$. According to Rosenberger and Lachin's simulation results, when $n = 50$, the average maximum imbalance of complete randomization is 8.88; 4.28 for the BCD(2/3); and 2.34 for the RBD(3, R); and 3.75 for the RBD(10, R).

In Section 4.1, we provide exact formulas of the variance of the terminal imbalance, $Var(D_n)$, for the RBD(B_{max}), PBD(B), and BSD(a). In Section 4.2, we investigate and compare this property for six restricted randomization procedures.

4.1 Variance of Terminal Imbalance

Theorem 8 and Theorem 9 provide $Var(|D_n|)$ of the RBD(B_{max}; R) and RBD(B_{max}; T), respectively. For the RBD with block sizes ranging from 2 to $2B_{max}$, the final imbalance D_n varies from $-B_{max}$ to B_{max}. A small variance of D_n indicates that repeating the RBD

produces similar values of imbalance.

Theorem 6. *Let $n \geq 1$. Let r, d and n have the same parity. Then the variance of imbalance of the RBD(B_{max}; R) is given by following.*
If $n \leq 2B_{max} + 2$,

$$Var(D_n) = 2 \sum_{b=1}^{B_{max}} \sum_{r=1}^{\min(2b,n)} \sum_{d=0}^{\min(r,2b-r)} d^2 \frac{\binom{r}{\frac{d+r}{2}}\binom{2b-r}{b-\frac{d+r}{2}}}{\binom{2b}{b}} \frac{1}{B_{max}} \left[\frac{1}{B_{max}+1} \left(1 + \frac{1}{B_{max}}\right)^{\frac{n-r}{2}} \right]^{\mathbb{I}_{(r<n)}}.$$

If $n \geq 2B_{max} + 3$,

$$Var(D_n) \approx 2 \sum_{b=1}^{B_{max}} \sum_{r=1}^{2b} \sum_{d=0}^{\min(r,2b-r)} d^2 \frac{\binom{r}{\frac{d+r}{2}}\binom{2b-r}{b-\frac{d+r}{2}}}{\binom{2b}{b}} \frac{2}{B_{max}(B_{max}+1)}.$$

Proof. The expectation of the terminal imblance if 0, $E(D_n) = 0$ because $E(D_n) = \sum_d (dP(D_n = d) + (-d)P(D_n = -d)) = 0$. Therefore, $Var(D_n) = E(D_n^2) = 2\sum_d d^2 P(D_n = d)$. Substituting Lemma 10 for $P(D_n = d)$ in Chapter 2, we obtain the variance of the terminal imbalance. □

Theorem 7. *Let $n \geq 1$. Let r, d and n have the same parity. Then the variance of imbalance of the RBD(B_{max}; T) is given by following. For $n \leq 2B_{max} + 2$,*

$$\begin{aligned}
Var(D_n) = &\sum_{b=1}^{B_{max}} \sum_{r=1}^{\min(b,n)} \sum_{d=0}^{r} d^2 \binom{r}{\frac{r+d}{2}} \frac{1}{2^{r-1}} \frac{1}{B_{max}} \left[\frac{1}{B_{max}+1} \left(1 + \frac{1}{B_{max}}\right)^{\frac{n-r}{2}} \right]^{\mathbb{I}_{(r<n)}} \\
&+ \sum_{b=1}^{B_{max}} \sum_{r=b+1}^{\min(2b,n)} \sum_{d=0}^{2b-r-1} \sum_{x=1}^{2b-r} d^2 \binom{r}{\frac{r+d}{2}} \binom{2b-x-1}{b-1} \frac{1}{2^{2b+r-x-2}} \frac{1}{B_{max}} \left[\frac{1}{B_{max}+1} \left(1 + \frac{1}{B_{max}}\right)^{\frac{n-r}{2}} \right]^{\mathbb{I}_{(r<n)}} \\
&+ \sum_{b=1}^{B_{max}} \sum_{r=b+1}^{\min(2b,n)} \sum_{x=2b-r+1}^{b} (2b-r)^2 \binom{2b-x-1}{b-1} \frac{1}{2^{2b-x-1}} \frac{1}{B_{max}} \left[\frac{1}{B_{max}+1} \left(1 + \frac{1}{B_{max}}\right)^{\frac{n-r}{2}} \right]^{\mathbb{I}_{(r<n)}}.
\end{aligned}$$

For $j \geq 2B_{max} + 3$,

$$Var(D_n) \approx 2 \sum_{b=1}^{B_{max}} \sum_{r=1}^{b} \sum_{d=0}^{r} d^2 \binom{r}{\frac{r+d}{2}} \frac{1}{2^r} \frac{2}{B_{max}(B_{max}+1)}$$

$$+ 2 \sum_{b=1}^{B_{max}} \sum_{r=b+1}^{2b} \sum_{d=0}^{2b-r-1} \sum_{x=1}^{2b-r} d^2 \binom{r}{\frac{r+d}{2}} \binom{2b-x-1}{b-1} \frac{1}{2^{2b+r-x-1}} \frac{2}{B_{max}(B_{max}+1)}$$

$$+ 2 \sum_{b=1}^{B_{max}} \sum_{r=b+1}^{2b} \sum_{x=2b-r+1}^{b} (2b-r)^2 \binom{2b-x-1}{b-1} \frac{1}{2^{2b-x}} \frac{2}{B_{max}(B_{max}+1)}.$$

Proof. The proof is same as the proof of Theorem 6 but substitute the Lemma 12 for $P(D_n = d)$. □

Theorem 8. *Let* $n \geq 1$. *Let* r, d *and* n *have the same parity. Then the variance of imbalance of the PBD(B; R) is given by*

$$Var(D_n) = 2 \sum_{d=0}^{\min(m,2B-m)} d^2 \frac{\binom{m}{\frac{m+d}{2}} \binom{2B-m}{B-\frac{m+d}{2}}}{\binom{2B}{B}},$$

where $\lfloor a \rfloor$ *denotes the integer part of* a *and* $m = n - 2B \left\lfloor \frac{n}{2B} \right\rfloor$.

Proof. From Lemma 9 in Chapter 1, the distribution of the terminal imbalance of the PBD (B; R) is as following by replacing R_j and B with m and B, respectively.

$$P(D_n = \pm d) = \frac{\binom{m}{\frac{m \pm d}{2}} \binom{2B-m}{B-\frac{m \pm d}{2}}}{\binom{2B}{B}}, d \in [0, \min(m, 2B-m)].$$

Hence the variance of D_n is calculated by $2 \sum_d d^2 P(D_n = d)$. □

Theorem 9. *Let $n \geq 1$. Let r, d and n have the same parity. Then the variance of imbalance of the PBD($B; T$) is given by following.*

$$Var(D_n) = \begin{cases} \sum_{d=0}^{m} d^2 \binom{m}{\frac{m+d}{2}} \frac{1}{2^{m-1}}, & 1 \leq m \leq B, \\ \sum_{d=0}^{2B-m-1} d^2 \binom{m}{\frac{m+d}{2}} \sum_{x=1}^{2B-m} \binom{2B-x-1}{B-1} \frac{1}{2^{2B+m-x-2}} & \\ \quad +(2B-m)^2 \sum_{x=2B-m+1}^{B} \binom{2B-x-1}{B-1} \frac{1}{2^{2B-x-1}}, & B+1 \leq m \leq 2B. \end{cases}$$

where $\lfloor a \rfloor$ denotes the integer part of a and $m = n - 2B \left\lfloor \frac{n}{2B} \right\rfloor$.

Proof. The distribution of D_n of the PBD(B; T) is as following by replacing R_j and B with m and B in Lemma 11, respectively.

$$P(D_n = \pm d) = \begin{cases} \binom{m}{\frac{m \pm d}{2}} \frac{1}{2^m}, & d \in [0, m], 1 \leq m \leq B, \\ \binom{m}{\frac{m \pm d}{2}} \sum_{x=1}^{2B-m} \binom{2B-x-1}{B-1} \frac{1}{2^{2B+m-x-1}}, & d \in [0, 2B-m), B+1 \leq m \leq 2B, \\ \sum_{x=2B-m+1}^{B} \binom{2B-x-1}{B-1} \frac{1}{2^{2B-x}}, & d = 2B-m, B+1 \leq m \leq 2B. \end{cases}$$

□

Chen investigated the balancing property of the BSD(a) by its asymptotic average imbalance. We have not seen the general formula of the variance of the final imbalance of the BSD(a). We provide it here.

Theorem 10. *Let $n \geq 1$. Let d and n have the same parity. For the BSD (a), when $a = 1$, $Var(D_n) = 1$ if n is odd, and $Var(D_n) = 0$ if n is even. When $a \geq 2$,*

$$Var(D_n) = \sum_{d=1}^{a-1} d^2 \left[\frac{2}{a} + \frac{2(-1)^{n+d}}{a} + \frac{4}{a} \sum_{l=1}^{a-1} \left(cos \frac{l\pi}{a} \right)^n cos \left(\frac{dl\pi}{a} \right) \right]$$

$$+ a + (-1)^{n+a}a + 2a\sum_{l=1}^{a-1}\left(cos\frac{l\pi}{a}\right)^n cosl\pi.$$

Proof. Formula (1.15) provides the high-order transition probabilities $P_{0,d}^{(n)}$ of D_n.

$$P(D_n = \pm d) = \eta_d + (-1)^{n+d}\eta_d + 2\eta_d\sum_{l=1}^{a-1}\left(cos\frac{l\pi}{a}\right)^n cos\left(\frac{dl\pi}{a}\right),$$

where $0 \leq d \leq \min(a, n)$, $\eta_0 = 1/(2a)$, $\eta_1 = \cdots = \eta_{a-1} = 1/a$, and $\eta_a = 1/(2a)$. Then $Var(D_n)$ can be obtained by calculating $E(D_n^2)$. Notice that when n and d do not have the same parity, $P(D_n = \pm d) = 0$. □

4.2 Computation and Comparison of the Terminal Imbalance

We compare $Var(D_n)$ for the six restricted randomization procedures, over a range of n and the parameter of each procedure. For the BCD(p), we use the result from Markaryan and Rosenberger and is given in equation (1.14). Figure 4.1 displays an overall view of $Var(D_n)$ for the six randomization procedures. Colors diverge from red to blue, indicating the increase of $Var(D_n)$. Note that the range of $Var(D_n)$ for each plot varies. While the RBD(B_{max}, R) has the smallest range from 0 to 5, the BCD(p) has the largest one from 0 to 100. We also plot the $Var(D_n)$ as a function of n for various values of the parameter of each design in Figure 4.2. In addition, $Var(D_n)$ of the RBD(B_{max}) and PBD(B) for different values of n, B_{max}, and B is provided in Table 4.2 and Table 4.3 for using the RAR and TBD within blocks, respectively.

From Figure 4.1 and Figure 4.2, we observe that the RBD(B_{max}, R) and RBD(B_{max}, T) have a similar pattern: $Var(D_n)$ is a monotone increasing function in B_{max} for each n, the curves only cross at $n = 1$; for a fixed B_{max}, as n increases, $Var(D_n)$ increases at the beginning and reaches its maximum value at $n = B_{max}$, then decreases to its local minimum

point at $n = 2B_{max}$. Then $Var(D_n)$ increases again and stays at two different values for odd and even n respectively. The difference between the steady state for odd and even n is larger using the RAR than the TBD within blocks.

The PBD(B, R) and PBD(B, T) behave similarly. The variances repeat themselves with every length of $2B$ as n increases. The value regresses to 0 at the end of every period. For each n, $Var(D_n)$ does not increase as B increases; some n is a multiple for various values of $2B$, then for those n, $Var(D_n)$ achieves to 0 at different values B. If n is a prime number, then $Var(D_n)$ never reach 0. The difference between the PBD(B, R) and PBD(B, T) is that for the PBD(B, R), the curve for the first half block is symmetric with the curve for the second half block. For the PBD(B, T), $Var(D_n)$ increases linearly in the first half block, then drops smoothly in the second half block. The slope of the linear part is constant across all values of B, which is 1.

BCD(0.5) is equivalent to complete randomization, hence it has a very large value of $Var(D_{100})$, which is 100. When p rises to 0.55, $Var(D_{100})$ drops greatly to 33.5. For the BCD(0.60), and BCD(2/3), $Var(D_{100})$ is 12.1 and 4.4, respectively. We see that $Var(D_n)$ is a decreasing function in p for each n, because as p gets larger, the randomization is more restricted, the probability of large imbalance gets smaller. For a fixed p, Figure 4.2 shows an increasing concave curve in n. The curve is smoother for smaller value of p. For large values of p, we see that $Var(D_n)$ converges to two different values for even and odd n.

The curves of $Var(D_n)$ for the BSD(a) look similar to those for the BCD(p). As a increases, $Var(D_n)$ increases. Since the allocation probabilities for the first a treatment assignments are 1/2, which is complete randomization, Figure 4.2(f) shows a linear relationship with slope 1 for n from 1 to a. From Figure 4.1(f) we notice that $Var(D_n)$ rises after a exceeds 10.

When $B_{max} = B = a$, RBD(B_{max}), PBD(B) and BSD(a) control the maximum imbalance throughout the trial at the same level, but the variabilities of the terminal imbalance are different. The sequence is RBD(B_{max}) $<$ PBD(B) $<$ BSD(a). For blocked randomization, using the RAR to assign patients within each block has a smaller variability of final

imbalance than using the TBD within each block.

Overall, the RBD(B_{max}, R) outperforms the other five restricted randomization procedures with respect to the variability of the terminal imbalance.

4.3 Conclusion

The variability of the terminal imbalance in treatment assignments for restricted randomization has been investigated in this chapter. Five theorems are provided to calculate the variance of final imbalances of five restricted randomization procedure. If the only consideration is the final imbalance in treatment assignments, the RBD(B_{max}, R) is the best procedure. If the total number of patients n is known, the clinical trialist can choose a PBD design with 0 variances for the terminal imbalance. Alternatively, if one can establish a fixed sample size that is divisible by all possible realizations of the block sizes from the RBD, one can eliminate the possibility of imbalance (Heussen, 2004).

Regarding the maximum imbalance in the course of a trial, one can find the simulation results of the average maximum imbalance of a trial with 50 patients for nine restricted randomization procedures in Rosenberger and Lachin (2002). Further simulation results of the maximum imbalance throughout a trial of size 100 for 20 restricted randomization procedures are provided in Table IV of Zhao, et al. (2011). It is expected that both the maximum imbalance and average maximum imbalance for the BSD(a) is a. The maximum imbalance for the RBD(B_{max}) and PBD(B) is B_{max} and B, respectively. It is nontrivial to establish the exact formula to calculate $E(\max_{1\leq j\leq n}|D_j|)$ for blocked randomization.

Table 4.1: Asymptotic distribution of $|D_n|$ for complete randomization

n	$[0, 10\%n)$	$[10\%n, 20\%n)$	$\geq 20\%n$
50	0.52	0.32	0.16
100	0.68	0.27	0.05
150	0.78	0.21	0.01
200	0.84	0.15	0.00
400	0.95	0.05	0.00
800	1.00	0.00	0.00

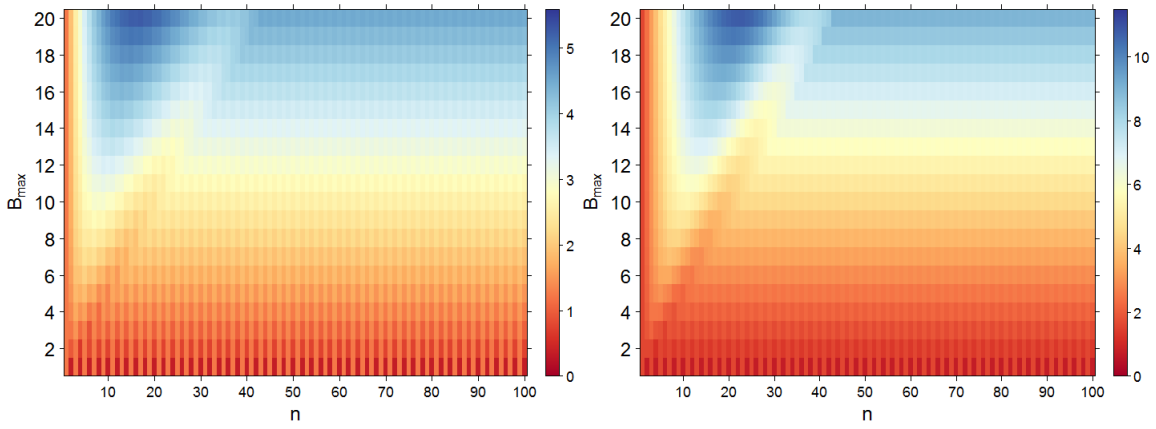
Table 4.2: Variance of imbalance of the RBD(B_{max}; R) and PBD(B; R) for different values of n, B_{max} and B

	RBD(B_{max}; R)						PBD(B; R)					
n	3	4	5	6	7	8	2	3	4	5	6	7
5	1.09	1.41	1.72	1.98	2.21	2.40	1.00	1.00	2.14	2.78	3.18	3.46
10	0.76	1.08	1.10	1.40	1.78	2.16	1.33	1.60	1.71	0.00	1.82	3.08
15	1.13	1.31	1.50	1.70	1.84	1.93	1.00	1.80	1.00	2.78	2.45	1.00
20	0.76	1.02	1.28	1.52	1.75	1.99	0.00	1.60	2.29	0.00	2.91	3.69
25	1.13	1.31	1.50	1.70	1.91	2.13	1.00	1.00	1.00	2.78	1.00	2.54
50	0.76	1.02	1.28	1.52	1.75	1.99	1.33	1.60	1.71	0.00	1.82	3.69
75	1.13	1.31	1.50	1.70	1.91	2.13	1.00	1.80	2.14	2.78	2.45	3.46
100	0.76	1.02	1.28	1.52	1.75	1.99	0.00	1.60	2.29	0.00	2.91	1.85

Table 4.3: Variance of imbalance of the RBD(B_{max}; T) and PBD(B; T) for different values of n, B_{max} and B

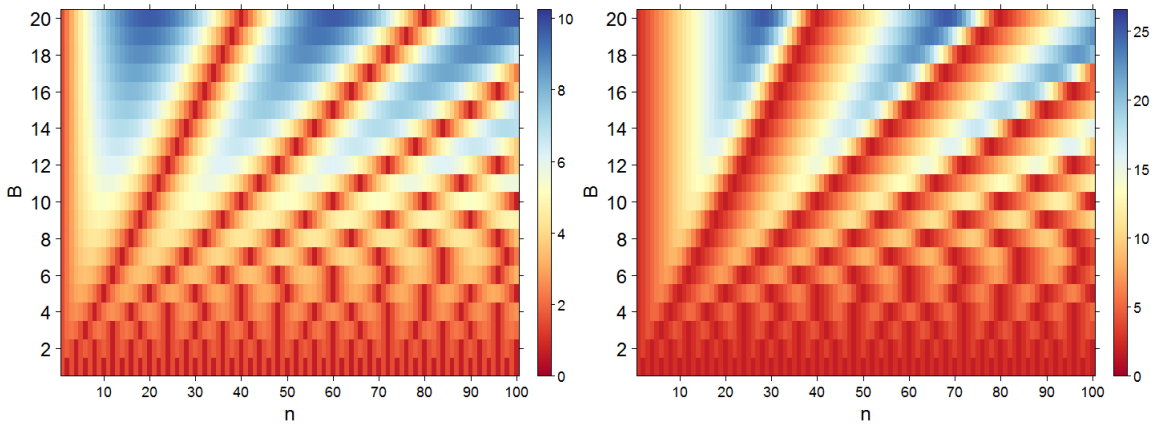
	RBD(B_{max}; T)						PBD(B; T)					
n	3	4	5	6	7	8	2	3	4	5	6	7
5	1.22	1.98	2.62	3.04	3.34	3.55	1.00	1.00	3.92	5.00	5.00	5.00
10	1.08	1.58	1.66	2.20	3.08	3.97	2.00	2.50	2.00	0.00	3.02	6.85
15	1.33	1.69	2.10	2.52	2.76	2.92	1.00	3.00	1.00	5.00	3.00	1.00
20	1.08	1.52	1.96	2.41	2.87	3.33	0.00	2.00	4.00	0.00	6.16	6.00
25	1.33	1.69	2.10	2.52	2.96	3.41	1.00	1.00	1.00	5.00	1.00	5.24
50	1.08	1.52	1.96	2.41	2.87	3.33	2.00	2.00	2.00	0.00	2.00	7.70
75	1.33	1.69	2.10	2.52	2.96	3.41	1.00	3.00	3.00	5.00	3.00	5.00
100	1.08	1.52	1.96	2.41	2.87	3.33	0.00	2.50	4.00	0.00	4.00	2.00

.



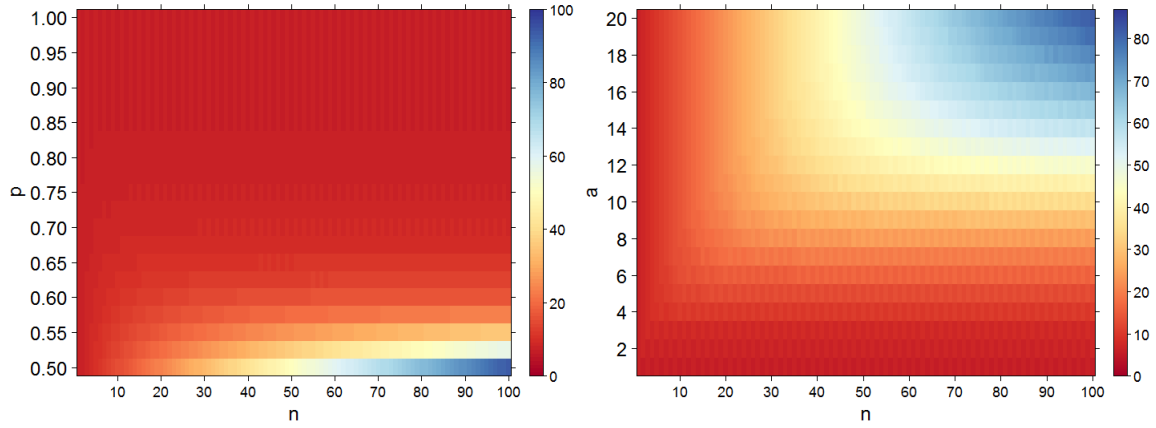
(a) RBD (B_{max}; R)

(b) RBD (B_{max}; T)



(c) PBD (B; R)

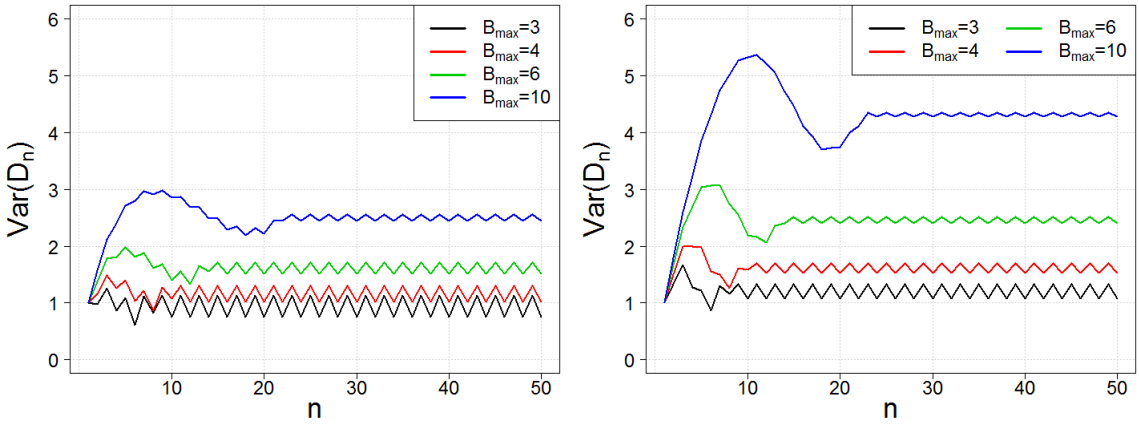
(d) PBD (B; T)



(e) BCD(p)

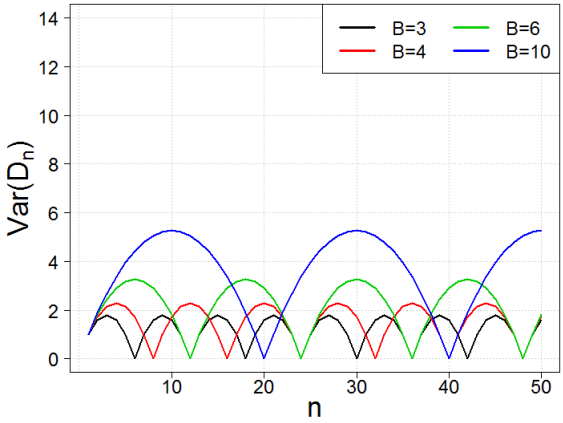
(f) BSD(a)

Figure 4.1: Variance of the final imbalance of six restricted randomization procedures. Note that each panel has different color scales.

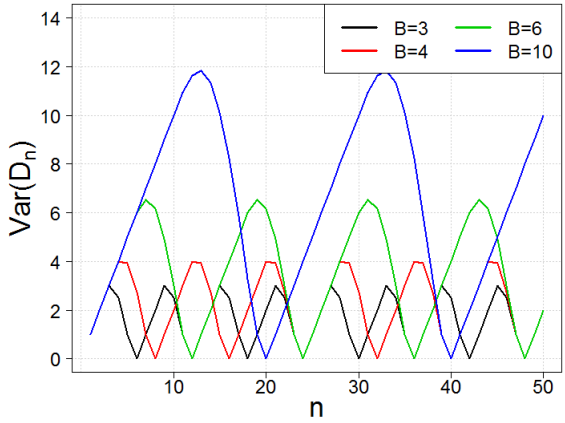


(a) RBD (B_{max}; R)

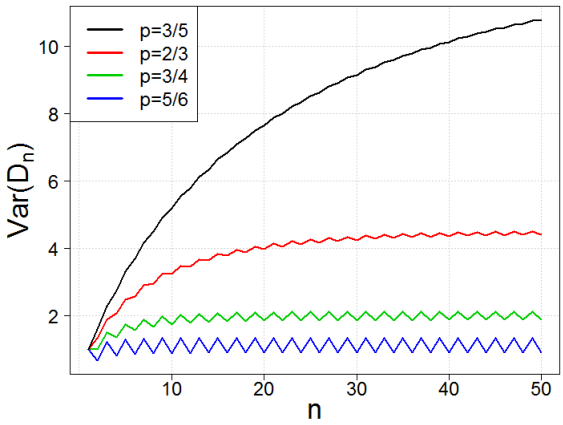
(b) RBD (B_{max}; T)



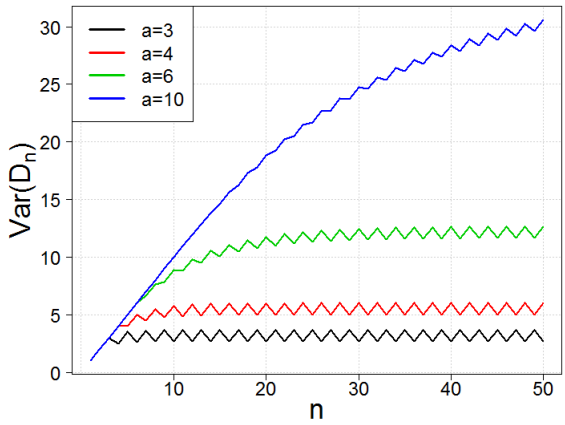
(c) PBD (B; R)



(d) PBD (B; T)



(e) BCD(p)



(f) BSD(a)

Figure 4.2: Variance of the final imbalance of six restricted randomization procedures. Note that each panel has different y-axis scales.

Chapter 5: Randomization Tests

One important advantage of randomization in clinical trials is that it provides an assumption-free statistical test for the treatment effect, which is known as the randomization test. In this chapter, we will investigate the performance of both the randomization test and the standard two-sample population-based test under restricted randomization procedures.

5.1 Background

The traditional and commonly used parametric statistical tests for comparing two treatments are based on random sampling from a population model. For example, in the two sample Student's t-test, it is assumed that the treatment groups are drawn randomly and independently of each other from their respective populations. Each individual in the population has an equal probability of being selected in the sample. This assumption may be inappropriate in a randomized clinical trial, where patients are not sampled randomly from two infinite homogeneous populations. First, there are no populations of patients on treatment A and B. Second, patients are selected non-randomly from non-randomly selected clinics; see Rosenberger and Lachin (2002, Chapter 7).

The null hypothesis of the randomization test is that the two treatments have equal effect. Under this null hypothesis, the observed response of each patient is independent of its treatment assignment. Therefore, for any given sequence of responses, one can tabulate all possible sequences of treatment assignments using the same randomization procedure, which is called the reference set. Each randomization sequence in the reference set generates a test statistic measuring the difference between the two treatment groups. Finally, the exact null distribution is obtained and the p-value of the test is the proportion that the test statistic is equal or greater than the observed one. Like all the basic hypothesis tests, the

p-value is compared with the nominal level, usually 0.05, and a smaller p-value indicates the extremeness of the observed value of the test statistic, hence a strong evidence of the difference between the two treatment groups. For the test statistic, the family of linear rank statistic is often used.

There are unconditional randomization test and conditional randomization test based on which reference set is used. The *unconditional reference set* contains all the possible permutation sequences, hence the number of sequences is 2^n. The *conditional reference set* only includes those sequences with the same number of patients assigned to treatment A and B as the observed values from the trial, hence the number of sequences is $\binom{n}{n_A}$. For the block design with fully filled blocks, since the treatments assignments have been forced to $n/2$, the sequences in unconditional reference set and conditional reference set are exactly the same, but with different probabilities. The difficulty of implementing a conditional randomization test is to determine the reference set and calculate the probability of each sequence. Plamadeala and Rosenberger (2012) provided a computational method to approximate conditional randomization test for those restricted randomization procedures that assign next patient with respect to previous treatment assignments. In particular, they derive the exact conditional distribution of $N_A(n)$ given $N_A(j)$ for the BCD(p), and also the conditional variance-covariate matrix of the randomization sequence. In this study, we only focus on the unconditional randomization test.

Today, the availability of fast computers has made randomization tests more feasible, even for large data sets. Monte Carlo simulation with sufficient sequences can be employed instead of enumerating all possible sequences (Zhang and Rosenberger, 2008). In a clinical trial of size n, a randomization procedure is employed to assign treatment to patients, and the response $x_1, x_2, \ldots, x_n$ are observed. A test statistic S_{obs} is calculated to measure the treatment difference between the two groups. To implement an unconditional Monte Carlo re-randomization test, one simulates L randomization sequences using the same randomization procedure as in the clinical trial and calculate the corresponding test statistic

$S_l, l = 1, 2, \ldots, L$, using the observations $x_1, x_2, \ldots, x_n$. The two-sided Monte Carlo p-value estimator is then defined as

$$\hat{p}_u = \frac{\sum_{l=1}^{L} I(|S_l| \geq |S_{obs}|)}{L}.$$

Galbete and Rosenberger (2015) compared the Monte Carlo re-randomization test with the exact and asymptotic test for some randomization procedures for which there is a known exact or asymptotic distribution of the test statistic. They conclude that Monte Carlo re-randomization test with sufficient sample size is very accurate. They find that L=15,000 is enough to produce a stable Monte Carlo distribution of the test statistic which is also close to the computation of an exact test. When the response of the study is binary and the sample size is small, the exact computation is theoretically the best and fastest to compute, but Monte Carlo re-randomization can be used when it is difficult to obtain the exact distribution. When the response of the study is continuous and the sample size is large, they suggest using the Monte Carlo simulation.

In this study, we investigate the type I error rate and power of both the Monte Carlo re-randomization test and standard two-sample t-test on restricted randomization procedures with continuous responses. The responses follow two models: one is a population model under which the assumption of the t-test is satisfied; the other one is a worse case scenario, which is under a vigorous linear time trend and the assumption of the t-test is violated. The restricted randomization procedures including the BCD(p), BSD(a), RBD(B_{max}, R), RBD(B_{max}, T), PBD(B, R), and PBD(B, T).

5.2 Simulation Algorithm

A two-sided randomization test and a two-sided two-sample t-test are simulated. The test statistic is the mean difference between two treatment groups. For each test, we choose $n = 50$ and significant level $\alpha = 0.05$. Each test is simulated 10,000 times. For each of the

10,000 tests, the response data is generated under two models:

1. Under H_0, $X_1, X_2, \ldots, X_{50} \sim$ i.i.d. $N(0,1)$.

   Under H_1, treatment A has a mean shift of 1.

2. Under H_0, $X_1, X_2, \ldots, X_{50}$ are distributed linearly on the interval $(-2, 2]$ plus a $N(0,1)$ random variable.

   Under H_1, treatment A has a mean shift of 1.

The treatment assignments are generated under the six restricted randomization procedures. We choose $p = 2/3, a = 3, 4, 6, 10, B_{max} = 3, 4, 10$, and $B = 2, 3$. For the t-test, we do not assume equal variance though the data is generated under equal variance. Hence, the degree of freedom of the test is calculated by the Welch-Satterthwaite equation. For each Monte Carlo re-randomization test, $L = 15,000$ sequences are used for computation.

5.3 Simulation Results

The simulation results of the nominal size and power are listed in Table 5.1 and Table 5.2, respectively. We see under model 1, both tests have size around 0.05 and power around 0.93, which indicates that both tests maintain size and power. The t-test is not more powerful than the randomization test for all the restricted randomization procedures. Moreover, there is no difference in the performance of the two tests among all the randomization procedures.

Under model 2, when the linear time trend violates the assumption of the t-test, the t-test fails to maintain nominal size and is very conservative except under the BSD(6) and BSD(10). However, the size of the randomization test remains around 0.05 under all the procedures.

The power of both tests is reduced under model 2. With the linear time trend, the power of the t-test is around 0.64 under all the randomization procedures. The power of the randomization test varies in procedures, with the lowest value under the BSD(10).

We see the power decreases as a increases under the BSD. All the blocked randomization procedures have similar power in randomization test.

5.4 Conclusion

We have conducted a comparison between the randomization test and t-test under 15 randomization procedures from six restricted randomization designs. It is expected that a parametric model would be more powerful than a nonparametric model if the data fits the population assumption of the parametric model. In our simulation results, the randomization test is comparable to the t-test under a population model for 15 randomization procedures. Moreover, the randomization test is robust to time trends in both size and power. The t-test fails to preserve the nominal size and has much smaller power comparing with that of the randomization test.

The type of randomization procedure does not affect the size and power under the population model. However, when there is a time trend, the power of randomization test depends on the randomization procedure. Using block randomization has a higher power in the randomization test than using the BCD(p) or BSD(a), but there is not much difference for the power between different block sizes. As a increases, the BSD(a) approaches to complete randomization. Hence, the t-test behaves as under complete randomization (Tamm and Hilgers, 2014).

Table 5.1: Simulated size of the randomization test and t-test under a single time shift and linear time trend. Each simulation based on 10,000 tests, $n = 50$. For the randomization test, $L = 15,000$

	Model 1		Model 2	
Procedures	Randomization Test	t-test	Randomization Test	t-test
BCD(2/3)	0.04	0.05	0.05	0.01
BSD(3)	0.05	0.05	0.05	0.00
BSD(4)	0.05	0.05	0.05	0.02
BSD(6)	0.05	0.05	0.06	0.04
BSD(10)	0.06	0.05	0.05	0.05
RBD(3,R)	0.05	0.04	0.05	0.00
RBD(3,T)	0.05	0.05	0.05	0.00
RBD(4,R)	0.04	0.04	0.05	0.00
RBD(4,T)	0.04	0.04	0.05	0.00
RBD(10,R)	0.05	0.05	0.05	0.01
RBD(10,T)	0.05	0.05	0.05	0.01
PBD(2,R)	0.05	0.04	0.05	0.00
PBD(2,T)	0.05	0.05	0.05	0.00
PBD(3,R)	0.05	0.05	0.05	0.00
PBD(3,T)	0.05	0.05	0.05	0.00

Table 5.2: Simulated power of the randomization test and t-test under a single time shift and linear time trend. Each simulation based on 10,000 tests, $n = 50$. For the randomization test, $L = 15,000$

	Model 1		Model 2	
Procedures	Randomization Test	t-test	Randomization Test	t-test
BCD(2/3)	0.92	0.93	0.78	0.64
BSD(3)	0.93	0.93	0.83	0.61
BSD(4)	0.94	0.94	0.82	0.64
BSD(6)	0.94	0.94	0.71	0.64
BSD(10)	0.94	0.94	0.63	0.63
RBD(3,R)	0.93	0.93	0.90	0.65
RBD(3,T)	0.93	0.93	0.92	0.64
RBD(4,R)	0.93	0.94	0.92	0.64
RBD(4,T)	0.94	0.94	0.91	0.64
RBD(10,R)	0.94	0.93	0.89	0.64
RBD(10,T)	0.94	0.94	0.85	0.64
PBD(2,R)	0.93	0.93	0.88	0.65
PBD(2,T)	0.93	0.94	0.91	0.64
PBD(3,R)	0.94	0.94	0.92	0.64
PBD(3,T)	0.93	0.94	0.91	0.64

Chapter 6: Graphical Comparisons with Bivariate Objectives

In practice, choosing an appropriate randomization procedure involves multiple criteria and competing goals. In previous sections, we have compared the six restricted randomization procedures with respect to the predictability, variability of terminal imbalance and performance of randomization tests separately. In this chapter, we will conduct graphical comparisons with respect to bivariate objectives.

Graphics has been used in many literatures for comparing many randomization designs simultaneously or choosing the appropriate design for a given sample size when two objectives are taking into account. Zhao, et al. (2011) plotted simulated correct guess probability against the maximum absolute imbalance for 260 randomization design scenarios. Atkinson (2014) plotted the simulated average expected selection bias and loss of precision for nine randomization procedures. Such graphs can also be found in Atkinson (2012) and Baldi Antognini and Zagoraiou (2014).

In this chapter, for both the predictability and imbalance metrics, we divide by n to simplify the comparisons. The two metrics become the degree of predictability per patient (ρ_{PRED}/n), which ranges from 0 to 0.5, and the average variance of terminal imbalance ($Var(D_n)/n$), which ranges from 0 to 1. We treat the predictability and imbalance equally in importance. Therefore, in our trade-off plot, the ratio of range of x-axis and y-axis is always 1:2, which ensures that same changes in x and y directions are equivalent in importance.

Note that it is not necessary that same change in y and x directions are equivalent in importance. We call this type of plot as a *trade-off plo*t, mainly because predictability and treatment imbalances are completing objectives. Any procedure appearing in the left bottom corner would be the best procedure for a given sample size with the lowest variance of terminal imbalance and lowest degree of predictability.

6.1 Predictability Versus Imbalance

In this section, we compare restricted randomization procedures with respect to terminal imbalance and treatment assignments predictability by plotting the average variance of terminal imbalance against the average degree of predictability.

We first explore the trade-off between predictability and imbalance for each design individually in Figure 6.1 - 6.4. We consider both cases when n is even and odd. Lines in different colors are for different values of sample size. Different symbols on the lines represent different values of the parameter for that randomization procedure. As observed from these figures, the predictability and imbalance metrics are in conflict. We also see all the randomization procedures have a similar pattern except the PBD(B). For the RBD(B_{max}, R) and RBD(B_{max}, T), as B_{max} increases, for a given sample size, the average variance of terminal imbalance decreases, the degree of predictability increases. A similar trend is observed for the BCD(p) when p is increasing. For the BSD(a), the average variance of imbalance decreases and the degree of predictability increases as a decreases. Moreover, each of these four designs behaves almost the same for even n and odd n. The PBD(B, R) and PBD(B, T) behave quite differently from the previous four designs. For a given sample size n, the PBD that has the largest divisor of n as its block size is the best design. For example, when $n = 200$ in Figure 6.2, among the three divisors ($2B = 4, 8, 20$) of 200, 20 is the largest one. The PBD(10) appears in the left bottom corner of each panel, with the smallest degree of predictability and perfect balance. However, the maximum absolute imbalance throughout the experiment is also the largest.

When B_{max} varies from 2 to 10, the difference of the average degree of predictability is about 0.1 under both the RBD(B_{max}, R) and RBD(B_{max}, T). But the difference of the average variance of terminal imbalance between $B_{max} = 2$ and $B_{max} = 10$ under the RBD(B_{max}, T) is much larger than that under the RBD(B_{max}, R). The same phenomenon is observed for the PBD.

The lines under the BSD are much steeper than under the RBD and BCD. The average

variance of terminal imbalance increases sharply as a increases, but the average degree of predictability declines slowly.

In Figure 6.5, we compare the following randomization procedures for a given sample size in a trade-off plot. For the BSD(p), we choose $p = 2/3$, which is recommended by Efron (1971), as well as $p = 3/5$, which is recommended by Pocock for a large trial (1983). For the BSD(a), we choose $a = 3, 4, 6$ and 10. For the RBD(B_{max}, R) and RBD(B_{max}, T), we let $B_{max} = 2, 3, 4, 10$. For the PBD(B, R) and PBD(B, T), we choose $B = 2, 3, 4, 10$ as well. We have seen that there is no significant difference between even and odd sample size, hence we only consider the even n here. In Figure 6.5, different colors represent different randomization designs. The labeling shown in each plot is the value of the design parameter. For example, a green 4 means the BSD(4), and a red 10 means the RBD(10, R).

When $n = 20$ (top left plot), PBD(10, T) and PBD(10, R) are closest to the origin with perfect balance and average degree of predictability around 0.1. Again, these two procedures suffer a downside of having a large maximum imbalance throughout the trial. it appears that in the rest procedures, the BSD(3) is closest one to the origin. The BSD(a) has a smaller average degree of predictability than the PBD(10, T) and PBD(10, R), and an average variance of terminal imbalance of 0.13. The BCD(2/3), PBD(4, T) and RBD(10, T) have similar predictability and imbalance. The other blocked randomization procedures have a very low variance of imbalance but a relatively high degree of predictability. When $n = 50$ in the plot (b), the BSD(3) appears to be the best design or the BSD(4) with a relatively high variance of imbalance and a smaller degree of predictability. When $n = 100$ and 200, the BSD(3), BSD(4) and PBD(10, T) appear in the left quadrant. For a same n, the PBD(10, T) has similar degree of predictability to the BSD(3), and generates perfect balanced treatment groups if n is divisible by 20.

We compare the RBD and PBD in Figure 6.6 to see whether the RBD is better than the PBD. As observed in all four plots, the RBD does not perform better than the PBD. There is a significant difference between even and odd sample size. Though the RBD have relatively smaller variances of imbalance, the degree of predictability is higher.

6.2 Predictability or Imablance Versus Type II Error

In order to show the best design in the left quadrant, we plot the type II error probability instead of power since the higher power the lower the type II error probability. Note that the best type II error probability is 0.07. We compare various restricted randomization procedures with respect to the type II error probability under a linear time trend and degree of predictability in Figure 6.7 and the type II error probability under a linear trend and variance of imbalance in 6.8. For the type II error probability, the simulation results in Chapter 5 are used here.

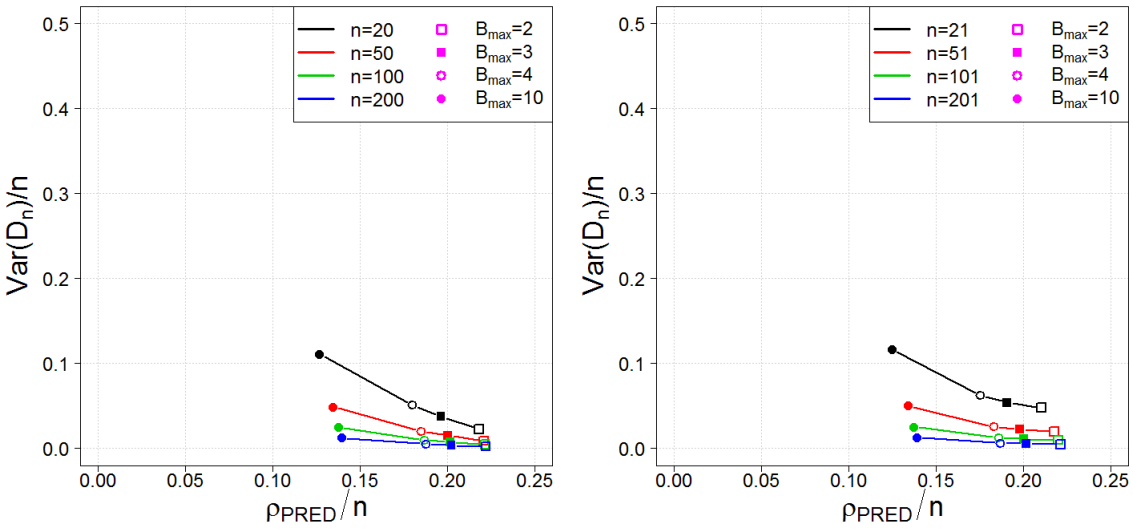
Figure 6.7 indicates that the type II error probability and degree of predictability are competing. For the BSD, a smaller a a smaller type II error probability a larger degree of predictability. For the blocked randomization, the type II error probability is around 0.1 for all the parameter selected except the RBD(10, T), which has a relative larger type II error probability. No procedure appears in the left bottom corner in Figure 6.5. The RBD(4, T), RBD(4, R), PBD(3, R) or PBD(3, T) have the smallest type II error probability with a degree of predictability slightly less than 0.2 while the BSD(3) and BSD(4) have the degree of predictability less than 0.1 but relatively higher type II error probability.

Figure 6.8 shows that the type II error probability and variance of imbalance behave consistently. All the blocked randomization procedures have a similar variance of imbalance and type II error probability except the RBD(10, T). The BSD(10) with a large variance of the imbalance and large type II error probability is almost in the middle of the graph.

6.3 Conclusion

In this chapter, we have conducted graphical comparison between six restricted randomization procedures with bivariate objectives: predictability versus imbalance, predictability versus type II error probability and imbalance versus type II error probability. The predictability is competing with the imbalance and type II error probability. A small degree of the predictability, a large variance of the terminal imbalance and a large type II error

probability. For the goal of minimizing predictability and achieving treatment balance, the BSD(3) appears to be the best procedure for a small size trial, and both the BSD(4) and BSD(3) perform best for a moderate and large trial. For the goal of minimizing predictability and preserving type II error probability under a linear time trend, when $n = 50$, the best procedure does not exist. One can choose the RBD(4, T), RBD(4, R), PBD(3, R) or PBD(3, T) if one prefers a smaller type II error probability, in other words, a higher power. If a lower degree of predictability is preferable, the BSD(3) and BSD(4) are good choices. Here we treat each objective equivalent; in practice, one can change the scaling if one objective is considered more important, but then the criteria will not be on the same scale.



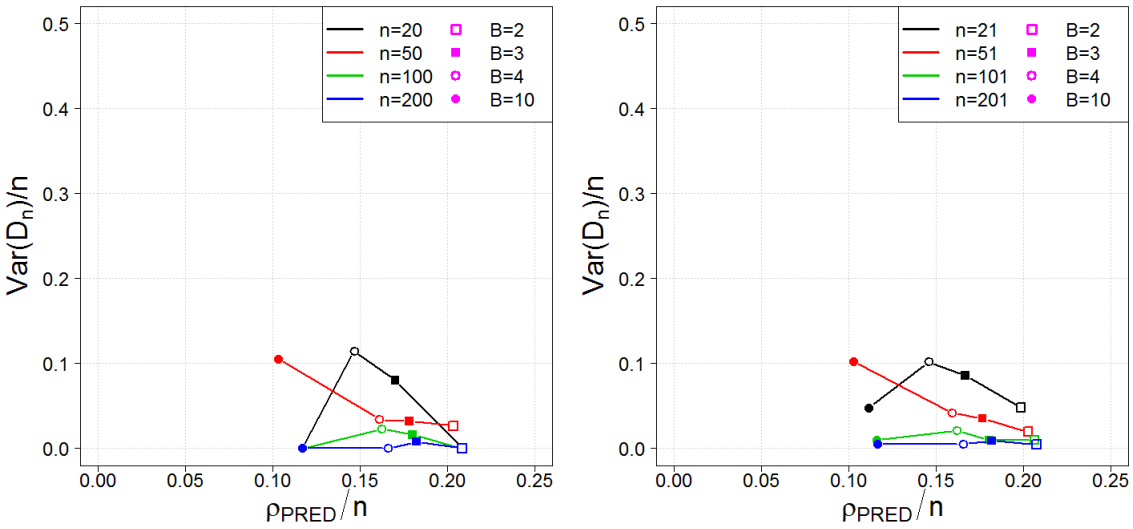
(a) RBD(B_{max}; R), n is even

(b) RBD(B_{max}; R), n is odd

(c) RBD(B_{max}; T), n is even

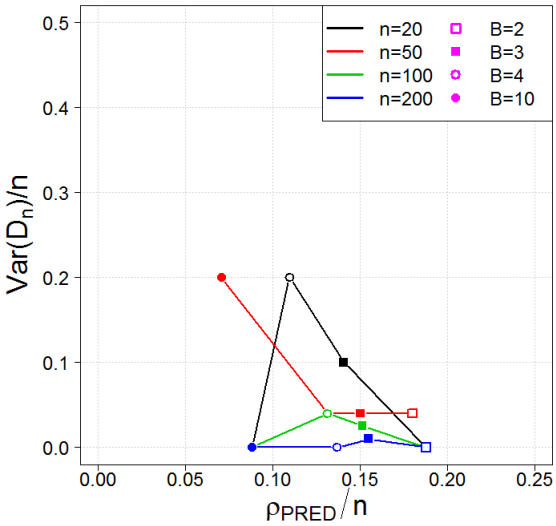
(d) RBD(B_{max}; T), n is odd

Figure 6.1: Trade-off plot for the RBD, comparing imbalance and predictability measures.

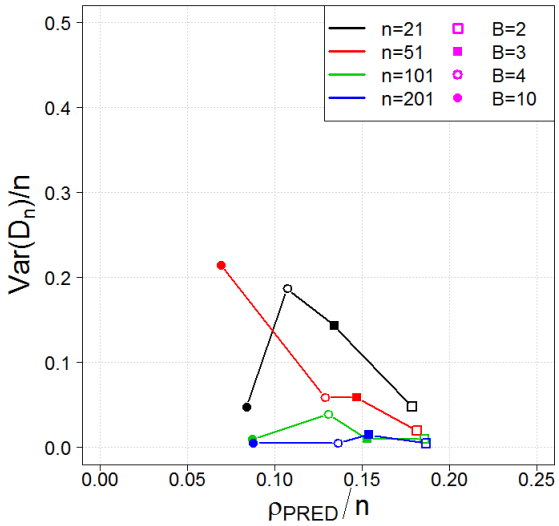


(a) PBD(B; R), n is even

(b) PBD(B; R), n is odd

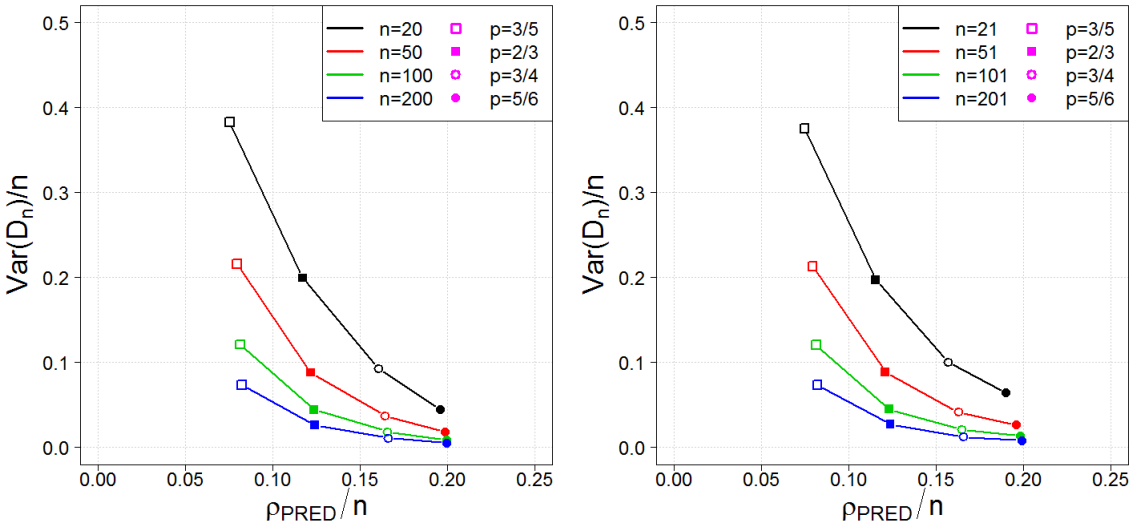


(c) PBD(B; T), n is even



(d) PBD(B; T), n is odd

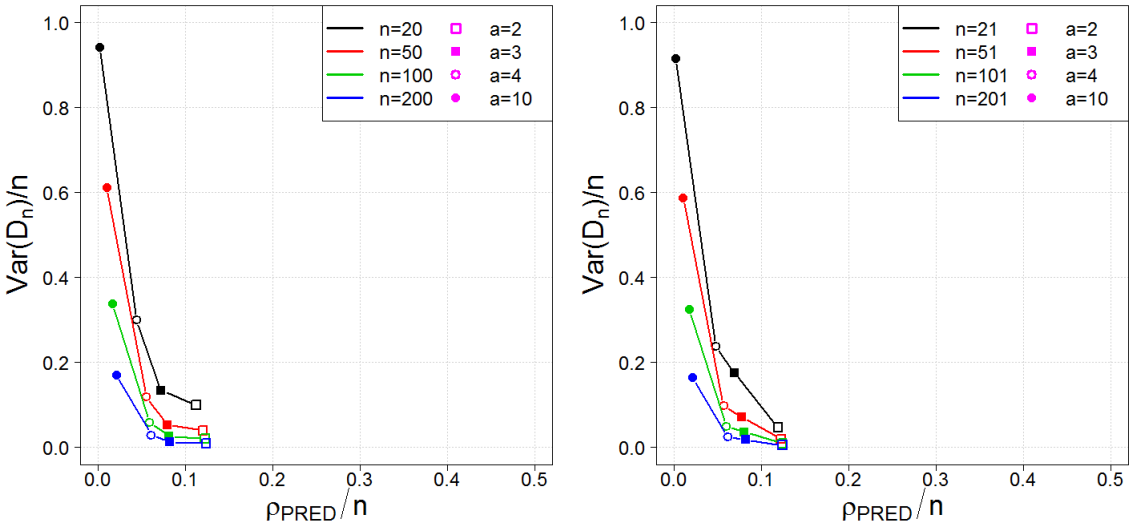
Figure 6.2: Trade-off plot for the PBD, comparing imbalance and predictability measures.



(a) BCD(p), n is even

(b) BCD(p), n is odd

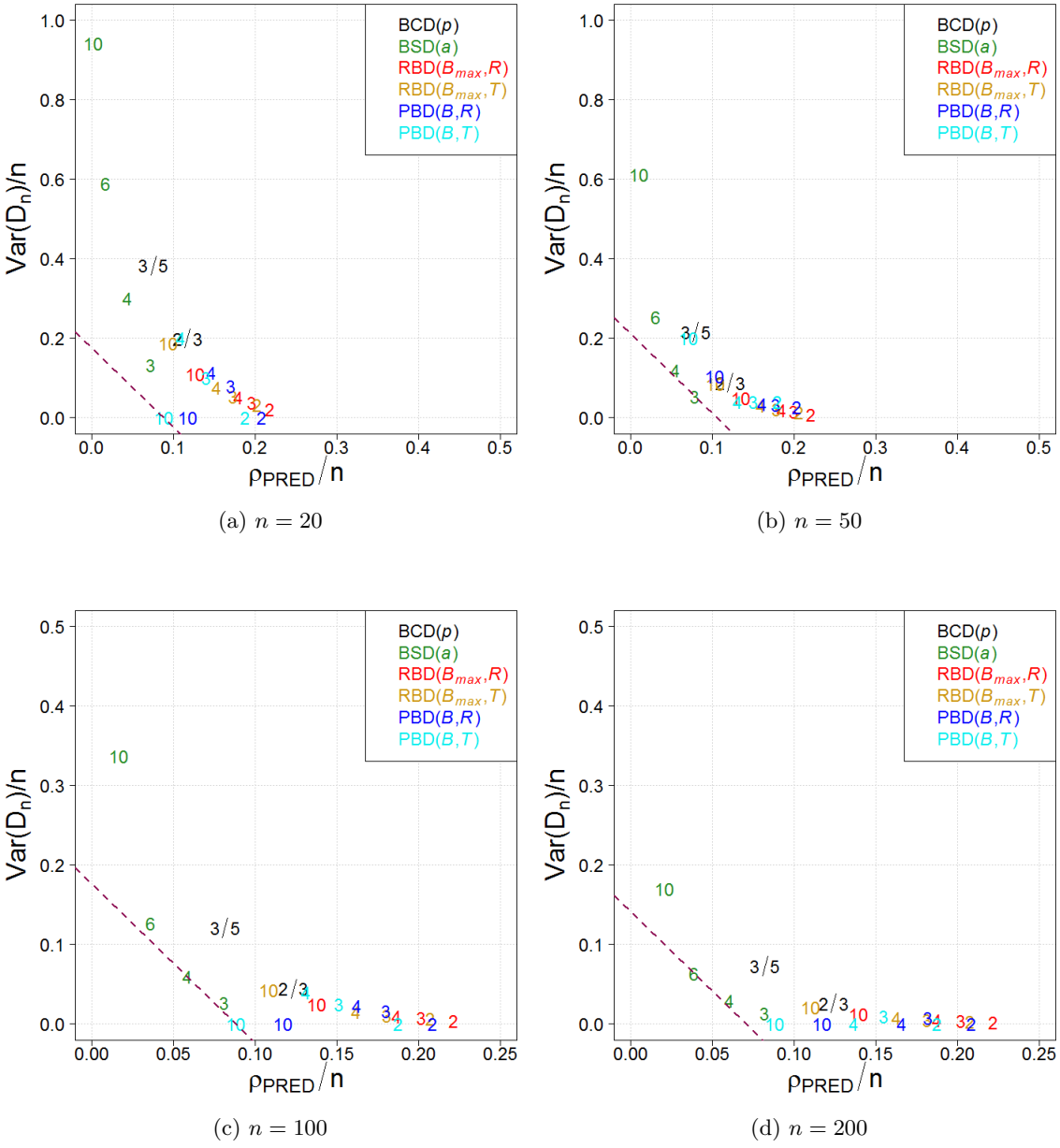
Figure 6.3: Trade-off plot for the BCD(p), comparing imbalance and predictability measures.



(a) BSD(a), n is even

(b) BSD(a), n is odd

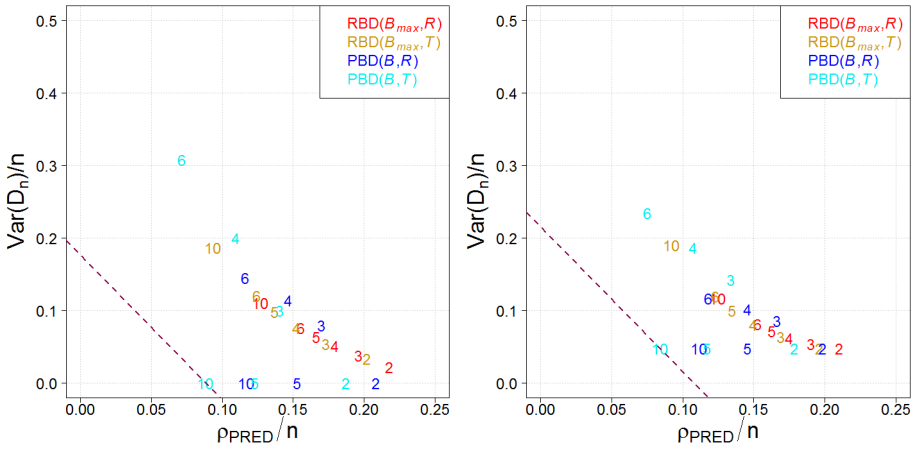
Figure 6.4: Trade-off plot for the BSD(a), comparing imbalance and predictability measures. Note that axes scales are different from Figure 6.1-6.3.



(a) $n = 20$ (b) $n = 50$

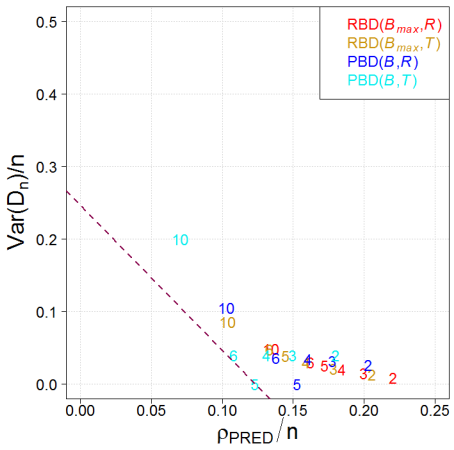
(c) $n = 100$ (d) $n = 200$

Figure 6.5: Trade-off plot for various restricted randomization procedures, comparing imbalance and predictability measures. Note that the upper panel and lower panel have different axes scales.

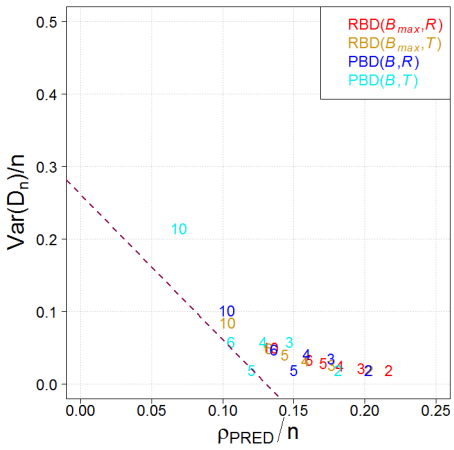


(a) $n = 20$

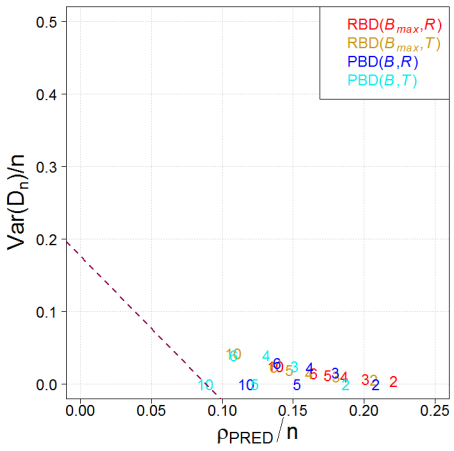
(b) $n = 21$



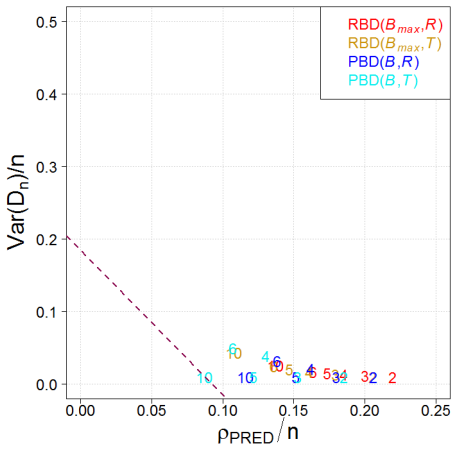
(c) $n = 50$



(d) $n = 51$



(e) $n = 100$



(f) $n = 101$

Figure 6.6: Trade-off plot for the RBD and PBD, comparing imbalance and predictability measures

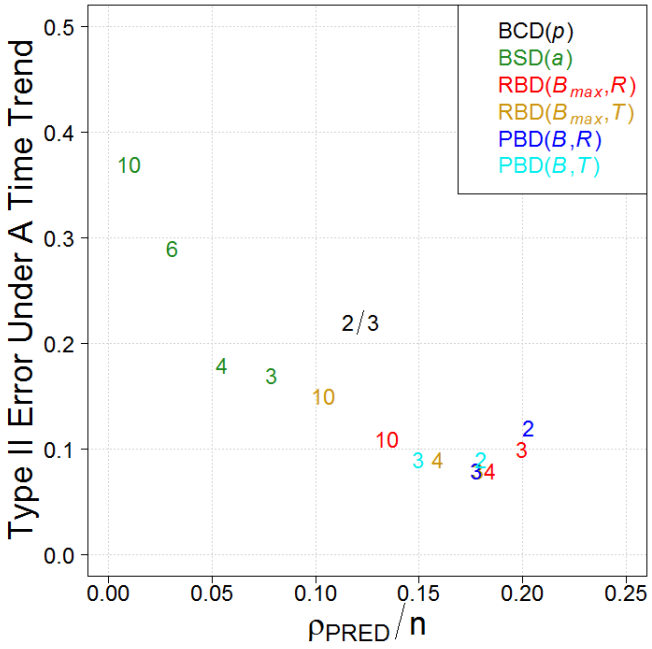


Figure 6.7: Trade-off plot for various restricted randomization procedures,$n = 50$, comparing predictability and type II error probability under a linear time trend

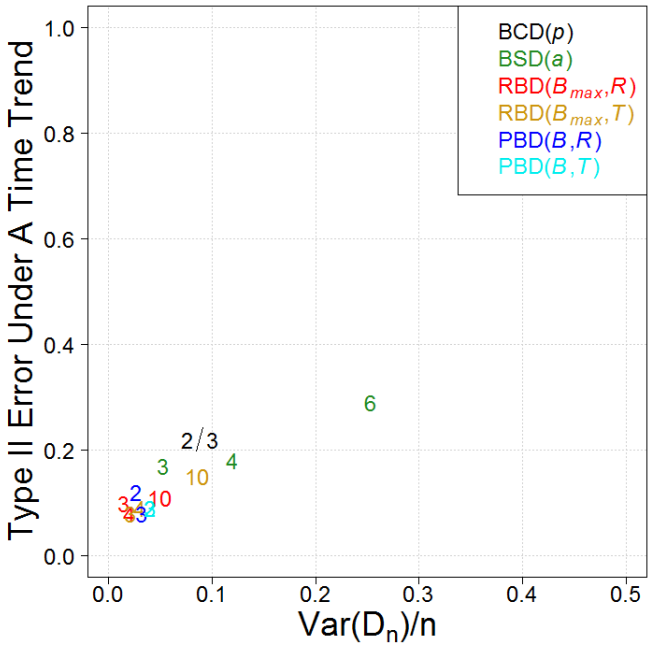


Figure 6.8: Trade-off plot for various restricted randomization procedures,$n = 50$, comparing imbalance and type II error probability under a linear time trend

Chapter 7: Conclusions and Future Work

In this chapter, we reiterate the contributions of this thesis.

The random block design is rarely carefully defined when it is employed in clinical trials. In this thesis, we have rigorously formalized the RBD by assuming a discrete uniform distribution for block size. Blocks sizes are randomly selected from even integers $2, 4, \ldots, 2B_{max}$ with equal probability, where B_{max} is predefined by investigators, and then patients are randomized within each block. Each block can be filled by either the RAR or the TBD. We suggest that when referencing the RBD, it is necessary to indicate the range of the block size and by which method each block is filled.

One of the contributions of this thesis is that we derived some important distributional properties of the RBD by combinatorics, such as the joint distribution of the block size and position number within the block, and the distribution of the terminal imbalance of treatment assignments for both the RBD(B_{max}, R) and RBD(B_{max}, T). We proposed the degree of predictability ρ_{PRED}, which is mathematically equivalent to the Blackwell-Hodges selection bias factor, to find the exact selection bias of restricted randomization procedures. With the distribution results of the RBD, we quantified ρ_{PRED} and balancing properties in closed-form formulas. This provides a statistical understanding of the RBD that can be used in comparison with other restricted randomization procedures.

Matt and Lachin (1988) calculated the selection bias for the PBD when each block is filled by the RAR and the last block is filled. We have not seen any research on the selection bias and imbalance of the PBD with unknown n under which the last block may not be fully filled. In this thesis, we have filled this gap by providing exact formulas to calculate the selection bias and variance of terminal imbalance for the PBD when the last block is unfilled. We have results for both the PBD using the RAR and the PBD using the TBD to randomize patients within each block.

We have conducted a comparison of the predictability and imbalance between the RBD and the PBD. We found the RBD does not reduce predictability as one might expect compared to the PBD. When $B_{max} = B$, the sequence of the degree of predictability is PBD(B, T) < PBD(B, R) $\approx$ RBD(B_{max}, T) < RBD(B_{max}, R). Regarding comparing the variance of the imbalance between the RBD and PBD, any general conclusion would depend on the value of n.

We also compared blocked randomization with other restricted randomization procedures whose exact distribution has been found, including the BCD and BSD, with respect to the selection bias and imbalance. Despite the fact that blocked randomization is the most commonly used procedure in clinical trials, for a moderate and large trial, both the BSD(3) and BSD(4) are better with comparative variances of imbalance to the blocked randomization but much smaller degree of predictability, and smaller maximum imbalances during the course of the trial. For a small trial, the BSD(3) performs better in balancing the predictability and imbalance.

Additionally, we investigated the properties of randomization tests in restricted randomization procedures. We found the randomization test is comparable to the two-sample Student's t-test under a population model, and robust to time trends in both size and power. Under a linear time trend, using blocked randomization has a higher power than using the BCD and BSD.

There are limitations in this study. While we claim we found the exact distribution for the RBD, we were unable to find the exact distribution for $j > 2B_{max} + 2$. In particular, in Lemma 6, we did not obtain the closed form for $\lambda_1, \ldots, \lambda_{B_{max}-1}$. However, one can use the recursive equation in Lemma 4 to calculate the exact distribution. When we quantify the predictability and imbalance for the RBD, for $j > 2B_{max} + 2$, we used the asymptotic approximation in Lemma 7, which simulation shows was quite accurate, even for a small sample. Another limitation is that we only investigated the performance of the randomization test under a few scenarios. A more explicit simulation is expected in the future, including more values of n and more restricted randomization procedures.

The following is the outline of future work:

1. When qualifying the balancing property of a randomization procedure, we want a lower value of both the variance of terminal imbalance and the average maximum imbalance throughout the trial. For example, a PBD with a block size 20 generates perfect balancing treatment group for a trial of size n that is divisible by 20, but there is a probability of having an imbalance of 10 during the trial, which may bias the study result if there is a drift in patient characteristics. We suggest quantifying the average maximum imbalance, $E(\max_{1\leq j\leq n}|D_j|)$, of the blocked randomization in the future, using the exact distributional results in Chapter 2.

2. One of the reasons that the RBD does not perform better than the PBD is that the RBD involves blocks of size 2. This greatly increases the degree of predictability. We propose the truncated random block design without a block of size 2, which is the RBD(B_{max}) with B_{max} starting at 2. We are working on the distribution of this truncated random block design. One also can consider varying the block sizes with unequal probabilities which puts a lower weight on the block size of 2.

3. The exact distribution of terminal imbalance of the BCDII defined in (1.5), as well as its selection bias, have been found by Chen (1999). However, there is little study on the performance of this design in the selection bias and balance. We tend to obtain the variance of the imbalance of the BCDII using the distribution of D_n and add this design into our comparison.

Bibliography

Bibliography

[1] Atkinson, A.C. (1982). Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* **69** 61-67.

[2] Atkinson, A.C. (1999). Optimum biased-coin designs for sequential treatment allocation with covariate information. *Statist. Med.* **18** 1741-1752.

[3] Atkinson, A.C. (2002). The comparison of designs for sequential clinical trials with covariate information. *J. R. Statist. Soc. C.* **165** 349-373.

[4] Atkinson, A.C. (2003). The distribution of loss in two-treatment biased-coin designs. *Biostat.* **4** 179-193.

[5] Atkinson, A.C. (2012). Bias and loss: the two sides of a biased coin. *Statist. Med.* **31** 3494-3503.

[6] Atkinson, A.C. (2014). Selecting a biased-coin design. *Statist. Sci.* **29** 144-163.

[7] Baldi Antognini, A. and Giovagnoli, A. (2004). A new 'biased coin design' for the sequential allocation of two treatments. *J. R. Statist. Soc. C* **53** 651-664.

[8] Baldi Antognini, A. and Zagoraiou, M. (2014). Balance and randomness in sequential clinical trials: the dominant biased coin design. *Pharm. Statist.* **13** 119-127.

[9] Berger, V. W. (2005). Quantifying the magnitude of baseline covariate imbalances result- ing from selection bias in randomized clinical trials. *Biom. J.* **47** 119-139.

[10] Berger, V. W. (2007). *Selection Bias and Covariate Imbalances in Randomized Clinical Trials.* John Wiley & Sons, New York.

[11] Berger, V. W. and Exner, D. V. (1999). Detecting selection bias in randomized clinical trials. *Control. Clin. Trials* **20** 319-327.

[12] Berger, V. W., Ivanova, A. and Deloria-Knoll, M. (2003). Minimizing predictability while retaining balance through the use of less restrictive randomization procedures. *Statist. Med.* **22** 3017-3028.

[13] Blackwell, D. and Hodges, J. L. Jr. (1957). Design for the control of selection bias. *Ann. Math. Statist.* **28** 449-460.

[14] Burman, C.F. (1996). On sequential treatment allocations in clinical trials. Technical Report, Department of Mathematics, Goteborg.

[15] Chalmers, T. (1990). Discussion of biostatistical collaboration in medical research by JH Ellenberg. *Biometrics* **46** 20-22.

[16] Chen, Y. P. (1999). Biased coin design with imbalance tolerance. *Comm. Statist. Stochastic Models* **15** 953-975.

[17] Chen, Y. P. (2000). Which design is better? Ehrenfest urn versus biased coin. *Advances in Applied Probability*, 738-749.

[18] Cornfield, J. (2012). Principles of research [reprinted from Amer. J. Mental Deficiency 64 (1959), no. 2, 240-252]. *Statist. Med.* **31** 2760-2768.

[19] Cox, D.R. and Miller, H.D. (1965) *The Theory of Stochastic Processes.* John Wiley & Sons, New York.

[20] Dupin–Spriet, T., Fermanian, J. and Spriet, A. (2004). Quantification of predictability in clinical trials using block randomization. *Drug. Inf. J.* **38** 127-133.

[21] Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58** 403-417.

[22] Fisher, R. A. (1935). *The Design of Experiments.* Oliver and Boyd, Edinburgh.

[23] Green, S. B. and Byar, D. P. (1984). Using observational data from registries to compare treatments: the fallacy of omnimetrics. *Statist. Med.* **3** 361-370.

[24] Heussen, N. (2004). Der Einfluss der Randomisierung in Blöcken zufälliger Länge auf die Auswertung klinischer Studien mittels Randomisationstest. RWTH Aachen University, Aachen, Germany.

[25] Hu, F. and Rosenberger, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials.* John Wiley & Sons, Hoboken, NJ.

[26] Karlin, S. (1968). *A First Course in Stochastic Processes.* New York: Academic Press.

[27] Markaryan, T. (2009). *Exact Distributional Properties of Efron's Biased Coin Design with Applications to Clinical Trials.* George Mason University. Fairfax, VA.

[28] Markaryan, T. and Rosenberger, W. F. (2010). Exact properties of Efron's biased coin randomization procedure. *Ann. Statist.* **38** 1546-1567.

[29] Matts, J. P. and Lachin, J. M. (1988). Properties of permuted-block randomization in clinical trials. *Control. Clin. Trials* **9** 327-344.

[30] McEntegart, D. J. (2003). The pursuit of balance using stratified and dynamic randomization techniques: an overview. *Drug. Inf. J.* **37** 293-308.

[31] Plamadeala, V. and Rosenberger, W. F. (2012). Sequential monitoring with conditional randomization tests. *Ann. Stats.* **40** 30-44.

[32] Pocock, S.J. (1983). *Clinical Trials: A Practical Approach.* John Wiley & Sons, New York.

[33] Proschan, M. (1994). Influence of selection bias on type I error rate under random permuted block designs. *Statist. Sinica* **4** 219-231.

[34] Rosenberger, W. F. and Lachin, J. M. (2002). *Randomization in Clinical Trials: Theory and practice.* John Wiley & Sons, New York .

[35] Salama, I., Ivanova, A. and Qaqish, B. (2008). Efficient generation of constrained block allocation sequences. *Statist. Med.* **27** 1421-1428.

[36] Smith, R.L. (1984a). Properties of biased coin designs in sequential clinical trials. *Ann. Statist.* **12** 1018-1034.

[37] Smith, R.L. (1984b). Sequential treatment allocation using biased coin designs. *J. R. Statist. Soc.* **46** 519-543.

[38] Soares, J.F. and Wu, C.F.J. (1983). Some restricted randomization rules in sequential designs. *Comm. Statist. A-Theor. Meth.* **12** 2017-2034.

[39] Tamm, M. and Hilgers, R.D. (2014). Chronological bias in randomized clinical trials arising from different types of unobserved time trends. *Methods Inf Med.* **53** 501-510.

[40] Wei, L.J. (1977). A class of designs for sequential clinical trials. *I. Amer. Statist. Assoc.* **72** 382-386.

[41] Wei, L.J. (1978). The adaptive biased coin design for sequential experiments. *Ann. Statist.* **6** 92-100.

[42] Zhang, L. and Rosenberger, W.F. (2011). Adaptive randomization in clinical trials. In *Design and Analysis of Experiments*, Vol. III. (K. Hinkelmann, ed.). Wiley, Hoboken.

[43] Zhang, Y. and Rosenberger, W.F. (2008). Sequential monitoring of conditional randomization tests: Generalized biased coin designs. *Seq. Anal.* **27** 234-253.

[44] Zhao, W., Weng, Y., Wu, Q. and Palesch, Y. (2011) Quantitative comparison of randomization designs in sequential clinical trials based on treatment balance and allocation randomness. *Pharm. Statist.* **11**, 39-48.

Curriculum Vitae

Hui Shao graduated from China Agriculture University in 2006 with a B.S. degree in Biology. She received her M.S. degree in Ecology from the University of Chinese Academy of Sciences in 2009. She obtained her M.S. degree in Statistics from the University of Toledo in 2012. In the same year, she joined the Ph.D program in Statistics at George Mason University. During her doctoral study, she has been both a graduate teaching assistant and a graduate research assistant, and a mathematical statistician summer intern at the U.S. Food and Drug Administration.