

# Statistical inference of covariate-adaptive randomized clinical trials

Wei Ma  
Qiqihar, Heilongjiang, P.R. China

M.S., University of Virginia, U.S.A., 2011  
B.S., Zhejiang University, P.R.China, 2009

A Dissertation Presented to the Graduate Faculty  
of University of Virginia in Candidacy for the Degree of  
Doctor of Philosophy

Department of Statistics

University of Virginia  
May, 2014

## Abstract

Covariate-adaptive designs are often implemented to balance important covariates in clinical trials. However, the theoretical properties of conventional statistical methods are usually unknown under covariate-adaptive randomized clinical trials. In literature, most studies are based on simulations. In this dissertation, we provide theoretical foundation of statistical inference under covariate-adaptive designs based on linear models and generalized linear models.

In Chapter 2, we derive asymptotic distributions of the test statistics of testing both treatment effects and significance of covariates under null and alternative hypotheses in the linear model framework. Under a large class of covariate-adaptive designs, we find that: (i) the hypothesis testing to compare treatment effects is usually conservative in terms of small Type I error; (ii) the hypothesis testing to compare treatment effects is usually more powerful than complete randomization; and (iii) the hypothesis testing for significance of covariates is still valid. The class includes most of the covariate-adaptive designs in literature, for example, Pocock and Simon's marginal procedure (Pocock and Simon, 1975), stratified permuted block design, etc. Numerical studies are also conducted to assess their corresponding finite sample properties.

In Chapter 3, theoretical properties of hypothesis testing under linear models are studied based on more general assumptions. In particular, the assumption used in Chapter 2 that all covariates are independent of each other is relaxed in this chapter by taking into consideration of the correlation between covariates. Under such general assumptions, we prove that the hypothesis testing to compare treatment effects is still conservative, while the estimators of covariate coefficients are biased under covariate-adaptive designs.

Covariate-adaptive designs are often used in clinical trials where outcome is not continuous. In these scenarios, generalized linear models can be used to perform statistical inference. For example, logistic regression is used when outcome is a binary variable. In Chapter 4, we concentrate on hypothesis testing for treatment effects based on logistic regression under covariate-adaptive designs. We propose a framework to derive theoretical properties of test statistic for stratified covariate-adaptive designs and conclude that the Type I error is also conservative. Numerical studies are also carried out to study power for several tests.

## Acknowledgements

I would like to gratefully and sincerely thank my advisor, Prof. Feifang Hu, for his guidance during my Ph.D. research and study at the University of Virginia. This dissertation is simply impossible without his mentorship and encouragement. His perpetual energy and passion in research had inspired all his students, including me. My research life would not have been so smooth and rewarding without his advice, understanding and patience. It was a great honor to work with him and I could not have imagined having a better advisor and mentor for my Ph.D. study.

I would like to thank Department of Statistics. I am especially thankful to Prof. Tingting Zhang, Prof. Dan Keenan and Prof. Guofen Yan for serving on my committee, devoting their time and providing constructive suggestions. I am grateful to Mrs. Karen Dalton, who was always accessible and willing to help. Additionally, thanks to all the students for making the department a convivial place to study. In particular, I would like to thank Hongjian Zhu, Yanqing Hu and Zhenjun Ma. They have been role models for me to follow and have always been my best counselors.

My deepest gratitude goes to my family for their unconditional love and support throughout my life. I am indebted to my mother, Yongli Ma, for her care and love. She worked industriously to provide me the best possible environment to grow up and receive education. I owe her everything and wish I could show her just how much I love and appreciate her. Last but not the least, I would like to thank my wonderful wife, Mingyue Hao. She was always there cheering me up and stood by me through the good times and the bad. Her love, sacrifice and support provided my inspiration and was my driving force to face challenges.

# Contents

Abstract . . . . .	ii
Acknowledgements . . . . .	iv
<b>1 Introduction</b>	<b>1</b>
1.1 Overview of covariate-adaptive randomized clinical trials . . . . .	1
1.1.1 Randomization . . . . .	1
1.1.2 Adaptive design . . . . .	3
1.1.3 Covariate-adaptive design . . . . .	5
1.2 Statistical inference for covariate-adaptive designs . . . . .	8
1.3 Motivation and organization of the dissertation . . . . .	11
<b>2 Statistical Inference for Linear Models with Independent Covari-</b>	
<b>ates</b>	<b>14</b>
2.1 Introduction . . . . .	14
2.2 Framework . . . . .	15
2.3 Theoretical properties . . . . .	20
2.4 Simulation study . . . . .	27
2.5 Conclusion . . . . .	39
2.6 Appendix: proof of theorems . . . . .	39

<b>3</b>	<b>Statistical Inference for Linear Models with Correlated Covariates</b>	<b>53</b>
3.1	Introduction . . . . .	53
3.2	Framework . . . . .	54
3.3	Theoretical properties . . . . .	61
3.4	Simulation study . . . . .	65
3.5	Conclusion . . . . .	70
3.6	Appendix: proof of theorems . . . . .	71
<b>4</b>	<b>Statistical Inference for Logistic Regression</b>	<b>85</b>
4.1	Introduction . . . . .	85
4.2	Framework . . . . .	86
4.3	Theoretical properties . . . . .	88
4.4	Simulation study . . . . .	90
4.5	Conclusion . . . . .	94
4.6	Appendix: proof of theorems . . . . .	95
<b>5</b>	<b>Conclusions</b>	<b>102</b>
	<b>Bibliography</b>	<b>104</b>

# Chapter 1

## Introduction

### 1.1 Overview of covariate-adaptive randomized clinical trials

#### 1.1.1 Randomization

In medical research and drug development studies, clinical trials are implemented to evaluate safety and efficacy of an intervention treatment, such as a medication, diagnostic test or medical device, and to compare the new treatment with an established treatment or placebo. It is generally agreed that an ideal clinical trial should be randomized. There are several advantages of randomization. For example, randomization can eliminate selection bias during assignment of patients to each treatment groups. Different randomization procedures are used in clinical trials. In general, we have three types of randomization, complete randomization, restricted randomization and adaptive randomization.

Complete randomization is a randomization method that assigns patients to each treatment group with equal probability. Suppose a clinical trial is studied with two

treatments 1 and 2. Let  $N$  be the total number of patients involved and a binary variable  $I_i$  indicating the treatment of the  $i$ th subject, where  $I_i = 1$  for treatment 1 and  $I_i = 0$  for treatment 2. Under these notations, the assignment  $I_i$ s of complete randomization are independent with  $\Pr(I_i = 1) = 1/2$  for  $i = 1, 2, \dots, N$ . Complete randomization is easy to implement, however, it is likely to lead to severe imbalance between different treatment groups, which limits its application in real clinical trials.

In clinical trials, balanced allocation is often desired so that any difference observed between two groups of patients can be attributed to treatment effect. If no covariates or prognostic factors are taken into consideration, balanced allocation can be achieved by restricted randomization. Restricted randomization imposes constraints on treatment assignments to obtain approximately equal number of patients in different treatment groups. Two popular restricted randomization methods are permuted block design and biased coin design proposed by Efron (1971). For permuted block design, with two treatments 1 and 2, the assignments of patients are conducted within a sequence of blocks with block size equal to  $2m$  ( $m \in \mathbb{Z}^+$ ). Every block consists of  $2m$  patients and the assignments are determined by a permutation of  $m$  A's and  $m$  B's. The perfect balance is achieved within each full block and imbalance may only happen if the last block in a clinical trial is not filled. Biased coin design was also proposed to achieve balance between the number of patients in two treatment. Let  $D_i$  is the difference between the number of patients assigned to treatment 1 and the number of patients assigned to treatment 2 after  $i$  subjects have been enrolled in a clinical trial, then the probability of the  $(i + 1)$ th subject



assigned to treatment 1 is,

$$\Pr(I_{(i+1)} = 1) = \begin{cases} p, & \text{if } D_i < 0 \\ 1 - p, & \text{if } D_i > 0 \\ 1/2, & \text{if } D_i = 0 \end{cases}$$

where  $1/2 < p < 1$  is the biased coin probability.

Besides complete randomization and restricted randomization, a broad family of randomization methods, adaptive randomization, are widely used in clinical trials. Since a clinical trial is a sequential process, one can take use of information from all previous patients already enrolled in the study when a new patient enters the trial. Because of the efficiency and flexibility of adaptive designs, they have gained more and more popularity both in literature and in practice in last few decades. An brief introduction to adaptive designs will be given in Section 1.1.2.

### 1.1.2 Adaptive design

According to Gallo, *et al.* (2006), an adaptive design is referred to “a clinical study design that uses accumulating data to decide how to modify aspects of the study as it continues, without undermining the validity and integrity of the trial”. Furthermore, they pointed out that “the goal of adaptive designs is to learn from the accumulating data and to apply what is learned as quickly as possible”. Based on the kinds of data and adaption used, there are various adaptive designs. One type of adaptive design sequentially changes assignment probabilities of subjects in order to achieve balance with respect to treatments and covariates and/or to increase efficiency and ethics. We mainly focus on this type of adaptive designs in this section. Other adaptive designs include adaptive dose finding, sample size re-estimation, seamless

Phase II/III designs and so on (Gallo, *et al.*, 2006).

Adaptive designs can employ accumulating information to adaptively change consequent treatment assignments. Several kinds of historical information can be obtained throughout the process of a clinical trial, including previous assignments of patients, previous responses of patients, and covariate information of previous patients and the current patient. Depending on the kinds of history information used for adaption, there are at least three types of adaptive designs, which are response-adaptive design, covariate-adaptive design and covariate-adjusted response-adaptive design (CARA) (Hu and Rosenberger, 2006).

*Response-adaptive designs* utilize response information from previous patients in the study to modify the allocation probabilities of consequent patients. The objectives of response-adaptive designs include enhancing ethics by assigning more patients to the superior treatment arm shown by all information so far for the ongoing trial, maximizing power to detect treatment difference, and achieving multiple optimal criteria simultaneously (Hu and Rosenberger, 2003). In particular, two types of response-response adaptive designs are popular in literature, urn model based design and doubly-adaptive biased coin design (DBCD). Urn models are extensively used in response-adaptive designs, which are originally developed from an ethical point of view, i.e., to assign more patients to the better treatment. Two typical urn-model-based methods are “randomized play-the-winner” model (Wei and Durham, 1978) and “drop-the-loser” model (Ivanova, 2003). For DBCD, the allocation probability of the next patient is determined by an allocation function with arguments of the current allocation proportion and the estimate of target allocation proportion. By doing this, the current allocation is forced to go towards the estimated target proportion that is usually derived based on a certain optimality criterion. DBCD was firstly proposed by Eisele (1994) and Eisele and Woodroffe (1995). Hu and

Zhang (2004) proposed a general family of allocation functions and established the asymptotic properties of DBCD based on widely satisfied conditions. More discussion about DBCD and optimal allocation proportion can be found in Melfi and Page (2000), Melfi, Page and Geraldles (2001), Hu and Rosenberger (2006), Tymofyeyev, Rosenberger and Hu (2007) and references therein.

*Covariate-adaptive designs* are randomization procedures that incorporate covariates of patients to balance treatment allocation over covariates. In clinical trials, covariates are referred as factors associated with outcomes of patients. Some typical covariates include gender, age, clinical center, blood pressure, stage of disease at the baseline and even gene expressions in some biomarker studies. During the process of a covariate-adaptive design, the assignment of the current patient depends on covariate information of all previous patients and the current patient's covariate information as well as the history of previous treatment assignments. Several covariate-adaptive designs are introduced in Section 1.1.3 in detail.

*Covariate-adjusted response-adaptive design* is proposed by Zhang, *et al.* (2007) and is a relatively new topic in literature. Instead of assuming homogeneity for patients in a treatment, CARA allows heterogeneity of patients based on their characteristics even though they belong to the same treatment group. Taking both covariate and response information into consideration, patients can be assigned to better treatment based on their own covariate profiles.

### 1.1.3 Covariate-adaptive design

In clinical trials, it is usually important to balance treatment arms with respect to key covariates. There are several advantages to apply covariate-adaptive randomization to clinical trials. As pointed out by Kundt (2009), a balanced trial can

give more efficient comparison of treatment effects from statistical point of view and deliver more convincing results to a general audience. In addition, achieving balance is also desirable in other situations, including trials with interim analysis, trials that terminate early and trials where the subgroup analysis is considered important (McEntegart, 2003; Toorawa *et al.*, 2009).

A natural idea to achieve balance over covariates is stratification. Strata are defined as different combinations of covariates' levels. To get balanced trial, we could apply separate restricted randomization within each stratum to obtain good balance within each stratum and further to obtain overall balance. Depending what restricted randomization is used, we have stratified permuted block design using permuted block design within strata and covariate-adaptive biased coin design using Efron's biased coin design within strata. Stratified permuted block design is the most popular method to balance covariates and is used in most of clinical trials, such as Iacono *et al.* (2006), Jakob *et al.* (2012) and so on. However, it only works well for a trial with a few strata and large number of patients, otherwise a large portion of incomplete strata would cause imbalance on stratum level and further on the overall level (Pocock, 1982).

To deal with many covariates, several marginal methods (also referred as minimization, dynamic allocation in literature) were proposed. Taves (1974) proposed a minimization method to deal with large number of covariates, but his method didn't involve randomness. Pocock and Simon (1975) generalized Taves' method by incorporating randomness, which has been more popular thereafter. Instead of attempting to eliminate imbalance within each stratum, their method achieves balance by reducing weighted sum of marginal imbalances. A simpler version of Pocock and Simon's marginal procedure with two treatments can be described as follows. Suppose  $N_{ijk}(n)$  is the number of patients on treatment  $k$  in level  $j$  of covariate

$Z_i$ ,  $i = 1, \dots, I$ ,  $j = 1, \dots, m_i$ ,  $k = 1, 2$ , after  $n$  patients are enrolled in the study. Let  $(z_1, \dots, z_I)$  be the covariate information for the next patient. Then the marginal imbalance with respect to covariate  $i$  is defined as  $D_i(n) = N_{iz_i1}(n) - N_{iz_i2}(n)$ . The next assignment is based on the weighted average of all marginal imbalances  $D(n) = \sum_{i=1}^I w_i D_i(n)$  by using the biased coin allocation,

$$\Pr(I_{n+1} = 1) = \begin{cases} p, & \text{if } D(n) < 0 \\ 1 - p, & \text{if } D(n) > 0 \\ 1/2, & \text{if } D(n) = 0 \end{cases}$$

where  $1/2 < p < 1$  is the biased coin probability.

The usage of Pocock and Simon's marginal procedure had increased in last decades. According to Taves (2010), Pocock and Simon's marginal procedure was implemented in over 400 clinical trials from 1989 to 2008. Some recent examples include Anderson *et al.* (2000), Gridelli *et al.* (2003), Krueger *et al.* (2007), Molander *et al.* (2007), Ohtori *et al.* (2012) and so on. However, despite its broad applications, the theoretical properties of Pocock and Simon's marginal procedure remain unknown ever since it was proposed. Recently, Hu and Zhang (2013) theoretically proved that, under Pocock and Simon's marginal procedure, the marginal and overall imbalances are bounded in probability, while the within-stratum imbalance increase with the rate of  $\sqrt{n}$  as the sample size increases. These conclusions provide foundation for us to further study theoretical properties of statistical inference under these randomization methods.

Recently, Hu and Hu (2012) proposed a new family of covariate-adaptive procedures which simultaneously eliminate imbalances at three different levels, including overall imbalance, marginal imbalance and within-stratum imbalance. When a new

patient enters the trial and is read for randomization, the assignment will be based on the weighted average of the above three imbalances,  $D(n) = w_1 D_{stratum}(n) + w_2 D_{marginal}(n) + w_3 D_{overall}(n)$ , where  $D_{marginal}(n)$  is defined as in Pocock and Simon's marginal procedure,  $D_{stratum}(n)$  and  $D_{overall}(n)$  are within-stratum imbalance and overall imbalance, respectively. Theoretically, they proved that overall imbalance, marginal imbalance and within-stratum imbalance are all bounded in probability under the new covariate-adaptive designs.

## 1.2 Statistical inference for covariate-adaptive designs

Even though many covariate-adaptive designs have been proposed and implemented in clinical trials, the discussion of statistical inference associated with those methods is limited. In practice, conventional tests are often employed without consideration of covariate-adaptive randomization scheme. It remains a concern if conventional tests are still valid under covariate-adaptive designs. It is now generally accepted that covariates used in trial design should also be incorporated in inference procedures. Forsythe (1987) suggested all covariates used in minimization method should be included into analysis to achieve a valid test through simulation studies. Shao, Yu and Zhong (2010) theoretically pointed out that "one way to obtain a valid test procedure is to use a correct model between outcomes and covariates, including those used in randomization".

However, in practice, not all covariate information used in randomization can be fully utilized in inference procedures. In a clinical trial described in Anderson *et al.* (2000), Pocock and Simon's marginal procedure is implemented to balance alloca-

tion over three covariates including clinical centers, performance status and disease extent. A continuous primary endpoint is compared between two treatment groups using the two sample  $t$ -test, without adjusting covariate effects at all. In practice, some randomization covariates are omitted in final analysis due to: (i) it is difficult to incorporate some covariates in the analysis model, for example, investigation sites, etc.; (ii) adjusting too many covariates usually means more complicated modeling techniques; and (iii) it requires correct model specification, which is usually unknown in practice.

There have been doubts about validity of statistical inference for covariate-adaptive designs, especially when covariates are fully or partially omitted in inference procedures. Birkett (1985) and Forsythe (1987) had raised concerns about validity of unadjusted analysis under covariate-adaptive designs. They found that the two sample  $t$ -test is conservative in terms of small type I error if Taves' minimization is used to allocate patients to treatments through simulation studies. They also found that the two sample  $t$ -test is less powerful for minimization than complete randomization for small treatment difference, but more powerful if larger treatment difference exists. In Shao, Yu and Zhong (2010) some theoretical work are done to study conservativeness of the two sample  $t$ -test. The following linear model with outcomes  $Y_{ij}$  for patient  $i$  under treatment  $j$ ,  $j = 0, 1$ , is assumed for covariate-adaptive biased coin design,

$$Y_{ij} = \mu_j + bZ_i + \varepsilon_{ij}$$

where  $Z_i$  is a univariate covariate,  $Z_i$ s are independent and identically distributed,  $\mu_j$  and  $b$  are unknown parameters and  $\varepsilon_{ij}$ s are independent and identically distributed random errors and independent of  $Z_i$ s. They theoretically proved that the

two sample  $t$ -test is conservative by assuming responses follow the above simple homogeneous linear model. Moreover, a bootstrap test is proposed to adjust Type I error under covariate-adaptive biased coin design.

Most work mentioned above is in the linear model framework where outcomes of clinical trials are continuous. When responses are binary, Feinstein and Landis (1976) and Green and Byar (1978) studied statistical problems comparing successful rate between two treatment groups on a special case where there are two strata and two treatment considered. Under this restricted assumption, they showed that Type I error is smaller than the nominal level under stratified randomization. Gail, Wieand and Piantadosi (1984) studied estimates of treatment effect in randomized experiments with nonlinear regression and omitted covariates. Gail (1988) studied properties of the score test for perfectly balanced studies across strata (not randomization) on a large family of generalized linear models. The following underlying model between response and treatment is assumed,

$$E(Y|T, X = i) = h(\alpha T + \beta_i),$$

where  $Y$  is the response variable,  $\alpha$  is the treatment effect, and  $\beta_i$  is the stratum parameter for  $X = i$ . This covers a large range of generalized linear model. For example,  $h(\eta) = \exp(\eta)/[1 + \exp(\eta)]$  in logistic regression. The properties of the score test is studied under perfectly balanced studies with no covariates included in analysis, so the working model is

$$E(Y|T, X = i) = h(\mu + \alpha T).$$

The properties of Type I error for several kinds of generalized linear models are



given under the studies that are perfectly balanced. For example, Type I error is shown to be conservative for logistic regression. However, the properties of the score test and other tests, such as the Wald test and the likelihood ratio test, are unknown for general covariate-adaptive designs. More discussions about inference for covariate-adaptive designs can be found in Simon (1979), Tu, Shalay and Pater (2000), Aickin (2009) and so forth.

### **1.3 Motivation and organization of the dissertation**

In literature, the results of statistical inference for covariate-adaptive designs are restricted in several aspects. (1) Conclusions are mainly drawn by simulations, theoretical work is very limited. Shao, Yu and Zhong (2010) proved the property of the two sample  $t$ -test based on covariate-adaptive biased coin design, which is a stratified design and less commonly used in practice. (2) When outcomes of clinical trials are continuous, only the two sample  $t$ -test is discussed, where no covariate information is incorporated in final analysis. In practice, it is often that a subset of randomization covariates are used in final statistical inference procedures. The corresponding theoretical properties remain unknown. (3) All studies focus on hypothesis testing for comparing treatment effects. There is very little, if any, discussion about inference of covariates under covariate-adaptive designs in literature. In view of the importance of inference of covariates in clinical and medical studies, for example, in some personalized medicine and biomarker finding studies, we also want to study inference properties for covariates under covariate-adaptive clinical trials. (4) The discussion of statistical inference under generalized linear models

with balanced covariates are very limited. The results are either based on a very restricted setting (Green and Byar, 1978) or only focused on some special types of studies, for example, studies that are not even randomized (Gail, 1988). Since logistic regression and other generalized linear models are commonly used to analyze clinical trial data when outcomes are not continuous, it is of importance to study the inference properties of more advanced models under covariate-adaptive designs.

Over the past several decades, scientists have identified more and more biomarkers (Ashley, *et al.*, 2010; Li, *et al.*, 2010; Lipkin, *et al.*, 2010; etc.) that may link with certain diseases in the fields of translational research (genomics, proteomics, and metabolomics). Based on these biomarkers, we would like to develop personalized medicine that helps patients to receive better treatment regimens based on their individual characteristics (which could be biomarkers or other covariates). Balancing treatment allocation for influential covariates has become more and more important in today's clinical trials (Hu, 2012). Therefore it is essential to study the theoretical behavior of testing hypotheses of both treatment effects and covariates under covariate-adaptive randomized clinical trials.

In view of importance to study statistical inference for covariate-adaptive randomized clinical trials and limitations of relevant discussions in literature, my dissertation endeavors to establish a theoretical foundation of statistical inference under both linear models and generalized linear models. For a general family of covariate-adaptive designs, the large sample theory of testing treatment effects and significance of covariates are derived. Our work provides a general framework to further study inference properties for covariate-adaptive designs and to address the concerns of how to conduct valid statistical inference when covariates are balanced.

In Chapter 2, we first study the testing hypotheses based on linear models for covariate-adaptive designs. Two types of hypothesis testing, comparing treatment

effects and testing significance of covariates, are examined under general conditions of covariate-adaptive designs. In particular, theoretical properties of these testing hypotheses are given for the scenario covariate information is fully or partially omitted from analysis procedures. We show (i) the hypothesis testing to compare treatment effects is usually conservative in terms of small Type I error; (ii) the hypothesis testing to compare treatment effects is usually more powerful (than complete randomization); and (iii) the hypothesis testing for significance of covariates is still valid. Extensive simulations are conducted to illustrate Type I error and power. Other topics, such as adjusting methods, model selection and model mis-specifications, are also discussed.

In Chapter 3, statistical inference is also studied under linear model framework but with more general assumptions. The results in Chapter 2 are based on the assumption all covariates are independent of each other, which does not necessarily hold in practice despite its importance in theory. By incorporating correlation structure between covariates, similar inference problems as Chapter 2 are studied. We show that the hypothesis testing to compare treatment effects is also conservative in terms of small Type I error. In fact, the conclusion of testing treatment effects in Chapter 2 can be considered as a special case of this chapter. However, the estimators of covariate effects are biased and the hypothesis testing for significance of covariates is not valid anymore.

In Chapter 4, hypothesis testing is studied for logistic regression with omitted covariates under covariate-adaptive designs. We mainly focus on the properties of the Wald test to compare treatment effect. The asymptotic distribution of the test statistic under the null hypothesis is given for the covariate-adaptive designs with within-stratum imbalances bounded in probability, such as stratified permuted block design. Both Type I error and power are also illustrated via simulations.

## Chapter 2

# Statistical Inference for Linear Models with Independent Covariates

### 2.1 Introduction

Statistical inference of covariate-adaptive randomized clinical trials is discussed for linear models in this chapter. For a large family of covariate-adaptive designs, we derive asymptotic distributions of test statistics for testing treatment effects and significance of covariates under null and alternative hypotheses. We find that: (i) the hypothesis testing to compare treatment effects is usually conservative in terms of small Type I error; (ii) the hypothesis testing to compare treatment effects is usually more powerful (than complete randomization); and (iii) the hypothesis testing for significance of covariates is still valid.

This chapter is organized as following. In Section 2.2, the general framework based on linear models is given to study hypothesis testing properties for a large

class of covariate-adaptive designs. In particular, both an underlying model and a working model are proposed to represent the situation that covariates used in randomization are fully or partially omitted in statistical inference. Two types of hypothesis testing are studied, comparing treatment effects between two randomization groups and testing covariate effects. How to handle continuous covariates and definition of imbalances of different levels are also discussed. Furthermore, in Section 2.3, theoretical results are presented for the two types of hypothesis testing. We show that comparing treatment effects is conservative under covariate-adaptive designs, while covariate effects remain valid. Extensive simulations are carried out in Section 2.4 to study Type I error and power. Other practical issues, such as model selection and model mis-specification, are also covered and discussed. In Section 2.6, conclusions are given and some possible future work is mentioned.

## 2.2 Framework

In this section, we study hypothesis testing based on a linear model framework for covariate-adaptive designs. Suppose two treatments 1 and 2 are studied under a covariate-adaptive randomized clinical trial,  $\mu_1$  and  $\mu_2$  are parameters measuring the main effects of treatment 1 and 2, respectively. Let  $N$  be the total number of patients enrolled in the study. Let  $I_i$  be the assignment of the  $i$ th patient, i.e.,  $I_i = 1$  for treatment 1 and  $I_i = 0$  for treatment 2,  $i = 1, 2, \dots, N$ . The following linear model is assumed for the response of the  $i$ th patient  $Y_i$ ,

$$Y_i = \mu_1 I_i + \mu_2 (1 - I_i) + \beta_1 X_{i,1} + \dots + \beta_p X_{i,p} + \gamma_1 Z_{i,1} + \dots + \gamma_q Z_{i,q} + \varepsilon_i, \quad (2.1)$$

where

- $X_{i,k}$ s and  $Z_{i,j}$ s are discrete or continuous covariates which are independent and identically distributed as  $X_k$  and  $Z_j$ ,  $k = 1, \dots, p$  and  $j = 1, \dots, q$ ;
- both  $X_{i,k}$ s and  $Z_{i,j}$ s are used in the randomization procedure, but only  $X_{i,k}$ s are used in final statistical inference,  $k = 1, \dots, p$  and  $j = 1, \dots, q$ ;
- all covariates are independent of each other, and  $\mathbf{E}X_k = 0$  and  $\mathbf{E}Z_j = 0$  for all  $k$  and  $j$ ,  $k = 1, \dots, p$  and  $j = 1, \dots, q$ ;
- $\varepsilon_i$ s are independent and identically distributed random errors with mean zero and variance  $\sigma_\varepsilon^2$  and independent of  $X_k$  and  $Z_j$ ,  $k = 1, \dots, p$  and  $j = 1, \dots, q$ .

Notice both  $X_{i,k}$  and  $Z_{i,j}$  are assumed to be scalars in model (2.1). If  $X_{i,k}$  (or  $Z_{i,j}$ ) is a discrete covariate,  $X_{i,k}$  (or  $Z_{i,j}$ ) is a scalar that can take several values corresponding to different categories. In practice, a vector is usually used to represent a discrete covariate with multiple categories. For example, a covariate that has 3 categories can be coded as a two-dimensional vector with values of  $(0, 0)$ ,  $(0, 1)$  and  $(0, 1)$ . In this chapter,  $X_{i,k}$  (or  $Z_{i,j}$ ) is assumed to be a scalar for simplicity, but all the results can be extended to the situation where discrete covariates with multiple categories are represented by vectors.

We define vectors  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^\top$ ,  $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)^\top$ ,  $\boldsymbol{\beta} = (\mu_1, \mu_2, \beta_1, \dots, \beta_p)^\top$ ,  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)^\top$ , and matrices

$$\mathbf{X} = \begin{bmatrix} I_1 & 1 - I_1 & X_{1,1} & \cdots & X_{1,p} \\ I_2 & 1 - I_2 & X_{2,1} & \cdots & X_{2,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ I_N & 1 - I_N & X_{N,1} & \cdots & X_{N,p} \end{bmatrix} \quad \text{and} \quad \mathbf{Z} = \begin{bmatrix} Z_{1,1} & \cdots & Z_{1,q} \\ \vdots & \ddots & \vdots \\ Z_{N,1} & \cdots & Z_{N,q} \end{bmatrix}.$$

With the matrix representation, the underlying model (2.1) can be written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}.$$

In this chapter, we study statistical inference for covariate-adaptive designs if only partial randomization covariates are implemented into the analysis step. The following working model is used to do statistical inference,

$$\mathbb{E}[Y_i] = \mu_1 I_i + \mu_2(1 - I_i) + \beta_1 X_{i,1} + \dots + \beta_p X_{i,p}. \quad (2.2)$$

or in the matrix form,

$$\mathbb{E}[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}.$$

Under the working model (2.2), the ordinary least squares (OLS) method is used to obtain the estimator of  $\boldsymbol{\beta}$ , which has the explicit form,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}).$$

When model (2.2) is constructed to study patients' data from a covariate-adaptive randomized clinical trial, the primary interest is usually to compare treatment effects between different groups. The following hypothesis testing is used to compare treatment effects of  $\mu_1$  and  $\mu_2$ .

$$H_0 : \mu_1 - \mu_2 = 0 \text{ versus } H_A : \mu_1 - \mu_2 \neq 0. \quad (2.3)$$

The test statistic for the above hypothesis testing (2.3) is

$$T = \frac{\mathbf{L}\hat{\boldsymbol{\beta}}}{(\hat{\sigma}^2 \mathbf{L}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{L}^\top)^{1/2}}, \quad (2.4)$$

where  $\mathbf{L} = (1, -1, 0, \dots, 0)$  and  $\hat{\sigma}^2 = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) / (N - p - 2)$ . If  $|T| > Z_{1-\alpha/2}$ , where  $Z_{1-\alpha/2}$  is  $(1 - \alpha/2)$  quantile of a standard normal distribution, we will reject the null hypothesis, otherwise accept the null hypothesis.

On the other hand, we consider general forms of hypothesis testing for significance of covariates. Let  $\mathbf{C}$  be an  $m \times (p + 2)$  matrix of rank  $m$  with  $m \leq p$ , where entries of the first two columns of  $\mathbf{C}$  are all zeros so that  $\mathbf{C}\boldsymbol{\beta}$  doesn't include any treatment effects. Then the hypothesis testing of interest is,

$$H_0 : \mathbf{C}\boldsymbol{\beta} = \boldsymbol{\xi}_0 \text{ versus } H_A : \mathbf{C}\boldsymbol{\beta} = \boldsymbol{\xi}_1 \neq \boldsymbol{\xi}_0. \quad (2.5)$$

The test statistic for hypothesis testing (2.5) is,

$$T^* = \frac{m^{-1}(\mathbf{C}\boldsymbol{\beta} - \boldsymbol{\xi}_0)^\top [\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top]^{-1} (\mathbf{C}\boldsymbol{\beta} - \boldsymbol{\xi}_0)}{\hat{\sigma}^2}. \quad (2.6)$$

If  $T^* > \chi_{m,(1-\alpha)}^2 / m$ , where  $\chi_{m,(1-\alpha)}^2$  is  $(1 - \alpha)$  quantile of a  $\chi^2$  distribution with degree of freedom  $m$ , we will reject the null hypothesis, otherwise accept the null hypothesis.

A special case of testing (2.5) is evaluating significance of a single covariate (biomarker). This is usually important in personalized medicine (Hu, 2012). Without loss of generality, we consider the hypothesis testing for  $\beta_1$ , the coefficient of  $X_1$ . To test the significance of  $\beta_1$ , the hypothesis is

$$H_0 : \beta_1 = 0 \text{ versus } H_A : \beta_1 \neq 0. \quad (2.7)$$



The test statistic for hypothesis testing (2.7) can be reduced to,

$$T_1 = \frac{\boldsymbol{\ell}\hat{\boldsymbol{\beta}}}{(\hat{\sigma}^2\boldsymbol{\ell}(\mathbf{X}^\top\mathbf{X})^{-1}\boldsymbol{\ell}^\top)^{1/2}}, \quad (2.8)$$

where  $\boldsymbol{\ell} = (0, 0, 1, 0, \dots, 0)$ . If  $|T_1| > Z_{1-\alpha/2}$ , where  $Z_{1-\alpha/2}$  is  $(1 - \alpha/2)$  quantile of a standard normal distribution, we will reject the null hypothesis, otherwise accept the null hypothesis.

In clinical trials, covariate-adaptive designs are usually based on discrete covariates (Taves, 2010). If a continuous covariate is to be used in randomization, a continuous-discrete conversion need be performed to breakdown the continuous covariate into a discrete variable with several subcategories. Let  $C = \{j \mid Z_j \text{ is continuous, } j = 1, \dots, q\}$  and  $C^* = \{k \mid X_k \text{ is continuous, } k = 1, \dots, p\}$ . If  $k \in C^*$  or  $j \in C$ , the covariate-adaptive design is applied with respect to discrete variables,  $d_k^*(X_k)$  or  $d_j(Z_j)$ , where  $d_k^*$ ,  $d_j$  are discrete functions. In such case, define  $\delta_{i,k}^* = X_{i,k} - \mathbf{E}[X_{i,k}|d_k^*(X_{i,k})]$  and  $\delta_{i,j} = Z_{i,j} - \mathbf{E}[Z_{i,j}|d_j(Z_{i,j})]$ . Let

$$\tilde{X}_k = \begin{cases} X_k, & \text{if } k \notin C^* \\ d_k^*(X_k), & \text{if } k \in C^* \end{cases}$$

and

$$\tilde{Z}_j = \begin{cases} Z_j, & \text{if } j \notin C \\ d_j(Z_j), & \text{if } j \in C \end{cases}.$$

Here,  $\tilde{X}_{i,k}$  and  $\tilde{Z}_{i,j}$  are  $i$ th observations of covariates  $\tilde{X}_k$  and  $\tilde{Z}_j$ ,  $k = 1, \dots, p$  and  $j = 1, \dots, q$ .  $\tilde{X}_{i,k}$  and  $\tilde{Z}_{i,j}$  are used in the covariate-adaptive randomization process. We further define three levels of imbalance between patients in two treatments.

Consider  $\tilde{X}_k$  have  $s_k^*$  levels and  $\tilde{Z}_j$  have  $s_j$  levels, resulting in  $\prod_{k=1}^p s_k^* \prod_{j=1}^q s_j$  strata in total. Let  $W_i = (\tilde{X}_{i,1}, \dots, \tilde{X}_{i,p}, \tilde{Z}_{i,1}, \dots, \tilde{Z}_{i,q})$  represents the covariate profile of the  $i$ th patient used in randomization, i.e.,  $W_i = (x_1^{t_1}, x_2^{t_2}, \dots, x_p^{t_p}, z_1^{r_1}, z_2^{r_2}, \dots, z_q^{r_q})$  if  $\tilde{X}_{i,k}$  is at level  $x_k^{t_k}$  and  $\tilde{Z}_{i,j}$  is at level  $z_j^{r_j}$ . For convenience, we use  $(t_1, t_2, \dots, t_p, r_1, r_2, \dots, r_q)$  to denote the stratum formed by patients who have the same covariate profile  $(x_1^{t_1}, x_2^{t_2}, \dots, x_p^{t_p}, z_1^{r_1}, z_2^{r_2}, \dots, z_q^{r_q})$ , and use  $(k; t_k)$  to denote the margin formed by patients with  $\tilde{X}_k = x_k^{t_k}$ , and similarly  $(j, r_j)$  to denote the margin formed by patients with  $\tilde{Z}_j = z_j^{r_j}$ . Then let

- $D_N$  be the difference between the numbers of patients in treatment group 1 and 2 as total, i.e., the number in group 1 minus the number in group 2;
- $D_N(k; t_k)$  and  $D_N(j; r_j)$  be the differences between the numbers of patients in the two treatment groups on the margin  $(k; t_k)$  and  $(j, r_j)$ , respectively;
- $D_N(t_1, t_2, \dots, t_p, r_1, r_2, \dots, r_q)$  be the difference between the numbers of patients in the two treatment groups within the stratum  $(t_1, t_2, \dots, t_p, r_1, r_2, \dots, r_q)$ .

These differences play important roles in properties of statistical inference for covariate-adaptive designs (see Section 2.3 for details).

## 2.3 Theoretical properties

Two types of hypothesis testing will be considered in this section, one is comparing main treatment effects between two groups and the other is testing significance of covariates. The testing hypotheses are conducted based on working model (2.2) when data are generated from the true model (2.1). Properties of those hypothesis testing are studied under both null hypothesis and alternative hypothesis. A test is

said to be (asymptotically) conservative, if its true Type I error is smaller than the significance level under the null hypothesis.

First, we consider the hypothesis tests for comparing treatment effects. We have the following main theorem.

**Theorem 2.3.1.** *Suppose that a covariate-adaptive design satisfies following two conditions:*

(A) *the overall imbalance is bounded, that is,  $D_N = O_p(1)$ ;*

(B) *the marginal imbalances for all covariates are bounded in probability, that is,*

$$D_N(k; t_k) = O_p(1) \text{ and } D_N(j; r_j) = O_p(1), \quad k = 1, 2, \dots, p, \quad j = 1, 2, \dots, q.$$

Then

(i) *under  $H_0 : \mu_1 - \mu_2 = 0$ ,*

$$T \xrightarrow{D} \mathbf{N}(0, \tau^2), \quad \tau^2 = \frac{\sigma_\varepsilon^2 + \sum_{j \in C} \gamma_j^2 \sigma_{\delta,j}^2}{\sigma_\varepsilon^2 + \sum_{j=1}^q \gamma_j^2 \text{Var}(Z_j)} = \frac{\sigma_\delta^2}{\sigma_z^2}. \quad (2.9)$$

where  $\sigma_z^2 = \sigma_\varepsilon^2 + \sum_{j=1}^q \gamma_j^2 \text{Var}(Z_j)$ ,  $\sigma_\delta^2 = \sigma_\varepsilon^2 + \sum_{j \in C} \gamma_j^2 \sigma_{\delta,j}^2$  and  $\sigma_{\delta,j}^2 = \mathbf{E}[\text{Var}(\delta_{i,j} | d_j(Z_{i,j}))]$ .

Hence,

(1) *If  $\gamma_j = 0$ ,  $j = 1, 2, \dots, q$ , then  $\tau = 1$ . Thus, when all covariates  $Z$ s are not related to  $Y$ , the hypothesis testing (3) can achieve correct Type I error.*

(2) *If at least one  $\gamma_j \neq 0$ ,  $j = 1, 2, \dots, q$ , then  $\tau < 1$ . In this case, the hypothesis testing (3) is conservative.*

(ii) *under  $H_A : \mu_1 - \mu_2 \neq 0$ , consider a sequence of local alternatives, i.e.,  $\mu_1 - \mu_2 = \delta/\sqrt{N}$  for a fixed  $\delta \neq 0$ , then*

$$T \xrightarrow{D} \mathbf{N}(\Delta, \tau^2), \quad \Delta = \frac{\delta}{2\sigma_z}. \quad (2.10)$$

*Hence, the power increases as more covariates are incorporated into model.*

In Theorem 2.3.1, theoretical properties of hypothesis testing for treatment effects are obtained under covariate-adaptive designs. In a covariate-adaptive randomized clinical trial, covariate information is incorporated in the design process to reduce imbalance of different levels (within-stratum, within-covariate-margin, and overall). Two mild conditions of covariate-adaptive designs are assumed to derive the asymptotic distribution of the test statistic for comparing treatment effects. Condition (A) states that the overall imbalance is bounded in probability and condition (B) requires that all marginal imbalances are all bounded in probability. These conditions are satisfied by various covariate-adaptive designs (see Corollary 2.3.2 for examples). Under these conditions, it can be shown in the proof that numerator of the test statistic,  $\mathbf{L}\hat{\boldsymbol{\beta}} (= \hat{\mu}_1 - \hat{\mu}_2)$ , has a smaller variance than the model-based variance estimator in the denominator if important covariates are omitted from the working model. Based on the asymptotic distributions of the test statistic under both the null hypothesis and the alternative hypothesis, Type I error is smaller than the nominal level if at least one  $\gamma_j \neq 0$ , and power performance can be discussed as well.

Now consider the power of the hypothesis testing (2.2), under the alternative hypothesis, the power is

$$\Pr(|T| > Z_{1-\alpha/2}) = \Phi\left(\frac{\delta}{2\sigma_\delta} - \frac{\sigma_z Z_{1-\alpha/2}}{\sigma_\delta}\right) + \Phi\left(-\frac{\delta}{2\sigma_\delta} - \frac{\sigma_z Z_{1-\alpha/2}}{\sigma_\delta}\right) + o(1).$$

The power of test (2.2) under complete randomization (based on the same setting as described in Section 2.2) is

$$\Pr(|T| > Z_{1-\alpha/2}) = \Phi\left(\frac{\delta}{2\sigma_z} - Z_{1-\alpha/2}\right) + \Phi\left(-\frac{\delta}{2\sigma_z} - Z_{1-\alpha/2}\right) + o(1).$$

From the power expressions for both covariate-adaptive designs and complete randomization above, we can conclude that the limiting power under covariate-adaptive designs is smaller than that under complete randomization when  $\delta$  is relatively small, and it is larger than complete randomization when  $\delta$  is large. This conclusion agrees with some simulation studies about two sample  $t$ -test in literature (Forsythe, 1987; Shao, Yu and Zhong, 2010) for certain covariate-adaptive designs. Our conclusion is more general, which can be applied to linear models under a large family of covariate-adaptive designs.

The following theorem shows that hypothesis tests regarding significance of covariates can still catch correct Type I error in covariate-adaptive designs, even though the power would be affected if not all covariates are incorporated in the analysis model.

**Theorem 2.3.2.** *Under the same conditions as in Theorem 2.3.1,*

(i) *under  $H_0 : \mathbf{C}\boldsymbol{\beta} = \boldsymbol{\xi}_0$ ,*

$$T^* \xrightarrow{D} \chi_{(m)}^2/m. \tag{2.11}$$

*Hence, the hypothesis testing (2.5) can achieve correct Type I error.*

(ii) *under  $H_A : \mathbf{C}\boldsymbol{\beta} = \boldsymbol{\xi}_1$ , consider a sequence of local alternatives, i.e.,  $(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_0) =$*

$\boldsymbol{\eta}/\sqrt{N}$  for a fixed  $\boldsymbol{\eta} \neq \mathbf{0}$ , then

$$T^* \xrightarrow{D} \chi_{(m)}^2(\lambda)/m, \quad \lambda = \boldsymbol{\eta}^\top [\mathbf{C}\mathbf{M}^{-1}\mathbf{C}^\top]^{-1}\boldsymbol{\eta}/\sigma_z^2, \quad (2.12)$$

where  $\mathbf{M} = \mathbf{diag}(1/2, 1/2, \text{Var}(X_1), \dots, \text{Var}(X_p))$  and  $\lambda$  is the noncentral parameter. Therefore, the power increases as more covariates are incorporated into model.

Theorem 2.3.1 and Theorem 2.3.2 imply that the overall difference and marginal imbalances play important roles in statistical inference for covariate-adaptive designs. For stratified permuted block design, the difference between the number of patients in two treatments within any stratum is the half of block size at maximum. Since the number of strata is finite for any covariate-adaptive design, the overall and marginal imbalance are less than a constant, thus the conditions (A) and (B) are satisfied. The theoretical properties for Pocock and Simon's marginal procedure remains unknown for decades and recently are derived by Hu and Zhang (2013). In their paper, the authors demonstrate the marginal imbalances and overall imbalance are bounded in probability for Pocock and Simon's marginal procedure, thus the conditions (A) and (B) are also satisfied. Furthermore, Hu and Hu (2012) proposed a large class of covariate-adaptive designs, which satisfy the conditions (A) and (B). Here we summarize these results in the following corollary.

**Corollary 2.3.1.** *Both Theorem 2.3.1 and Theorem 2.3.2 hold under the following covariate-adaptive designs:*

- (i) *Pocock and Simon's marginal procedures (Pocock and Simon, 1975);*
- (ii) *stratified permuted block designs; and*
- (iii) *the class of covariate-adaptive designs proposed by Hu and Hu (2012).*

REMARK 2.3.1. *If we consider complete randomization as a special case of covariate-adaptive design, it does not satisfy the conditions in Theorem 2.3.1, because the marginal imbalance  $D_N(j; r_j) = O_p(N^{1/2})$  for complete randomization. The numerical study in the next section shows that the test of treatment effect under complete randomization is not conservative.*

Based on Theorem 2.3.2, we can see that hypothesis testing of covariates is still valid in the sense of Type I error under covariate-adaptive designs. A linear regression model can be directly used to test significance of prognostic factors with a working model only containing partial covariate information. On the other hand, however, the power may be harmed by omitting important covariates in the working model. Consider the non-central parameter in (2.12), it increases with  $\sigma_z^2$  reduced, so the power increases with fewer important covariates omitted from the model. Therefore, it is helpful of incorporating more important covariates, if possible, to obtain a more powerful test.

Corollary 2.3.2 gives an important special case of testing covariates, where only a single coefficient is considered.

**Corollary 2.3.2.** *Under the same conditions as in Theorem 2.3.1,*

(i) *under  $H_0 : \beta_1 = 0$ ,*

$$T_1 \xrightarrow{D} \mathbf{N}(0, 1). \quad (2.13)$$

*Hence, the hypothesis testing (2.7) can achieve correct Type I error.*

(ii) *under  $H_A : \beta_1 \neq 0$ , consider a sequence of local alternatives, i.e.,  $\beta_1 = \delta_{\beta_1}/\sqrt{N}$*

for a fixed  $\delta_{\beta_1}$ , then

$$T_1 \xrightarrow{D} \mathbf{N}(\Delta_{\beta_1}, 1), \quad \Delta_{\beta_1} = \frac{\delta_{\beta_1} \sigma_1}{\sigma_z},$$

where  $\sigma_1^2 = \text{Var}(X_1)$ . Hence, the power increases as more covariates are incorporated into model.

According to Theorem 2.3.1 and Theorem 2.3.2, a model with only influential covariates can achieve valid tests. It is known that too many unnecessary variables in the model will increase variations of estimators and affect statistical results. Hence, if only influential variables are incorporated in the model, it will not only reduce unnecessary variations, but also give valid inference. Some numerical studies are performed in Section 2.4 about model selections.

To prove Theorem 2.3.1, the asymptotic properties of both the numerator and the denominator of test statistic  $T$  are studied. In the Appendix, we show that, under the conditions in Theorem 2.3.1,

$$\sqrt{N}[\mathbf{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})] = \sqrt{N}[(\hat{\mu}_1 - \hat{\mu}_2) - (\mu_1 - \mu_2)] \xrightarrow{D} \mathbf{N}(0, 4(\sum_{j \in C} \gamma_j^2 \sigma_{\delta,j}^2 + \sigma_\varepsilon^2)), \quad (2.14)$$

where  $\sigma_{\delta,j}^2 = \mathbf{E}[\text{Var}(\delta_{i,j}|d_j(Z_{i,j}))]$ ,  $j \in C$ .

On the other hand, for the denominator part of  $T$ ,

$$\hat{\sigma}^2 \mathbf{L}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{L}^\top = \frac{4}{N}(\sigma_\varepsilon^2 + \sum_{j=1}^q \gamma_j^2 \text{Var}(Z_j)) + o_p\left(\frac{1}{N}\right). \quad (2.15)$$

Combining the results of (2.14) and (2.15), the asymptotic distributions (2.9) and (2.10) can be obtained. Further notice  $\gamma_j^2 \text{Var}(Z_j) > \gamma_j^2 \mathbf{E}[\text{Var}(\delta_{i,j}|d_j(Z_{i,j}))]$  if  $\gamma_j \neq 0$



and  $j \in C$ , we have

$$\Pr(|T| > Z_{1-\alpha/2}) \rightarrow 2\Phi\left(-\frac{\sigma_z Z_{1-\alpha/2}}{\sigma_\delta}\right) < \alpha \quad (2.16)$$

where  $\Phi$  is the cumulative distribution function of a standard normal distribution, which proves the conservativeness of the test if not all  $\gamma_j$ s are zeros. Detailed proofs of Theorem 2.3.1 and Theorem 2.3.2 are in the Appendix.

## 2.4 Simulation study

*Case 1: Testing Treatment Effects.* First, we consider simulations to study Type I error of the hypothesis testing for comparing treatment effects under three designs: Pocock and Simon's marginal procedure, stratified permuted block design and complete randomization. For each type of design, both continuous case and discrete case are considered. The following linear model including two covariates  $Z_1$  and  $Z_2$  is assumed for responses  $Y_i$ ,

$$Y_i = \mu_1 I_i + \mu_2 (1 - I_i) + \beta_1 Z_{i,1} + \beta_2 Z_{i,2} + \varepsilon_i,$$

where  $\varepsilon_i$  is distributed as  $\mathbf{N}(0, 1)$ ,  $\beta_1 = \beta_2 = 1$ . No difference in treatment effects is assumed to study Type I error, i.e.,  $\mu_1 = \mu_2$ . For the discrete case,  $Z_1$  follows Bernoulli( $p_1$ ) and  $Z_2$  follows Bernoulli( $p_2$ ); for the continuous case, both  $Z_1$  and  $Z_2$  follow normal distributions  $\mathbf{N}(0, 1)$ . If covariates  $Z_1$  and  $Z_2$  are continuous, they are discretized into bernoulli variables  $Z'_1$  and  $Z'_2$  with the probabilities  $p_1$  and  $p_2$  in order to be used in randomization. More specifically, if  $Z_1 < Z_{(p_1)}$ , where  $Z_{(p_1)}$  is  $p_1$  quantile of a standard normal distribution, then  $Z'_1 = 0$ , otherwise  $Z'_1 = 1$ . Original variables (without discretization) are used in statistical inference procedures.

To carry out simulations, the biased coin probability 0.75 and equal weights are used for Pocock and Simon's marginal procedure, and the block size 4 is used for stratified permuted block design. The significance level is  $\alpha = 0.05$  and sample size  $N$  is 100, 200 or 500. The hypothesis tests include the two sample  $t$ -test ( $t$ -test), the linear model with a single covariate  $Z_1$  ( $lm(Z_1)$ ), the linear model with a single covariate  $Z_2$  ( $lm(Z_2)$ ) and the linear model with both covariate  $Z_1$  and  $Z_2$  ( $lm(Z_1, Z_2)$ ). By choosing  $(p_1, p_2) = (0.5, 0.5)$ , the simulation results for Pocock and Simon's marginal procedure, stratified permuted block design and complete randomization are demonstrated in Table 2.1.

In each simulation, Type I error of covariate-adaptive randomization methods is also examined with the bootstrap  $t$ -test described in Shao, Yu and Zhong (2010). To do the test,  $B$  bootstrap samples  $(Y_1^{*b}, Z_{1,1}^{*b}, Z_{1,2}^{*b}), \dots, (Y_N^{*b}, Z_{N,1}^{*b}, Z_{N,2}^{*b})$ ,  $b = 1, 2, \dots, B$ , are generated independently as simple random samples with replacement from  $(Y_1, Z_{1,1}, Z_{1,2}), \dots, (Y_N, Z_{N,1}, Z_{N,2})$ . The covariate-adaptive procedure on the original data is applied on the covariates of each bootstrap sample  $(Z_{1,1}^{*b}, Z_{1,2}^{*b}), \dots, (Z_{N,1}^{*b}, Z_{N,2}^{*b})$ , from which the bootstrap analogues of treatment assignments,  $I_1^{*b}, \dots, I_N^{*b}$ , can be obtained. Define

$$\bar{Y}_1 - \bar{Y}_2 = \frac{1}{n_1} \sum_{i=1}^N I_i Y_i - \frac{1}{n_2} \sum_{i=1}^N (1 - I_i) Y_i, \quad n_1 = \sum_{i=1}^N I_i, \quad n_2 = \sum_{i=1}^N (1 - I_i),$$

and

$$\hat{\theta}^*(b) = \frac{1}{n_1^{*b}} \sum_{i=1}^N I_i^{*b} Y_i^{*b} - \frac{1}{n_2^{*b}} \sum_{i=1}^N (1 - I_i^{*b}) Y_i^{*b}, \quad n_1^{*b} = \sum_{i=1}^N I_i^{*b}, \quad n_2^{*b} = \sum_{i=1}^N (1 - I_i^{*b}).$$

The bootstrap estimator of the variance of  $\bar{Y}_1 - \bar{Y}_2$  is then the sample variance of  $\hat{\theta}^*(b)$ ,  $b = 1, 2, \dots, B$ , represented by  $v_B$ . Then the bootstrap  $t$ -test has the form of

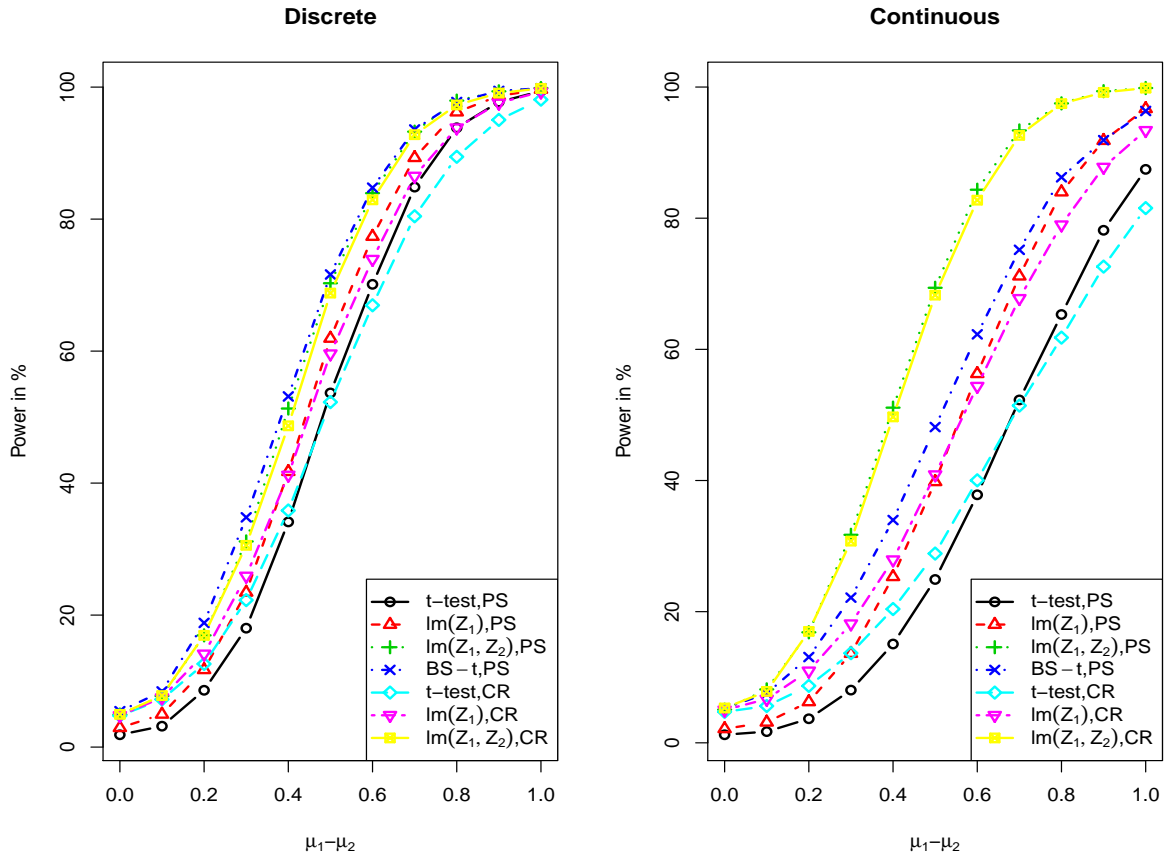
Table 2.1: Simulated Type I error for Pocock and Simon’s marginal procedure (PS), stratified permuted block design (SPB) and complete randomization (CR) in %. Simulations based on 10000 runs.

$Z$	Method	$N$	$t$ -test	$lm(Z_1)$	$lm(Z_2)$	$lm(Z_1, Z_2)$	$BS$ - $t$
Discrete	<b>PS</b>	100	1.75	3.05	3.09	5.21	5.18
		200	1.62	2.78	2.86	4.99	4.88
		500	1.66	2.81	2.77	4.87	4.90
	<b>SPB</b>	100	1.85	2.86	3.05	5.29	5.67
		200	1.54	2.69	2.73	4.84	4.95
		500	1.55	2.77	2.65	4.84	5.60
	<b>CR</b>	100	5.04	5.27	5.11	5.31	-
		200	5.00	4.95	5.12	5.21	-
		500	4.73	4.83	4.68	4.77	-
Continuous	<b>PS</b>	100	1.43	2.15	2.02	4.98	5.16
		200	1.07	1.74	1.80	4.53	5.62
		500	0.91	1.72	1.73	4.72	4.79
	<b>SPB</b>	100	1.22	1.83	2.05	5.01	5.68
		200	0.98	1.86	1.77	5.08	5.19
		500	1.15	1.98	1.84	5.48	5.61
	<b>CR</b>	100	5.20	5.31	4.82	4.92	-
		200	5.06	5.14	4.85	5.46	-
		500	4.87	5.05	4.71	4.77	-

$T_B = (\bar{Y}_1 - \bar{Y}_2)/\hat{v}_B^{1/2}$ . In Shao, Yu and Zhong (2010), it is shown that the bootstrap  $t$ -test can maintain nominal Type I error under covariate-adaptive biased coin design.  $B = 500$  is used in all following simulations.

Several conclusions can be drawn from Table 2.1. First, the Type I error is close to 5% under the full model  $lm(Z_1, Z_2)$  for covariate-adaptive designs. This is consistent with the theoretical results in Section 2.3, when no randomization covariate is omitted to construct the final analysis model. Secondly, under both Pocock and Simon’s marginal procedure and stratified permuted block design, the two sample  $t$ -test,  $lm(Z_1)$  and  $lm(Z_2)$  are all conservative. Among these three tests, the two sample  $t$ -test is the most conservative one with the least Type I error.

Figure 2.1: Simulated power for Pocock and Simon’s marginal procedure (PS) and complete randomization (CR) under discrete and continuous cases. Simulation based on 10000 runs and number of patients  $N = 100$ .



Thirdly, the Type I error of the bootstrap  $t$ -test ( $BS-t$ ) is close to the nominal level 5% under both Pocock and Simon’s marginal procedure and stratified permuted block design. Furthermore, under complete randomization, the Type I error is close to 5% for all four tests. We also tried different  $(p_1, p_2)$ , similar results are obtained and are not shown here.

*Case 2: Power Comparison.* Now we compare power for different hypothesis testing methods under Pocock and Simon’s marginal procedure and complete randomization. The same model as in Case 1 is used, except that difference exists

between treatment effects  $\mu_1$  and  $\mu_2$ , i.e.,  $\mu_1 - \mu_2 \neq 0$ . Sample size  $N = 100$  and  $(p_1, p_2) = (0.5, 0.5)$  are used in simulations. Several hypothesis testing methods for treatment effects are compared under covariate-adaptive randomization. All the results of power are given in Figure 2.1, from which several conclusions can be made. The two sample  $t$ -test is less powerful than  $(lm(Z_1))$  and  $(lm(Z_2))$ , and all those three methods are less powerful than  $(lm(Z_1, Z_2))$  under Pocock and Simon's marginal procedure. The bootstrap  $t$ -test has similar power performance to the full model  $(lm(Z_1, Z_2))$  with correct model specification when covariates are discrete. However, the bootstrap  $t$ -test is less powerful than  $(lm(Z_1, Z_2))$  if covariates are continuous. Furthermore, the power of each test can also be compared between covariate-adaptive randomization and complete randomization. For example, the two sample  $t$ -test has smaller power under Pocock and Simon's marginal procedure than that under complete randomization when  $|\mu_1 - \mu_2|$  is relatively small due to conservativeness, but has larger power when  $|\mu_1 - \mu_2|$  becomes larger. We also tried stratified permuted block randomization (not reported here), all tests considered have similar power performance as Pocock and Simon's marginal procedure. Based on Table 2.1 and Figure 2.1, we can see that the theoretical properties in Theorem 2.3.1 hold for sample size around 100.

We further consider the power comparison for small sample sizes 32 and 64. The simulated power is reported in Table 2.2 under the same model as in Case 1. In Pocock and Simon's marginal procedure, the probability of biased coin assignment is 0.8 and equal weights are assigned on two covariates. The hypothesis testing is based on the model  $(lm(Z_1, Z_2))$  with both covariates  $Z_1$  and  $Z_2$ . Table 2.2 indicates the advantages of covariate-adaptive designs over complete randomization for small sample size.

*Case 3: Significance of Covariates.* Type I error of testing (2.7) for significance

Table 2.2: Power Comparison for Pocock and Simon’s marginal procedure and Complete Randomization. Simulation based on 10000 runs and sample size  $N = 32, 64$ .

$\mu_1 - \mu_0$	$N = 32$		$N = 64$	
	<b>CR</b>	<b>PS</b>	<b>CR</b>	<b>PS</b>
0.0	4.96	5.03	5.17	5.08
0.1	5.86	6.35	6.50	6.48
0.2	7.81	8.51	12.12	12.68
0.3	12.34	13.26	22.06	22.21
0.4	18.15	19.44	34.46	34.76
0.5	25.28	26.54	47.99	49.35
0.6	33.96	36.98	63.04	65.53
0.7	44.16	47.83	76.72	78.13
0.8	53.74	57.28	86.97	87.95
0.9	64.30	68.70	93.14	94.24
1.0	73.63	77.10	97.02	97.51

of covariates is studied under Pocock and Simon’s marginal procedure and complete randomization. As the same model in Case 1, we set  $\beta_1 = 0$ ,  $\beta_2 = 1$  and  $\mu_1 = \mu_2$ . To run simulations, the biased coin probability 0.75 and equal weights are used for Pocock and Simon’s marginal procedure. The significance level is  $\alpha = 0.05$  and sample size  $N$  is 100, 200 or 500. Hypothesis testing of the significance of  $\beta_1$  are conducted based on the two models with or without  $Z_2$ , i.e.,  $lm(Z_1, Z_2)$  and  $lm(Z_1)$ . The results of Type I error are given in Table 2.3, from which it can be seen that the tests of  $\beta_1$  are valid in terms of Type I error for both Pocock and Simon’s marginal procedure and complete randomization. Figure 2.2 reports the power of testing  $\beta_1$  under two randomization methods, which indicates that the model  $lm(Z_1, Z_2)$  with  $Z_2$  in the analysis is more powerful than  $lm(Z_1)$ . This agrees with Theorem 2.3.2.

*Case 4: Model Selection.* So far, all simulation results are based on linear models with up to two covariates. Now we consider a linear model with more than two covariates. In this situation, variable selection techniques can be applied to select a subset of influential covariates on outcomes to be used in inference procedures.

Table 2.3: Simulated Type I error for  $H_0 : \beta_1 = 0$  versus  $H_A : \beta_1 \neq 0$  for Pocock and Simon's marginal procedure (PS) and complete randomization (CR) in %. Simulations based on 10000 runs.

$Z$	$N$	$lm(Z_1)$		$lm(Z_1, Z_2)$	
		<b>PS</b>	<b>CR</b>	<b>PS</b>	<b>CR</b>
Discrete	100	4.96	4.98	4.93	4.88
	200	5.35	5.28	4.76	4.82
	500	5.55	5.55	5.22	5.20
Continuous	100	4.98	4.90	5.10	4.98
	200	5.14	5.14	5.05	5.18
	500	5.11	5.15	5.28	5.29

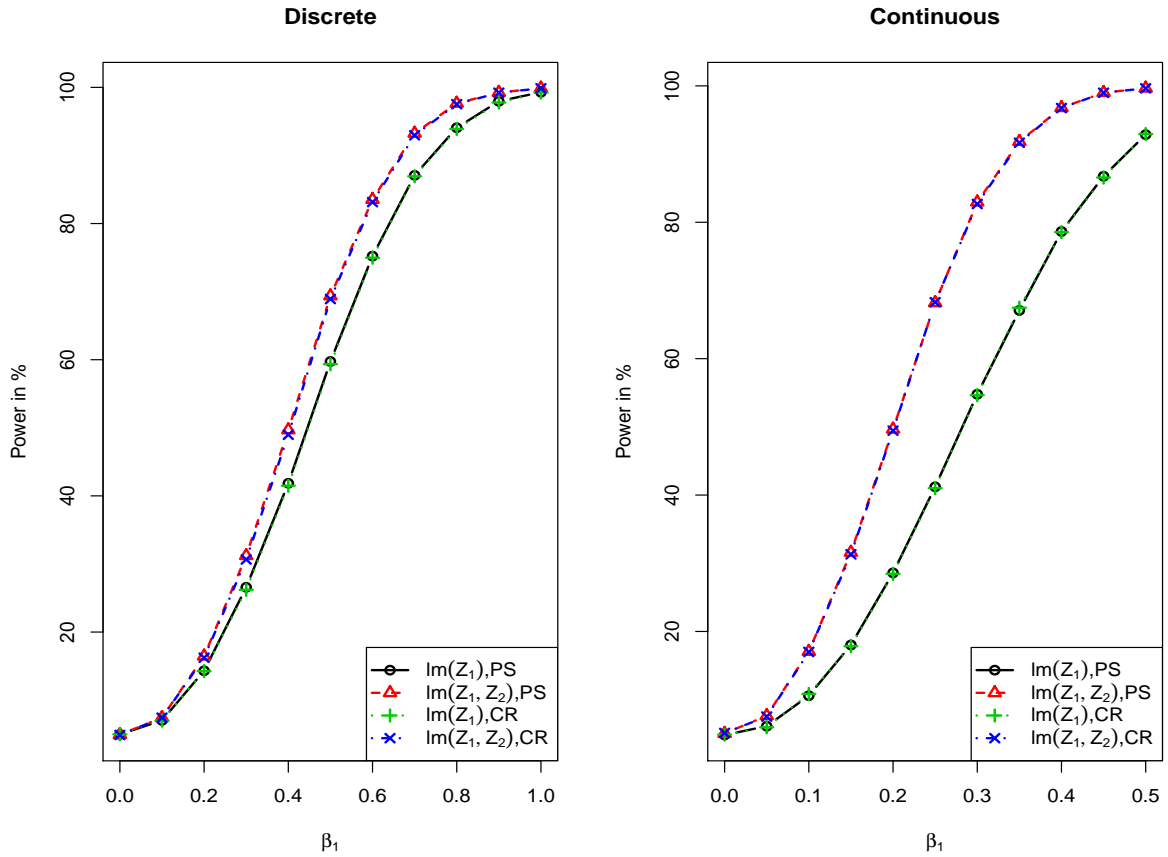
Suppose outcomes  $Y_i$  follows the following model with five covariates,

$$Y_i = \mu_1 I_i + \mu_2(1 - I_i) + X_{i,1}\beta_1 + X_{i,2}\beta_2^\top + X_{i,3}\beta_3 + Z_{i,1}\gamma_1^\top + Z_{i,2}\gamma_2 + \varepsilon_i, \quad (2.17)$$

where  $X_1$  is a binary covariate with the probability of 0.5 to take 0 or 1,  $X_2$  is a discrete variable with four possible values (0, 0, 0), (1, 0, 0), (0, 1, 0) and (0, 0, 1) for equal probabilities 0.25,  $Z_1$  is a discrete covariate with three possible values (0, 0), (0, 1) and (1, 0) for equal probabilities 1/3,  $X_3$  and  $Z_2$  are standard normal distributed variables, which are discretized to bernoulli variables with the probability 0.5 for randomization process. Also, it is assumed  $\beta_1 = 3$ ,  $\beta_2 = (2, 3, 4)$ ,  $\beta_3 = 2$ ,  $\beta_4 = (0, 0)$ ,  $\beta_5 = 0$ , so that only  $X_1$ ,  $X_2$  and  $X_3$  have effects on the outcome  $Y$ .  $\varepsilon_i$  is normally distributed with a mean of 0 and a standard deviation of 2. Here only Pocock and Simon's marginal procedure is considered, since there are too many strata if stratified permuted block design is implemented.

To adjust conservative Type I error, one way is incorporating all randomization covariates into the analysis model ( $lm5$ ), where there is no loss of information used in hypothesis testing compared to randomization. A more efficient approach ac-

Figure 2.2: Simulated power for  $H_0 : \beta_1 = 0$  versus  $H_A : \beta_1 \neq 0$  for Pocock and Simon’s marginal procedure (PS) and complete randomization (CR) in %. Simulations based on 10000 runs and number of patients  $N = 100$ .



According to Theorem 2.3.1 is only incorporating the covariates those are influential on outcomes, i.e., constructing the model with  $X_1$ ,  $X_2$  and  $X_3$  (*lm3*). In Table 2.4, the model selection with BIC by stepwise algorithm is used to select the analysis model to do statistical inference. The algorithm is realized with “stepAIC” function in **R** by specifying  $k = \log(N)$ . Stepwise selection can be implemented backward, forward or with both directions. Here we only report the results for the backward selection (*backward*). Similar results are obtained from the other two stepwise selection methods (not shown). These methods are able to automatically achieve a



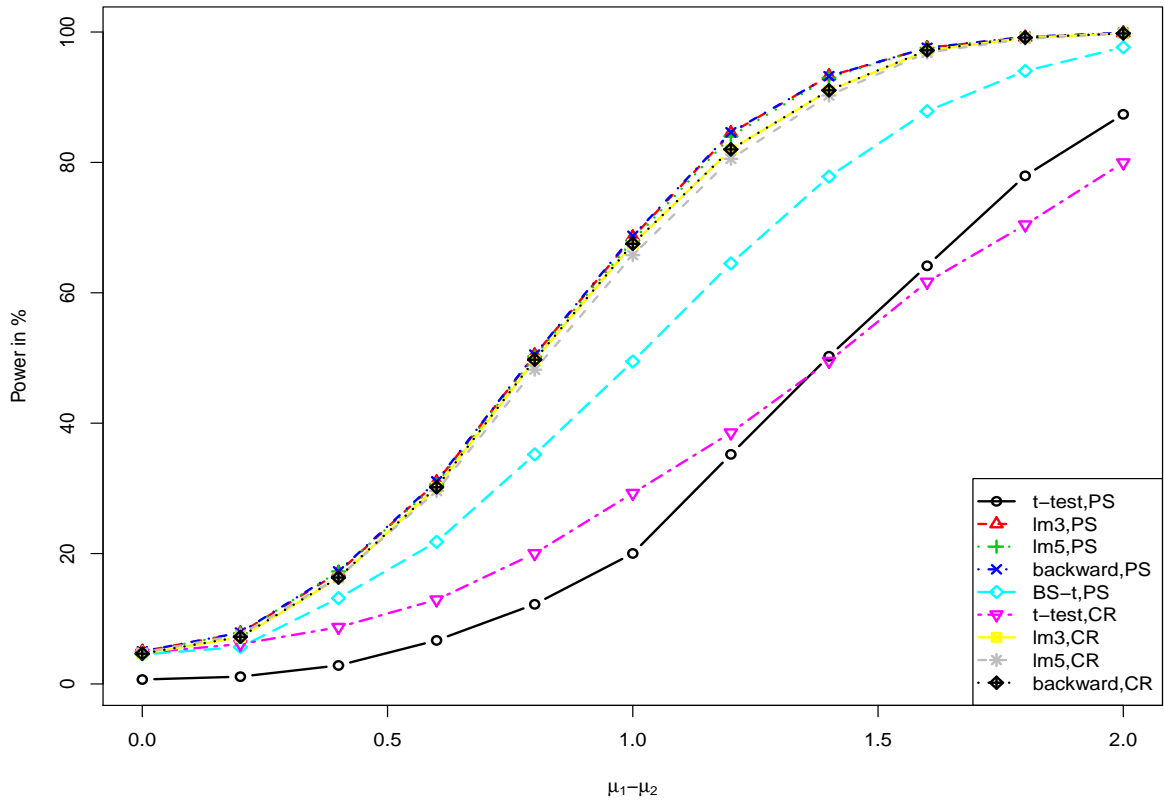
Table 2.4: Simulated Type I error under model (2.17) with backward model selection techniques for Pocock and Simon’s marginal procedure (PS) and complete randomization (CR). The number in parenthesis is the percentage of correct selection. Simulation based on 5000 runs.

	$N$	$t$ -test	$lm3$	$lm5$	$backward$	$BS$ - $t$
<b>PS</b>	50	1.12	4.64	4.74	4.88(84.34%)	5.44
	100	0.74	4.98	5.06	5.02(94.30%)	5.08
	200	0.40	4.96	4.96	4.94(96.70%)	4.70
	500	0.38	5.24	5.18	5.26(98.84%)	5.06
<b>CR</b>	50	5.32	4.94	4.94	5.66(83.60%)	-
	100	5.54	5.26	5.44	5.36(94.40%)	-
	200	5.06	4.98	4.82	5.02(97.12%)	-
	500	4.80	4.92	4.94	4.92(98.76%)	-

final model from a bunch of candidate models with different combinations of multiple covariates, based on which treatment effects can be compared. The results of Type I error of  $lm3$ ,  $lm5$  and  $backward$  are given in Table 2.4. The results of the two sample  $t$ -test ( $t$ -test) and the bootstrap  $t$ -test ( $BS$ - $t$ ) are also included. Power comparison results are shown in Figure 2.3. From the results of Table 2.4 and Figure 2.3, hypothesis testing for treatment effects based on conventional stepwise model selection techniques has valid Type I error and has similar power performance compared to the hypothesis testing with all randomization covariates used in the analysis, as long as the model selection techniques are able to identify the subset of all influential covariates.

*Case 5: Model Mis-specification.* In Section 2.2, the underlying response model (2.1) is assumed to have a form of linearly additive covariate effects in addition to treatment effects. Based on the model, theoretical properties of hypothesis testing are studied under covariate-adaptive designs. It is shown in Theorem 2.3.1 that valid Type I error can be obtained by utilizing a linear model incorporating all randomization covariates (the full model). This is within expectation since the full

Figure 2.3: Simulated power under model (2.17) with backward model selection techniques for Pocock and Simon's marginal procedure (PS) and complete randomization (CR). Simulation based on 5000 runs and number of patients  $N = 100$ .



model has the correct model specification that coincides with the assumed model. However, the underlying response model is usually unknown in practice and it is possible that covariate effects are not linearly additive on responses. For example, a covariate may have a non-linear form or have interaction effect with other covariates. Under these scenarios, the full model with all randomization covariates in a linearly additive pattern no longer has the correct model specification. In this section, a nonlinear response model with covariate effects in an exponential function is assumed to investigate properties of hypothesis testing for comparing treatment effects. Other

Table 2.5: Simulated Type I error under model (2.18) for Pocock and Simon’s marginal procedure (PS), stratified permuted block design (SPB) and complete randomization (CR). Simulation based on 10000 runs.

	$N$	$t$ -test	$lm(Z_1, Z_2)$	$BS$ - $t$
<b>PS</b>	100	1.70	3.87	3.90
	200	2.28	4.01	4.70
	500	2.32	4.11	4.25
<b>SPB</b>	100	1.59	3.66	4.96
	200	1.70	3.27	4.46
	500	1.98	3.55	4.70
<b>CR</b>	100	3.33	3.90	-
	200	4.01	4.03	-
	500	4.47	4.11	-

situations of model mis-specification can be studied similarly and will be left for future research projects. The following model is assumed for the responses  $Y$ . Two covariates  $Z_1$  and  $Z_2$  have a non-linear effect on the response  $Y$  in addition to treatment effects.

$$Y_i = \mu_1 I_i + \mu_2 (1 - I_i) + \exp\{Z_{i,1} + Z_{i,2}\} + \varepsilon_i, \quad (2.18)$$

where  $\varepsilon_i$  is distributed as  $\mathbf{N}(0, 4)$ .  $Z_1$  and  $Z_2$  are distributed as normal distributions  $\mathbf{N}(0, 1)$ , which are discretized into bernoulli variables with the probability 0.5 in order to be used in randomization procedures. Three randomization methods are investigated, including Pocock and Simon’s marginal procedure, stratified permuted block design and complete randomization. The two sample  $t$ -test ( $t$ -test), the linear model incorporating both covariates ( $lm(Z_1, Z_2)$ ) and the bootstrap  $t$ -test ( $BS$ - $t$ ) are used to compare treatment effects. The results of Type I error and power are given in Table 2.5 and 2.6, respectively.

From Table 2.5, the two sample  $t$ -test is conservative under covariate-adaptive

Table 2.6: Simulated power under model (2.18) for Pocock and Simon’s marginal procedure (PS), stratified permuted block design (SPB) and complete randomization (CR). Simulation based on 10000 runs and number of patients  $N = 100$ .

$\mu_1 - \mu_0$	<b>PS</b>			<b>SPB</b>			<b>CR</b>	
	<i>t-test</i>	$lm(Z_1, Z_2)$	<i>BS-t</i>	<i>t-test</i>	$lm(Z_1, Z_2)$	<i>BS-t</i>	<i>t-test</i>	$lm(Z_1, Z_2)$
0.0	1.87	3.67	4.14	1.40	3.13	4.92	3.60	4.37
0.4	4.21	8.31	7.71	3.01	6.35	7.57	5.69	7.73
0.8	10.71	19.56	16.85	9.26	17.63	17.14	12.66	18.39
1.2	21.91	35.81	29.90	19.79	34.39	30.61	21.85	33.71
1.6	35.79	53.98	44.25	34.73	52.37	46.14	35.60	51.25
2.0	50.37	66.52	57.94	49.97	66.99	60.75	48.32	64.83
2.4	62.81	77.09	69.23	62.34	76.80	70.67	60.48	75.32
2.8	71.56	83.22	76.54	71.30	83.59	77.96	69.56	82.37
3.2	79.88	87.99	83.16	79.09	88.32	83.93	76.86	87.27
3.6	84.69	91.36	87.39	84.22	91.25	87.85	82.97	91.06
4.0	88.47	93.30	90.37	87.87	93.25	90.43	86.97	92.48

designs. Due to model mis-specification, the Type I error of the linear model ( $lm(Z_1, Z_2)$ ) is deviated from the nominal level 5% , which is particularly obvious under stratified permuted block design. However, the bootstrap *t*-test tend to be robust under covariate-adaptive designs with the Type I error closer to 5%. Under complete randomization, the Type I error should be close to 5% as sample size increase based on the central limit theorem. Regarding the power comparison in Table 2.6, the model ( $lm(Z_1, Z_2)$ ) has better performance than the two sample *t*-test by utilizing more covariate information under all these three design methods. In addition, the power of the model ( $lm(Z_1, Z_2)$ ) is slightly larger than the bootstrap *t*-test under covariate-adaptive designs.

## 2.5 Conclusion

Covariate-adaptive designs have shown several advantages in applications to clinical trials, but the randomization scheme that is different from complete randomization may have influences on corresponding statistical inference results. In particular, the treatment comparison test based on linear models is shown to be conservative for a large class of covariate-adaptive designs, including Pocock and Simon’s marginal procedure and stratified permuted block design, if an influential randomization covariate is not incorporated in the final analysis. The results in this chapter provide a way to better understand statistical inference for covariate-adaptive designs in theory and allow us to implement covariate-adaptive designs appropriately to achieve valid statistical conclusions in practice.

The theoretical results in this chapter are based on several assumptions, among which all covariates are assumed to be independent of each other. However, there is no good way to guarantee that all covariates used in randomization are not correlated in practice. In view of that, similar problems will be studied with the independence assumption relaxed in Chapter 3. Moreover, generalized linear models are also used for covariate-adaptive designs when outcomes are not continuous. For example, logistic regression is used if responses are binary variables. These topics will be covered in Chapter 4.

## 2.6 Appendix: proof of theorems

To prove Theorem 2.3.1 and Theorem 2.3.2, we first prove the following two lemmas.

**Lemma 2.6.1.** *Under the assumptions of Theorem 2.3.1, then*

1.  $\sum I_i/N \xrightarrow{P} 1/2, \sum(1 - I_i)/N \xrightarrow{P} 1/2;$

$$2. \sum I_i X_{i,k}/N \xrightarrow{P} \mathbf{E}X_k/2, \sum (1 - I_i) X_{i,k}/N \xrightarrow{P} \mathbf{E}X_k/2, \sum I_i Z_{i,j}/N \xrightarrow{P} \mathbf{E}Z_j/2, \\ \text{and } \sum (1 - I_i) Z_{i,j}/N \xrightarrow{P} \mathbf{E}Z_j/2.$$

*Proof.* To prove part 1, it is easily seen that

$$\frac{1}{N} \sum I_i = \frac{1}{2} + \frac{\sum (2I_i - 1)}{2N} = \frac{1}{2} + \frac{1}{2} \frac{D_N}{N} \xrightarrow{P} \frac{1}{2}.$$

Now, we prove part 2. We first discuss the case when  $X_k$  is a discrete variable. Then

$$\frac{1}{N} \sum I_i X_{i,k} = \frac{1}{2} \left[ \frac{1}{N} \sum X_{i,k} + \frac{1}{N} \sum (2I_i - 1) X_{i,k} \right].$$

Notice that

$$\frac{1}{N} \sum (2I_i - 1) X_{i,k} = \frac{1}{N} \sum_{t_k=1}^{s_k^*} D(k; t_k) x_k^{t_k}.$$

The marginal imbalances  $D(k; t_k) = O_p(1)$  and  $x_k^{t_k}$ s are discrete values, which together with the weak law of large numbers implies that

$$\frac{1}{N} \sum I_i X_{i,k} \xrightarrow{P} \frac{1}{2} \mathbf{E}X_k.$$

If  $X_k$  is a continuous variable, then  $\delta_{i,k}^* = X_{i,k} - \mathbf{E}[X_{i,k} | d_k^*(X_{i,k})]$ , thus

$$\frac{1}{N} \sum I_i X_{i,k} = \frac{1}{N} \sum I_i \delta_{i,k}^* + \frac{1}{N} \sum I_i \mathbf{E}[X_{i,k} | d_k^*(X_{i,k})].$$

Since  $\mathbf{E}(X_{i,k} | d_k^*(X_{i,k}))$  is discrete, according to the proof of discrete case above,

$$\frac{1}{N} \sum I_i \mathbf{E}(X_{i,k} | d_k^*(X_{i,k})) \xrightarrow{P} \frac{1}{2} \mathbf{E}X_k.$$

To prove  $\sum I_i X_{i,k}/N \xrightarrow{P} \frac{1}{2} \mathbf{E} X_k$  for a continuous covariate, we also need show

$$\frac{1}{N} \sum I_i \delta_{i,k}^* \xrightarrow{P} 0.$$

Define

$$\tilde{X} = \{\tilde{X}_{i,k}, i = 1, 2, \dots, N, k = 1, 2, \dots, p\}.$$

Notice  $\delta_{i,k}^*$  and  $I_i$  are independent given  $\tilde{X}$ ,

$$\mathbf{E}(I_i \delta_{i,k}^* | \tilde{X}) = \mathbf{E}(I_i | \tilde{X}) \mathbf{E}(\delta_{i,k}^* | \tilde{X}) = \mathbf{E}(I_i | \tilde{X}) \mathbf{E}[\delta_{i,k}^* | d_k^*(X_{i,k})] = 0.$$

By the law of large numbers, for any  $\eta$ ,

$$\lim_{N \rightarrow \infty} \Pr\left[\left|\frac{1}{N} \sum I_i \delta_{i,k}^* - 0\right| > \eta | \tilde{X}\right] = 0$$

almost surely. By the dominated convergence theorem,

$$\frac{1}{N} \sum I_i \delta_{i,k}^* \xrightarrow{P} 0,$$

which finishes the proof of  $\sum I_i X_{i,k}/N \xrightarrow{P} \mathbf{E} X_k/2$ . By symmetry,  $\sum (1 - I_i) X_{i,k} \xrightarrow{P} \frac{1}{2} \mathbf{E} X_k/2$ . The proofs on  $Z_j$  are similar.  $\square$

**Lemma 2.6.2.** *Under the assumptions of Theorem 2.3.1,  $\hat{\beta}$  is a consistent estimator of  $\beta$ .*

*Proof.* By the definition of  $\hat{\beta}$ ,

$$\hat{\beta} = \beta + \left(\frac{\mathbf{X}^\top \mathbf{X}}{N}\right)^{-1} \frac{\mathbf{X}^\top \mathbf{Z} \boldsymbol{\gamma}}{N} + \left(\frac{\mathbf{X}^\top \mathbf{X}}{N}\right)^{-1} \frac{\mathbf{X}^\top \boldsymbol{\varepsilon}}{N}.$$

Notice  $\mathbf{E}X_k = \mathbf{E}Z_j = 0$  for all  $k$  and  $j$ , by the weak law of large numbers and independence of covariates,

$$\frac{1}{N}\mathbf{X}^\top\mathbf{X} \xrightarrow{P} \text{diag}\left(\frac{1}{2}, \frac{1}{2}, \text{Var}(X_1), \dots, \text{Var}(X_p)\right)$$

and

$$\frac{1}{N}\mathbf{X}^\top\mathbf{Z} = \frac{1}{N} \begin{bmatrix} \sum I_i Z_{i,1} & \cdots & \sum I_i Z_{i,q} \\ \sum (1 - I_i) Z_{i,1} & \cdots & \sum (1 - I_i) Z_{i,q} \\ \sum X_{i,1} Z_{i,1} & \cdots & \sum X_{i,1} Z_{i,q} \\ \vdots & \ddots & \vdots \\ \sum X_{i,p} Z_{i,1} & \cdots & \sum X_{i,p} Z_{i,q} \end{bmatrix} \xrightarrow{P} \begin{bmatrix} \frac{1}{2}\mathbf{E}Z_1 & \cdots & \frac{1}{2}\mathbf{E}Z_q \\ \frac{1}{2}\mathbf{E}Z_1 & \cdots & \frac{1}{2}\mathbf{E}Z_q \\ \mathbf{E}X_1\mathbf{E}Z_1 & \cdots & \mathbf{E}X_1\mathbf{E}Z_q \\ \vdots & \ddots & \vdots \\ \mathbf{E}X_p\mathbf{E}Z_1 & \cdots & \mathbf{E}X_p\mathbf{E}Z_q \end{bmatrix}.$$

Further, by the independence of  $I_i$  and  $\varepsilon_i$ , for  $i = 1, \dots, N$ ,

$$\frac{1}{N}\mathbf{X}^\top\boldsymbol{\varepsilon} = \frac{1}{N} \begin{bmatrix} \sum I_i \varepsilon_i \\ \sum (1 - I_i) \varepsilon_i \\ \sum X_{i,1} \varepsilon_i \\ \vdots \\ \sum X_{i,p} \varepsilon_i \end{bmatrix} \xrightarrow{P} \mathbf{0}.$$

Hence,

$$\left(\frac{\mathbf{X}^\top\mathbf{X}}{N}\right)^{-1} \frac{\mathbf{X}^\top\mathbf{Z}\boldsymbol{\gamma}}{N} + \left(\frac{\mathbf{X}^\top\mathbf{X}}{N}\right)^{-1} \frac{\mathbf{X}^\top\boldsymbol{\varepsilon}}{N} \xrightarrow{P} \mathbf{0}.$$

Then it follows that,

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \xrightarrow{P} \mathbf{0}.$$



□

*Proof of Theorem 2.3.1.* The following test statistic is used for hypothesis testing (2.3),

$$\frac{\mathbf{L}\hat{\boldsymbol{\beta}}}{(\hat{\sigma}^2 \mathbf{L}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{L}^\top)^{1/2}},$$

where  $\mathbf{L} = (1, -1, 0, \dots, 0)$  and  $\hat{\sigma}^2 = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) / (N - p - 2)$ ,  $p$  is the total number of independent variables in the model (2.2) besides  $\mu_1$  and  $\mu_2$ .

The numerator is equal to

$$\mathbf{L}\hat{\boldsymbol{\beta}} = \hat{\mu}_1 - \hat{\mu}_2 = \mu_1 - \mu_2 + \mathbf{L}\left(\frac{\mathbf{X}^\top \mathbf{X}}{N}\right)^{-1} \frac{\mathbf{X}^\top \mathbf{Z}\boldsymbol{\gamma}}{N} + \mathbf{L}\left(\frac{\mathbf{X}^\top \mathbf{X}}{N}\right)^{-1} \frac{\mathbf{X}^\top \boldsymbol{\varepsilon}}{N}.$$

We have shown that in the proof of Lemma 2.6.2,

$$\frac{1}{N} \mathbf{X}^\top \mathbf{X} \xrightarrow{P} \text{diag}\left(\frac{1}{2}, \frac{1}{2}, \text{Var}(X_1), \dots, \text{Var}(X_p)\right) \doteq \mathbf{M}.$$

Let

$$\mathbf{A} = \mathbf{L}\mathbf{M}^{-1} \left[ \frac{\mathbf{X}^\top \mathbf{Z}\boldsymbol{\gamma}}{N} + \frac{\mathbf{X}^\top \boldsymbol{\varepsilon}}{N} \right]$$

and

$$\mathbf{B} = \mathbf{L} \left[ \left(\frac{\mathbf{X}^\top \mathbf{X}}{N}\right)^{-1} - \mathbf{M}^{-1} \right] \left[ \frac{\mathbf{X}^\top \mathbf{Z}\boldsymbol{\gamma}}{N} + \frac{\mathbf{X}^\top \boldsymbol{\varepsilon}}{N} \right].$$

By some matrix calculations,

$$\begin{aligned} \mathbf{A} &= \frac{2}{N} \left( \sum_j \sum_i (2I_i - 1) \gamma_j Z_{i,j} + \sum_i (2I_i - 1) \varepsilon_i \right) = \frac{2}{N} \left( \sum_{j \in C} \sum_i (2I_i - 1) \gamma_j \delta_{i,j} \right. \\ &\quad \left. + \sum_{j \in C} \sum_i (2I_i - 1) \gamma_j \mathbf{E}[Z_{i,j} | d_j(Z_{i,j})] + \sum_{j \notin C} \sum_i (2I_i - 1) \gamma_j Z_{i,j} + \sum_i (2I_i - 1) \varepsilon_i \right). \end{aligned}$$

Apply condition (B) to all margins with respect to each covariate  $Z_j, j = 1, 2, \dots, q$ , we have

$$\sum_i (2I_i - 1) \gamma_j Z_{i,j} = \sum_{r_j=1}^{s_j} D(j; r_j) \gamma_j z_j^{r_j} = O_p(1), \quad j \notin C,$$

and

$$\sum_i (2I_i - 1) \gamma_j \mathbf{E}(Z_{i,j} | d_j(Z_{i,j})) = \sum_{r_j=1}^{s_j} D(j; r_j) \gamma_j \mathbf{E}[Z_j | d_j(Z_j) = z_j^{r_j}] = O_p(1), \quad j \in C.$$

It follows that

$$\sum_{j \in C} \sum_i (2I_i - 1) \gamma_j \mathbf{E}[Z_{i,j} | d_j(Z_{i,j})] + \sum_{j \notin C} \sum_i (2I_i - 1) \gamma_j Z_{i,j} = O_p(1).$$

Then

$$\frac{2}{N} \left( \sum_{j \in C} \sum_i (2I_i - 1) \gamma_j \mathbf{E}[Z_{i,j} | d_j(Z_{i,j})] + \sum_{j \notin C} \sum_i (2I_i - 1) \gamma_j Z_{i,j} \right) = o_p(N^{-1/2}).$$

Define  $\tilde{Z} = \{\tilde{Z}_{i,j}, i = 1, 2, \dots, N, j = 1, 2, \dots, q\}$  and  $\tilde{I} = \{I_i, i = 1, \dots, N\}$ , then  $(\delta_{i,j}, \varepsilon_i)$  are conditionally independent of  $\tilde{I}$  given  $\tilde{Z}$ . Notice  $\mathbf{E}(\varepsilon_i | \tilde{Z}) = \mathbf{E}(\varepsilon_i) = 0$ ,

and  $\mathbf{E}(\delta_{i,j}|\tilde{Z}) = \mathbf{E}[\delta_{i,j}|d_j(Z_{i,j})] = 0$  by the definition of  $\delta_{i,j}$ , then

$$\frac{2}{N}\mathbf{E}\left(\sum_{j \in \mathcal{C}} \sum_i (2I_i - 1)\gamma_j \delta_{i,j} + \sum_i (2I_i - 1)\varepsilon_i | \tilde{Z}\right) = 0.$$

Further, by the conditional independence of  $(\delta_{i,j}, \varepsilon_i)$  and  $\tilde{I}$  given  $\tilde{Z}$ ,

$$\text{Cov}\left(\frac{2}{N} \sum_{j \in \mathcal{C}} \sum_i (2I_i - 1)\gamma_j \delta_{i,j}, \frac{2}{N} \sum_i (2I_i - 1)\varepsilon_i | \tilde{Z}\right) = 0$$

and notice  $(2I_i - 1)^2 = 1$ ,

$$\text{Var}\left(\frac{2}{N} \sum_i (2I_i - 1)\varepsilon_i | \tilde{Z}\right) = \mathbf{E}\left(\frac{4}{N^2} \sum_i (2I_i - 1)^2 \varepsilon_i^2 | \tilde{Z}\right) = \frac{4\sigma_\varepsilon^2}{N}$$

and

$$\text{Var}\left(\frac{2}{N} \sum_{j \in \mathcal{C}} \sum_i (2I_i - 1)\gamma_j \delta_{i,j} | \tilde{Z}\right) = \frac{4}{N^2} \mathbf{E}\left(\sum_{j \in \mathcal{C}} \sum_i (2I_i - 1)^2 \gamma_j^2 \delta_{i,j}^2 | \tilde{Z}\right) = \frac{4}{N} \sum_{j \in \mathcal{C}} \gamma_j^2 V_{\delta,j},$$

where  $V_{\delta,j} = N^{-1} \sum \text{Var}[\delta_{i,j}|d_j(Z_{i,j})]$ . Therefore,

$$\text{Var}\left[\frac{2}{N} \left(\sum_{j \in \mathcal{C}} \sum_i (2I_i - 1)\gamma_j \delta_{i,j} + \sum_i (2I_i - 1)\varepsilon_i\right) | \tilde{Z}\right] = \frac{4\left(\sum_{j \in \mathcal{C}} \gamma_j^2 V_{\delta,j} + \sigma_\varepsilon^2\right)}{N}.$$

By the central limit theorem, given  $(\tilde{I}, \tilde{Z})$ ,

$$\frac{2}{\sqrt{N}} \left\{ \sum_{j \in \mathcal{C}} \sum_i (2I_i - 1)\gamma_j \delta_{i,j} + \sum_i (2I_i - 1)\varepsilon_i \right\}$$

is asymptotically normal with mean zero and variance  $4\left(\sum_{j \in \mathcal{C}} \gamma_j^2 V_{\delta,j} + \sigma_\varepsilon^2\right)$ , which converges to  $4\left(\sum_{j \in \mathcal{C}} \gamma_j^2 \sigma_{\delta,j}^2 + \sigma_\varepsilon^2\right)$ , where  $\sigma_{\delta,j}^2 = \mathbf{E}[\text{Var}(\delta_{i,j}|d_j(Z_{i,j}))]$ . Since the limit-

ing distribution is independent of  $(\tilde{I}, \tilde{Z})$ , the convergence in distribution also holds unconditionally. Hence, by Slutsky's theorem,

$$\sqrt{N}\mathbf{A} \xrightarrow{D} \mathbf{N}(0, 4(\sum_{j \in C} \gamma_j^2 \sigma_{\delta,j}^2 + \sigma_\varepsilon^2)).$$

For part  $\mathbf{B}$ , we will show  $\sqrt{N}\mathbf{B} \xrightarrow{P} 0$ . To prove this, it suffices to show

$$\frac{\mathbf{X}^\top \mathbf{Z} \boldsymbol{\gamma}}{\sqrt{N}} + \frac{\mathbf{X}^\top \boldsymbol{\varepsilon}}{\sqrt{N}} = O_p(1),$$

since we have already known

$$\left(\frac{\mathbf{X}^\top \mathbf{X}}{N}\right)^{-1} - \mathbf{M}^{-1} \xrightarrow{P} \mathbf{0}.$$

Notice that

$$\frac{\mathbf{X}^\top \mathbf{Z} \boldsymbol{\gamma}}{\sqrt{N}} + \frac{\mathbf{X}^\top \boldsymbol{\varepsilon}}{\sqrt{N}} = \frac{1}{\sqrt{N}} \begin{bmatrix} \sum_j \sum_i I_i Z_{i,j} \gamma_j + \sum_i I_i \varepsilon_i \\ \sum_j \sum_i (1 - I_i) Z_{i,j} \gamma_j + \sum_i (1 - I_i) \varepsilon_i \\ \sum_j \sum_i X_{i,1} Z_{i,j} \gamma_j + \sum_i X_{i,1} \varepsilon_i \\ \vdots \\ \sum_j \sum_i X_{i,p} Z_{i,j} \gamma_j + \sum_i X_{i,p} \varepsilon_i \end{bmatrix}.$$

By the central limit theorem, for any  $k$ ,  $k = 1, 2, \dots, p$

$$\frac{1}{\sqrt{N}} \left( \sum_j \sum_i X_{i,k} Z_{i,j} \gamma_j + \sum_i X_{i,k} \varepsilon_i \right) = O_p(1).$$

Also,

$$\begin{aligned} \frac{1}{\sqrt{N}} \left( \sum_j \sum_i I_i Z_{i,j} \gamma_j + \sum_i I_i \varepsilon_i \right) &= \frac{1}{2} \left[ \frac{1}{\sqrt{N}} \sum_j \sum_i Z_{i,j} \gamma_j + \frac{1}{\sqrt{N}} \sum_i \varepsilon_i \right. \\ &\quad \left. + \frac{1}{\sqrt{N}} \sum_j \sum_i (2I_i - 1) Z_{i,j} \gamma_j + \frac{1}{\sqrt{N}} \sum_i (2I_i - 1) \varepsilon_i \right]. \end{aligned}$$

Again by the central limit theorem,

$$\frac{1}{\sqrt{N}} \sum_j \sum_i Z_{i,j} \gamma_j + \frac{1}{\sqrt{N}} \sum_i \varepsilon_i = O_p(1).$$

Notice that

$$\frac{1}{\sqrt{N}} \sum_j \sum_i (2I_i - 1) Z_{i,j} \gamma_j + \frac{1}{\sqrt{N}} \sum_i (2I_i - 1) \varepsilon_i = \frac{\sqrt{N}}{2} \mathbf{A}.$$

It follows from  $\sqrt{N} \mathbf{A}$  converges to a normal distribution that,

$$\frac{1}{\sqrt{N}} \sum_j \sum_i (2I_i - 1) Z_{i,j} \gamma_j + \frac{1}{\sqrt{N}} \sum_i (2I_i - 1) \varepsilon_i = O_p(1).$$

Hence,

$$\frac{1}{\sqrt{N}} \left( \sum_j \sum_i I_i Z_{i,j} \gamma_j + \sum_i I_i \varepsilon_i \right) = O_p(1).$$

By symmetry,

$$\frac{1}{\sqrt{N}} \left( \sum_j \sum_i (1 - I_i) Z_{i,j} \gamma_j + \sum_i (1 - I_i) \varepsilon_i \right) = O_p(1).$$

It follows that

$$\frac{\mathbf{X}^\top \mathbf{Z} \boldsymbol{\gamma}}{\sqrt{N}} + \frac{\mathbf{X}^\top \boldsymbol{\varepsilon}}{\sqrt{N}} = O_p(1).$$

Hence, we can conclude  $\sqrt{N} \mathbf{B} \xrightarrow{P} 0$ , which together with Slutsky's theorem implies that,

$$\sqrt{N} [(\hat{\mu}_1 - \hat{\mu}_2) - (\mu_1 - \mu_2)] \xrightarrow{D} \mathbf{N}(0, 4(\sum_{j \in \mathcal{C}} \gamma_j^2 \sigma_{\delta,j}^2 + \sigma_\varepsilon^2)),$$

On the other side, we show  $\hat{\sigma}^2 \mathbf{L}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{L}^\top$  is an inflated estimator of the variance of  $\mathbf{L} \hat{\boldsymbol{\beta}}$ , which means it is larger than  $4(\sum_{j \in \mathcal{C}} \gamma_j^2 \sigma_{\delta,j}^2 + \sigma_\varepsilon^2)$ . Notice that

$$\mathbf{L}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{L}^\top = \frac{1}{N} \mathbf{L} \left( \frac{\mathbf{X}^\top \mathbf{X}}{N} \right)^{-1} \mathbf{L}^\top = \frac{4}{N} + o_p\left(\frac{1}{N}\right)$$

and

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{N-p-2} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}})^\top (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}) = \frac{1}{N} (\mathbf{Y} - \mathbf{X} \boldsymbol{\beta})^\top (\mathbf{Y} - \mathbf{X} \boldsymbol{\beta}) \\ &+ \frac{1}{N} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{X}^\top \mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \frac{2}{N} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \mathbf{X}^\top (\mathbf{Y} - \mathbf{X} \boldsymbol{\beta}) + o_p(1). \end{aligned}$$

By consistency of  $\hat{\boldsymbol{\beta}}$  proved in Lemma 2.6.2,

$$\frac{1}{N} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{X}^\top \mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{P} 0$$

and

$$\frac{2}{N} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \mathbf{X}^\top (\mathbf{Y} - \mathbf{X} \boldsymbol{\beta}) = 2(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \frac{\mathbf{X}^\top \mathbf{Z} \boldsymbol{\gamma}}{N} + 2(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \frac{\mathbf{X}^\top \boldsymbol{\varepsilon}}{N} \xrightarrow{P} 0.$$

It follows from the law of large numbers and independence of  $\mathbf{Z}$  and  $\boldsymbol{\varepsilon}$  that,

$$\frac{1}{N}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \frac{1}{N}(\mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon})^\top(\mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}) \xrightarrow{P} \sigma_\varepsilon^2 + \sum_{j=1}^q \gamma_j^2 \text{Var}(Z_j).$$

Hence,

$$\hat{\sigma}^2 \xrightarrow{P} \sigma_\varepsilon^2 + \sum_{j=1}^q \gamma_j^2 \text{Var}(Z_j)$$

and thus,

$$\hat{\sigma}^2 \mathbf{L}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{L}^\top = \frac{4}{N}(\sigma_\varepsilon^2 + \sum_{j=1}^q \gamma_j^2 \text{Var}(Z_j)) + o_p\left(\frac{1}{N}\right).$$

Then under  $H_0 : \mu_1 - \mu_2 = 0$ ,

$$T \xrightarrow{D} \mathbf{N}(0, \tau^2), \quad \tau^2 = \frac{\sigma_\varepsilon^2 + \sum_{j \in C} \gamma_j^2 \sigma_{\delta,j}^2}{\sigma_\varepsilon^2 + \sum_{j=1}^q \gamma_j^2 \text{Var}(Z_j)} = \frac{\sigma_\delta^2}{\sigma_z^2}.$$

where  $\sigma_z^2 = \sigma_\varepsilon^2 + \sum_{j=1}^q \gamma_j^2 \text{Var}(Z_j)$  and  $\sigma_\delta^2 = \sigma_\varepsilon^2 + \sum_{j \in C} \gamma_j^2 \sigma_{\delta,j}^2$ .

Notice, for  $\gamma_j \neq 0$  and  $j \in C$ ,

$$\gamma_j^2 \text{Var}(Z_j) > \gamma_j^2 \mathbf{E}[\text{Var}(\delta_{i,j} | d_j(Z_{i,j}))] = \gamma_j^2 \sigma_{\delta,j}^2,$$

it follows, when  $N \rightarrow \infty$ ,

$$\Pr(|T| > Z_{1-\alpha/2}) \rightarrow 2\Phi\left(-\frac{\sigma_z Z_{1-\alpha/2}}{\sigma_\delta}\right) < \alpha,$$

where  $\Phi$  is the cumulative distribution function of standard normal distribution.

Similarly, under  $H_A : \mu_1 - \mu_2 \neq 0$  with a sequence of local alternatives, i.e.,  $\mu_1 - \mu_2 = \delta/\sqrt{N}$  for a fixed  $\delta \neq 0$ ,

$$T \xrightarrow{D} \mathbf{N}(\Delta, \tau^2), \quad \Delta = \frac{\delta}{2\sigma_z}.$$

which completes the proof of Theorem 2.3.1. □

*Proof of Theorem 2.3.2.* The test statistic can be written as

$$T^* = \frac{m^{-1}\sqrt{N}(\mathbf{C}\boldsymbol{\beta} - \boldsymbol{\xi}_0)^\top [\mathbf{C}(\mathbf{X}^\top \mathbf{X}/N)^{-1} \mathbf{C}^\top]^{-1} \sqrt{N}(\mathbf{C}\boldsymbol{\beta} - \boldsymbol{\xi}_0)}{\hat{\sigma}^2}.$$

Under  $H_0 : \mathbf{C}\boldsymbol{\beta} = \boldsymbol{\xi}_0$ , it holds

$$\sqrt{N}(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\xi}_0) = \sqrt{N}\mathbf{C}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}),$$

and  $H_A : \mathbf{C}\boldsymbol{\beta} = \boldsymbol{\xi}_1$ , it holds

$$\sqrt{N}(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\xi}_0) = \sqrt{N}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{C}\boldsymbol{\beta} + \boldsymbol{\xi}_1 - \boldsymbol{\xi}_0) = \sqrt{N}\mathbf{C}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \sqrt{N}(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_0).$$

Notice

$$\frac{1}{N}\mathbf{X}^\top \mathbf{X} \xrightarrow{P} \mathbf{M} = \text{diag}\left(\frac{1}{2}, \frac{1}{2}, \text{Var}(X_1), \dots, \text{Var}(X_p)\right),$$



we have

$$\begin{aligned}
\sqrt{N}\mathbf{C}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \sqrt{N}\mathbf{C}\left[\left(\frac{\mathbf{X}^\top\mathbf{X}}{N}\right)^{-1}\frac{\mathbf{X}^\top\mathbf{Z}\boldsymbol{\gamma}}{N} + \left(\frac{\mathbf{X}^\top\mathbf{X}}{N}\right)^{-1}\frac{\mathbf{X}^\top\boldsymbol{\varepsilon}}{N}\right] \\
&= \sqrt{N}\mathbf{C}\mathbf{M}^{-1}\left[\frac{\mathbf{X}^\top\mathbf{Z}\boldsymbol{\gamma}}{N} + \frac{\mathbf{X}^\top\boldsymbol{\varepsilon}}{N}\right] + o_p(1) \\
&= \mathbf{C}\mathbf{M}^{-1}\frac{1}{\sqrt{N}}\begin{bmatrix} \sum_j \sum_i I_i Z_{i,j} \gamma_j + \sum_i I_i \varepsilon_i \\ \sum_j \sum_i (1 - I_i) Z_{i,j} \gamma_j + \sum_i (1 - I_i) \varepsilon_i \\ \sum_j \sum_i X_{i,1} Z_{i,j} \gamma_j + \sum_i X_{i,1} \varepsilon_i \\ \vdots \\ \sum_j \sum_i X_{i,p} Z_{i,j} \gamma_j + \sum_i X_{i,p} \varepsilon_i \end{bmatrix} + o_p(1).
\end{aligned}$$

Let  $\mathbf{C} = [\mathbf{0}_{m \times 2}, \tilde{\mathbf{C}}]$ , where  $\mathbf{0}_{m \times 2}$  is an  $(m \times 2)$  matrix with all entries zero, and  $\mathbf{M} = \text{diag}(1/2, 1/2, \tilde{\mathbf{M}})$ , then

$$\sqrt{N}\mathbf{C}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \tilde{\mathbf{C}}\tilde{\mathbf{M}}^{-1}\frac{1}{\sqrt{N}}\begin{bmatrix} \sum_j \sum_i X_{i,1} Z_{i,j} \gamma_j + \sum_i X_{i,1} \varepsilon_i \\ \vdots \\ \sum_j \sum_i X_{i,p} Z_{i,j} \gamma_j + \sum_i X_{i,p} \varepsilon_i \end{bmatrix} + o_p(1).$$

By the central limit theorem, together with fact  $\mathbf{C}\mathbf{M}^{-1}\mathbf{C}^\top = \tilde{\mathbf{C}}\tilde{\mathbf{M}}^{-1}\tilde{\mathbf{C}}^\top$ ,

$$\sqrt{N}\mathbf{C}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} \mathbf{N}(\mathbf{0}, \sigma_z^2 \mathbf{C}\mathbf{M}^{-1}\mathbf{C}^\top).$$

Moreover, it can shown that,

$$\frac{[\mathbf{C}(\mathbf{X}^\top\mathbf{X}/N)^{-1}\mathbf{C}^\top]^{-1}}{\hat{\sigma}^2} \xrightarrow{P} [\sigma_z^2 \mathbf{C}\mathbf{M}^{-1}\mathbf{C}^\top]^{-1}.$$

Hence, under  $H_0 : \mathbf{C}\boldsymbol{\beta} = \boldsymbol{\xi}_0$ ,

$$T^* \xrightarrow{D} \chi_m^2/m,$$

and under  $H_A : \mathbf{C}\boldsymbol{\beta} = \boldsymbol{\xi}_1$  with a sequence of local alternatives, i.e.,  $(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_0) = \boldsymbol{\eta}/\sqrt{N}$  for a fixed  $\boldsymbol{\eta} \neq \mathbf{0}$ ,

$$T^* \xrightarrow{D} \chi_{(m)}^2(\lambda)/m, \quad \lambda = \boldsymbol{\eta}^\top [\mathbf{C}\mathbf{M}^{-1}\mathbf{C}^\top]^{-1} \boldsymbol{\eta} / \sigma_z^2,$$

which finishes the proof of Theorem 2.3.2. □

## Chapter 3

# Statistical Inference for Linear Models with Correlated Covariates

### 3.1 Introduction

In Chapter 2, the properties of statistical inference are studied for linear models under a large family of covariate-adaptive designs. Among several assumptions in the framework to derive theoretical results, it is assumed that all the covariates used in randomization are independent of each other. Based on this critical assumption of independence and other assumptions, we proved that the hypothesis testing to compare treatment effects between two groups is conservative and that the hypothesis testing for a linear combination of covariates remains valid. However, despite its importance in theory, the assumption that all covariates are independent is very strong and may be not satisfied in practice. Therefore, there are concerns that the properties of statistical inference may be different from the independence case and whether the conclusions established in the last chapter still hold if covariates are correlated. The objective of this chapter is to address these concerns regarding

statistical inference for linear models with correlated covariates.

Under complete randomization, patients are allocated to each treatment group of a clinical trial with equal probability. It can be shown that, if covariates are correlated under complete randomization, the hypothesis testing for comparing treatment effects remains valid, however, the estimators for coefficients of covariates are biased if important covariates are omitted and are not included in the analysis model. The corresponding properties of statistical inference under covariate-adaptive designs will be studied in this chapter. In particular, biasedness and consistency will be given with regards to the estimators of treatment effects and covariate effects (Theorem 3.3.1). Moreover, the properties of hypothesis testing of treatment effects will also be established (Theorem 3.3.2).

In Section 3.2, the framework to study statistical inference for linear models with correlated covariates is described in details. In Section 3.3, theoretical properties of estimators and testing hypotheses are given and discussed for covariate-adaptive designs. Type I error and power are evaluated and are compared with complete randomization via simulations in Section 3.4. Conclusions and remarks are given in Section 3.5, and technical proofs of theorems are attached in the end of this chapter.

## 3.2 Framework

In this section, the general framework is proposed to study inference properties of linear models with correlated covariates for covariate-adaptive randomized clinical trials. By incorporating a covariance matrix into the underlying model, the assumption of independent covariates is relaxed and the covariates are allowed to be correlated with each other in the new framework, which enables us to study inference properties under a more general setting. Analogous to the independence case,

an underlying model and a working model are given to represent the situation that only partial covariate information used in randomization is employed to construct the final analysis. The main difference from Chapter 2 is that the covariance matrix of covariates can be any symmetric positive semi-definite matrix, instead of being restricted to a diagonal matrix.

Similar to Chapter 2, it is supposed that there are two treatment groups in a covariate-adaptive randomized clinical trial, which are treatment 1 and treatment 2. The main treatment effects are measured by  $\mu_1$  and  $\mu_2$  for treatment 1 and 2, respectively. Let  $N$  be the total number of subjects enrolled in the study,  $I_i$  be the assignment of the  $i$ th subject, where  $I_i = 1$  for treatment 1 and  $I_i = 0$  for treatment 2,  $i = 1, 2, \dots, N$ . The following linear model is assumed to be the underlying model,

$$Y_i = \mu_1 I_i + \mu_2 (1 - I_i) + \beta_1 X_{i,1} + \dots + \beta_p X_{i,p} + \gamma_1 Z_{i,1} + \dots + \gamma_q Z_{i,q} + \varepsilon_i, \quad (3.1)$$

where

- $X_{i,k}$ s and  $Z_{i,j}$ s are discrete or continuous covariates which are independent and identically distributed as  $X_k$  and  $Z_j$ ,  $k = 1, \dots, p$  and  $j = 1, \dots, q$ ;
- both  $X_{i,k}$ s and  $Z_{i,j}$ s are used in the randomization procedure, but only  $X_{i,k}$ s are used in final statistical inference,  $k = 1, \dots, p$  and  $j = 1, \dots, q$ ;
- all covariates have means equal to zero, i.e.,  $\mathbf{E}X_k = 0$  and  $\mathbf{E}Z_j = 0$  for all  $k$  and  $j$ ,  $k = 1, \dots, p$  and  $j = 1, \dots, q$ ;
- $\varepsilon_i$ s are independent and identically distributed random errors with mean zero and variance  $\sigma_\varepsilon^2$  and independent of  $X_k$  and  $Z_j$ ,  $k = 1, \dots, p$  and  $j = 1, \dots, q$ .
- let  $\mathcal{X} = (X_1, \dots, X_p)^\top$  and  $\mathcal{Z} = (Z_1, \dots, Z_q)^\top$ , then the covariance matrix of the

covariate vector  $(\mathcal{X}^\top, \mathcal{Z}^\top)^\top = (X_1, \dots, X_p, Z_1, \dots, Z_q)^\top$  is

$$\begin{aligned} \Sigma &= \begin{bmatrix} \text{Cov}(X_1, X_1) & \cdots & \text{Cov}(X_1, X_p) & \text{Cov}(X_1, Z_1) & \cdots & \text{Cov}(X_1, Z_q) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_p, X_1) & \cdots & \text{Cov}(X_p, X_p) & \text{Cov}(X_p, Z_1) & \cdots & \text{Cov}(X_p, Z_q) \\ \text{Cov}(Z_1, X_1) & \cdots & \text{Cov}(Z_1, X_p) & \text{Cov}(Z_1, Z_1) & \cdots & \text{Cov}(Z_1, Z_q) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(Z_q, X_1) & \cdots & \text{Cov}(Z_q, X_p) & \text{Cov}(Z_q, Z_1) & \cdots & \text{Cov}(Z_q, Z_q) \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_{x,x} & \Sigma_{x,z} \\ \Sigma_{z,x} & \Sigma_{z,z} \end{bmatrix}, \end{aligned} \quad (3.2)$$

where  $\Sigma_{x,x} = \text{Var}(\mathcal{X})$ ,  $\Sigma_{z,z} = \text{Var}(\mathcal{Z})$ ,  $\Sigma_{x,z} = \text{Cov}(\mathcal{X}, \mathcal{Z})$  and  $\Sigma_{z,x} = \Sigma_{x,z}^\top$ .

REMARK 3.2.1. *The assumptions of the underlying model (3.1) are similar to those of the independence case in Chapter 2, except that a dependence structure of covariates is defined in (3.2), which allows us to study inference properties of covariate-adaptive designs for a more general covariate profile. This assumption is much closer to the situation in practice where covariates considered in a clinical trial are not necessarily independent. In addition, similar to Chapter 2, all covariates are assumed to be a scalar, however, our results can be easily extended to the case where discrete covariates are represented by vectors.*

Further, define  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^\top$ ,  $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)^\top$ ,  $\boldsymbol{\beta} = (\mu_1, \mu_2, \beta_1, \dots, \beta_p)^\top$ ,  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)^\top$ ,  $\mathbf{I} = (I_1, \dots, I_N)^\top$ , and  $\mathbf{1} = (1, \dots, 1)^\top$  with all of  $N$  entries equal to

1. In addition, define matrices

$$\mathbf{Z} = \begin{bmatrix} Z_{1,1} & \cdots & Z_{1,q} \\ Z_{2,1} & \cdots & Z_{2,q} \\ \vdots & \ddots & \vdots \\ Z_{N,1} & \cdots & Z_{N,q} \end{bmatrix}$$

and

$$\mathbf{X} = \begin{bmatrix} I_1 & 1 - I_1 & X_{1,1} & \cdots & X_{1,p} \\ I_2 & 1 - I_2 & X_{2,1} & \cdots & X_{2,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ I_N & 1 - I_N & X_{N,1} & \cdots & X_{N,p} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & (\mathbf{1} - \mathbf{I}) & \underline{\mathbf{X}} \end{bmatrix},$$

where

$$\underline{\mathbf{X}} = \begin{bmatrix} X_{1,1} & \cdots & X_{1,p} \\ X_{2,1} & \cdots & X_{2,p} \\ \vdots & \ddots & \vdots \\ X_{N,1} & \cdots & X_{N,p} \end{bmatrix}.$$

Therefore, the matrix  $\mathbf{Z}$  includes the covariate information which is used in randomization but is not used in analysis. The matrix  $\mathbf{X}$  are the data used to construct the analysis model, including treatment assignments of all patients  $\mathbf{I}$  and covariate information  $\underline{\mathbf{X}}$  that is employed both in randomization and inference.

Under these notations, it is straightforward to rewrite the underlying true model

with the matrix representation, which is equivalent to model (3.1),

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon},$$

Our interest is to study the properties of statistical inference for a covariate-adaptive design if all or partial randomization covariates are omitted from the linear model for final analysis. For this purpose, a working model (analysis model) with partial covariate information is defined to do statistical inference based on the data obtained from a clinical trial. The working model to do analysis is assumed to be,

$$\mathbf{E}[Y_i] = \mu_1 I_i + \mu_2(1 - I_i) + \beta_1 X_{i,1} + \dots + \beta_p X_{i,p}, \quad (3.3)$$

or with the matrix representation,

$$\mathbf{E}[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}.$$

In the working model (3.3), there are in total  $(p + 2)$  unknown parameters to estimate, including the treatment effects  $\mu_1$  and  $\mu_2$ , as well as  $p$  coefficients of covariates  $\beta_1, \dots, \beta_p$ . The ordinary least squares (OLS) method can be used to obtain the estimate of  $\boldsymbol{\beta}$ , which has the closed expression,  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$ , in matrix form.

When we studied statistical inference for linear models with independent covariates under covariate-adaptive designs in Chapter 2, two types of hypothesis testing are considered. One is comparing treatment effects, which is usually the primary goal of a clinical study, the other one is to test if a covariate is influential on outcomes of patients in the study. Both testing hypotheses are studied based on the facts that the corresponding estimators, the treatment difference and covariate effects, are unbiased if covariates are independent to each other. Under the assumption of



correlated covariates in this chapter, however, we will show in the next section that, even though the treatment difference is still unbiased, the covariate effects are not unbiased anymore. Therefore, we only consider the hypothesis testing to compare treatment effects between  $\mu_1$  and  $\mu_2$ , which is described below.

$$H_0 : \mu_1 - \mu_2 = 0 \text{ versus } H_A : \mu_1 - \mu_2 \neq 0. \quad (3.4)$$

The corresponding test statistic for (3.4) is

$$T = \frac{\mathbf{L}\hat{\boldsymbol{\beta}}}{(\hat{\sigma}^2 \mathbf{L}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{L}^\top)^{1/2}}, \quad (3.5)$$

where  $\mathbf{L} = (1, -1, 0, \dots, 0)$  and  $\hat{\sigma}^2 = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) / (N - p - 2)$ . If  $|T| > Z_{1-\alpha/2}$ , where  $Z_{1-\alpha/2}$  is  $(1 - \alpha/2)$  quantile of a standard normal distribution, we will reject the null hypothesis, otherwise accept the null hypothesis.

The imbalances of different levels (within-stratum imbalance, marginal imbalance, overall imbalance) are defined exactly the same as in Chapter 2. Let  $C = \{j \mid Z_j \text{ is continuous}, j = 1, \dots, q\}$  and  $C^* = \{k \mid X_k \text{ is continuous}, k = 1, \dots, p\}$ . If  $k \in C^*$  or  $j \in C$ , the covariate-adaptive design is applied with respect to discrete variables,  $d_k^*(X_k)$  or  $d_j(Z_j)$ , where  $d_k^*$ ,  $d_j$  are discrete functions. In such case, define  $\delta_{i,k}^* = X_{i,k} - \mathbb{E}[X_{i,k} | d_k^*(X_{i,k})]$  and  $\delta_{i,j} = Z_{i,j} - \mathbb{E}[Z_{i,j} | d_j(Z_{i,j})]$ . Let

$$\tilde{X}_k = \begin{cases} X_k, & \text{if } k \notin C^* \\ d_k^*(X_k), & \text{if } k \in C^* \end{cases}$$

and

$$\tilde{Z}_j = \begin{cases} Z_j, & \text{if } j \notin C \\ d_j(Z_j), & \text{if } j \in C \end{cases}.$$

Here,  $\tilde{X}_{i,k}$  and  $\tilde{Z}_{i,j}$  are  $i$ th observations of covariates  $\tilde{X}_k$  and  $\tilde{Z}_j$ ,  $k = 1, \dots, p$  and  $j = 1, \dots, q$ .  $\tilde{X}_{i,k}$  and  $\tilde{Z}_{i,j}$  are used in the covariate-adaptive randomization process. We further define three levels of imbalance between patients in two treatments. Consider  $\tilde{X}_k$  have  $s_k^*$  levels and  $\tilde{Z}_j$  have  $s_j$  levels, resulting in  $\prod_{k=1}^p s_k^* \prod_{j=1}^q s_j$  strata in total. Let  $W_i = (\tilde{X}_{i,1}, \dots, \tilde{X}_{i,p}, \tilde{Z}_{i,1}, \dots, \tilde{Z}_{i,q})$  represents the covariate profile of the  $i$ th patient used in randomization, i.e.,  $W_i = (x_1^{t_1}, x_2^{t_2}, \dots, x_p^{t_p}, z_1^{r_1}, z_2^{r_2}, \dots, z_q^{r_q})$  if  $\tilde{X}_{i,k}$  is at level  $x_k^{t_k}$  and  $\tilde{Z}_{i,j}$  is at level  $z_j^{r_j}$ . For convenience, we use  $(t_1, t_2, \dots, t_p, r_1, r_2, \dots, r_q)$  to denote the stratum formed by patients who have the same covariate profile  $(x_1^{t_1}, x_2^{t_2}, \dots, x_p^{t_p}, z_1^{r_1}, z_2^{r_2}, \dots, z_q^{r_q})$ , and use  $(k; t_k)$  to denote the margin formed by patients with  $\tilde{X}_k = x_k^{t_k}$ , and similarly  $(j, r_j)$  to denote the margin formed by patients with  $\tilde{Z}_j = z_j^{r_j}$ . Then let

- $D_N$  be the difference between the numbers of patients in treatment group 1 and 2 as total, i.e., the number in group 1 minus the number in group 2;
- $D_N(k; t_k)$  and  $D_N(j; r_j)$  be the differences between the numbers of patients in the two treatment groups on the margin  $(k; t_k)$  and  $(j, r_j)$ , respectively;
- $D_N(t_1, t_2, \dots, t_p, r_1, r_2, \dots, r_q)$  be the difference between the numbers of patients in the two treatment groups within the stratum  $(t_1, t_2, \dots, t_p, r_1, r_2, \dots, r_q)$ .

### 3.3 Theoretical properties

In this section, the theoretical properties of statistical inference are given under the framework described in Section 3.2 where covariates are assumed to be correlated. In particular, point estimators and hypothesis testing are studied based on the working model (3.3) where outcomes of patients follow the underlying true model (3.1) under covariate-adaptive designs. The inference of both treatment effects and covariate effects are of interest. The following theorem concerns about point estimators.

**Theorem 3.3.1.** *Suppose that a covariate-adaptive design satisfies following two conditions:*

(A) *the overall imbalance is bounded, that is,  $D_N = O_p(1)$ ;*

(B) *the marginal imbalances for all covariates are bounded in probability, that is,*

$$D_N(k; t_k) = O_p(1) \text{ and } D_N(j; r_j) = O_p(1), \quad k = 1, 2, \dots, p, \quad j = 1, 2, \dots, q.$$

*Then*

(i)  $\hat{\mu}_1 - \hat{\mu}_2$  *is a consistent estimator of  $\mu_1 - \mu_2$ , i.e.,*

$$\hat{\mu}_1 - \hat{\mu}_2 \xrightarrow{P} \mu_1 - \mu_2.$$

(ii)  $(\hat{\beta}_1, \dots, \hat{\beta}_p)^\top$  *is a biased estimator of  $(\beta_1, \dots, \beta_p)^\top$  with bias  $\Sigma_{x,x}^{-1} \Sigma_{x,z} \gamma$ , i.e.,*

$$(\hat{\beta}_1, \dots, \hat{\beta}_p)^\top - (\beta_1, \dots, \beta_p)^\top \xrightarrow{P} \Sigma_{x,x}^{-1} \Sigma_{x,z} \gamma.$$

**REMARK 3.3.1.** *Theorem 3.3.1 implies that the estimator of the difference between treatment effects is unbiased no matter covariates are correlated or not. In this*

sense, we can always achieve consistent estimator of treatment effects despite the fact that we may not fully understand the correlation structure between covariates and just omit several important covariates in analysis.

**REMARK 3.3.2.** *On the other hand, the estimators of covariates' coefficients are biased if only partial covariate information is used under covariate-adaptive designs by Theorem 3.3.1. For some studies where the interest is to identify influential covariates on outcomes, such as biomarker finding studies, we need hesitate to make conclusions about estimates of covariate effects due to bias. Notice that the estimators are also biased under complete randomization if the omitted covariates are correlated with the covariates in the analysis model. In other word, the estimators of covariates have similar behaviors under covariate-adaptive designs and complete randomization.*

**Theorem 3.3.2.** *Suppose that a covariate-adaptive design satisfies following two conditions:*

(A) *the overall imbalance is bounded, that is,  $D_N = O_p(1)$ ;*

(B) *the marginal imbalances for all covariates are bounded in probability, that is,*

$$D_N(k; t_k) = O_p(1) \text{ and } D_N(j; r_j) = O_p(1), \quad k = 1, 2, \dots, p, \quad j = 1, 2, \dots, q.$$

Then

(i) *under  $H_0 : \mu_1 - \mu_2 = 0$ ,*

$$T \xrightarrow{D} \mathbf{N}(0, \tau^2), \quad \tau^2 = \frac{\sigma_\varepsilon^2 + \mathbb{E}[\text{Var}((\mathcal{Z}^\top - \mathcal{X}^\top \Sigma_{x,x}^{-1} \Sigma_{x,z}) \gamma | \tilde{\mathbf{D}})]}{\sigma_\varepsilon^2 + \text{Var}[(\mathcal{Z}^\top - \mathcal{X}^\top \Sigma_{x,x}^{-1} \Sigma_{x,z}) \gamma]} = \frac{\sigma_1^2}{\sigma_2^2}, \quad (3.6)$$

where  $\sigma_1^2 = \sigma_\varepsilon^2 + \mathbb{E}[\text{Var}((\mathcal{Z}^\top - \mathcal{X}^\top \Sigma_{x,x}^{-1} \Sigma_{x,z}) \gamma | \tilde{\mathbf{D}})]$ ,  $\sigma_2^2 = \sigma_\varepsilon^2 + \text{Var}[(\mathcal{Z}^\top - \mathcal{X}^\top \Sigma_{x,x}^{-1} \Sigma_{x,z}) \gamma]$  and  $\tilde{\mathbf{D}} = \{\tilde{X}_k, \tilde{Z}_j, k = 1, 2, \dots, p, j = 1, 2, \dots, q\}$ .

(ii) under  $H_A : \mu_1 - \mu_2 \neq 0$ , consider a sequence of local alternatives, i.e.,  $\mu_1 - \mu_2 = \delta/\sqrt{N}$  for a fixed  $\delta \neq 0$ , then

$$T \xrightarrow{D} \mathbf{N}(\Delta, \tau^2), \quad \Delta = \frac{\delta}{2\sigma_2}. \quad (3.7)$$

REMARK 3.3.3. Notice that  $\sigma_1^2 < \sigma_2^2$  in general, the test statistic under the null hypothesis has an asymptotic normal distribution with a mean zero and a standard deviation smaller than 1. Hence, the hypothesis testing to compare treatment effects is conservative in terms of a smaller Type I error than the nominal level, which means it is more difficult to reject the null hypothesis under the null hypothesis. This conclusion is consistent with the result of Theorem 2.3.1.

REMARK 3.3.4. Theorem 3.3.2 includes Theorem 2.3.1 as a special case. If the covariance matrix  $\Sigma$  is a diagonal matrix, then the results of Theorem 3.3.2 is equivalent to Theorem 2.3.1 where all covariates are independent. In fact, it can be shown that the asymptotic distributions of testing statistics derived in this chapter are identical to the distributions in Theorem 2.3.1 either under the null hypothesis or the alternative hypothesis.

Based on the asymptotic distribution under the alternative hypothesis, the power of hypothesis testing (3.4) can be studied and compared to the power under complete randomization. Similar to the power comparison discussed in Section 2.3, the power of hypothesis testing (3.4) under covariate-adaptive designs is,

$$\Pr(|T| > Z_{1-\alpha/2}) = \Phi\left(\frac{\delta}{2\sigma_1} - \frac{\sigma_2 Z_{1-\alpha/2}}{\sigma_1}\right) + \Phi\left(-\frac{\delta}{2\sigma_1} - \frac{\sigma_2 Z_{1-\alpha/2}}{\sigma_1}\right) + o(1),$$

and the power of hypothesis testing (3.4) under complete randomization (with the

same setting as described in Section 3.2) is,

$$\Pr(|T| > Z_{1-\alpha/2}) = \Phi\left(\frac{\delta}{2\sigma_2} - Z_{1-\alpha/2}\right) + \Phi\left(-\frac{\delta}{2\sigma_2} - Z_{1-\alpha/2}\right) + o(1).$$

These power expressions are similar to those in Chapter 2, despite the definitions of  $\sigma_1$  and  $\sigma_2$  are somehow different. Therefore, it can be concluded that the power of hypothesis testing (3.4) under covariate-adaptive designs is smaller than that under complete randomization when the difference of treatment effects is relatively small, which becomes larger than complete randomization when the difference of treatment effects is relatively large. Simulation studies are carried out in the next section to illustrate these conclusions numerically.

Theorem 3.3.1 and Theorem 3.3.2 hold under the same conditions stated in Theorem 2.3.1: condition (A) the overall imbalance is bounded in probability; condition (B) all marginal imbalances are all bounded in probability. Then we have the following corollary similar to Corollary 2.3.1.

**Corollary 3.3.1.** *Both Theorem 3.3.1 and Theorem 3.3.2 holds under the following covariate-adaptive designs:*

- (i) *Pocock and Simon's marginal procedures (Pocock and Simon, 1975);*
- (ii) *stratified permuted block designs; and*
- (iii) *the class of covariate-adaptive designs proposed by Hu and Hu (2012).*

The proofs of Theorem 3.3.1 and Theorem 3.3.2 are similar to Theorem 2.3.1 and its corresponding lemmas, but need consider the correlation structure in covariates. See the Appendix for the detailed proofs.

### 3.4 Simulation study

*Case 1: Testing Treatment Effects.* In this case, Type I error of hypothesis testing (3.4) for comparing treatment effects is studied under three randomization design methods: Pocock and Simon's marginal procedure, stratified permuted block design and complete randomization. Both continuous case and discrete case are considered under each design. The following linear model with two covariates  $Z_1$  and  $Z_2$  is assumed to be the underlying model,

$$Y_i = \mu_1 I_i + \mu_2(1 - I_i) + \beta_1 Z_{i,1} + \beta_2 Z_{i,2} + \varepsilon_i,$$

where  $\varepsilon_i$  is distributed as  $\mathbf{N}(0, 1)$ ,  $\beta_1 = \beta_2 = 1$ . No difference in treatment effects is assumed here to study Type I error, i.e.,  $\mu_1 = \mu_2$ . For the discrete case,  $Z_1$  and  $Z_2$  follow Bernoulli(0.5) with the correlation between  $Z_1$  and  $Z_2$  equal to  $\rho$ ; for the continuous case,  $Z_1$  and  $Z_2$  follow a standard normal distribution  $\mathbf{N}(0, 1)$  with the correlation between  $Z_1$  and  $Z_2$  equal to  $\rho$ . When covariates  $Z_1$  and  $Z_2$  are continuous, they are discretized to bernoulli variables  $Z'_1$  and  $Z'_2$  with the probability 0.5 to be used in randomization. Specifically, if  $Z_1 < Z_{(0.5)}$ , where  $Z_{(0.5)}$  is the 0.5 quantile of a standard normal distribution, then  $Z'_1 = 0$ , otherwise  $Z'_1 = 1$ . Similar process applies on  $Z_2$  to obtain the discrete variable  $Z'_2$ , then  $Z'_1$  and  $Z'_2$  are used in the corresponding covariate-adaptive design. Original variables  $Z_1$  and  $Z_2$  (without discretization) are used in statistical inference procedures.

To carry out simulations, the same setting of parameters is used as in Chapter 2. In Pocock and Simon's marginal procedures, equal weights are assigned to two covariates and the biased coin probability is equal to 0.75. The block size 4 is used for stratified permuted block design. The significance level is  $\alpha = 0.05$  and sample

Table 3.1: Simulated Type I error for Pocock and Simon’s marginal procedure (PS), stratified permuted block design (SPB) and complete randomization (CR) in % with  $\rho = 0.5$ . Simulations based on 10000 runs.

$Z$	Method	$N$	$t$ -test	$lm(Z_1)$	$lm(Z_2)$	$lm(Z_1, Z_2)$	$BS$ - $t$
Discrete	<b>PS</b>	100	1.07	3.34	3.40	4.90	5.24
		200	0.82	3.27	3.23	4.98	5.17
		500	0.82	3.12	3.27	4.88	4.89
	<b>SPB</b>	100	1.10	3.42	3.63	5.15	5.92
		200	1.07	3.36	3.37	5.11	5.57
		500	0.86	3.38	3.40	5.13	5.69
	<b>CR</b>	100	5.05	4.97	4.99	5.08	-
		200	5.27	5.10	5.44	5.36	-
		500	4.90	4.98	4.89	4.67	-
Continuous	<b>PS</b>	100	0.50	2.80	2.85	5.08	5.60
		200	0.48	2.63	2.71	4.69	5.24
		500	0.44	2.81	2.61	5.24	5.09
	<b>SPB</b>	100	0.44	2.69	2.73	4.85	5.35
		200	0.41	2.58	2.61	5.23	5.84
		500	0.48	2.71	2.65	5.35	5.27
	<b>CR</b>	100	5.12	5.24	4.88	4.92	-
		200	4.69	5.14	5.07	5.46	-
		500	4.73	4.95	4.78	4.77	-

size  $N$  is 100, 200 or 500. The hypothesis tests are based on the two sample  $t$ -test ( $t$ -test), the linear model with a single covariate  $Z_1$  ( $lm(Z_1)$ ), the linear model with a single covariate  $Z_2$  ( $lm(Z_2)$ ), the linear model with both covariate  $Z_1$  and  $Z_2$  ( $lm(Z_1, Z_2)$ ) and the Bootstrap  $t$ -test ( $BS$ - $t$ ) with  $B = 500$ . The simulation results of  $\rho = 0.5$  are presented in Table 3.1 for Pocock and Simon’s marginal procedure, stratified permuted block design and complete randomization.

Several conclusions can be drawn from Table 3.1. (1) The Type I error is close to 5% under the full model  $lm(Z_1, Z_2)$  for all three randomization. Such result is expected because  $lm(Z_1, Z_2)$  captures the underlying true model. (2) Under the two types of covariate-adaptive designs considered, the two sample  $t$ -test,  $lm(Z_1)$



Table 3.2: Simulated Type I error for Pocock and Simon's marginal procedure (PS) and complete randomization (CR) in % with different values of  $\rho$ . Simulations based on 10000 runs and sample size  $N = 100$ .

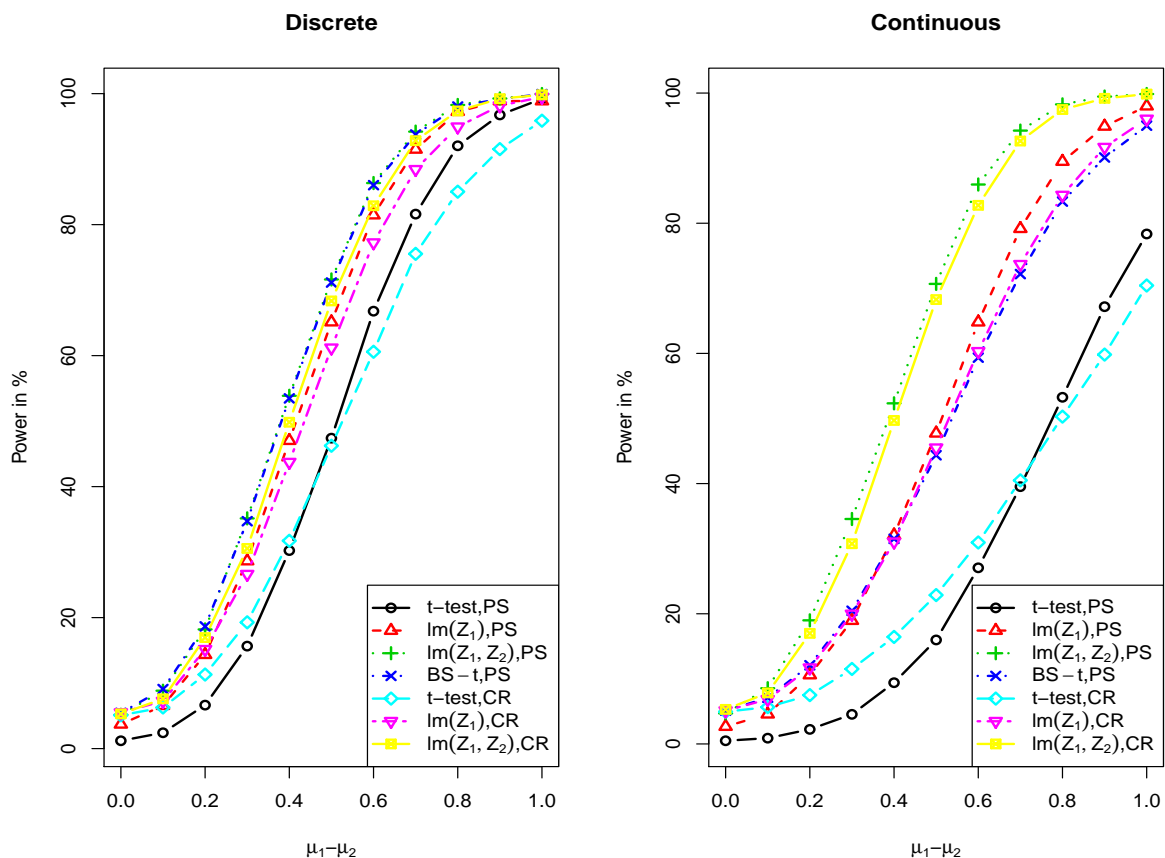
$Z$	Method	$\rho$	$t$ -test	$lm(Z_1)$	$lm(Z_2)$	$lm(Z_1, Z_2)$	$BS$ - $t$
Discrete	<b>PS</b>	0.8	0.71	4.25	4.22	4.98	5.23
		0.5	1.12	3.72	3.51	5.00	5.40
		0.2	1.46	3.05	2.93	4.78	5.16
		0	1.98	2.95	3.13	5.11	5.39
		-0.2	2.37	3.04	3.23	5.21	5.41
		-0.5	3.32	3.65	3.63	5.24	5.57
		-0.8	4.19	4.22	4.19	4.83	5.21
	<b>CR</b>	0.8	5.21	4.99	5.02	4.90	-
		0.5	4.94	4.87	4.86	4.67	-
		0.2	4.90	4.60	5.20	4.98	-
		0	4.93	4.77	4.97	4.76	-
		-0.2	4.60	4.80	4.84	5.02	-
		-0.5	5.03	5.06	5.04	5.24	-
		-0.8	4.82	4.84	4.94	5.02	-
Continuous	<b>PS</b>	0.8	0.34	4.22	4.23	5.01	5.38
		0.5	0.53	2.43	2.86	4.90	5.15
		0.2	0.78	2.28	2.07	4.71	5.05
		0	0.12	1.98	2.00	5.21	5.17
		-0.2	1.66	2.00	1.96	4.75	5.27
		-0.5	2.83	2.88	2.89	5.19	5.44
		-0.8	4.13	4.08	4.05	4.78	5.45
	<b>CR</b>	0.8	4.83	4.70	4.95	4.79	-
		0.5	5.06	5.02	5.35	5.04	-
		0.2	5.25	5.04	5.06	4.99	-
		0	4.66	4.76	4.89	4.69	-
		-0.2	4.87	4.72	5.11	4.94	-
		-0.5	4.99	4.95	5.15	5.36	-
		-0.8	4.92	4.84	4.79	4.85	-

and  $lm(Z_2)$  are all conservative with Type I error less than 5%. These findings are consistent to Theorem 3.3.2 with regards to hypothesis testing for linear models with omitted covariates. (3) Under covariate-adaptive designs, the two sample  $t$ -test is the most conservative one with the least Type I error. However, the bootstrap  $t$ -test has a Type I error close to the nominal level 5%, which provides a way to capture the true Type I error when the covariates are correlated. (4) Under complete randomization, the Type I error is close to 5% for all four tests.

Type I error for more values of  $\rho$  ( $0, \pm 0.2, \pm 0.5, \pm 0.8$ ) are examined in Table 3.2, in which all the results are similar to Table 3.1 and are consistent with Theorem 3.3.2. The results of stratified permuted block design are similar to Pocock and Simon's marginal procedure and are not reported here. In Case 2 with regards to power comparison, only  $\rho = 0.5$  is used to study power of hypothesis testing based on linear models with correlated covariates under covariate-adaptive designs.

*Case 2: Power Comparison.* Now we compare power for different hypothesis testing methods under Pocock and Simon's marginal procedure and complete randomization. The same model is used as in Case 1, but difference exists between treatment effects  $\mu_1$  and  $\mu_2$ , i.e.,  $\mu_1 - \mu_2 \neq 0$ . Sample size  $N = 100$  and  $\rho = 0.5$  are used in simulations. The results of power comparison are given in Figure 3.1, and the conclusions are summarized below. (1) Under Pocock and Simon's marginal procedure, the two sample  $t$ -test is less powerful than  $(lm(Z_1))$  or  $(lm(Z_2))$ , and all those three methods are less powerful than  $(lm(Z_1, Z_2))$ . Therefore, including more covariates into analysis leads to larger power and allows us to detect difference between treatment effects more efficiently. (2) The bootstrap  $t$ -test has similar power performance as the full model when covariates are discrete, since they are using the same amount information of covariates. However, the bootstrap  $t$ -test is less powerful than  $(lm(Z_1, Z_2))$  if covariates are continuous because there is loss of

Figure 3.1: Simulated power for Pocock and Simon’s marginal procedure (PS) and complete randomization (CR) under discrete case and continuous case with  $\rho = 0.5$ . Simulation based on 10000 runs and number of patients  $N = 100$ .



information due to discretization. (3) Compared to complete randomization, the two sample  $t$ -test under Pocock and Simon’s marginal procedure has smaller power when the treatment difference is relatively small, and has larger power when the treatment difference becomes larger. We also tried stratified permuted block randomization (not reported here), all tests considered have similar power performance as Pocock and Simon’s marginal procedure. Combined the results in Figure 3.1 and Table 3.1, 3.2, the theoretical properties in Theorem 3.3.2 hold for sample size around 100.

### 3.5 Conclusion

In this chapter we proposed theoretical properties of statistical inference for linear models with correlated covariates. Under the assumption that the covariates are correlated instead of independent of each other, the asymptotic distributions of testing statistics are derived under both the null hypothesis and the alternative hypothesis. Based on the asymptotic distributions, we show that the hypothesis testing for comparing treatment effects is conservative with smaller Type I error than the nominal level if partial covariate information is not incorporated in analysis. This conclusion is similar to the results of independent case in Chapter 2. Theorem 3.3.2, together with Theorem 2.3.1 provides theoretical basis to understand the hypothesis testing for treatment effects under covariate-adaptive designs.

The properties of covariate effects are affected by whether covariates are independent or not. If covariates used in randomization are correlated and important covariates are omitted from analysis, the estimators of covariates' coefficients are biased, which cause the hypothesis testing not valid anymore. Therefore, we should be careful when we conduct statistical inference for covariates under covariate-adaptive designs because we may not be aware of whether covariates are independent or not in practice.

In this chapter, the same conditions are used as in Chapter 2 to discuss statistical inference for covariate-adaptive designs, which are (A) overall imbalance is bounded in probability; (B) marginal imbalances are bounded in probability. These conditions are relatively mild and satisfied by a broad family of covariate-adaptive designs, including commonly used designs such as stratified permuted block design and Pocock and Simon's marginal procedure. Our work in Chapter 2 and 3 provides guidance on applications of linear models under covariate-adaptive designs.

### 3.6 Appendix: proof of theorems

**Lemma 3.6.1.** *Under the assumptions of Theorem 3.3.1, then*

1.  $\sum I_i/N \xrightarrow{P} 1/2$ ,  $\sum(1 - I_i)/N \xrightarrow{P} 1/2$ ;
2.  $\sum I_i X_{i,k}/N \xrightarrow{P} \mathbf{E}X_k/2$ ,  $\sum(1 - I_i)X_{i,k}/N \xrightarrow{P} \mathbf{E}X_k/2$ ,  $\sum I_i Z_{i,j}/N \xrightarrow{P} \mathbf{E}Z_j/2$ ,  
and  $\sum(1 - I_i)Z_{i,j}/N \xrightarrow{P} \mathbf{E}Z_j/2$ .

*Proof.* The proof is the same as for the proof of Lemma 2.6.1 in Chapter 2.  $\square$

*Proof of Theorem 3.3.1.* By the definition of the OLS estimator,

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left(\frac{\mathbf{X}^\top \mathbf{X}}{N}\right)^{-1} \frac{\mathbf{X}^\top \mathbf{Z} \boldsymbol{\gamma}}{N} + \left(\frac{\mathbf{X}^\top \mathbf{X}}{N}\right)^{-1} \frac{\mathbf{X}^\top \boldsymbol{\varepsilon}}{N}.$$

Notice  $\mathbf{E}X_k = \mathbf{E}Z_j = 0$  for all  $k$  and  $j$ , by the weak law of large numbers and Lemma 3.6.1,

$$\begin{aligned} \frac{1}{N} \mathbf{X}^\top \mathbf{X} &= \frac{1}{N} \begin{bmatrix} \sum I_i & 0 & \sum I_i X_{i,1} & \cdots & \sum I_i X_{i,p} \\ 0 & \sum(1 - I_i) & \sum(1 - I_i)X_{i,1} & \cdots & \sum(1 - I_i)X_{i,p} \\ \sum X_{i,1}I_i & \sum X_{i,1}(1 - I_i) & \sum X_{i,1}X_{i,1} & \cdots & \sum X_{i,1}X_{i,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum X_{i,p}I_i & \sum X_{i,p}(1 - I_i) & \sum X_{i,p}X_{i,1} & \cdots & \sum X_{i,p}X_{i,p} \end{bmatrix} \\ &\xrightarrow{P} \text{diag}\left(\frac{1}{2}, \frac{1}{2}, \boldsymbol{\Sigma}_{x,x}\right) \end{aligned}$$

and

$$\frac{1}{N} \mathbf{X}^\top \mathbf{Z} = \frac{1}{N} \begin{bmatrix} \sum I_i Z_{i,1} & \cdots & \sum I_i Z_{i,q} \\ \sum (1 - I_i) Z_{i,1} & \cdots & \sum (1 - I_i) Z_{i,q} \\ \sum X_{i,1} Z_{i,1} & \cdots & \sum X_{i,1} Z_{i,q} \\ \vdots & \ddots & \vdots \\ \sum X_{i,p} Z_{i,1} & \cdots & \sum X_{i,p} Z_{i,q} \end{bmatrix} \xrightarrow{P} \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{\Sigma}_{x,z} \end{bmatrix}.$$

Further, by the independence of  $I_i$  and  $\varepsilon_i$ , for  $i = 1, \dots, N$ ,

$$\frac{1}{N} \mathbf{X}^\top \boldsymbol{\varepsilon} = \frac{1}{N} \begin{bmatrix} \sum I_i \varepsilon_i \\ \sum (1 - I_i) \varepsilon_i \\ \sum X_{i,1} \varepsilon_i \\ \vdots \\ \sum X_{i,p} \varepsilon_i \end{bmatrix} \xrightarrow{P} \mathbf{0}.$$

Hence,

$$\left(\frac{\mathbf{X}^\top \mathbf{X}}{N}\right)^{-1} \frac{\mathbf{X}^\top \mathbf{Z} \boldsymbol{\gamma}}{N} + \left(\frac{\mathbf{X}^\top \mathbf{X}}{N}\right)^{-1} \frac{\mathbf{X}^\top \boldsymbol{\varepsilon}}{N} \xrightarrow{P} \begin{bmatrix} 0 \\ 0 \\ \boldsymbol{\Sigma}_{x,x}^{-1} \boldsymbol{\Sigma}_{x,z} \boldsymbol{\gamma} \end{bmatrix},$$

that is

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \xrightarrow{P} \begin{bmatrix} 0 \\ 0 \\ \boldsymbol{\Sigma}_{x,x}^{-1} \boldsymbol{\Sigma}_{x,z} \boldsymbol{\gamma} \end{bmatrix}.$$

It follows that,

$$\hat{\mu}_1 - \hat{\mu}_2 \xrightarrow{P} \mu_1 - \mu_2$$

and

$$(\hat{\beta}_1, \dots, \hat{\beta}_p)^\top - (\beta_1, \dots, \beta_p)^\top \xrightarrow{P} \Sigma_{x,x}^{-1} \Sigma_{x,z} \boldsymbol{\gamma},$$

then the proof is complete.  $\square$

**Lemma 3.6.2.** *Suppose  $\mathbf{a} = (a_1, \dots, a_q)^\top$  is a constant vector consisting of  $q$  elements,  $\mathbf{b} = (b_1, \dots, b_p)^\top$  is a constant vector consisting of  $p$  elements, and  $c$  is a constant number, then*

$$\frac{1}{\sqrt{N}} (2\mathbf{I} - \mathbf{1})^\top (\mathbf{Z}\mathbf{a} + \mathbf{X}\mathbf{b} + c\boldsymbol{\varepsilon}) \xrightarrow{D} \mathbf{N}(0, \mathbf{E}[\text{Var}(\mathcal{Z}^\top \mathbf{a} + \mathcal{X}^\top \mathbf{b} | \tilde{\mathbf{D}})] + c^2 \sigma_\varepsilon^2),$$

*Proof.*

$$\begin{aligned} (2\mathbf{I} - \mathbf{1})(\mathbf{Z}\mathbf{a} + \mathbf{X}\mathbf{b} + c\boldsymbol{\varepsilon}) &= \sum_{j \in C} \sum_i (2I_i - 1) a_j \delta_{i,j} + \sum_{j \in C} \sum_i (2I_i - 1) a_j \mathbf{E}[Z_{i,j} | d_j(Z_{i,j})] \\ &+ \sum_{j \notin C} \sum_i (2I_i - 1) a_j Z_{i,j} + \sum_{k \in C^*} \sum_i (2I_i - 1) b_k \delta_{i,k}^* + \sum_{k \in C^*} \sum_i (2I_i - 1) b_k \mathbf{E}[X_{i,k} | d_k^*(X_{i,k})] \\ &+ \sum_{k \notin C^*} \sum_k (2I_i - 1) b_k X_{i,k} + \sum_i (2I_i - 1) c \varepsilon_i \end{aligned}$$

Apply condition (B) to all margins with respect to each covariate  $Z_j, j = 1, 2, \dots, q$ ,

we have

$$\sum_i (2I_i - 1) a_j Z_{i,j} = \sum_{r_j=1}^{s_j} D(j; r_j) a_j z_j^{r_j} = O_p(1), \quad j \notin C,$$

and

$$\sum_i (2I_i - 1)a_j \mathbb{E}[Z_{i,j}|d_j(Z_{i,j})] = \sum_{r_j=1}^{s_j} D(j; r_j) a_j \mathbb{E}[Z_j|d_j(Z_j) = z_j^{r_j}] = O_p(1), \quad j \in C.$$

It follows that

$$\sum_{j \in C} \sum_i (2I_i - 1)a_j \mathbb{E}[Z_{i,j}|d_j(Z_{i,j})] + \sum_{j \notin C} \sum_i (2I_i - 1)a_j Z_{i,j} = O_p(1).$$

Similarly,

$$\sum_{k \in C^*} \sum_i (2I_i - 1)b_k \mathbb{E}[X_{i,k}|d_j^*(X_{i,k})] + \sum_{k \notin C^*} \sum_i (2I_i - 1)b_k X_{i,k} = O_p(1).$$

Define  $\tilde{\mathbf{D}}_N = \{\tilde{X}_{i,k}, \tilde{Z}_{i,j}, i = 1, 2, \dots, N, j = 1, 2, \dots, q, k = 1, 2, \dots, p\}$  and  $\tilde{I} = \{I_i, i = 1, \dots, N\}$ , then  $(\delta_{i,j}, \delta_{i,k}^*, \varepsilon_i)$  are conditionally independent of  $\tilde{I}$  given  $\tilde{\mathbf{D}}_N$ . In addition,  $\mathbb{E}(\varepsilon_i|\tilde{\mathbf{D}}_N) = \mathbb{E}(\varepsilon_i) = 0$ ,  $\mathbb{E}(\delta_{i,j}|\tilde{\mathbf{D}}_N) = \mathbb{E}[\delta_{i,j}|d_j(Z_{i,j})] = 0$  by the definition of  $\delta_{i,j}$ , and similarly  $\mathbb{E}(\delta_{i,k}^*|\tilde{\mathbf{D}}_N) = 0$ . Then

$$\frac{1}{\sqrt{N}} \mathbb{E}\left(\sum_{j \in C} \sum_i (2I_i - 1)a_j \delta_{i,j} + \sum_{k \in C^*} \sum_i (2I_i - 1)b_k \delta_{i,k}^* + \sum_i (2I_i - 1)c \varepsilon_i | \tilde{\mathbf{D}}_N\right) = 0.$$

Further, notice the fact that  $(\delta_{i,j}, \delta_{i,k}^*, \varepsilon_i)$  and  $\tilde{I}$  are independent of each other given  $\tilde{\mathbf{D}}_N$  and  $(2I_i - 1)^2 = 1$ ,

$$\text{Cov}\left(\frac{1}{\sqrt{N}} \left(\sum_{j \in C} \sum_i (2I_i - 1)a_j \delta_{i,j} + \sum_{k \in C^*} \sum_i (2I_i - 1)b_k \delta_{i,k}^*\right), \frac{1}{\sqrt{N}} \sum_i (2I_i - 1)c \varepsilon_i | \tilde{\mathbf{D}}_N\right) = 0$$



and

$$\text{Var}\left(\frac{1}{\sqrt{N}} \sum_i (2I_i - 1)c\varepsilon_i | \tilde{\mathbf{D}}_N\right) = \text{E}\left(\frac{1}{N} \sum_i (2I_i - 1)^2 c^2 \varepsilon_i^2 | \tilde{\mathbf{D}}_N\right) = c^2 \sigma_\varepsilon^2$$

and

$$\begin{aligned} & \text{Var}\left(\frac{1}{\sqrt{N}} \left( \sum_{j \in C} \sum_i (2I_i - 1)a_j \delta_{i,j} + \sum_{k \in C^*} \sum_i (2I_i - 1)b_k \delta_{i,k}^* \right) | \tilde{\mathbf{D}}_N\right) \\ &= \frac{1}{N} \text{E}\left( \sum_{j \in C} \sum_{j' \in C} \sum_i (2I_i - 1)^2 a_j a_{j'} \delta_{i,j} \delta_{i,j'} + \sum_{k \in C^*} \sum_{k' \in C^*} \sum_i (2I_i - 1)^2 b_k b_{k'} \delta_{i,k}^* \delta_{i,k'}^* \right. \\ & \quad \left. + \sum_{j \in C} \sum_{k \in C^*} \sum_i (2I_i - 1)^2 a_j b_k \delta_{i,j} \delta_{i,k}^* | \tilde{\mathbf{D}}_N\right) = V, \end{aligned}$$

where  $V = N^{-1} \sum_i \text{Var}\left(\sum_j \delta_{i,j} a_j + \sum_k \delta_{i,k}^* b_k | \tilde{\mathbf{D}}_N\right)$ . Therefore,

$$\text{Var}\left(\frac{1}{\sqrt{N}} \left( \sum_{j \in C} \sum_i (2I_i - 1)\gamma_j \delta_{i,j} + \sum_{k \in C^*} \sum_i (2I_i - 1)b_k \delta_{i,k}^* + \sum_i (2I_i - 1)c\varepsilon_i \right) | \tilde{\mathbf{D}}_N\right) = c^2 \sigma_\varepsilon^2 + V$$

By the central limit theorem, given  $(\tilde{I}, \tilde{\mathbf{D}}_N)$ ,

$$\frac{1}{N^{1/2}} \left\{ \sum_{j \in C} \sum_i (2I_i - 1)a_j \delta_{i,j} + \sum_{k \in C^*} \sum_i (2I_i - 1)b_k \delta_{i,k}^* + \sum_i (2I_i - 1)c\varepsilon_i \right\}$$

is asymptotically normal with mean zero and variance  $(c^2 \sigma_\varepsilon^2 + V)$ , which converges to  $(c^2 \sigma_\varepsilon^2 + \text{E}[\text{Var}(\mathbf{Z}^\top \mathbf{a} + \mathbf{X}^\top \mathbf{b} | \tilde{\mathbf{D}})])$ , where  $\tilde{\mathbf{D}} = \{\tilde{X}_k, \tilde{Z}_j, j = 1, 2, \dots, q, k = 1, 2, \dots, p\}$ . Since the limiting distribution is independent of  $(\tilde{I}, \tilde{\mathbf{D}}_N)$ , the convergence in distribution also holds unconditionally. Hence, by Slutsky's Theorem,

$$\frac{1}{\sqrt{N}} (\mathbf{2I} - \mathbf{1})^\top (\mathbf{Z}\mathbf{a} + \mathbf{X}\mathbf{b} + c\varepsilon) \xrightarrow{D} \mathbf{N}(0, \text{E}[\text{Var}(\mathbf{Z}^\top \mathbf{a} + \mathbf{X}^\top \mathbf{b} | \tilde{\mathbf{D}})] + c^2 \sigma_\varepsilon^2),$$

which completes the proof of Lemma 3.6.2. □

*Proof of Theorem 3.3.2.* The following test statistic is used for hypothesis testing (3.4),

$$\frac{\mathbf{L}\hat{\boldsymbol{\beta}}}{(\hat{\sigma}^2\mathbf{L}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{L}^\top)^{1/2}},$$

where  $\mathbf{L} = (1, -1, 0, \dots, 0)$  and  $\hat{\sigma}^2 = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})/(N - p - 2)$ ,  $p$  is the total number of independent variables in the model (3.3) besides  $\mu_1$  and  $\mu_2$ .

The numerator is equal to

$$\mathbf{L}\hat{\boldsymbol{\beta}} = \hat{\mu}_1 - \hat{\mu}_2 = \mu_1 - \mu_2 + \mathbf{L}\left(\frac{\mathbf{X}^\top\mathbf{X}}{N}\right)^{-1}\frac{\mathbf{X}^\top\mathbf{Z}\boldsymbol{\gamma}}{N} + \mathbf{L}\left(\frac{\mathbf{X}^\top\mathbf{X}}{N}\right)^{-1}\frac{\mathbf{X}^\top\boldsymbol{\varepsilon}}{N}.$$

We have shown that in the proof of Theorem 3.3.2,

$$\frac{1}{N}\mathbf{X}^\top\mathbf{X} \xrightarrow{P} \text{diag}\left(\frac{1}{2}, \frac{1}{2}, \boldsymbol{\Sigma}_{x,x}\right) \doteq \mathbf{M}.$$

Let

$$\mathbf{S}_1 = \mathbf{L}\mathbf{M}^{-1}\left[\frac{\mathbf{X}^\top\mathbf{Z}\boldsymbol{\gamma}}{N} + \frac{\mathbf{X}^\top\boldsymbol{\varepsilon}}{N}\right]$$

and

$$\mathbf{S}_2 = \mathbf{L}\left[\left(\frac{\mathbf{X}^\top\mathbf{X}}{N}\right)^{-1} - \mathbf{M}^{-1}\right]\left[\frac{\mathbf{X}^\top\mathbf{Z}\boldsymbol{\gamma}}{N} + \frac{\mathbf{X}^\top\boldsymbol{\varepsilon}}{N}\right].$$

By some matrix calculations,

$$\mathbf{S}_1 = \frac{2}{N}\left(\sum_j \sum_i (2I_i - 1)\gamma_j Z_{i,j} + \sum_i (2I_i - 1)\varepsilon_i\right).$$

For  $\mathbf{S}_2$ , notice that

$$\frac{\mathbf{X}^\top \mathbf{Z} \boldsymbol{\gamma}}{N} + \frac{\mathbf{X}^\top \boldsymbol{\varepsilon}}{N} = \frac{1}{N} \begin{bmatrix} \sum_j \sum_i I_i Z_{i,j} \gamma_j + \sum_i I_i \varepsilon_i \\ \sum_j \sum_i (1 - I_i) Z_{i,j} \gamma_j + \sum_i (1 - I_i) \varepsilon_i \\ \sum_j \sum_i X_{i,1} Z_{i,j} \gamma_j + \sum_i X_{i,1} \varepsilon_i \\ \vdots \\ \sum_j \sum_i X_{i,p} Z_{i,j} \gamma_j + \sum_i X_{i,p} \varepsilon_i \end{bmatrix},$$

then  $\mathbf{S}_2$  can be written as,

$$\begin{aligned} \mathbf{S}_2 &= \mathbf{L} \left[ \left( \frac{\mathbf{X}^\top \mathbf{X}}{N} \right)^{-1} - \mathbf{M}^{-1} \right] \begin{bmatrix} 0 \\ 0 \\ \underline{\mathbf{XZ}} \boldsymbol{\gamma} / N \end{bmatrix} \\ &+ \mathbf{L} \left[ \left( \frac{\mathbf{X}^\top \mathbf{X}}{N} \right)^{-1} - \mathbf{M}^{-1} \right] \frac{1}{N} \begin{bmatrix} \sum_j \sum_i I_i Z_{i,j} \gamma_j + \sum_i I_i \varepsilon_i \\ \sum_j \sum_i (1 - I_i) Z_{i,j} \gamma_j + \sum_i (1 - I_i) \varepsilon_i \\ \sum_i X_{i,1} \varepsilon_i \\ \vdots \\ \sum_i X_{i,p} \varepsilon_i \end{bmatrix} \\ &= \mathbf{S}_{2,1} + \mathbf{S}_{2,2}. \end{aligned}$$

where

$$\mathbf{S}_{2,1} = \mathbf{L} \left[ \left( \frac{\mathbf{X}^\top \mathbf{X}}{N} \right)^{-1} - \mathbf{M}^{-1} \right] \begin{bmatrix} 0 \\ 0 \\ \underline{\mathbf{XZ}} \boldsymbol{\gamma} / N \end{bmatrix}$$

and

$$\mathbf{S}_{2,2} = \mathbf{L} \left[ \left( \frac{\mathbf{X}^\top \mathbf{X}}{N} \right)^{-1} - \mathbf{M}^{-1} \right] \frac{1}{N} \begin{bmatrix} \sum_j \sum_i I_i Z_{i,j} \gamma_j + \sum_i I_i \varepsilon_i \\ \sum_j \sum_i (1 - I_i) Z_{i,j} \gamma_j + \sum_i (1 - I_i) \varepsilon_i \\ \sum_i X_{i,1} \varepsilon_i \\ \vdots \\ \sum_i X_{i,p} \varepsilon_i \end{bmatrix}.$$

For part  $\mathbf{S}_{2,1}$ , to get the inverse of  $\mathbf{X}^\top \mathbf{X}/N$  in  $\mathbf{S}_{2,1}$ , let

$$\begin{aligned} \frac{1}{N} \mathbf{X}^\top \mathbf{X} &= \frac{1}{N} \begin{bmatrix} \sum I_i & 0 & \sum I_i X_{i,1} & \cdots & \sum I_i X_{i,p} \\ 0 & \sum (1 - I_i) & \sum (1 - I_i) X_{i,1} & \cdots & \sum (1 - I_i) X_{i,p} \\ \sum X_{i,1} I_i & \sum X_{i,1} (1 - I_i) & \sum X_{i,1} X_{i,1} & \cdots & \sum X_{i,1} X_{i,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum X_{i,p} I_i & \sum X_{i,p} (1 - I_i) & \sum X_{i,p} X_{i,1} & \cdots & \sum X_{i,p} X_{i,p} \end{bmatrix} \\ &= \begin{bmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{bmatrix}, \end{aligned}$$

where

$$\mathcal{A} = \begin{bmatrix} \sum I_i/N & 0 \\ 0 & \sum (1 - I_i)/N \end{bmatrix}$$

$$\mathcal{B} = \mathcal{C}^\top = \begin{bmatrix} \sum I_i X_{i,1}/N & \cdots & \sum I_i X_{i,p}/N \\ \sum (1 - I_i) X_{i,1}/N & \cdots & \sum (1 - I_i) X_{i,p}/N \end{bmatrix}$$

and

$$\mathcal{D} = \frac{\underline{\mathbf{X}}^\top \underline{\mathbf{X}}}{N}.$$

By matrix blockwise inverse,

$$\begin{bmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathcal{A} - \mathcal{B}\mathcal{D}^{-1}\mathcal{C})^{-1} & -(\mathcal{A} - \mathcal{B}\mathcal{D}^{-1}\mathcal{C})^{-1}\mathcal{B}\mathcal{D}^{-1} \\ -\mathcal{D}^{-1}\mathcal{C}(\mathcal{A} - \mathcal{B}\mathcal{D}^{-1}\mathcal{C})^{-1} & \mathcal{D}^{-1} + \mathcal{D}^{-1}\mathcal{C}(\mathcal{A} - \mathcal{B}\mathcal{D}^{-1}\mathcal{C})^{-1}\mathcal{B}\mathcal{D}^{-1} \end{bmatrix}$$

Therefore, by matrix calculations,

$$\mathbf{S}_{2,1} = \mathbf{L} \left[ \left( \frac{\underline{\mathbf{X}}^\top \underline{\mathbf{X}}}{N} \right)^{-1} - \mathcal{M}^{-1} \right] \begin{bmatrix} 0 \\ 0 \\ \underline{\mathbf{X}}\mathbf{Z}\boldsymbol{\gamma}/N \end{bmatrix} = (1, -1) [ -(\mathcal{A} - \mathcal{B}\mathcal{D}^{-1}\mathcal{C})^{-1}\mathcal{B}\mathcal{D}^{-1} ] \frac{\underline{\mathbf{X}}\mathbf{Z}\boldsymbol{\gamma}}{N}$$

It is easy to show

$$\mathcal{A} - \mathcal{B}\mathcal{D}^{-1}\mathcal{C} \xrightarrow{P} \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix} - \mathbf{0}\boldsymbol{\Sigma}_{x,x}^{-1}\mathbf{0} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}.$$

In addition, notice that

$$\frac{\sum X_{i,k}}{\sqrt{N}} = O_p(1)$$

by the central limit theorem and

$$\frac{\sum (2I_i - 1)X_{i,k}}{\sqrt{N}} = O_p(1)$$

according to Lemma 3.6.2 by letting  $\mathbf{a} = \mathbf{0}$ ,  $c = 0$  and  $\mathbf{b}$  is a vector with the  $k$ th

element equal to 1 and all other elements equal to 0, then

$$\sqrt{N}\mathcal{B} = O_p(1).$$

It follows that

$$-(\mathcal{A} - \mathcal{B}\mathcal{D}^{-1}\mathcal{C})^{-1}\mathcal{B}\mathcal{D}^{-1} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}^{-1} \mathcal{B}\Sigma_{x,x}^{-1} + o_p(N^{-1/2})$$

and further,

$$\mathbf{S}_{2,1} = \frac{2}{N} \left( \sum_i (2I_i - 1)X_{i,1}, \dots, \sum_i (2I_i - 1)X_{i,k} \right) \Sigma_{x,x}^{-1} \Sigma_{x,z} \boldsymbol{\gamma} + o_p(N^{-1/2}).$$

For part  $\mathbf{S}_{2,2}$ , we will show  $\mathbf{S}_{2,2} = o_p(N^{-1/2})$ . To prove this, it suffices to show

$$\frac{1}{\sqrt{N}} \begin{bmatrix} \sum_j \sum_i I_i Z_{i,j} \gamma_j + \sum_i I_i \varepsilon_i \\ \sum_j \sum_i (1 - I_i) Z_{i,j} \gamma_j + \sum_i (1 - I_i) \varepsilon_i \\ \sum_i X_{i,1} \varepsilon_i \\ \vdots \\ \sum_i X_{i,p} \varepsilon_i \end{bmatrix} = O_p(1),$$

since we have already known

$$\left( \frac{\mathbf{X}^\top \mathbf{X}}{N} \right)^{-1} - \mathcal{M}^{-1} \xrightarrow{P} \mathbf{0}.$$

By the central limit theorem, for any  $k$ ,  $k = 1, 2, \dots, p$

$$\frac{1}{\sqrt{N}} \sum_i X_{i,k} \varepsilon_i = O_p(1).$$

Also, we can show

$$\begin{aligned} \frac{1}{\sqrt{N}} \left( \sum_j \sum_i I_i Z_{i,j} \gamma_j + \sum_i I_i \varepsilon_i \right) &= \frac{1}{2} \left[ \frac{1}{\sqrt{N}} \sum_j \sum_i Z_{i,j} \gamma_j + \frac{1}{\sqrt{N}} \sum_i \varepsilon_i \right. \\ &\quad \left. + \frac{1}{\sqrt{N}} \sum_j \sum_i (2I_i - 1) Z_{i,j} \gamma_j + \frac{1}{\sqrt{N}} \sum_i (2I_i - 1) \varepsilon_i \right]. \end{aligned}$$

Again by the central limit theorem,

$$\frac{1}{\sqrt{N}} \sum_j \sum_i Z_{i,j} \gamma_j + \frac{1}{\sqrt{N}} \sum_i \varepsilon_i = O_p(1).$$

Apply Lemma 3.6.2 by letting  $\mathbf{a} = (1, 1, \dots, 1)$ ,  $\mathbf{b} = \mathbf{0}$  and  $c = 1$ ,

$$\frac{1}{\sqrt{N}} \sum_j \sum_i (2I_i - 1) Z_{i,j} \gamma_j + \frac{1}{\sqrt{N}} \sum_i (2I_i - 1) \varepsilon_i = O_p(1).$$

Hence,

$$\frac{1}{\sqrt{N}} \left( \sum_j \sum_i I_i Z_{i,j} \gamma_j + \sum_i I_i \varepsilon_i \right) = O_p(1).$$

By symmetry,

$$\frac{1}{\sqrt{N}} \left( \sum_j \sum_i (1 - I_i) Z_{i,j} \gamma_j + \sum_i (1 - I_i) \varepsilon_i \right) = O_p(1).$$

It follows that

$$\frac{1}{\sqrt{N}} \begin{bmatrix} \sum_j \sum_i I_i Z_{i,j} \gamma_j + \sum_i I_i \varepsilon_i \\ \sum_j \sum_i (1 - I_i) Z_{i,j} \gamma_j + \sum_i (1 - I_i) \varepsilon_i \\ \sum_i X_{i,1} \varepsilon_i \\ \vdots \\ \sum_i X_{i,p} \varepsilon_i \end{bmatrix} = O_p(1),$$

hence,

$$\mathbf{S}_{2,2} = o_p(N^{-1/2}).$$

Therefore,

$$\mathbf{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \frac{2}{N} (2\mathbf{I} - \mathbf{1})^\top (\mathbf{Z}\boldsymbol{\gamma} - \underline{\mathbf{X}}\boldsymbol{\Sigma}_{x,x}^{-1}\boldsymbol{\Sigma}_{x,z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}) + o_p(N^{-1/2}).$$

By Lemma 3.6.2,

$$\sqrt{N}\mathbf{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} \mathbf{N}(0, 4\{\mathbb{E}[\text{Var}((\mathbf{Z}^\top - \mathbf{X}^\top \boldsymbol{\Sigma}_{x,x}^{-1} \boldsymbol{\Sigma}_{x,z})\boldsymbol{\gamma} | \tilde{\mathbf{D}})] + \sigma_\varepsilon^2\}),$$

On the other hand, we show  $\hat{\sigma}^2 \mathbf{L}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{L}^\top$  is an inflated estimator of the variance of  $\mathbf{L}\hat{\boldsymbol{\beta}}$ , which is larger  $4\{\mathbb{E}[\text{Var}((\mathbf{Z} - \mathbf{X}\boldsymbol{\Sigma}_{x,x}^{-1}\boldsymbol{\Sigma}_{x,z})\boldsymbol{\gamma} | \tilde{\mathbf{D}})] + \sigma_\varepsilon^2\}$ . Notice that

$$\mathbf{L}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{L}^\top = \frac{1}{N} \mathbf{L} \left( \frac{\mathbf{X}^\top \mathbf{X}}{N} \right)^{-1} \mathbf{L}^\top = \frac{4}{N} + o_p\left(\frac{1}{N}\right)$$



and

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{N-p-2}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \frac{1}{N}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &+ \frac{1}{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{X}^\top \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \frac{2}{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{X}^\top(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + o_p(1).\end{aligned}$$

By Theorem 3.3.2,

$$\begin{aligned}\frac{1}{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{X}^\top \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &\xrightarrow{P} (0, 0, \boldsymbol{\gamma}^\top \boldsymbol{\Sigma}_{x,z}^\top \boldsymbol{\Sigma}_{x,x}^{-1}) \text{diag}\left(\frac{1}{2}, \frac{1}{2}, \boldsymbol{\Sigma}_{x,x}\right) \begin{bmatrix} 0 \\ 0 \\ \boldsymbol{\Sigma}_{x,x}^{-1} \boldsymbol{\Sigma}_{x,z} \boldsymbol{\gamma} \end{bmatrix} \\ &= \boldsymbol{\gamma}^\top \boldsymbol{\Sigma}_{x,z}^\top \boldsymbol{\Sigma}_{x,x}^{-1} \boldsymbol{\Sigma}_{x,z} \boldsymbol{\gamma},\end{aligned}$$

and similarly,

$$\frac{2}{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{X}^\top(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = 2(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \frac{\mathbf{X}^\top \mathbf{Z} \boldsymbol{\gamma}}{N} + 2(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \frac{\mathbf{X}^\top \boldsymbol{\varepsilon}}{N} \xrightarrow{P} 2\boldsymbol{\gamma}^\top \boldsymbol{\Sigma}_{x,z}^\top \boldsymbol{\Sigma}_{x,x}^{-1} \boldsymbol{\Sigma}_{x,z} \boldsymbol{\gamma}.$$

It follows from the law of large numbers and independence of  $\mathbf{Z}$  and  $\boldsymbol{\varepsilon}$  that,

$$\frac{1}{N}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \frac{1}{N}(\mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon})^\top(\mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}) \xrightarrow{P} \sigma_\varepsilon^2 + \boldsymbol{\gamma}^\top \boldsymbol{\Sigma}_{z,z} \boldsymbol{\gamma}.$$

Therefore,

$$\hat{\sigma}^2 \xrightarrow{P} \sigma_\varepsilon^2 + \boldsymbol{\gamma}^\top (\boldsymbol{\Sigma}_{z,z} - \boldsymbol{\Sigma}_{x,z}^\top \boldsymbol{\Sigma}_{x,x}^{-1} \boldsymbol{\Sigma}_{x,z}) \boldsymbol{\gamma} = \sigma_\varepsilon^2 + \text{Var}[(\mathbf{Z}^\top - \mathbf{X}^\top \boldsymbol{\Sigma}_{x,x}^{-1} \boldsymbol{\Sigma}_{x,z}) \boldsymbol{\gamma}]$$

and

$$\hat{\sigma}^2 \mathbf{L}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{L}^\top = \frac{4}{N}(\sigma_\varepsilon^2 + \text{Var}[(\mathbf{Z}^\top - \mathbf{X}^\top \boldsymbol{\Sigma}_{x,x}^{-1} \boldsymbol{\Sigma}_{x,z}) \boldsymbol{\gamma}]) + o_p\left(\frac{1}{N}\right).$$

Then under  $H_0 : \mu_1 - \mu_2 = 0$ ,

$$T \xrightarrow{D} \mathbf{N}(0, \tau^2), \quad \tau^2 = \frac{\sigma_\varepsilon^2 + \mathbb{E}[\text{Var}((\mathbf{Z}^\top - \mathbf{X}^\top \Sigma_{x,x}^{-1} \Sigma_{x,z}) \boldsymbol{\gamma} | \tilde{\mathbf{D}})]}{\sigma_\varepsilon^2 + \text{Var}[(\mathbf{Z}^\top - \mathbf{X}^\top \Sigma_{x,x}^{-1} \Sigma_{x,z}) \boldsymbol{\gamma}]} = \frac{\sigma_1^2}{\sigma_2^2},$$

where  $\sigma_1^2 = \sigma_\varepsilon^2 + \mathbb{E}[\text{Var}((\mathbf{Z}^\top - \mathbf{X}^\top \Sigma_{x,x}^{-1} \Sigma_{x,z}) \boldsymbol{\gamma} | \tilde{\mathbf{D}})]$  and  $\sigma_2^2 = \sigma_\varepsilon^2 + \text{Var}[(\mathbf{Z}^\top - \mathbf{X}^\top \Sigma_{x,x}^{-1} \Sigma_{x,z}) \boldsymbol{\gamma}]$ .

Notice that

$$\text{Var}[(\mathbf{Z}^\top - \mathbf{X}^\top \Sigma_{x,x}^{-1} \Sigma_{x,z}) \boldsymbol{\gamma}] > \mathbb{E}[\text{Var}((\mathbf{Z}^\top - \mathbf{X}^\top \Sigma_{x,x}^{-1} \Sigma_{x,z}) \boldsymbol{\gamma} | \tilde{\mathbf{D}})],$$

which leads to, when  $N \rightarrow \infty$ ,

$$\Pr(|T| > Z_{1-\alpha/2}) \rightarrow 2\Phi\left(-\frac{\sigma_2 Z_{1-\alpha/2}}{\sigma_1}\right) < \alpha,$$

where  $\Phi$  is the cumulative distribution function of standard normal distribution.

Similarly, under  $H_A : \mu_1 - \mu_2 \neq 0$  with a sequence of local alternatives, i.e.,  $\mu_1 - \mu_2 = \delta/\sqrt{N}$  for a fixed  $\delta \neq 0$ ,

$$T \xrightarrow{D} \mathbf{N}(\Delta, \tau^2), \quad \Delta = \frac{\delta}{2\sigma_2}.$$

which completes proof of Theorem 3.3.2. □

## Chapter 4

# Statistical Inference for Logistic Regression

### 4.1 Introduction

In Chapter 2 and 3, statistical inference for linear models are studied under covariate-adaptive designs. The discussion concentrates on the scenario of continuous responses on which linear models can be applied. However, the response variable of interest is not limited to continuous variable in clinical trials. For example, the outcome of a clinical trial can be a binary variable indicating whether a patient is a responder or non-responder to a newly developed therapy. In this case, a logistic regression can be utilized to model the relationship between patients' responses and treatment assignments adjusted by covariates, if any. Other types of responses in clinical trials include count data or time-to-event data where Poisson model or survival model should be appropriately applied. Considering wide applications of clinical trials with response variables that are not continues, it is necessary to study the inference properties of generalized linear models and even more advanced models

under covariate-adaptive randomized clinical trials.

In this chapter our discussion focuses on statistical inference for logistic regression, which is one of the most important and commonly used analysis model in the family of generalized linear models. When responses are binary in a covariate-adaptive randomized clinical trial, logistic regression can be utilized to test the significance of treatment effects or covariates. Thus a natural question arises whether the conclusions regarding treatment effects in Chapter 2 and 3 still hold for logistic regression. Corresponding theoretical properties are explored in this chapter.

In Section 4.2, a general framework is proposed to study properties of the Wald test if logistic regression is used for analysis with all randomization covariates omitted in the working model. In Section 4.3, theoretical properties of hypothesis testing for comparing treatment effects are proposed. Simulations are carried out in Section 4.4 to investigate Type I error and power. In Section 4.5, conclusion remarks are given and possible future directions are also discussed.

## 4.2 Framework

In this section, we study the properties of the Wald test on treatment effect for logistic regression under covariate-adaptive randomized clinical trials. Response  $Y$  is a binary variable with possible values 1 (success) and 0 (failure). Suppose there are two treatments, treatment 1 and treatment 2, and in total  $N$  patients are enrolled in the study. Covariate profile of patients is denoted by  $X$ , which can be either univariate or multivariate. The covariate-adaptive randomization is implemented directly on  $X$  if  $X$  is discrete, or on  $D(X)$  if  $X$  is continuous, where  $D(\cdot)$  is a discrete function. The treatment assignment  $I$  can take values of 1 and 0, corresponding to treatments 1 and 2, respectively. The following model is assumed

to be the underlying true model,

$$p = h(\mu + \alpha I + \beta X) = h(\eta), \quad (4.1)$$

where  $p$  is the success probability given treatment assignment and covariate information, which is equal to  $\mathbf{E}(Y|I, X)$  since  $Y$  is a bernoulli variable.  $h(\cdot)$  is the function  $h(\eta) = \exp(\eta)/[1 + \exp(\eta)]$ . The treatment effect is represented by  $\alpha$ , which is equal to the log odds ratio of success for a patient assigned to treatment 1 compared to treatment 2. The underlying model (4.1) is equivalent to,

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \mu + \alpha I + \beta X.$$

In this chapter, the discussion is focused on testing treatment effect for logistic regression under covariate-adaptive designs. Covariate effects are not our main interest, and we only consider the scenario where all covariate information of  $X$  is omitted in analysis. Then the following model is assumed to be the working model for statistical analysis,

$$p = h(\mu + \alpha I). \quad (4.2)$$

Based on the working model (4.2), there are several ways to test if the treatment effect is significant, including the Wald test, the score test and the likelihood ratio test. Among these three tests, the Wald test is most widely used and is available in most statistical package, so we only consider the Wald test in this chapter. Under the working model, the Wald test has the following form to test the treatment effect.

$$H_0 : \alpha = 0 \text{ versus } H_A : \alpha \neq 0. \quad (4.3)$$

The test statistic for (4.3) is

$$T = \frac{\hat{\alpha}}{\hat{se}(\alpha)}, \quad (4.4)$$

where  $\hat{se}(\alpha)$  is the estimated standard deviation of  $\alpha$  derived from the inverse of the estimated information matrix. If  $|T| > Z_{1-\alpha/2}$ , where  $Z_{1-\alpha/2}$  is  $(1 - \alpha/2)$  quantile of a standard normal distribution, we will reject the null hypothesis, otherwise accept the null hypothesis.

### 4.3 Theoretical properties

The hypothesis testing for treatment effect is discussed based on the linear model framework in Chapter 2 and 3 (Theorem 2.3.1, Theorem 3.3.2). In this section, the similar problem is studied for logistic regression under covariate-adaptive designs. In particular, the Type I error of hypothesis testing (4.3) based on the working model (4.2) is evaluated if the underlying model (4.1) holds. Compared to the conditions (A) and (B) in Theorem 2.3.1 (Theorem 3.3.2), a relatively strong condition is assumed to derive asymptotic properties. The main results are given below.

**Theorem 4.3.1.** *Suppose that a covariate-adaptive design satisfies the condition that all within-stratum imbalances are  $O_p(1)$ , then under  $H_0 : \alpha = 0$ ,*

$$T \xrightarrow{D} \mathbf{N}\left(0, \frac{\mathbf{E}[\mathbf{Var}(Y|D(X))]}{\mathbf{Var}(Y)}\right). \quad (4.5)$$

*Hence, the hypothesis testing (4.3) is conservative.*

**REMARK 4.3.1.** *Since  $\mathbf{E}[\mathbf{Var}(Y|D(X))]$  is generally smaller than  $\mathbf{Var}(Y)$ , the asymptotic distribution of test statistic  $T$  under the null hypothesis is not a standard normal*

distribution as under complete randomization. Instead, it is a normal distribution with a mean zero and a standard deviation smaller than 1. Based on the decision rule of hypothesis testing (4.3), the Type I error is smaller than the nominal level. In the other word, the hypothesis testing for treatment effect is conservative, causing more difficult to reject the null hypothesis even though the null hypothesis is true.

REMARK 4.3.2. *The condition regarding within-stratum imbalances in Theorem 4.3.1 is satisfied by several covariate-adaptive randomization methods. A few typical examples include covariate-adaptive designs balancing covariates by stratification, such as stratified permuted block design and covariate-adaptive biased coin design. In addition, by choosing appropriate weights in the new designs developed by Hu and Hu (2012), the condition in Theorem 4.3.1 can also be satisfied. Moreover, this condition is stronger than the conditions in Theorem 2.3.1 or Theorem 3.3.2. A covariate-adaptive design with all within-stratum imbalances  $O_p(1)$  will automatically have marginal imbalances  $O_p(1)$  and overall imbalance  $O_p(1)$ .*

REMARK 4.3.3. *Pocock and Simon's marginal procedure doesn't satisfy the condition in Theorem 4.3.1. Hu and Zhang (2013) proved that, even though marginal imbalances and overall imbalance are  $O_p(1)$  for Pocock and Simon's marginal procedure, within-stratum imbalances increases with the order of  $O_p(N^{1/2})$ . However, all within-stratum imbalances are  $O_p(1)$  is only a sufficient condition, but not necessary, to derive the conservativeness of hypothesis testing (4.3). Even though Theorem 4.3.1 cannot be applied, we will further look at the properties for Pocock and Simon's marginal procedure by conducting simulations in Section 4.4.*

In Theorem 4.3.1, we only consider a special scenario that all randomization covariates are omitted in the working model. Theoretical properties remain unknown if only partial covariate information is excluded from analysis based on the logistic

regression, even though simulations show that the hypothesis testing for treatment effect is also conservative. Moreover, the asymptotic distribution of test statistic (4.3) under the alternative hypothesis is not given due to technical difficulties and will be left as a future research topic. However, the power of logistic regression under covariate-adaptive designs are studied via simulations in the next section.

To prove Theorem 4.3.1, the asymptotic behaviors of the numerator  $\hat{\alpha}$  and the denominator  $\hat{se}(\alpha)$  in test statistic (4.4) are studied. Refer to the Appendix in Section 4.6 for detailed proofs.

## 4.4 Simulation study

*Case 1: Type I error.* Simulations are carried out to study Type I error of the hypothesis testing of treatment effect when a logistic regression model is used. Two types of covariate-adaptive designs, including stratified permuted block design and Pocock and Simon's marginal procedure, are considered in this case. Results for complete randomization are also attached for comparison. For each type of design, both continuous and discrete cases are studied. The following model including two covariates  $Z_1$  and  $Z_2$  is assumed to be the underlying true model,

$$\text{logit}(p_i) = \mu_1 I_i + \mu_2 (1 - I_i) + \beta_1 Z_{i,1} + \beta_2 Z_{i,2},$$

where  $\beta_1 = \beta_2 = 1$ . To evaluate Type I error, treatment effects are assumed to be identical between two groups, i.e.,  $\mu_1 = \mu_2 = 0$ . For the discrete case,  $Z_1$  and  $Z_2$  are independent and follow Bernoulli(0.5); for the continuous case,  $Z_1$  and  $Z_2$  are independent and follow a standard normal distributions  $\mathbf{N}(0, 1)$ . If covariates  $Z_1$  and  $Z_2$  are continuous, they are discretized to bernoulli variables  $Z'_1$  and  $Z'_2$  with



the probability 0.5 to be used in randomization. More specifically, if  $Z_1 < Z_{(p_1)}$ , where  $Z_{(p_1)}$  is  $p_1$  quantile of a standard normal distribution, then  $Z'_1 = 0$ , otherwise  $Z'_1 = 1$ . Original covariates without discretization are used to do analysis.

In stratified permuted blocked designs, the block size 4 is used. In Pocock and Simon's marginal procedure, the probability of biased coin assignment is 0.75 and equal weights are assigned on two covariates. Other parameters in simulations are: the significance level is  $\alpha = 0.05$  and sample size  $N$  is 100, 200 or 500. The analysis models studied here include a logistic model with no covariate ( $glm0$ ), a logistic model with a single covariate  $Z_1$  ( $glm(Z_1)$ ), a logistic model with a single covariate  $Z_2$  ( $glm(Z_2)$ ) and a logistic model with both covariate  $Z_1$  and  $Z_2$  ( $glm(Z_1, Z_2)$ ). The simulation results for stratified permuted block design, Pocock and Simon's marginal procedure and complete randomization are given in Table 4.1.

Several conclusions can be obtained from Table 4.1. (1) The logistic regression with no covariates included ( $glm0$ ) under stratified permuted block design has Type I error less than 5%. This is consistent to the conclusion of Theorem 4.3.1 that the hypothesis testing for treatment effect is conservative if within-stratum imbalances are bounded in probability. (2) Under stratified permuted block design, the results of Type I error of  $glm(Z_1)$  and  $glm(Z_2)$  are also conservative, where only partial covariate information is implemented into analysis. (3) The Type I error of the full model ( $glm(Z_1, Z_2)$ ) is close to 5% under all three randomization methods, which makes sense since the full model is the underlying true model. (4) In Table 4.1,  $glm0$ ,  $glm(Z_1)$  and  $glm(Z_2)$  all have Type I error smaller than the nominal level 5% under Pocock and Simon's marginal procedure. Notice that within-stratum imbalances of Pocock and Simon's marginal procedure are  $O_p(N^{1/2})$  and thus don't satisfy the condition in Theorem 4.3.1. It is still not clear about theoretical properties of testing hypothesis for Pocock and Simon's marginal procedure and further

Table 4.1: Simulated Type I error for stratified permuted block design (SPB), Pocock and Simon’s marginal procedure (PS) and complete randomization (CR) in %. Simulations based on 10000 runs.

$Z$	Method	$N$	$glm0$	$glm(Z_1)$	$glm(Z_2)$	$glm(Z_1, Z_2)$	
Discrete	<b>SPB</b>	100	3.73	4.43	4.51	5.18	
		200	4.05	4.46	4.38	5.01	
		500	3.90	4.36	4.31	4.85	
	<b>PS</b>	100	3.80	4.53	4.48	4.97	
		200	4.08	4.58	4.44	5.00	
		500	4.04	4.55	4.56	5.09	
	<b>CR</b>	100	4.51	4.55	4.62	4.73	
		200	5.08	5.00	5.14	5.19	
		500	4.92	4.80	4.75	4.74	
	Continuous	<b>SPB</b>	100	3.28	3.59	3.63	4.67
			200	3.57	3.89	3.91	5.16
			500	3.14	3.44	3.41	4.55
<b>PS</b>		100	3.65	4.24	3.90	5.11	
		200	3.78	4.06	4.09	5.23	
		500	3.27	3.93	3.72	5.08	
<b>CR</b>		100	5.21	5.83	4.73	5.04	
		200	5.15	4.67	5.16	4.93	
		500	5.22	4.99	5.64	5.47	

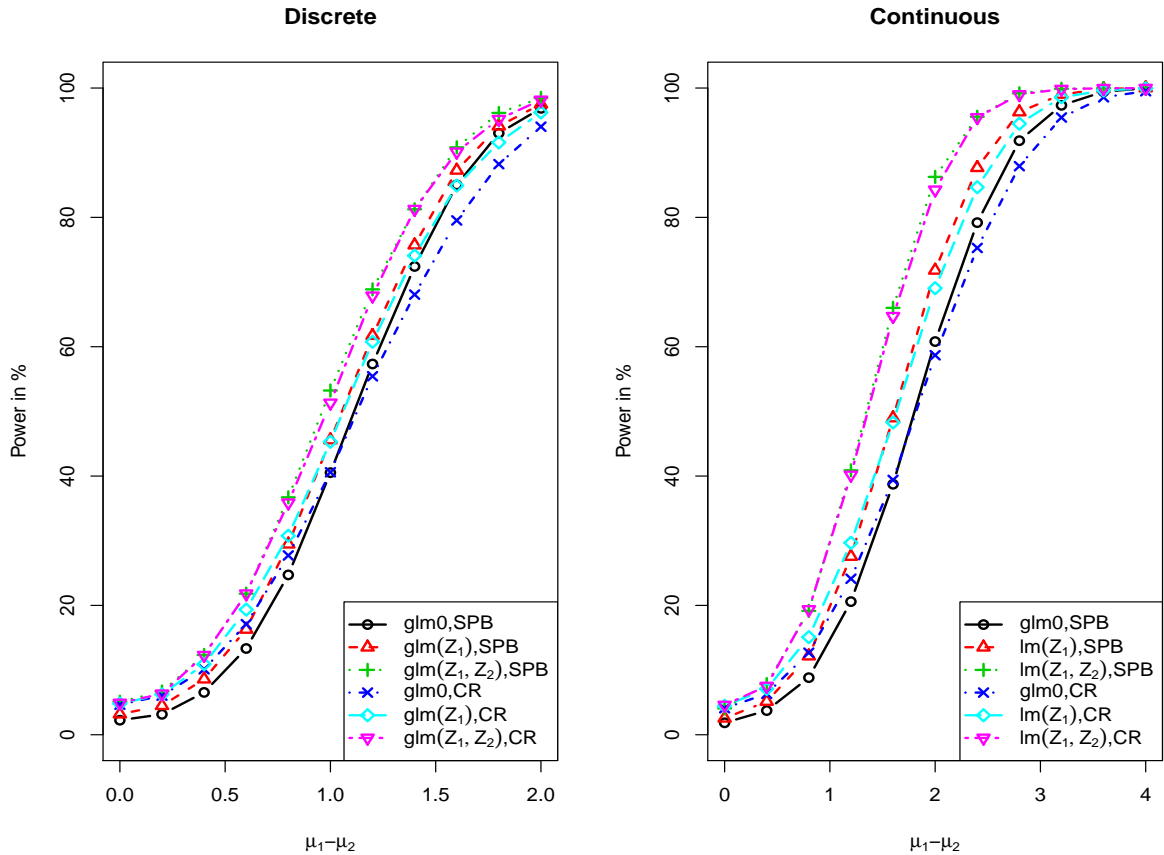
theoretical results are desired. (5) The Type I error of all hypothesis testing are around 5% and valid under complete randomization.

*Case 2: Power Comparison.* In this case, power are compared for different hypothesis testing methods under stratified permuted block design and complete randomization. The same model as in Case 1 is used, however, it is assumed that difference exists between treatment effects  $\mu_1$  and  $\mu_2$ , i.e.,  $\mu_1 - \mu_2 \neq 0$ . The following model including two covariates  $Z_1$  and  $Z_2$  is used as the underlying model,

$$\text{logit}(p_i) = \mu_1 I_i + \mu_2 (1 - I_i) + \beta_1 Z_{i,1} + \beta_2 Z_{i,2},$$

where  $\beta_1 = \beta_2 = 2$  and  $\mu_2 = -3$ . Treatment effects are assumed to be different to

Figure 4.1: Simulated power for stratified permuted block design (SPB) and complete randomization (CR) under discrete and continuous cases. Simulation based on 10000 runs and number of patients  $N = 100$ .



study power, i.e.  $\mu_1 \neq \mu_2$ . Other parameter settings are the same as Case 1. The power results are presented in Figure 4.1.

Figure 4.1 implies a few conclusions which are summarized below. (1) Among all hypothesis testing methods considered,  $glm(Z_1, Z_2)$  is the most powerful one under either stratified permuted block design or complete randomization.  $glm0$  and  $glm(Z_1)$  are less powerful than  $glm(Z_1, Z_2)$ , and  $glm0$  is the least powerful one. These conclusions are not surprising since more covariates in the analysis model allow us to better detect difference between treatment effects. (2) The power of

the full model  $glm(Z_1, Z_2)$  under stratified permuted block design is slightly larger than that under complete randomization. (3) The power of  $glm0$  under stratified permuted block design is smaller than that under complete randomization when treatment effect is relatively small. However, the power of stratified permuted block design becomes larger than that of complete randomization when treatment effect is getting larger. The power of  $glm(Z_1)$  has similar pattern under stratified permuted block design and complete randomization.

## 4.5 Conclusion

In this chapter, we studied the properties of statistical inference for logistic regression under covariate-adaptive designs. We showed that the hypothesis testing of treatment effect is conservative when covariates are fully omitted from inference procedures. The asymptotic distribution of test statistic is given for stratified permuted block design and other covariate-adaptive designs with within-stratum imbalances bounded in probability. However, the theoretical properties remain unknown for covariate-adaptive designs which don't satisfy the conditions in Theorem 4.3.1. In view of that, simulations are carried out to study Type I error under Pocock and Simon's marginal procedure. The results imply that the test of treatment effect is also conservative for Pocock and Simon's marginal procedure if covariates used in covariate-adaptive design are omitted from the analysis model. Corresponding theoretical work is a promising future research topic.

The discussion in this chapter focuses on statistical inference solely for one type of generalized linear model, logistic regression whose outcome is a binary variable. Other kinds of outcomes often occur in clinical trials as well. For example, Poisson model and Cox model also have broad applications in practice. Gail (1988) examined

the properties of the score test for a large family of generalized linear models based on perfectly balanced studies, but the conclusion cannot be applied to most of covariate-adaptive randomized clinical trials. The theoretical properties of statistical inference (such as the Wald test and the likelihood ratio test) for generalized linear models are mostly unexplored under covariate-adaptive designs. Our work in this chapter provides a framework for further research in this area.

## 4.6 Appendix: proof of theorems

In this section, Theorem 4.3.1 will be proved based on the following three lemmas.

**Lemma 4.6.1.** *Under the assumptions of Theorem 4.3.1, then*

$$\frac{1}{N} \sum I_i \xrightarrow{P} \frac{1}{2}. \quad (4.6)$$

*Proof.* Let the overall imbalance, denoted by  $D_N$ , be the difference between the numbers of patients in treatment group 1 and 2 as total, then  $D_N = \sum(2I_i - 1)$ . Since the overall imbalance  $D_N$  is the sum of all within-stratum imbalance, then the overall imbalance is also bounded in probability, i.e.,  $D_N = O_p(1)$ , if all within-stratum imbalances are all bounded in probability. It follows that,

$$\frac{1}{N} \sum I_i = \frac{1}{2} + \frac{\sum(2I_i - 1)}{2N} = \frac{1}{2} + \frac{1}{2} \frac{D_N}{N} \xrightarrow{P} \frac{1}{2}.$$

The proof of Lemma 4.6.1 is completed. □

**Lemma 4.6.2.** *Under the assumptions of Theorem 4.3.1, if the null hypothesis*

$H_0 : \alpha = 0$  is true, then

$$\frac{1}{\sqrt{N}} \sum (2I_i - 1)Y_i \xrightarrow{D} \mathbf{N}(0, \mathbf{E}[h(\mu + \beta X)(1 - h(\mu + \beta X))]). \quad (4.7)$$

*Proof.* Define  $\mathcal{D} = \{D(X_i), i = 1, 2, \dots, N\}$  and  $\mathcal{I} = \{I_i, i = 1, 2, \dots, N\}$ . Then

$$\sum (2I_i - 1)Y_i = \sum (2I_i - 1)[Y_i - \mathbf{E}(Y_i|\mathcal{D})] + \sum (2I_i - 1)\mathbf{E}(Y_i|\mathcal{D})$$

Since  $\mathbf{E}(Y_i|\mathcal{D}) = \mathbf{E}(Y_i|D(X_i))$  can only take a finite number of values corresponding to each strata and within-stratum imbalances are all bounded in probability, then

$$\sum (2I_i - 1)\mathbf{E}(Y_i|\mathcal{D}) = o_p(N^{1/2}).$$

Further, notice that  $I_i$  and  $(Y_i - \mathbf{E}(Y_i|\mathcal{D}))$  are independent given  $\mathcal{D}$  under the null hypothesis, then

$$\mathbf{E}\left[\frac{1}{N} \sum (2I_i - 1)(Y_i - \mathbf{E}(Y_i|\mathcal{D}))|\mathcal{D}\right] = \frac{1}{N} \sum \mathbf{E}[(2I_i - 1)|\mathcal{D}]\mathbf{E}[Y_i - \mathbf{E}(Y_i|\mathcal{D})|\mathcal{D}] = 0,$$

and using the fact  $(2I_i - 1)^2 = 1$ ,

$$\mathbf{Var}\left[\frac{1}{N} \sum (2I_i - 1)(Y_i - \mathbf{E}(Y_i|\mathcal{D}))|\mathcal{D}\right] = \frac{\sum \mathbf{E}[(2I_i - 1)^2(Y_i - \mathbf{E}(Y_i|\mathcal{D}))^2|\mathcal{D}]}{N^2} = \frac{\sum \mathbf{Var}[Y_i|\mathcal{D}]}{N^2}.$$

By the central limit theorem, the conditional distribution of

$$\frac{1}{\sqrt{N}} \sum (2I_i - 1)(Y_i - \mathbf{E}(Y_i|\mathcal{D}))$$

given  $(\mathcal{D}, \mathcal{I})$ , is asymptotically normal with mean zero and variance  $\sum \mathbf{Var}[Y_i|\mathcal{D}]/N$ , which converges to  $\mathbf{E}[\mathbf{Var}[Y|D(X)]]$  by the law of large numbers. Thus, it also holds

unconditionally that

$$\frac{1}{\sqrt{N}} \sum (2I_i - 1)(Y_i - \mathbf{E}(Y_i | \mathcal{D})) \xrightarrow{D} \mathbf{N}(0, \mathbf{E}[\text{Var}[Y | D(X)]]).$$

Using the fact  $\mathbf{E}[\text{Var}[Y | D(X)]] = \mathbf{E}[h(\mu + \beta X)(1 - h(\mu + \beta X))]$  and Slutsky's theorem,

$$\frac{1}{\sqrt{N}} \sum (2I_i - 1)Y_i \xrightarrow{D} \mathbf{N}(0, \mathbf{E}[h(\mu + \beta X)(1 - h(\mu + \beta X))]),$$

which completes the proof.  $\square$

**Lemma 4.6.3.** *Under the assumptions of Theorem 4.3.1, if the null hypothesis  $H_0 : \alpha = 0$  is true, then*

$$\frac{1}{N} \sum I_i Y_i \xrightarrow{P} \frac{1}{2} \mathbf{E}Y. \quad (4.8)$$

*Proof.* It is easy to show

$$\frac{1}{N} \sum I_i Y_i = \frac{\sum Y_i}{2N} + \frac{\sum (2I_i - 1)Y_i}{2N}.$$

It follows from the law of large numbers that

$$\frac{\sum Y_i}{2N} \xrightarrow{P} \frac{1}{2} \mathbf{E}Y.$$

Further, by Lemma 4.6.2,  $\sum (2I_i - 1)Y_i / \sqrt{N}$  is bounded in probability since it follows a normal distribution asymptotically, then

$$\frac{\sum (2I_i - 1)Y_i}{2N} = \frac{1}{2\sqrt{N}} \frac{\sum (2I_i - 1)Y_i}{\sqrt{N}} \xrightarrow{P} 0.$$

By Slutsky's theorem, the proof of Lemma 4.6.3 is completed.  $\square$

*Proof of Theorem 4.3.1.* Under the working model (4.2), the log-likelihood for  $Y$ , conditional on  $\eta$ , is given as

$$l = \sum_{i=1}^n [Y_i \eta_i - \log(1 + \exp(\eta_i))].$$

Thus the derivative of the log-likelihood  $l$  with respect to  $\mu$  and  $\alpha$  are,

$$\frac{\partial l}{\partial \mu} = \sum Y_i - \sum h(\mu + \alpha I_i)$$

and

$$\frac{\partial l}{\partial \alpha} = \sum I_i Y_i - \sum I_i h(\mu + \alpha I_i).$$

Then the MLE estimators of  $\mu$  and  $\alpha$  can be obtained by letting  $\partial l / \partial \mu = \partial l / \partial \alpha = 0$ .

Define  $n_1 = \sum I_i$  and  $n_2 = \sum (1 - I_i) = N - n_1$ , then

$$\hat{\mu} = h^{-1}\left(\frac{\sum (1 - I_i) Y_i}{n_2}\right)$$

and

$$\hat{\alpha} = h^{-1}\left(\frac{\sum I_i Y_i}{n_1}\right) - h^{-1}\left(\frac{\sum (1 - I_i) Y_i}{n_2}\right).$$

Using Taylor expansions,

$$h^{-1}\left(\frac{\sum I_i Y_i}{n_1}\right) = h^{-1}(\mathbf{E}Y) + h^{-1(1)}(\mathbf{E}Y)\left(\frac{\sum I_i Y_i}{n_1} - \mathbf{E}Y\right) + \frac{1}{2}h^{-1(2)}(\xi)\left(\frac{\sum I_i Y_i}{n_1} - \mathbf{E}Y\right)^2,$$

where  $h^{-1(n)}$  denotes the  $n$ th derivative of  $h$ ,  $n = 1, 2$ , and  $\xi$  is between  $\sum I_i Y_i / n_1$  and  $\mathbf{E}Y$ .



Notice that

$$\frac{\sum I_i Y_i}{n_1} - \mathbf{E}Y = \frac{2}{N} \sum I_i Y_i - \mathbf{E}Y + \sum I_i Y_i \left( \frac{1}{\sum I_i} - \frac{2}{N} \right)$$

and

$$\sum I_i Y_i \left( \frac{1}{\sum I_i} - \frac{2}{N} \right) = \frac{\sum I_i Y_i (n_1 - n_2)/N}{N \sum I_i/N} = o_p(N^{-1/2}),$$

since  $\sum I_i/N \xrightarrow{P} 1/2$ ,  $\sum I_i Y_i/N \xrightarrow{P} \mathbf{E}Y/2$  by Lemma 4.6.1 and Lemma 4.6.3, and the overall imbalance  $n_1 - n_2 = D_N$  is  $O_p(1)$ , bounded in probability.

It follows from the law of large numbers and  $\sum (2I_i - 1)Y_i/\sqrt{N} = O_p(1)$  by Lemma 4.6.2 that

$$\frac{\sum I_i Y_i}{n_1} - \mathbf{E}Y = \frac{\sum Y_i}{N} - \mathbf{E}Y + \frac{\sum (2I_i - 1)Y_i}{N} + o_p(N^{-1/2}) = O_p(1).$$

Also, by the central limit theorem and Lemma 4.6.2,

$$\sqrt{N} \left( \frac{\sum I_i Y_i}{n_1} - \mathbf{E}Y \right) = \frac{\sum Y_i - N\mathbf{E}Y}{\sqrt{N}} + \frac{\sum (2I_i - 1)Y_i}{\sqrt{N}} + o_p(1) = O_p(1).$$

Thus,

$$h^{-1} \left( \frac{\sum I_i Y_i}{n_1} \right) = h^{-1}(\mathbf{E}Y) + h^{-1(1)}(\mathbf{E}Y) \left( \frac{2}{N} \sum I_i Y_i - \mathbf{E}Y \right) + o_p(N^{-1/2})$$

Similarly,

$$h^{-1} \left( \frac{\sum (1 - I_i)Y_i}{n_1} \right) = h^{-1}(\mathbf{E}Y) + h^{-1(1)}(\mathbf{E}Y) \left( \frac{2}{N} \sum (1 - I_i)Y_i - \mathbf{E}Y \right) + o_p(N^{-1/2})$$

Thus,

$$\hat{\alpha} = h^{-1(1)}(\mathbf{E}Y) \frac{2 \sum (2I_i - 1)Y_i}{N} + o_p(N^{-1/2}).$$

Notice that  $h^{-1(1)}(\mathbf{E}Y) = 1/[\mathbf{E}Y(1 - \mathbf{E}Y)] = 1/\mathbf{Var}(Y)$ , together with Lemma 4.6.2,

$$N^{1/2}\hat{\alpha} \xrightarrow{D} \mathbf{N}\left(0, \frac{4\mathbf{E}[\mathbf{Var}(Y|D(X))]}{\mathbf{Var}(Y)^2}\right)$$

Next, we consider the asymptotic property of  $\hat{s}e(\hat{\alpha})$ , which is obtained from inverse of the estimated information matrix  $\hat{I}(\mu, \alpha)$ . Notice that

$$\frac{\partial^2 l}{\partial \mu^2} = - \sum h'(\mu + \alpha I_i),$$

$$\frac{\partial^2 l}{\partial \alpha^2} = - \sum h'(\mu + \alpha I_i) I_i^2,$$

and

$$\frac{\partial^2 l}{\partial \mu \partial \alpha} = - \sum h'(\mu + \alpha I_i) I_i.$$

Substituting  $\mu$  and  $\alpha$  by their MLE estimators  $\hat{\mu}$  and  $\hat{\alpha}$ , we can show that

$$\begin{aligned} \sum h'(\hat{\mu} + \hat{\alpha} I_i) &= n_1 h'(\hat{\mu} + \hat{\alpha}) + n_2 h'(\hat{\mu}) \\ &= n_1 \left( \frac{\sum I_i Y_i}{n_1} \right) \left( 1 - \frac{\sum I_i Y_i}{n_1} \right) + n_2 \left( \frac{\sum (1 - I_i) Y_i}{n_1} \right) \left( 1 - \frac{\sum (1 - I_i) Y_i}{n_1} \right) \\ &\xrightarrow{P} \frac{N}{2} \mathbf{E}Y(1 - \mathbf{E}Y) + \frac{N}{2} \mathbf{E}Y(1 - \mathbf{E}Y) = N \mathbf{Var}(Y). \end{aligned}$$

Similarly,

$$\sum h'(\hat{\mu} + \hat{\alpha}I_i)I_i^2 = \sum h'(\hat{\mu} + \hat{\alpha}I_i)I_i = \frac{N}{2}\text{Var}(Y).$$

Then we have

$$\hat{I}(\mu, \alpha) \xrightarrow{P} N\text{Var}(Y) \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/2 \end{bmatrix},$$

which implies that

$$\hat{I}^{-1}(\mu, \alpha) \xrightarrow{P} \frac{\text{Var}(Y)^{-1}}{N} \begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix}.$$

It follows that

$$\hat{se}(\hat{\alpha}) = \frac{2}{\sqrt{N}}\text{Var}(Y)^{-1/2}.$$

Finally, by Slutsky's theorem,

$$T = \frac{\hat{\alpha}}{\hat{se}(\alpha)} \xrightarrow{D} \mathbf{N}\left(0, \frac{\mathbf{E}[\text{Var}(Y|D(X))]}{\text{Var}(Y)}\right),$$

which completes the proof of Theorem 4.3.1. □

## Chapter 5

### Conclusions

In this dissertation, we studied theoretical properties of statistical inference under general covariate-adaptive designs based on linear models and logistic regression. In Chapter 2 and 3, we derived the corresponding asymptotic distributions of several test statistics under both null and alternative hypotheses. Instead of focusing on a specific covariate-adaptive design, we studied the problem from the angle of imbalance measure of different levels (overall, marginal, within-stratum). So the conclusions can be applied to a broad range of covariate-adaptive designs, including stratified permuted block design and Pocock and Simon's marginal procedure. For example, to apply Theorems 2.3.1 and 2.3.2 to a specific covariate-adaptive randomized clinical trial, one just need to check the conditions (A) and (B) to see if they are satisfied. Furthermore, Chapter 3 allows us to study inference properties based on more general assumptions that covariates used in randomization are correlated with each other. The results in this dissertation provide new insights about balance and efficiency of clinical trials, and the framework can be used to study other statistical methods under covariate-adaptive designs.

Based on Theorems 2.3.1 and 2.3.2, incorporating all randomization covariates

in the final analysis model can achieve valid hypothesis testing for treatment effects and covariates. In practice, model selection techniques can also be used as in Case 4 of Section 2.4 to target the subset of all influential covariates, based on which valid tests can be obtained. However, sometimes not all important randomization covariates are used in the inference step for a covariate-adaptive design. Then the actual Type I error may be not equal to the nominal level when comparing treatment effects and adjustment is necessary to achieve a valid test in this situation. If no covariate is incorporated in the final analysis model, a bootstrap  $t$ -test as described in Shao, Yu and Zhong (2010) can be implemented to restore Type I error, and this method is shown to be more powerful than the conventional two sample  $t$ -test under covariate-adaptive designs. Similar bootstrap adjustment methods can be considered to correct the variance estimation and to make  $\tau = 1$  in Theorems 2.3.1 and 3.3.2. We leave this as a future research project.

The linear regression model is used to study hypothesis testing in Chapter 2 and 3, where response of patients is a continuous variable. In reality, the endpoints of clinical trials may be other types of variables. For instance, a clinical trial can be designed to compare success rates between a new medication and a standard treatment, where the response is a binary variable. In literature, Feinstein and Landis (1976) and Green and Byar (1978) studied the properties of unadjusted tests under stratified randomization for binary response based on a very special setting, and concluded that Type I error would decrease when stratified randomization is used rather than unstratified randomization. In Chapter 4, we extended their results to logistic regression under the assumption within-stratum imbalances are bounded in probability for covariate-adaptive designs. It is shown that the hypothesis testing for comparing treatment effect is conservative when all covariates are omitted from the analysis procedure. However, inference properties are not fully clear in the-

ory for the case only partial covariate information is used in analysis even though simulations show that Type I error is also conservative. The theoretical results for other types of covariate-adaptive designs, including Pocock and Simon's marginal procedure, remains unknown. As future research projects, we will study behaviors of conventional statistical inference for different kinds of responses based on generalized linear models, survival models and so on. There are some other kinds of covariate-adaptive randomization procedures in literature, including Zelen (1974), Wei (1978), Begg and Iglewicz (1980), and Atkinson (1982). Theorems developed in this dissertation may not apply to these designs, because it is unknown whether the conditions in this dissertation (Theorem 2.3.1, Theorem 3.3.1, Theorem 4.3.1) remain true for these designs.

The proposed properties of hypothesis testing for covariate-adaptive designs can be generalized in several ways. (1) All covariate-adaptive designs considered in this dissertation are based on discrete covariates. We may consider covariate-adaptive designs (Lin and Su, 2012; Ma and Hu, 2013; etc.) that directly utilize continuous covariates without discretization. However, related theoretical work is limited in literature. (2) The covariate effects are assumed to be additive in true models to derive theoretical results. We may apply the similar idea to more complicated cases, such as interaction terms, higher order terms. (3) We studied hypothesis testing for logistic regression by assuming conditions on within-stratum imbalances. More general properties may be studied under conditions on marginal imbalances similar to the linear model framework. Moreover, properties of other generalized linear model are also desired. (4) The proposed properties are based on clinical trials with two treatments, which can be generalized to multiple treatments (Tymofyeyev, Rosenberger and Hu, 2007). Those topics are left for future research.

## Bibliography

- [1] Aickin, M. (2009). A simulation study of the validity and efficiency of design-adaptive allocation to two groups in the regression situation. *The International Journal of Biostatistics*, **5**, 1, 14.
- [2] Anderson, H., Hopwood, P., Stephens, R. J., Thatcher, N., Cottier, B., Nicholson, M., Milroy, R., Maughan, T. S., Falk, S. J., Bond, M. G., Burt, P. A., Connolly, C. K., McIllmurray, M. B. and Carmichael, J. (2009). Gemcitabine plus best supportive care (BSC) vs BSC in inoperable non-small cell lung cancer—a randomized trial with quality of life as the primary outcome. *British Journal of Cancer*, **83**, 4, 447–453.
- [3] Ashley, E. A., Butte, A. J., Wheeler, M. T., Chen, R., Klein, T. E., Dewey, F. E., Dudley, J. T., Ormond, K. E., Pavlovic, A., Hudgins, L., Gong, L., Hodges, L. M., Berlin, D. S., Thorn, C. F., Sangkuhl, K., Hebert, J. M., Woon, M., Sagreiya, H., Whaley, R., Morgan, A. A., Pushkarev, D., Neff, N. F., Knowles, J. W., Chou, M., Thakuria, J., Rosenbaum, A., Zaranek, A. W., Church, G., Greely, H. T., Quake, S. R. and Altman, R. B. (2010). Clinical evaluation incorporating a personal genome. *The Lancet*, **375**, 1525–1535.
- [4] Atkinson, A. C. (1982). Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika*, **69**, 61–67.

- [5] Begg, C. B. and Iglewicz, B. (1980). A treatment allocation procedure for sequential clinical trials. *Biometrics*, **36**, 81–90.
- [6] Birkett, N. J. (1985). Adaptive allocation in randomized controlled trials. *Controlled Clinical Trials*, **6**, 146–155.
- [7] Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika*, **58**, 403–417.
- [8] Eisele, J. (1994). The doubly adaptive biased coin design for sequential clinical trials. *Journal of Statistical Planning and Inference*, **38**, 249–262.
- [9] Eisele, J. and Woodroffe, M. (1995). Central limit theorems for doubly adaptive biased coin designs. *The Annals of Statistics*, **23**, 234–254.
- [10] Forsythe, A. B. (1987). Validity and power of tests when groups have been balanced for prognostic factors. *Computational Statistics Data Analysis*, **5**, 193–200.
- [11] Feinstein, A. R. and Landis, J. R. (1976). The role of prognostic stratification in preventing the bias permitted by random allocation of treatment. *Journal of Chronic Diseases*, **29**, 277–284.
- [12] Gail, M. H., Wieand, S., Piantadosi, S. (1984). Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*, **71**, 3, 431–444.
- [13] Gail, M. H. (1988). The effect of pooling across strata in perfectly balanced studies. *Biometrics*, **44**, 1, 151–162.
- [14] Gallo, P., Chuang-Stein, C., Dragalin, V., Gaydos, B., Krams, M. and Pinheiro, J. (2006). Adaptive designs in clinical drug development executive summary



- of the phrma working group. *Journal of Biopharmaceutical Statistics* **16** 275–283.
- [15] Green, S. B. and Byar, D. P. (1978). The effect of stratified randomization on size and power of statistical tests in clinical trials. *Journal of Chronic Diseases*, **31**, 445–454.
- [16] Gridelli, C., Gallo, C., Shepherd, F. A., Illiano, A., Piantedosi, F., Robbiati, S. F., Manzione, L., Barbera, S., Frontini, L., Veltri, E., Findlay, B., Cigolari, S., Myers, R., Ianniello, G. P., Gebbia, V., Gasparini, G., Fava, S., Hirsh, V., Bezzjak, A., Seymour, L. and Perrone, F. (2003). Gemcitabine plus vinorelbine compared with cisplatin plus vinorelbine or cisplatin plus gemcitabine for advanced non-small-cell lung cancer: a phase III trial of the Italian GEMVIN Investigators and the National Cancer Institute of Canada Clinical Trials Group. *Journal of Clinical Oncology*, **21**, 16, 3025–3034.
- [17] Hu, F. (2012). Statistical issues in trial design and personalized medicine. *Clinical Investigation*, **2**, 121–124.
- [18] Hu, Y. and Hu, F. (2012). Asymptotic properties of covariate-adaptive randomization. *Annals of Statistics*, **40**, 1794–1815.
- [19] Hu, F. and Rosenberger, W. F. (2003). Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. *Journal of the American Statistical Association* **98**, 671–678.
- [20] Hu, F. and Rosenberger, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. John Wiley and Sons. Wiley Series in Probability and Statistics.

- [21] Hu, F. and Zhang, L. (2013). On the theory of covariate-adaptive designs. *Annals of Statistics*, submitted.
- [22] Hu, F. and Zhang L. (2004). Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. *The Annals of Statistics*, **32**, 268–301.
- [23] Iacono, A. T., Johnson, B. A., Grgurich, W. F., Youssef, J. G., Corcoran, T. E., Seiler, D. A., Dauber, J. H., Smaldone, G. C., Zeevi, A., Yousem, S. A., Fung, J. J., Burckart, G. J., McCurry, K. R. and Griffith, B. P. (2006). A randomized trial of inhaled cyclosporine in lung-transplant recipients. *The New England Journal of Medicine*, **354**, 2, 141–150.
- [24] Ivanova, A. V. (2003). A play-the-winner type urn model with reduced variability. *Metrika*, **58**, 1–13.
- [25] Krueger, G. G., Langley, R. G., Leonardi, C., Yeilding, N., Guzzo, C, Wang, Y., Dooley, L. T. and Lebwohl, M. (2007). A human interleukin-12/23 monoclonal antibody for the treatment of psoriasis. *The New England Journal of Medicine*, **356**, 6, 580–592.
- [26] Jakob, S. M., Ruokonen, E., Grounds, R. M., Sarapohja, T., Garratt, C., Pocock, S. J., Bratty, J. R. and Takala, J. (2012). Dexmedetomidine vs midazolam or propofol for sedation during prolonged mechanical ventilation: two randomized controlled trials. *Journal of American Medical Association*, **307**, 11, 1151–1160.
- [27] Kundt, G. (2009). Comparative evaluation of balancing properties of stratified randomization procedures. *Methods Inf Med*, **48**, 129–134.

- [28] Li, Y., Sheu, C.-C., Ye, Y., Andrade, M. D., Wang, L., Chang, S.-C., Aubry, M. C., Aakre, J. A., Allen, M. S., Chen, F., Cunningham, J. M., Deschamps, C., Jiang, R., Lin, J., Marks, R. S., Pankratz, V. S., Su, L., Li, Y., Sun, Z., Tang, H., Vasmatazis, G., Harris, C. C., Spitz, M. R., Jen, J., Wang, R., Zhang, Z.-F., Christiani, D. C., Wu, X. and Yang, P. (2010). Genetic variants and risk of lung cancer in never smokers: a genome-wide association study. *Lancet Oncology*, **11**, 321–330.
- [29] Lin, Y. and Su, Z. (2012) Balancing continuous and categorical baseline covariates in sequential clinical trials using the area between empirical cumulative distribution functions. *Statistics in Medicine*, **31**, 1961–1971.
- [30] Lipkin, S. M., Chao, E. C., Moreno, V., Rozek, L. S., Rennert, H., Pinchev, M., Dizon, D., Rennert, G., Kopelovich, L. and Gruber, S. B. (2010). Genetic variation in 3-hydroxy-3-methylglutaryl CoA reductase modifies the chemopreventive activity of statins for colorectal cancer. *Cancer Prevention Research*, **3**, 597–603.
- [31] Ma, Z., Hu, F. (2013). Balancing continuous covariates based on Kernel densities. *Contemporary Clinical Trials*, **34**, 2, 262–269.
- [32] Melfi, V., Page, C. (2000). Estimation after adaptive allocation. *Journal of Statistical Planning and Inference*, **87**, 353–363.
- [33] Melfi, V., Page, C. and Geraldès, M (2001). An adaptive randomized design with application to estimation. *Canad. J. Statist*, **29**, 107–116.
- [34] McEntegart D. J. (2003). The pursuit of balance using stratified and dynamic randomization techniques: an overview. *Drug Information Journal*, **37**, 2, 293–308.

- [35] Molander, A., Warfvinge, J., Reit, C. and Kvist, T. (2007). Clinical and radiographic evaluation of one- and two-visit endodontic treatment of asymptomatic necrotic teeth with apical periodontitis: a randomized clinical trial. *Journal of Endodontics*, **33**, 10, 1145–1148.
- [36] Ohtori, S., Miyagi, M., Eguchi, Y., Inoue, G., Orita, S., Ochiai, N., Kishida, S., Kuniyoshi, K., Nakamura, J., Aoki, Y., Ishikawa, T., Arai, G., Kamoda, H., Suzuki, M., Takaso, M., Furuya, T., Kubota, G., Sakuma, Y., Oikawa, Y., Toyone, T. and Takahashi, K. (2012). Efficacy of epidural administration of anti-interleukin-6 receptor antibody onto spinal nerve for treatment of sciatica. *European Spine Journal*, **21**, 10, 2079–2084.
- [37] Pocock, S. J. and Simon, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*, **3**, 103–115.
- [38] Pocock, S. J. (1982). Statistical aspects of clinical trial design. *The Statistician*, **31**, 1–18.
- [39] Rosenberger, W. F. and Sverdlov, O. (2008). Handling covariates in the design of clinical trials. *Statistical Science*, **23**, 404–419.
- [40] Shao, J., Yu, X. and Zhong, B. (2010). A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika*, **97**, 2, 347–360.
- [41] Simon, R (1979). Restricted randomization designs in clinical trials. *Biometrics*, **35**, 2, 503–512.
- [42] Taves, D. R. (1974). Minimization: a new method of assigning patients to treatment and control groups. *Clin Pharmacol Therap*, **15**, 443–453.

- [43] Taves, D. R. (2010). The use of minimization in clinical trials. *Contemporary Clinical Trials*, **31**, 180–184.
- [44] Toorawa, R., Adena, M., Donovan, M., Jones, S. and Conlon, J. (2009). Use of simulation to compare the performance of minimization with stratified blocked randomization. *Pharmaceutical Statistics*, **8**, 264–278.
- [45] Tymofyeyev, Y., Rosenberger, W.F. and Hu, F. (2007). Implementing optimal allocation in sequential binary response experiments. *Journal of the American Statistical Association*, **102**, 224–234.
- [46] Tu, D., Shalay, K and Pater, J. (2000). Adjustment of treatment effect for covariates in clinical trials: statistical and regulatory issues. *Drug Information Journal*, **31**, 180–184.
- [47] Wei, L. J. (1978). An application of an urn model to the design of sequential controlled clinical trials. *Journal of American Statistical Association*, **34**, 2, 511–523.
- [48] Wei, L. J. and Durham, S. D. (1978). The randomized play-the-winner rule in medical trials. *Journal of the American Statistical Association*, **73**, 840–843.
- [49] Zelen, M. (1974). The randomization and stratification of patients to clinical trials. *Journal of Chronic Diseases*, **27**, 365–375.
- [50] Zhang, L., Hu, F., Cheung, S, H. and Chan, W. S. (2007). Asymptotic properties of covariate-adjusted response-adaptive designs. *The Annals of Statistics*, **35**, 1166–1182.