Adaptive Designs Utilizing Covariates for Precision Medicine and Their Statistical Inference

by Wanying Zhao

B.S. in Statistics, June 2011, Dongbei University of Finance and Economics
M.S. in Statistics, May 2013, University of Illinois Urbana-Champaign

A Dissertation submitted to

The Faculty of
The Columbian College of Arts and Sciences
of The George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

May 21, 2017

Dissertation directed by

Feifang Hu
Professor of Statistics

The Columbian College of Arts and Sciences of The George Washington University certifies that Wanying Zhao has passed the Final Examination for the degree of Doctor of Philosophy as of March 1, 2017. This is the final and approved form of the dissertation.

Adaptive Designs Utilizing Covariates for Precision Medicine and Their Statistical Inference

Wanying Zhao

Dissertation Research Committee:

Feifang Hu, Professor of Statistics, Dissertation Director

Hosam M. Mahmoud, Professor of Statistics, Committee Member

Hua Liang, Professor of Statistics, Committee Member

Dedication

To my parents for their unconditional love and support. I really appreciate their sacrifices for putting me through the best education possible. I would not be who I am today without them.

Acknowledgments

I would like to express the most sincere and deepest gratitude to my dissertation advisor, Professor Feifang Hu, for his encouragement and guidance on my study and research, for his patience and immense knowledge. His enthusiasm and aspiration in research motivated me to keep moving forward. Whenever I met difficulties, he was always there to lead me and provide insightful advice. The work would not have been possible without him. Also, his positive attitude and humor made every meeting with him very enjoyable. It has been an honor to be his first Ph.D. student in the George Washington University. I could not have imagined having a better advisor and mentor for my Ph.D. study.

I would also like to thank Professors Hosam Mahmoud, Hua Liang, Judy Wang and Jianhui Zhou for being my thesis committee members, providing critical comments and constructive suggestions.

My sincere thanks also go to my Ph.D. peers for the time we were working together before final exams, and for all the fun we have had in the last four years. I am especially thankful to Fan Wang for countless discussions on simulations, and Xin Li for correcting grammar and notations in my writings. Moreover, I would like to thank Wei Ma and Yanqing Hu. They provided not only great help for the dissertation, but also valuable advice for my career after graduation.

Last but not the least, I would like to thank my parents, Weihong Wang and Jiji Zhao, for their constant love and support throughout my life. Jingyao Hou, my boyfriend, was always encouraging and accompanying me through tough times.

Abstract of Dissertation

Adaptive Designs Utilizing Covariates for Precision Medicine and Their Statistical Inference

Precision medicine takes account of individual characteristics for disease treatment and prevention. To develop precision medicine, more covariates of patients are under consideration in clinical trials. Based on different roles played in clinical studies, covariates can be categorized into two types: (1) prognostic covariates, which are used to balance treatment allocations for a simple treatment comparison, and (2) predictive covariates, which are used to select a suitable treatment based on efficiency and ethics.

In the literature, covariate-adaptive designs have been proposed to deal with prognostic covariates, while covariate-adjusted response-adaptive designs utilize information of predictive covariates. Theoretical properties of statistical inference under these two procedures are provided based on linear models in Chapter 2 and Chapter 3. We show that, when the covariate, which is used in the randomization procedure, is excluded from the working model of inference, the hypothesis testing to compare treatment effects under CARA designs is not always valid, depending on the unknown parameters in the employed model and the choice of the allocation function. While, under covariate-adaptive designs, the hypothesis testing is usually conservative when prognostic covariates, which are balanced in randomization, are omitted in statistical inference.

In Chapter 4, we propose a general framework of new CARA designs, which can incorporate both prognostic and predictive covariates in the randomization procedure simultaneously. Similar problems of hypothesis testing are studied under new designs when prognostic covariates are excluded from the final analysis. Some mild conditions for imbalances and target allocation proportions need to be satisfied to derive

asymptotic properties of test statistics. It is proved that the test for comparing treatment effects is usually conservative under new designs. One possible solution to this problem is the bootstrap method. New CARA designs have advantages leading to improvement on average outcomes, but still allowing high power.

Continuous covariates are often used in clinical trials. In Chapter 5, the discrete assumption for prognostic covariates is relaxed and similar inference problems are studied under the CAR design and the new CARA design. Asymptotic distributions of test statistics are obtained under both the null and alternative hypotheses. We show that the tests for comparing treatment effects are conservative under both designs when prognostic covariates are omitted for statistical inference regardless of whether they are discrete or continuous.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview of Randomization Designs in Clinical Trials

1.1.1 Randomization

Clinical trials are prospective research studies of human subjects, which are designed to answer specific questions about biomedical or behavioral interventions, including new treatments, such as novel drugs, vaccines or medical devices, and known interventions. They are used to generate data on safety and efficacy of new biomedical or behavioral interventions for further study and comparison. Randomized clinical trials are considered the "gold standard" of measuring the effects of the intervention. They have been in practice since 1929 when Ronald Fisher implemented the first randomization scheme, which was the flip of a coin, in a tuberculosis trial [9]. The advantage of randomization is that it enables one to evaluate whether the intervention itself, as opposed to other factors, causes the observed outcomes. There is persuasive evidence that randomized clinical trials are superior to other study designs in measuring the true effect of an intervention. In general, randomization procedures can be categorized into three types: complete randomization, restricted randomization and adaptive randomization.

Complete randomization is simplest in terms of data analysis and convenience, similar

to "repeatedly tossing a fair coin". Under this design, subjects are randomly assigned to each treatment with equal probability. Suppose there are two treatments 1 and 2 in a clinical trial. Let N be the total number of subjects involved and $I_m, m = 1, \ldots, N$, index the treatment (1 if treatment 1; 0 if treatment 2) for the m-th subject. Assignments I_m's are independently assigned with $\mathbb{P}(I_m = 1) = 1/2,\ m = 1, \ldots, N$. Complete randomization is easy to implement, flexible and robust against both selection and accidental biases. Any number of treatment can be investigated. However, severe imbalance between different treatment arms may exist, especially when the number of randomized subjects is small [29], which will lead to larger experimental error compared with other designs.

In clinical trials, a balanced number of subjects for each treatment is desirable so that any observed difference between treatment groups can be attributed to treatment effects. Without considering covariates, *restricted randomization* refers to any procedure used with random assignments to achieve approximate balance between different treatment groups in size. The two most popular restricted randomization procedures are permuted block design and biased coin design proposed by Efron [11]. The randomization scheme of *permuted block design* consists of a sequence of blocks such that each block contains a pre-specified number of treatment assignments in a random order, so that the balanced allocation can be achieved at the completion of each block. Imbalance only occurs when the last block is not filled. Suppose permuted block design is conducted in a two-armed trial (treatments 1 and 2) with block size $2m,\ m \in \mathbb{Z}^+$. Then, each block consists of $2m$ subjects and the assignments are determined by a permutation of exactly m ones and m twos. The major disadvantage of permuted block design is that even if the block size is large and randomly varied, the procedure can lead to selection bias [29]. *Biased coin design* is introduced as a procedure to minimize selection bias while maintaining allocation balance throughout the trial. For two treatments 1 and 2, the imbalance D_m is defined as the difference

between the number of subjects assigned to treatment 1 and the number of subjects assigned to treatment 2 after m assignments, i.e., $D_m = N_{1,m} - N_{2,m}$. Then, the probability of the $(m+1)$-st subject randomized to treatment 1 is

$$\mathbb{P}\left(I_{m+1} = 1\right) = \begin{cases} p^*, & \text{if } D_m < 0; \\ 1/2, & \text{if } D_m = 0; \\ 1 - p^*, & \text{if } D_m > 0, \end{cases}$$

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

In any clinical trials, subjects are enrolled sequentially. *Adaptive randomization designs* are constructed based on information from all previous subjects already in the trial to achieve some goals, such as to maintain the integrity and validity of the trial, while providing flexibility in identifying the optimal treatment. Because of the efficiency and ethicality of adaptive designs, they have gained more popularity in the last few decades, both in the literature and in practice. More discussion of adaptive designs will be given in Section 1.1.2.

1.1.2 Adaptive Designs

The Pharmaceutical Research and Manufacturers of America (PhRMA) Working Group on Adaptive Design refers to *an adaptive design* as a clinical trial design that uses accumulating data to decide on how to modify aspects of the study as it continues, without undermining the validity and integrity of the trial [16]. According to Chow and Chang [9], adaptive designs are very attractive due to the following reasons. "First, it reflects medical practice in the real world. Second, it is ethical with respect to both efficacy and safety (toxicity) of the test treatment under investigation. Third, it is not only flexible, but also efficient in the early and late phase

of clinical development." There are various adaptive designs depending on the data and adaptations used, such as group sequential, sample size re-estimation, adaptive dose finding, adaptive treatment-switching, adaptive seamless phase II/III trial, and so on. In this work, we only focus on adaptive designs that sequentially modify the treatment allocation probabilities of subjects based on accrued data in order to achieve balance with respect to treatments and covariates and/or increase efficiency and safety.

The most commonly used historical information from the on-going clinical trials by adaptive designs includes treatment assignments, responses and covariate profiles of previous subjects and the covariate profile of the current subject. Then, adaptive designs can be categorized into three major types, based on the history information, which are response-adaptive designs (RAR), covariate-adaptive designs (CAR) and covariate-adjusted response-adaptive designs (CARA).

Response-adaptive designs were developed to provide an allocation scheme for the incoming subject based on responses and treatment assignments of previous subjects, who were already in the trial. There are two main families of response-adaptive designs in the literature: urn model based designs and doubly-adaptive biased coin designs (DBCD). The preliminary idea of urn model based designs is to assign more subjects to the better treatment. Two typical urn model based designs are randomized play-the-winner model and drop-the-loser model. For a clinical trial with two treatments A and B, the treatment assignment of *randomized play-the-winner* [52] is determined by an urn, that initially contains α balls of each type. A ball is randomly selected with replacement and the corresponding treatment is allocated to the current subject in the trial. Add β balls of the same type after a success response, otherwise add β balls of the other type. The procedure repeats until all subjects are assigned to a treatment. Under *drop-the-loser design* [26], the inferior treatment

detected from subjects' responses will be dropped based on some pre-specified criteria in the on-going trial. However, these are ad-hoc designs and subject to high variability. The doubly-adaptive biased coin design (*DBCD*) was first proposed by Eisele [12] and Eisele and Woodroofe [13]. It takes account of the proportion of subjects assigned to each treatment and the current estimate of the target allocation proportion to determine the allocation probability for the next subject. For clinical trials with two treatments 1 and 2, let $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ be the corresponding unknown parameters from the model employed and $\rho\left(\boldsymbol{\theta}_1,\ \boldsymbol{\theta}_2\right)$ is the desired allocation proportion for treatment 1. At the beginning of the trial, $2n_0$ subjects are assigned to the two treatments using some restricted randomization procedures, such as the permuted block design, to make sure that both treatment arms have equal number of subjects. After m assignments ($m \geq 2n_0$), estimators of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ can be obtained as $\hat{\boldsymbol{\theta}}_{1,m}$ and $\hat{\boldsymbol{\theta}}_{2,m}$, respectively. DBCD proposed by Eisele [12] and Eisele and Woodroofe [13] assigns the $(m+1)$-st subject to treatment 1 with probability $g\left(N_{1,m}/m,\ \hat{\rho}_m\right)$, where g is an allocation function, $N_{1,m}$ is the number of subjects randomized to treatment 1 after m assignments and $\hat{\rho}_m = \rho\left(\hat{\boldsymbol{\theta}}_{1,m}, \hat{\boldsymbol{\theta}}_{2,m}\right)$. Later, Hu and Zhang [22] extended the idea of DBCD and proposed a general family of allocation functions. Under this design, the $(m+1)$-st subject will be randomized to treatment 1 with probability $g^{(\alpha)}\left(N_{1,m}/m,\ \hat{\rho}_m\right)$, where $\alpha \geq 0$ and $g(x,y)$ is the allocation function satisfying:

$$g^{(\alpha)}(0,y) = 1, \quad g^{(\alpha)}(1,y) = 0,$$

$$g^{(\alpha)}(x,y) = \frac{y(y/x)^{\alpha}}{y(y/x)^{\alpha} + (1-y)((1-y)/(1-x))^{\alpha}}.$$

The asymptotic properties were established based on widely satisfied conditions. This new family of designs was proved to be less variable than both the randomized play-the-winner [52] and the adaptive randomization design proposed by Melfi, Page and

Geraldes [36], and it can target any given allocation proportion. More discussion about DBCD and optimal allocation proportion can be found in Melfi and Page [35], Hu and Rosenberger [21], Tymofyeyev, Rosenberger and Hu [51].

Covariate-adaptive designs incorporate covariates of subjects in clinical trials to balance covariates across treatment groups. Here, covariates refer to factors that have influences on the response of subjects to a treatment, such as gender, age, investigation centers, blood pressure, a particular genotype, and so on. Under covariate-adaptive designs, the treatment assignment of the current subject is determined by covariate profiles of all previous subjects and the current subject's covariates, together with all treatment assignments of previous subjects. More detailed covariate-adaptive designs are discussed in Section 1.1.3.

Covariate-adjusted response-adaptive designs only have a brief history in the literature. Unlike response-adaptive designs, which assume homogeneity for subjects in a treatment group, CARA designs allow heterogeneity of subjects based on their characteristics even though they have the same treatment assignment. They sequentially assign subjects to treatments with probabilities based on previous treatment assignments, responses, covariates and the current subject's covariates. They can preserve the advantages of response-adaptive designs with the incorporation of covariates. Under CARA designs, subjects can get a better treatment based on their personal information. Details of CARA designs are introduced in Section 1.1.4.

1.1.3 Covariate-Adaptive Designs

In most cases, some covariates of subjects can have a deterministic influence on the responses of subjects to a treatment. For example, the efficacy of a cancer treatment may depend on age, and whether the patient is a smoker or not. A well-balanced design over covariates of interest not only is preferred in the viewpoint of statistical

efficiency, but also can enhance trials by increasing the credibility of the trial, precision of subgroup or interim analysis, and robustness to model mis-specification [34]. Therefore, it has become more and more important to balance treatment allocation with respect to key covariates at the design stage [28].

A natural idea to achieve balance over covariates is stratification. Strata are constructed based on different combinations of covariate levels. Restricted randomization designs can be separately applied to each stratum to obtain balance within each stratum, and further to obtain the good overall balance. Two commonly used restricted randomization methods within strata are permuted block design and Efron's biased coin design. *Stratified permuted block design* is the most popular covariate-adaptive randomization procedure currently used in clinical trial practice. It can be easily implemented without sophisticated technologies and computation. However, if there are too many strata in relation to the target sample size, typically due to too many covariates or too many levels within individual covariates, some of the strata will be empty or sparse. This would lead to severe allocation imbalance on stratum level and further on the overall level [37]. Moreover, it is lack of randomness on treatment allocation and vulnerable to selection bias [33].

Several marginal methods (also referred as *minimization* in the literature) were introduced as alternatives to stratified permuted block designs to deal with a large number of covariates. Taves [47] proposed the first method, which was deterministic to allocate treatments in order to minimize imbalances over important covariates. Pocock and Simon [38] extended Taves' idea by incorporating randomness and established a generalization of minimization to control treatment imbalance over covariate margins. In a clinical trial with two treatments 1 and 2, suppose there are J covariates. Let $N_{jz_jk}(n)$ denote the number of subjects in treatment k on level z_j of the j-th covariate, $j = 1, \ldots, J,\ k = 1, 2$, after assigning n sub-

jects. Let $(z_1, \ldots, z_J)$ be the covariate information for the $(n+1)$-st subject. Then, the marginal imbalance among treatment groups with respect to the j-th covariate is defined as $D_j(n) = N_{jz_j1}(n) - N_{jz_j2}(n)$. A weighted sum of marginal imbalances $D(n) = \sum_{j=1}^{J} \omega_j D_j(n)$ is computed to determine the treatment allocation for the $(n+1)$-st subject by using Efron's biased coin design,

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

where $1/2 < p^* < 1$ is the biased coin probability. However, large imbalances tend to exist within an individual stratum [28, 45]. Indeed, its asymptotic properties were shown by Ma, Hu and Zhang [32] that the overall imbalance and marginal imbalances are bounded in probability, however, the imbalances within strata increase at the rate of $\sqrt{N}$, as the sample size increases. Another generalization of minimization was proposed by Wei [52] based on Friedman's urn model. Let U_{jz_j} represent the urn of the j-th covariate with level z_j. At the beginning of a trial, each urn contains α_1 type 1 balls and α_2 type 2 balls. Suppose the $(n+1)$-st subject has covariate information $(z_1, \ldots, z_J)$. Let $Y_{jz_jk}(n)$ denote the number of type k balls in the urn U_{jz_j} after n assignments, $k = 1, 2$. Then, the imbalance with respect to the urn U_{jz_j} is defined as $D_{jz_j}(n) = (Y_{jz_j1}(n) - Y_{jz_j2}(n))/(Y_{jz_j1}(n) + Y_{jz_j2}(n))$. The urn with the largest absolute value of imbalance is used to determine the treatment assignment. A ball is randomly draw from that urn and replaced. If the selected ball is type k, then the subject is assigned to treatment k, $k = 1, 2$. Also, α_k type k balls and $\beta_k \geq 0$ balls of the other type will be added to each urn with observations. More extended versions of minimization can be found in Signorini *et al.* [45], Heritier, Gebski and Pillai [18], Russell *et al.* [42], Lebowitsch *et al.* [31], etc. The minimization method is

increasingly used due to its reduction on the marginal and overall imbalances found by simulation studies [28, 49, 53]. Taves [48] pointed out that Pocock and Simon's marginal procedure was implemented in over 400 clinical trials from 1989 to 2008. However, the advantages of these covariate-adaptive designs have only been substantiated by simulation results. As indicated in Rosenberger and Sverdlov [40], there is no theoretical justification about these designs.

Hu and Hu [24] proposed a general family of covariate-adaptive designs, which includes some new designs and many existing designs as special cases, such as stratified permuted block designs, minimization, and Efron's biased coin design. In particular, they employ the Efron's biased coin design to assign the new subject with higher probability to the treatment that has a smaller weighted sum of imbalances at three different levels

$$Imb_n^{(k)} = \omega_0 \left[D_n^{(k)}\right]^2 + \sum_{i=1}^{J} \omega_{m,j} \left[D_n^{(k)}(j; z_j)\right]^2 + \omega_s \left[D_n^{(k)}(z_1, \ldots, z_J)\right]^2,$$

where $D_n^{(k)}$, $D_n^{(k)}(j; z_j)$ and $D_n^{(k)}(z_1, \ldots, z_J)$ are the overall imbalance, marginal imbalance and within-stratum imbalance of treatment k, respectively. The procedure simultaneously eliminates imbalances at all three levels. They also obtained the theoretical properties for the new family of covariate-adaptive designs under certain conditions, and proved that the overall imbalance, marginal imbalances and within-stratum imbalances are all bounded in probability under this family of designs, hence it can yield more balanced allocations.

1.1.4 Covariate-Adjusted Response-Adaptive Designs

Randomization designs depending on responses of subjects have been proved by Rosenberger and Hu [39] to have a significant advantage on minimizing the number of

subjects assigned to the inferior treatment, but still allowing statistical inference with high power. However, they do not account for covariates in the randomization procedure, which may have strong impact on responses. Therefore, covariate-adjusted response-adaptive (CARA) designs arose in clinical trials to assign subjects to the most appropriate treatment based on their covariate profiles. Under a CARA design, the treatment allocation probability for the current subject is based on all history information of previously assigned subjects, together with the covariate information of the current subject.

The history of CARA designs is very short, and only a few sources of literature focus on this direction. Bandyopadhyay and Biswas [5] considered a linear model to utilize covariate information with continuous responses for comparing two treatments. Under the assumption that larger response corresponds to higher efficacy, the new subject will be randomized to treatment 1 with probability $\mathbb{P}(I_{n+1} = 1) = \Phi(d_n/c)$, where d_n is the difference between treatment means from the previous n subjects, c is a pre-specified scale constant and Φ is the standard normal CDF. However, performance of this procedure depends on the choice of c, and smaller values of c may lead to severe treatment allocation imbalances, hence great power losses. Moreover, covariates of the new subject are not incorporated in this procedure. Atkinson [2] considered adaptive biased coin designs for K treatments based on a linear model. Atkinson and Biswas [3, 4] improved the allocation rule of Bandyopadhyay and Biswas [5] by proposing adaptive biased coin designs and Bayesian adaptive biased coin designs for clinical trials with normal responses. Their procedures combined both efficiency and ethical considerations based on a weighted D_A-optimal criterion. Rosenberger, Vidyashankar and Agarwal [41] considered a CARA design for binary responses that used a logistic regression model $\mathsf{logit}\left(\mathbb{P}(Y_k = 1|\boldsymbol{X} = \boldsymbol{x})\right) = \boldsymbol{\theta}_k'\boldsymbol{x}$, where $\boldsymbol{\theta}_k$ is the vector of model parameters for treatment k, $k = 1, 2$. Then, the allocation probability of

assigning the new subject to treatment 1 is

$$\mathbb{P}(I_{n+1}=1)=F\left(\left(\hat{\boldsymbol{\theta}}_{n1}-\hat{\boldsymbol{\theta}}_{n2}\right)'\boldsymbol{x}_{n+1}\right),$$

where F is the standard logistic CDF and $\hat{\boldsymbol{\theta}}_n$'s are MLEs of model parameters calculated from the previous n subjects. A simulation shows that their procedure significantly reduced the percentage of treatment failures. But various weaknesses are presented in the above methods, including high variability, ignoring covariate information of the current subject, limited to specific cases of responses or two treatments, lack of theoretical justification and foundations for statistical inference study [23].

Zhang *et al.* [54] laid out a general framework of CARA randomization procedures for a very broad class of models with the form $\mathbb{E}\left(Y_k|\boldsymbol{X}\right)=p_k\left(\boldsymbol{\theta}_k,\boldsymbol{X}\right)$ for K treatments, where $p_k(\cdot,\cdot)$ are known functions and $\boldsymbol{\theta}_k$ are unknown parameters, $k=1,\ldots,K$. This family of CARA designs includes the logistic regression model and the normal linear regression as special cases. When a new subject enters the trial and is ready for randomization, the assignment will be based on all history information, including treatment assignments, covariate profiles and responses of previous subjects, as well as the covariate information of the new subject. The new subject is randomized to treatment k with probability

$$\mathbb{P}\left(I_{n+1}=k\right)=\pi_k(\hat{\boldsymbol{\theta}}_k,\boldsymbol{X}_{n+1}),$$

where $\pi_k(\cdot,\cdot)$ are some given allocation functions satisfying $\pi_1+\cdots+\pi_K\equiv 1$. By choosing different allocation functions, this family of CARA designs can achieve different goals. The asymptotic properties of estimators and allocation proportions were comprehensively studied under certain widely satisfied conditions, which provides a theoretical foundation for comparing treatment effects under this randomization

method. Recently, Hu, Zhu and Hu [23] extended this procedure by simultaneously taking efficiency and medical ethics into consideration. They proposed a new family of CARA designs based on two general measurements of efficiency and ethics. The proposed family also unified several well-known designs, such as DBCD by Eisele and Woodroofe [13] and Hu and Zhang [22] as special cases. Asymptotic properties were established on the allocation proportion and the parameter estimates for binary covariates and it remains unknown for the case of continuous covariates.

1.2 Statistical Inference of Randomization Designs Utilizing Covariates

As covariates have gained increasing importance in clinical trials, it is essential to study the theoretical properties of hypothesis testings for both treatment effects and covariate effects. Although various of adaptive designs utilizing covariates have been proposed and used in clinical trials, the discussion of statistical inference for these designs is very limited. In practice, covariate information is not considered in conventional tests. Whether the most classical tests, such as statistical methods based on maximum likelihood estimators or moment estimators, are still valid remains a concern with the incorporation of covariates in randomization procedures [19]. It is now generally accepted that a valid test can be obtained, if all covariates, that are used in randomization, are incorporated in the inference procedures. This was suggested by Forsythe [15] through simulation studies, and theoretically pointed out by Shao, Yu and Zhong [44].

However, not all covariates, which are used in the randomization procedure, are included in the final analysis in practice for the following three main reasons [32]: (i) it is very difficult to incorporate some covariates in the analysis model, for example, in the viewpoint of credibility of a trial, investigation sites, that are needed to be balanced across treatment groups, should not be included in the inference pro-

cedure; (ii) including too many covariates usually requires much more complicated modeling techniques, such as model selection, methods dealing with dependent covariates; and (iii) correct model specification is required, which is usually unknown in practice.

There have been doubts about validity of statistical inference for randomization designs incorporating covariates, especially when covariates are fully or partially omitted in inference procedures. Birkett [7] and Forsythe [15] had raised concerns about the validity of unadjusted analysis under covariate-adaptive designs for continuous normal responses. Both of them found out the conservativeness of testing treatment effects using a two-sample t-test under covariate-adaptive designs, and there is no obvious power improvement, when the nominal critical value is used. Due to the complex inter-dependence among covariates, treatment assignments and responses introduced by the allocation scheme, investigation of this concern has been mostly done under restricted conditions or by limited simulation studies.

The first attempt of theoretical results was made by Shao, Yu and Zhong [44] under a simple linear model for two treatments $Y_{ij} = \mu_j + bZ_i + \epsilon_{ij}$, where Y_{ij} is the response for subject i with treatment j, Z_i's are independent and identically distributed univariate covariates, μ_j and b are unknown parameters, and ϵ_{ij}'s are independent and identically distributed random errors and independent of Z_i's. They theoretically proved that a two-sample t-test without covariates included in the final analysis is conservative under stratified randomization designs, and a bootstrap t-test is proposed to adjust the estimator of variance and then achieve nominal Type I error. Results of power comparison depend on the magnitude of treatment effect. However, these results are only applicable to a covariate-adaptive biased coin design, which is a stratified design to apply the biased coin method to each stratum but less commonly used in practice.

Barbáchano and Coad [6] have also studied the analysis of covariance t-test under the linear model framework, when covariate-adaptive randomization is used. They proposed that the nominal significance level can be maintained and the test is slightly more powerful compared with complete randomization, especially when the number of covariates is large.

With a linear model, Ma, Hu and Zhang [32] showed the conservativeness of classical tests to compare treatment effects under more general covariate-adaptive designs. Two assumed conditions that the overall imbalance and marginal imbalances are all bounded in probability can be easily satisfied by various of covariate-adaptive designs, including stratified permuted block designs and Pocock and Simon's minimization. Similar bootstrap method, as in Shao, Yu and Zhong [44], can be employed to achieve nominal Type I error. As many influential covariates are identified to be linked with certain disease and are often included in the regression model, they also showed that the test for significance of covariates is still valid.

Most of the existing work in the literature on covariate-adaptive randomization is based on linear models, where the balance with respect to important covariates across treatment groups leads to a more efficient trial. However, for nonlinear models, balance may yield a less efficient trial [10]. In the case of dichotomous responses, Feinstein and Landis [14] and Green and Byar [17] studied hypothesis tests comparing successful rate between two treatment groups on a special case of two strata and two treatments. Under this restricted assumption, they showed that the overestimation of the variance of estimated treatment effects under stratified randomization results in conservative hypothesis tests. A theoretical study of asymptotic results for covariate-adaptive randomization is provided by Shao and Yu [43] under generalized linear model with possibly unknown link functions, $g\left[\mathbb{E}(Y_{ij}|\boldsymbol{Z}_i)\right] = \mu_j + \boldsymbol{\beta}'\boldsymbol{Z}_i$, where $g(\cdot)$ is a link function, $\boldsymbol{Z}_i$ is the covariate vector for subject i, $\boldsymbol{\beta}$ is a vector of unknown

parameters measuring the effects of covariates, Y_{ij} and μ_j are defined same as in the linear model [44]. Both binary and continuous responses are considered. It is proved that the simple t-test without using any covariates is conservative, and a valid test can be constructed using bootstrap. This bootstrap test asymptotically has same power as the t-test which correctly uses covariates in the final analysis. More discussion can be found in Tu, Shalay and Pater [50], Aickin [1], and so on.

1.3 Motivation and Organization

Over the past several decades, scientists have identified many new biomarkers /covariates that may be associated with certain diseases as the development of translational research (genomics, proteomics and metabolomics). Precision medicine is developed allowing physicians to tailor a treatment regimen based on an individual patient's characteristics of interest (which could be biomarkers or other covariates). To develop precision medicine, more covariates are under consideration in a clinical trial. New designs that balance treatment assignments with respect to key covariates of interest as well as minimize the number of subjects randomized to the inferior treatment for the sake of ethics are needed. At the same time, covariates can be categorized into two types according to their different roles played in clinical studies: (i) *prognostic covariates* that are balanced to provide a valid comparison, and (ii) *predictive covariates* that are used to select a suitable treatment based on efficiency and ethics. The increasing importance of the roles of these covariates can be exemplified by the following clinical studies.

The HER2 gene is identified to have a significant effect on increasing the aggressiveness of tumor in breast cancer. In the clinical trial of Slamon *et al.* [46], it was shown that the addition of trastuzumab, which is an antibody against HER2, to chemotherapy has a great impact on reducing the risk of death and extending the median

survival for patients with overexpressed HER2. Hence, HER2 is highly recommended for every invasive breast cancer to be tested as a predictive covariate, since the result can significantly impact the treatment selection. Age and Karnofsky score of patients are balanced between treatment groups to provide a valid comparison, which can be considered as prognostic covariates. In phase II of the "Basket" study of Hyman *et al.* [25], patients with BRAF V600 Mutant were found out to have an impressively good response to the Venmurafenib treatment. The presence of BRAF V600 Mutant can be used as an important predictive covariate allowing patients to be subgrouped for the Venmurafenib treatment with complementary treatments. In the "Basket" study, characteristics of patients (such as age and sex) are used for the balancing purpose to form comparable groups and ensure the credibility of the trial. So, they can be considered as prognostic covariates. Additional examples can be found in many other trials, such as Buyse *et al.* [8], Larkin *et al.* [30] and Krisam and Kieser [27].

In the literature, many covariate-adaptive designs have been proposed to balance prognostic covariates for a simple treatment comparison. While covariate-adjusted response-adaptive designs have been proposed to deal with predictive covariates for both efficiency and ethical reasons. Unfortunately, there is no design thus far that is available to deal with both prognostic covariates and predictive covariates simultaneously, since the incorporation of covariate information is complex due to the existence of two dissimilar types of covariates which play different roles in a clinical study. Therefore, the ultimate objectives of this work are to propose new designs which incorporate both types of covariates in the randomization procedure and study the hypothesis testings under the new designs. Ma, Hu and Zhang [32] had proved the theoretical properties of statistical inference based on a large class of covariate-adaptive designs. Before investigating problems of new designs, it is crucial to study hypothesis testings of comparing treatment effects under CARA designs, especially when covariates are fully or partially omitted in the final analysis, since not all these

covariates used in randomization can be included in the inference procedure due to complexity in practice.

In Chapter 2, a theoretical foundation of statistical inference under CARA designs is established with a linear model framework for clinical trials with two treatments and continuous responses. The hypothesis testing to compare treatment effects is examined under widely satisfied conditions of CARA designs. The asymptotic distributions of the test statistic under the null and alternative hypotheses are derived without considering covariates in the final analysis. We show that the hypothesis testing to compare treatment effects is not always conservative, but possible to be liberal, depending on the unknown parameters in the employed model and the choice of allocation function. Simulations with multiple scenarios of sample size and parameter settings are conducted to illustrate Type I error and power.

In Chapter 3, with a linear model, we study the statistical inference under the covariate-adaptive designs with a fixed target allocation proportion, which is an extension of covariate-adaptive designs proposed by Hu and Hu [24] and a special case of the new designs. Also, we consider testing the significance of covariates in this case. We show that the hypothesis testing to compare treatment effects is usually conservative in terms of small Type I error, which agrees with the conclusion of the special case using $1/2$ as the fixed target allocation proportion by Ma, Hu and Zhang [32]. The limiting power reaches its maximum when the allocation proportion is $1/2$, and it is larger than that under complete randomization when the difference between treatment effects is relatively large. The hypothesis testing for significance of covariates is still valid.

In Chapter 4, the innovative new CARA designs are proposed, which can incorporate both prognostic and predictive covariates in the randomization procedure simultaneously. Hypothesis testings of comparing treatment effects and significance of covari-

ates are studied under the new designs when prognostic covariates are excluded from the final analysis. The asymptotic distributions of test statistics under both the null and alternative hypotheses are given for the new designs with some mild conditions for imbalances and target allocation proportions. Simulation studies are conducted to explore the imbalances at all three levels, and to study Type I error and power under complete randomization, the covariate-adaptive design in Chapter 3, the CARA design in Chapter 2 and the new design.

In Chapter 5, the discrete assumption for prognostic covariates is relaxed and similar inference problems are studied under the covariate-adaptive design in Chapter 3 and the new CARA design in Chapter 4. Note that the new CARA design is applied using fixed target allocation proportions in the randomization procedure instead of adaptively estimated ones. Asymptotic distributions of test statistics are obtained under the same conditions used in Chapter 3. We show that the test for comparing treatment effects is conservative under both designs, when prognostic covariates are omitted for statistical inference regardless whether they are discrete or continuous.

Chapter 2

Statistical Inference under CARA Designs

2.1 Introduction

Statistical inference of CARA randomization designs is discussed under a linear model with two treatments and a univariate covariate in this chapter. For a general family of CARA designs, we derive the asymptotic distributions of test statistics for comparing treatment effects under both the null and alternative hypotheses. We find out that the hypothesis testing to compare treatment effects is not always valid when the key covariate used in randomization is omitted in the final analysis; it depends on the unknown parameters in the employed model and the choice of the allocation function. Proper adjustment of the variance of the test statistic, such as the bootstrap t-test, can achieve nominal Type I error. The full model is used in the inference procedure for power comparison between CARA designs and complete randomization in order to obtain an unbiased estimator of the difference between two treatment effects. CARA designs can significantly improve average responses, but still allowing high power.

This chapter is organized as follows. In Section 2.2, the general framework of CARA designs under a linear model is introduced to study hypothesis testing properties for comparing treatment effects. Both an underlying model and a working model are proposed to represent the situation that the univariate covariate used in randomiza-

tion is omitted in statistical inference. Theoretical results for the test statistic under both the null and alternative hypotheses are presented in Section 2.3. In this section, we show that the two-sample t-test comparing treatment effects is not always valid when the covariate is omitted in the final analysis. Further, simulation studies are carried out in Section 2.4 to study Type I error and power of CARA designs under multiple scenarios compared with complete randomization. Conclusions are given in Section 2.5, and salient points of theoretical results are shown in Section 2.6.

2.2 Framework

In this section, we study hypothesis testings based on a linear model framework for a general family of CARA designs. The following notations and definitions are introduced to help describe the randomized treatment allocation scheme of CARA designs.

Consider a clinical trial with two treatments studied under a CARA design. Let N be the total number of subjects enrolled in the trial. For the m-th subject, I_m represents the treatment assignment, i.e., $I_m = 1$ assigned to treatment 1 and $I_m = 0$ to treatment 2, where $m = 1, 2, \ldots, N$. And N_k is the number of subjects assigned to treatment k, $k = 1, 2$, where $N_1 = \sum_{m=1}^{N} I_m$ and $N = N_1 + N_2$. Let Y_m denote the response of the m-th subject. Assume that the covariate information of subjects can be observed in the trial. Then, let X_m be the univariate covariate of the m-th subject. Let $\mathcal{I}_m = \sigma(I_1, \ldots, I_m)$, $\mathcal{Y}_m = \sigma(Y_1, \ldots, Y_m)$ and $\mathcal{X}_m = \sigma(X_1, \ldots, X_m)$ be the corresponding sigma fields for treatment assignments, responses and covariates. In addition, let $\mathcal{F}_m = \mathcal{I}_m \otimes \mathcal{Y}_m \otimes \mathcal{X}_m$ be the sigma field of history information after m assignments.

Conditional on the treatment assignment I_m, the following linear model is assumed

for the response of the m-th subject Y_m,

$$Y_m = \mu_1 I_m + \mu_2(1 - I_m) + \gamma X_m + \epsilon_m, \tag{2.1}$$

where

- X_m's, $m = 1, \ldots, N$ can be discrete or continuous covariates, which are independent and identically distributed as X with $\mathbb{E}(X) = 0$ and $\mathsf{Var}(X) = \sigma_x^2$;
- X_m's are used in the randomization procedure, but not used in final statistical inference, $m = 1, \ldots, N$;
- ϵ_m's, $m = 1, \ldots, N$ are independent and identically distributed random errors with mean zero and variance σ_ϵ^2, and independent of the covariate X;
- $\boldsymbol{\theta} = (\mu_1, \mu_2, \gamma)$ are unknown parameters and the parameter space $\boldsymbol{\Theta}$ is a bounded domain in $\mathbb{R}^3$.

Notice that under CARA designs, we do not need to restrict the covariate X to be discrete. Unlike covariate-adaptive designs under which studies are usually conducted based on discrete covariates only, we can directly incorporate continuous covariates without discretization. Also, we assume the covariate effect γ to be same for the two treatments.

2.2.1 CARA Designs

A general CARA design is defined by

$$\psi_m = \mathbb{E}\left(I_m \mid \mathcal{F}_{m-1}, X_m\right) = \mathbb{E}\left(I_m \mid \mathcal{I}_{m-1}, \mathcal{Y}_{m-1}, \mathcal{X}_m\right),$$

the conditional probabilities of assigning the m-th subject to treatment 1, conditioning on the entire history of all previous $m-1$ subjects, including their treatment

assignments, responses and covariate information, as well as the covariate information of the current subject.

The allocation scheme of CARA designs is as follows:

1. Assign m_0 subjects randomly to each treatment by using some restricted randomization design.

2. Assume that $n-1$ assignments have been finished ($n > 2m_0$). Responses $\{Y_i,\ i = 1, \ldots, n-1\}$ and corresponding covariates $\{X_i,\ i = 1, \ldots, n-1\}$ are observed. Let $\hat{\boldsymbol{\theta}}_{n-1} = (\hat{\mu}_{1,n-1}, \hat{\mu}_{2,n-1}, \hat{\gamma}_{n-1})$ be the estimators of $\boldsymbol{\theta} = (\mu_1, \mu_2, \gamma)$ based on the observed $(n-1)$-size sample $\{(Y_i, I_i, X_i),\ i = 1, \ldots, n-1\}$.

3. When the n-th subject comes into the trial and the corresponding covariate X_n is observed, we assign the subject to treatment 1 with probability

$$\mathbb{P}\left(I_n = 1 | \mathcal{F}_{n-1}, X_n\right) = \pi\left(\hat{\boldsymbol{\theta}}_{n-1}, X_n\right), \tag{2.2}$$

   where $\mathcal{F}_{n-1} = \mathcal{I}_{n-1} \otimes \mathcal{Y}_{n-1} \otimes \mathcal{X}_{n-1}$ is the sigma field of the history, and $\pi(\cdot, \cdot)$ is some given allocation function. Therefore, the probability that the subject is assigned to treatment 2 is $1 - \pi\left(\hat{\boldsymbol{\theta}}_{n-1}, X_n\right)$.

4. Repeat step 2 and 3 until each subject is assigned to a treatment.

Different choices of allocation functions can generate different possible classes of useful designs. Under the linear model framework, the allocation function can be defined as

$$\pi\left(\boldsymbol{\theta}, X\right) = \frac{G\left(\mu_1 + \gamma X\right)}{G\left(\mu_1 + \gamma X\right) + G\left(\mu_2 + \gamma X\right)}, \tag{2.3}$$

where G is some positive monotone smooth real function defined in $\mathbb{R}$. Then, the allocation function π satisfies $0 < \pi < 1,\ \pi = 1/2$, when $\mu_1 = \mu_2$. Suppose a larger response is preferred, then G should be an increasing function and a subject will

have a larger probability to be randomized to the treatment with a larger treatment effect μ for all X.

The following example gives a more detailed illustration of the allocation scheme of CARA designs under the linear model framework.

EXAMPLE. Suppose the target sample size of a two-armed clinical trial is 100. We choose the standard normal cumulative density function Φ as the function G in (2.3) to construct the allocation function π for the CARA design. At the beginning, we assign 20 subjects to the two treatments using permuted block design with block size 4. After each assignment, the covariate information, treatment assignment and response of the subject are recorded. Further, we use least squares estimation method to update the estimators $\hat{\boldsymbol{\theta}}_{n-1} = (\hat{\mu}_{1,n-1}, \hat{\mu}_{2,n-1}, \hat{\gamma}_{n-1})$ by using data collected from the first $n-1$ subjects ($n > 20$). When the n-th subject enters the trial and the covariate X_n is recorded, we assign the subject to treatment 1 with probability

$$\pi\left(\hat{\boldsymbol{\theta}}_{n-1}, X_n\right) = \frac{\Phi(\hat{Y}_{1,n})}{\Phi(\hat{Y}_{1,n})+\Phi(\hat{Y}_{2,n})} = \frac{\Phi\left(\hat{\mu}_{1,n-1} + \hat{\gamma}_{n-1}X_n\right)}{\Phi\left(\hat{\mu}_{1,n-1} + \hat{\gamma}_{n-1}X_n\right) + \Phi\left(\hat{\mu}_{2,n-1} + \hat{\gamma}_{n-1}X_n\right)}.$$

The update of these estimators and the randomization procedure repeat until the completion of assigning all 100 subjects.

2.2.2 Hypothesis Testing

In this section, we study statistical inference under CARA designs when the randomization covariate is omitted in the final analysis. The working model of inference would be

$$\mathbb{E}(Y_m) = \mu_1 I_m + \mu_2(1 - I_m). \tag{2.4}$$

Our primary interest is to compare treatment effects μ_1 and μ_2, and the following

hypothesis testing is used

$$H_0 : \mu_1 - \mu_2 = 0 \text{ versus } H_A : \mu_1 - \mu_2 \neq 0. \tag{2.5}$$

The test statistic is

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{S^2_{Y_1}}{N_1} + \frac{S^2_{Y_2}}{N_2}}}, \tag{2.6}$$

where $\bar{Y}_k$ and $S^2_{Y_k}$ are sample mean and sample variance for responses in treatment k, $k = 1, 2$. The following statistics,

$$N_1 = \sum_{m=1}^{N} I_m,\ N_2 = \sum_{m=1}^{N}(1 - I_m),\ \bar{Y}_1 = \frac{\sum_{m=1}^{N} Y_m I_m}{N_1},\ \bar{Y}_2 = \frac{\sum_{m=1}^{N} Y_m(1 - I_m)}{N_2}$$
$$S^2_{Y_1} = \frac{\sum_{m=1}^{N} Y_m^2 I_m - N_1 \bar{Y}_1^2}{N_1 - 1},\ S^2_{Y_2} = \frac{\sum_{m=1}^{N} Y_m^2(1 - I_m) - N_2 \bar{Y}_2^2}{N_2 - 1},$$

are computed to construct the test statistic in (2.6).

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 it.

2.3 Theoretical Properties

The hypothesis testing of comparing two treatment effects is considered in this section. It is conducted based on the working model (2.4), when data are generated from the underlying true model (2.1). Asymptotic properties of the test statistic (2.6) are studied under both the null and alternative hypotheses in the following theorem.

THEOREM 2.3.1. Suppose that a CARA design has a linear model framework (2.1) and the allocation function satisfies the following two conditions:

(A) For any fixed X, the allocation function $\pi(\boldsymbol{\theta}, X)$ is a continuous function of $\boldsymbol{\theta}$;

(B) The allocation function $\pi(\boldsymbol{\theta}, X)$ is differentiable with respect to $\boldsymbol{\theta}$ under the expectation, and there is a $\delta > 0$ such that

$$g(\boldsymbol{\theta}^*) = g(\boldsymbol{\theta}) + (\boldsymbol{\theta}^* - \boldsymbol{\theta}) \left(\frac{dg}{d\boldsymbol{\theta}^*}\bigg|_{\boldsymbol{\theta}} \right)^T + o_p\left(||\boldsymbol{\theta}^* - \boldsymbol{\theta}||^{1+\delta} \right),$$

where $\dfrac{dg}{d\boldsymbol{\theta}^*} = \left(\dfrac{dg}{d\mu_1^*}, \dfrac{dg}{d\mu_2^*}, \dfrac{dg}{d\gamma^*} \right)$.

Then, we have

(i) under $H_0: \ \mu_1 - \mu_2 = 0$, the test statistic (2.6) has the following asymptotic distribution:

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\dfrac{S_{Y_1}^2}{N_1} + \dfrac{S_{Y_2}^2}{N_2}}} \xrightarrow{D} \mathcal{N}\left(0, \ \frac{\sigma_0^2}{\sigma_a^2}\right), \tag{2.7}$$

where

$$\sigma_0^2 = 4\left(\gamma^2\sigma_x^2 + \sigma_\epsilon^2\right) + 32\gamma^2 \left(\frac{d\phi}{d\boldsymbol{\theta}}\right) \boldsymbol{V}_0 \left(\frac{d\phi}{d\boldsymbol{\theta}}\right)^T + 16\gamma\sigma_\epsilon^2 \left(\frac{d\phi}{d\mu_1} - \frac{d\phi}{d\mu_2}\right),$$
$$\sigma_a^2 = 4\left(\gamma^2\sigma_x^2 + \sigma_\epsilon^2\right), \ \phi(\cdot) = \mathbb{E}\left[X \cdot \pi(\cdot, X)\right], \ \boldsymbol{V}_0 = \sigma_\epsilon^2 \cdot \mathsf{diag}\left(2, \ 2, \ \frac{1}{\sigma_x^2}\right).$$

The hypothesis testing (2.5) is not always valid if omitting the covariate X in statistical inference. Type I error of the test depends on unknown parameters in the employed model and the choice of the allocation function π. When

$$\frac{d\phi}{d\mu_1} \in \left[-\frac{1}{4\gamma} \vee 0, -\frac{1}{4\gamma} \wedge 0 \right],$$

the test is asymptotically conservative, i.e., Type I error is smaller than nominal significance level α; Otherwise, the test is liberal, i.e., Type I error is larger than α.

(ii) under $H_a: \ \mu_1 - \mu_2 \neq 0$, consider a sequence of local alternatives, that is,

$\mu_1 - \mu_2 = \delta/\sqrt{N}$ for some fixed $\delta \neq 0$, then the test statistic (2.6) has the following asymptotic distribution:

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{S^2_{Y_1}}{N_1} + \frac{S^2_{Y_2}}{N_2}}} \xrightarrow{D} \mathcal{N}\left(\frac{1}{\sigma_b} \cdot \left[\delta + \frac{\sqrt{N}\gamma\phi(\boldsymbol{\theta})}{g(\boldsymbol{\theta})[1-g(\boldsymbol{\theta})]}\right], \frac{\sigma_1^2}{\sigma_b^2}\right), \tag{2.8}$$

where $\eta(\cdot) = \mathbb{E}[X^2 \cdot \pi(\cdot, X)]$, $\boldsymbol{V} = \sigma_\epsilon^2 \begin{pmatrix} \frac{1}{g(\boldsymbol{\theta})} & 0 & \frac{\phi(\boldsymbol{\theta})}{g(\boldsymbol{\theta})\sigma_x^2} \\ & \frac{1}{1-g(\boldsymbol{\theta})} & -\frac{\phi(\boldsymbol{\theta})}{(1-g(\boldsymbol{\theta}))\sigma_x^2} \\ & & \frac{1}{\sigma_x^2} \end{pmatrix}$,

$$\begin{aligned}\sigma_1^2 =& \frac{\gamma^2\sigma_x^2}{[1-g(\boldsymbol{\theta})]^2} + \frac{\sigma_\epsilon^2}{g(\boldsymbol{\theta})[1-g(\boldsymbol{\theta})]} + \gamma^2\eta(\boldsymbol{\theta}) \cdot \left[\frac{1}{g^2(\boldsymbol{\theta})} - \frac{1}{[1-g(\boldsymbol{\theta})]^2}\right] \\ &+ \frac{2\gamma\sigma_\epsilon^2}{g^2(\boldsymbol{\theta})[1-g(\boldsymbol{\theta})]^2} \cdot \left[\frac{d\phi}{d\mu_1} \cdot (1-g(\boldsymbol{\theta})) - \frac{d\phi}{d\mu_2} \cdot g(\boldsymbol{\theta}) + \frac{d\phi}{d\gamma} \cdot \frac{\phi(\boldsymbol{\theta})}{\sigma_x^2}\right] \\ &+ \frac{\gamma^2}{g^2(\boldsymbol{\theta})[1-g(\boldsymbol{\theta})]^2} \cdot \left[2\left(\frac{d\phi}{d\boldsymbol{\theta}}\right)\boldsymbol{V}\left(\frac{d\phi}{d\boldsymbol{\theta}}\right)^T - \phi^2(\boldsymbol{\theta})\right],\end{aligned}$$

$$\begin{aligned}\sigma_b^2 =& \frac{\gamma^2\sigma_x^2}{[1-g(\boldsymbol{\theta})]^2} + \frac{\sigma_\epsilon^2}{g(\boldsymbol{\theta})[1-g(\boldsymbol{\theta})]} + \gamma^2\eta(\boldsymbol{\theta}) \cdot \left[\frac{1}{g^2(\boldsymbol{\theta})} - \frac{1}{[1-g(\boldsymbol{\theta})]^2}\right] \\ &- \gamma^2\phi^2(\boldsymbol{\theta})\left[\frac{1}{g^3(\boldsymbol{\theta})} - \frac{1}{[1-g(\boldsymbol{\theta})]^3}\right].\end{aligned}$$

Theoretical properties of hypothesis testing to compare two treatment effects are obtained under CARA designs in Theorem 2.3.1. Covariate information is utilized in the randomization procedure to help assign subjects to a better treatment based on their own characteristics. Two widely satisfied conditions of the allocation function are assumed to derive the asymptotic distributions of test statistic (2.6). Under these conditions, we show in the proof that the asymptotic variance of test statistic can be smaller or larger than 1, when the covariate is omitted in the working model (2.4),

i.e., the numerator of the test statistic, $\bar{Y}_1 - \bar{Y}_2$, has variance smaller or larger than the model-based variance estimator in the denominator. This result comes from the difference between the variability increase introduced by estimating allocation probabilities and the variability reduction due to the incorporation of covariate information in randomization, depending on the unknown parameters $\boldsymbol{\theta} = (\mu_1, \mu_2, \gamma)$ in model settings and the choice of the allocation function π.

2.4 Simulation Study

Case 1: Testing treatment effects. We first conduct simulations to study Type I error of the hypothesis testing for comparing two treatment effects under the CARA design and complete randomization. The linear model in (2.1) is assumed. The covariate X_m and random error ϵ_m are independently distributed as $\mathcal{N}(0, 1)$. To study Type I error, we assume there is no difference between two treatment effects, i.e., we have $\mu_1 = \mu_2$.

In the simulations, the standard normal distribution CDF Φ is selected to construct the allocation function π. The following three scenarios of parameters are used to study Type I error: (i) $(\mu_1, \mu_2, \gamma) = (0, 0, 1)$, (ii) $(\mu_1, \mu_2, \gamma) = (3, 3, 1)$, and (iii) $(\mu_1, \mu_2, \gamma) = (0, 0, 2)$. The hypothesis tests include the two-sample t-test (t-test) and the test for comparing treatment effects under the full model (2.1) ($lm(X)$). The significance level is $\alpha = 0.05$, and sample size N is 100, 200 or 500. For the CARA design, $2m_0 = 20$ subjects are assigned to the two treatments using permuted block design with block size 4 at the beginning of allocation. Simulation results demonstrated in Table 2.1 are based on 10,000 runs.

Several conclusions can be drawn from Table 2.1. First, Type I error of test using the full model (2.1) ($lm(X)$) is close to 5%. Therefore, the hypothesis testings of comparing treatment effects are valid when the randomization covariate is included

Table 2.1: Simulated Type I error under the CARA design (CARA) and complete randomization (CR) in % using a linear model (2.1). Parameters are (μ_1, μ_2, γ) with three scenarios. Simulations are based on 10,000 runs.

Method	N	$(0,0,1)$		$(3,3,1)$		$(0,0,2)$	
		t-test	$lm(X)$	t-test	$lm(X)$	t-test	$lm(X)$
CARA	100	2.27	5.20	4.73	5.36	6.48	5.38
	200	2.22	5.09	4.37	4.93	9.94	5.44
	500	2.02	5.28	4.36	5.01	13.54	5.16
CR	100	5.50	5.61	5.09	5.32	4.83	4.93
	200	4.85	5.06	4.94	5.00	4.95	5.15
	500	5.02	4.97	5.10	5.24	4.93	5.14

in the working model of the final analysis. Second, Type I error of the two-sample t-test (t-test) under the CARA design depends on the unknown parameters $\boldsymbol{\theta}$. This is consistent with the theoretical results in Section 2.3. We can calculate $d\phi/d\mu_1$ for these three scenarios, which are -0.1486, -0.0139, and -0.2751. Hence, we can see that $d\phi/d\mu_1$ of the first two scenarios are within the range $[-0.25, 0]$, while that of the last scenario is out of the range $[-0.125, 0]$. Therefore, the test statistics under the first two scenarios have variance smaller than 1, resulting in Type I error smaller than 5%, while that under the last scenario has variance greater than 1, hence the test is liberal. As the effect of covariate on response weakens, compared with treatment effects, Type I error would be closer to the nominal level α. In the last scenario, Type I error increases as the sample size gets larger, approaching the theoretical Type I error 17.31%. Further, under complete randomization (CR), Type I error is always close to 5% for all scenarios of parameters and sample size. We also try different allocation functions and combinations of parameters, similar results can be obtained and are not shown here.

Since the two-sample t-test is not always valid under CARA designs when the co-

variate is omitted in the inference analysis, proper adjustment on variance of the test statistic can be conducted to achieve nominal Type I error. One possible approach is the bootstrap t-test. In each simulation, original covariates $X_1, \ldots, X_N$, are collected after CARA randomization. Then, B bootstrap samples $X_1^b, \ldots, X_N^b,\ b = 1, \ldots, B$ are generated independently as random samples with replacement. The CARA allocation scheme described in Section 2.2.1, estimated parameters and residuals from the original N samples are applied to each bootstrap sample to obtain treatment assignments $I_1^b, \ldots, I_N^b$ and responses $Y_1^b, \ldots, Y_N^b$. Define

$$\bar{Y}_1 - \bar{Y}_2 = \frac{1}{N_1}\sum_{m=1}^{N} Y_m I_m - \frac{1}{N_2}\sum_{m=1}^{N} Y_m(1 - I_m),\ N_1 = \sum_{m=1}^{N} I_m,\ N_2 = \sum_{m=1}^{N}(1 - I_m),$$

and

$$\hat{D}(b) = \frac{1}{N_1^b}\sum_{m=1}^{N} Y_m^b I_m^b - \frac{1}{N_2^b}\sum_{m=1}^{N} Y_m^b(1 - I_m^b),\ N_1^b = \sum_{m=1}^{N} I_m^b,\ N_2^b = \sum_{m=1}^{N}(1 - I_m^b).$$

The bootstrap estimator of variance for $\bar{Y}_1 - \bar{Y}_2$ is the sample variance of $\hat{D}(b)$, $b = 1, \ldots, B$, denoted as ν_B. Then, the test statistic of the bootstrap t-test is

$$T_B = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\hat{\nu}_B}}.$$

The number of bootstrap samples used in the following simulations is $B = 500$, and results are shown in Table 2.2. We can see that the bootstrap t-test (BS-t) can still achieve nominal Type I error under CARA designs.

Case 2: Power comparison. Next, we compare power under CARA designs (CARA) and complete randomization (CR). Under the alternative hypothesis, there is a difference between treatment effects μ_1 and μ_2, i.e., we have $\mu_1 - \mu_2 \neq 0$. From the proof, the equation (2.22) of Theorem 2.3.1, we can see that $\bar{Y}_1 - \bar{Y}_2$ is no longer an unbiased estimator of $\mu_1 - \mu_2$ under the alternative hypothesis, when the covariate X is not

Table 2.2: Simulated Type I error of the bootstrap t-test (BS-t) under CARA designs in % with a linear model (2.1). Parameters are (μ_1, μ_2, γ) with three scenarios. Simulations are based on 1,000 runs.

Test	N	$(0,0,1)$	$(3,3,1)$	$(0,0,2)$
	100	4.7	5.1	5.5
BS-t	200	5.1	5.4	5.4
	500	4.9	5.2	5.3

included in final inference. It is important to use the full model ($lm(X)$) in Case 1 under CARA design to conduct power comparison. Simulations are studied with sample size $N = 100$. All results of power are given in Figure 2.1, from which several conclusions can be drawn. The hypothesis testing of comparing treatment effects using the full model (2.1) under the CARA design (CARA, $lm(X)$) has very similar power performance to that under complete randomization (CR, $lm(X)$). However, the CARA design (CARA) obtains better average responses in the right plot, and the improvement becomes more obvious, when the difference between two treatment effects $\mu_1 - \mu_2$ gets larger. Based on Figure 2.1, we can see that CARA designs have advantages of assigning more subjects to a more suitable treatment based on their characteristics and achieving better outcomes.

Further, we explore the power for small sample size 32 and 64. The simulated power and average responses are demonstrated in Table 2.3 using the same model and parameters as in Figure 2.1. And the results show the advantages of CARA designs (CARA) over complete randomization (CR) on both power and average responses for small sample size.

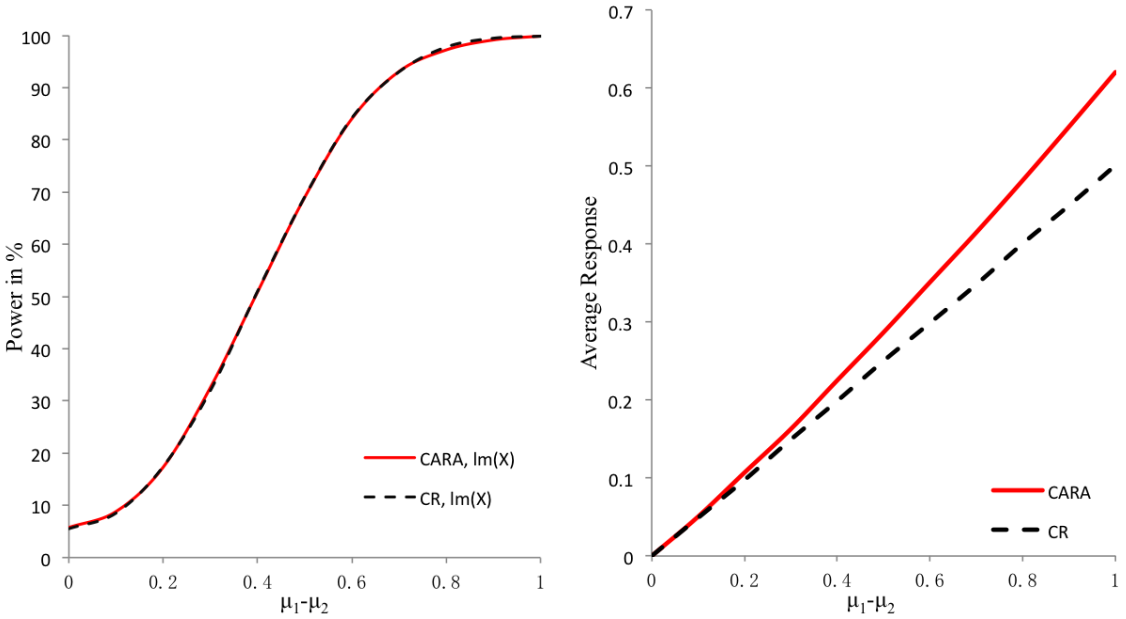

Figure 2.1: Simulated power (left) and average responses (right) for the CARA design (CARA) and complete randomization (CR) with $\mu_2 = 0$ and $\gamma = 1$. Simulations are based on 10,000 runs.

Table 2.3: Simulated power (average response) of full model $lm(X)$ for CARA design and complete randomization (CR) in %. Simulations are based on 10,000 runs.

	$N = 32$		$N = 64$	
$\mu_1 - \mu_2$	CR	CARA	CR	CARA
0.0	5.10 (0.002)	5.75 (0.002)	5.12 (-0.002)	5.20 (0.001)
0.1	5.89 (0.048)	6.70 (0.054)	6.26 (0.050)	6.92 (0.050)
0.2	7.72 (0.010)	9.49 (0.109)	11.31 (0.099)	12.73 (0.104)
0.3	12.81 (0.149)	13.14 (0.164)	20.76 (0.147)	21.79 (0.162)
0.4	18.05 (0.199)	19.50 (0.223)	34.03 (0.204)	35.68 (0.222)
0.5	25.90 (0.247)	26.92 (0.282)	49.24 (0.250)	49.23 (0.282)
0.6	36.21 (0.302)	36.92 (0.343)	63.48 (0.298)	64.86 (0.341)
0.7	44.76 (0.352)	46.33 (0.401)	76.87 (0.348)	78.15 (0.407)
0.8	56.03 (0.399)	57.37 (0.467)	87.52 (0.401)	86.88 (0.472)
0.9	67.63 (0.447)	67.54 (0.535)	93.48 (0.452)	93.74 (0.534)
1.0	75.35 (0.503)	75.69 (0.601)	97.23 (0.501)	97.37 (0.601)

2.5 Conclusion

CARA designs incorporate covariates, responses and treatment assignments into the randomization scheme, which have influence on statistical inference results of conventional hypothesis testings. In this chapter, we show that, based on a linear model, the hypothesis testing for comparing two treatment effects is not always valid, when the covariate used in the randomization procedure is omitted in the inference analysis under a family of CARA designs. The performance of the two-sample t-test depends on the parameters in the employed model, as well as the choice of the allocation function for treatment assignments. Proper adjustment of the variance of the test statistic is necessary to achieve nominal Type I error. CARA designs using the full model in the final analysis can obtain better average responses compared with complete randomization, but still allowing statistical inference with high power. The results in this chapter provide a theoretical foundation for statistical inference under CARA designs and a better understanding on the proper implementation of CARA designs to achieve valid statistical conclusions in practice.

Theoretical results in this chapter show the influence of the incorporation of predictive covariates in the randomization procedure under CARA designs on conventional hypothesis testings. However, prognostic covariates, which can be used for balancing across treatment arms to provide a valid comparison, are not considered in the underlying model and the randomization scheme. In Chapter 3, we study similar problems with only prognostic covariates incorporated in randomization. Further, new designs utilizing both types of covariates and their statistical inference will be covered in Chapter 4.

2.6 Appendix: Proof of Theorems

Proof of Theorem 2.3.1.

It follows from (2.1) that $\sum_{m=1}^{N} Y_m I_m = N_1\mu_1 + \sum_{m=1}^{N} (\gamma X_m + \epsilon_m) I_m$. Note that $X_m I_m = X_m I_m - \mathbb{E}(X_m I_m|\mathcal{F}_{m-1}) + \phi\left(\hat{\boldsymbol{\theta}}_{m-1}\right)$, where $\phi(\boldsymbol{\theta}) = \mathbb{E}[X \cdot \pi(\boldsymbol{\theta}, X)]$, then we have

$$\sum_{m=1}^{N} (\gamma X_m + \epsilon_m) I_m = \gamma \sum_{m=1}^{N} [X_m I_m - \mathbb{E}(X_m I_m|\mathcal{F}_{m-1})] + \sum_{m=1}^{N} \epsilon_m I_m + \gamma \sum_{m=1}^{N} \phi\left(\hat{\boldsymbol{\theta}}_{m-1}\right). \tag{2.9}$$

It is straightforward that the first two terms of (2.9) are martingales, and the last term can be approximated by two other martingales. Write $\Delta U_m = \epsilon_m I_m$ and $\Delta M_m = \gamma [X_m I_m - \mathbb{E}(X_m I_m|\mathcal{F}_{m-1})]$. Then, let $M_N = \sum_{m=1}^{N} \Delta M_m$ and $U_N = \sum_{m=1}^{N} \Delta U_m$. Under the two conditions of the allocation function, we employ Taylor expansion to get

$$\phi\left(\hat{\boldsymbol{\theta}}_m\right) = \phi(\boldsymbol{\theta}) + \left(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}\right)\left(\frac{d\phi}{d\boldsymbol{\theta}}\right)^T + o_p\left(||\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}||^{1+\delta}\right), \tag{2.10}$$

for some $\delta > 0$. Suppose we have

$$\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta} = \frac{1}{N}\sum_{m=1}^{N} \boldsymbol{h}(Y_m, X_m)(1 + o(1)) + o\left(N^{-1/2}\right), \quad \text{a.s.}, \tag{2.11}$$

where $\boldsymbol{h}$ is some function with $\mathbb{E}[\boldsymbol{h}(Y, X)|X] = \mathbf{0}$, and $\mathbb{E}||\boldsymbol{h}(Y, X)||^2 < \infty$ depending on estimation methods. Under the linear model framework, we use the least squares method to estimate parameters and define

$$\boldsymbol{h}(Y, X) = \left(\frac{\epsilon \cdot I}{g(\boldsymbol{\theta})}, \frac{\epsilon \cdot (1 - I)}{1 - g(\boldsymbol{\theta})}, \frac{\epsilon \cdot X}{\sigma_x^2}\right). \tag{2.12}$$

By combining (2.10), (2.11) and (2.12), we can get

$$\phi\left(\hat{\boldsymbol{\theta}}_m\right) = \phi(\boldsymbol{\theta}) + \frac{\boldsymbol{T}_m(1+o(1))}{m}\left(\frac{d\phi}{d\boldsymbol{\theta}}\right)^T + o\left(m^{-1/2}\right), \quad \text{a.s.},$$

where

$$\Delta\boldsymbol{T}_i = \boldsymbol{h}\left(Y_i, X_i\right) \text{ and } \boldsymbol{T}_m = \sum_{i=1}^{m}\Delta\boldsymbol{T}_i = \left(\sum_{i=1}^{m}\frac{\epsilon_i\cdot I_i}{g(\boldsymbol{\theta})}, \sum_{i=1}^{m}\frac{\epsilon_i\cdot(1-I_i)}{1-g(\boldsymbol{\theta})}, \sum_{i=1}^{m}\frac{\epsilon_i\cdot X_i}{\sigma_x^2}\right).$$

Therefore, the last term of (2.9) can be approximated by

$$\begin{aligned}\gamma\sum_{m=1}^{N}\phi\left(\hat{\boldsymbol{\theta}}_{m-1}\right) &= \gamma N\phi(\boldsymbol{\theta}) + \sum_{m=1}^{N}\frac{\boldsymbol{T}_{m-1}(1+o(1))}{m-1}\left(\frac{d\phi}{d\boldsymbol{\theta}}\right)^T + o\left(\sqrt{N}\right) \quad \text{a.s.}\\ &= \gamma N\phi(\boldsymbol{\theta}) + \gamma\sum_{m=1}^{N}\frac{\boldsymbol{T}_{m-1}}{m-1}\left(\frac{d\phi}{d\boldsymbol{\theta}}\right)^T + o_p\left(\sqrt{N}\right).\end{aligned}$$

Write $\Delta D_m = \gamma\dfrac{\boldsymbol{T}_{m-1}}{m-1}\left(\dfrac{d\phi}{d\boldsymbol{\theta}}\right)^T$ and $D_N = \sum_{m=1}^{N}\Delta D_m$, we have

$$\sum_{m=1}^{N}Y_mI_m = N_1\mu_1 + \gamma N\phi(\boldsymbol{\theta}) + M_N + U_N + D_N + o_p\left(\sqrt{N}\right). \tag{2.13}$$

In the same way, we get

$$\sum_{m=1}^{N}Y_m(1-I_m) = N_2\mu_2 + \sum_{m=1}^{N}(\gamma X_m + \epsilon_m) - \left[\gamma N\phi(\boldsymbol{\theta}) + M_N + U_N + D_N\right] + o_p\left(\sqrt{N}\right). \tag{2.14}$$

And by letting $S_N = \sum_{m=1}^{N}(\gamma X_m + \epsilon_m)$, we have

$$\bar{Y}_1 - \bar{Y}_2 = \mu_1 - \mu_2 + \left(\frac{1}{N_1} + \frac{1}{N_2}\right)\left[\gamma N\phi(\boldsymbol{\theta}) + M_N + U_N + D_N\right] - \frac{S_N}{N_2} + o_p\left(\frac{1}{\sqrt{N}}\right). \tag{2.15}$$

Applying the law of the iterated logarithm for martingales to (2.15) together with

$\phi(\boldsymbol{\theta}) = \mathbb{E}\left[X \cdot \pi(X, \boldsymbol{\theta})\right] = \mathbb{E}(X)/2 = 0$ under H_0, we have

$$\bar{Y}_1 - \bar{Y}_2 = O\left(\sqrt{\frac{\log\log N}{N}}\right), \quad a.s.$$

Next, we consider the asymptotic normality for martingales. Note that $M_N, U_N, \boldsymbol{T}_N$ and S_N are all sums of martingale differences. With the asymptotic properties of allocation proportions, it is easy to verify that the Lindeberg's condition is satisfied. To complete the proof, it suffices to derive the variance. Define $\eta(\boldsymbol{\theta}) = \mathbb{E}[X^2 \cdot \pi(\boldsymbol{\theta}, X)]$. The conditional variances of martingale differences $\{\Delta M_n, \Delta S_n, \Delta \boldsymbol{T}_n, \Delta U_n\}$ satisfy

$$\begin{aligned}
\mathbb{E}\left[(\Delta M_n)^2 | \mathcal{F}_{n-1}\right] &= \gamma^2 \mathbb{E}\left[X_n^2 I_n - 2X_n I_n \mathbb{E}(X_n I_n | \mathcal{F}_{n-1}) + \mathbb{E}^2(X_n I_n | \mathcal{F}_{n-1}) | \mathcal{F}_{n-1}\right] \\
&= \gamma^2 \left[\eta(\hat{\boldsymbol{\theta}}) - \phi^2(\hat{\boldsymbol{\theta}})\right] \;\to\; \gamma^2 \left[\eta(\boldsymbol{\theta}) - \phi^2(\boldsymbol{\theta})\right] \qquad \text{in } L_1, \\
\mathbb{E}\left[(\Delta U_n)^2 | \mathcal{F}_{n-1}\right] &= \mathbb{E}\left(\epsilon_n^2 I_n | \mathcal{F}_{n-1}\right) = \sigma_\epsilon^2 g\left(\hat{\boldsymbol{\theta}}\right) \;\to\; \sigma_\epsilon^2 g\left(\boldsymbol{\theta}\right) \qquad \text{in } L_1, \\
\mathbb{E}\left[(\Delta S_n)^2 | \mathcal{F}_{n-1}\right] &= \mathbb{E}\left[\left(\gamma X_n + \epsilon_n\right)^2 | \mathcal{F}_{n-1}\right] = \gamma^2 \sigma_x^2 + \sigma_\epsilon^2,
\end{aligned}$$

$$\mathbb{E}\left[\Delta \boldsymbol{T}_n^T \Delta \boldsymbol{T}_n | \mathcal{F}_{n-1}\right] = \begin{pmatrix}
\frac{\mathbb{E}\left(\epsilon_n^2 I_n | \mathcal{F}_{n-1}\right)}{g^2(\boldsymbol{\theta})} & \frac{\mathbb{E}\left(\epsilon_n^2 I_n (1 - I_n) | \mathcal{F}_{n-1}\right)}{g(\boldsymbol{\theta})[1 - g(\boldsymbol{\theta})]} & \frac{\mathbb{E}\left(\epsilon_n^2 X_n I_n | \mathcal{F}_{n-1}\right)}{g(\boldsymbol{\theta})\sigma_x^2} \\
 & \frac{\mathbb{E}\left(\epsilon_n^2 (1 - I_n) | \mathcal{F}_{n-1}\right)}{[1 - g(\boldsymbol{\theta})]^2} & \frac{\mathbb{E}\left(\epsilon_n^2 X_n (1 - I_n) | \mathcal{F}_{n-1}\right)}{[1 - g(\boldsymbol{\theta})]\sigma_x^2} \\
 & & \frac{\mathbb{E}\left(\epsilon_n^2 X_n^2 | \mathcal{F}_{n-1}\right)}{\sigma_x^4}
\end{pmatrix}$$

$$\rightarrow \ \sigma_\epsilon^2 \begin{pmatrix} \frac{1}{g(\boldsymbol{\theta})} & 0 & \frac{\phi(\boldsymbol{\theta})}{g(\boldsymbol{\theta})\sigma_x^2} \\ & \frac{1}{1-g(\boldsymbol{\theta})} & -\frac{\phi(\boldsymbol{\theta})}{(1-g(\boldsymbol{\theta}))\sigma_x^2} \\ & & \frac{1}{\sigma_x^2} \end{pmatrix} = \boldsymbol{V} \quad \text{in } L_1.$$

The conditional covariances of martingale differences are calculated as follows:

$$\begin{aligned}
\mathbb{E}\left(\Delta M_n \Delta S_n | \mathcal{F}_{n-1}\right) &= \gamma^2 \mathbb{E}\left[X_n^2 I_n - X_n \mathbb{E}\left(X_n I_n | \mathcal{F}_{n-1}\right) | \mathcal{F}_{n-1}\right] \\
&= \gamma^2 \eta\left(\hat{\boldsymbol{\theta}}\right) \ \rightarrow \ \gamma^2 \eta(\boldsymbol{\theta}) \quad \text{in } L_1, \\
\mathbb{E}\left(\Delta U_n \Delta \boldsymbol{T}_n | \mathcal{F}_{n-1}\right) &= \left(\frac{\mathbb{E}\left(\epsilon_n^2 I_n | \mathcal{F}_{n-1}\right)}{g(\boldsymbol{\theta})}, \ 0, \ \frac{\mathbb{E}\left(\epsilon_n^2 X_n I_n | \mathcal{F}_{n-1}\right)}{\sigma_x^2}\right) \\
&= \sigma_\epsilon^2 \left(\frac{1}{g(\boldsymbol{\theta})} \cdot g\left(\hat{\boldsymbol{\theta}}\right), \ 0, \ \frac{1}{\sigma_x^2} \cdot \phi\left(\hat{\boldsymbol{\theta}}\right)\right) \rightarrow \ \sigma_\epsilon^2 \left(1, \ 0, \ \frac{\phi(\boldsymbol{\theta})}{\sigma_x^2}\right) \ \text{ in } L_1, \\
\mathbb{E}\left(\Delta U_n \Delta S_n | \mathcal{F}_{n-1}\right) &= \mathbb{E}\left(\epsilon_n^2 I_n | \mathcal{F}_{n-1}\right) = \sigma_\epsilon^2 g\left(\hat{\boldsymbol{\theta}}\right) \ \rightarrow \ \sigma_\epsilon^2 g(\boldsymbol{\theta}) \quad \text{in } L_1, \\
\mathbb{E}\left(\Delta \boldsymbol{T}_n \Delta S_n | \mathcal{F}_{n-1}\right) &= \left(\frac{\mathbb{E}\left(\epsilon_n^2 I_n | \mathcal{F}_{n-1}\right)}{g(\boldsymbol{\theta})}, \ \frac{\mathbb{E}\left(\epsilon_n^2 (1-I_n) | \mathcal{F}_{n-1}\right)}{1-g(\boldsymbol{\theta})}, \ 0\right) \ \rightarrow \ \sigma_\epsilon^2 (1,1,0) \ \text{ in } L_1.
\end{aligned}$$

Since X and ϵ are independent, $\mathbb{E}(\Delta M_n \Delta U_n | \mathcal{F}_{n-1}) = \mathbb{E}(\Delta M_n \Delta \boldsymbol{T}_n | \mathcal{F}_{n-1}) = 0$. It follows that

$$\begin{aligned}
\sigma_d^2 =& \mathsf{Var}\left[\sqrt{N}(\bar{Y}_1 - \bar{Y}_2)\right] \\
=& N\left(\frac{1}{N_1} + \frac{1}{N_2}\right)^2 \cdot \left[\mathsf{Var}\left(M_N\right) + \mathsf{Var}\left(U_N\right) + \mathsf{Var}\left(D_N\right) + 2\mathsf{Cov}\left(U_N, \ D_N\right)\right] \\
& - \frac{2N}{N_2}\left(\frac{1}{N_1} + \frac{1}{N_2}\right) \cdot \left[\mathsf{Cov}\left(M_N, \ S_N\right) + \mathsf{Cov}\left(U_N, \ S_N\right) + \mathsf{Cov}\left(D_N, \ S_N\right)\right] \\
& + \frac{N}{N_2^2} \cdot \mathsf{Var}\left(S_N\right)
\end{aligned}$$

$$
\begin{aligned}
=&N\left(\frac{1}{N_1}+\frac{1}{N_2}\right)^2\cdot\Big[N\gamma^2\left(\eta(\boldsymbol{\theta})-\phi^2(\boldsymbol{\theta})\right)+N\sigma_\epsilon^2 g(\boldsymbol{\theta})+2N\gamma^2\left(\frac{d\phi}{d\boldsymbol{\theta}}\right)\boldsymbol{V}\left(\frac{d\phi}{d\boldsymbol{\theta}}\right)^T\\
&+2N\gamma\sigma_\epsilon^2\cdot\left(\frac{d\phi}{d\mu_1}+\frac{d\phi}{d\gamma}\cdot\frac{\phi(\boldsymbol{\theta})}{\sigma_x^2}\right)+o(N)\Big]+\frac{N^2}{N_2^2}\cdot\left(\gamma^2\sigma_x^2+\sigma_\epsilon^2\right)\\
&-\frac{2N}{N_2}\left(\frac{1}{N_1}+\frac{1}{N_2}\right)\cdot\left[N\gamma^2\eta(\boldsymbol{\theta})+N\sigma_\epsilon^2 g(\boldsymbol{\theta})+N\gamma\sigma_\epsilon^2\left(\frac{d\phi}{d\mu_1}+\frac{d\phi}{d\mu_2}\right)+o(N)\right].
\end{aligned}
\tag{2.16}
$$

Hence, under $H_0:\mu_1-\mu_2=0$, we have

$$
\sigma_d^2\ \to\ 4\left(\gamma^2\sigma_x^2+\sigma_\epsilon^2\right)+32\gamma^2\left(\frac{d\phi}{d\boldsymbol{\theta}}\right)\boldsymbol{V}_0\left(\frac{d\phi}{d\boldsymbol{\theta}}\right)^T+16\gamma\sigma_\epsilon^2\left(\frac{d\phi}{d\mu_1}-\frac{d\phi}{d\mu_2}\right)=\sigma_0^2,\quad \text{a.s.}
\tag{2.17}
$$

By the central limit theorem for martingales, under $H_0:\mu_1=\mu_2$, it follows that

$$
\sqrt{N}\left(\bar{Y}_1-\bar{Y}_2\right)\ \xrightarrow{D}\ \mathcal{N}\left(0,\ \sigma_0^2\right).
\tag{2.18}
$$

Next, we look at the denominator in the test statistic (2.6). The sample variance of responses in treatment 1 can be written as

$$
S_{Y_1}^2=\frac{1}{N_1-1}\sum_{m=1}^{N}Y_m^2I_m-\frac{N_1}{N_1-1}\bar{Y}_1^2.
$$

It follows from (2.13) that $\bar{Y}_1\ \xrightarrow{P}\ \mu_1$, hence we have $\bar{Y}_1^2\ \xrightarrow{P}\ \mu_1^2$ under H_0. Also, we have

$$
\begin{aligned}
\sum_{m=1}^{N}Y_m^2I_m&=\sum_{m=1}^{N}\left(\mu_1+\gamma X_m+\epsilon_m\right)^2I_m\\
&=N_1\mu_1^2+2\mu_1\sum_{m=1}^{N}\left(\gamma X_m+\epsilon_m\right)I_m+\sum_{m=1}^{N}\left(\gamma X_m+\epsilon_m\right)^2I_m.
\end{aligned}
\tag{2.19}
$$

It has been verified that $\sum_{m=1}^{N}\left(\gamma X_m+\epsilon_m\right)I_m=N\gamma\phi(\boldsymbol{\theta})+O_p\left(\sqrt{N\log\log N}\right)$ based on (2.13) and the law of the iterated logarithm for martingales. Now, we only look

at the last term of (2.19), then we have

$$\sum_{m=1}^{N} \left(\gamma X_m + \epsilon_m\right)^2 I_m = \gamma^2 \sum_{m=1}^{N} X_m^2 I_m + \sum_{m=1}^{N} \epsilon_m^2 + 2\gamma \sum_{m=1}^{N} X_m \epsilon_m I_m. \tag{2.20}$$

It is very straightforward that the last term of (2.20) is a martingale. In the following proof, we will show that the first two terms can be both approximated by martingales and some constants. Notice that $X_m^2 I_m = X_m^2 I_m - \mathbb{E}\left(X_m^2 I_m | \mathcal{F}_{m-1}\right) + \eta\left(\hat{\boldsymbol{\theta}}_{m-1}\right)$, together with Taylor expansion, we have

$$\begin{aligned}
\sum_{m=1}^{N} X_m^2 I_m &= \sum_{m=1}^{N} \left[X_m^2 I_m - \mathbb{E}(X_m^2 I_m | \mathcal{F}_{m-1})\right] + \sum_{m=1}^{N} \eta\left(\hat{\boldsymbol{\theta}}_{m-1}\right) \\
&= \sum_{m=1}^{N} \left[X_m^2 I_m - \mathbb{E}(X_m^2 I_m | \mathcal{F}_{m-1})\right] + \sum_{m=1}^{N} \frac{\boldsymbol{T}_{m-1}}{m-1} \left(\frac{d\eta}{d\boldsymbol{\theta}}\right)^T \\
&\quad + N\eta(\boldsymbol{\theta}) + o_p\left(\sqrt{N}\right) \\
\text{and } \sum_{m=1}^{N} \epsilon_m^2 I_m &= \sum_{m=1}^{N} \epsilon_m^2 \left[I_m - \mathbb{E}(I_m | \mathcal{F}_{m-1})\right] + \sum_{m=1}^{N} g\left(\hat{\boldsymbol{\theta}}_{m-1}\right) \epsilon_m^2 \\
&= \sum_{m=1}^{N} \epsilon_m^2 \left[I_m - \mathbb{E}(I_m | \mathcal{F}_{m-1})\right] + \sum_{m=1}^{N} \epsilon_m^2 \frac{\boldsymbol{T}_{m-1}}{m-1} \left(\frac{dg}{d\boldsymbol{\theta}}\right)^T \\
&\quad + Ng(\boldsymbol{\theta}) \sum_{m=1}^{N} \epsilon_m^2 + o_p\left(\sqrt{N}\right).
\end{aligned}$$

Combination of the above two equations with the last term in (2.20) yields

$$\sum_{m=1}^{N} \left(\gamma X_m + \epsilon_m\right)^2 I_m = N\gamma^2 \eta(\boldsymbol{\theta}) + g(\boldsymbol{\theta}) \sum_{m=1}^{N} \epsilon_m^2 + O\left(\sqrt{N \log\log N}\right), \quad \text{a.s.}$$

Therefore, we have

$$\sum_{m=1}^{N} Y_m^2 I_m = N_1\mu_1^2 + 2\mu_1 \sum_{m=1}^{N} (\gamma X_m + \epsilon_m) I_m + \sum_{m=1}^{N} (\gamma X_m + \epsilon_m)^2 I_m$$
$$= N_1\mu_1^2 + 2N\mu_1\gamma\phi(\boldsymbol{\theta}) + N\gamma^2\eta(\boldsymbol{\theta}) + g(\boldsymbol{\theta}) \sum_{m=1}^{N} \epsilon_m^2 + O\left(\sqrt{N \log\log N}\right), \quad \text{a.s.}$$

Hence, we get $S_{Y_1}^2 \xrightarrow{P} \gamma^2\sigma_x^2+\sigma_\epsilon^2$ under H_0. And similarly we can get $S_{Y_2}^2 \xrightarrow{P} \gamma^2\sigma_x^2+\sigma_\epsilon^2$. Then, under H_0, we get

$$\sqrt{N}\sqrt{\frac{S_{Y_1}^2}{N_1} + \frac{S_{Y_2}^2}{N_2}} \xrightarrow{P} 2\sqrt{\gamma^2\sigma_x^2 + \sigma_\epsilon^2} = \sigma_a.$$

Finally, under $H_0 : \mu_1 = \mu_2$, we have, by Slutsky's Theorem, that

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{S_{Y_1}^2}{N_1} + \frac{S_{Y_2}^2}{N_2}}} \xrightarrow{D} \mathcal{N}\left(0,\ \frac{\sigma_0^2}{\sigma_a^2}\right), \tag{2.21}$$

where $\sigma_0^2 = 4\left(\gamma^2\sigma_x^2 + \sigma_\epsilon^2\right) + 32\gamma^2 \left(\frac{d\phi}{d\boldsymbol{\theta}}\right) \boldsymbol{V}_0 \left(\frac{d\phi}{d\boldsymbol{\theta}}\right)^T + 16\gamma\sigma_\epsilon^2 \left(\frac{d\phi}{d\mu_1} - \frac{d\phi}{d\mu_2}\right)$, $\sigma_a^2 = 4\left(\gamma^2\sigma_x^2 + \sigma_\epsilon^2\right)$ and $\phi(\cdot) = \mathbb{E}[X \cdot \pi(\cdot, X)]$.

From the definition of the allocation function in (2.3) and the two conditions in Theorem 2.3.1, it follows that $\frac{d\phi}{d\mu_1} = -\frac{d\phi}{d\mu_2}$ and $\frac{d\phi}{d\gamma} = 0$. The last two terms of σ_0^2 represent the difference between the variability increase introduced by estimating allocation probablities and the variability reduction due to the incorporation of covariate information in the process of allocation. The sum of last two terms in σ_0^2 can be written as

$$V_{diff} = 32\gamma^2 \left(\frac{d\phi}{d\boldsymbol{\theta}}\right) \boldsymbol{V}_0 \left(\frac{d\phi}{d\boldsymbol{\theta}}\right)^T + 16\gamma\sigma_\epsilon^2 \left(\frac{d\phi}{d\mu_1} - \frac{d\phi}{d\mu_2}\right) = 32\gamma\sigma_\epsilon^2 \cdot \frac{d\phi}{d\mu_1} \cdot \left(4\gamma \cdot \frac{d\phi}{d\mu_1} + 1\right)$$

$$\Longrightarrow V_{diff} \begin{cases} \leq 0 \text{ if } \dfrac{d\phi}{d\mu_1} \in \left[-\dfrac{1}{4\gamma} \vee 0, -\dfrac{1}{4\gamma} \wedge 0 \right]; \\ > 0 \text{ otherwise.} \end{cases}$$

If V_{diff} is negative, then the asymptotic variance of the test statistic T under H_0 is smaller than 1, resulting in a conservative test in terms of smaller Type I error. If positive, we will get a liberal test. $\square$

Consider a sequence of local alternatives, that is, $\mu_1 - \mu_2 = \delta/\sqrt{N}$ for some fixed $\delta \neq 0$, we can write (2.15) as

$$\bar{Y}_1 - \bar{Y}_2 = \mu_1 - \mu_2 + \frac{\gamma\phi(\boldsymbol{\theta})}{g(\boldsymbol{\theta})[1-g(\boldsymbol{\theta})]} + O_p\left(\sqrt{\frac{\log\log N}{N}}\right). \tag{2.22}$$

And it follows from (2.16) that

$$\begin{aligned} \sigma_d^2 \ \to\ & \frac{\gamma^2\sigma_x^2}{[1-g(\boldsymbol{\theta})]^2} + \frac{\sigma_\epsilon^2}{g(\boldsymbol{\theta})[1-g(\boldsymbol{\theta})]} + \gamma^2\eta(\boldsymbol{\theta}) \cdot \left(\frac{1}{g^2(\boldsymbol{\theta})} - \frac{1}{[1-g(\boldsymbol{\theta})]^2} \right) \\ & + \frac{2\gamma\sigma_\epsilon^2}{g^2(\boldsymbol{\theta})[1-g(\boldsymbol{\theta})]^2} \cdot \left[\frac{d\phi}{d\mu_1} \cdot (1-g(\boldsymbol{\theta})) - \frac{d\phi}{d\mu_2} \cdot g(\boldsymbol{\theta}) + \frac{d\phi}{d\gamma} \cdot \frac{\phi(\boldsymbol{\theta})}{\sigma_x^2} \right] \\ & + \frac{\gamma^2}{g^2(\boldsymbol{\theta})[1-g(\boldsymbol{\theta})]^2} \cdot \left[2 \cdot \left(\frac{d\phi}{d\boldsymbol{\theta}} \right) \boldsymbol{V} \left(\frac{d\phi}{d\boldsymbol{\theta}} \right)^T - \phi^2(\boldsymbol{\theta}) \right] = \sigma_1^2, \qquad \text{a.s.} \end{aligned}$$

Therefore, by the central limit theorem for martingales, we have that under H_a : $\mu_1 - \mu_2 = \delta/\sqrt{N}$,

$$\sqrt{N}\left(\bar{Y}_1 - \bar{Y}_2\right) \ \xrightarrow{D}\ \mathcal{N}\left(\delta + \frac{\sqrt{N}\gamma\phi(\boldsymbol{\theta})}{g(\boldsymbol{\theta})[1-g(\boldsymbol{\theta})]},\ \sigma_1^2 \right). \tag{2.23}$$

Looking at the denominator under the local alternatives, in the similar way, we

have

$$
\begin{aligned}
S^2_{Y_1} &\xrightarrow{P} \frac{\gamma^2\eta(\boldsymbol{\theta})}{g(\boldsymbol{\theta})} - \frac{\gamma^2\phi^2(\boldsymbol{\theta})}{g^2(\boldsymbol{\theta})} + \sigma^2_\epsilon, \\
S^2_{Y_2} &\xrightarrow{P} \frac{\gamma^2[\sigma^2_x - \eta(\boldsymbol{\theta})]}{1-g(\boldsymbol{\theta})} - \frac{\gamma^2\phi^2(\boldsymbol{\theta})}{[1-g(\boldsymbol{\theta})]^2} + \sigma^2_\epsilon.
\end{aligned}
$$

Hence, we get

$$
\begin{aligned}
N \cdot \left[\frac{S^2_{Y_1}}{N_1} + \frac{S^2_{Y_2}}{N_2}\right] \xrightarrow{P} & \frac{\gamma^2\sigma^2_x}{[1-g(\boldsymbol{\theta})]^2} + \frac{\sigma^2_\epsilon}{g(\boldsymbol{\theta})[1-g(\boldsymbol{\theta})]} + \gamma^2\eta(\boldsymbol{\theta}) \cdot \left[\frac{1}{g^2(\boldsymbol{\theta})} - \frac{1}{[1-g(\boldsymbol{\theta})]^2}\right] \\
& - \gamma^2\phi^2(\boldsymbol{\theta}) \left[\frac{1}{g^3(\boldsymbol{\theta})} - \frac{1}{[1-g(\boldsymbol{\theta})]^3}\right] = \sigma^2_b. \qquad (2.24)
\end{aligned}
$$

By Slutsky's Theorem, we have

$$
T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{S^2_{Y_1}}{N_1} + \frac{S^2_{Y_2}}{N_2}}} \xrightarrow{D} \mathcal{N}\left(\frac{1}{\sigma_b} \cdot \left[\delta + \frac{\sqrt{N}\gamma\phi(\boldsymbol{\theta})}{g(\boldsymbol{\theta})[1-g(\boldsymbol{\theta})]}\right],\ \frac{\sigma^2_1}{\sigma^2_b}\right) \qquad (2.25)
$$

under local alternatives, $H_a : \mu_1 - \mu_2 = \delta/\sqrt{N}$ for some fixed $\delta \neq 0$. This completes the proof of Theorem 2.3.1. □

Chapter 3

Statistical Inference under CAR Designs with a Fixed Allocation Proportion

3.1 Introduction

In this chapter, we extend the Hu and Hu procedure [24] by using a fixed allocation proportion, not limited to 1/2, in a two-armed clinical trial. Unequal allocations are preferred in many situations, for example, an active treatment is favored over placebo, or one treatment is extremely expensive in comparison to the other treatment. This design can also be considered as a special case of new CARA designs since it only balances prognostic covariates across treatment arms in the randomization procedure, without considering predictive covariates for treatment selection. Statistical inference under this covariate-adaptive design is investigated with a linear model. Asymptotic distributions of test statistics for comparing treatment effects and for testing the significance of covariates, especially the interaction term for covariate and treatment, are derived under both the null and alternative hypotheses. Several conclusions can be drawn: The hypothesis testing for comparing treatment effects is usually conservative when prognostic covariates balanced in randomization are omitted in the final

analysis, while the one for testing the significance of the interaction term between the covariate X and treatment I can still achieve nominal Type I error.

This chapter is organized as follows. In Section 3.2, the framework of the covariate-adaptive design with a fixed allocation proportion under a linear model is introduced to study hypothesis testing properties for testing treatment effects and significance of covariates. Both an underlying full model and a working model are proposed to represent the situation that prognostic covariates used in randomization are omitted in statistical inference. Theoretical results for test statistics under both the null and alternative hypotheses are presented in Section 3.3. Simulation studies are conducted in Section 3.4 to study Type I error and power of the covariate-adaptive design under multiple scenarios compared with complete randomization. Conclusions are given in Section 3.5, and major proof of theoretical results is shown in Section 3.6.

3.2 Framework

Consider a two-armed clinical trial studied under the covariate-adaptive design. Suppose the target allocation proportion to treatment 1 is ρ. Let $(X_m,\ Z_{m,1},\ Z_{m,2})$ be the covariate profile of the m-th subject. Conditional on the treatment assignment I_m, the following linear model is assumed for the response of the m-th subject Y_m,

$$Y_m = \mu_1 I_m + \mu_2 (1 - I_m) + \beta_1 X_m + \beta_2 X_m I_m + \gamma_1 Z_{m,1} + \gamma_2 Z_{m,2} + \epsilon_m, \qquad (3.1)$$

where

- X_m's, $Z_{m,1}$'s and $Z_{m,2}$'s, $m = 1, \ldots, N$ are discrete variables, which are independent and identically distributed as X, Z_1 and Z_2;
- Z_1 and Z_2 can take multiple values, and X only takes two values a and b with probability p and $1 - p$ respectively;

- all covariates are independent of each other with mean zero, and $\mathsf{Var}(X) = \sigma_x^2$, $\mathsf{Var}(Z_1) = \sigma_1^2$ and $\mathsf{Var}(Z_2) = \sigma_2^2$;

- both $Z_{m,1}$'s and $Z_{m,2}$'s are used to be balanced in the randomization procedure, but only X_m's are used in statistical inference;

- ϵ_i's are independent and identically distributed random errors with mean zero and variance σ_ϵ^2, and are independent of X, Z_1 and Z_2.

Notice that we only consider discrete covariates in this chapter. Extension to the situation, where X is dichotomous and Z_1, Z_2 are continuous, is considered in Chapter 5. Also, an interaction term between the covariate X and treatment assignment I is included in the full model (3.1) to differentiate the covariate effects of X for two treatment groups, while the covariate effects of Z_1 and Z_2 have no difference across treatment groups respectively. Since Z_1 and Z_2 are incorporated in the randomization procedure for balancing purpose, they are considered as prognostic covariates.

We define vectors $\boldsymbol{Y} = (Y_1, \ldots, Y_N)^T$, $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_N)^T$, $\boldsymbol{\beta} = (\mu_1, \mu_2, \beta_1, \beta_2)^T$, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)^T$ and matrices

$$\boldsymbol{X} = \begin{pmatrix} I_1 & 1 - I_1 & X_1 & X_1 I_1 \\ \vdots & \vdots & \vdots & \vdots \\ I_N & 1 - I_N & X_N & X_N I_N \end{pmatrix} \quad \text{and} \quad \boldsymbol{Z} = \begin{pmatrix} Z_{1,1} & Z_{1,2} \\ \vdots & \vdots \\ Z_{N,1} & Z_{N,2} \end{pmatrix}.$$

Write the underlying full model (3.1) into a matrix form, then we get

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

3.2.1 CAR Designs

A general covariate-adaptive design is defined by

$$\psi_m = \mathbb{E}\left(I_m | \mathcal{I}_{m-1}, \mathcal{Z}_m\right), \ \mathcal{Z}_m = \sigma\left(Z_{1,1}, \ldots, Z_{m,1}, Z_{1,2}, \ldots, Z_{m,2}\right),$$

the conditional probabilities of assigning the m-th subject to treatment 1, conditioning on the treatment assignments and prognostic covariates of the previous $m-1$ subjects, as well as prognostic covariates of the current subject. The target allocation proportion to treatment 1 is supposed to be ρ.

For convenience, we use (z_1, z_2) to denote the *stratum* which consists of subjects with the covariate profile $(Z_1, Z_2) = (z_1, z_2)$. $(j; z_j)$ denotes the *margin* formed by subjects with $Z_j = z_j, \ j = 1, 2$.

The allocation scheme of the covariate-adaptive design is as follows:

1. Assign m_0 subjects randomly to each treatment by using some restricted randomization design.

2. Assume that $n-1$ assignments have been finished ($n > 2m_0$). Corresponding covariates $\{(Z_{i,1}, Z_{i,2}), \ i = 1, \ldots, n-1\}$ are observed. The m-th subject has covariate profile $(Z_{n,1}, Z_{n,2}) = (z_1^*, z_2^*)$. Then, the n-th subject belongs to the stratum (z_1^*, z_2^*) and margins $(j; z_j^*), \ j = 1, 2$.

3. If we assign the n-th subject to treatment 1,

   - let $D_n^{(1)} = N_{n1} - \rho \cdot n$, where N_{n1} is the number of subjects in treatment 1 after n assignments.

   - similarly, let $D_n^{(1)}(j; z_j^*) = N_{n1}(j; z_j^*) - \rho N_n(j; z_j^*)$ be the differences between the actual and target numbers of subjects in treatment 1 on the margins $(j; z_j^*), \ j = 1, 2$, and $D_n^{(1)}(z_1^*, z_2^*) = N_{n1}(z_1^*, z_2^*) - \rho N_n(z_1^*, z_2^*)$ be

the difference within the stratum (z_1^*, z_2^*).

4. Define an imbalance measure $Imb_n^{(1)}$ by

$$Imb_n^{(1)} = \omega_o \left[D_n^{(1)}\right]^2 + \sum_{j=1}^{2} \omega_{m,j} \left[D_n^{(1)}(j; z_j^*)\right]^2 + \omega_s \left[D_n^{(1)}(z_1^*, z_2^*)\right]^2, \tag{3.2}$$

which is the weighted sum of imbalances at the overall, marginal and within-stratum levels caused by assigning the n-th subject to treatment 1. Nonnegative weights $\omega_o,\ \omega_{m,j},\ j = 1, 2$ and ω_s are placed on the overall, marginal and within-stratum levels, respectively. Without loss of generality, we assume

$$\omega_o + \sum_{j=1}^{2} \omega_{m,j} + \omega_s = 1.$$

5. If we assign the n-th subject to treatment 2, the imbalance measure $Imb_n^{(2)}$ can be calculated in the same way, which represents the weighted sum of imbalances caused by assigning the n-th subject to treatment 2.

6. Then conditioning on treatment assignments of the previous $n-1$ subjects and prognostic covariate profiles of all first n subjects, we assign the n-th subject to treatment 1 with probability

$$\mathbb{P}\left(I_n = 1 | \mathcal{I}_{n-1}, \mathcal{Z}_n\right) = \begin{cases} p^*, & \text{if } Imb_n^{(1)} < Imb_n^{(2)}; \\ 1/2, & \text{if } Imb_n^{(1)} = Imb_n^{(2)}; \\ q^*, & \text{if } Imb_n^{(1)} > Imb_n^{(2)}, \end{cases} \tag{3.3}$$

where p^* is the biasing probability satisfying $0 < q^* < p^* < 1,\ p^* + q^* = 1$.

7. Repeat step 2 - step 6 until each subject is assigned to a treatment.

Without incorporating the covariate X in randomization, this covariate-adaptive de-

sign is a special case of new CARA designs; if the target allocation proportion $\rho = 1/2$, this procedure reduces to the design proposed by Hu and Hu [24].

3.2.2 Hypothesis Testing

In this section, we study statistical properties of hypothesis testings under the covariate-adaptive design when prognostic covariates Z_1 and Z_2 are excluded from the final analysis. The working model of inference would be

$$\mathbb{E}(Y_m) = \mu_1 I_m + \mu_2(1 - I_m) + \beta_1 X_m + \beta_2 X_m I_m, \tag{3.4}$$

and in the matrix form $\mathbb{E}(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta}$. Under the working model (3.4), the estimator of $\boldsymbol{\beta}$ using least square estimation is

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} = \boldsymbol{\beta} + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}).$$

The primary interest is to compare the two treatment effects μ_1 and μ_2, and the hypothesis testing (2.5) is conducted. The corresponding test statistic is

$$T = \frac{\boldsymbol{L}\hat{\boldsymbol{\beta}}}{\sqrt{\hat{\sigma}^2\boldsymbol{L}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}^T}}, \tag{3.5}$$

where $\boldsymbol{L} = (1, -1, 0, 0)$ and $\hat{\sigma}^2 = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})/(N - 4)$. 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 it.

We are also interested in the coefficient of the interaction term for the covariate X and treatment I. It indicates that subjects potentially have different responses for the two treatments. Then, the covariate X may contribute to treatment selection, and we can develop a procedure with incorporation of the covariate X in randomization, resulting

in the new CARA design that deals with both types of covariates simultaneously. Therefore, another hypothesis testing of interest is

$$H_0 : \ \beta_2 = 0 \text{ versus } H_A : \ \beta_2 \neq 0, \tag{3.6}$$

and the corresponding test statistic is

$$T_* = \frac{\boldsymbol{l}\hat{\boldsymbol{\beta}}}{\sqrt{\hat{\sigma}^2 \boldsymbol{l}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{l}^T}}, \tag{3.7}$$

where $\boldsymbol{l} = (0, 0, 0, 1)$. 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 it.

3.3 Theoretical Properties

Two hypothesis testings are considered in this section, one is to compare the two treatment effects μ_1 and μ_2, and the other is to test the significance of the interaction term in (3.1). These hypotheses are conducted based on the working model (3.4) while the data are generated from the underlying full model (3.1). Asymptotic properties of test statistics are studied under both the null and alternative hypotheses in the following theorems.

THEOREM 3.3.1. Suppose that the covariate-adaptive design in Section 3.2.1 satisfies the following two conditions:

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

(B) the marginal imbalances for Z_1 and Z_2 are bounded in probability, that is, $D_N(j; z_j) = O_p(1), \ j = 1, 2.$

Then, we have

(i) under $H_0: \ \mu_1 - \mu_2 = 0$, the test statistic (3.5) has the following asymptotic distribution:

$$T = \frac{\boldsymbol{L}\hat{\boldsymbol{\beta}}}{\sqrt{\hat{\sigma}^2 \boldsymbol{L}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}^T}} \ \xrightarrow{D} \ \mathcal{N}\left(0, \ \frac{\sigma_\epsilon^2}{\sigma^2}\right), \tag{3.8}$$

where $\sigma^2 = \gamma_1^2\sigma_1^2 + \gamma_2^2\sigma_2^2 + \sigma_\epsilon^2$.

Therefore,

(1) if $\gamma_j = 0, \ j = 1, 2$, then $\sigma_\epsilon^2/\sigma^2 = 1$. Thus, when prognostic covariates Z_1 and Z_2 are completely not associated with the response Y, the hypothesis testing (2.5) can achieve nominal Type I error. However, if at least one $\gamma_j \neq 0, \ j = 1, 2$, then $\sigma_\epsilon^2/\sigma^2 < 1$, resulting in a conservative test.

(2) The asymptotic distribution of T under H_0 does not depend on the target allocation proportion ρ used in randomization, treatment effects μ_1 and μ_2, nor on the covariate X and its related parameters in (3.1).

(ii) under $H_a: \ \mu_1 - \mu_2 \neq 0$, consider a sequence of local alternatives, that is, $\mu_1 - \mu_2 = \delta/\sqrt{N}$ for some fixed $\delta \neq 0$, then the test statistic (3.5) has the following asymptotic distribution:

$$T = \frac{\boldsymbol{L}\hat{\boldsymbol{\beta}}}{\sqrt{\hat{\sigma}^2 \boldsymbol{L}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}^T}} \ \xrightarrow{D} \ \mathcal{N}\left(\frac{\delta \cdot \sqrt{\rho(1-\rho)}}{\sigma}, \ \frac{\sigma_\epsilon^2}{\sigma^2}\right). \tag{3.9}$$

Thus, the limiting power reaches its maximum when $\rho = 1/2$, and it still does not depend on X and its related parameters in (3.1).

Under the alternative hypothesis, the power under the covariate-adaptive design is

$$\mathbb{P}\left(|T| > Z_{1-\alpha/2}\right) = \Phi\left(\frac{\delta\sqrt{\rho(1-\rho)}}{\sigma_\epsilon} - \frac{\sigma}{\sigma_\epsilon}Z_{1-\alpha/2}\right) + \Phi\left(-\frac{\delta\sqrt{\rho(1-\rho)}}{\sigma_\epsilon} - \frac{\sigma}{\sigma_\epsilon}Z_{1-\alpha/2}\right) + o(1).$$

It is straightforward to see that the limiting power under the covariate-adaptive design

reaches the maximum when $\rho = 1/2$. The power under complete randomization is

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

The limiting power under the covariate-adaptive design is smaller than that under complete randomization when δ is relatively small, and it is larger than complete randomization when δ is large.

The asymptotic properties of hypothesis testing for significance of the interaction term (3.6) are shown in the following theorem.

THEOREM 3.3.2. Suppose that the covariate-adaptive design in Section 3.2.1 satisfies that all within-stratum imbalances are $O_p(1)$, then we have

(i) under $H_0 : \ \beta_2 = 0$,

$$T_* = \frac{\boldsymbol{l}\hat{\boldsymbol{\beta}}}{\sqrt{\hat{\sigma}^2 \boldsymbol{l}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{l}^T}} \xrightarrow{D} \mathcal{N}(0,\ 1). \tag{3.10}$$

Therefore, the hypothesis testing (3.6) can achieve nominal Type I error.

(ii) under $H_a : \ \beta_2 \neq 0$, consider a sequence of local alternatives, that is, $\beta_2 = \delta_{\beta_2}/\sqrt{N}$ for some fixed $\delta_{\beta_2} \neq 0$, then the test statistic (3.7) has the following asymptotic distribution:

$$T_* = \frac{\boldsymbol{l}\hat{\boldsymbol{\beta}}}{\sqrt{\hat{\sigma}^2 \boldsymbol{l}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{l}^T}} \xrightarrow{D} \mathcal{N}\left(\frac{\delta_{\beta_2} \cdot \sigma_x \sqrt{\rho(1-\rho)}}{\sigma},\ 1\right). \tag{3.11}$$

Thus, the limiting power reaches its maximum when $\rho = 1/2$ and it depends on variance of X, but not on parameters related with X.

Theoretical properties of hypothesis testings to compare two treatment effects and to test significance of the interaction term are obtained under covariate-adaptive designs in Theorem 3.3.1 and Theorem 3.3.2, respectively. Prognostic covariates Z_1

and Z_2 are used in the randomization procedure to balance treatment allocations for credibility of the trial. And the design deals with three levels of imbalances (overall, marginal and within-stratum) simultaneously. To derive the asymptotic distributions of test statistic for comparing treatment effects under both the null and alternative hypotheses, we assume that the overall and all marginal imbalances are bounded in probability. With these conditions satisfied, we show that the asymptotic variance of test statistic is smaller than 1 when prognostic covariates is omitted in the working model, i.e., the numerator of the test statistic, $\boldsymbol{L}\hat{\boldsymbol{\beta}} = \hat{\mu}_1 - \hat{\mu}_2$, has variance smaller than the model-based variance estimator in the denominator, which leads to a conservative test. Larger effects of prognostic covariates on responses will result in a more conservative test. While a valid test for significance of the interaction term can still be obtained given all within-stratum imbalances are bounded in probability. The limiting power of both hypothesis testings depends on the target allocation proportion ρ used in randomization, and reaches its maximum using equal allocation proportion, i.e., $\rho = 1/2$.

3.4 Simulation Study

Case 1: Testing treatment effects. Simulations are conducted to study Type I error and power of the hypothesis testing for comparing two treatment effects under the covariate-adaptive design (CAR) and complete randomization (CR). The data are generated from the underlying model (3.1). The two prognostic covariates Z_1 and Z_2 are distributed independently with equal probability to take value 1 or -1. We set a in the distribution of the covariate X to be 1 with different p's, then these two values can determine the distribution of X. The random error ϵ follows a standard normal distribution. To study Type I error, we assume there is no difference between two treatment effects, i.e., we have $\mu_1 = \mu_2$.

The hypothesis tests include the test ($lm(X)$) under the working model (3.4) and the corresponding bootstrap test (BS), and the test under the full model (3.1) ($lm(X, Z)$). Sample size is $N = 500$ and the significance level is $\alpha = 0.05$. For the covariate-adaptive design, $2m_0 = 20$ subjects are assigned to the two treatments using permuted block design with block size 4 at the beginning of allocation. Simulation results demonstrated in Table 3.1 are based on 10,000 simulations. We use equal weights, i.e., $\omega_o = \omega_{m,1} = \omega_{m,2} = \omega_s = 1/4$, for three levels of imbalances. The following scenarios of parameters are used to study Type I error.

Scenario	μ_1	μ_2	β_1	β_2	γ_1	γ_2	p	ρ
S_1	0	0	1	1	1/2	1/2	1/2	1/2
S_2	0	0	1/2	-1/3	1/2	1/2	1/4	1/2
S_3	1	1	1	1	1/2	1/2	1/2	1/2
S_4	0	0	1	1	1/2	1/2	1/2	2/3
S_5	0	0	1	1	1	1	1/2	1/2
S_6	2	2	1/2	1/3	1/2	-1/2	1/3	1/4

Table 3.1: Simulated Type I error for comparing treatment effects under the covariate-adaptive design (CAR) and complete randomization (CR) in %. Simulations are based on 10,000 runs and sample size $N = 500$.

Scenario	CAR		CR	
	$lm(X)$	$lm(X, Z)$	$lm(X)$	$lm(X, Z)$
S_1	1.69	4.73	5.30	5.11
S_2	1.62	4.79	5.03	4.99
S_3	1.54	5.20	5.20	5.23
S_4	1.78	5.00	-	-
S_5	0.06	5.28	4.71	4.79
S_6	1.94	5.10	4.98	4.71

Several conclusions can be drawn from Table 3.1. First, Type I error is affected under the covariate-adaptive design (CAR) when prognostic covariates are omitted in the

working model of inference ($lm(X)$). And results agree with theoretical properties that the test is conservative in terms of smaller Type I error. From the results of scenarios $S_1 - S_4$, we can see that the performance of test does not depend on the distribution of X and its related parameters β_1 and β_2, the treatment effects μ_1 and μ_2, nor the allocation proportion ρ. While results of S_1, S_5 and S_6 indicate that the larger the covariate effects of Z_1 and Z_2, the more conservative the test under the working model ($lm(X)$). Second, the test under the covariate-adaptive design (CAR) using the full model (3.1) ($lm(X, Z)$) can achieve nominal Type I error as expected. Third, Type I error under complete randomization (CR) is always close to 5% for the two hypothesis testing models.

Case 2: Power comparison. Next we compare the power for the two hypothesis testings, $lm(X)$ and $lm(X, Z)$, under three procedures, namely the covariate-adaptive design with the allocation proportion $\rho = 1/2$ (CAR1), $\rho = 2/3$ (CAR2) and complete randomization (CR). The same model in Case 1 of Section 3.4 is applied, but there exists a difference between two treatment effects, i.e., we have $\mu_1 - \mu_2 \neq 0$. The results of power comparison are given in Figure 3.1, and the conclusions are summarized as follows.

First, for the hypothesis testing using the working model ($lm(X)$), the covariate-adaptive designs with both allocation proportions (CAR1 and CAR2) are less powerful than that under complete randomization (CR) when the difference between two treatment effects is relatively small, and have larger power when the difference becomes larger. Moreover, the power of the covariate-adaptive design with $\rho = 1/2$ (CAR1) is always slightly larger than that with $\rho = 2/3$ (CAR2), which is consistent with theoretical results. Second, the hypothesis testing using the full model ($lm(X, Z)$) is more powerful than that using the working model ($lm(X)$) under all three procedures. CAR1 and complete randomization (CR) have very similar power

performance using the full model ($lm(X, Z)$), while CAR2 loses power a little. Third, CAR1 does not improve average responses much compared with complete randomization (CR) since it still uses equal allocation proportion. However, by assigning more subjects to the better treatment, CAR2 can obtain significantly better average responses.

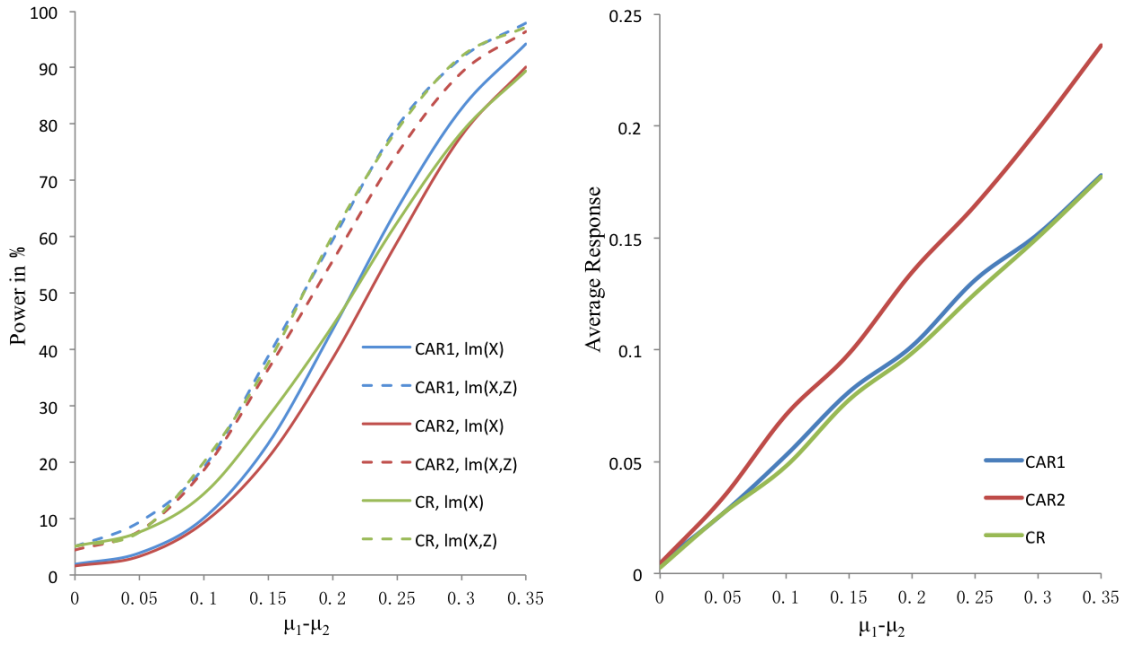

Figure 3.1: Simulated power of comparing treatment effects (left) and average responses (right) under the covariate-adaptive design with $\rho = 1/2$ (CAR1), $\rho = 2/3$ (CAR2) and complete randomization (CR). Simulations are based on sample size $N = 500$ and 1000 runs.

Case 3: Significance of interaction term. To explore the interaction effect of the potential predictive covariate X and the treatment I, we study the hypothesis testings for the significance of the interaction term based on both the working model and the full model, i.e., $lm(X)$ and $lm(X, Z)$, under the covariate-adaptive design (CAR) and complete randomization (CR). Sample size is $N = 500$ and the significance level is $\alpha = 0.05$. Four different scenarios of parameters are used to study Type I error, and the results are given in Table 3.2.

From Table 3.2, it can be seen that tests of β_2 using the working model ($lm(X)$) and

Scenario	μ_1	μ_2	β_1	β_2	γ_1	γ_2	p	ρ
S_1	1	1	1	0	1/2	1/2	1/2	1/2
S_2	1	1	1/2	0	1/2	1/2	1/3	1/2
S_3	1	1	1	0	1/2	1/2	1/2	2/3
S_4	1/2	0	1	0	-1/2	1/3	1/2	1/2

Table 3.2: Simulated Type I error for the significance of interaction term under the covariate-adaptive design (CAR) and complete randomization (CR) in %. Simulations are based on 10,000 runs and sample size $N = 500$.

Scenario	CAR		CR	
	$lm(X)$	$lm(X, Z)$	$lm(X)$	$lm(X, Z)$
S_1	5.33	5.40	5.11	5.26
S_2	5.02	4.84	4.84	4.99
S_3	4.94	5.23	-	-
S_4	5.02	4.99	5.28	5.24

the full model ($lm(X, Z)$) are valid in terms of Type I error under the two procedures, regardless of the distribution of X, the allocation proportion ρ, and parameters in the employed model, which is consistent with Theorem 3.3.2. Results in Table 3.3 indicate that the power of testing β_2 based on the full model ($lm(X, Z)$) is more powerful than that based on the working model ($lm(X)$) under all three procedures, and the covariate-adaptive design with $\rho = 2/3$ (CAR2) slightly loses power.

3.5 Conclusion

The covariate-adaptive design incorporates prognostic covariates in the randomization procedure for balancing purpose to provide a valid comparison for different treatments, which may have an influence on statistical inference. In this chapter, we study two types of hypothesis testings under the covariate-adaptive design with prognostic covariates excluded from the final analysis. Asymptotic distributions of test statistics

Table 3.3: Simulated power for the significance of interaction term using the working model ($lm(X)$) and the full model ($lm(X, Z)$) under the covariate-adaptive design with $\rho = 1/2$ (CAR1), $\rho = 2/3$ (CAR2) and complete randomization (CR) in %. Simulations are based on 1,000 runs and sample size $N = 500$.

	CAR1		CAR2		CR	
β_2	$lm(X)$	$lm(X, Z)$	$lm(X)$	$lm(X, Z)$	$lm(X)$	$lm(X, Z)$
0.00	5.5	5.4	4.9	5.5	4.8	5
0.05	7.4	8.9	7.1	7.9	7.0	8.8
0.10	14.7	20.7	13.8	17.5	14.8	20.4
0.15	27.7	36.5	25.4	34.2	27.2	36.9
0.20	44.9	61.2	40.6	55.8	44.6	60.1
0.25	63.2	78.3	57.5	75.7	62.8	80.1
0.30	77.7	90.5	72.7	88.3	77.8	89.6
0.35	88.7	97.6	86.0	94.9	88.6	96.9
0.40	95.2	99.5	93.2	98.9	94.8	99.5
0.45	98.4	99.7	97.3	99.7	98.2	99.9
0.50	99.5	100.0	98.9	99.8	99.5	100.0

under both the null and alternative hypotheses are derived under some mild conditions for the overall, marginal and within-stratum imbalances. The test for comparing treatment effects is usually conservative under the covariate-adaptive design if important prognostic covariates are omitted, while a valid test can still be obtained for testing the significance of the interaction term. In the case that an allocation proportion in favor of the better treatment is used in a clinical study, the extended covariate-adaptive design can significantly improve average responses with only slight power loss.

In this chapter, we restrict prognostic covariates to be discrete in the underlying model (3.1). However, many continuous covariates, that need to be balanced, are used in clinical trial practice, such as age, blood pressure, etc. Continuous-discrete

conversion can be applied for the implementation of the covariate-adaptive design, but it will lose information of covariates. Similar problems with direct use of continuous prognostic covariates in the underlying full model are studied in Chapter 5. Moreover, the covariate X, which may have different effects on responses depending on treatment arms, is not used in the randomization scheme. It can work as a predictive covariate to facilitate treatment selection. The new CARA design utilizing both predictive and prognostic covariates in randomization and its statistical properties of hypothesis testings will be investigated in Chapter 4.

3.6 Appendix: Proof of Theorems

Proof of Theorem 3.3.1.

According to the condition (A) in Theorem 3.3.1 and the definition of the overall imbalance in treatment 1, we have

$$\frac{1}{N}\sum_{i=1}^{N} I_i = \frac{1}{N}\left[D_N + \rho N\right] = \rho + \frac{1}{N}\cdot D_N \xrightarrow{P} \rho.$$

The condition (B) assumes that marginal imbalances $D_N(j; z_j) = O_p(1),\ j = 1, 2$ and Z_j take finite discrete values, together with weak law of large numbers, they imply that

$$\frac{1}{N}\sum_{i=1}^{N} I_i Z_{i,j} = \frac{1}{N}\sum_{z_j} z_j N_1(Z_j = z_j) = \frac{1}{N}\sum_{z_j} z_j D_N(j; z_j) + \frac{\rho}{N}\sum_{i=1}^{N} Z_{i,j} \xrightarrow{P} 0.$$

Since X and I are independent, it follows from weak law of large numbers that

$$\frac{1}{N}\sum_{i=1}^{N} I_i X_i \ \xrightarrow{P}\ 0 \ \text{ and } \ \frac{1}{N}\sum_{i=1}^{N} I_i X_i^2 \ \xrightarrow{P}\ \rho\sigma_x^2.$$

Hence, we have

$$\frac{\boldsymbol{X}^t\boldsymbol{X}}{N} = \frac{1}{N}\begin{pmatrix} \sum I_i & \sum I_i(1-I_i) & \sum X_i I_i & \sum X_i I_i \\ & \sum (1-I_i)^2 & \sum X_i(1-I_i) & \sum X_i I_i(1-I_i) \\ & & \sum X_i^2 & \sum X_i^2 I_i \\ & & & \sum X_i^2 I_i \end{pmatrix}$$

$$\xrightarrow{P} \begin{pmatrix} \rho & & & \\ & 1-\rho & & \\ & & \sigma_x^2 & \rho\sigma_x^2 \\ & & \rho\sigma_x^2 & \rho\sigma_x^2 \end{pmatrix} = \boldsymbol{M}.$$

Let

$$A = \boldsymbol{L}\boldsymbol{M}^{-1}\left(\frac{\boldsymbol{X}^t\boldsymbol{Z}\boldsymbol{\gamma}}{N} + \frac{\boldsymbol{X}^t\boldsymbol{\epsilon}}{N}\right), \tag{3.12}$$

and

$$B = \boldsymbol{L}\left[\left(\frac{\boldsymbol{X}^t\boldsymbol{X}}{N}\right)^{-1} - \boldsymbol{M}^{-1}\right]\left(\frac{\boldsymbol{X}^t\boldsymbol{Z}\boldsymbol{\gamma}}{N} + \frac{\boldsymbol{X}^t\boldsymbol{\epsilon}}{N}\right), \tag{3.13}$$

where $\boldsymbol{L} = (1, -1, 0, 0)$. After some matrix calculations, we have

$$\begin{aligned} A =& \frac{1}{N}\cdot\left\{\gamma_1\left[\frac{\sum_{i=1}^N I_i Z_{i,1}}{\rho} - \frac{\sum_{i=1}^N (1-I_i)Z_{i,1}}{1-\rho}\right] + \gamma_2\left[\frac{\sum_{i=1}^N I_i Z_{i,2}}{\rho} - \frac{\sum_{i=1}^N (1-I_i)Z_{i,2}}{1-\rho}\right]\right. \\ &\left. + \left[\frac{\sum_{i=1}^N I_i\epsilon_i}{\rho} - \frac{\sum_{i=1}^N (1-I_i)\epsilon_i}{1-\rho}\right]\right\} = \frac{1}{N}\cdot(A_1 + A_2 + A_3). \end{aligned}$$

We already know that $\sum_{i=1}^N I_i Z_{i,j} = \sum_{z_j} z_j D_N(j; z_j) + \rho\sum_{i=1}^N Z_{i,j}$ and $\sum_{i=1}^N (1 - I_i)Z_{i,j} = (1-\rho)\sum_{i=1}^N Z_{i,j} - \sum_{z_j} z_j D_N(j; z_j)$. Since marginal imbalances are $O_p(1)$,

we have

$$\frac{A_j}{\sqrt{N}} = \frac{\sum_{z_j} z_j D_N(j; z_j)}{\sqrt{N}\rho(1-\rho)} = o_p(1),\ j = 1, 2.$$

Define $\tilde{I} = (I_1, \ldots, I_N)$. According to the convergence of $\sum_{i=1}^N I_i/N$ and the independence between I and ϵ, we have

$$\mathsf{Var}\left(\frac{1}{\sqrt{N}} \sum_{i=1}^N I_i \epsilon_i \middle| \tilde{I} \right) = \sigma_\epsilon^2 \cdot \frac{\sum_{i=1}^N I_i}{N} \to \rho\sigma_\epsilon^2,$$

$$\mathsf{Cov}\left(\frac{\sum_{i=1}^N \epsilon_i}{\sqrt{N}},\ \frac{\sum_{i=1}^N I_i \epsilon_i}{\sqrt{N}} \middle| \tilde{I} \right) = \sigma_\epsilon^2 \cdot \frac{\sum_{i=1}^N I_i}{N} \to \rho\sigma_\epsilon^2.$$

Hence, by the central limit theorem, given $\tilde{I}$, $A_3/\sqrt{N}$ is asymptotically normal with mean zero and variance $\dfrac{\sigma_\epsilon^2}{\rho(1-\rho)}$. Since the limiting distribution is independent of $\tilde{I}$, this also holds unconditionally. Then, by Slutsky's theorem, we have

$$\sqrt{N}A \ \xrightarrow{D}\ \mathcal{N}\left(0,\ \frac{\sigma_\epsilon^2}{\rho(1-\rho)}\right). \tag{3.14}$$

We already prove that $\left(\dfrac{\boldsymbol{X}^t\boldsymbol{X}}{N}\right)^{-1} - \boldsymbol{M}^{-1} = o_p(1)$. It is straightforward that

$$\frac{\boldsymbol{X}^T\boldsymbol{Z}\boldsymbol{\gamma}}{\sqrt{N}} + \frac{\boldsymbol{X}^T\boldsymbol{\epsilon}}{\sqrt{N}} = O_p(1).$$

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

$$\sqrt{N}\left[(\hat{\mu}_1 - \hat{\mu}_2) - (\mu_1 - \mu_2)\right] \ \xrightarrow{D}\ \mathcal{N}\left(0,\ \frac{\sigma_\epsilon^2}{\rho(1-\rho)}\right).$$

Notice that

$$\boldsymbol{L}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}^T = \frac{1}{N}\ \boldsymbol{L}\left(\frac{\boldsymbol{X}^T\boldsymbol{X}}{N}\right)^{-1}\boldsymbol{L}^T = \frac{1}{N\rho(1-\rho)} + o_p\left(\frac{1}{N}\right)$$

and

$$\begin{aligned}\hat{\sigma}^2 =& \frac{1}{N-4}\left(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}\right)^T\left(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}\right)\\ =& \frac{1}{N}\left(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\right)^T\left(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\right)+\frac{1}{N}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)^T\boldsymbol{X}^T\boldsymbol{X}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\\ &+\frac{2}{N}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\boldsymbol{X}^T\left(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}\right)+o_p(1). \end{aligned} \tag{3.15}$$

Because of the consistency of $\hat{\boldsymbol{\beta}}$, it follows from weak law of large numbers that $\hat{\sigma}^2 \xrightarrow{P} \gamma_1^2\sigma_1^2+\gamma_2^2\sigma_2^2+\sigma_\epsilon^2=\sigma^2$. Hence, we have

$$\hat{\sigma}^2\boldsymbol{L}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{L}^T=\frac{\sigma^2}{N\rho(1-\rho)}+o_p\left(\frac{1}{N}\right).$$

Then under H_0: $\mu_1-\mu_2=0$, by Slutsky's theorem, we get

$$T \xrightarrow{D} \mathcal{N}\left(0,\ \frac{\sigma_\epsilon^2}{\sigma^2}\right),\ \text{where } \sigma^2=\gamma_1^2\sigma_1^2+\gamma_2^2\sigma_2^2+\sigma_\epsilon^2. \tag{3.16}$$

When prognostic covariates balanced in randomization are excluded from the working model of inference, the test of comparing treatment effects is usually conservative in terms of smaller Type I error. Neither the allocation proportion ρ, treatment effects μ_1 and μ_2, nor the covariate X and its related parameters have effects on the performance of the test. □

Similarly, under H_a : $\mu_1-\mu_2\neq 0$ with a sequence of local alternatives, i.e., we have $\mu_1-\mu_2=\delta/\sqrt{N}$ for a fixed $\delta\neq 0$, the test statistics has the following asymptotic distribution

$$T \xrightarrow{D} \mathcal{N}\left(\frac{\delta\sqrt{\rho(1-\rho)}}{\sigma},\ \frac{\sigma_\epsilon^2}{\sigma^2}\right). \tag{3.17}$$

The limiting power under the covariate-adaptive design reaches its maximum using $\rho=1/2$, and decreases as ρ becomes more extreme. This completes the proof of Theorem 3.3.1. □

Proof of Theorem 3.3.2.

According to the condition of Theorem 3.3.2 that all within-stratum imbalances are $O_p(1)$, we automatically have the overall imbalance and all marginal imbalances are also $O_p(1)$. In the similar way, let

$$A_* = \boldsymbol{l}\boldsymbol{M}^{-1}\left(\frac{\boldsymbol{X}^t\boldsymbol{Z}\boldsymbol{\gamma}}{N} + \frac{\boldsymbol{X}^t\boldsymbol{\epsilon}}{N}\right), \tag{3.18}$$

and

$$B_* = \boldsymbol{l}\left[\left(\frac{\boldsymbol{X}^t\boldsymbol{X}}{N}\right)^{-1} - \boldsymbol{M}^{-1}\right]\left(\frac{\boldsymbol{X}^t\boldsymbol{Z}\boldsymbol{\gamma}}{N} + \frac{\boldsymbol{X}^t\boldsymbol{\epsilon}}{N}\right), \tag{3.19}$$

where $\boldsymbol{l} = (0,0,0,1)$. After some matrix calculation, we have

$$\begin{aligned} A_* =& \frac{1}{(1-\rho)\sigma_x^2}\cdot\frac{1}{N}\Bigg\{-\left[\gamma_1\sum_{i=1}^N X_iZ_{i,1} + \gamma_2\sum_{i=1}^N X_iZ_{i,2} + \sum_{i=1}^N X_i\epsilon_i\right] \\ &+ \frac{1}{\rho}\cdot\left[\gamma_1\sum_{i=1}^N I_iX_iZ_{i,1} + \gamma_2\sum_{i=1}^N I_iX_iZ_{i,2} + \sum_{i=1}^N I_iX_i\epsilon_i\right]\Bigg\} \\ =& \frac{1}{(1-\rho)\sigma_x^2}\cdot\left(\frac{A_{1*}}{N} + \frac{A_{2*}}{N}\right). \end{aligned}$$

Define $\tilde{Z} = \{Z_{i,j},\ i = 1,\ldots,N,\ j = 1,2\}$. Given $\tilde{Z}$, it is easy to get from weak law of large numbers that

$$\mathsf{Var}\left(\frac{A_{1*}}{\sqrt{N}}\middle|\,\tilde{Z}\right) = \sigma_x^2\cdot\left(\frac{\gamma_1^2\sum_{i=1}^N Z_{i,1}^2}{N} + \frac{\gamma_2^2\sum_{i=1}^N Z_{i,1}^2}{N} + \sigma_\epsilon^2\right) \to \sigma_x^2\sigma^2.$$

According to the condition that marginal imbalances are $O_p(1)$, we have

$$\frac{1}{N}\sum_{i=1}^N I_iZ_{i,j}^2 = \frac{1}{N}\sum_{z_j} z_j^2 D_N(j;z_j) + \frac{\rho}{N}\sum_{i=1}^N Z_{i,j}^2 \xrightarrow{P} \rho\sigma_j^2,\ j = 1,2.$$

And the condition that within-stratum imbalances are $O_p(1)$ indicates

$$\frac{1}{N}\sum_{i=1}^{N} I_i Z_{i,1} Z_{i,2} = \frac{1}{N}\sum_{z_1,z_2} z_1 z_2 D_N(z_1,z_2) + \frac{\rho}{N}\sum_{i=1}^{N} Z_{i,1}Z_{i,2} \xrightarrow{P} 0,\ j=1,2.$$

Therefore, given $(\tilde{I},\tilde{Z})$, we have

$$\begin{aligned}
\mathsf{Var}\left(\frac{A_{2*}}{\sqrt{N}}\middle|\tilde{I},\tilde{Z}\right) =& \sigma_x^2\cdot\left[\frac{\gamma_1^2}{\rho^2}\frac{\sum_{i=1}^N I_i Z_{i,1}^2}{N} + \frac{\gamma_2^2}{\rho^2}\frac{\sum_{i=1}^N I_i Z_{i,2}^2}{N} + \frac{\sigma_\epsilon^2}{\rho^2}\frac{\sum_{i=1}^N I_i}{N}\right.\\
&\left. + \frac{2\gamma_1\gamma_2}{\rho}\frac{\sum_{i=1}^N I_i Z_{i,1}Z_{i,2}}{N}\right] \to \frac{\sigma_x^2\sigma^2}{\rho},\\
\mathsf{Cov}\left(\frac{A_{1*}}{\sqrt{N}},\ \frac{A_{2*}}{\sqrt{N}}\middle|\tilde{I},\tilde{Z}\right) =& -\frac{\sigma_x^2}{\rho}\cdot\left[\frac{\gamma_1^2\sum_{i=1}^N I_i Z_{i,1}^2}{N} + \frac{\gamma_2^2\sum_{i=1}^N I_i Z_{i,2}^2}{N} + \frac{\sigma_\epsilon^2\sum_{i=1}^N I_i}{N}\right.\\
&\left. + \frac{2\gamma_1\gamma_2\sum_{i=1}^N I_i Z_{i,1}Z_{i,2}}{N}\right] \to -\sigma_x^2\sigma^2.
\end{aligned}$$

By the central limit theorem, given $(\tilde{I},\tilde{Z})$, $\frac{A_{1*}}{\sqrt{N}} + \frac{A_{2*}}{\sqrt{N}}$ is asymptotically normal with mean zero and variance $\frac{1-\rho}{\rho}\cdot\sigma_x^2\sigma^2$. Since the limiting distribution does not depend on $(\tilde{I},\tilde{Z})$, the convergence in distribution also holds unconditionally. So, by Slutsky's theorem, we get

$$\sqrt{N}A_* \xrightarrow{D} \mathcal{N}\left(0,\ \frac{\sigma^2}{\rho(1-\rho)\sigma_x^2}\right). \tag{3.20}$$

As in the proof of Theorem 3.3.1, we can conclude $\sqrt{N}B_* \xrightarrow{P} 0$, which together with Slutsky's theorem implies that

$$\sqrt{N}\left(\hat{\beta}_2 - \beta_2\right) \xrightarrow{D} \mathcal{N}\left(0,\ \frac{\sigma^2}{\rho(1-\rho)\sigma_x^2}\right).$$

It is straightfoward to get $\hat{\sigma}^2\boldsymbol{l}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{l}^t \ \to\ \frac{\sigma^2}{N\rho(1-\rho)\sigma_x^2} + o_p\left(\frac{1}{N}\right)$. Then under $H_0:\ \beta_2 = 0$, we have

$$T_* \xrightarrow{D} \mathcal{N}(0,\ 1).$$

Therefore, the hypothesis testing for the significance of the interaction term is still valid in terms of Type I error under the covariate-adaptive design, even when prognostic covariates are not used for statistical inference. □

Under $H_a: \ \beta_2 \neq 0$ with a sequence of local alternatives, i.e., we have $\beta_2 = \delta_{\beta_2}/\sqrt{N}$ for a fixed $\delta_{\beta_2} \neq 0$, the test statistic has the following asymptotic distribution

$$T_* \ \xrightarrow{D} \ \mathcal{N}\left(\frac{\delta_{\beta_2} \cdot \sigma_x \sqrt{\rho(1-\rho)}}{\sigma}, \ 1\right),$$

which finishes the proof of Theorem 3.3.2. □

Chapter 4

Statistical Inference under New CARA Designs

4.1 Introduction

Hu and Hu, *et al.* [20] stated that precision medicine is to use information associated with an individual patient systematically to select or optimize that patient's preventative and therapeutic care. Covariates of patients can be classified into two types according to their different roles played in clinical studies, which are prognostic covariates and predictive covariates. Correspondingly, prognostic covariates should be balanced to provide a simple treatment comparison, while predictive covariates may be used for the selection of suitable treatments.

A superior design of a clinical trial of precision medicine relies on whether patients' characteristics (covariates, especially predictive covariates) could be incorporated in the design protocol. In the literature, covariate-adaptive designs have been proposed to balance prognostic covariates, while covariate-adjusted response-adaptive designs deal with predictive covariates. Unfortunately, there is no design thus far that can simultaneously deal with both types of covariates. In this chapter, we will establish an innovative class of covariate-adjusted response-adaptive designs with incorporation of both prognostic and predictive covariates in their randomization schemes. Statistical

inference under the new procedure is studied under a linear model. We find out that the hypothesis testing to compare treatment effects is usually conservative under the new design if prognostic covariates used in the randomization procedure are excluded from the final analysis. Moreover, the test for the significance of the interaction term between the predictive covariate and treatment is not valid any more. It can be conservative or liberal, depending on the definition of target allocation proportions, unknown parameters in the underlying model and the predictive covariate.

This chapter is organized as follows. In Section 4.2, the general framework of the new CARA design is introduced. Similar hypothesis testings for comparing treatment effects and significance of covariates are conducted based on an underlying full model and a working model. Theoretical results for test statistics under both null and alternative hypotheses are provided in Section 4.3. Simulation studies are conducted in Section 4.4 to compare the finite sample properties of four randomization procedures, namely, complete randomization (CR), the covariate-adaptive design in Chapter 3 (CAR), the covariate-adjusted response-adaptive design in Chapter 2 (CARA) and the new design (NEW), in terms of imbalances, Type I error and power. Conclusions are summarized in Section 4.5, and major proof of theoretical results is shown in Section 4.6.

4.2 Framework

A clinical trial with two treatment arms is considered under the new CARA design. Let $(X_m, Z_{m,1}, Z_{m,2})$ be the covariate profile of the m-th subject. The *margin* and *stratum* defined by Z_1 and Z_2 are same as described in Section 3.2.1. The target allocation proportion depends on the covariate X. Therefore, (Z_1, Z_2) work as prognostic covariates, while X as a predictive covariate. Suppose the target allocation to treatment 1 is $\rho(X)$. We assume the same underlying full model (3.1) and working

model of inference (3.4) as in Section 3.2.

The new CARA design can be generally defined by

$$\psi_m = \mathbb{E}(I_m|\mathcal{I}_{m-1}, \mathcal{Y}_{m-1}, \mathcal{X}_m, \mathcal{Z}_m),$$

the conditional probabilities of assigning the m-th subject to treatment 1, conditioning on the treatment assignments, responses, predictive and prognostic covariates of the previous $m-1$ subjects, as well as both types of covariates of the current subject.

The allocation scheme of the new procedure is defined as follows:

1. The first $2m_0$ subjects are assigned to treatment 1 and 2 with restricted randomization (m_0 subjects in each treatment).

2. Suppose $(n-1)$ subjects have been assigned to treatments ($n > 2m_0$) and we observed their responses. The n-th subject has covariate $(X_n, Z_{n,1}, Z_{n,2})$, where $(Z_{n,1}, Z_{n,2})$ falls within stratum (z_1^*, z_2^*). Based on $(\boldsymbol{Y}_{n-1}, \boldsymbol{I}_{n-1}, \boldsymbol{X}_n)$, we estimate the target allocation $\hat{\rho}(X_n)$. (similar to the estimation of allocation probability under CARA designs in Section 2.2.1).

3. If the n-th subject were assigned to treatment 1,

   - let $D_n^{(1)}(X_n) = N_{n1}(X_n) - \hat{\rho}(X_n)N_n(X_n)$, where $N_n(X_n)$ is the total number of subjects with covariate X_n for the first n subjects. Further, $N_{n1}(X_n)$ is the number of subjects with covariate X_n in treatment 1.

   - similarly, let $D_n^{(1)}(j; z_j^*; X_n) = N_{n1}(j; z_j^*; X_n) - \hat{\rho}(X_n)N_n(j; z_j^*; X_n)$ and $D_n^{(1)}(z_1^*, z_2^*; X_n) = N_{n1}(z_1^*, z_2^*; X_n) - \hat{\rho}(X_n)N_n(z_1^*, z_2^*; X_n)$ be the differences between the actual and target number of subjects in treatment 1 on the margin $(j; z_j^*)$, $j = 1, 2$, and within the stratum (z_1^*, z_2^*), respectively.

4. Define an imbalance measure $Imb_n^{(1)}(X_n)$ by

$$Imb_n^{(1)}(X_n) = w_o \left[D_n^{(1)}(X_n)\right]^2 + \sum_{j=1}^{2} w_{m,j} \left[D_n^{(1)}(j; z_j^*; X_n)\right]^2 + w_s \left[D_n^{(1)}(z_1^*, z_2^*; X_n)\right]^2, \tag{4.1}$$

where $w_o + w_s + \sum_{j=1}^{2} w_{m,j} = 1$. If the n-th subject were assigned to treatment 2, we can define $Imb_n^{(2)}(X_n)$ in a similar fashion. Then, we assign the n-th subject to treatment 1 with probability based on Efron's biased coin function, which is similar with (3.3).

5. Repeat step 2 - step 4 until each subject is assigned to a treatment.

At the end of the trial, we can obtain estimated imbalances $\hat{D}_N(X) = N_1(X) - \hat{\rho}_N(X)N(X)$ and true imbalances $D_N(X) = N_1(X) - \rho(X)N(X)$ in treatment group 1 at overall, marginal and within-stratum levels, respectively. Here, $\hat{\rho}_N(X)$ is the estimated allocation proportion to treatment 1 from N samples, while $\rho(X)$ is the theoretical value.

Without considering the covariate X, the new design reduces to the covariate-adaptive design in Chapter 3; without prognostic covariates Z_1 and Z_2, the new design reduces to the CARA design in Chapter 2.

4.3 Theoretical Properties

Two hypothesis testings are considered in this section. Data are simulated from the underlying full model (3.1) and the hypothesis testings are conducted based on the working model of inference (3.4). The primary interest is still to compare treatment effects μ_1 and μ_2 using the hypothesis testing (2.5). Asymptotic properties of the test statistic are studied under both the null and alternative hypotheses in Theorem 4.3.1. We are also interested in the significance of the interaction term for the predictive covariate X and treatment assignment I, which indicates the different effects of X

on responses in different treatment groups. Then, the hypothesis testing (3.6) is conducted, and asymptotic properties of the corresponding test statistic are provided in Theorem 4.3.2.

THEOREM 4.3.1. Suppose that a new CARA design with adaptive allocation probabilities satisfies the following conditions:

(A) The estimated overall imbalance is bounded in probability, that is, $\hat{D}_N(X) = O_p(1)$ and the true overall imbalance $D_N(X)$ is of order $O_p(\sqrt{N})$;

(B) The estimated marginal imbalances for Z_1 and Z_2 are bounded in probability, that is, $\hat{D}_N(j; z_j; X) = O_p(1)$, and the true marginal imbalances are of order $O_p(\sqrt{N})$, that is, $D_N(j; z_j; X) = O_p(\sqrt{N}),\ j = 1, 2$;

(C) The estimated allocation proportion converges in probability to the true one, i.e., $\hat{\rho}_N(X) \xrightarrow{P} \rho(X)$. And $\sqrt{N}\left[\hat{\rho}_N(X) - \rho(X)\right]$ has an asymptotic normal distribution with mean zero and a finite variance.

Then, we have

(i) under $H_0 :\ \mu_1 - \mu_2 = 0$, the test statistic (3.5) has the following asymptotic distribution:

$$T = \frac{\boldsymbol{L}\hat{\boldsymbol{\beta}}}{\sqrt{\hat{\sigma}^2 \boldsymbol{L}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}^T}} \xrightarrow{D} \mathcal{N}\left(0,\ \frac{\sigma_\epsilon^2}{\sigma^2}\right), \tag{4.2}$$

where $\sigma^2 = \gamma_1^2\sigma_1^2 + \gamma_2^2\sigma_2^2 + \sigma_\epsilon^2$.

The test statistic for comparing treatment effects using the working model (3.4) under the new CARA design has the same asymptotic distribution with that under the covariate-adaptive design in Chapter 3. The definition of target allocation proportions, treatment effects and the predictive covariate X do not affect the performance of the test.

(ii) under $H_a :\ \mu_1 - \mu_2 \neq 0$, consider a sequence of local alternatives, that is,

$\mu_1 - \mu_2 = \delta/\sqrt{N}$ for some fixed $\delta \neq 0$, then the test statistic (3.5) has the following asymptotic distribution:

$$T = \frac{\boldsymbol{L}\hat{\boldsymbol{\beta}}}{\sqrt{\hat{\sigma}^2 \boldsymbol{L}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}^T}} \xrightarrow{D} \mathcal{N}\left(\frac{\delta}{\sigma} \cdot \sqrt{\frac{\rho_a(1-\rho_a)\cdot\rho_b(1-\rho_b)}{(1-p)\rho_a(1-\rho_a)+p\rho_b(1-\rho_b)}},\ \frac{\sigma_\epsilon^2}{\sigma^2}\right), \tag{4.3}$$

where $\rho_x = \rho(X = x),\ x = a, b$. Therefore, the limiting power can be affected by the definition of target allocation proportions, unknown parameters in the employed model and the predictive covariate X.

THEOREM 4.3.2. Suppose that the new CARA design with adaptive allocation probabilities satisfies all conditions in Theorem 4.3.1, then we have

(i) under $H_0 :\ \beta_2 = 0$,

$$T^* \xrightarrow{D} \mathcal{N}\left(0,\ \frac{\sigma_c^2 \cdot (1-p)^3}{\sigma_d^2}\right), \tag{4.4}$$

where

$$\begin{aligned}
&\sigma_d^2 = a^6 p^3 \sigma^2 \cdot \rho_a\rho_b(1-\rho_a)(1-\rho_b)\left[p\rho_a(1-\rho_a)+(1-p)\rho_b(1-\rho_b)\right],\\
&\rho_1 = p\rho_a + (1-p)\rho_b,\ \rho_2 = ap(\rho_a-\rho_b),\ \rho_3 = a^2\rho_a\cdot p + b^2\rho_b\cdot(1-p),\\
&c_{1*} = \rho_2\left[\rho_2^2 - (1-\rho_1)(\sigma_x^2-\rho_3)\right],\ c_{2*} = \rho_2(\rho_2^2-\rho_1\rho_3),\\
&c_{3*} = (1-\rho_1)(\rho_2^2-\rho_1\rho_3),\ c_{4*} = \rho_1(1-\rho_1)\sigma_x^2 - \rho_2^2,\\
&\sigma_c^2 = (c_{1*}^2 + c_{3*}^2\sigma_x^2)\cdot\sigma^2 + 2(c_{1*}c_{2*}\rho_1 + c_{1*}c_{4*}\rho_2 + c_{2*}c_{3*}\rho_2 + c_{3*}c_{4*}\rho_3)\cdot\sigma^2\\
&\qquad + c_{2*}^2\cdot\left\{\left[p\rho_a^2+(1-p)\rho_b^2\right]\cdot\sigma_z^2 + \rho_1\sigma_\epsilon^2\right\} + 2c_{2*}c_{4*}\cdot\left\{ap(\rho_a^2-\rho_b^2)\sigma_z^2 + \rho_2\sigma_\epsilon^2\right\}\\
&\qquad + c_{4*}^2\cdot\left\{\left[pa^2\rho_a^2+(1-p)b^2\rho_b^2\right]\cdot\sigma_z^2 + \rho_3\sigma_\epsilon^2\right\}.
\end{aligned}$$

When $\rho_a = \rho_b = \rho$, i.e., we have $\mu_1 = \mu_2$ and $\beta_2 = 0$, the variance of the asymptotic distribution for T^* can be simplified as $\sigma_\epsilon^2/\sigma^2$, resulting in a conservative test. However, when $\rho_a \neq \rho_b$, the performance of the test depends on

unknown parameters in the employed model, the definition of target allocation proportions and the predictive covariate X. This test can be conservative or liberal.

(ii) under $H_a: \beta_2 \neq 0$, consider a sequence of local alternatives, that is, $\beta_2 = \delta_{\beta_2}/\sqrt{N}$ for some fixed $\delta_{\beta_2} \neq 0$, then the test statistic (3.7) has the following asymptotic distribution:

$$T^* \xrightarrow{D} \mathcal{N}\left(\frac{\delta_{\beta_2}}{\sigma} \cdot \sqrt{\frac{a^2 p \cdot \rho_a \rho_b (1-\rho_a)(1-\rho_b)}{(1-p)\left[p\rho_a(1-\rho_a) + (1-p)\rho_b(1-\rho_b)\right]}}\ ,\ \frac{\sigma_c^2}{\sigma_d^2}\right). \quad (4.5)$$

Under new CARA designs, theoretical properties of hypothesis testings to compare two treatment effects and to test significance of the interaction term are obtained in Theorem 4.3.1 and Theorem 4.3.2, respectively. The predictive covariate is used for estimating allocation proportions and prognostic covariates are utilized for balancing purpose. Conditions of imbalances are assumed to derive asymptotic distributions of test statistics. Under these conditions, it has been shown that the test statistic for comparing treatment effect under the new CARA design has asymptotic same normal distribution with that under the covariate-adaptive design, and the test is usually conservative when prognostic covariates are omitted in the working model. However, the test for significance of the interaction term is not valid any more. Its performance is affected by unknown parameters in the full model (3.1), the definition of target allocation proportions and the predictive covariate. And, so is the limiting power of both hypothesis testings.

4.4 Simulation Study

Case 1: Checking imbalances. Simulations are conducted to explore imbalances at all three different levels and check the conditions for imbalances, on which the theoret-

ical properties in Section 4.3 are based, under four randomization designs, namely, complete randomization (CR), the covariate-adaptive design (CAR), the covariate-adjusted response-adaptive design (CARA) and the new design (NEW).

Data are generated from the underlying model (3.1) with $\mu_1 = \mu_2 = 1/2$, $\beta_1 = 1$, $\beta_2 = -1/2$, $\gamma_1 = \gamma_2 = 1/2$. All covariates Z_1, Z_2 and X are distributed independently with equal probability to take value 1 or -1. The random error ϵ follows a standard normal distribution. For the CARA design and the new design, we use the following optimal allocations defined by the standard normal CDF Φ:

$$\rho(X=1) = \frac{\Phi(\mathbb{E}(Y|I=1,X=1))}{\Phi(\mathbb{E}(Y|I=1,X=1)) + \Phi(\mathbb{E}(Y|I=0,X=1))} = 0.4741$$

and

$$\rho(X=-1) = \frac{\Phi(\mathbb{E}(Y|I=1,X=-1))}{\Phi(\mathbb{E}(Y|I=1,X=-1)) + \Phi(\mathbb{E}(Y|I=0,X=-1))} = 0.6184.$$

Therefore, the overall allocation proportion to treatment 1 under these two designs is $\rho = \mathbb{P}(X=1)\rho(X=1) + \mathbb{P}(X=-1)\rho(X=-1) = 0.5463$.

In the simulation, allocation proportions to treatment 1 are estimated adaptively under the CARA design and the new design. While for complete randomization (CR) and the covariate-adaptive design (CAR), the allocation proportions assigning to treatment 1 are fixed to be $1/2$, regardless of the value of X. After each subject is assigned to a treatment, estimated allocation proportions to treatment 1 can be obtained as shown in Table 4.1. We can clearly see that estimated allocation proportions for all procedures converge to their theoretical values, and only those under the CARA design (CARA) and the new design (NEW) depend on the value of X.
At the end of the trial, imbalances $D_N = N_1 - \rho N$ and $D_N(X=1) = N_1(X=1) - \rho(X=1)N(X=1)$ at three different levels are obtained using the theoretical

Table 4.1: Estimated allocation proportions to treatment 1 under the four procedures. Results are based on sample size $N = 1000$ and 5000 simulations.

Method	$\hat{\rho}(X = 1)$	$\hat{\rho}(X = -1)$
CR	0.5003	0.5000
CAR	0.5003	0.5000
CARA	0.4748	0.6164
NEW	0.4743	0.6176

target proportions ρ and $\rho(X = 1)$, respectively. By the symmetric characteristic of the designs, theoretical means of D_N and $D_N(X = 1)$ at each level is 0. Standard deviations of these two types of imbalances are reported in Table 4.2 based on 5000 simulations and sample size 200, 400, 1000. For simplicity, only the results of 2 strata and 2 margins are presented. The first column shows the overall imbalances, the second and third columns are marginal imbalances and the last two are within-stratum imbalances.

Results in Table 4.2 indicate that the imbalances in treatment 1, i.e., D_N, at all three different levels under complete randomization (CR), the CARA design (CARA) and the new design (NEW) are all of order $O_p\left(\sqrt{N}\right)$. The new design (NEW) has relatively smaller standard deviations compared with those under the CARA design (CARA) and complete randomization (CR). Under the covariate-adaptive design (CAR), D_N at all three levels are bounded in probability, i.e. $D_N = O_p(1)$, hence their standard deviations do not increase as the sample size becomes larger.

The imbalances in treatment 1 when $X = 1$, i.e., $D_N(X = 1)$, for three levels are all of order $O_p\left(\sqrt{N}\right)$ under these four designs. Those under complete randomization (CR) and the CARA design (CARA) have similar performance, and those under the covariate-adaptive design (CAR) have smaller standard deviations since prognostic covariates are balanced. However, balancing prognostic covariates using different

Table 4.2: Standard deviations of imbalances in treatment 1 D_N (standard deviations of imbalances in treatment 1 when $X = 1$, i.e., $D_N(X = 1)$) for different levels. Results are based on sample size $N = 200,\ 400,\ 1000$ and 5000 simulations.

Method	N	Overall	$Z_1 = 1$	$Z_2 = -1$	$(1, 1)$	$(-1, -1)$
CR	200	7.07 (5.03)	5.05 (3.56)	4.97 (3.54)	3.51 (2.51)	3.51 (2.50)
	400	10.11 (7.18)	7.07 (5.03)	7.03 (5.01)	4.97 (3.56)	4.94 (3.51)
	1000	15.75 (11.32)	11.12 (7.96)	11.22 (7.95)	7.77 (5.48)	7.82 (5.56)
CAR	200	0.73 (3.58)	0.73 (2.53)	0.73 (2.57)	0.74 (1.82)	0.74 (1.82)
	400	0.72 (4.98)	0.73 (3.55)	0.74 (3.52)	0.74 (2.52)	0.75 (2.49)
	1000	0.72 (7.95)	0.74 (5.59)	0.73 (5.71)	0.74 (3.92)	0.73 (4.01)
CARA	200	8.67 (5.08)	5.88 (3.64)	5.39 (3.51)	3.95 (2.56)	3.55 (2.45)
	400	13.50 (7.26)	8.59 (5.08)	8.27 (4.94)	5.60 (3.61)	5.41 (3.53)
	1000	22.53 (11.65)	13.91 (8.21)	13.60 (8.08)	8.95 (5.67)	8.76 (5.65)
NEW	200	5.34 (1.42)	3.18 (0.98)	2.84 (0.94)	2.09 (0.82)	1.86 (0.81)
	400	7.78 (1.84)	4.22 (1.16)	4.08 (1.10)	2.50 (0.86)	2.40 (0.85)
	1000	11.80 (2.73)	6.16 (1.52)	6.14 (1.52)	3.37 (0.98)	3.40 (1.00)

allocation proportions based on the values of X under the new design (NEW) results in significant reduction on standard deviations of imbalances compared with the covariate-adaptive design (CAR).

Case 2: Testing treatment effects. We carry out simulations to study Type I error and power of the hypothesis testing for comparing two treatment effects under the four procedures. To study Type I error, we use the same parameter settings in Case 1 of Section 4.4. The hypothesis tests include the test using the working model (3.4) ($lm(X)$), the bootstrap test for $lm(X)$ (BS) and the test using the full model (3.1) ($lm(X, Z)$). The significance level is $\alpha = 0.05$ and sample size N is $200, 400$ or 1000. The results are based on 1000 simulations and in each simulation, 200 bootstrap samples are randomly generated with replacement.

Several conclusions can be drawn from Table 4.3: (1) Type I error using the working

model (3.4) ($lm(X)$) is affected by omitting prognostic covariates under adaptive designs (CAR, CARA, NEW), while a valid test can still be obtained under complete randomization (CR); (2) with the working model (3.4), tests $lm(X)$ are conservative under both the covariate-adaptive design (CAR) and the new design (NEW), but it can be liberal under the CARA design (CARA), which agree with theoretical results; (3) after adjustment on the variance of test statistic, bootstrap tests (BS) can restore nominal error rate under these three adaptive designs (CAR, CARA, NEW); (4) tests using the full model (3.1) ($lm(X, Z)$) have Type I error close to 5% for all four designs.

Table 4.3: Simulated Type I error of three hypothesis tests in % under four randomization procedures based on sample sizes $N = 200, 400, 1000$ and 1000 simulations.

Method	SS	$lm(X)$	BS	$lm(X, Z)$
CR	200	5.0	-	5.4
	400	5.3	-	5.0
	1000	5.0	-	4.8
CAR	200	2.3	4.7	5.2
	400	1.7	4.8	4.9
	1000	1.3	5.2	5.0
CARA	200	5.8	5.1	5.1
	400	5.4	5.2	5.0
	1000	5.5	5.2	4.9
NEW	200	2.2	5.4	5.3
	400	1.9	5.4	5.1
	1000	1.5	4.8	4.7

Case 3: Power comparison. Next we compare the power for the three hypothesis testings, namely $lm(X)$, BS and $lm(X, Z)$, under the four procedures. The same model in Case 1 of Section 4.4 is applied, but there exists a difference between two treatment effects, i.e., we have $\mu_1 - \mu_2 \neq 0$.

All the results of power and average responses are given in Figure 4.1, from which several conclusions can be made. First, it is shown in the left plot that tests using the working model (3.4) ($lm(X)$) are less powerful than those using the full model (3.1) ($lm(X, Z)$) under all these four methods. Second, both the covariate-adaptive design (CAR) and the new design (NEW) have smaller power for $lm(X)$ compared with complete randomization (CR) when the difference between μ_1 and μ_2 is relatively small due to the conservativeness, but they have larger power then the difference becomes larger. Third, from the middle plot, we can see that bootstrap tests for $lm(X)$ (BS) under the covariate-adaptive design (CAR) and the NEW design (NEW) significantly improve the power, and perform very similar with those using the full model ($lm(X, Z)$). However, the bootstrap test under the CARA design (CARA) still has smaller power than that using the full model ($lm(X, Z)$). This may be because that $\hat{\mu}_1 - \hat{\mu}_2$ under the working model (3.4) is not an unbiased estimator of $\mu_1 - \mu_2$ anymore, and correction of the variance of the test statistic can not solve this problem. Last, the CARA design (CARA) and the new design (NEW) show their advantages on assigning more subjects to a better treatment and obtaining significantly higher average responses in the right plot, compared with complete randomization (CR) and the covariate-adaptive design (CAR), by incorporating responses and predictive covariates in the randomization procedure.

Case 4: Significance of interaction term. The interaction term between the predictive covariate X and the treatment I contributes to the difference between two allocation proportions $\rho(X = 1)$ and $\rho(X = -1)$. Therefore, we study the hypothesis testings for the significance of the interaction term based on both the working model (3.4) and the full model (3.1), i.e., $lm(X)$ and $lm(X, Z)$, under the new design (NEW) and complete randomization (CR). Sample size $N = 500$ and the significance level is $\alpha = 0.05$. Four different scenarios of parameters are used to study Type I error.

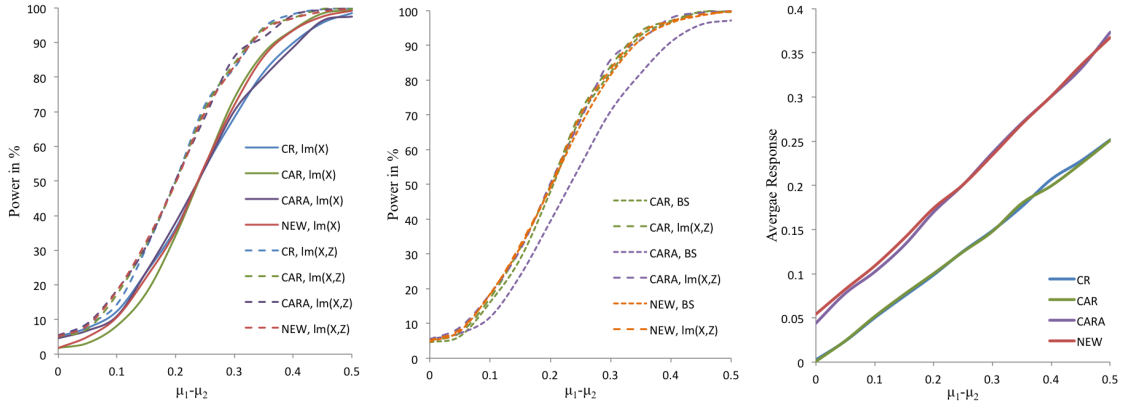

Figure 4.1: Simulated power of comparing treatment effects for $lm(X)$ and $lm(X, Z)$ (left), simulated power for $lm(X, Z)$ and BS (middle), and average responses (right) under four procedures. Simulations are based on sample size $N = 400$, bootstrap samples $B = 200$ and 1000 runs.

Scenario	μ_1	μ_2	β_1	γ_1	γ_2	p
S_1	0	0	1	0.5	0.5	0.5
S_2	1	0	0.5	0.5	0.5	0.5
S_3	1	0	0.5	1	-1	0.5
S_4	0.5	0	1	0.1	0.1	0.7

Table 4.4: Simulated Type I error for significance of interaction term under the new design (NEW) and complete randomization (CR) in %. Simulations are based on 10,000 runs.

Scenario	NEW		CR	
	$lm(X)$	$lm(X, Z)$	$lm(X)$	$lm(X, Z)$
S_1	1.58	4.73	5.06	4.93
S_2	1.84	4.79	4.96	4.85
S_3	0.85	4.89	5.09	5.05
S_4	6.26	4.83	5.20	5.31

From Table 4.4, it can be seen that the test for significance of the interaction term is not valid under the new design (NEW) when prognostic covariates are omitted in the final analysis. The performance of the test under the working model ($lm(X)$) can be conservative or liberal depending on the unknown parameters in the employed model,

the definition of target allocation proportion and the predictive covariate X.

Table 4.5: Simulated power for significance of interaction term using the working model ($lm(X)$) and the full model ($lm(X, Z)$) under the new design (NEW) and complete randomization (CR) in %. Simulations are based on 1000 runs and sample size 500.

	NEW		CR	
β_2	$lm(X)$	$lm(X, Z)$	$lm(X)$	$lm(X, Z)$
0.00	2.0	5.0	4.8	5.0
0.05	3.6	8.0	7.7	8.3
0.10	10.0	20.4	14.9	20.6
0.15	22.9	37.1	27.4	37.9
0.20	42.7	59.8	44.1	59.9
0.25	63.6	77.9	62.2	79.9
0.30	81.5	91.6	78.5	91.9
0.35	93.9	97.2	88.5	96.8
0.40	97.7	99.5	95.0	99.9
0.45	99.4	99.9	98.7	100.0
0.50	99.8	100.0	99.5	100.0

Results in Table 4.5 show the power comparison under the new design (NEW) and complete randomization (CR) using $\mu_1 = 1/2,\ \mu_2 = 0,\ \beta_1 = 1/2,\ \gamma_1 = \gamma_2 = 1/2$ in the underlying model and $p = 1/2$ in the distribution of X. The power of testing β_2 based on the full model ($lm(X, Z)$) is usually more powerful than that based on the working model ($lm(X)$) under both procedures. The new design (NEW) can achieve higher power than complete randomization (CR) when β_2 gets larger under the working model ($lm(X)$), while they have very similar performance under the full model ($lm(X, Z)$).

4.5 Conclusion

New CARA designs incorporate both types of covariates in the randomization procedure. They can balance treatment allocations with respect to prognostic covariates and assign subjects based on their predictive covariates. From simulation results, we can see that although the standard deviations of imbalances in treatment 1, conditional on the value of X for all three levels, increase as the sample size gets larger under the new CARA designs, they have been significantly reduced compared with those under other designs. Type I error is affected by the incorporation of covariates under new CARA designs when prognostic covariates, balanced in the randomization procedure, are omitted in the final analysis. Simulation results agree with theoretical properties that the test is usually conservative. One possible solution for this problem is to use a bootstrap test, and it can restore nominal error rate. New CARA designs also show their advantage on assigning more subjects to a better treatment by targeting different allocation proportions based on the value of the predictive covariate X. Average responses of subjects can be greatly improved under the new designs, but still allowing high power when the difference between treatment effects is relatively large.

Theoretical results in this chapter are based on several assumptions, among which all covariates are assumed to be discrete. A continuous-discrete transformation is required when covariates are continuous, which will lose important information of covariates. By now, there is no good way to deal with continuous predictive covariates directly. Similar problems are studied with the discrete assumption for prognostic covariates relaxed in Chapter 5. However, in this case, we can only apply the new CARA design using theoretical target allocation proportions in the randomization scheme. Asymptotic properties under the new designs with continuous prognostic covariates and adaptively estimated allocation proportions in randomization still remain un-

known.

4.6 Appendix: Proof of Theorems

Proof of Theorem 4.3.1.

According to the definition of estimated overall imbalances, we have

$$\hat{D}_N(X) = N_1(X) - \hat{\rho}(X) \cdot N(X).$$

Let $\hat{\rho}_x = \hat{\rho}_N(X = x),\ x = a, b$. Note that $\sum_{i=1}^N I_i = N_1(a) + N_1(b)$ and $\sum_{i=1}^N I_i X_i = a \cdot N_1(a) + b \cdot N_1(b)$. Then, it follows that

$$\begin{aligned}
\sum_{i=1}^N I_i &= \hat{D}_N(a) + \hat{D}_N(b) + \hat{\rho}_a \cdot N(a) + \hat{\rho}_b \cdot N(b), \\
\sum_{i=1}^N I_i X_i &= a \cdot \hat{D}_N(a) + b \cdot \hat{D}_N(b) + a\hat{\rho}_a \cdot N(a) + b\hat{\rho}_b \cdot N(b).
\end{aligned}$$

By weak law of large numbers, we have $N(a)/N \xrightarrow{P} p$, and $N(b)/N \xrightarrow{P} 1 - p$. It implies from conditions (A) and (C) that

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^N I_i &\xrightarrow{P} p \cdot \rho_a + (1-p) \cdot \rho_b = \rho_1, \\
\frac{1}{N}\sum_{i=1}^N I_i X_i &\xrightarrow{P} ap \cdot \rho_a + b(1-p) \cdot \rho_b = ap \cdot (\rho_a - \rho_b) = \rho_2.
\end{aligned}$$

Similar techniques can be applied to obtain

$$\frac{1}{N}\sum_{i=1}^N I_i X_i^2 \xrightarrow{P} a^2\rho_a \cdot p + b^2\rho_b \cdot (1-p) = \rho_3.$$

Based on the definition of estimated marginal imbalances, together with conditions

(B) and (C), we get

$$\frac{\sum_{i=1}^{N} I_i Z_{i,j}}{N} = \frac{\sum_{z_j} [z_j \cdot N_1(j; z_j)]}{N} = \sum_x \left[\hat{\rho}_x \cdot \frac{\sum_{i=1}^{N} Z_{i,j} \mathbf{1}(X_i = x)}{N} \right] + O_p\left(\frac{1}{N}\right),$$

where $x = a, b$ and $j = 1, 2$. By weak law of large numbers and the independence between X and (Z_1, Z_2), we have

$$\frac{1}{N} \sum_{i=1}^{N} I_i Z_{i,j} \xrightarrow{P} 0, \ j = 1, 2.$$

The proof for $\sum_{i=1}^{N} I_i X_i Z_{i,j}/N, \ j = 1, 2$ are similar.

First, we look at the denominator, which is the model-based variance estimator. By weak law of large numbers, the independence between covariates and the convergences above, we can get $\frac{\boldsymbol{X}^t \boldsymbol{Z}}{N} \xrightarrow{P} \mathbf{0}$ and

$$\frac{\boldsymbol{X}^t \boldsymbol{X}}{N} \xrightarrow{P} \begin{pmatrix} \rho_1 & 0 & \rho_2 & \rho_2 \\ 0 & 1-\rho_1 & -\rho_2 & 0 \\ \rho_2 & -\rho_2 & \sigma_x^2 & \rho_3 \\ \rho_2 & 0 & \rho_3 & \rho_3 \end{pmatrix} = \boldsymbol{M}, \ |\boldsymbol{M}| = \frac{a^4 p^2 \cdot \rho_a \rho_b (1-\rho_a)(1-\rho_b)}{(1-p)^2},$$

where $\rho_1 = p \cdot \rho_a + (1-p) \cdot \rho_b, \ \rho_2 = ap \cdot (\rho_a - \rho_b), \ \rho_3 = a^2 \rho_a \cdot p + b^2 \rho_b \cdot (1-p)$. Further, by the independence between (I, X) and ϵ, we have $\frac{\boldsymbol{X}^t \boldsymbol{\epsilon}}{N} \xrightarrow{P} \mathbf{0}$.

Because of the consistency of $\hat{\boldsymbol{\beta}}$, the second and third terms of $\hat{\sigma}^2$ in (3.15) converge to 0 in probability. It follows from the law of large numbers and independence

between $(Z_1,\, Z_2)$ and ϵ that $\hat{\sigma}^2 \xrightarrow{P} \gamma_1^2\sigma_1^2 + \gamma_2^2\sigma_2^2 + \sigma_\epsilon^2 = \sigma^2$. Hence, we have

$$\hat{\sigma}^2 \boldsymbol{L}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{L}^t = \frac{\hat{\sigma}^2}{N}\, \boldsymbol{L}\left(\frac{\boldsymbol{X}^t\boldsymbol{X}}{N}\right)^{-1}\boldsymbol{L}^t = \frac{\sigma^2}{N}\cdot\left[\frac{p}{\rho_a(1-\rho_a)} + \frac{1-p}{\rho_b(1-\rho_b)}\right] + o_p\left(\frac{1}{N}\right).$$

Next, we look at the numerator. After some matrix calculations for A defined in (3.12), we have

$$\begin{aligned} A =& \frac{1}{N|\boldsymbol{M}|}\cdot\left\{(\sigma_x^2-\rho_3)(\rho_2^2-\rho_1\rho_3)\cdot\left(\gamma_1\sum_{i=1}^N Z_{i,1} + \gamma_2\sum_{i=1}^N Z_{i,2} + \sum_{i=1}^N \epsilon_i\right)\right. \\ &+ \left[(\rho_3-\rho_2^2)\cdot\sigma_x^2 - \rho_3^2\right]\cdot\left(\gamma_1\sum_{i=1}^N I_i Z_{i,1} + \gamma_2\sum_{i=1}^N I_i Z_{i,2} + \sum_{i=1}^N I_i\epsilon_i\right) \\ &+ \rho_2(\rho_2^2-\rho_1\rho_3)\cdot\left(\gamma_1\sum_{i=1}^N X_i Z_{i,1} + \gamma_2\sum_{i=1}^N X_i Z_{i,2} + \sum_{i=1}^N X_i\epsilon_i\right) \\ &\left.+ \rho_2\left[\rho_3 - (1-\rho_1)\cdot\sigma_x^2\right]\cdot\left(\gamma_1\sum_{i=1}^N I_i X_i Z_{i,1} + \gamma_2\sum_{i=1}^N I_i X_i Z_{i,2} + \sum_{i=1}^N I_i X_i\epsilon_i\right)\right\}. \end{aligned} \tag{4.6}$$

Let $A_1 = \gamma_1\sum_{i=1}^N Z_{i,1} + \gamma_2\sum_{i=1}^N Z_{i,2} + \sum_{i=1}^N \epsilon_i$ and $A_3 = \sum_{i=1}^N X_i\epsilon_i + \gamma_1\sum_{i=1}^N X_i Z_{i,1} + \gamma_2\sum_{i=1}^N X_i Z_{i,2}$. Define $\tilde{X} = \{X_1,\ldots,X_N\}$ and $\tilde{Z} = \{Z_{i,j}, i = 1,\ldots,N, j = 1,2\}$. Then by weak law of large numbers, given $(\tilde{X},\tilde{Z})$, we can easily get

$$\mathsf{Var}\left(\frac{A_1}{\sqrt{N}}\middle|\,\tilde{Z}\right) \to \sigma^2,\ \mathsf{Var}\left(\frac{A_3}{\sqrt{N}}\middle|\,\tilde{X},\tilde{Z}\right) \to \sigma_x^2\sigma^2,\ \mathsf{Cov}\left(\frac{A_1}{\sqrt{N}},\ \frac{A_3}{\sqrt{N}}\middle|\,\tilde{X},\tilde{Z}\right) \to 0.$$

$$\begin{aligned} \text{Let} \qquad S_j =& \sum_{z_j}\sum_{x}\left[z_j D_N(j;z_j;x)\right] \\ =& \sum_{z_j}\sum_{x}\left[z_j\hat{D}_N(j;z_j;x)\right] + \sum_{x}\left[\sqrt{N}(\hat{\rho}_x-\rho_x)\cdot\frac{\sum_{i=1}^N Z_{i,j}\mathbf{1}(X_i=x)}{\sqrt{N}}\right], \end{aligned}$$

where $x = a, b$ and $j = 1, 2$. We know from conditions (B) and (C) that $\hat{D}_N(j;z_j;x) = O_p(1)$ and $\sqrt{N}(\hat{\rho}_x - \rho_x) = O_p(1)$. Since $X,\ Z_1$ and Z_2 only take finite number of

values, we have

$$\sum_{z_j}\sum_{x}\left[z_j\hat{D}_N(x;z_j)\right]=O_p(1).$$

Further by weak law of large numbers and the independence between X and (Z_1,Z_2), we have $\frac{\sum_{i=1}^N Z_{i,j}\mathbf{1}(X_i=x)}{N}\xrightarrow{P}0$ for $j=1,2$. Hence, we get $S_j=o_p\left(\sqrt{N}\right)$, $j=1,2$. Since we have

$$\sum_{i=1}^N I_iZ_{i,j}=S_j+\sum_x\left[\rho_x\cdot\sum_{i=1}^N Z_{i,j}\mathbf{1}(X_i=x)\right],$$

we can write

$$\begin{aligned}&\gamma_1\sum_{i=1}^N I_iZ_{i,1}+\gamma_2\sum_{i=1}^N I_iZ_{i,2}+\sum_{i=1}^N I_i\epsilon_i\\=&\sum_j\sum_x\left[\rho_x\cdot\sum_{i=1}^N\gamma_jZ_{i,j}\mathbf{1}(X_i=x)\right]+\sum_{i=1}^N I_i\epsilon_i+o_p\left(\sqrt{N}\right).\end{aligned}$$

Let $A_2=\sum_j\sum_x\left[\rho_x\cdot\sum_{i=1}^N\gamma_jZ_{i,j}\mathbf{1}(X_i=x)\right]+\sum_{i=1}^N I_i\epsilon_i$. Define $\tilde{I}=\{I_1,\ldots,I_N\}$. Then, given $(\tilde{X},\tilde{Z},\tilde{I})$, by weak law of large numbers and the convergence of $\sum_{i=1}^N I_i/N$ above, we have

$$\mathsf{Var}\left(\frac{A_2}{\sqrt{N}}\middle|\tilde{X},\tilde{Z},\tilde{I}\right)\to\left[p\rho_a^2+(1-p)\rho_b^2\right]\sigma_z^2+\rho_1\sigma_\epsilon^2,\text{ where }\sigma_z^2=\gamma_1^2\sigma_1^2+\gamma_2^2\sigma_2^2.$$

Similarly, let

$$\begin{aligned}W_j=&\sum_{z_j}\sum_x\left[x\cdot z_jD_N(j;z_j;x)\right]=\sum_{z_j}\sum_x\left[x\cdot z_j\hat{D}_N(j;z_j;x)\right]\\&+\sum_x\left[x\cdot\sqrt{N}(\hat{\rho}_x-\rho_x)\cdot\frac{\sum_{i=1}^N Z_{i,j}\mathbf{1}(X_i=x)}{\sqrt{N}}\right],\end{aligned}$$

where $x=a,b$ and $j=1,2$. Then, we have

$$\sum_{i=1}^N I_iX_iZ_{i,j}=W_j+\sum_x\left[x\rho_x\cdot\sum_{i=1}^N Z_{i,j}\mathbf{1}(X_i=x)\right].$$

Since we have $W_j = o_p(\sqrt{N})$, we can write

$$\begin{aligned}
&\gamma_1 \sum_{i=1}^{N} I_i X_i Z_{i,1} + \gamma_2 \sum_{i=1}^{N} I_i X_i Z_{i,2} + \sum_{i=1}^{N} I_i X_i \epsilon_i \\
=& \sum_j \sum_x \left[x\rho_x \cdot \sum_{i=1}^{N} \gamma_j Z_{i,j} \mathbf{1}(X_i = x) \right] + \sum_{i=1}^{N} I_i X_i \epsilon_i + o_p\left(\sqrt{N}\right).
\end{aligned}$$

Let $A_4 = \sum_j \sum_x \left[x\rho_x \cdot \sum_{i=1}^{N} \gamma_j Z_{i,j} \mathbf{1}(X_i = x) \right] + \sum_{i=1}^{N} I_i X_i \epsilon_i$. By weak law of large numbers and the convergence of $\sum_{i=1}^{N} I_i X_i^2 / N$, we get

$$\mathsf{Var}\left(\frac{A_4}{\sqrt{N}} \middle| \tilde{X}, \tilde{Z}, \tilde{I} \right) \to \left[p \cdot a^2 \rho_a^2 + (1-p) \cdot b^2 \rho_b^2 \right] \sigma_z^2 + \rho_3 \sigma_\epsilon^2.$$

Covariances between these terms can be obtained in the same way. Given $(\tilde{X}, \tilde{Z}, \tilde{I})$, we have

$$\begin{aligned}
\mathsf{Cov}\left(\frac{A_1}{\sqrt{N}}, \frac{A_2}{\sqrt{N}} \middle| \tilde{X}, \tilde{Z}, \tilde{I} \right) &\to \rho_1 \sigma^2, \mathsf{Cov}\left(\frac{A_1}{\sqrt{N}}, \frac{A_4}{\sqrt{N}} \middle| \tilde{X}, \tilde{Z}, \tilde{I} \right) \to \rho_2 \sigma^2, \\
\mathsf{Cov}\left(\frac{A_2}{\sqrt{N}}, \frac{A_3}{\sqrt{N}} \middle| \tilde{X}, \tilde{Z}, \tilde{I} \right) &\to \rho_2 \sigma^2, \mathsf{Cov}\left(\frac{A_3}{\sqrt{N}}, \frac{A_4}{\sqrt{N}} \middle| \tilde{X}, \tilde{Z}, \tilde{I} \right) \to \rho_3 \sigma^2, \\
\mathsf{Cov}\left(\frac{A_2}{\sqrt{N}}, \frac{A_4}{\sqrt{N}} \middle| \tilde{X}, \tilde{Z}, \tilde{I} \right) &\to ap(\rho_a^2 - \rho_b^2)\sigma_z^2 + \rho_2 \sigma_\epsilon^2.
\end{aligned}$$

From (4.6) we know that

$$\sqrt{N} A = \frac{1}{|\boldsymbol{M}|} \left(c_1 \cdot \frac{A_1}{\sqrt{N}} + c_2 \cdot \frac{A_2}{\sqrt{N}} + c_3 \cdot \frac{A_3}{\sqrt{N}} + c_4 \cdot \frac{A_4}{\sqrt{N}} \right) + o_p(1),$$

where

$$\begin{aligned}
c_1 &= (\sigma_x^2 - \rho_3)(\rho_2^2 - \rho_1 \rho_3), \quad c_2 = (\rho_3 - \rho_2^2)\sigma_x^2 - \rho_3^2, \\
c_3 &= \rho_2(\rho_2^2 - \rho_1 \rho_3), \quad c_4 = \rho_2 \left[\rho_3 - (1 - \rho_1)\sigma_x^2 \right].
\end{aligned}$$

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

$$\frac{1}{|\boldsymbol{M}|}\left(c_1 \cdot \frac{A_1}{\sqrt{N}} + c_2 \cdot \frac{A_2}{\sqrt{N}} + c_3 \cdot \frac{A_3}{\sqrt{N}} + c_4 \cdot \frac{A_4}{\sqrt{N}}\right)$$

is asymptotically normal with mean zero and variance $\sigma_\epsilon^2 \cdot \left[\frac{p}{\rho_a(1-\rho_a)} + \frac{1-p}{\rho_b(1-\rho_b)}\right]$. The limiting distribution is independent of $(\tilde{X}, \tilde{Z}, \tilde{I})$, so it also holds unconditionally. Therefore, by Slutsky's theorem, we have

$$\sqrt{N}A \xrightarrow{D} \mathcal{N}\left(0,\ \sigma_\epsilon^2 \cdot \left[\frac{p}{\rho_a(1-\rho_a)} + \frac{1-p}{\rho_b(1-\rho_b)}\right]\right).$$

We already proved that $\left(\frac{\boldsymbol{X}^t\boldsymbol{X}}{N}\right)^{-1} - \boldsymbol{M}^{-1} = o_p(1)$ and $\frac{\boldsymbol{X}^t\boldsymbol{Z}\boldsymbol{\gamma}}{\sqrt{N}} + \frac{\boldsymbol{X}^t\boldsymbol{\epsilon}}{\sqrt{N}} = O_p(1)$. Hence, $\sqrt{N}B \xrightarrow{P} 0$, where B is defined as (3.13). Together with Slutsky's theorem, it follows that

$$\sqrt{N}\left[(\hat{\mu}_1 - \hat{\mu}_2) - (\mu_1 - \mu_2)\right] \xrightarrow{D} \mathcal{N}\left(0,\ \sigma_\epsilon^2 \cdot \left[\frac{p}{\rho_a(1-\rho_a)} + \frac{1-p}{\rho_b(1-\rho_b)}\right]\right).$$

Therefore, under $H_0:\ \mu_1 - \mu_2 = 0$, we have

$$T \xrightarrow{D} \mathcal{N}\left(0,\ \frac{\sigma_\epsilon^2}{\sigma^2}\right),\ \text{where } \sigma^2 = \gamma_1^2\sigma_1^2 + \gamma_2^2\sigma_2^2 + \sigma_\epsilon^2. \tag{4.7}$$

The test statistic under the new CARA design using the working model (3.4) has the same asymptotic normal distribution with that under the covariate-adaptive design in Chapter 3. The incorporation of the predictive covariate X in the randomization scheme does not affect the performance of the test when there is no difference between two treatment effects. □

Similarly, under $H_a:\ \mu_1 - \mu_2 \neq 0$ with a sequence of local alternatives, i.e., we have

$\mu_1 - \mu_2 = \delta/\sqrt{N}$ for a fixed $\delta \neq 0$, we get

$$T \xrightarrow{D} \mathcal{N}\left(\frac{\delta}{\sigma} \cdot \sqrt{\frac{\rho_a(1-\rho_a)\cdot\rho_b(1-\rho_b)}{(1-p)\rho_a(1-\rho_a)+p\rho_b(1-\rho_b)}}\ ,\ \frac{\sigma_\epsilon^2}{\sigma^2}\right). \tag{4.8}$$

The limiting power is affected by unknown parameters in the employed model, the definition of target allocation proportion and the predictive covariate X. This finishes the proof of Theorem 4.3.1. □

Proof of Theorem 4.3.2.

It is straightforward to get

$$\hat{\sigma}^2 \boldsymbol{l}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{l}^t = \frac{\sigma^2}{N}\cdot\frac{1-p}{a^2p}\cdot\left[\frac{1-p}{\rho_a(1-\rho_a)}+\frac{p}{\rho_b(1-\rho_b)}\right]+o_p\left(\frac{1}{N}\right).$$

After some matrix calculation for A_* defined in (3.18), we have

$$\sqrt{N}A_* = \frac{1}{|\boldsymbol{M}|}\left(c_{1*}\cdot\frac{A_1}{\sqrt{N}}+c_{2*}\cdot\frac{A_2}{\sqrt{N}}+c_{3*}\cdot\frac{A_3}{\sqrt{N}}+c_{4*}\cdot\frac{A_4}{\sqrt{N}}\right)+o_p(1),$$

where

$$c_{1*} = \rho_2\left[\rho_2^2-(1-\rho_1)(\sigma_x^2-\rho_3)\right],\quad c_{2*} = \rho_2(\rho_2^2-\rho_1\rho_3),$$
$$c_{3*} = (1-\rho_1)(\rho_2^2-\rho_1\rho_3),\quad c_{4*} = \rho_1(1-\rho_1)\sigma_x^2-\rho_2^2.$$

By the central limit theorem and Slutsky's theorem, under $H_0:\ \beta_2 = 0$, we have

$$T^* \xrightarrow{D} \mathcal{N}\left(0,\ \frac{\sigma_c^2\cdot(1-p)^3}{\sigma_d^2}\right),$$

where

$$\sigma_d^2 = a^6 p^3 \sigma^2 \cdot \rho_a \rho_b (1-\rho_a)(1-\rho_b) \left[p\rho_a(1-\rho_a) + (1-p)\rho_b(1-\rho_b) \right],$$

$$\rho_1 = p\rho_a + (1-p)\rho_b,\ \rho_2 = ap(\rho_a - \rho_b),\ \rho_3 = a^2\rho_a \cdot p + b^2\rho_b \cdot (1-p),$$

$$\begin{aligned}\sigma_c^2 = {} & (c_{1*}^2 + c_{3*}^2\sigma_x^2) \cdot \sigma^2 + 2(c_{1*}c_{2*}\rho_1 + c_{1*}c_{4*}\rho_2 + c_{2*}c_{3*}\rho_2 + c_{3*}c_{4*}\rho_3) \cdot \sigma^2 \\ & + c_{2*}^2 \cdot \left\{ \left[p\rho_a^2 + (1-p)\rho_b^2 \right] \cdot \sigma_z^2 + \rho_1\sigma_\epsilon^2 \right\} + 2c_{2*}c_{4*} \cdot \left\{ ap(\rho_a^2 - \rho_b^2)\sigma_z^2 + \rho_2\sigma_\epsilon^2 \right\} \\ & + c_{4*}^2 \cdot \left\{ \left[pa^2\rho_a^2 + (1-p)b^2\rho_b^2 \right] \cdot \sigma_z^2 + \rho_3\sigma_\epsilon^2 \right\}.\end{aligned}$$

When $\rho_a = \rho_b = \rho$, i.e., $\mu_1 = \mu_2$ and $\beta_2 = 0$, the variance of the asymptotic distribution for T^* can be simplified as $\sigma_\epsilon^2/\sigma^2$, resulting in a conservative test. However, when $\rho_a \neq \rho_b$, the performance of test depends on parameter settings in the employed model, the definition of target allocation proportion and the predictive covariate of X. The test can be conservative or liberal. □

Under $H_a:\ \beta_2 \neq 0$ with a sequence of local alternatives, i.e., $\beta_2 = \delta_{\beta_2}/\sqrt{N}$ for a fixed $\delta_{\beta_2} \neq 0$, we have

$$T^* \xrightarrow{D} \mathcal{N}\left(\frac{\delta_{\beta_2}}{\sigma} \cdot \sqrt{\frac{a^2 p \cdot \rho_a\rho_b(1-\rho_a)(1-\rho_b)}{(1-p)\left[p\rho_a(1-\rho_a) + (1-p)\rho_b(1-\rho_b)\right]}},\ \frac{\sigma_c^2 \cdot (1-p)^3}{\sigma_d^2} \right).$$

The proof of Theorem 4.3.2 is finished. □

Chapter 5

Statistical Inference for Continuous Prognostic Covariates

5.1 Introduction

Covariates of patients used in a clinical trial can be categorical (e.g. sex, investigative site) and/or continuous (e.g. age, weight, blood pressure, body mass index). However, prognostic covariates incorporated in the underlying model of the covariate-adaptive design in Chapter 3 and the new design in Chapter 4 are restricted to be discrete variables. Therefore, if a continuous covariate is to be used in the randomization procedure, discretization should be performed to split its range into several distinct intervals, within each of which a subcategory is defined. Usually the choice of discretization for a continuous covariate is from a clinical perspective or depends on the model assumption of the covariate distribution. In the study of adaptive stratification in an acute stroke clinical trial by Weir and Lees [53], twelve variables were selected to be potentially included. Five continuous variables were categorized before stratification. Based on the historical information of covariate distributions, age and mean arterial pressure were separated into quintiles, and plasma glucose level was first split by 8mmol/l and then the above values were further subdivided into quartiles; delay to trial enrollment from stroke onset and Glasgow coma scale on admission were split into two subcategories by critical values from clinical experience, respectively.

In this chapter, we study the asymptotic properties of statistical inference for compar-

ing treatment effects in the case that prognostic covariates Z_1 and Z_2 are continuous variables. In Section 5.2, theoretical results of hypothesis testings are given under the extended covariate-adaptive design and the new CARA design using theoretical target allocation proportions in randomization. Simulations in Section 5.3 are conducted for the evaluation of Type I error and power comparison with complete randomization. Conclusions and technical proof of theorems are given in Section 5.3 and 5.4, respectively.

5.2 Theoretical Properties

The same underlying full mode (3.1) and working model of inference (3.4) are assumed. In the case that the two prognostic covariates Z_1 and Z_2 are continuous, the covariate-adaptive design and the new CARA design are applied with respect to the discrete variables $d_1(Z_1)$ and $d_2(Z_2)$, where $d_1(\cdot)$ and $d_2(\cdot)$ are discrete functions. Define $\delta_{i,j} = Z_{i,j} - \mathbb{E}\left[Z_{i,j}|d_j(Z_{i,j})\right],\ i = 1, \ldots, N,\ j = 1, 2$. For simplicity, we still use z_1 and z_2 to denote the values of $d_1(Z_1)$ and $d_2(Z_2)$. Note that for the new CARA design, we use the theoretical allocation proportion $\rho(X)$ to calculate imbalances at different levels in the randomization procedure, instead of using the adaptively estimated one $\hat{\rho}(X)$.

To compare the two treatment effects, the hypothesis testing (2.5) is conducted based on the working model (3.4), and data are simulated from the underlying full model (3.1). Asymptotic properties of the test statistic under the covariate-adaptive design are studied under both the null and alternative hypotheses and provided in Theorem 5.2.1, while those under the new CARA design in Theorem 5.2.2.

THEOREM 5.2.1. Suppose that the covariate-adaptive design in Section 3.2.1 satisfies the same two conditions in Theorem 3.3.1. Then, we have

(i) under $H_0: \ \mu_1 - \mu_2 = 0$, the test statistic (3.5) has the following asymptotic distribution:

$$T = \frac{\boldsymbol{L}\hat{\boldsymbol{\beta}}}{\sqrt{\hat{\sigma}^2 \boldsymbol{L}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}^T}} \ \xrightarrow{D} \ \mathcal{N}\left(0, \ \frac{\sigma_\delta^2}{\sigma^2}\right), \tag{5.1}$$

where

$$\sigma^2 = \gamma_1^2\sigma_1^2 + \gamma_2^2\sigma_2^2 + \sigma_\epsilon^2, \ \sigma_\delta^2 = \gamma_1^2\sigma_{\delta,1}^2 + \gamma_2^2\sigma_{\delta,2}^2 + \sigma_\epsilon^2, \ \sigma_{\delta,j}^2 = \mathbb{E}\left[\mathsf{Var}\left(\delta_{i,j}|d_j(Z_{i,j})\right)\right].$$

Note that for $\gamma_j \neq 0$, $\mathsf{Var}(Z_j) = \sigma_j^2 > \mathbb{E}\left[\mathsf{Var}\left(\delta_{i,j}|d_j(Z_{i,j})\right)\right] = \sigma_{\delta,j}^2$. Therefore, $\sigma_\delta^2/\sigma^2 < 1$, still resulting in a conservative test. But the direct incorporation of continuous prognostic covariates in the full model (3.1) can improve the conservativeness compared with the discrete case. And the variance of the limiting distribution depends on the discretization functions $d_j(\cdot)$.

(ii) under $H_a: \ \mu_1 - \mu_2 \neq 0$, consider a sequence of local alternatives, that is, $\mu_1 - \mu_2 = \delta/\sqrt{N}$ for some fixed $\delta \neq 0$, then the test statistic (3.5) has the following asymptotic distribution:

$$T = \frac{\boldsymbol{L}\hat{\boldsymbol{\beta}}}{\sqrt{\hat{\sigma}^2 \boldsymbol{L}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}^T}} \ \xrightarrow{D} \ \mathcal{N}\left(\frac{\delta \cdot \sqrt{\rho(1-\rho)}}{\sigma}, \ \frac{\sigma_\delta^2}{\sigma^2}\right). \tag{5.2}$$

Thus, the limiting power reaches its maximum when $\rho = 1/2$.

THEOREM 5.2.2. Suppose that the new CARA design in Section 3.2.1 using theoretical allocation proportions in the randomization scheme satisfies the following two conditions:

(A) The overall imbalances are bounded in probability, that is, $D_N(X) = O_p(1)$;

(B) The marginal imbalances for Z_1 and Z_2 are bounded in probability, that is, $D_N(j; z_j; X) = O_p(1)$, where $j = 1, 2$;

Then, we have

(i) under $H_0: \ \mu_1 - \mu_2 = 0$, the test statistic (3.5) has the following asymptotic distribution:

$$T = \frac{\boldsymbol{L}\hat{\boldsymbol{\beta}}}{\sqrt{\hat{\sigma}^2 \boldsymbol{L}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}^T}} \xrightarrow{D} \mathcal{N}\left(0, \ \frac{\sigma_\delta^2}{\sigma^2}\right), \tag{5.3}$$

where

$$\sigma^2 = \gamma_1^2\sigma_1^2 + \gamma_2^2\sigma_2^2 + \sigma_\epsilon^2, \ \sigma_\delta^2 = \gamma_1^2\sigma_{\delta,1}^2 + \gamma_2^2\sigma_{\delta,2}^2 + \sigma_\epsilon^2, \ \sigma_{\delta,j}^2 = \mathbb{E}\left[\mathsf{Var}\left(\delta_{i,j}|d_j(Z_{i,j})\right)\right].$$

Therefore, the test statistic under the new design has the same asymptotic distribution with that under the covariate-adaptive design. The test is conservative in terms of smaller Type I error, and its performance depends on the discretization functions $d_j(\cdot)$ and the two prognostic covariates.

(ii) under $H_a: \ \mu_1 - \mu_2 \neq 0$, consider a sequence of local alternatives, that is, $\mu_1 - \mu_2 = \delta/\sqrt{N}$ for some fixed $\delta \neq 0$, then the test statistic (3.5) has the following asymptotic distribution:

$$T = \frac{\boldsymbol{L}\hat{\boldsymbol{\beta}}}{\sqrt{\hat{\sigma}^2 \boldsymbol{L}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}^T}} \xrightarrow{D} \mathcal{N}\left(\frac{\delta}{\sigma} \cdot \sqrt{\frac{\rho_a(1-\rho_a)\rho_b(1-\rho_b)}{p\rho_a(1-\rho_a) + (1-p)\rho_b(1-\rho_b)}}, \ \frac{\sigma_\delta^2}{\sigma^2}\right). \tag{5.4}$$

The limiting power is affected by the predictive covariate X and the definition of target allocation proportions.

Continuous prognostic covariates are included in the full model (3.1) directly, and utilized in randomization after discretization. Asymptotic distributions of test statistics for comparing treatment effects are obtained in Theorem 5.2.1 under the covariate-adaptive design and in Theorem 5.2.2 under the new design using fixed allocation proportions in randomization. In the proof, we show that tests using the working model (3.4) are usually conservative in terms of smaller Type I error under both designs. Variances of the limiting distributions depend on the choice of discretiza-

tion methods, and become larger compared with discrete cases in Chapter 3 and Chapter 4, resulting in less conservative tests. Conclusions for limiting powers under both designs are similar with those in discrete cases. Under the covariate-adaptive design, it is affected by the target allocation proportion; while under the new design, it is affected by the two target allocation proportions as well as the predictive covariate.

5.3 Simulation Study

Case 1: Testing treatment effects. Similar simulations are conducted to study Type I error and power of hypothesis testings for comparing two treatment effects under the covariate-adaptive design and the new CARA design using theoretical target allocation proportions, respectively. Data are generated from the underlying model (3.1). The predictive covariate X and random error have same distributions as in Case 1 of Section 3.4. The two prognostic covariates Z_1 and Z_2 are distributed independently from $Uniform(-1, 1)$. The discrete variables $d_j(Z_j)$, where

$$d_j(Z_j) = \begin{cases} 1, & \text{if } Z_j \geq 0; \\ -1, & \text{if } Z_j < 0, \end{cases} \qquad j = 1, 2,$$

are used in the randomization schemes of the two adaptive designs. Two target allocation proportions used under the new design are defined same as those in Case 1 of Section 4.4.

To study Type I error, we assume two treatment effects are same, i.e., we have $\mu_1 = \mu_2$. Same simulation scenarios and hypothesis testings as in Case 1 of Section 3.4 are used for the covariate-adaptive design, and results are presented in Table 5.1. The conclusions obtained from Table 5.1 are similar to those from Table 3.1 and

consistent with theoretical results in Theorem 5.2.1. Incorporation of continuous prognostic covariates directly in the underlying model will increase the variance of the test statistic compared with discrete ones, resulting in less conservative tests under the covariate-adaptive designs. Both tests under complete randomization (CR) and the test based on the full model under the covariate-adaptive design (CAR, $lm(X, Z)$) can still achieve nominal Type I error.

Table 5.1: Simulated Type I error for comparing treatment effects with continuous prognostic covariates under the covariate-adaptive design (CAR) and complete randomization (CR) in %. Simulations are based on 10,000 runs and sample size $N = 500$.

Scenario	CAR		CR	
	$lm(X)$	$lm(X, Z)$	$lm(X)$	$lm(X, Z)$
S_1	3.72	4.58	5.24	4.98
S_2	3.86	4.74	5.40	5.22
S_3	3.80	5.18	5.30	5.08
S_4	3.76	4.70	-	-
S_5	1.86	5.02	5.12	4.96
S_6	3.82	5.02	4.52	4.88

Similar simulation studies are conducted for the new CARA design using theoretical target allocation proportions under the following scenarios of parameters. Several conclusions can be drawn from Table 5.2. First, none of treatment effects, the predictive covariate X and its related parameters and the two theoretical target allocation proportions have effects on the performance of the test under the new CARA design (NEW); larger effects of prognostic covariates on responses lead to a more conservative test. Second, under the null hypothesis, the test under the new CARA design (NEW) with theoretical $\rho(X)$'s performs asymptotically same with that under the covariate-adaptive design (CAR), which agrees with theoretical results. Third, tests under complete randomization (CR) and the test based on the full model under

the new design (NEW, $lm(X, Z)$) have Type I error close to 5%.

Scenario	μ_1	μ_2	β_1	β_2	γ_1	γ_2	p	ρ_a	ρ_b
C_1	0	0	1	0	0.5	0.5	0.5	0.5	0.5
C_2	0.5	0.5	1	-0.5	0.5	0.5	0.3	0.5	0.5
C_3	0.5	0.5	1	-0.5	0.5	0.5	0.5	0.4741	0.6184
C_4	1	1	0.5	0.5	1	1	0.5	0.5115	0.4197
C_5	1	1	0.5	0.5	0.5	-0.5	0.5	0.5115	0.4197

Table 5.2: Simulated Type I error for comparing treatment effects with continuous prognostic covariates under the new CARA design using theoretical target allocation proportions (NEW) and complete randomization (CR) in %. Simulations are based on 10,000 runs.

Scenario	NEW		CR	
	$lm(X)$	$lm(X, Z)$	$lm(X)$	$lm(X, Z)$
C_1	3.80	5.32	5.36	4.48
C_2	3.86	4.84	4.58	4.56
C_3	3.78	5.04	4.90	5.20
C_4	1.94	4.98	4.96	4.96
C_5	3.66	4.92	4.98	4.64

Case 2: Power comparison. First, we compare the power for the two hypothesis testings, $lm(X)$ and $lm(X, Z)$, under the covariate-adaptive design with allocation proportion $\rho = 1/2$ (CAR1), $\rho = 2/3$ (CAR2) and complete randomization (CR). Results of power comparison are given in Figure 5.1. For the new CARA design, according to the theoretical results, if we use $p = 1/2$ for the distribution of X, then the power will reach its maximum when $\rho_a = \rho_b = 1/2$, and the design reduces to CAR1. Therefore, we compare the power and average response under the new design (NEW) with target allocation proportions defined in Case 1 of Section 4.4, CAR1 and complete randomization (CR) in Figure 5.2. Same pattern with those of the discrete case in Chapter 3 and 4 can be observed.

In both left plots of Figure 5.1 and Figure 5.2, the tests under the full model ($lm(X, Z)$) is more powerful than those under the working model ($lm(X)$) for all procedures. When the hypothesis testing is conducted using the full model ($lm(X, Z)$), all procedures have similar power performance. However, when prognostic covariates are excluded from the working model of inference, the power of each procedure can also be compared between complete randomization. The tests under adaptive designs (CAR1, CAR2, NEW) have smaller power than that under complete randomization (CR) when the difference between two treatment effects is relatively small due to conservativeness, but have larger power when the difference becomes larger. Further among these three adaptive designs utilizing covariates, the covariate-adaptive design with $\rho = 1/2$ (CAR1) is most powerful compared with that with $\rho = 2/3$ (CAR2) and the new design (NEW).

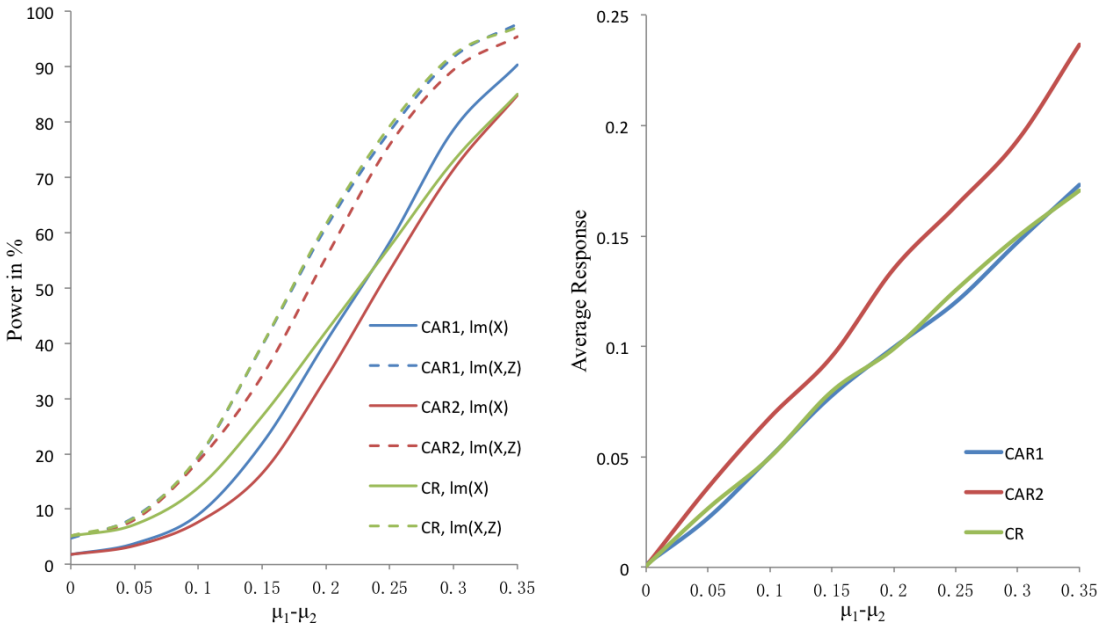

Figure 5.1: Simulated power of comparing treatment effects (left) and average response (right) comparing the covariate-adaptive design with $\rho = 1/2$ (CAR1), $\rho = 2/3$ (CAR2) and complete randomization (CR). Simulation based on sample size $N = 500$ and 1,000 runs.

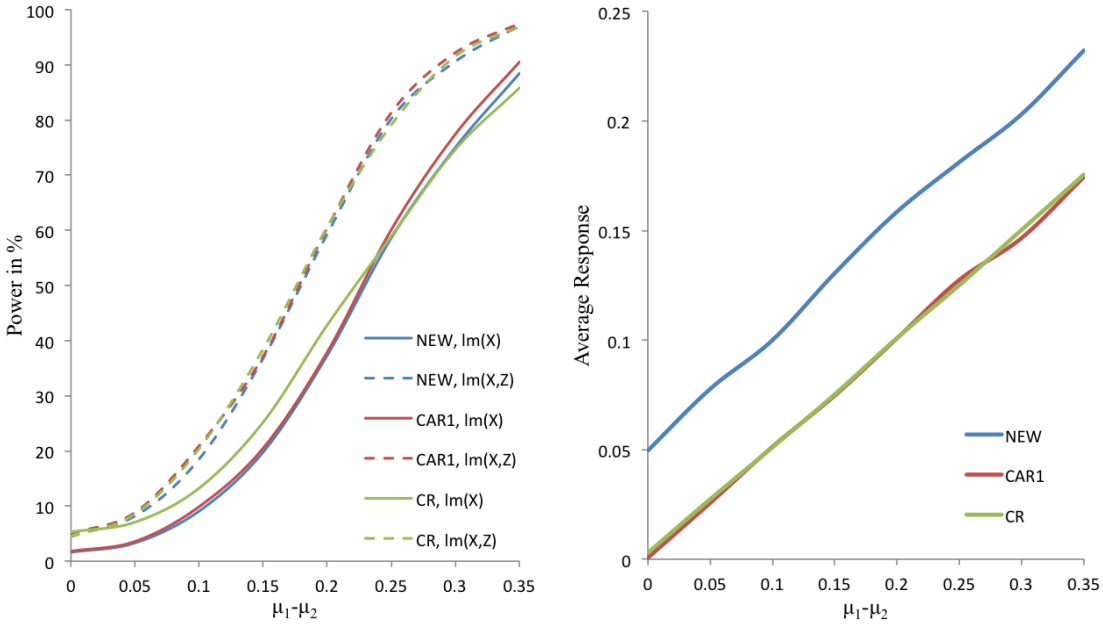

Figure 5.2: Simulated power of comparing treatment effects (left) and average response (right) comparing the new procedure (NEW), the covariate-adaptive design with $\rho = 1/2$ (CAR1) and complete randomization (CR). Simulation based on sample size $N = 500$ and 1,000 runs.

5.4 Conclusion

In this chapter, we proposed theoretical properties of statistical inference under covariate-adaptive designs and new CARA designs with continuous prognostic covariates. Note that under the new designs, allocation proportions used in the randomization procedure are fixed to be theoretical ones instead of adaptively estimated ones. Two conditions about imbalances are used to obtain theoretical results under these two designs, which are (A) the overall imbalance is bounded in probability; (B) marginal imbalances are bounded in probability. These conditions are relatively mild and can be satisfied by a broad family of adaptive designs utilizing covariates. The asymptotic distributions of test statistics are derived under both the null and alternative hypotheses, from which we can conclude that the hypothesis testing for comparing treatment effects under both designs is conservative in terms of smaller Type I error if prognostic covariates are not incorporated in analysis. This conclusion

is similar to the results of discrete cases in Chapter 3 and Chapter 4. Therefore, Type I error will be affect by omitting prognostic covariates in the final analysis regardless of whether they are discrete or continuous.

5.5 Appendix: Proof of Theorems

Proof of Theorem 5.2.1.

Define $\tilde{Z} = \{d_j(Z_{i,j}),\ i = 1, \ldots, N,\ j = 1, 2\}$ and $\tilde{I} = \{I_i,\ i = 1, \ldots, N\}$. Then, by definition and weak law of large numbers, $\mathbb{E}\left(\delta_{i,j}|\tilde{Z}\right) = 0$ and

$$\frac{1}{N}\mathsf{Var}\left(\sum_{i=1}^{N}\delta_{i,j}\middle|\tilde{Z}\right) = \frac{1}{N}\sum_{i=1}^{N}\mathsf{Var}\left(\delta_{i,j}|\tilde{Z}\right) \to \sigma^2_{\delta,j},$$

where $\sigma^2_{\delta,j} = \mathbb{E}\left[\mathsf{Var}\left(\delta_{i,j}|d_j(Z_{i,j})\right)\right],\ j = 1, 2.$

After some matrix calculations for (3.12) based on the condition that marginal imbalances are $O_p(1)$, we have

$$\begin{aligned} A =& \frac{1}{N}\left\{\gamma_1\left[\frac{\sum_{i=1}^{N} I_i\delta_{i,1}}{\rho(1-\rho)} - \frac{\sum_{i=1}^{N}\delta_{i,1}}{1-\rho}\right] + \gamma_2\left[\frac{\sum_{i=1}^{N} I_i\delta_{i,2}}{\rho(1-\rho)} - \frac{\sum_{i=1}^{N}\delta_{i,2}}{1-\rho}\right]\right. \\ &\left. + \left[\frac{\sum_{i=1}^{N} I_i\epsilon_i}{\rho(1-\rho)} - \frac{\sum_{i=1}^{N}\epsilon_i}{1-\rho}\right]\right\} + o_p\left(\frac{1}{\sqrt{N}}\right) \\ =& \frac{1}{N}(A_1 + A_2 + A_3) + o_p\left(\frac{1}{\sqrt{N}}\right). \end{aligned}$$

Given $\tilde{Z}$, $\delta_{i,j}$ and $\tilde{I}$ are independent. Since $d_j(Z_{i,j})$'s take finite discrete values and marginal imbalances are $O_p(1)$, it follows from weak law of large numbers that

$$\mathbb{E}\left(\frac{\sum_{i=1}^{N} I_i\delta_{i,j}}{N}\middle|\tilde{I}, \tilde{Z}\right) = \frac{1}{N}\sum_{i=1}^{N} I_i\mathbb{E}\left(\delta_{i,j}|\tilde{Z}\right) = 0,$$

$$\mathsf{Var}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} I_i\delta_{i,j}\middle|\tilde{I},\tilde{Z}\right) = \frac{\sum_{i=1}^{N}\left[I_i\mathsf{Var}\left(\delta_{i,j}|\tilde{Z}\right)\right]}{N} \to \ \rho\sigma_{\delta,j}^2.$$

In the same way, we can get

$$\mathsf{Cov}\left(\frac{\sum_{i=1}^{N} I_i\delta_{i,j}}{\sqrt{N}},\ \frac{\sum_{i=1}^{N}\delta_{i,j}}{\sqrt{N}}\middle|\tilde{I},\tilde{Z}\right) \to \ \rho\sigma_{\delta,j}^2 \text{ and } \mathsf{Cov}\left(\frac{\sum_{i=1}^{N} I_i\delta_{i,1}}{\sqrt{N}},\ \frac{\sum_{i=1}^{N} I_i\delta_{i,2}}{\sqrt{N}}\middle|\tilde{I},\tilde{Z}\right) = 0.$$

The proof for

$$\mathsf{Cov}\left(\frac{\sum_{i=1}^{N} I_i\delta_{i,1}}{\sqrt{N}},\ \frac{\sum_{i=1}^{N}\delta_{i,2}}{\sqrt{N}}\middle|\tilde{I},\tilde{Z}\right) = \mathsf{Cov}\left(\frac{\sum_{i=1}^{N}\delta_{i,1}}{\sqrt{N}},\ \frac{\sum_{i=1}^{N} I_i\delta_{i,2}}{\sqrt{N}}\middle|\tilde{I},\tilde{Z}\right) = 0$$

are similar. Therefore, we have

$$\mathsf{Var}\left(\frac{A_j}{\sqrt{N}}\middle|\tilde{I},\tilde{Z}\right) \to \frac{\gamma_j^2\sigma_{\delta,j}^2}{\rho(1-\rho)} \text{ and } \mathsf{Cov}\left(\frac{A_1}{\sqrt{N}},\ \frac{A_2}{\sqrt{N}}\middle|\tilde{I},\tilde{Z}\right) = 0.$$

By the independence between ϵ and (δ, I) given $\tilde{Z}$ and the convergence of $\sum_{i=1}^{N} I_i/N$, it is straightforward that $\mathsf{Var}\left(\frac{A_3}{\sqrt{N}}\middle|\tilde{I},\tilde{Z}\right) \to \frac{\sigma_\epsilon^2}{\rho(1-\rho)}$ and

$$\mathsf{Cov}\left(\frac{A_1}{\sqrt{N}},\ \frac{A_3}{\sqrt{N}}\middle|\tilde{I},\tilde{Z}\right) = \mathsf{Cov}\left(\frac{A_2}{\sqrt{N}},\ \frac{A_3}{\sqrt{N}}\middle|\tilde{I},\tilde{Z}\right) = 0.$$

Therefore, by the central limit theorem, given $(\tilde{I},\tilde{Z})$, $(A_1+A_2+A_3)/\sqrt{N}$ is asymptotically normal with mean zero and variance $\frac{\gamma_1^2\sigma_{\delta,1}^2+\gamma_2^2\sigma_{\delta,2}^2+\sigma_\epsilon^2}{\rho(1-\rho)}$. Since the limiting distribution is independent of $(\tilde{I},\tilde{Z})$, it also holds unconditionally. Hence, by Slutsky's theorem, we get

$$\sqrt{N}A \ \xrightarrow{D}\ \mathcal{N}\left(0,\ \frac{\gamma_1^2\sigma_{\delta,1}^2+\gamma_2^2\sigma_{\delta,2}^2+\sigma_\epsilon^2}{\rho(1-\rho)}\right).$$

The conclusion that $\sqrt{N}B \xrightarrow{P} 0$, where B is defined same as (3.13), can be easily

obtained. Together with Slutsky's theorem, it implies that

$$\sqrt{N}\left[(\hat{\mu}_1 - \hat{\mu}_2) - (\mu_1 - \mu_2)\right] \xrightarrow{D} \mathcal{N}\left(0,\ \frac{\gamma_1^2\sigma_{\delta,1}^2 + \gamma_2^2\sigma_{\delta,2}^2 + \sigma_\epsilon^2}{\rho(1-\rho)}\right).$$

The proof for the convergence of denominator does not change from the discrete case. Therefore under $H_0:\ \mu_1 - \mu_2 = 0$, we have

$$T \xrightarrow{D} \mathcal{N}\left(0,\ \frac{\sigma_\delta^2}{\sigma^2}\right),\ \text{where } \sigma_\delta^2 = \gamma_1^2\sigma_{\delta,1}^2 + \gamma_2^2\sigma_{\delta,2}^2 + \sigma_\epsilon^2 \text{ and } \sigma^2 = \gamma_1^2\sigma_1^2 + \gamma_2^2\sigma_2^2 + \sigma_\epsilon^2.$$

This completes the proof of Theorem 5.2.1. □

Proof of Theorem 5.2.2.

Denote the two target allocation proportions used in the randomization of the new CARA design as $\rho_a = \rho(X = a)$ and $\rho_b = \rho(X = b)$. Define $\tilde{Z} = \{d_j(Z_{i,j}),\ i = 1, \ldots, N,\ j = 1, 2\}$, $\tilde{X} = \{X_i,\ i = 1, \ldots, N\}$ and $\tilde{I} = \{I_i,\ i = 1, \ldots, N\}$. After some matrix calculations for (4.6), we have

$$\begin{aligned} A =& \frac{1}{N|\boldsymbol{M}|} \cdot \left\{ c_1 \cdot \left(\gamma_1 \sum_{i=1}^N Z_{i,1} + \gamma_2 \sum_{i=1}^N Z_{i,2} + \sum_{i=1}^N \epsilon_i \right) \right. \\ &+ c_2 \cdot \left(\gamma_1 \sum_{i=1}^N I_i Z_{i,1} + \gamma_2 \sum_{i=1}^N I_i Z_{i,2} + \sum_{i=1}^N I_i \epsilon_i \right) \\ &+ c_3 \cdot \left(\gamma_1 \sum_{i=1}^N X_i Z_{i,1} + \gamma_2 \sum_{i=1}^N X_i Z_{i,2} + \sum_{i=1}^N X_i \epsilon_i \right) \\ &\left. + c_4 \cdot \left(\gamma_1 \sum_{i=1}^N I_i X_i Z_{i,1} + \gamma_2 \sum_{i=1}^N I_i X_i Z_{i,2} + \sum_{i=1}^N I_i X_i \epsilon_i \right) \right\}, \end{aligned}$$

where $c_1 - c_4$ are corresponding coefficients in (4.6).

Let $A_1 = \gamma_1 \sum_{i=1}^N Z_{i,1} + \gamma_2 \sum_{i=1}^N Z_{i,2} + \sum_{i=1}^N \epsilon_i$ and $A_3 = \sum_{i=1}^N X_i\epsilon_i + \gamma_1 \sum_{i=1}^N X_i Z_{i,1} +$

$\gamma_2 \sum_{i=1}^N X_i Z_{i,2}$. It is obvious that

$$\mathsf{Var}\left(\frac{A_1}{\sqrt{N}}\middle|\tilde{Z}\right) \to \sigma^2,\ \mathsf{Var}\left(\frac{A_3}{\sqrt{N}}\middle|\tilde{X},\tilde{Z}\right) \to \sigma_x^2\sigma^2 \text{ and } \mathsf{Cov}\left(\frac{A_1}{\sqrt{N}},\ \frac{A_3}{\sqrt{N}}\middle|\tilde{Z},\tilde{X}\right) \to 0,$$

by weak law of large numbers. Also, we have

$$\begin{aligned}\frac{\sum_{i=1}^N I_i Z_{i,j}}{\sqrt{N}} =& \frac{1}{\sqrt{N}}\left[\sum_{i=1}^N I_i\delta_{i,j} + \sum_{i=1}^N I_i \mathbb{E}(Z_{i,j}|d_j(Z_{i,j}))\right] \\ =& \frac{\sum_{i=1}^N I_i\delta_{i,j}}{\sqrt{N}} + \frac{\sum_x\left[\rho_x \sum_{i=1}^N \mathbb{E}(Z_{i,j}|d_j(Z_{i,j}))\mathbf{1}(X_i=x)\right]}{\sqrt{N}} + o_p(1),\end{aligned}$$

where $x = a,\ b$.

Let $\frac{A_2}{\sqrt{N}} = \sum_j \gamma_j \left[\frac{\sum_{i=1}^N I_i\delta_{i,j}}{\sqrt{N}} + \frac{\sum_x[\rho_x \sum_{i=1}^N \mathbb{E}(Z_{i,j}|d_j(Z_{i,j}))\mathbf{1}(X_i=x)]}{\sqrt{N}}\right] + \frac{\sum_{i=1}^N I_i\epsilon_i}{\sqrt{N}}$.
Next, we look at the conditional variance of $A_2/\sqrt{N}$.

Similar to the continuous case of the covariate-adaptive design with allocation proportion ρ, we have $\mathsf{Var}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N I_i\delta_{i,j}\middle|\tilde{I},\tilde{Z}\right) \to \rho_1\sigma^2_{\delta,j}$. It is easy to get from weak law of large numbers that

$$\begin{aligned}\mathsf{Var}\left[\frac{\rho_a \sum_{i=1}^N \mathbb{E}(Z_{i,j}|d_j(Z_{i,j}))\mathbf{1}(X_i=a)}{\sqrt{N}}\middle|\tilde{Z},\tilde{X}\right] &\to p\rho_a^2\tau_j^2, \\ \mathsf{Var}\left[\frac{\rho_b \sum_{i=1}^N \mathbb{E}(Z_{i,j}|d_j(Z_{i,j}))\mathbf{1}(X_i=b)}{\sqrt{N}}\middle|\tilde{Z},\tilde{X}\right] &\to (1-p)\rho_b^2\tau_j^2,\end{aligned}$$

where $\tau_j^2 = \mathsf{Var}\left[\mathbb{E}(Z_{i,j}|d_j(Z_{i,j}))\right]$.

Because of the independence between $\delta_{i,j}$ and $\tilde{I}$ given $\tilde{Z}$, we get

$$\mathsf{Cov}\left(\frac{\sum_{i=1}^N I_i\delta_{i,j}}{\sqrt{N}},\ \frac{\rho_x}{\sqrt{N}}\sum_{i=1}^N \mathbb{E}(Z_{i,j}|d_j(Z_{i,j}))\mathbf{1}(X_i=x)\middle|\tilde{I},\tilde{Z},\tilde{X}\right) = 0.$$

The proof of other conditional covariances between terms of $A_2/\sqrt{N}$ is similar. Then,

we have

$$\mathsf{Var}\left(\frac{A_2}{\sqrt{N}}\middle|\tilde{I},\tilde{Z},\tilde{X}\right)\to\ \rho_1\sigma_\delta^2+\left[p\rho_a^2+(1-p)\rho_b^2\right]\sigma_{z*}^2,$$

where $\rho_1=p\rho_a+(1-p)\rho_b$, $\sigma_\delta^2=\gamma_1^2\sigma_{\delta,1}^2+\gamma_2^2\sigma_{\delta,2}^2+\sigma_\epsilon^2$ and $\sigma_{z*}^2=\gamma_1^2\tau_1^2+\gamma_2^2\tau_2^2$. Similarly, we have

$$\frac{\sum_{i=1}^N I_iX_iZ_{i,j}}{\sqrt{N}}=\sum_x\left[x\rho_x\cdot\frac{\sum_{i=1}^N\mathbb{E}(Z_{i,j}|d_j(Z_{i,j}))\mathbf{1}(X_i=x)}{\sqrt{N}}\right]+\frac{\sum_{i=1}^N I_iX_i\delta_{i,j}}{\sqrt{N}}+o_p(1),$$

and let

$$\frac{A_4}{\sqrt{N}}=\sum_j\gamma_j\left\{\sum_x\left[x\rho_x\cdot\frac{\sum_{i=1}^N\mathbb{E}(Z_{i,j}|d_j(Z_{i,j}))\mathbf{1}(X_i=x)}{\sqrt{N}}\right]+\frac{\sum_{i=1}^N I_iX_i\delta_{i,j}}{\sqrt{N}}\right\}+\frac{\sum_{i=1}^N I_iX_i\epsilon_i}{\sqrt{N}},$$

where $x=a,\ b$ and $j=1,2$. The proof for the conditional variance of $A_4/\sqrt{N}$ is in the same fashion.

Like the discrete case of the new CARA design, by weak law of large numbers, we have

$$\mathsf{Var}\left[\frac{a\rho_a}{\sqrt{N}}\cdot\sum_{i=1}^N\mathbb{E}(Z_{i,j}|d_j(Z_{i,j}))\mathbf{1}(X_i=a)\middle|\tilde{Z},\tilde{X}\right]\to pa^2\rho_a^2\tau_j^2,$$
$$\mathsf{Var}\left[\frac{b\rho_b}{\sqrt{N}}\cdot\sum_{i=1}^N\mathbb{E}(Z_{i,j}|d_j(Z_{i,j}))\mathbf{1}(X_i=b)\middle|\tilde{Z},\tilde{X}\right]\to(1-p)b^2\rho_b^2\tau_j^2.$$

By the independence of $\delta_{i,j}$ and $(\tilde{I},X)$ given $\tilde{Z}$ and the condition that marginal imbalances are $O_p(1)$, we have

$$\mathsf{Var}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N I_iX_i\delta_{i,j}\middle|\tilde{I},\tilde{Z},\tilde{X}\right)=\frac{1}{N}\ \left\{\sum_{i=1}^N I_iX_i^2\mathbb{E}\left[\delta_{i,j}^2|\tilde{Z}\right]\right\}\to\rho_3\sigma_{\delta,j}^2.$$

Therefore, we get $\mathsf{Var}\left(\frac{A_4}{\sqrt{N}}\middle|\tilde{I},\tilde{Z},\tilde{X}\right)\to\rho_3\sigma_\delta^2+\left[a^2\rho_a^2p+b^2\rho_b^2(1-p)\right]\sigma_{z*}^2$.

Next, we look at the conditional covariances between terms of $\sqrt{N}A$. By weak law

of large numbers, we have

$$\begin{aligned}
&\mathsf{Cov}\left(\frac{\sum_{i=1}^N Z_{i,j}}{\sqrt{N}},\ \frac{\sum_{i=1}^N I_i Z_{i,j}}{\sqrt{N}}\middle|\tilde{I},\tilde{Z}\right)\\
=&\frac{1}{N}\left[\mathsf{Cov}\left(\sum_{i=1}^N \mathbb{E}(Z_{i,j}|d_j(Z_{i,j})),\ \sum_{i=1}^N I_i\mathbb{E}(Z_{i,j}|d_j(Z_{i,j})\middle|\tilde{I},\tilde{Z}\right)\right.\\
&\left.+\mathsf{Cov}\left(\sum_{i=1}^N \delta_{i,j},\ \sum_{i=1}^N I_i\delta_{i,j}\middle|\tilde{I},\tilde{Z}\right)\right]\to \rho_1(\sigma_{\delta,j}^2+\tau_j^2)=\rho_1\sigma_j^2.
\end{aligned}$$

Therefore, $\mathsf{Cov}\left(\frac{A_1}{\sqrt{N}},\ \frac{A_2}{\sqrt{N}}\middle|\tilde{I},\tilde{Z}\right)\ \to\ \rho_1\sigma^2$. In the same way, we get

$$\begin{aligned}
&\mathsf{Cov}\left(\frac{A_1}{\sqrt{N}},\ \frac{A_4}{\sqrt{N}}\middle|\tilde{I},\tilde{Z},\tilde{X}\right)\ \to\ \rho_2\sigma^2,\ \mathsf{Cov}\left(\frac{A_2}{\sqrt{N}},\ \frac{A_3}{\sqrt{N}}\middle|\tilde{I},\tilde{Z},\tilde{X}\right)\ \to\ \rho_2\sigma^2,\\
&\mathsf{Cov}\left(\frac{A_2}{\sqrt{N}},\ \frac{A_4}{\sqrt{N}}\middle|\tilde{I},\tilde{Z},\tilde{X}\right)\ \to\ \rho_2\sigma_\delta^2+ap(\rho_a^2-\rho_b^2)\sigma_{z*}^2,\\
&\mathsf{Cov}\left(\frac{A_3}{\sqrt{N}},\ \frac{A_4}{\sqrt{N}}\middle|\tilde{I},\tilde{Z},\tilde{X}\right)\ \to\ \rho_3\sigma^2.
\end{aligned}$$

Hence, by the central limit theorem, given $(\tilde{I},\tilde{Z},\tilde{X})$, we have that

$$\frac{1}{\sqrt{N}|\boldsymbol{M}|}(c_1A_1+c_2A_2+c_3A_3+c_4A_4)$$

is asymptotically normal with mean zero and variance $\dfrac{\sigma_{A*}^2}{|M|^2}$,

$$\begin{aligned}
\text{where } \sigma_{A*}^2=&\sigma_A^2+(\gamma_1^2\sigma_{\delta,1}^2+\gamma_2^2\sigma_{\delta,2}^2)\cdot\Big\{c_4^2\cdot\left[pa^2\rho_a(1-\rho_a)+(1-p)b^2\rho_b(1-\rho_b)\right]\\
&+c_2^2\cdot[p\rho_a(1-\rho_a)+(1-p)\rho_b(1-\rho_b)]+2c_2c_4\cdot ap(\rho_a-\rho_b)(1-\rho_a-\rho_b)\Big\}.
\end{aligned}$$

As the limiting distribution does not depend on $(\tilde{I},\tilde{Z},\tilde{X})$, it also holds unconditionally. Further, by Slutsky's theorem, we have

$$\sqrt{N}A\ \xrightarrow{D}\ \mathcal{N}\left(0,\ \frac{\sigma_{A*}^2}{|M|^2}\right).$$

The proof for the convergence of denominator does not change from the discrete case. Together with Slutsky's theorem and simplification, the test statistic, under the null hypothesis $H_0 : \ \mu_1 - \mu_2 = 0$, has the following asymptotic distribution

$$T \ \xrightarrow{D} \ \mathcal{N}\left(0, \ \frac{\sigma_\delta^2}{\sigma^2}\right), \text{ where } \sigma_\delta^2 = \gamma_1^2\sigma_{\delta,1}^2 + \gamma_2^2\sigma_{\delta,2}^2 + \sigma_\epsilon^2 \text{ and } \sigma^2 = \gamma_1^2\sigma_1^2 + \gamma_2^2\sigma_2^2 + \sigma_\epsilon^2. \ \square$$

Chapter 6

Conclusions and Discussions

6.1 Conclusion Remarks

Precision medicine has become a very important part of healthcare, and "an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person" according to the National Institutes of Health. Development in the fields of translational research allows scientists to identify more and more biomarkers/covariates, which are associated with certain diseases and can be incorporated in clinical trials to select or optimize therapeutic care. Based on the different roles played in clinical studies, we categorize covariates of patients into two types: the prognostic covariates for balancing purpose and the predictive covariates for treatment selection.

In the literature, many adaptive randomization designs handling covariates have been proposed. Note that when covariates and/or responses are used in randomization, the validity of statistical inference can be affected, and proper adjustment may be needed for Type I error and power. In this work, we studied theoretical properties of statistical inference under adaptive designs utilizing covariates based on linear models. In Chapter 2 and 3, we derived the corresponding asymptotic distributions of test statistics under both null and alternative hypotheses for CARA designs, which only deal with predictive covariates for efficiency and ethics reasons, and covariate-adaptive designs, which only balance prognostic covariates across treatment groups to provide a valid comparison. And, we established an innovative class of new CARA

designs in Chapter 4, which incorporate both types of covariates in the randomization procedure simultaneously. We also provided theoretical results of statistical inference under the new design.

The covariate-adjusted response-adaptive (CARA) designs update allocation probabilities based on history information of responses and predictive covariates to facilitate treatment selection. Better average outcomes can be obtained without compromising power if all covariates used in randomization are included for inference as well. However, the allocation scheme can have a great impact on statistical inference. Based on Theorem 2.3.1, the test for comparing treatment effects is not always valid if the covariates used in the randomization procedure are fully or partially omitted in the inference analysis, and its performance depends on the choice of the allocation function and unknown parameters in the employed model. Clinicians should properly implement CARA designs to obtain a valid test. And, one possible solution for this problem is bootstrap methods. Prognostic covariates are not considered in CARA designs, therefore large imbalances may exist between treatment groups, even compared with complete randomization.

The covariate-adaptive designs (CAR) improve balance of treatment allocations to provide more comparable treatment groups, leading to a more credible trial. With correct model specification for responses and covariates, a valid test can be obtained if all covariates balanced in the randomization procedure are also included in the final analysis. However, according to Theorem 3.3.1, if these prognostic covariates are fully or partially omitted in the inference procedure, the classical tests for comparing treatment effects are usually conservative due to the overestimation of variance of the test statistic. Covariate-adaptive designs only incorporate prognostic covariates in randomization, without considering predictive covariates and responses for ethical issues.

The proposed new CARA design deals with both prognostic covariates and predictive covariates simultaneously, which makes it different from all classical covariate-adaptive designs and CARA designs. The target allocation proportions depend on the predictive covariates, and prognostic covariates are balanced at the overall, marginal and within-stratum levels at the same time. It has been shown by simulations that the new CARA design can significantly reduce imbalances at all three levels and has advantage on assigning more patients to a more appropriate treatment by targeting different allocation proportions based on the predictive covariates of subjects. Classical tests of comparing treatment effects are usually conservative when prognostic covariates are excluded from the inference based on Theorem 4.3.1, and the bootstrap method can be applied to restore the nominal error rate.

In Chapter 5, we extended the theoretical results for comparing treatment effects under the covariate-adaptive design in Chapter 3 and the new CARA design with fixed allocation proportions by incorporating continuous prognostic covariates directly in the underlying model. Under certain conditions, similar conclusions with discrete cases can be obtained. Theorem 5.2.1 and Theorem 5.2.2 showed that the hypothesis testings are conservative under both designs when prognostic covariates are omitted from the analysis procedure.

6.2 Further Discussions

Due to the complex inter-dependence among covariates, treatment assignments and responses introduced by the allocation scheme, theoretical results in this work are obtained under restricted conditions and have some limitations. The proposed properties of hypothesis testings for new CARA designs can be generalized in several ways.

(1) Both types of covariates incorporated in the new design are restricted to be dis-

crete. If covariates are continuous, the continuous-discrete conversion has to be performed, which will lose information of covariates. In Chapter 5, we extended the discrete assumption for prognostic covariates. However, theoretical results of the hypothesis testing for comparing treatment effects can only be derived when new CARA designs use fixed target allocation proportions in the randomization procedure. Statistical inference using adaptively estimated allocation proportions still remains unknown. Also, utilizing continuous predictive covariates directly is next to be considered.

(2) All proposed properties in this work are based on clinical trials with two treatments, which can be generalized to multiple treatment arms. Also, only one predictive covariate and two prognostic covariates are included in the underlying full model. Similar problems can be investigated in a more general case with p predictive covariates and q prognostic ones.

(3) Hypothesis testings are studied under a linear framework for continuous responses. Dichotomous responses are very common in clinical practice, hence properties for logistic regressions are also desired. Since there are no closed forms of parameter estimation in logistic regressions, the study for hypothesis testings of comparing treatment effects will be much more complicated.

(4) CARA designs in Chapter 2 and new designs in Chapter 4 both assume that responses of subjects can be observed immediately after assignments. However, this barely happens in clinical practice. Until the response of a subject is observed, the information cannot contribute to adjust treatment allocations of later subjects. Incorporation of time-delayed responses into the randomization scheme needs to be considered.

(5) In this work, the underlying response model 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 randomization procedures. Moreover, we assume that the full model has correct model specification that coincides with the underlying model. However, the true underlying model is usually unknown in practice and it is possible that covariate effects are not linearly additive. Therefore, the situation of model mis-specification is another topic that is 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] Atkinson, A. (2004). Adaptive biased-coin designs for clinical trials with several treatments. *Discussiones Mathematicae - Probability and Statistics*, 24:85–108.

[3] Atkinson, A. and Biswas, A. (2005a). Adaptive biased-coin designs for skewing the allocaiton proportion in clinical trials with normal responses. *Statistics in Medicine*, 24:2477–2492.

[4] Atkinson, A. and Biswas, A. (2005b). Bayesian adaptive biased-coin designs for clinical trials with normal responses. *Biometrics*, 61:118–125.

[5] Bandyopadhyay, U. and Biswas, A. (2001). Adaptive designs for normal responses with prognostic factors. *Biometrika*, 88:409–419.

[6] Barbáchano, Y. and Coad, D. (2013). Inference following designs which adjust for imbalances in prognostic factors. *Clinical Trials*, 10:540–551.

[7] Birkett, N. (1985). Adaptive allocation in randomized controlled trials. *Controlled Clinical Trials*, 6:146–155.

[8] Buyse, G., Goemans, N., Van den Hauwe, M., Thijs, D., de Groot, I., Schara, U., Ceulemans, B., Meier, T., and Mertens, L. (2011). Idebenone as a novel, therapeutic approach for duchenne muscular dystrophy: Results from a 12 month, double-blind, randomized placebo-controlled trial. *Neuromuscular Disorders*, 21:396–405.

[9] Chow, S. and Chang, M. (2006). *Adaptive design methods in clinical trials*. Chapman and Hall/CRC Press, Taylor and Francis, New York, NY.

[10] Cox, D. and McCullagh, P. (1982). Some aspects of analysis of covariance (with discussion). *Biometrics*, 38:541–561.

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

[12] Eisele, J. (1994). The doubly adaptive biased coin design for sequential clinical trials. *Journal of Statistical Planning and Inference*, 38:249–262.

[13] Eisele, J. and Woodroofe, M. (1995). Central limit theorems for doubly adaptive biased coin designs. *The Annals of Statistics*, 23:234–254.

[14] Feinstein, A. and Landis, J. (1976). The role of prognostic stratification in preventing the bias permitted by random allocation of treatment. *Journal of Chronic Diseases*, 29:277–284.

[15] Forsythe, A. (1987). Validity and power of tests when groups have been balanced for prognostic factors. *Computational Statistics Data Analysis*, 5:193–200.

[16] Gallo, P., Chuang-Stein, C., Dragalin, V., Gaydos, B., Krams, M., and Pin-heiro, J. (2006). Adaptive design in clinical drug development an executive summary of the phrma working group (with discussions). *Journal of Biopharmaceutical Statistics*, 16:275–283.

[17] Green, S. and Byar, D. (1978). The effect of stratified randomization on size and power of statistical tests in clinical trials. *Journal of Chronic Diseases*, 31:445–454.

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

[19] Hu, F. (2012). Statistical issues in trial design and personalized medicine. *Clinical Investigation*, 2:121–124.

[20] Hu, F., Hu, Y., Ma, W., Zhang, L., and Zhu, H. (2015a). Statistical inference of adaptive randomized clinical trials for personalized medicine. *Clinical Investigation*, 5:415–425.

[21] Hu, F. and Rosenberger, W. (2006). The theory of response-adaptive ran- domization in clinical trials. *John Wiley and Sons. Wiley Series in Probability and Statistics*.

[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] Hu, J., Zhu, H., and Hu, F. (2015b). A unified family of covariate-adjusted response-adaptive designs based on efficiency and ethics. *Journal of the American Statistical Association*, 110:357–367.

[24] Hu, Y. and Hu, F. (2012). Asymptotic properties of covariate-adaptive randomization. *The Annals of Statistics*, 40:1794–1815.

[25] Hyman, D., Puzanov, I., Subbiah, V., Faris, J., Chau, I., Blay, J., Wolf, J., Raje, N., Diamond, E., Hollebecque, A., Gervais, R., Elez-Fernandez, M., Italiano, A., Hofheinz, R., Hidalgo, M., Chan, E., Schuler, M., Lasserre, S., Makrutzki, M., Sirzen, F., Veronese, M., Tabernero, J., and Baselga, J. (2015). Vemurafenib in multiple nonmelanoma cancers with braf v600 mutations. *New England Journal of Medicine*, 373:726–736.

[26] Ivanova, A. (2003). A play-the-winner type urn model with reduced variability. *Metrika*, 58:1–13.

[27] Krisam, J. and Kieser, M. (2015). Optimal decision rules for biomarker-based subgroup selection for a targeted therapy in oncology. *International journal of molecular sciences*, 16:10354–10375.

[28] Kundt, G. (2009). Comparative evaluation of balancing properties of stratified randomization procedures. *Methods of Information in Medicine*, 48:129–134.

[29] Lachin, J., Matts, J., and Wei, L. (1988). Randomization in clinical trials: conclusions and recommendations. *Controlled Clinical Trials*, 9:365–374.

[30] Larkin, J., Chiarion-Sileni, V., Gonzalez, R., Grob, J., Cowey, C., Lao, C., Schadendorf, D., Dummer, R., Smylie, M., Rutkowski, P., and Ferrucci, P. (2015). Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. *New England Journal of Medicine*, 373:23–34.

[31] Lebowitsch, J., Ge, Y., Young, B., and Hu, F. (2012). Generalized multidimensional dynamic allocation method. *Statistics in Medicine*, 31:3537–3544.

[32] Ma, W., Hu, F., and Zhang, L. (2015). Testing hypotheses of covariate-adaptive randomized clinical trials. *Journal of the American Statistical Association*, 110:669–680.

[33] Matts, J. and Lachin, J. (1988). Properties of permuted-block randomization in clinical trials. *Controlled Clinical Trials*, 9:327–344.

[34] McEntegart, D. (2003). The pursuit of balance using stratified and dynamic randomization techniques: an overview. *Drug Information Journal*, 37:293–308.

[35] Melfi, V. and Page, C. (2000). Estimation after adaptive allocation. *Journal of Statistical Planning and Inference*, 87:353–363.

[36] Melfi, V., Page, C., and Geraldes, M. (2001). An adaptive randomized design with application to estimation. *Canadian Journal of Statististics*, 29:107–116.

[37] Pocock, S. (1982). Statistical aspects of clinical trial design. *The Statistician*, 31:1–18.

[38] Pocock, S. and Simon, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*, 3:103–115.

[39] Rosenberger, W. and Hu, F. (2004). Asymptotic properties of covariate-adaptive randomization. *Clinical Trials*, 1:141–147.

[40] Rosenberger, W. and Sverdlov, O. (2008). Handling covariates in the design of clinical trials. *Statistical Science*, 23:404–419.

[41] Rosenberger, W., Vidyashankar, A., and Agarwal, D. (2002). Covariate-adjusted response-adaptive designs for binary response. *Journal of Biopharmaceutical Statistics*, 11:227–236.

[42] Russell, D., Hoare, Z., Whitaker, R., Whitaker, C., and I.T., R. (2011). Generalized method for adaptive randomization in clinical trials. *Statistics in Medicine*, 30:922–934.

[43] Shao, J. and Yu, X. (2013). Validity of tests under covariate-adaptive biased coin randomization and generalized linear models. *Biometrics*, 69:960–969.

[44] Shao, J., Yu, X., and Zhong, B. (2010). A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika*, 97:347–360.

[45] Signorini, D., Leung, O., Simes, R., Beller, E., Gebski, V., and Callaghan, T. (1993). Dynamic balanced randomization for clinical trials. *Statistics in medicine*, 12:2343–2350.

[46] Slamon, D., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M., Baselga, J., and Norton, L. (2001). Use of chemotherapy plus a monoclonal antibody against her2 for metastatic breast cancer that overexpresses her2. *New England Journal of Medicine*, 344:783–792.

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

[48] Taves, D. (2010). The use of minimization in clinical trials. *Contemporary Clinical Trials*, 31:180–184.

[49] Toorawa, R., Adena, M., Donovan, M., Jones, S., and Conlon, J. (2003). Use of simulation to compare the performance of minimization with stratified blocked randomization. *Pharmaceutical Statistics*, 8:264–278.

[50] Tu, D., Shalay, K., and Pater, J. (2000). Adjustment of treatment effect for covariates in clinical trials: Statistical and regulatory issues. *Drug Information Journal*, 34:511–523.

[51] Tymofyeyev, Y., Rosenberger, W., and Hu, F. (2007). Implementing optimal allocation in sequential binary response experiments. *Journal of the American Statistical Association*, 102:224–234.

[52] Wei, L. and Durham, S. (1978). The randomized play-the-winner rule in medical trials. *Journal of the American Statistical Association*, 73:840–843.

[53] Weir, C. and Lees, K. (2003). Comparison of stratification and adaptive methods for treatment allocation in an acute stroke clinical trial. *Statistics in Medicine*, 22:705–726.

[54] Zhang, L., Hu, F., Cheung, S., and Chan, W. (2007). Asymptotic properties of covariate-adjusted response-adaptive designs. *The Annals of Statistics*, 35:1166–1182.