Statistical Inference for Adaptive Designs in Multi-Center Clinical Trials

by

© **Selvakkadunko Selvaratnam**

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Department of Mathematics and Statistics

Memorial University

January 2018

St. John's, Newfoundland and Labrador, Canada

Abstract

We discuss methods for comparing effects of two treatments A and B. We investigate the performance of response-adaptive (RA) and covariate-adjusted response-adaptive (CARA) designs in multi-center clinical trials. First, we discuss applying RA designs to maximize the well-being of participating patients in multi-center clinical trials. We assume that the centers are selected from a large population of centers and develop a generalized linear mixed model (GLMM) to examine the treatment effect. The asymptotic properties of the maximum likelihood (ML) estimators of model parameters are derived using the influence function method. We verified their theoretical properties through simulation studies. The techniques are then applied to a real data that were obtained from a multi-center clinical trial designed to compare two cream preparations (active drug/control) for treating an infection. Secondly, we investigate the efficiency for estimates of model parameters and ethics for participating patients among RA, CARA, and completely randomized (CR) designs for a generalized linear model (GLM). We consider the logit model to measure efficiency and ethics. Furthermore, we showed that ML estimators of GLM parameters are consistent and asymptotically follow multivariate normal distribution for adaptive designs. A simulation study was conducted to verify these theoretical results. Finally, we provide a justification of why asymptotic results for Wald-type tests for adaptive designs can be used. We proved that the choice of adaptive designs affects the statistical power of hypothesis testing via these quantities: the target allocation proportion, the bias of the randomization procedure from the target, and the variability induced by the randomization process. Moreover, we showed that the statistical power increases when the design variability decreases for a covariate in a logit model. Our theoretical findings are verified by simulation results.

This thesis is dedicated to my father, Selvaratnam Kandiah, and my mother, Parameswary Selvaratnam, for their love, endless support and encouragement.

Acknowledgements

First, I would like to express my sincere gratitude to my advisor Dr. Yanqing Yi and co-advisor Dr. Alwell Julius Oyet for their guidance, supervision, patience, enthusiasm, motivation, positivity and immense knowledge during my Ph.D program. They have provided constant support during my Ph.D. research and guided me throughout the preparation of this thesis.

I am also very appreciative of the efforts of the members of my supervisory committee, Dr. J Concepción Loredo-Osti, and Dr. Zhaozhi Fan, and their concern throughout the completion of my program. Also, their valuable feedbacks and comments have helped to improve this thesis.

My sincere thanks also goes to Dr. Veeresh Gadag, for useful discussions and valuable comments during the initial stage of this research.

I would like to thank Dr. Alwell Julius Oyet, Dr. Asokan Mulayath Variyath, Dr. Brajendra Sutradhar, Dr. Hong Wang, Dr. J Concepción Loredo-Osti, and Dr. Zhaozhi Fan who have instructed me over the years and guided my education in Statistics.

I sincerely acknowledge the financial support provided by the School of Graduate Studies, the Department of Mathematics & Statistics, and Center for Fisheries Ecosystems Research (CFER) at the Marine Institute, Memorial University of Newfoundland in the forms of Graduate Fellowship, Teaching Assistants, and Graduate Assistants. Also, I gratefully acknowledge the financial support provided by Dr. Yi through her Natural Sciences and Engineering Research Council of Canada discovery grant and the Dean Rourke Transition Fund from the Faculty of Medicine.

I would like to express my sincere gratitude to Dr. Noel Cadigan and administrative Director Thomas Brown, the Center for Fisheries Ecosystems Research (CFER) at the Marine Institute, Memorial University of Newfoundland, for the job opportunity as a graduate assistant. This was a remarkable assistance towards the completion of this thesis.

It is my great pleasure to thank my friends who encouraged and helped me during the Ph.D program.

Finally, my deep and sincere gratitude to my parents: Selvaratnam Kandiah and Parameswary Selvaratnam, my sisters: Senthamilselvi and Nanthini, my brother: Iraivan, my wife: Ajitha, my children: Nithulan and Nilaneka, for their continuous and unparalleled love, help, support, and sacrifices.

Table of contents

List of tables

List of figures

List of symbols

A	Treatment A
B	Treatment B
$C(\mathbf{z_i}) =:$	$(\mathbf{z'_1}, \mathbf{z'_2}, \cdots, \mathbf{z'_i})'$
d	The rank of a matrix $\mathbf{D}$
d^*	The number of sample points in Gauss-Hermite approximation
$\mathbf{d}_0$	The column matrix
$\mathbf{D}$	A matrix of a full rank
H_0	The null hypothesis
H_A	The alternative hypothesis
$F^*(\mathbf{t})$	The distribution function used in the influence function method
$\mathbf{F}_n(\boldsymbol{\theta})$	The observed Fisher information matrix
$(1/n)\mathbf{F}_n(\boldsymbol{\theta})$	The average observed Fisher information matrix
g	An allocation function in doubly adaptive biased coin design
$g^*(.)$	A regression function
G^*	An arbitrary distribution function used in the influence function method
$\mathbf{G}$	An open ball with center $\boldsymbol{\theta_0}$
$\mathbf{I}^*$	Identity matrix
$\mathbf{I}(\boldsymbol{\theta})$	Fisher information matrix
J	The number of medical centers
K	The number of covariates other than treatment
$L_k + 1$	The number of levels in the covariate v_k^*; $k = 1, 2, \cdots, K$
m	The total number of covariate configuration levels
n_j^*	The total number of patients in center j; $j = 1, 2, \cdots, J$
n_h	The total number of patients in the configuration level h; $h = 1, 2, \cdots, m$
n	The total number of patients
p	The number of dummy variables for covariates in the logit model
P_{AS}	The true success probability of treatment A
P_{BS}	The true success probability of treatment B
$q :=$	$2(p+1)$

List of symbols

S	The number of simulations
$\mathbf{s}_n(\boldsymbol{\theta})$:	The score function
T_W	The Wald-type of test statistic
$\mathbf{X}$	The design matrix
X_{ijA}	A random variable that patient i in center j is assigned to treatment A
X_{iA}	A random variable that patient i is assigned to treatment A when a common randomization for treatment assignments is applied to all centers
Y_{ij}	A random variable that is the response of patient i in center j
Y_i	A random variable that is the response of patient i when a common randomization for treatment assignments is applied to all centers
u	The initial number of balls in the randomized play the winner
u_j	The effect by center j
$u_j^* :=$	$\frac{u_j}{\sqrt{2}\sigma}$
v_k^*	The covariate k; $k = 1, 2, \cdots, K$
$\mathbf{v}_h$	The covariate configuration level h or stratum h
$\mathbf{w}_i$: $=$	$(x_{iA}, 1, \mathbf{z}_i', x_{iA}\mathbf{z}_i')'$
$\mathbf{w}_{iR}$: $=$	$(x_{iA}, 1, \mathbf{z}_i')'$
$\mathbf{z_{ij}}$	The vector of covariates of patient i in center j
$\mathbf{z_i}$	The vector of covariates of patient i when a common randomization for treatment assignments is applied to all centers
ρ_h:=	$P(\mathbf{Z} = \mathbf{v}_h)$
$\mathcal{X}_{ij}$:=	$\sigma(X_{1jA}, X_{2jA}, \cdots, X_{ijA})$ is the sigma algebra of treatment assignments in center j
$\mathcal{X}_i$:=	$\sigma(X_{1A}, X_{2A}, \cdots, X_{iA})$ is the sigma algebra of treatment assignments when a common randomization for treatment assignments is applied to all centers
$\mathcal{Y}_{ij}$:=	$\sigma(Y_{1j}, Y_{2j}, \cdots, Y_{ij})$ is the sigma algebra of responses in center j
$\mathcal{Y}_i$:=	$\sigma(Y_1, Y_2, \cdots, Y_i)$ is the sigma algebra of responses when a common randomization for treatment assignments is applied to all centers
$\mathcal{F}_{ij}$:=	$\sigma(\mathcal{X}_{ij}, \mathcal{Y}_{ij})$
$\mathcal{F}_i$:=	$\sigma(\mathcal{X}_i, \mathcal{Y}_i)$

List of symbols

α	It is used in the randomized play the winner
β	It is used in the randomized play the winner
α^*	The size of the test
$\boldsymbol{\beta}$:=	(γ_0, β_A)
σ	The variance of center effects in GLMM
β_A	The effect of treatment A compared to treatment B
$\boldsymbol{\delta}$	The interaction effects
γ_0	Intercept term
$\boldsymbol{\gamma}$	The main effects of covariates
Λ_{ij} :=	$P(Y_{ij} = 1 \mid x_{ijA}, \mathbf{z_{ij}})$
Λ_i :=	$P(Y_i = 1 \mid x_{iA}, \mathbf{z_i})$
	when a common randomization for treatment assignments is applied to all centers
η_i:=	$\mathbf{w}_i'\boldsymbol{\theta}$
η_{ij}:=	$\mathbf{w}_{ij}'\boldsymbol{\theta}$
$\boldsymbol{\theta}$:=	$(\beta_A, \gamma_0, \boldsymbol{\gamma}', \boldsymbol{\delta}')'$
$\boldsymbol{\theta_0}$:=	$(\beta_{A0}, \gamma_{00}, \boldsymbol{\gamma}_0', \boldsymbol{\delta}_0')'$ is the q-dimensional vector of true unknown parameters
$\boldsymbol{\Omega}(\boldsymbol{\theta_0})$	An admissible set
$\phi^{(a)}$:=	$n[\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0]'[\mathbf{D}\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{D}']^{-1}[\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0]$ is a non-centrality parameter
ϕ:=	$n[\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0]'\{\mathbf{D}[(1/n)\mathbf{F}_n(\boldsymbol{\theta_0})]^{-1}\mathbf{D}'\}^{-1}[\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0]$
π_{ijA}	The probability that patient i in center j is getting treatment A.
π_{iA}	The probability that patient i is getting treatment A
	when a common randomization for treatment assignments is applied to all centers
Φ	The cumulative distribution function of the standard normal distribution
$\Psi(\Lambda_i)$:=	η_i
$\boldsymbol{\theta_0}$	True parameter vector
ξ	One of the doubly adaptive biased coin design parameters
ρ	A desired allocation proportion in doubly adaptive biased coin design
υ	The proportion of patients assigned to treatment A

List of abbreviations

AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
CA	Covariate-Adaptive
CARA	Covariate-Adjusted Response-Adaptive
CR	Completely Randomized
CRD	Completely Randomized Design
DBCD	Doubly adaptive biased coin design
FDA	Food and Drug Administration
GLM	Generalized Linear Model
GLMM	Generalized Linear Mixed Model
GQL	Generalized Quasi-Likelihood
ML	Maximum Likelihood
MM	Method of Moments
PW	Play the winner
QAIC	Quasi Akaike Information Criterion
QL	Quasi-Likelihood
RA	Response-Adaptive
RACA	Response-Adaptive Covariate-Adjusted
RPW	Randomized play the winner

Chapter 1

An Overview of Statistical Modelling and Adaptive Designs in Clinical Trials

In the past several decades, much research has been conducted in developing therapy methods and drug development through clinical trials. Researchers have been interested in acquiring an efficient procedure for comparing new treatments with existing ones. These procedures include the design criteria, which consist of treatment assignments to patients, and comparisons of treatment methods characterized by statistical approaches for identifying the best treatment. At the initial stage of a clinical trial, the selection of a suitable design for treatment assignment is of primary concern. It is also important to identify a suitable model that will be used for data analysis at the initial stage of the experiment. Thus is however challenging due to lack of sufficient data at the beginning of an experiment. In fact, sufficient data usually is available at the end of a clinical trial. Given sufficient data for statistical inference, one can choose a simple and an efficient model, for future usage, from a set of candidate models. For

example, the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) can be used to choose a model to describe a given data set. Furthermore, the Quasi-AIC (QAIC) can be applied as a model selection criterion when the estimation of regression parameters are based on Quasi-Likelihood (QL) methods [Burnham and Anderson (2002)]. Thus, the absence of sufficient data at the initial stage of the trial makes the identification of the best model a challenging task.

In this thesis, we investigate some designs for treatment assignment and develop a new approach for conducting statistical inference for the purpose of comparing two treatments, say A and B, in multi-center clinical trials. In particular, we will investigate the performance of Response-Adaptive (RA), and Covariate-Adjusted Response-Adaptive (CARA) designs in maximizing the well-being of participating patients while collecting responses and associated covariates and assigning treatments to patients at participating medical centers. The purpose is to optimise the resources while having an efficient statistical inference procedure at the end of the clinical trial. Throughout this thesis, treatment A will be considered an experimental treatment and treatment B will be an existing treatment. Burnham and Anderson (2002) notes that although an ideal data set cannot be collected to explain the behavior of treatments A and B, the experimenter should be cautious when collecting data for this purpose. For instance, an experimenter has to be cautious with the type of response (binary, count, continuous, and longitudinal) and covariates to be collected, since the correct identification of the behavior of treatments A and B depends on the data collected.

According to Sverdlov (2016), clinical trials typically have several goals, which can be

divided into **two main objectives**:

objective 1 : have an efficient statistical inference at the end of a clinical trial, and

objective 2 : respect the well-being and dignity of participating patients. (1.1)

According to Sverdlov (2016), efficiency generally refers to the power of testing a research hypothesis in clinical trials, while ethics often concerns patients assigned to unsafe or inferior treatments. An efficient statistical inference is necessary for the well-being of future patients. Sverdlov (2016) refers to the conflict between these two objectives as "individual versus collective ethics". It is clear that statistical power increases when sample size is larger. However, increasing the sample size results in the following: (i) raising the cost of an experiment and (ii) increasing the number of patients in randomization.

Suppose a clinical trial is conducted in J medical centers, which are randomly selected from a large number of medical facilities. Furthermore, suppose that the responses are binary and denoted by Y_{ij}. Let n_j^* be the total number of patients who are recruited and assigned to one and only one of the two treatments in center j, $j = 1, 2, \cdots, J$ and $n = \sum_{j=1}^{J} n_j^*$, where n is the total number of participating patients. Define the response Y_{ij} and treatment assignment X_{ijA} to patient i in center j, for $i = 1, 2, \cdots, n_j^*$ and $j = 1, 2, \cdots, J$ by

$$Y_{ij} = \begin{cases} 1 & \text{if treatment is} \\ & \text{a success,} \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad X_{ijA} = \begin{cases} 1 & \text{if patient } i \text{ in center } j \\ & \text{is assigned to treatment } A, \\ 0 & \text{otherwise.} \end{cases} \tag{1.2}$$

Although the comparison of treatments effect is the primary goal, treatment effect can

be efficiently compared when suitable covariates are included in a proper model. Some typical covariates are gender, smoking status, age, cholesterol level, chronic conditions, and so on. The smoking status of a patient can influence the response of medical care for cancer, whereas the patients' chronic disease, for instance high blood glucose level or cholesterol level, can affect the response of medical care for hypertension. Suppose that K number of categorical covariates other than treatment, say $v_1^*, v_2^*, \cdots, v_K^*$, are collected in the clinical trial, and each covariate has a finite number of levels, where v_k^* has $L_k + 1$ levels: say $v_{k0}^*, v_{k1}^*, \cdots, v_{kL_k}^*$ for $k = 1, 2, \cdots, K$. Then, we can define a set of dummy variables based on the reference category v_{k0}^*, $k = 1, 2, \cdots, K$, for each covariate. Without loss of generality, let $\mathbf{Z'_{ij}} = (Z_{ij1}, \cdots, Z_{ijp})$ be the p dimensional vector of dummy variables corresponding to the covariates of patient i in center j, where $p = \sum_{k=1}^{K} L_k$ and each component of $\mathbf{Z_{ij}}$ has binary levels, for $i = 1, 2, \cdots, n_j^*$ and $j = 1, 2, \cdots, J$. These dummy variables represent the characteristic of patients. Thus, $\mathbf{Z_{ij}}$ is the covariates of patient i in center j.

Sverdlov (2016) notes that one can assume a starting model for a clinical trial. In what follows, we add center effects to Sverdlov (2016) initial model because the experiment in this thesis is conducted at multiple clinics. Therefore, we assume that Y_{ij}, conditional on x_{ijA}, $\mathbf{z_{ij}}$, and u_j follows the statistical model,

$$E(Y_{ij}|x_{ijA}, \mathbf{z_{ij}}, u_j) \quad = \quad g^*(\boldsymbol{\theta}, x_{ijA}, \mathbf{z_{ij}}, u_j), \tag{1.3}$$

for $i = 1, 2, \cdots, n_j^*$ and $j = 1, 2, \cdots, J$, where $g^*(.)$ is a regression function; u_j is the effect of center j; $\boldsymbol{\theta}$ is a vector of model parameters which includes the main effect of treatment A compared to treatment B, the main effects of other covariates, and the effects of treatment by covariate interactions.

In this thesis, we wish to conduct efficient statistical inference of the treatment effect, in the following areas:

(i) **The selection of an appropriate regression function g^*.**

The function g^* in (1.3) can be selected from existing binary link functions such as logit, probit, cauchit, and complementary log-log [See McCullagh and Nelder (1989)]. The maximum likelihood (ML) method for parameter estimation rely on a link function. On the other hand, identifying the behavior of collected data using either the Quasi-Likelihood (QL) or the Generalized Quasi-Likelihood (GQL) method does not require a full model assumption to conduct statistical inference. Also, a nonparametric method could be used to conduct inference at end of the clinical trials [Sverdlov (2016)].

(ii) **The inclusion of variables in linear predictor.**

In our analysis, we will determine the types of variables, such as covariate and center effect, that should be incorporated into the linear predictor. Depending on the variables included in the model, we will use the Generalized Linear Model (GLM), or the Generalized Linear Mixed Model (GLMM) or the Generalized Linear Model with fixed center effects to analyze the data set.

(iii) **The estimation of model parameters.**

A suitable method will be implemented to estimate the model parameters. Some methods in literature are Maximum Likelihood (ML), Method of Moments (MM), Quasi-Likelihood (QL) [see Wedderburn (1974); McCullagh and Nelder (1983)], and Generalized Quasi-Likelihood (GQL) [see Sutradhar (2003)]. In this thesis, we have applied the ML method. We also introduced a new approach in § 2.3 for computing MLE based on the concept of influence functions.

(iv) **Hypothesis testing.**

After estimating the model parameters, we will conduct hypothesis testing to identify whether treatment A is significantly different from treatment B for a given significance level. For example, a Likelihood-Ratio test, a Score test, or a Wald test will be implemented to test regression parameters. Furthermore, the power of the test for a given size of the test will be computed and checked with the experimenter's predetermined value of power. Also, if the computed power does not attain this threshold value, then we will increase the sample size until this threshold value is achieved.

(v) **Interpretation of results.**

Finally, the results will be interpreted after reaching the experimenter's threshold value of power. Also, the best treatment will be identified at the end of the clinical trial.

1.1 Designs in Clinical Trials

Friedman et al. (2015) define "a clinical trial as a prospective study comparing the effects and value of intervention(s) against an existing treatment in human beings". For instance, in a clinical trial to compare a new treatment, say A against an existing treatment, B, treatment A is considered the intervention and treatment B as the existing. In statistics, the variable treatment is an example of a controllable variable or factor since the values or levels of the treatments can be set by the experimenter. Thus, we define controllable variables or factors as any variable that might influence the response whose values the experimenter can set. That is, by controllable variable we mean that the experimenter can decide the type of randomization to apply in selecting a design. In contrast, covariates are variables that might affect the outcome

but experimenters cannot control, nonetheless, these covariates can be measured.

The concept of selection of designs was introduced by James Lind in a clinical trial, which was conducted in 1747 [Dunn (1997)] involving six groups of food, which were suspected by Lind, to cure scurvy. However, in this experiment, there was no evidence that the designs were chosen with a pre-specified objective of conducting valid inference. According to Oyet (1997), the principle of optimal designs was first proposed by Smith (1918). In these optimal designs, a criteria is applied to choose designs for the purpose of conducting efficient statistical inference with minimum sample size, which will reduce the cost of experimentation.

When constructing optimal designs, randomization is commonly applied to lessen experimental bias. In fact, the principle of randomization introduced by Ronald Fisher in 1926 was systematically applied in agriculture. Thus, completely randomized (CR) designs or equal allocations is a randomization method used to avoid selection or experimenter bias during treatment allocation [Shao and Yu (2013)]. Furthermore, although some pivotal covariates are unknown to the experimenter, we may be able to estimate the treatment effect efficiently by applying randomization because randomization reduces experimental bias. In fact, complete randomization is used in equal allocation. However, "equal allocation may result in severe imbalance not only between the treatment groups but also across covariates" [Shao and Yu (2013)]. But, balances will be asymptotically achieved between the treatment groups as well as across covariates. In fact, reducing experimental bias is the pre-selected objective of equal allocation. Instead of the concept of complete randomization, it is possible to apply other randomization methods to quickly achieve **objective 1** in (1.1). For instance, accrued information is utilized in adaptive designs to avoid complete

randomization. So, we discuss the concept of adaptive randomization in the next section.

1.1.1 Adaptive Designs

Adaptive designs were introduced to overcome disadvantages of equal allocation without completely eliminating the principle of randomization. So, these designs impose some restrictions in the selection process while maintaining the spirit of randomization. The history of treatment assignments in an experiment is used to select the next assignment in restricted randomization, which is a member of the family of adaptive designs. In adaptive designs, accumulated data are used to create these restrictions. As a result, restricted randomization is an approach to maintain balance between treatment groups when sample size is small. Bailey (1987) had noted that the concept of restricted randomization was developed by Yates (1948) and Youden (1972). The main purpose of restricted randomization is to achieve **objective 1** in (1.1). However, this randomization was not based on the concept of optimal designs.

Contrary to the objective of restricted randomization, the concept of adaptive designs, first introduced by Thompson (1933) was to obtain data for the purpose of respecting the well-being and dignity of participating patients. In 2010, the US Food and Drug Administration (FDA) recommended some guidelines for applying adaptive designs in clinical trials. The FDA (2010) notes that modifications can be made to the design based on interim analysis of already collected data before collecting data for final analysis. Modification can be made to the following:

(i) randomization procedure,

(ii) total sample size of the study (including early termination), and

(iii) analytic methods to evaluate the endpoints (e.g., covariates of final analysis, statistical methodology, Type I error control).

These adaptive designs are not only applied to choose the best treatment, but they are also used to find the best dose level for a group of patients having specific covariates, because small dosage amounts may not significantly improve the disease status of a patient. On the other hand, a large dosage may produce dangerous side effects [FDA (2010)]. Adaptive designs are also used in the development of medical device [FDA (2015)].

There are currently a wide variety of adaptive designs in the literature. The differences in these designs are determined by their pre-selected objectives. For instance, Covariate-Adaptive (CA), Response-Adaptive (RA), Covariate-Adjusted Response-Adaptive (CARA), Response-Adaptive Covariate-Adjusted (RACA) designs are members of the family of adaptive designs. Rosenberger et al. (2012), states that "an important class of clinical trial designs is adaptive randomization, which is a change in randomization probabilities during the course of the trial to promote multiple experimental objectives, while protecting the study from bias and preserving inferential validity of the results". Next, we describe these adaptive designs based on the accruing data and randomization of treatment assignments.

A sigma algebra is a set of all possible information generated by random variables. For $i = 1, 2, \cdots, n_j^*$, let us assume that $\mathcal{X}_{ij} = \sigma(X_{1jA}, X_{2jA}, \cdots, X_{ijA})$, and $\mathcal{Y}_{ij} = \sigma(Y_{1j}, Y_{2j}, \cdots, Y_{ij})$ are the sigma algebras generated by treatment assignments, and responses respectively in center j, $j = 1, 2, \cdots, J$. Thus, for $i = 1, 2, \cdots, n_j^*$, $\sigma(\mathcal{X}_{ij}, \mathcal{Y}_{ij})$ is the sigma algebra in center j, $j = 1, 2, \cdots, J$. That is, $\sigma(\mathcal{X}_{ij}, \mathcal{Y}_{ij})$ is a set of all possible information generated by $\{X_{1jA}, X_{2jA}, \cdots, X_{ijA}, Y_{1j}, Y_{2j}, \cdots, Y_{ij}\}$. For

$i = 1, 2, \cdots, n_j^*$, define $C(\mathbf{z_{ij}}) = (\mathbf{z'_{1j}}, \mathbf{z'_{2j}}, \cdots, \mathbf{z'_{ij}})'$ is the history of covariates in center j, $j = 1, 2, \cdots, J$ and $\mathcal{F}_{ij} = \sigma(\mathcal{X}_{ij}, \mathcal{Y}_{ij})$ for $j = 1, 2, \cdots, J$. Let π_{ijA} be the probability that patient i in center j will receive treatment A, for $i = 1, 2, \cdots, n_j^*$ and $j = 1, 2, \cdots, J$. We also define similar sigma algebras when we apply a common randomization for treatment assignments to all centers. In that case, for $i = 1, 2, \cdots, n$, $\mathcal{X}_i = \sigma(X_{1A}, X_{2A}, \cdots, X_{iA})$, and $\mathcal{Y}_i = \sigma(Y_1, Y_2, \cdots, Y_i)$ are the sigma algebras generated by treatment assignments, and responses respectively with $\sigma(\mathcal{X}_i, \mathcal{Y}_i)$ as the common sigma algebra. For $i = 1, 2, \cdots, n$, define $C(\mathbf{z_i}) = (\mathbf{z'_1}, \mathbf{z'_2}, \cdots, \mathbf{z'_i})'$ is the history of covariates and $\mathcal{F}_i = \sigma(\mathcal{X}_i, \mathcal{Y}_i)$. Here, π_{iA} is the probability that patient i is getting treatment A for $i = 1, 2, \cdots, n$. In Chapter 2, response adaptive designs were only applied to each center independently. The adaptive designs in Chapters 3 and 4, however, implemented a common randomization to all centers. In the next subsections, the definitions of adaptive designs are discussed for a common randomization to all centers.

Restricted Randomization

When applying restricted randomization, treatment assignments of previous patients are taken into account when choosing a treatment for a new patient. Thus, $\pi_{iA} = 1/2$ and $\pi_{iA} = P(X_{iA} = 1|\mathcal{X}_{i-1})$ for $i = 1, 2, \cdots, n$.

Response-Adaptive (RA) Designs

In RA designs, treatment assignments and available responses of previous patients are used in randomization of treatment assignments. Therefore, $\pi_{1A,RA} = 1/2$ and $\pi_{iA,RA} = P[X_{iA} = 1|\sigma(\mathcal{X}_{i-1}, \mathcal{Y}_{i-1})]$ for $i = 2, 3, \cdots, n$, where $\sigma(\mathcal{X}_{i-1}, \mathcal{Y}_{i-1})$ is the sigma algebra of treatment assignments and responses.

Response-adaptive designs have become a desirable treatment allocation procedure in clinical trials because they commonly lead to the assignment of more patients to the potentially better treatment. Beginning with Zelen (1969), several authors have proposed a variety of adaptive procedures for allocating treatments to patients in clinical trials. Zelen (1969) introduced an urn based procedure commonly referred to as the play the winner (PW) rule for comparing pairs of treatments, say A and B. In this procedure, the first treatment assignment to patient 1 is usually made based on the outcome of tossing a fair coin. According to Zelen (1969), a success on a particular treatment generates a future trial on the same treatment with a new patient; a failure on a treatment generates a future trial on the alternate treatment. Wei and Durham (1978) noted that the time it takes to observe the response of a patient in a clinical trial may be much longer than the time between entry of new patients for treatment assignment. Thus, they proposed a modification to the PW rule called the randomized play the winner (RPW) rule. They recommended placing an initial number of u balls of each type in the urn. In their procedure, β balls of type A and α ($\beta \geq \alpha \geq 0$) balls of type B are added to the urn if the response to treatment A is a success. The type of balls added to the urn is reversed if the response to treatment B is a success, that is, β balls of type B and α ($\beta \geq \alpha \geq 0$) balls of type A are added to the urn. The randomized play the winner rule is thus usually denoted by RPW(u, α, β).

It is clear that adaptive designs based on the urn model can only be applied in clinical trials with binary responses. If an experimenter has a target proportion of allocation for, say treatment A in mind, the PW and RPW rules cannot be applied. As a result, Eisele (1994) and Eisele and Woodroofe (1995) proposed a doubly adaptive biased coin design (DBCD) which uses an allocation function $g(.)$ to target any specified allocation proportion ρ for treatment A. The design is said to be doubly adaptive because the

procedure requires estimating the value of ρ, the desired allocation proportion, after each trial and also takes into account the current proportion of subjects assigned to each treatment. In our simulation studies, we have used the allocation proportion

$$\rho(\hat{P}_{AS}, \hat{P}_{BS}) = \frac{\sqrt{\hat{P}_{AS}}}{(\sqrt{\hat{P}_{AS}} + \sqrt{\hat{P}_{BS}})}, \tag{1.4}$$

proposed by Rosenberger, Stallard, Ivanova, Harper and Ricks (2001), to compute estimates of ρ where $\hat{P}_{AS}$ and $\hat{P}_{BS}$ are the proportions of successes in the group of patients assigned to treatment A and treatment B respectively. The expression (1.4) is commonly referred to as the RSIHR allocation proportion. However, the approach of Eisele (1994) and Eisele and Woodroofe (1995) is more complicated to achieve the desired allocation proportion, ρ [see Rosenberger and Lachin (2016)]. Recently, Hu and Zhang (2004) developed a family of allocation functions defined for all $\xi \geq 0$ by

$$\begin{aligned} g^{(\xi)}(0,\rho) &= 1, \quad g^{(\xi)}(1,\rho) = 0, \\ g^{(\xi)}(\upsilon,\rho) &= \frac{\rho(\rho/\upsilon)^{\xi}}{\rho(\rho/\upsilon)^{\xi} + (1-\rho)((1-\rho)/(1-\upsilon))^{\xi}}, \end{aligned} \tag{1.5}$$

where υ is the proportion of patients assigned to treatment A, and ξ is nonnegative integer. for assigning treatments in DBCD and generalized the concept to more than two treatments. They also studied the asymptotic properties of the proportion of patients υ assigned to treatment A under certain regularity conditions. Hu et al. (2006) then derived a lower bound for the asymptotic variance of the allocation proportions for response-adaptive procedures under the assumption of normality. They discussed the issue of how to choose the best adaptive design procedure for a particular experiment and showed that the DBCD approach of Hu and Zhang (2004) is asymptotically best under certain conditions. More recently, Baldi Antognini and Zagoraiou (2012)

highlighted the importance of incorporating covariates in models for generating adaptive designs since the effectiveness of a treatment typically depends on the profile of a patient. Thus, they introduced the so-called reinforced DBCD. Nevertheless, they noted that further research is needed in analyzing the data obtained through response-adaptive the case of generalized linear models (GLM). Additional reviews of various types of response adaptive designs can be found in Rosenberger and Lachin (2002), and Chow and Chang (2008).

Covariate-Adaptive (CA) Designs

If the treatment selection for a new patient is based on the previous history of patients' treatment assignments and covariate profiles, as well as the covariate profile of the new patient, then the design criterion is called the CA designs in clinical trials. In CA designs, treatment assignments and covariates of previous patients as well as the covariates of the current patient are employed in the selection of treatments. It follows that $\pi_{iA,CA}(\boldsymbol{z}_i) = P[X_{iA} = 1|\mathcal{X}_{i-1}, C(\mathbf{z_{i-1}}), \boldsymbol{z}_i]$ for $i = 1, 2, \cdots, n$.

Rosenberger and Sverdlov (2008) note that "the goal of CA designs is to adaptively balance the covariate profiles of patients randomized to treatments". The objective of balancing treatments include achieving overall balance, balance within covariate margin, and balance within stratum, which is a combination of the levels of covariates. The advantage of achieving overall balance of treatment assignment is an increase in the power of hypothesis testing. If pre-stratification of covariates is possible at the initial stage of a clinical trial, then separate restricted randomization can be implemented to assign treatments within each stratum to attain these balancing goals. Such a design method is called the stratified permuted block (SPB) design. However, when the number of strata is large, then SPB design is impractical [Hu and Hu

(2012)]. Thus, CA designs were introduced to achieve these balancing goals when the number of strata is large. These CA designs are called minimization procedures [Rosenberger and Sverdlov (2008)]. Moreover, many authors such as Pocock and Simon (1975), Wei (1978), Hu and Hu (2012), and Lin and Su (2012) developed CA designs to achieve the balancing of covariates to treatments in sequential clinical trials.

The equal allocation method for treatment assignments does not depend on covariate profiles. Therefore, there is a chance that under equal allocation procedures, all treatment assignments to some covariate profiles will be of one category, say treatment A. When this happens, there is treatment imbalance within covariate margin or stratum. This will then decrease the power of the test for significance of all model parameters. However, the overall balancing goals ($n_A/n \to 0.5$ in probability) of treatment assignments can still be achieved by applying the equal allocation, where n_A is the number of patients assigned to treatment A from a total number of n patients.

The family of CA designs can be further partitioned into two sub families. The first family of CA designs is based on balancing treatment assignments over the covariate profiles by defining the measure of treatment imbalances. For instance, Pocock and Simon (1975) established a measure of marginal treatment imbalances that can be constructed by using differences between the number of treatments within the levels of covariates and appropriate weights. Hu and Hu (2012) developed a function for the measure of imbalance that includes three types of treatment imbalances: overall, marginal, and within stratum. Lin and Su (2012) developed a measure of treatment imbalance with empirical cumulative distribution functions by using observed covariates. Then, the minimization method can be applied to minimize these measures of treatment imbalances to achieve treatment balancing goals. In the minimization

method, when a new patient becomes available for treatment assignment, the selection probability p^* is chosen in such a way that there is a higher chance of assigning the treatment that will minimize the measure of treatment imbalances over covariate profiles. Moreover, an experimenter will decide on a value greater than 1/2 and less than or equal to one to assign to the selection probability p^* [see Chapter 9 of Rosenberger and Lachin (2016)]. In fact, if the experimenter picks the higher chance to be equal to one, then the allocation becomes deterministic.

Covariate-Adjusted Response-Adaptive (CARA) Designs

In CARA designs, the information on treatment assignments, covariates, and responses of previous patients as well as the covariates of the current patient are utilized in the new patient's treatment assignment. Thus, $\pi_{iA,CARA}(\boldsymbol{z}_i) = P[X_{iA} = 1|\mathcal{F}_{i-1}, C(\mathbf{z_{i-1}}), \boldsymbol{z}_i]$ for $i = 1, 2, \cdots, n$, where $\mathcal{F}_{i-1} = \sigma(\mathcal{X}_{i-1}, \mathcal{Y}_{i-1})$.

In recent years, clinical trials are mostly conducted in five different phases namely, phase 0 to phase IV. Also, a large number of subjects participate in phase III clinical trials. Thus, ethicists are consulted by a clinical trial research team to maintain human ethical standards [see Rosenberger, Vidyashankar and Agarwal (2001)]. High standard of ethics imply that more patients have to be treated by the best treatment. Response-adaptive designs are known to be effective in assigning more patients to the best treatment.

Consider the multi-center clinical trial for Stroke Prevention in Atrial Fibrillation Study [see Hart et al. (2003), Stroke Prevention in Atrial Fibrillation Investigators (1990)]. According to Hu et al. (2015), "Had the researchers ignored the factor of patients anticoagulation status, which can be used as a covariate in the statistical

model, their results would have produced a misleading conclusion that aspirin was generally more effective than placebo in preventing the occurrence of stroke". Thus, either aspirin or placebo is the globally best treatment to reduce the number of strokes in patients having atrial fibrillation. That is, aspirin is a better treatment only for a subgroup of patients and this clinical trial is an evidence for existence of drug by drug interactions. In fact, researchers have been exploring the invention of personalized medicine due to increasing availability of biomarkers and the observed heterogeneity of patients' responses to treatment [see Sverdlov (2016)]. Therefore, how does covariates in randomization of treatment assignments help to achieve **objective 2** in (1.1)? CARA design has been shown to be a method to attain this objective [Hu (2012)]. More details about CARA designs can be found in Rosenberger, Vidyashankar and Agarwal (2001), Zhang et al. (2007), and Hu et al. (2015).

Rosenberger, Vidyashankar and Agarwal (2001) considered the application of CARA designs based on a logit model. Zhang et al. (2007) extended the CARA designs to more than two treatments under the framework of a generalized linear model and discussed a detailed formulation of the design under linear and logistic regression models. But they considered separate models for each treatment while applying CARA design to select treatment assignments. Recently, Zhu (2015) noticed a drawback of the treatment assignment procedure of Zhang et al. (2007). Zhu (2015) notes that "it assumes that there are no common parameters for the two treatments of interest, thus estimating every parameter based on the data from just one treatment, which excludes many commonly used models". The model of Rosenberger, Vidyashankar and Agarwal (2001) has been implemented user-friendly statistical software, RStudio. Here, we will investigate the large sample behavior of the Maximum Likelihood Estimate (MLE) of model parameters in the model of Rosenberger, Vidyashankar and

Agarwal (2001) for CARA designs.

According to Hu (2012), "covariate information plays an important role in the design and analysis of clinical trials to develop personalized medicine". In fact, covariate information is used in the randomization of CA and CARA designs. A pre-selected objective of CA designs is achieving statistical efficiency; whereas achieving participating patients' ethics is the pre-selected objective of CARA designs. Achieving efficiency and ethics goals may be stand alone objective in these adaptive designs, however, in recent years, researchers have been interested in how to incorporate achieving both efficiency and ethics aims in a design. For example, Response-Adaptive Covariate-Adjusted (RACA) designs have incorporated components of efficiency and ethics in a design. This design is described in the next section.

Response-Adaptive Covariate-Adjusted (RACA) Designs

Response-Adaptive Covariate-Adjusted (RACA) designs were introduced by Ning and Huang (2010). In RACA design, the information on treatment assignments, covariates, and available responses of previous patients as well as the covariates of the current patient are used in the new patient's treatment assignment [Ning and Huang (2010)], however noted that the mechanism of CARA and RACA designs are completely different. To be specific the probability of assigning a new patient to treatment A is

$$\begin{aligned} \pi_{iA,RACA}(\boldsymbol{z}_i) &= P[X_{iA}=1|\mathcal{F}_{i-1}, C(\mathbf{z_{i-1}}), \boldsymbol{z}_i] \\ &= \frac{[\pi_{iA,RA}]^{\tau_1}[\pi_{iA,CA}]^{\tau_2}}{[\pi_{iA,RA}]^{\tau_1}[\pi_{iA,CA}]^{\tau_2} + [1-\pi_{iA,RA}]^{\tau_1}[1-\pi_{iA,CA}]^{\tau_2}} \end{aligned} \tag{1.6}$$

where $\pi_{iA,RA} = P[X_{iA}=1|\sigma(\mathcal{X}_{i-1}, \mathcal{Y}_{i-1})]$, and $\pi_{iA,CA}(\boldsymbol{z}_i) = P[X_{iA}=1|\mathcal{X}_{i-1}, C(\mathbf{z_{i-1}}), \boldsymbol{z}_i]$

for $i = 2, 3, \cdots, n$, τ_1 and τ_2 are the tuning parameters. It is clear that if $\tau_1 = 0$ and $\tau_2 = 1$, we get the pure CA design; if $\tau_1 = 1$ and $\tau_2 = 0$, we get the pure RA design. Thus, RACA designs are the combination of RA and CA designs. Also, the measures of efficiency and ethics can be controlled by the tuning parameters in RACA designs [see Lin et al. (2016)]. A difference of CARA whereas an initial model is required for its application, an initial model is not needed to apply RACA designs. According to Yuan and Liu (2011), when we incorporate CA design into a group sequential RA design, the resulting design combines the advantages of CA and RA design.

Response-adaptive designs create more severe treatment imbalances compared to equal allocation over covariate profiles. However, the focus of RACA designs is to achieve simultaneously **objective 1** and **objective 2** in (1.1). That is, measures of efficiency and ethics are accounted in RACA designs. In fact, recent research have focused on achieving efficiency and ethics in a design. For example Hu et al. (2015) developed a unified family of CARA designs using the components of efficiency and ethics in a design.

1.2 Statistical Models

It is well known that physical, psychological, and genetic factors can contribute to differences between patients in a clinical trial. Human beings are different among others with respect to several factors such as physical, psychological, and genetic factors. Thus, one can expect patient population to be heterogeneous. Therefore, statistical modelling is required to identify the effect of treatments on the response by including these heterogeneous factors, which are known as covariates. In what follows, we will discuss some of these candidate models, which can be used in the analysis of

a given data. The models are Generalized Linear Models, and Generalized Linear Mixed Effect Models.

1.2.1 Generalized Linear Models (GLMs)

The GLMs have been implemented in various disciplines such as agriculture, economics, engineering, medicine, and social sciences [Lindsey (1997)]. These class of models were first introduced by Nelder and Wedderburn (1972). Moreover, GLMs can be seen as an extension of classical linear models [McCullagh and Nelder (1983)].

Components of GLMs

A GLM has three components, namely, the probability distribution, the linear predictor, and the link function [McCullagh and Nelder (1983)]. We provide a description of these components only for binary responses, which are success or failure. Define, the response Y_i and treatment assignment X_{iA} of patient i for $i = 1, 2, \cdots, n$ by

$$Y_i = \begin{cases} 1 & \text{if treatment is} \\ & \text{a success,} \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad X_{iA} = \begin{cases} 1 & \text{if patient } i \text{ is assigned} \\ & \text{to treatment } A, \\ 0 & \text{otherwise.} \end{cases} \tag{1.7}$$

Let, $\mathbf{z_i'} = (z_{i1}, \cdots, z_{ip})$ be the p dimensional vector of observed covariates of patient i. We note that each component of $\mathbf{z_i}$ is either 0 or 1, for $i = 1, 2, \cdots, n$.

Following McCullagh and Nelder (1989) it is common to define a generalized linear model, in general, as

$$E(Y_i|x_{iA}, \mathbf{z_i}) \;=\; g^*(\boldsymbol{\theta}, x_{iA}, \mathbf{z_i}), \quad i = 1, 2, \cdots, n. \tag{1.8}$$

Let $\Lambda_i = E(Y_i|x_{iA}, \mathbf{z_i})$. We note that the GLM in (1.8) is similar to (1.3) without center effects. In this thesis, we consider the response to be binary. Therefore, we have

$$E(Y_i|x_{iA}, \mathbf{z_i}) \quad = \quad P(Y_i = 1|x_{iA}, \mathbf{z_i}).$$

Our purpose is to develop the relationship between the probability of the response, Λ, the treatment assignment and covariates of a patient, and treatment by covariate interactions $\mathbf{w}_i' = (x_{iA}, 1, \mathbf{z_i}, x_{iA}\mathbf{z_i'})$. We now describe the three components of a GLM.

1. **The probability distribution of** Y

   When we apply equal allocation for treatment assignments, the responses are independent. These responses are however not identically distributed because the distribution of Y_i depends on $\mathbf{w_i}$, where $\mathbf{w}_i' = (x_{iA}, 1, \mathbf{z_i'}, x_{iA}\mathbf{z_i'})$. Suppose that conditional on $\mathbf{w_i}$, the binary response Y_i follows the bernoulli distribution with probability mass function

   $$P(Y_i = y_i|\mathbf{w_i}) \quad = \quad \Lambda_i^{y_i}(1 - \Lambda_i)^{(1-y_i)}, \text{ for } y_i = 0, 1, \tag{1.9}$$

   where $\Lambda_i = E(Y_i|x_{iA}, \mathbf{z_i})$. Now when an adaptive design is applied to select treatment assignments, it creates dependency among responses because the treatment assignment will depend on accumulating data. However, the distribution assumption in (1.9) is still valid when adaptive designs are used as the treatment selection criteria.

2. **The linear predictor**

   We will assume that Λ is influenced through a linear combination of treatment

assignments and covariates given by

$$\eta_i \;=\; \mathbf{w}_i'\boldsymbol{\theta},$$

where $\boldsymbol{\theta}$ is a vector of model parameters including the main effect of treatments.

3. **Link function, Ψ**

   To examine the relationship between Λ and η, we need a function to construct the relationship. It is clear that, Λ takes values between 0 and 1, whereas, η can take values between $-\infty$ and ∞. The functions which define the relationship between Λ and η are called link functions in GLMs. In general, Λ and η are connected through a link function

$$\Psi(\Lambda_i) \;=\; \eta_i.$$

   Some candidate link functions are

   (i) the logit function

$$\Psi_1(\Lambda) \;=\; \ln[\Lambda/(1-\Lambda)],$$

   (ii) the probit function

$$\Psi_2(\Lambda) \;=\; \Phi^{-1}(\Lambda),$$

   (iii) the cauchit function

$$\Psi_3(\Lambda) \;=\; \tan\left[\pi\Lambda - \frac{\pi}{2}\right],$$

(iv) the complementary log-log function

$$\Psi_4(\Lambda) \;=\; \ln[-\ln(1-\Lambda)].$$

1.2.2 Generalized Linear Mixed Effect Models (GLMMs)

Generalized linear mixed models are very popular statistical models which have been used extensively in many areas of applications such as biomedical, clinical trials [Agresti and Hartzel (2000); Yaseri et al. (2014)], social science and agricultural science. A useful discussion on the theories and applications of GLMM can be found in, for instance Jiang (2007) and McCulloch, Searle and Neuhaus (2008).

When clinical trials are conducted in multiple centers, there is a possibility that center effects might influence the response. In fact, some unobservable causes may also affect responses of patients. For example, according to Kahan (2014), some possible latent effects are

(i) effect of variation in surgeons' skill between centers

(ii) effect of differences in guidelines of centers

(iii) these multiple centers might be selected from different countries.

It is clear that under equal allocation criterion for treatment assignment, responses of patients within a center might be correlated; but responses of patients between centers are independent. However, we can assume that conditional on center effect, responses of patients within a center are independent [Jiang (2007)]. Since J centers in the experiment are randomly selected from a large number of medical centers, we will assume that center effects are random and these center effects are heterogeneous

between centers. On the other hand, the fixed effects of covariates are common across centers. Thus, the GLMM approach we have developed in this thesis will be used to account for three features: (i) within center correlation, (ii) between center heterogeneity, and (iii) common fixed effects across centers [Tuerlinckx et al. (2006)]. In what follows, we will assume that these center effects follow a known distribution.

1.3 Statistical inference

In a clinical trial, treatment comparisons include the following: (i) finding the best dosage level of a drug for a group of patients who have certain characteristics or covariates, (ii) comparing a new drug with existing drugs for a disease. Also, an efficient treatment comparison through hypothesis testing is important for future patients. In general, an experimenter may claim that a new treatment is more effective than an existing treatment. Now, if there is no interaction between treatment and covariates, the experimenter may test significance for overall effect of treatment A compared to treatment B through the hypothesis,

$$H_0 : \beta_{A0} = 0 \qquad H_A : \beta_{A0} > 0, \tag{1.10}$$

where β_{A0} is the true effect of the treatment A (the new treatment) compared to treatment B (an existing treatment). Suppose the experimenter decides to reject H_0 when, in fact, H_0 is true. Then, though the new treatment does not lead to any improvement in the responses of patients compared to the existing treatment, the experimenter will recommend the new treatment. This type of error is called Type I error. It is important to verify that hypothesis testing procedures are able to control the Type I error. Thus, we will examine the size of the test through simulation studies

in Chapter 4.

Concerning the power of test, there exist three different methods, (i) exact method, (ii) approximation method, and (iii) simulation method, commonly applied for power analysis in statistical models and tests [Castelloe (2000)]. Castelloe (2000) notes that there is no standard procedure for power analysis in Generalized Linear Models. An approximation method for power analysis computation of logistic regression was first introduced by Whittemore (1981). However, this approach is only suitable for binary responses with rare events such as disease or death and covariates that are discrete or continuous. These covariates were also assumed to have a joint probability distribution function. Later, Self and Mauritsen (1988) developed an approximation procedure to power analysis based on score tests for GLMs. They implemented their approach to categorical covariates with a finite number of distinct covariate configurations. Later, Self et al. (1992) established a tool for power computation based on the likelihood ratio test. They compared their method with the method of Self and Mauritsen (1988) through simulation studies. Shieh (2000) carried out a simulation study to compare the method of Whittemore (1981) and Self et al. (1992) with various combination of response probabilities and covariate distribution in logistic regression models. Later, Shieh (2005) proposed a method for power computation based on Wald statistic. In fact, his/her method accommodates multiple parameters, and the flexibility of covariates configurations within the framework of GLMs. Lyles et al. (2007) developed a method for estimating conditional power for binary, ordinal, or count responses in GLMs.

Recently, Yi and Wang (2011) introduced the generalized score statistic method which is an extension of Rao's score test to response-adaptive designs. They demonstrated

that the generalized score statistic method performs well compared to the score test when applying adaptive design as the design criteria. We observe that the design criteria influences the power of chi-squared tests through the non-centrality parameter [see Hu and Rosenberger (2003)]. They identified three major influence factors: (i) the target allocation proportion, (ii) the randomization bias from target proportion, and (iii) the variance of randomization from target proportion. Yi and Wang (2009) examined the performance of response-adaptive designs for the assignments of patients to the best treatment and the power of the statistical test using a variance-penalized criterion. Implementing the transition probability procedure of a Markov chain, Yi (2013) established a method to compute the exact statistical power for the general class of response-adaptive designs. According to Chow and Chang (2008), a major or significant adaptation leads to the moving target population rather than the fixed target population. An effective statistical inference can be conducted considering these strategies: (i) sample size adjustment at interim, (ii) sample size allocation to treatments, (iii) delete, add, or change treatment arms, and (iv) change in study endpoints.

Recently, Ma et al. (2015) established a theoretical foundation for hypothesis testing for parameters in linear models under a large class of CA designs, which includes Pocock and Simon (1975) marginal method and stratified permuted block design. Also, they used the ordinary least squares method to estimate their model parameters. We will however develop a theoretical foundation for hypothesis testing when responses are binary, and the design criteria are CARA designs. Furthermore, we will apply the maximum likelihood method to estimate the parameters in the logit model. The objectives of CA designs are to adaptively balance the covariate profiles of patients randomized to treatments. As a result, the power of the hypothesis testing is maximized. On the other hand, the objectives of CARA designs are to minimize

the assignments of patients to inferior treatment. Therefore, the objectives of these two designs are completely different. That is, CA designs are less ethical than CARA designs but CA designs are more efficient than CARA designs. In fact, CA designs are more efficient than equal allocations [see Ma et al. (2015)].

Although Ma et al. (2015) used all the important covariates at the design stage, they dropped some covariate information in the final statistical inference. This will however lead to estimators of parameters that are generally inconsistent and biased. This will also affect the derivation for the distribution of the test statistic.

1.4 Motivation and contribution of this thesis

1.4.1 Motivation of this thesis

According to Hu et al. (2006), properties of the statistical methods under RA designs are well established under the assumption of a simple homogeneous parametric structure [see Wei (1978); Ivanova (2003); Eisele (1994); Hu and Zhang (2004); Yi and Wang (2007); Yi and Wang (2011); Rosenberger et al. (1997)]. The diversity of patients' characteristics were not considered in the investigation of RA designs. Rosenberger and Hu (2002) provided some conditions for the asymptotic normality of regression parameters in Generalized Linear Models (GLMs) that includes covariates of patients when sequential designs are applied to treatment assignments. When we add center effects in GLM, we have two modelling approaches, namely, GLM with fixed grouping effects and Generalized Linear Mixed Effect Model (GLMM) [see Agresti and Hartzel (2000) and Broström and Holmberg (2001)]. In this thesis, we examine the performance of RA randomizations when a Generalized Linear Mixed Effect Model (GLMM) is the parametric model in multi-center clinical trials.

Yi and Wang (2007) provided conditions for consistency and asymptotic normality of ML estimators for a class of adaptive designs under the assumption that only treatment influences the response. Rosenberger, Vidyashankar and Agarwal (2001) demonstrated that CARA designs reduces the number of patients assigned to inferior treatments through simulation studies. However, the large sample behavior of regression parameters of the model in Rosenberger, Vidyashankar and Agarwal (2001) was not discussed in the literature for CARA designs [see Basak et al. (2009)]. Thus, one of the objectives of this thesis is to establish the conditions for consistency and asymptotic normality of ML estimators for CARA designs.

Rosenberger, Vidyashankar and Agarwal (2001) discussed the most natural mapping, defined in § 3.7, with model based odds ratio for reducing the number of patients to inferior treatments. In fact, we can efficiently minimize the number of patients to inferior treatments if the initial model for applying CARA design is correctly specified. In other words, the assignment of patients to better treatments might be inefficient if an initial model is misspecified. However, when we apply the RA designs using proportions, the initial model is not required. In what follows, we will investigate the efficiency and ethics between RA and CARA designs.

Even though data collected from RA designs are dependent among responses, Yi and Wang (2011) justified that Wald, score, and the likelihood ratio tests can be used when the sample size is large. Moreover, they introduced the generalized score statistic for RA designs and concluded that the performance of Wald test is better than the score test, the generalized score test, and the likelihood ratio test. In fact, this

result motivated us to discuss the Wald test for adaptive designs considering the heterogeneity of patients' characteristics in the logit model.

Covariate-adaptive designs are usually applied to quickly achieve **objective 1** in (1.1). Furthermore, RA designs generate more severe imbalances of covariates over treatment arms when compared to equal allocation [Ning and Huang (2010)]. Hence, Ning and Huang (2010) introduced the Response-Adaptive Covariate-Adjusted (RACA) designs to achieve **objective 1** and **objective 2** in (1.1) simultaneously. Now, if treatment by covariate interactions exist, then CARA designs will reduce the number of patients assigned to inferior treatments. However, Ning and Huang (2010) discussed that "the identification of such interaction terms in regression models is not feasible unless the sample size is large". This observation motivated us to compare the performance of CARA designs with RA designs through simulation studies.

Hu and Rosenberger (2003) explored the relationship between the non-centrality parameter of the usual chi-square test for binary responses and the design's quantities: the target allocation proportion, the bias of the randomization procedure from that target, and the variability induced by the randomization process. However, they derived an expression for the non-centrality parameter under the assumption of a simple homogeneous parametric structure. These results motivated us to consider deriving the non-centrality parameter when we relax the assumption of simple uniform parametric structure and prove the relationship between the non-centrality parameter and target allocation proportion when a covariate is in the logit model.

1.4.2 Contribution of this thesis

In this section, we discuss our contribution to the literature through this thesis. In Chapter 2, we propose an approach to investigate the large sample theory of regression parameters of Generalized Linear Mixed Models (GLMMs) with sequential designs via the influence function method for familial data that was developed by Zhang and Oyet (2014). The performance of RA designs was investigated for GLMMs. Moreover, the influence function of ML estimates was derived and used to obtain a closed form expression of the asymptotic covariance of ML estimates, which does not currently exist in the literature. A new searching method for estimating the model parameters is introduced based on the influence function method. Also, we verified that this new iteration method works better than the Hessian matrix searching method through simulation studies and application to real data. The main results of Chapter 2 are outlined in a recent paper by Selvaratnam, Oyet, Yi and Gadag (2017).

In Chapter 3, we discuss the logit model for a general class of adaptive designs. The consistency and asymptotic normality of ML estimators of regression parameters of logit model was examined for adaptive designs. We reduced the strong regularity assumption that Fisher information and observed Fisher information matrices are positive definite matrices within a neighborhood that is close to vector of true parameters to a weaker assumption. This weak assumption is that the Fisher information matrix is positive define matrix at the vector of true parameters. We consider the odds-ratio-based limiting allocation that was introduced by Basak et al. (2009). We apply the Doubly adaptive Biased Coin Design (DBCD) to target this limiting allocation. Furthermore, the performance of this RA randomization is compared with CARA designs and equal allocation by simulation studies.

In Chapter 4, we investigate the power computation methods for a general class of adaptive designs. We examine the asymptotic distribution of the Wald test statistic in hypothesis testing under null and alternative hypothesis for adaptive designs considering the logit model as a true model. We investigate the performance of three designs: the RA design, the CARA design, and equal allocation. In this investigation, we examine the quantities: the number of patients assigned to the inferior treatment, the design variability, statistical power for testing hypotheses, and Type I error rates. We examine the non-centrality parameter of the Wald test for binary responses with the inclusion of heterogeneous patients' characteristic in a logit model. We proved that this non-centrality parameter is a function of the design proportions. Furthermore, we demonstrated that this function is concave when we assume only one covariate in the logit model.

Chapter 2

Estimation of a Generalized Linear Mixed Model for Response-Adaptive Designs in Multi-Center Clinical Trials

2.1 Introduction

The objective of comparing the effectiveness of two treatments in a clinical trial is not only gathering information about the relative effectiveness of the treatments but also assigning treatments to patients in a way that consider the wellbeing of patients; that is **objective 2** in (1.1). The response-adaptive designs are generally discussed in the literature to achieve this objective in a clinical trial assuming a simple homogeneous parametric structure. In this chapter, we examine the performance of response-adaptive designs when we assume the generalized linear mixed effect model (GLMM) is an ideal model.

The construction of likelihood functions under GLMM has largely assumed that covariates are fixed. However, generating treatment assignments through adaptive designs create dependency among responses. Consequently, we follow the approach of Rosenberger et al. (1997), Hu et al. (2006), and Yi and Wang (2007) to construct the likelihood function for GLMM with response-adaptive randomization by using the idea of likelihood function for sequential decision process. One difficulty commonly associated with using GLMMs is the problem of obtaining closed form expressions for the asymptotic variance of MLEs of the model parameters because the likelihood function contains integrals which cannot be solved analytically. Thus, asymptotic results in the literature have been based on the inverse of the Hessian matrix obtained from the likelihood function. In this chapter, we avoid the complications introduced by the integrals that cannot be solved analytically by using a Gauss-Hermite quadrature method to approximate the integrals in the likelihood function. This novel approach then allows us to exploit influence function techniques, as in Zhang and Oyet (2014), to establish the asymptotic properties of consistency and normality and to derive a closed form expression for the asymptotic covariance matrix of the MLEs. These results are outlined in Theorem 2.3.1 of § 2.3.

In § 2.2, we introduce the likelihood function for GLMM with response-adaptive randomization and apply the Gauss-Hermite quadrature to obtain the MLEs of the model parameters. The Gauss-Hermite quadrature has also been used by other authors to approximate integrals in generalized linear mixed models. See for instance Agresti and Hartzel (2000), Fahrmeir and Tutz (2001), Broström and Holmberg (2001), and Fortin (2013). We derive the influence function of the MLEs in § 2.3 and use this result to discuss the consistency and asymptotic normality of the MLEs. The results

of a simulation study and an application to real data are discussed in § 2.4 and § 2.5 respectively. We conclude this chapter with some remarks in § 2.6.

2.2 The Model and Parameter Estimation

Suppose that patients are recruited sequentially into a center and each patient is treated with one and only one of two treatments A and B. Suppose the number of patients within center j, n_j^*, is assumed to be fixed. Also, responses and treatment assignments of all n patients are defined in (1.2). Furthermore, let $\mathbf{X}_{ij} = (1, X_{ijA})'$ as the covariate associated with the binary response Y_{ij}. In what follows, we will assume that the response Y_{ij} is generated from a GLMM given by

$$\text{logit}[P(Y_{ij} = 1|u_j, x_{ijA})] = \mathbf{x}'_{ij}\boldsymbol{\beta} + u_j \tag{2.1}$$

where $\boldsymbol{\beta} = (\gamma_0, \beta_A)'$ and the random center effect u_j, $j = 1, 2, \cdots, J$, are independent normal random variables with mean zero and common variance σ^2.

For $i \geq 2$, let $\pi_{ijA} = P[X_{ijA} = 1|(x_{1jA}, y_{1j}), \cdots, (x_{(i-1)jA}, y_{(i-1)j})]$ and $\pi_{ijB} = 1 - \pi_{ijA}, 1 \leq j \leq J$. In response-adaptive designs, the random allocation rule $\boldsymbol{\pi}_j = \{\boldsymbol{\pi}_{ij}, i = 1, 2, \cdots n_j^*\}$ typically consists of a sequence of vector of probabilities, where $\boldsymbol{\pi}_{ij} = (\pi_{ijA}, \pi_{ijB})$. It is common to pre-specify the value of the allocation probability for the first patient in center j, $j = 1, 2, \cdots, J$, $\pi_{1jA} = P(X_{1jA} = 1)$ to, say a value of 1/2. Clearly, the fact that each $\boldsymbol{\pi}_{ij}$, $i \geq 2$ depends on previous treatment assignments and responses induce some dependency among the collected data. Thus, following Yi and Wang (2007), the unconditional likelihood function for

$\{(y_{ij}, x_{ijA});\ j = 1, 2, \cdots, J \text{ and } i = 1, 2, \cdots, n_j^*\}$ can be written as

$$\begin{aligned} L_R(\boldsymbol{\theta}) &= \prod_{j=1}^{J} \int_{-\infty}^{\infty} \left\{ \prod_{i=1}^{n_j^*} \left[\pi_{ijA}^{x_{ijA}} \pi_{ijB}^{(1-x_{ijA})} P(Y_{ij} = y_{ij} | u_j, x_{ijA}) \right] \frac{1}{\sqrt{2\pi}\sigma} \exp(-u_j^2/2\sigma^2) \right\} du_j \\ &= h(\tilde{\pi}) \prod_{j=1}^{J} \int_{-\infty}^{\infty} \left\{ \left[\prod_{i=1}^{n_j^*} \left[1 + \exp(-\mathbf{x}_{ij}'\boldsymbol{\beta} - u_j)\right]^{-y_{ij}} \left[1 + \exp(\mathbf{x}_{ij}'\boldsymbol{\beta} + u_j)\right]^{-(1-y_{ij})} \right] \right. \\ &\quad \left. \frac{1}{\sqrt{2}\sigma} \exp(-u_j^2/2\sigma^2) \right\} du_j, \end{aligned} \tag{2.2}$$

where $h(\tilde{\pi}) = \prod_{j=1}^{J} (1/\sqrt{\pi}) \prod_{i=1}^{n_j^*} \pi_{ijA}^{x_{ijA}} \pi_{ijB}^{(1-x_{ijA})}$ and $\boldsymbol{\theta}' = (\boldsymbol{\beta}', \sigma)$. In the special case of equal allocation, $h(\tilde{\pi}) \propto (1/2)^n$.

Next, we let $r = 1$ if a patient receives treatment A and $r = 0$ otherwise and apply the transformation $u_j^* = \frac{u_j}{\sqrt{2}\sigma}$ to (2.2), to obtain

$$\begin{aligned} L_R(\boldsymbol{\theta}) &= h(\tilde{\pi}) \prod_{j=1}^{J} \int_{-\infty}^{\infty} \left\{ \left(\prod_{r=0}^{1} \left[1 + \exp(-\gamma_0 - \beta_A r - \sqrt{2}\sigma u_j^*)\right]^{-n_{jrS}^*} \right. \right. \\ &\quad \left. \left. \left[1 + \exp(\gamma_0 + \beta_A r + \sqrt{2}\sigma u_j^*)\right]^{-n_{jrF}^*} \right) \exp(-[u_j^*]^2) \right\} du_j^*. \end{aligned} \tag{2.3}$$

In (2.3), n_{jrS}^* and n_{jrF}^* are the number of successes and failures respectively, in center j under treatment r. Given $\mathbf{y}_j = (y_{1j}, y_{2j}, \cdots, y_{n_j j})'$, the observed vector of responses from center j, the log-likelihood function of $\boldsymbol{\theta}$ can be written as

$$l_R(\boldsymbol{\theta}) = \ln h(\tilde{\pi}) + \sum_{j=1}^{J} \ln \int_{-\infty}^{\infty} f_{\mathbf{y}_j}(u_j^*, \boldsymbol{\theta}) e^{-[u_j^*]^2} du_j^*,$$

where $f_{\mathbf{y}_j}(u^*, \boldsymbol{\theta}) = \prod_{r=0}^{1} \left[1 + \exp(-\boldsymbol{\theta}'\mathbf{V}_r)\right]^{-n_{jrS}^*} \left[1 + \exp(\boldsymbol{\theta}'\mathbf{V}_r)\right]^{-n_{jrF}^*}$ with $\mathbf{V}_r' = (1, r, \sqrt{2}u^*)$.

We note that the integral in $l_R(\boldsymbol{\theta})$ cannot be solved analytically. Therefore, we have used the Gauss-Hermite quadrature method to approximate the integral in the following way. Let d^* be the number of sample points to be used in the approximation

and s_h the roots of the Hermite polynomial $Q_{d^*}(s)$ $(h = 1, 2, \cdots, d^*)$ with associated weights w_h. Then, by applying the Gauss-Hermite approximation to the integral in the log-likelihood function $l_R(\boldsymbol{\theta})$ we have

$$l_R(\boldsymbol{\theta}) \approx \ln h(\tilde{\pi}) + \sum_{j=1}^{J} \ln \left[\sum_{h=1}^{d^*} w_h f_{\mathbf{y}_j}(s_h, \boldsymbol{\theta}) \right]. \tag{2.4}$$

By Theorem 5.1.9 of Brass and Petras (2011), for each $j, j = 1, \cdots, J$, the Gauss-Hermite approximation converges to the exact integral as $d^* \to \infty$.

In our simulation studies and in our application to real data, the *optimx* function in R software was applied to solve the maximum likelihood estimating equation $\frac{\partial l_R(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0$, where

$$\frac{\partial l_R(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \approx \sum_{j=1}^{J} \left\{ \left[\sum_{h=1}^{d^*} w_h f_{\mathbf{y}_j}(s_h, \boldsymbol{\theta}) \right]^{-1} \left[\sum_{h=1}^{d^*} w_h \frac{\partial f_{\mathbf{y}_j}(s_h, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] \right\}, \tag{2.5}$$

with $\frac{\partial f_{\mathbf{y}_j}(u^*, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = f_{\mathbf{y}_j}(u^*, \boldsymbol{\theta}) \frac{\partial \ln f_{\mathbf{y}_j}(u^*, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$. Now,

$$\ln f_{\mathbf{y}_j}(u^*, \boldsymbol{\theta}) = -\sum_{r=0}^{1} \{ n^*_{jrS} \ln[1 + \exp(-\boldsymbol{\theta}' \mathbf{V}_r)] + n^*_{jrF} \ln[1 + \exp(\boldsymbol{\theta}' \mathbf{V}_r)] \},$$

$$\begin{aligned} \frac{\partial \ln f_{\mathbf{y}_j}(u^*, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= -\sum_{r=0}^{1} \left\{ n^*_{jrS} \frac{\exp(-\boldsymbol{\theta}' \mathbf{V}_r)}{[1 + \exp(-\boldsymbol{\theta}' \mathbf{V}_r)]} (-1) \mathbf{V}_r + n^*_{jrF} \frac{\exp(\boldsymbol{\theta}' \mathbf{V}_r)}{[1 + \exp(\boldsymbol{\theta}' \mathbf{V}_r)]} \mathbf{V}_r \right\} \\ &= \sum_{r=0}^{1} \left\{ n^*_{jrS} \frac{1}{[1 + \exp(\boldsymbol{\theta}' \mathbf{V}_r)]} \mathbf{V}_r - n^*_{jrF} \frac{1}{[1 + \exp(-\boldsymbol{\theta}' \mathbf{V}_r)]} \mathbf{V}_r \right\} \\ &= \sum_{r=0}^{1} \left\{ n^*_{jrS} [1 + \exp(\boldsymbol{\theta}' \mathbf{V}_r)]^{-1} \mathbf{V}_r - n^*_{jrF} [1 + \exp(-\boldsymbol{\theta}' \mathbf{V}_r)]^{-1} \mathbf{V}_r \right\}. \end{aligned} \tag{2.6}$$

The inputs to the *optimx* function in R were the log-likelihood function (2.4), the

gradient vector in (2.5) and the Hessian matrix $\frac{\partial^2 l_R(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$, given by

$$\frac{\partial^2 l_R(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \approx \sum_{j=1}^{J}\Bigg\{\left[\sum_{h=1}^{d^*} w_h f_{\mathbf{y}_j}(s_h,\boldsymbol{\theta})\right]^{-1}\left[\sum_{h=1}^{d^*} w_h \frac{\partial^2 f_{\mathbf{y}_j}(s_h,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] - \left[\sum_{h=1}^{d^*} w_h f_{\mathbf{y}_j}(s_h,\boldsymbol{\theta})\right]^{-2}\left[\sum_{h=1}^{d^*} w_h \frac{\partial f_{\mathbf{y}_j}(s_h,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]\left[\sum_{h=1}^{d^*} w_h \frac{\partial f_{\mathbf{y}_j}(s_h,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right]\Bigg\}, \tag{2.7}$$

where

$$\frac{\partial f_{\mathbf{y}_j}(u^*,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} = \left[\frac{\partial f_{\mathbf{y}_j}(u^*,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]'$$

$$\frac{\partial^2 f_{\mathbf{y}_j}(u^*,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = \frac{\partial \ln f_{\mathbf{y}_j}(u^*,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial f_{\mathbf{y}_j}(u^*,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} + \frac{\partial^2 \ln f_{\mathbf{y}_j}(u^*,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} f_{\mathbf{y}_j}(u^*,\boldsymbol{\theta}),$$

and

$$\frac{\partial^2 \ln f_{\mathbf{y}_j}(u^*,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = -\sum_{r=0}^{1}\left\{n^*_{jrS}\exp(\boldsymbol{\theta}'\mathbf{V}_r)[1+\exp(\boldsymbol{\theta}'\mathbf{V}_r)]^{-2} + n^*_{jrF}\exp(-\boldsymbol{\theta}'\mathbf{V}_r)[1+\exp(-\boldsymbol{\theta}'\mathbf{V}_r)]^{-2}\right\}\mathbf{V}_r\mathbf{V}_r'.$$

2.3 Asymptotic Properties

We mentioned earlier that the presence of integrals, which are functions of the unknown parameter vector $\boldsymbol{\theta}$, in the likelihood function for GLMMs has limited the ability of previous authors to verify conditions that are necessary for an appropriate central limit theorem to be valid. In this section, we avoid this difficulty by first deriving the influence function of the MLEs and then using the result to obtain the asymptotic properties of the MLEs. Zhang and Oyet (2014) applied a similar approach

to derive the asymptotic properties of the generalized quasi likelihood estimators of the parameters of a branching process model. In statistics, the influence function is the effect on an estimator of changing one point of the sample. The influence function of an estimate also indicates the sensitivity of the estimate to the observations Shen (1995). We note that in general, the number of patients within each center may not be equal. However, we will assume that the number of patients in each of the J centers is fixed but the number of centers J can be increased as many as possible.

Now, for fixed $i, i = 1, 2, \cdots, n_0^*$ where $n_0^* = \min(n_1^*, n_2^*, \cdots, n_J^*)$, let $\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_J$ be a sequence of independent and identically distributed random vectors with joint distribution function $F^*(\mathbf{t}), \mathbf{t} = (t_1, t_2, \cdots, t_{n_0^*})'$. Define the empirical distribution function of the observed responses $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_J$ as

$$F_J^*(\mathbf{t}) = \frac{1}{J} \sum_{j=1}^{J} \delta_{\mathbf{y}_j}(\mathbf{t}),$$

where $\delta_{\mathbf{y}_j}(\mathbf{t})$ is the indicator function

$$\delta_{\mathbf{y}_j}(\mathbf{t}) = \begin{cases} 1 & \text{if } y_{1j} \leq t_1, y_{2j} \leq t_2, \cdots, y_{n_0^* j} \leq t_{n_0^*}, \\ 0 & \text{otherwise.} \end{cases}$$

For an arbitrary distribution function G^* and $\epsilon \geq 0$, define $F_\epsilon^* = (1-\epsilon)F^* + \epsilon G^*$ to be the ϵ-contaminated distribution function of F^*. Then, using (2.5), the ML estimating equation $\dfrac{\partial l_R(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ can be written as

$$L^*(\mathbf{y}, \boldsymbol{\theta}, F_J^*) = \frac{\partial l_R(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \int \left[\sum_{h=1}^{d^*} w_h f_{\mathbf{y}}(s_h, \boldsymbol{\theta}(F_J^*)) \right]^{-1} \left[\sum_{h=1}^{d^*} w_h H_{\mathbf{y}}(s_h, \boldsymbol{\theta}(F_J^*)) \right] dF_J^*(\mathbf{y}) = \mathbf{0}, \tag{2.8}$$

where $H_{\mathbf{y}}(s_h, \boldsymbol{\theta}) = \dfrac{\partial f_{\mathbf{y}}(s_h, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$. That is, $\boldsymbol{\theta}(F_J^*) = \hat{\boldsymbol{\theta}}$ is a solution to (2.8). To simplify

notations, we let

$$A_{\epsilon\mathbf{y}} = \left[\sum_{h=1}^{d^*} w_h f_{\mathbf{y}}(s_h, \boldsymbol{\theta}(F_\epsilon^*))\right]^{-1} \text{and}, B_{\epsilon\mathbf{y}} = \left[\sum_{h=1}^{d^*} w_h H_{\mathbf{y}}(s_h, \boldsymbol{\theta}(F_\epsilon^*))\right]. \tag{2.9}$$

Then, with

$$A_{0\mathbf{y}} = A_{\epsilon\mathbf{y}}|_{\epsilon=0} \text{ and } B_{0\mathbf{y}} = B_{\epsilon\mathbf{y}}|_{\epsilon=0}, \tag{2.10}$$

we have that

$$\begin{aligned} L^*(\mathbf{y}, \boldsymbol{\theta}, F^*) &= \int \left[\sum_{h=1}^{d^*} w_h f_{\mathbf{y}}(s_h, \boldsymbol{\theta}(F^*))\right]^{-1} \left[\sum_{h=1}^{d^*} w_h H_{\mathbf{y}}(s_h, \boldsymbol{\theta}(F^*))\right] dF^* \\ &= \int A_{0\mathbf{y}} B_{0\mathbf{y}} dF^* = \mathbf{0}, \end{aligned} \tag{2.11}$$

since the true value of $\boldsymbol{\theta}$, $\boldsymbol{\theta}(F^*)$ is also a solution to the estimating equation. By definition, the influence function of $\hat{\boldsymbol{\theta}}$ at F^* is the Gâteaux derivative of $\boldsymbol{\theta}(F_\epsilon^*)$ at $\epsilon = 0$ and $G^* = \delta_{\mathbf{y}_j}(\mathbf{y})$.

Now, at F_ϵ^*, since(2.11) we have

$$\begin{aligned} L^*(\mathbf{y}, \boldsymbol{\theta}, F_\epsilon^*) &= \int \left[\sum_{h=1}^{d^*} w_h f_{\mathbf{y}}(s_h, \boldsymbol{\theta}(F_\epsilon^*))\right]^{-1} \left[\sum_{h=1}^{d^*} w_h H_{\mathbf{y}}(s_h, \boldsymbol{\theta}(F_\epsilon^*))\right] dF_\epsilon^* \\ &= \int A_{\epsilon\mathbf{y}} B_{\epsilon\mathbf{y}} dF_\epsilon^* = \mathbf{0} \end{aligned} \tag{2.12}$$

Therefore, the partial derivative of $L(\mathbf{y}, \boldsymbol{\theta}, F_\epsilon^*)$ with respect to ϵ is given by

$$
\begin{aligned}
&\frac{\partial L^*(\mathbf{y}, \boldsymbol{\theta}, F_\epsilon^*)}{\partial \epsilon} \\
&= \frac{\partial}{\partial \epsilon} \int A_{\epsilon\mathbf{y}} B_{\epsilon\mathbf{y}} dF_\epsilon^*(\mathbf{y}) \\
&= \int \frac{\partial A_{\epsilon\mathbf{y}} B_{\epsilon\mathbf{y}} dF_\epsilon^*(\mathbf{y})}{\partial \epsilon} \\
&= \int \frac{\partial A_{\epsilon\mathbf{y}} B_{\epsilon\mathbf{y}} dF_\epsilon^*(\mathbf{y})}{\boldsymbol{\theta}'(F_\epsilon^*)} \frac{\partial \boldsymbol{\theta}(F_\epsilon^*)}{\partial \epsilon} \\
&= \int \left[\frac{\partial A_{\epsilon\mathbf{y}}}{\partial \boldsymbol{\theta}'(F_\epsilon^*)} B_{\epsilon\mathbf{y}} dF_\epsilon^*(\mathbf{y}) + A_{\epsilon\mathbf{y}} \frac{\partial B_{\epsilon\mathbf{y}}}{\partial \boldsymbol{\theta}'(F_\epsilon^*)} dF_\epsilon^*(\mathbf{y}) + A_{\epsilon\mathbf{y}} B_{\epsilon\mathbf{y}} \frac{\partial dF_\epsilon^*(\mathbf{y})}{\partial \boldsymbol{\theta}'(F_\epsilon^*)} \right] \frac{\partial \boldsymbol{\theta}(F_\epsilon^*)}{\partial \epsilon} \\
&= \int \left[\frac{\partial A_{\epsilon\mathbf{y}}}{\partial \boldsymbol{\theta}'(F_\epsilon^*)} B_{\epsilon\mathbf{y}} \right] \frac{\partial \boldsymbol{\theta}(F_\epsilon^*)}{\partial \epsilon} dF_\epsilon^*(\mathbf{y}) + \int \left[A_{\epsilon\mathbf{y}} \frac{\partial B_{\epsilon\mathbf{y}}}{\partial \boldsymbol{\theta}'(F_\epsilon^*)} \right] \frac{\partial \boldsymbol{\theta}(F_\epsilon^*)}{\partial \epsilon} dF_\epsilon^*(\mathbf{y}) + \int \left[A_{\epsilon\mathbf{y}} B_{\epsilon\mathbf{y}} \right] d \frac{\partial F_\epsilon^*(\mathbf{y})}{\partial \epsilon} \\
&= \int \frac{\partial A_{\epsilon\mathbf{y}}}{\partial \boldsymbol{\theta}'(F_\epsilon^*)} B_{\epsilon\mathbf{y}} \frac{\partial \boldsymbol{\theta}(F_\epsilon^*)}{\partial \epsilon} dF_\epsilon^* + \int A_{\epsilon\mathbf{y}} \frac{\partial B_{\epsilon\mathbf{y}}}{\partial \boldsymbol{\theta}'(F_\epsilon^*)} \frac{\partial \boldsymbol{\theta}(F_\epsilon^*)}{\partial \epsilon} dF_\epsilon^* + \int A_{\epsilon\mathbf{y}} B_{\epsilon\mathbf{y}} (dG^* - dF^*),
\end{aligned}
$$

It follows that at $\epsilon = 0$, we have

$$
\begin{aligned}
&\left. \frac{\partial L^*(\mathbf{y}, \boldsymbol{\theta}, F_\epsilon^*)}{\partial \epsilon} \right|_{\epsilon=0} \\
&= \int \left[\frac{\partial A_{0\mathbf{y}}}{\partial \boldsymbol{\theta}'(F^*)} B_{0\mathbf{y}} \right] I(\boldsymbol{\theta}(G^*)) dF^* + \int \left[A_{0\mathbf{y}} \frac{\partial B_{0\mathbf{y}}}{\partial \boldsymbol{\theta}'(F^*)} \right] I(\boldsymbol{\theta}(G^*)) dF^* + \int A_{0\mathbf{y}} B_{0\mathbf{y}} (dG^* - dF^*) \\
&= \int \left[\frac{\partial A_{0\mathbf{y}}}{\partial \boldsymbol{\theta}'(F^*)} B_{0\mathbf{y}} \right] I(\boldsymbol{\theta}(G^*)) dF^* + \int \left[A_{0\mathbf{y}} \frac{\partial B_{0\mathbf{y}}}{\partial \boldsymbol{\theta}'(F^*)} \right] I(\boldsymbol{\theta}(G^*)) dF^* + \int A_{0\mathbf{y}} B_{0\mathbf{y}} dG^*,
\end{aligned}
$$

where $\frac{\partial A_{0\mathbf{y}}}{\partial \boldsymbol{\theta}'(F^*)} = \left. \frac{\partial A_{\epsilon\mathbf{y}}}{\partial \boldsymbol{\theta}'(F_\epsilon^*)} \right|_{\epsilon=0}$, $\frac{\partial B_{0\mathbf{y}}}{\partial \boldsymbol{\theta}'(F^*)} = \left. \frac{\partial B_{\epsilon\mathbf{y}}}{\partial \boldsymbol{\theta}'(F_\epsilon^*)} \right|_{\epsilon=0}$, and $I(\boldsymbol{\theta}(G^*)) = \left. \frac{\partial \boldsymbol{\theta}(F_\epsilon^*)}{\partial \epsilon} \right|_{\epsilon=0}$. Now to obtain the influence function for a given distribution function G^* and

$$
\begin{aligned}
I(\boldsymbol{\theta}(G^*)) &= -\left\{ \int \left[B_{0\mathbf{y}} \frac{\partial A_{0\mathbf{y}}}{\partial \boldsymbol{\theta}'(F^*)} + A_{0\mathbf{y}} \frac{\partial B_{0\mathbf{y}}}{\partial \boldsymbol{\theta}'(F^*)} \right] dF^* \right\}^{-1} \left[\int A_{0\mathbf{y}} B_{0\mathbf{y}} \, dG^* \right], \\
&= E_{G^*}[-C^{-1} B_{0\mathbf{y}} A_{0\mathbf{y}}] \qquad (2.13)
\end{aligned}
$$

where $C = \int \left[B_{0\mathbf{y}} \frac{\partial A_{0\mathbf{y}}}{\partial \boldsymbol{\theta}'(F^*)} + A_{0\mathbf{y}} \frac{\partial B_{0\mathbf{y}}}{\partial \boldsymbol{\theta}'(F^*)} \right] dF^*$. From (2.11) and (2.13), it is easy to see that at $G^* = F^*$ we have, $I(\boldsymbol{\theta}(F^*)) = E_{F^*}[-C^{-1} B_{0\mathbf{y}} A_{0\mathbf{y}}] = \mathbf{0}$. We also have that, at $G^* = \delta_{y_j}(\mathbf{y})$ the influence function of $\boldsymbol{\theta}$ at F^* becomes $IF(y, \boldsymbol{\theta}, F^*) = -C^{-1} B_{0\mathbf{y}} A_{0\mathbf{y}}$. In

order to derive the asymptotic properties of the MLE, $\boldsymbol{\theta}(F_J^*)$ of $\boldsymbol{\theta}$, we first use the Taylor series expansion to obtain the first order approximation of $\boldsymbol{\theta}(F_\epsilon^*)$, as

$$\begin{aligned}\boldsymbol{\theta}(F_\epsilon^*) &= \boldsymbol{\theta}(F_\epsilon^*)|_{\epsilon=0} + \left.\frac{\partial \boldsymbol{\theta}(F_\epsilon^*)}{\partial \epsilon}\right|_{\epsilon=0} \epsilon + \cdots \\ &= \boldsymbol{\theta}(F^*) + E_{G^*}[IF(\mathbf{y}, \boldsymbol{\theta}, F^*)]\ \epsilon + \cdots,\end{aligned}$$

for any ϵ and G^*. In particular, at $\epsilon = 1$ and $G^* = F_J^*$ we have

$$\boldsymbol{\theta}(F_J^*) - \boldsymbol{\theta}(F^*) \approx \frac{1}{J}\sum_{j=1}^{J} IF(\mathbf{y}_j, \boldsymbol{\theta}, F^*),$$

which can also be written as

$$\sqrt{J}(\boldsymbol{\theta}(F_J^*) - \boldsymbol{\theta}(F^*)) \approx \frac{1}{\sqrt{J}}\sum_{j=1}^{J} IF(\mathbf{y}_j, \boldsymbol{\theta}, F^*).$$

Therefore, by the strong law of large numbers $\frac{1}{J}\sum_{j=1}^{J} IF(\mathbf{y}_j, \boldsymbol{\theta}, F^*) \xrightarrow{a.s} \mathbf{0}$, since $E_{F^*}[IF(y, \boldsymbol{\theta}, F^*)] = \mathbf{0}$. Furthermore, $\frac{1}{\sqrt{J}}\sum_{j=1}^{J} IF(\mathbf{y}_j, \boldsymbol{\theta}, F^*) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$, by the central limit theorem where $\boldsymbol{\Sigma} = E_{F^*}[IF(y, \boldsymbol{\theta}, F^*)IF(y, \boldsymbol{\theta}, F^*)']$. These results are summarized in the following theorem.

Theorem 2.3.1. *Suppose $n_j^* = n_0^*$ for $j = 1, 2, \cdots, J$ and $\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_J$ be a sequence of independent and identically distributed random variables with distribution function $F^*(\mathbf{y})$. Let $F_J^*(\mathbf{y})$ be the empirical distribution function of the observed responses $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_J$. Then, using the Gauss-Hermite approximation defined in (2.4) such that $d^* \to \infty$ where d^* is the number of sample points in the Gauss-Hermite approximation, we have that*

(a) the influence function of $\boldsymbol{\theta}$ at F^* is $IF(y, \boldsymbol{\theta}, F^*) = -C^{-1}B_{0\mathbf{y}}A_{0\mathbf{y}}$,

(b) by the strong law of large numbers, $\boldsymbol{\theta}(F_J^*) \xrightarrow{a.s} \boldsymbol{\theta}(F^*)$ as $J \to \infty$, and

(c) by the central limit theorem, $\sqrt{J}(\boldsymbol{\theta}(F_J^*) - \boldsymbol{\theta}(F^*)) \xrightarrow{d} N\{\mathbf{0}, \boldsymbol{\Sigma}\}$ *as* $J \to \infty$*, with* $\boldsymbol{\Sigma} = E_{F^*}[IF(y, \boldsymbol{\theta}, F^*)IF(y, \boldsymbol{\theta}, F^*)']$.

In our simulation studies, we have used the plug-in estimate of the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ to compute an estimate of the variance of the MLE of $\boldsymbol{\theta}$. It can be easily verified that the plug-in estimate is given by

$$\begin{aligned} \hat{V}(\hat{\boldsymbol{\theta}}) &= \frac{1}{J} E_{F_J^*}[IF(\mathbf{y}, \boldsymbol{\theta}, F^*)IF(\mathbf{y}, \boldsymbol{\theta}, F^*)'], \\ &= \frac{1}{J^2} \sum_{j=1}^{J} IF(\mathbf{y}_j, \boldsymbol{\theta}, F^*)IF'(\mathbf{y}_j, \boldsymbol{\theta}, F^*), \end{aligned} \tag{2.14}$$

where $IF(\mathbf{y}_j, \boldsymbol{\theta}, F^*) = -C_S^{-1} B_{0\mathbf{y}_j} A_{0\mathbf{y}_j}$ and

$$C_S = \frac{1}{J} \sum_{j=1}^{J} \left[B_{0\mathbf{y}_j} \frac{\partial A_{0\mathbf{y}_j}}{\partial \boldsymbol{\theta}'(F^*)} + A_{0\mathbf{y}_j} \frac{\partial B_{0\mathbf{y}_j}}{\partial \boldsymbol{\theta}'(F^*)} \right]. \tag{2.15}$$

2.4 Simulation Studies

In this section, we examine the performance of the ML estimation procedure and the asymptotic results we derived in (2.2) and (2.3) respectively, through simulation studies. For this purpose, we have chosen $J = 25$, 50, and 100 centers with $n_0^* = n_j^* = 15$ patients in each of the centers in order to examine the performance of the estimates for small ($J = 25$), moderate ($J = 50$) and large ($J = 100$) values of J. Each data set used in computing the parameter estimates and values of the asymptotic variances was generated using various combinations of values of the parameters γ_0, β_A and σ. The values of the parameters we considered were (a) $\gamma_0 = -2, \beta_A = 3$, and $\sigma = 0.8$; (b) $\gamma_0 = -1, \beta_A = 4$, and $\sigma = 1$; and (c) $\gamma_0 = 0.5, \beta_A = 1$, and $\sigma = 0.8$. It is clear, from our discussions that the assignment of treatments to patients in each center will

depend on the response-adaptive design used in the clinical trials. The choice of design for treatment assignment will also affect the performance of the parameter estimates. Therefore, in our simulation studies we will compare the performance of the estimates under the completely randomized design (CRD), the RPW rule introduced by Wei and Durham (1978), and the doubly adaptive biased coin design targeting the RSIHR desired allocation to treatment A (Doubly$_\xi$).

The data generation procedure we applied starts with an initial assignment of treatment A to patient 1 and treatment B to patient 2. This initial assignment automatically determines the values of the covariate X_{ijA} to be $x_{1jA} = 1$ and $x_{2jA} = 0$. We then used equation (2.17) to compute the conditional probability that the response of the treatments assigned to patient i is a success given that they were treated in center j, j fixed. Suppose we denote this conditional probability by τ^*. Then, the response for patient i in center j was obtained by generating a Bernoulli observation with probability of success τ^*. In order to generate the response for the next patient in the same center, we first used one of the designs (CRD, RPW or Doubly$_\xi$) to assign a treatment to the patient. As noted earlier, once the treatment has been assigned, the value of the covariate X_{ijA} is known. The process of generating the ith response and treatment assignment for a fixed design was then repeated until all patients in a given center j have been treated. The entire data generating process was repeated to generate data for each of the 25, 50 or 100 centers.

Under CRD, patients have an equal chance of receiving either treatment A or treatment B. That is, for a fixed center $j = 1, 2, \cdots, J$, $P(X_{ijA}) = 1/2$, $i = 3, 4, \cdots, n_j^*$. So, to determine the treatment for the next available patient we started by generating a Bernoulli observation with probability of success 1/2. We then assigned treatment

Table 2.1: Proportion of patients assigned to treatment A in all centers (υ_A) based on RPW, CRD and Doubly$_\xi$ procedures from 3000 simulations. Simulated means (SM) and simulated standard errors (SSE), estimated standard errors based on the Hessian matrix (ESE), estimated standard errors based on the influence function (IESE), mean squared error (MSE) and 95% coverage probability [C95%] based on normal distribution for the MLEs of model parameters, with covariate effects $\gamma_0 = -2$, $\beta_A = 3$, $\sigma = 0.8$, $J = 100$ and $n_0^* = 15$.

Design	Quantity	$\hat{\gamma}_0$	$\hat{\beta}_A$	$\hat{\sigma}$	υ_A
RPW	SM	-2.0110	3.0068	0.7914	0.6514
	SSE	0.1612	0.1669	0.1034	0.0128
	ESE	0.1617	0.1664	0.1047	-
	IESE	0.1610	0.1649	0.1033	-
	MSE	0.0523	0.0556	0.0218	-
	C95%	0.9490	0.9530	0.9487	-
CRD	SM	-2.0043	3.0051	0.7850	0.5002
	SSE	0.1430	0.1535	0.1051	0.0120
	ESE	0.1423	0.1565	0.1066	-
	IESE	0.1418	0.1550	0.1048	-
	MSE	0.0408	0.0481	0.0227	-
	C95%	0.9493	0.9523	0.9527	-
Doubly$_0$	SM	-2.0064	3.0090	0.7906	0.5876
	SSE	0.1516	0.1631	0.1070	0.0120
	ESE	0.1518	0.1613	0.1054	-
	IESE	0.1511	0.1600	0.1042	-
	MSE	0.0462	0.0527	0.0227	-
	C95%	0.9483	0.9470	0.9473	-
Doubly$_3$	SM	-2.0100	3.0080	0.7880	0.6172
	SSE	0.1587	0.1668	0.1048	0.0067
	ESE	0.1549	0.1626	0.1046	-
	IESE	0.1541	0.1612	0.1032	-
	MSE	0.0494	0.0544	0.0221	-
	C95%	0.9433	0.9420	0.9513	-
Doubly$_7$	SM	-2.0108	3.0086	0.7858	0.6221
	SSE	0.1580	0.1650	0.1036	0.0059
	ESE	0.1555	0.1630	0.1044	-
	IESE	0.1547	0.1618	0.1029	-
	MSE	0.0493	0.0539	0.0218	-
	C95%	0.9490	0.9507	0.9520	-

Table 2.2: Proportion of patients assigned to treatment A in all centers (υ_A) based on RPW, CRD and Doubly$_\xi$ DBCD procedures from 3000 simulations. Simulated means (SM) and simulated standard errors (SSE), estimated standard errors based on the Hessian matrix (ESE), estimated standard errors based on the influence function (IESE), mean squared error (MSE) and 95% coverage probability [C95%] based on normal distribution for the MLEs of model parameters, with covariate effects $\gamma_0 = 0.5$, $\beta_A = 1$, $\sigma = 0.8$, $J = 50$ and $n_0^* = 15$.

Design	Quantity	$\hat{\gamma}_0$	$\hat{\beta}_A$	$\hat{\sigma}$	υ_A
RPW	SM	0.4992	1.0058	0.7849	0.5615
	SSE	0.1736	0.1907	0.1482	0.0280
	ESE	0.1688	0.1858	0.1441	-
	IESE	0.1687	0.1834	0.1409	-
	MSE	0.0589	0.0710	0.0431	-
	C95%	0.9450	0.9407	0.9427	-
CRD	SM	0.4983	1.0027	0.7818	0.5001
	SSE	0.1606	0.1807	0.1455	0.0167
	ESE	0.1607	0.1807	0.1426	-
	IESE	0.1610	0.1780	0.1394	-
	MSE	0.0519	0.0653	0.0420	-
	C95%	0.9530	0.9527	0.9433	-
Doubly$_0$	SM	0.5007	1.004	0.7801	0.5242
	SSE	0.1650	0.1816	0.1429	0.0182
	ESE	0.1627	0.1813	0.1429	-
	IESE	0.1630	0.1788	0.1398	-
	MSE	0.0540	0.0659	0.0414	-
	C95%	0.9440	0.9480	0.9470	-
Doubly$_3$	SM	0.5018	1.006	0.7823	0.5314
	SSE	0.1651	0.1775	0.1440	0.0098
	ESE	0.1628	0.1790	0.1431	-
	IESE	0.1632	0.1768	0.1400	-
	MSE	0.0540	0.0636	0.0416	-
	C95%	0.9387	0.9493	0.9457	-
Doubly$_7$	SM	0.5022	1.005	0.7828	0.5320
	SSE	0.1647	0.1795	0.1443	0.0084
	ESE	0.1629	0.1787	0.1430	-
	IESE	0.1634	0.1766	0.1396	-
	MSE	0.0539	0.0642	0.0417	-
	C95%	0.9400	0.9490	0.9450	-

Table 2.3: Proportion of patients assigned to treatment A in all centers (υ_A) based on RPW, CRD and Doubly$_\xi$ DBCD procedures from 3000 simulations. Simulated means (SM) and simulated standard errors (SSE), estimated standard errors based on the Hessian matrix (ESE), estimated standard errors based on the influence function (IESE), mean squared error (MSE) and 95% coverage probability [C95%] based on normal distribution for the MLEs of model parameters, with covariate effects $\gamma_0 = -2$, $\beta_A = 3$, $\sigma = 0.8$, $J = 25$ and $n_0^* = 15$.

Design	Quantity	$\hat{\gamma}_0$	$\hat{\beta}_A$	$\hat{\sigma}$	υ_A
RPW	SM	-2.0326	3.0291	0.7569	0.6514
	SSE	0.3353	0.3478	0.2228	0.0263
	ESE	0.3262	0.3375	0.2180	-
	IESE	0.3194	0.3281	0.2048	-
	MSE	0.2217	0.2370	0.1013	-
	C95%	0.9500	0.9570	0.9710	-
CRD	SM	-2.0200	3.0250	0.7629	0.4997
	SSE	0.2820	0.3182	0.2300	0.0244
	ESE	0.2862	0.3165	0.2250	-
	IESE	0.2821	0.3065	0.2097	-
	MSE	0.1633	0.2027	0.1103	-
	C95%	0.9520	0.9527	0.9787	-
Doubly$_0$	SM	-2.0237	3.0257	0.7617	0.5870
	SSE	0.3083	0.3284	0.2278	0.0235
	ESE	0.3054	0.3263	0.2205	-
	IESE	0.2990	0.3144	0.2060	-
	MSE	0.1905	0.2159	0.1045	-
	C95%	0.9520	0.9550	0.9737	-
Doubly$_3$	SM	-2.0281	3.0292	0.7617	0.6168
	SSE	0.3137	0.3304	0.2253	0.0131
	ESE	0.3123	0.3293	0.2186	-
	IESE	0.3060	0.3180	0.2050	-
	MSE	0.1983	0.2194	0.1029	-
	C95%	0.9513	0.9537	0.9680	-
Doubly$_7$	SM	-2.0315	3.0296	0.7583	0.6220
	SSE	0.3163	0.3278	0.2223	0.0116
	ESE	0.3136	0.3302	0.2175	-
	IESE	0.3070	0.3187	0.2048	-
	MSE	0.2010	0.2184	0.0999	-
	C95%	0.9507	0.9570	0.9667	-

A to the patient if the outcome was 1. In this case, $x_{ijA} = 1$. Alternatively, we assigned treatment B to the patient if the outcome was 0. In this case, $x_{ijA} = 0$. In order to determine the treatment allocation under DBCD targeting the RSIHR desired allocation to treatment A (Doubly$_\xi$), we used available information and the expression (1.4) proposed by Rosenberger, Stallard, Ivanova, Harper and Ricks (2001) to compute an estimate of the desired allocation proportion ρ to treatment A. We then used the estimated desired allocation proportion ρ to calculate the DBCD treatment allocation function (1.5) for treatment A for each value of the parameter $\xi =$ 0, 3 and 7, where υ is the proportion of patients assigned to treatment A. In a sequential process, the values of υ and ρ in (1.4) were separately updated for each center.

The results in Tables 2.1 - 2.5 show that the ML estimating procedure we outlined in 2.2 performed well by consistently estimating the model parameters. In addition to estimating the parameters, we also computed the standard errors of the estimators in three ways for the purpose of comparison and validating our theoretical results. First, we computed the variance of the estimates obtained from the 3000 simulations. We have denoted the standard errors obtained by taking the square root of the variance computed from this method by SSE. Secondly, we used the inverse of the Hessian matrix obtained at the final stage of the maximization process to compute the standard errors and finally, we used the plug-in estimate of variance based on the influence function in (2.14) to compute the standard errors. The estimates of the standard errors shown in Tables 2.1 - 2.5 are very similar in magnitude. In the tables, the estimated standard errors denoted by ESE are based on the inverse of the Hessian matrix and the influence function based estimated standard errors are denoted by IESE. Bias is calculated by the difference between ML estimate and true parameter. Then, MSE is computed by the sum of squared bias and ESE. Since the values of the

estimated standard errors from all three methods are very similar, we used the ESE to construct 95% confidence intervals for the parameters γ_0, β_A, and σ based on the assumption of normality. These results are also shown in Tables 2.1 - 2.5. In all cases, the estimated coverage probability under the assumption of normality was found to be approximately 95% under each of the designs we considered. There was however some slight deviations in the coverage probability estimate for the variance $\hat{\sigma}$ when $J = 25$ (Table 2.3). These simulation results validate the influence function based proof of asymptotic normality of the parameter estimates we outlined in the section 2.3.

The values of υ_A in Tables 2.1 - 2.5 show that the two response-adaptive designs namely, RPW and DBCDs, consistently assigned the treatment with high probability of success to more patients which is a morally and ethically desirable outcome. That is, υ_A is larger than 50% in both cases, whereas $\upsilon_A = 50\%$ under CRD. We had mentioned earlier that the simulated mean squared errors for all response-adaptive designs considered in this chapter were similar in value. Our simulation results also show that the design variability under DBCD becomes increasingly smaller than those under RPW and CRD as ξ becomes larger in magnitude. For instance, in Table 2.1, the design variability under RPW and CRD were 0.0128 and 0.0120 respectively, whereas, the design variability under DBCD for $\xi = 0, 3, 7$ were 0.0120, 0.0067 and 0.0059 respectively. A similar pattern can be seen in the values of the design variability in Tables 2.2 - 2.5. These results agree with the conclusion of Hu et al. (2006) who noted that the design variability of the allocation proportion can attain its lower bound for larger values of ξ. Taken together, these simulation results demonstrate that the Gauss-Hermite quadrature method was effective in approximating the integral in the log-likelihood function.

Table 2.4: Proportion of patients assigned to treatment A in all centers (υ_A) based on RPW, CRD and Doubly$_\xi$ DBCD procedures from 3000 simulations. Simulated means (SM) and simulated standard errors (SSE), estimated standard errors based on the Hessian matrix (ESE), estimated standard errors based on the influence function (IESE), mean squared error (MSE) and 95% coverage probability [C95%] based on normal distribution for the MLEs of model parameters, with covariate effects $\gamma_0 = -1$, $\beta_A = 4$, $\sigma = 1$, $J = 100$ and $n_0^* = 15$.

Design	Quantity	$\hat{\gamma}_0$	$\hat{\beta}_A$	$\hat{\sigma}$	υ_A
RPW	SM	-0.9977	3.9989	0.9853	0.6958
	SSE	0.1569	0.2177	0.1364	0.0156
	ESE	0.1586	0.2163	0.1363	-
	IESE	0.1580	0.2142	0.1333	-
	MSE	0.0499	0.0944	0.0374	-
	C95%	0.9477	0.9530	0.9523	-
CRD	SM	-1.0033	4.0092	0.9902	0.4998
	SSE	0.1354	0.2082	0.1288	0.0121
	ESE	0.1365	0.2088	0.1261	-
	IESE	0.1364	0.2069	0.1242	-
	MSE	0.0371	0.0872	0.0326	-
	C95%	0.9530	0.9540	0.9487	-
Doubly$_0$	SM	-1.0051	4.0099	0.9902	0.5846
	SSE	0.1424	0.2094	0.1330	0.0124
	ESE	0.1440	0.2106	0.1301	-
	IESE	0.1436	0.2082	0.1278	-
	MSE	0.0412	0.0884	0.0348	-
	C95%	0.9610	0.9507	0.9420	-
Doubly$_3$	SM	-1.0049	4.0090	0.9913	0.6104
	SSE	0.1453	0.2088	0.1338	0.0071
	ESE	0.1460	0.2108	0.1312	-
	IESE	0.1455	0.2085	0.1285	-
	MSE	0.0426	0.0883	0.0352	-
	C95%	0.9460	0.9533	0.9460	-
Doubly$_7$	SM	-1.0051	4.0084	0.9904	0.6140
	SSE	0.1449	0.2093	0.1362	0.0064
	ESE	0.1463	0.2110	0.1313	-
	IESE	0.1458	0.2088	0.1288	-
	MSE	0.0426	0.0886	0.0360	-
	C95%	0.9523	0.9533	0.9380	-

Table 2.5: Proportion of patients assigned to treatment A in all centers (υ_A) based on RPW, CRD and Doubly$_\xi$ DBCD procedures from 3000 simulations. Simulated means (SM) and simulated standard errors (SSE), estimated standard errors based on the Hessian matrix (ESE), estimated standard errors based on the influence function (IESE), mean squared error (MSE) and 95% coverage probability [C95%] based on normal distribution for the MLEs of model parameters, with covariate effects $\gamma_0 = 0.5$, $\beta_A = 1$, $\sigma = 0.8$, $J = 100$ and $n_0^* = 20$.

Design	Quantity	$\hat{\gamma}_0$	$\hat{\beta}_A$	$\hat{\sigma}$	υ_A
RPW	SM	0.5023	0.9997	0.7922	0.5717
	SSE	0.1124	0.1151	0.0895	0.0190
	ESE	0.1116	0.1136	0.0901	-
	IESE	0.1117	0.1130	0.0894	-
	MSE	0.0251	0.0261	0.0162	-
	C95%	0.9467	0.9500	0.9453	-
CRD	SM	0.5012	1.0012	0.7920	0.4999
	SSE	0.1076	0.1108	0.0900	0.0106
	ESE	0.1064	0.1101	0.0892	-
	IESE	0.1064	0.1092	0.0881	-
	MSE	0.0229	0.0244	0.0162	-
	C95%	0.9447	0.9487	0.9480	-
Doubly$_0$	SM	0.5035	1.0010	0.7920	0.5261
	SSE	0.1091	0.1120	0.0914	0.0116
	ESE	0.1077	0.1105	0.0896	-
	IESE	0.1078	0.1095	0.0886	-
	MSE	0.0236	0.0247	0.0165	-
	C95%	0.9447	0.9507	0.9407	-
Doubly$_3$	SM	0.5040	1.0002	0.7919	0.5328
	SSE	0.1090	0.1067	0.0914	0.0063
	ESE	0.1076	0.1092	0.0897	-
	IESE	0.1078	0.1085	0.0887	-
	MSE	0.0235	0.0233	0.0165	-
	C95%	0.9473	0.9527	0.9413	-
Doubly$_7$	SM	0.5034	1.0009	0.7919	0.5335
	SSE	0.1086	0.1057	0.0898	0.0052
	ESE	0.1076	0.1090	0.0897	-
	IESE	0.1078	0.1083	0.0888	-
	MSE	0.0234	0.0231	0.0162	-
	C95%	0.9460	0.9560	0.9467	-

2.5 Application to Real Data

Through simulation, we have shown in § 2.4, that the asymptotic standard errors of the estimated parameters based on the influence function (IESE) and the Hessian Matrix (ESE), were similar in value to the simulated standard errors (SSE) (see Tables 2.1 - 2.5). In this section, we apply the techniques proposed in this chapter to real data obtained from a multi-center clinical trial. The trials were conducted at eight different centers for the purpose of comparing two cream preparations (active drug, control) for treating an infection (Beitler and Landis (1985)). GLMM can be used conceptually in a multi-center clinical trial when a sample of centers is selected from a large number of centers and the number of centers selected is reasonably large. Grizzle (1987) suggested that occasionally, it is preferable to consider center effects as random effects rather than fixed effects even if the sample of centers was not randomly chosen. He also noted that the number of centers selected has to be large enough for consistent estimation of the model parameters in GLMM. Though the data in this application was collected from only eight medical centers, a total of 273 patients participated in the study. The data, taken from Agresti and Hartzel (2000) are summarized in Table 2.6. Agresti and Hartzel (2000) had considered several models for estimating the treatment effect. However, we will discuss only the GLM with fixed center effects and the GLMM. The model with fixed center effects we considered is given by,

$$\text{logit}[P(Y_{ij}=1|x_{ijA})] = \beta_F x_{ijA} + u_j^F,\ i = 1, 2, \cdots, n_j^* \text{ and } j = 1, 2, \cdots, J, \quad (2.16)$$

where the center effects $u_j^F,\ j = 1, 2, \cdots, J$ are assumed to be fixed and β_F is the treatment effect that is assumed to be constant over centers. We also considered a

GLMM with logit link function defined by

$$\text{logit}[P(Y_{ij}=1|u_j, x_{ijA})] \;=\; \gamma_0 + \beta_A x_{ijA} + u_j,\; i = 1, 2, \cdots, n_j^* \text{ and } j = 1, 2, \cdots, J, \tag{2.17}$$

where γ_0 is the overall intercept, β_A is the treatment effect that is assumed to be constant over centers, and the random center effect u_j, $j = 1, 2, \cdots, J$, are independent normal random variables with mean zero and common center variance σ^2.

Table 2.6: Clinical trial relating treatment to response for eight centres.

Centre	Treatment	Response		Total
		Success	Failure	
1	Drug	11	25	36
	Control	10	27	37
2	Drug	16	4	20
	Control	22	10	32
3	Drug	14	5	19
	Control	7	12	19
4	Drug	2	14	16
	Control	1	16	17
5	Drug	6	11	17
	Control	0	12	12
6	Drug	1	10	11
	Control	0	10	10
7	Drug	1	4	5
	Control	1	8	9
8	Drug	4	2	6
	Control	6	1	7
Total	Drug	55	75	130
	Control	47	96	143

Though models (2.16) and (2.17) appear to be similar in structure the methods for estimating the model parameters are completely different. In fact, the overall intercept

γ_0 in (2.17) cannot be estimated separately in (2.16) because the overall intercept is included in center effects. Indeed, our main objective is to estimate the treatment effect efficiently and not how to estimate the overall intercept and center effects separately. We will apply two separate iteration methods for estimating the parameters of the models. The first method is the usual approach based on the gradient vector and the Hessian matrix defined in the section 2.2 of this chapter. We introduce a new approach, in this section, based entirely on the influence function given by

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{new}} \;=\; \hat{\boldsymbol{\theta}}_{\boldsymbol{old}} + \frac{1}{J}\sum_{j=1}^{J} IF(\mathbf{y}_j, \boldsymbol{\theta}, F^*)|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\boldsymbol{old}}}, \tag{2.18}$$

where $\frac{1}{J}\sum_{j=1}^{J} IF(\mathbf{y}_j, \boldsymbol{\theta}, F^*) = -(JC_S)^{-1}\sum_{j=1}^{J} B_{0\mathbf{y}_j} A_{0\mathbf{y}_j}$ with C_S given by (2.15) and $A_{0\mathbf{y}_j}$, and $B_{0\mathbf{y}_j}$ defined by (2.10).

We observe that in Table 2.6, the responses of all patients assigned to the control group in Centers 5 and 6 were failures. As a result, following Agresti and Hartzel (2000) we will examine whether centers with 0 successes have any influence on the results of descriptive and inferential analyses. To investigate the effect which centers with 0 successes for treatment has on inference, we will examine the $p-$values of the Wald test for treatment effect. More specifically, the Wald Statistic (see Lyles, Lin and Williamson (2007)) $\chi_c^2 = \hat{\beta}_A^2/Var(\hat{\beta}_A)$ and the corresponding $p-$value for testing $H_0 : \beta_A = 0$ versus the alternative $H_a : \beta_A > 0$ will be computed under three scenarios, namely, (a) **DATA 1** which is the original data shown in Table 2.6; (b) **DATA 2** which excludes data from centers with zero responses in the control group; and (c) **DATA 3** which combines the data from Centers 5, 6 and 7. The results from this analysis are shown in Table 2.7.

Table 2.7: Values of the $D-$optimality criterion, the Wald Statistic and their correponding $p-$values for testing the significance of treatment effect computed using **DATA 1**, **DATA 2** and **DATA 3**, and estimated parameters and their standard errors based on (a) the Hessian matrix denoted by Estimate and ESE respectively; (b) the influence function approach in (2.18) denoted by IEstimate and IESE respectively for the parameters in model (2.17) and the parameters in model (2.16) denoted by FEstimate and FESE respectively.

Data	Quantity	$\hat{\gamma}_0$	$\hat{\beta}_A/\hat{\beta}_F$	$\hat{\sigma}$	$D-$optimality	Wald Statistics	$p-$value
DATA 1	(a) Estimate	-1.1894	0.7385	1.3971	203.5655	6.044	0.014
	ESE	0.5803	0.3004	0.4196			
	(b) IEstimate	-1.1894	0.7385	1.3971	564.1102	5.806	0.016
	IESE	0.6367	0.3065	0.2419			
	FEstimate	-	0.7766	-	-	-	-
	FESE		0.3067				
DATA 2	(a) Estimate	-0.6605	0.5548	1.2720	166.1339	3.102	0.078
	ESE	0.5834	0.3150	0.4372			
	(b) IEstimate	-0.6605	0.5549	1.2720	829.5909	3.970	0.046
	IESE	0.5819	0.2785	0.2238			
	FEstimate	-	0.5754	-	-	-	-
	FESE		0.3205				
DATA 3	(a) Estimate	-0.8096	0.7440	1.3297	166.2647	6.150	0.013
	ESE	0.6167	0.3000	0.4338			
	(b) IEstimate	-0.8097	0.7441	1.3297	656.4324	5.427	0.020
	IESE	0.6852	0.3194	0.2102			
	FEstimate	-	0.7774	-	-	-	-
	FESE		0.3050				

We had mentioned earlier that our main objective is to estimate treatment effect efficiently while maximizing the number of patients assigned to the best treatment. The results in Table 2.7 show that estimates of the treatment effects, overall intercept and the standard deviation σ obtained from the two methods are about the same in magnitude within **DATA 1**, **DATA 2** and **DATA 3**. That is, the presence of zero response does not appear to affect the parameter estimates. However, the standard errors of the estimates obtained from the two methods are, in most cases, different in value with noticeably larger differences in $SE(\hat{\gamma}_0)$ and $SE(\hat{\sigma})$. For instance, $ESE(\hat{\gamma}_0)$ = 0.5803 and $IESE(\hat{\gamma}_0)$ = 0.6367 and $ESE(\hat{\sigma})$ = 0.4196 and $IESE(\hat{\sigma})$ = 0.2419, whereas $ESE(\hat{\beta}_A/\hat{\beta}_F)$ = 0.3004 and $IESE(\hat{\beta}_A/\hat{\beta}_F)$ = 0.3065 for **DATA 1**. The same pattern is replicated in the results for **DATA 3**. In these cases, we observe that $IESE(\hat{\gamma}_0) > ESE(\hat{\gamma}_0)$ while $IESE(\hat{\sigma}) < ESE(\hat{\sigma})$. These results are reasonable since $\gamma_0 + u_j = u_j^F$ for $j = 1, 2, \cdots, J$. Also, the estimation method for GLMM estimates these effects separately while these effects cannot be separated in the model (2.16). It is interesting to note that the patterns in the standard errors of $\hat{\gamma}_0$ and $\hat{\beta}_A/\hat{\beta}_F$ described above are reversed when the data for centers with zero responses in the control group are excluded (**DATA 2**) when estimating the model parameters. In this case, $IESE(\hat{\gamma}_0)$ = 0.5819 is slightly less than $ESE(\hat{\gamma}_0)$ = 0.5834 and $IESE(\hat{\beta}_A/\hat{\beta}_F)$ = 0.2785 is also less than $ESE(\hat{\beta}_A/\hat{\beta}_F)$ = 0.3150. Given the small number of centers for which data was available, the change in pattern may be attributed to the cumulative reduction in the total number of responses in **DATA 2** from 130 to 102 in the treatment group and from 143 to 121 in the control group. More specifically, the number of successes and failures reduced from 55 to 48 and from 75 to 54 respectively in the treatment group and from 96 to 74 failures in the control group. We note that the patterns were the same in **DATA 1** and in **DATA 3** where there was no reduction in the overall number of responses.

Given the differences between the standard error estimates obtained from the two methods of estimation we examined the D-optimality criterion in order to compare the covariance matrices of the estimates. Also, this D-optimality is computed by the determinant of Hessian matrix. The results in Table 2.7 show that at the final stage of the iteration process, estimates based on the influence function approach consistently maximized the determinant of the information matrix. For instance the value of the D-optimality criterion based on the Hessian matrix approach and the influence function approach were 203.56 and 564.11 respectively for **DATA 1** and 166.13 and 829.59 respectively for **DATA 2**. These results indicate that when compared to the method based on the Hessian matrix, the influence function iteration method searches for solutions in a neighborhood that is closer to the vector of true parameters. Thus, the influence function estimate will, in general, be closer to the true parameter. The results of our simulation studies with a large number of centers shown in Tables 2.1 - 2.5 also confirms this conclusion since the values of IESE are, in general, smaller when compared with the values of ESE for all parameter estimates. Concerning the Wald test for significance of treatment effect, we observe that the p-values are about the same in magnitude for **DATA 1** and **DATA 3** irrespective of the method of iteration used in estimating the parameters. However, when the estimation is based on the Hessian matrix, the p-values change rapidly from 0.014, under the original data (**DATA 1**), to 0.078 once Centers 5 and 6 with 0 successes were deleted from the data (**DATA 2**). This then leads to contradictory interpretations of the effect of treatments at 5% level of significance. On the other hand, the effect of excluding centers with 0 successes from the data on the the p-values is not as severe and lead to the same conclusion for **DATA 1**, **DATA 2** and **DATA 3** when the estimation is based on the influence function. Thus, for this data under consideration the influence function

approach appear to lead to a more stable result irrespective of whether centers with 0 successes are included or excluded from the data. The approach based on the Hessian matrix is however not so stable. These results clearly highlight the need for further research on the influence function method of estimation in generalized linear models.

2.6 Conclusion

In this chapter, we have discussed a generalized linear mixed model for analyzing data arising from the application of response-adaptive designs in multi-center clinical trials. We have shown that the estimators of the model parameters are consistent and asymptotically normally distributed. We have also introduced a new iteration method based entirely on the influence function of the parameter estimates. Previously, the computation of the asymptotic variance of the regression parameter estimators have been based entirely on the inverse of the Hessian matrix obtained from the likelihood function. We have now provided an alternative, in this chapter, by deriving a closed form expression for the asymptotic variance of the regression parameter estimators based on an influence function approach. To our knowledge, such a closed form expression does not currently exist in the literature. In fact, our asymptotic approach does not depend on the selection criteria for treatment assignments. These selection criteria for treatment assignments include the family of response-adaptive designs and the completely randomized design. In our simulation studies we have demonstrated that estimates of the asymptotic variance computed from the closed form expression we derived compare favourably well with the true values obtained directly from simulation. Thus, in practice the IESE can be used to estimate the asymptotic variance of the regression parameter estimators. We have also demonstrated through simulation that response-adaptive designs are more ethically and morally desirable because they

assign the potentially better treatment to more patients and that the design variability improves with appropriate choice of the parameter in the allocation function of Hu and Zhang (2004). Moreover, the estimated 95% coverage probabilities based on the normal distribution were shown to be unbiased. This clearly indicates that the maximum likelihood approach with the Gauss-Hermite quadrature approximation performs very well in estimating the parameters of the GLMM. Finally, we note that the Gauss-Hermite quadrature for integral approximation as well as the influence function technique for deriving the asymptotic properties can be easily extended to generalized linear mixed models based on the exponential family.

Chapter 3

Estimation of a Generalized Linear Model for Adaptive Designs in Multi-Center Clinical Trials

3.1 Introduction

In Chapter 2, we assumed the GLMM as an ideal model and investigated the efficiency and ethics of participating patients between RA designs and equal allocation. Also, we applied separate randomization to each center. In this chapter, we assume that center effects are fixed effects. Therefore, we consider the generalized linear model (GLM) is an ideal model. The common randomization for treatment assignments is applied to all medical centers. We examine the asymptotic theories for a general class of adaptive designs assuming the GLM as an ideal model. Also, we compare the efficiency and ethics among RA, CARA, and Completely Randomized (CR) design under this assumption.

In this chapter, we introduce the logit model which includes treatment by covariate interactions and discuss estimation method of model parameters in § 3.2. We provide a procedure for Covariate-Adjusted Response-Adaptive designs in § 3.3. The § 3.4 deals with proposed conditions for asymptotic properties of parameter estimates for CARA designs. We discuss a odds-ratio-based target allocation proportion for Response-Adaptive design in § 3.5. The theoretical results in § 3.4 are validated through simulation studies in § 3.6. Finally, conclusions are provided in § 3.7.

3.2 The Logit Model and Parameter Estimation

In this section, we describe the logit model and the estimation method for model parameters for comparing two treatments, which are treatment A and treatment B, when other associated categorical covariates are considered. The logit model is a well-known model for constructing the relationship between binary responses and associated covariates. Moreover, it is easy to interpret estimates in the logit model if there is no interaction effects in the model. However, it is not easy to interpret the estimates if the interaction terms are significant in the logit model [Ai and Norton (2003)] . In this chapter, we assume that responses are binary, which is either success or failure, so the logit model was chosen to establish a linkage between the responses and covariates. Let n be the total number of patients who were recruited and assigned one and only one treatment from two treatments at the end of a clinical trial. The response and treatment assignment of patient i are defined by (1.7).

Also, let $\mathbf{Z_i'} = (Z_{i1}, \cdots, Z_{ip})$ be the p dimensional vector of covariate information of patient i, $i = 1, 2, \cdots, n$. Here, one of the covariates in the p-dimensional vector $\mathbf{Z_i}$ may represent centers to account for center effects. We define $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m\}$

as the mutually exclusive configuration levels of $\mathbf{Z}$. That is, a typical patient belongs to one and only from these categories. When patients enter sequentially into the clinical trial, we assume that the experimenter knows the value of the random variable $\mathbf{Z}$ before entering a patient into a medical clinic and we assume that the joint probability mass function of $\mathbf{Z}$ follows multinomial distribution with unknown parameters, but these parameters do not depend on model parameters which is defined in (3.1). Let us assume that the probability value of $\mathbf{v}_h$, $P(\mathbf{Z} = \mathbf{v}_h) = \rho_h$; $h = 1, 2, \cdots, m$, then $\sum_{h=1}^{m} \rho_h = 1$. Also, $\mathbf{z_i}$ is the observed value of covariate vector of patient i. Furthermore, let $\mathcal{X}_i = \sigma(X_{1A}, X_{2A}, \cdots, X_{iA})$ and $\mathcal{Y}_i = \sigma(Y_1, Y_2, \cdots, Y_i)$ be the sigma algebras generated by treatment assignments and responses respectively. Define $C(\mathbf{z_i}) = (\mathbf{z_1'}, \mathbf{z_2'}, \cdots, \mathbf{z_i'})'$ as the history of covariates and $\mathcal{F}_i = \sigma(\mathcal{X}_i, \mathcal{Y}_i)$.

Consider the logit model:

$$\text{logit}[P(Y_i = 1|x_{iA}, \mathbf{z}_i)] = x_{iA}\beta_A + \gamma_0 + \mathbf{z}_i'\boldsymbol{\gamma} + x_{iA}\mathbf{z}_i'\boldsymbol{\delta}, \text{ for } i = 1, 2, \cdots, n, \tag{3.1}$$

where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \cdots, \gamma_p)'$ are the main effects of covariates and $\boldsymbol{\delta} = (\delta_1, \delta_2, \cdots, \delta_p)'$ are the treatment by covariates interaction effects. Also, β_A is the effect of treatment A compared to treatment B, γ_0 is the intercept term in this model. Therefore, the probability mass function for the random variables $\{Y_i, i = 1, 2, \cdots, n\}$, is $P(Y_i = y_i|x_{iA}, \mathbf{z}_i) = [1 + \exp(-\mathbf{w}_i'\boldsymbol{\theta})]^{-y_i}[1 + \exp(\mathbf{w}_i'\boldsymbol{\theta})]^{-(1-y_i)}$, where $\mathbf{w}_i = (x_{iA}, 1, \mathbf{z}_i', x_{iA}\mathbf{z}_i')'$, $\boldsymbol{\theta} = (\beta_A, \gamma_0, \boldsymbol{\gamma}', \boldsymbol{\delta}')'$ is a $q := 2(p + 1)$ dimensional vector. Also we assume that $\boldsymbol{\theta}$ belongs to an admissible set $\boldsymbol{\Omega}(\boldsymbol{\theta_0}) \subseteq \Re^q$; $\boldsymbol{\Omega}(\boldsymbol{\theta_0})$ is open and convex in $\Re^q$, where $\boldsymbol{\theta_0} = (\beta_{A0}, \gamma_{00}, \boldsymbol{\gamma}_0', \boldsymbol{\delta}_0')'$ is the q-dimensional vector of true unknown parameters, which we can estimate by the method of ML. For the logit model in (3.1), it can be shown that

$E(Y_i|x_{iA}, \mathbf{z}_i) = [1 + \exp(-\mathbf{w}_i'\boldsymbol{\theta})]^{-1}$, and $Var(Y_i|x_{iA}, \mathbf{z}_i) = \exp(-\mathbf{w}_i'\boldsymbol{\theta})[1 + \exp(-\mathbf{w}_i'\boldsymbol{\theta})]^{-2}$. The likelihood function can be written as

$$\mathcal{L}_n(\boldsymbol{\theta}) = \prod_{i=1}^{n} [\pi_{iA}(\mathbf{z}_i)]^{x_{iA}} [1 - \pi_{iA}(\mathbf{z}_i)]^{(1-x_{iA})} [1 + \exp(-\mathbf{w}_i'\boldsymbol{\theta})]^{-y_i} [1 + \exp(\mathbf{w}_i'\boldsymbol{\theta})]^{-(1-y_i)}, \quad (3.2)$$

where $\pi_{iA}(\mathbf{z}_i) = P[X_{iA} = 1|\mathcal{F}_{i-1}, C(\mathbf{z_{i-1}}), \mathbf{z}_i]$. Then, the log-likelihood function $\ell_n(\boldsymbol{\theta})$ and the score function of the log-likelihood are given by

$$\ell_n(\boldsymbol{\theta}) = \ln h(\tilde{\pi}) - \sum_{i=1}^{n} y_i \ln[1 + \exp(-\mathbf{w}_i'\boldsymbol{\theta})] - \sum_{i=1}^{n} (1 - y_i) \ln[1 + \exp(\mathbf{w}_i'\boldsymbol{\theta})], \quad (3.3)$$

and

$$\mathbf{s}_n(\boldsymbol{\theta}) = \ell_n'(\boldsymbol{\theta}) = \sum_{i=1}^{n} y_i \mathbf{w}_i - \sum_{i=1}^{n} [1 + \exp(-\mathbf{w}_i'\boldsymbol{\theta})]^{-1} \mathbf{w}_i, \quad (3.4)$$

where $h(\tilde{\pi}) = \prod_{i=1}^{n} [\pi_{iA}(\mathbf{z}_i)]^{x_{iA}} [1 - \pi_{iA}(\mathbf{z}_i)]^{(1-x_{iA})}$ and $\ell_n'(\boldsymbol{\theta}) = \dfrac{\partial \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$. We then solve the ML estimating equation

$$\ell_n'(\boldsymbol{\theta}) = \mathbf{0}, \quad (3.5)$$

to obtain the ML estimates of the model parameters.

The second derivative of the log-likelihood function called the observed Fisher information matrix

$$\mathbf{F}_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \exp(-\mathbf{w}_i'\boldsymbol{\theta})[1 + \exp(-\mathbf{w}_i'\boldsymbol{\theta})]^{-2} \mathbf{w}_i \mathbf{w}_i'. \quad (3.6)$$

is used in the Newton-Raphson iteration procedure

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{new}} = \hat{\boldsymbol{\theta}}_{\boldsymbol{old}} + [\mathbf{F}_n(\boldsymbol{\theta})]^{-1}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\boldsymbol{old}}} [\mathbf{s}_n(\boldsymbol{\theta})]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\boldsymbol{old}}},$$

to obtain the MLE $\hat{\boldsymbol{\theta}}_n = (\hat{\beta}_{A,n}, \hat{\gamma}_{0,n}, \hat{\boldsymbol{\gamma}}'_n, \boldsymbol{\delta}'_n)'$ of $\boldsymbol{\theta_0} = (\beta_{A0}, \gamma_{00}, \boldsymbol{\gamma}'_0, \boldsymbol{\delta}'_0)'$ which is the vector of true model parameters. The asymptotic distribution of the MLE $\hat{\boldsymbol{\theta}}_n$ will be discussed in § 3.4.

3.3 The Covariate-Adjusted Response-Adaptive Design (CARA)

Several authors have noted that minimum information is required to get ML estimates [Silvapulle (1981), Albert (1984), and Santner and Duffy (1986)]. If sequential trial gets minimum information, then we are able to compute the inverse of the Hessian matrix. Thus, we need initial information to obtain ML estimates of model parameters in (3.1) as a result of the application of CARA design. Before getting minimum information for which $\frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}$ is estimable, equal allocation is applied for treatment assignments. Suppose the first n_0 number of patients' information is enough to solve the equation in (3.5) and $\frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}$ is estimable for all $n \geq n_0$.

We evaluate some quantities for patient $(n+1)$ having the covariate information $\mathbf{z}_{n+1}$. If this patient is assigned to treatment A, the expected success probability of this patient becomes $P_{n+1,A} = [1 + \exp(-\hat{\beta}_{A,n} - \hat{\gamma}_{0,n} - \mathbf{z}'_{n+1}\hat{\boldsymbol{\gamma}}_n - \mathbf{z}'_{n+1}\hat{\boldsymbol{\delta}}_n)]^{-1}$. Similarly, if this patient is assigned to treatment B, the expected success probability of this patient, $P_{n+1,B} = [1 + \exp(-\hat{\gamma}_{0,n} - \mathbf{z}'_{n+1}\hat{\boldsymbol{\gamma}}_n]^{-1}$. Then, the model based odds ratio for comparing treatment A to treatment B is $x_m = \frac{P_{n+1,A}}{(1-P_{n+1,A})}\frac{(1-P_{n+1,B})}{P_{n+1,B}} = \exp(\hat{\beta}_{A,n} + \mathbf{z}'_{n+1}\hat{\boldsymbol{\delta}}_n)$. Rosenberger, Vidyashankar and Agarwal (2001) applied the most natural mapping to obtain the treatment assignment function, $f : [0, \infty) \to [0, 1]$, defined by

$$f(x) = \frac{x}{x+1}. \tag{3.7}$$

Using this function, the allocation probability to treatment A becomes

$$\pi_{[n+1]A}(\mathbf{z}_{n+1}) = \{1 + \exp[-(\hat{\beta}_{A,n} + \mathbf{z}'_{n+1}\hat{\boldsymbol{\delta}}_n)]\}^{-1}. \tag{3.8}$$

As an example, suppose that patient i belongs to the $\mathbf{v}_h$ group. We thus allocate treatment A to patient i with the following probability:

$$\pi_{iA}(\mathbf{v}_h) = P(X_{iA} = 1|\mathcal{F}_{i-1}, C(\mathbf{z_{i-1}}), \mathbf{z}_i = \mathbf{v}_h) = \{1 + \exp[-(\hat{\beta}_{A,(i-1)} + \mathbf{v}'_h\hat{\boldsymbol{\delta}}_{i-1})]\}^{-1}, \tag{3.9}$$

where $i > n_0$ and n_0 is the initial number of patients to whom treatments were assigned using equal allocation. We note that this is one example of a CARA design.

3.4 Asymptotics of Parameter Estimates for CARA

Define the subgroup of patient indices having the $\mathbf{v}_h$ $(h = 1, 2, \cdots, m)$ covariates configuration level as

$$J_n^{(h)} = \{i : \mathbf{z}_i = \mathbf{v}_h; \quad i = 1, 2, \cdots, n\}.$$

Since the clinical trial in this study is sequential trial and patients arrive sequentially to the clinic, we define

$$J_\infty^{(h)} = \{i : \mathbf{z}_i = \mathbf{v}_h; \quad i = 1, 2, \cdots\cdots\};$$

Next, we describe the subgroup of patients belongs to stratum h. We define a set

$$J^{(h)} = \{(i, h) : \mathbf{z}_i = \mathbf{v}_h; \quad i = 1, 2, \cdots n\},$$

where (i, h) is an ordered pair indicating that patient i belongs to stratum h. For ease of notation we assign the first element of $J^{(h)}$ as 1, the second element as 2, $\cdots$ and the last element of $J^{(h)}$ as n_h. $J^{(h)}$ can be written as

$$J^{(h)} = \{1, 2, \cdots, n_h\}.$$

For instance suppose, $n = 8$ and $m = 4$, $\mathbf{z}_1 = \mathbf{v}_2, \mathbf{z}_2 = \mathbf{v}_1, \mathbf{z}_3 = \mathbf{v}_1, \mathbf{z}_4 = \mathbf{v}_3, \mathbf{z}_5 = \mathbf{v}_4, \mathbf{z}_6 = \mathbf{v}_2, \mathbf{z}_7 = \mathbf{v}_4, \mathbf{z}_8 = \mathbf{v}_2$. Then, the set

$$J^{(1)} = \{(2, 1), (3, 1)\},$$

identifies the patient 2 and patient 3 belong to stratum 1, $n_1 = 2$, and $J^{(1)} = \{1, 2\}$. The set $J^{(h)}$ contains n_h number of patient indices with $n = \sum_{h=1}^{m} n_h$. We have $\pi_{iA}(\mathbf{v}_h)$ is the probability of patient i assigned to treatment A if patient i belongs to $\mathbf{v}_h$ group, for $i = 1, 2, \cdots, n$ and $h = 1, 2, \cdots, m$. In fact, we do not know the format of $\pi_{iA}(\mathbf{v}_h)$. In the next Lemma, we discuss the average of these $\pi_{iA}(\mathbf{v}_h)$ converges almost surely the average of patients assigned to treatment A when we fix h, $h = 1, 2, \cdots, m$.

Lemma 3.4.1. For a group of patients having $\mathbf{v}_h$ covariates configuration level, we have $\frac{N_{Ah}(n)}{n_h} - \frac{1}{n_h} \sum_{i \in J_n^{(h)}} \pi_{iA}(\mathbf{v}_h) \xrightarrow{a.s.} 0$ as $n_h \to \infty$, where $N_{Ah}(n) = \sum_{i \in J_n^{(h)}} X_{iA}$ is the number of patients assigned to treatment A in $\mathbf{v}_h$ group, $N_{Bh}(n) = n_h - N_{Ah}(n)$, and n_h is the number of patients in $\mathbf{v}_h$ group for $h = 1, 2, \cdots, m$. *Also, $\pi_{iA}(\boldsymbol{v}_h)$ is the probability of patient i assigned to treatment A if patient i belongs to $\boldsymbol{v}_h$ group, for $i = 1, 2, \cdots, n$ and $h = 1, 2, \cdots, m$.*

Proof. For a fixed $h = 1, 2, \cdots, m$, define

$S_n(\mathbf{v}_h) = \sum_{i}^{n} I[\mathbf{z}_i = \mathbf{v}_h]\{X_{iA} - \pi_{iA}(\mathbf{v}_h)\} = \sum_{i \in J_n^{(h)}} \{X_{iA} - \pi_{iA}(\mathbf{v}_h)\}$, where I is an indicator function. Let us consider

$S_{n+1}(\mathbf{v}_h) - S_n(\mathbf{v}_h) = I[\mathbf{z}_{n+1} = \mathbf{v}_h]\{X_{(n+1)A} - \pi_{(n+1)A}(\mathbf{z}_{n+1})\}$.

Case 1: If $\mathbf{z}_{n+1} = \mathbf{v}_h$,

$$\begin{aligned} E[S_{n+1}(\mathbf{v}_h) - S_n(\mathbf{v}_h)|\mathcal{F}_n, C(\mathbf{z_n}), \mathbf{z}_{n+1}] &= E\{X_{(n+1)A}|\mathcal{F}_n, C(\mathbf{z_n}), \mathbf{z}_{n+1} = \mathbf{v}_h\} - \pi_{(n+1)A}(\mathbf{v}_h) \\ &= P\{X_{(n+1)A} = 1|\mathcal{F}_n, C(\mathbf{z_n}), \mathbf{z}_{n+1} = \mathbf{v}_h\} - \pi_{(n+1)A}(\mathbf{v}_h) \\ &= 0. \end{aligned}$$

Case 2: If $\mathbf{z}_{n+1} \neq \mathbf{v}_h$, then $S_{n+1}(\mathbf{v}_h) - S_n(\mathbf{v}_h) = 0$. So that

$E[S_{n+1}(\mathbf{v}_h) - S_n(\mathbf{v}_h)|\mathcal{F}_n, C(\mathbf{z_n}), \mathbf{z}_{n+1}] = 0.$

Therefore, the sequence of partial sums $\{S_n(\mathbf{v}_h)\}$ is a martingale. We know that $0 < |X_{nA} - \pi_{nA}(\mathbf{v}_h)| < 1$ for $n \in J_\infty^{(h)}$. It follows that $\sum_{n \in J_\infty^{(h)}} n_h^{-2} E\{[X_{nA} - \pi_{nA}(\mathbf{v}_h)]^2|\mathcal{F}_{n-1}, C(\mathbf{z_{n-1}}), \mathbf{z}_n = \mathbf{v}_h\} < \infty$. Therefore, by Theorem 1 of Csörgö (1968) {the strong law of large numbers for martingales}, we obtain $\lim_{n_h \to \infty} \frac{1}{n_h} S_{n_h}(\mathbf{v}_h) = 0$. Hence, the Lemma holds. □

The observed Fisher information matrices are random matrices. If the average observed Fisher information matrices converges to a non-random matrix, then we can discuss the consistency of ML estimators. We require a condition for the average observed Fisher information matrices converges to a non-random matrix. Therefore, we initially will set a condition for this purpose. We describe this condition in the following Assumption 3.4.1.

Assumption 3.4.1. *For each* $h = 1, 2, \cdots, m$, $(1/n_h) \sum_{i \in J_n^{(h)}} \pi_{iA}(\boldsymbol{v}_h) \xrightarrow{a.s.} \pi_A(\boldsymbol{v}_h)$ *as* $n_h \to \infty$, *where* $0 < \pi_A(\boldsymbol{v}_h) < 1$, *where,* $\pi_{iA}(\boldsymbol{v}_h)$ *is the treatment assignment function.*

We now introduce some new notations for elements of the observed Fisher information matrix for mathematical convenience. Let $\frac{1}{n}\mathbf{F}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{H}_i(\boldsymbol{\theta})$, where

$$\mathbf{H}_i(\boldsymbol{\theta}) = g(\mathbf{w}_i'\boldsymbol{\theta}) \begin{bmatrix} x_{iA} & x_{iA} & x_{iA}\mathbf{z}_i' & x_{iA}\mathbf{z}_i' \\ x_{iA} & 1 & \mathbf{z}_i' & x_{iA}\mathbf{z}_i' \\ x_{iA}\mathbf{z}_i & \mathbf{z}_i & \mathbf{z}_i\mathbf{z}_i' & x_{iA}\mathbf{z}_i\mathbf{z}_i' \\ x_{iA}\mathbf{z}_i & x_{iA}\mathbf{z}_i & x_{iA}\mathbf{z}_i\mathbf{z}_i' & x_{iA}\mathbf{z}_i\mathbf{z}_i' \end{bmatrix} \text{ and} \tag{3.10}$$

$g(\mathbf{w}_i'\boldsymbol{\theta}) = \exp(-\mathbf{w}_i'\boldsymbol{\theta})[1+\exp(-\mathbf{w}_i'\boldsymbol{\theta})]^{-2}$. Furthermore, define the functions of $\boldsymbol{\theta}$: $\lambda_{Ah}(\boldsymbol{\theta})$ and $\lambda_{Bh}(\boldsymbol{\theta})$ as

$$\begin{aligned} \lambda_{Ah}(\boldsymbol{\theta}) &= \exp(-\beta_A - \gamma_0 - \mathbf{v}_h'\boldsymbol{\gamma} - \mathbf{v}_h'\boldsymbol{\delta})[1+\exp(-\beta_A - \gamma_0 - \mathbf{v}_h'\boldsymbol{\gamma} - \mathbf{v}_h'\boldsymbol{\delta})]^{-2} \\ \lambda_{Bh}(\boldsymbol{\theta}) &= \exp(-\gamma_0 - \mathbf{v}_h'\boldsymbol{\gamma})[1+\exp(-\gamma_0 - \mathbf{v}_h'\boldsymbol{\gamma})]^{-2}. \end{aligned}$$

Conditional on $\mathcal{F}_{i-1}$, $C(\mathbf{z_{i-1}})$, and $\mathbf{z}_i$, X_{iA} follows a bernoulli distribution with the unknown parameter. The recursive expectations can be taken from step 1 to step n to obtain the unconditional expectation of the elements of the average observed information matrix which is defined in (3.10). If this average expectation of the observed Fisher information matrix converges almost surely to a non-random matrix, then this non-random matrix is called the Fisher information matrix. In this section, the Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$ is defined in Lemma 3.4.2. Lemma 3.4.2 also describes the existence of the Fisher information matrix under some conditions.

Lemma 3.4.2. If Assumption 3.4.1 and $\frac{n_h}{n} \xrightarrow{a.s} \rho_h$ are satisfied, then we have $(1/n)\sum_{i=1}^{n} E_{i-1}\{\mathbf{H}_i(\boldsymbol{\theta})\} \xrightarrow{a.s.} \mathbf{I}(\boldsymbol{\theta})$ as $n \to \infty$, where the conditional expectation defined by

$E_{i-1}[\bullet] = E_{x_{iA}}[\bullet|\mathcal{F}_{i-1}, C(\mathbf{z_{i-1}}), \mathbf{z}_i]$, $\mathbf{I}(\boldsymbol{\theta})$ is the Fisher information matrix that is a non-random matrix

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} I_{11}(\boldsymbol{\theta}) & I_{12}(\boldsymbol{\theta}) & \mathbf{I}_{13}(\boldsymbol{\theta}) & \mathbf{I}_{14}(\boldsymbol{\theta}) \\ I'_{12}(\boldsymbol{\theta}) & I_{22}(\boldsymbol{\theta}) & \mathbf{I}_{23}(\boldsymbol{\theta}) & \mathbf{I}_{24}(\boldsymbol{\theta}) \\ \mathbf{I}'_{13}(\boldsymbol{\theta}) & \mathbf{I}'_{23}(\boldsymbol{\theta}) & \mathbf{I}_{33}(\boldsymbol{\theta}) & \mathbf{I}_{34}(\boldsymbol{\theta}) \\ \mathbf{I}'_{14}(\boldsymbol{\theta}) & \mathbf{I}'_{24}(\boldsymbol{\theta}) & \mathbf{I}'_{34}(\boldsymbol{\theta}) & \mathbf{I}_{44}(\boldsymbol{\theta}) \end{bmatrix}, \tag{3.11}$$

where $\boldsymbol{\theta} \in \boldsymbol{\Omega}(\boldsymbol{\theta_0})$. Moreover, $\mathbf{H}_i(\boldsymbol{\theta})$ is defined in (3.10) for $i = 1, 2, \cdots, n$ and n_h is the number of patients in $\mathbf{v}_h$ group.

Proof. First we note that $I_{12}(\boldsymbol{\theta}) = I_{11}(\boldsymbol{\theta})$, $I_{14}(\boldsymbol{\theta}) = I_{24}(\boldsymbol{\theta}) = I_{13}(\boldsymbol{\theta})$, and $I_{44}(\boldsymbol{\theta}) = I_{34}(\boldsymbol{\theta})$. Therefore, we have the following results for the components of $\mathbf{I}(\boldsymbol{\theta})$.

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} E_{i-1}\{g(\mathbf{w}_i'\boldsymbol{\theta})x_{iA}\} &= \frac{1}{n}\sum_{i=1}^{n} E\{g(\mathbf{w}_i'\boldsymbol{\theta})x_{iA}|\mathcal{F}_{i-1}, C(\mathbf{z_{i-1}}), \mathbf{z}_i\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{h=1}^{m} I[\mathbf{z}_i = \mathbf{v}_h]\lambda_{Ah}(\boldsymbol{\theta})\pi_{iA}(\mathbf{z}_i) \\
&= \sum_{h=1}^{m}\sum_{i=1}^{n} \frac{1}{n} I[\mathbf{z}_i = \mathbf{v}_h]\lambda_{Ah}(\boldsymbol{\theta})\pi_{iA}(\mathbf{z}_i) \\
&= \sum_{h=1}^{m} \frac{n_h}{n}\lambda_{Ah}(\boldsymbol{\theta}) \left[\frac{1}{n_h}\sum_{i\in J_n^{(h)}} \pi_{iA}(\mathbf{v}_h)\right] \\
&\xrightarrow{a.s.} \sum_{h=1}^{m} \rho_h \lambda_{Ah}(\boldsymbol{\theta})\pi_A(\mathbf{v}_h) = I_{11}(\boldsymbol{\theta}) \quad \text{[because of } \frac{n_h}{n} \xrightarrow{a.s} \rho_h, \text{ and} \\
&\quad \text{the Assumption 3.4.1.]} \\
\frac{1}{n}\sum_{i=1}^{n} E_{i-1}\{g(\mathbf{w}_i'\boldsymbol{\theta})\} &= \frac{1}{n}\sum_{i=1}^{n} E\{g(\mathbf{w}_i'\boldsymbol{\theta})|\mathcal{F}_{i-1}, C(\mathbf{z_{i-1}}), \mathbf{z}_i\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{h=1}^{m} I[\mathbf{z}_i = \mathbf{v}_h]\{\lambda_{Ah}(\boldsymbol{\theta})\pi_{iA}(\mathbf{z}_i) + \lambda_{Bh}(\boldsymbol{\theta})[1 - \pi_{iA}(\mathbf{z}_i)]\} \\
&= \sum_{h=1}^{m} \frac{n_h}{n}\left\{\lambda_{Ah}(\boldsymbol{\theta})\left[\frac{1}{n_h}\sum_{i\in J_n^{(h)}} \pi_{iA}(\mathbf{v}_h)\right] + \lambda_{Bh}(\boldsymbol{\theta})\left[1 - \frac{1}{n_h}\sum_{i\in J_n^{(h)}} \pi_{iA}(\mathbf{v}_h)\right]\right\} \\
&\xrightarrow{a.s.} \sum_{h=1}^{m} \rho_h\{\lambda_{Ah}(\boldsymbol{\theta})\pi_A(\mathbf{v}_h) + \lambda_{Bh}(\boldsymbol{\theta})[1 - \pi_A(\mathbf{v}_h)]\} = I_{22}(\boldsymbol{\theta}).
\end{aligned}$$

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} E_{i-1}\{g(\mathbf{w}_i'\boldsymbol{\theta})x_{iA}\mathbf{z}_i'\} &= \frac{1}{n}\sum_{i=1}^{n} E\{g(\mathbf{w}_i'\boldsymbol{\theta})x_{iA}\mathbf{z}_i'|\mathcal{F}_{i-1}, C(\mathbf{z_{i-1}}), \mathbf{z}_i\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{h=1}^{m} I[\mathbf{z}_i = \mathbf{v}_h]\lambda_{Ah}(\boldsymbol{\theta})\mathbf{z}_i'\pi_{iA}(\mathbf{z}_i) \\
&= \sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta})\frac{1}{n}\sum_{i=1}^{n} I[\mathbf{z}_i = \mathbf{v}_h]\mathbf{z}_i'\pi_{iA}(\mathbf{z}_i) \\
&= \sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta})\frac{n_h}{n}\left[\frac{1}{n_h}\sum_{i\in J_n^{(h)}}\pi_{iA}(\mathbf{v}_h)\right]\mathbf{v}_h' \\
&\xrightarrow{a.s.} \sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta})\rho_h\pi_A(\mathbf{v}_h)\mathbf{v}_h' = \mathbf{I}_{13}(\boldsymbol{\theta}).
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} E_{i-1}\{g(\mathbf{w}_i'\boldsymbol{\theta})\mathbf{z}_i'\} &= \frac{1}{n}\sum_{i=1}^{n} E\{g(\mathbf{w}_i'\boldsymbol{\theta})\mathbf{z}_i'|\mathcal{F}_{i-1}, C(\mathbf{z_{i-1}}), \mathbf{z}_i\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{h=1}^{m} I[\mathbf{z}_i = \mathbf{v}_h]\{\lambda_{Ah}(\boldsymbol{\theta})\pi_{iA}(\mathbf{z}_i) + \lambda_{Bh}(\boldsymbol{\theta})[1-\pi_{iA}(\mathbf{z}_i)]\}\mathbf{v}_h' \\
&= \sum_{h=1}^{m}\frac{n_h}{n}\left\{\lambda_{Ah}(\boldsymbol{\theta})\left[\frac{1}{n_h}\sum_{i\in J_n^{(h)}}\pi_{iA}(\mathbf{v}_h)\right] + \lambda_{Bh}(\boldsymbol{\theta})\left[1-\frac{1}{n_h}\sum_{i\in J_n^{(h)}}\pi_{iA}(\mathbf{v}_h)\right]\right\}\mathbf{v}_h' \\
&\xrightarrow{a.s.} \sum_{h=1}^{m}\rho_h\{\lambda_{Ah}(\boldsymbol{\theta})\pi_A(\mathbf{v}_h) + \lambda_{Bh}(\boldsymbol{\theta})[1-\pi_A(\mathbf{v}_h)]\}\mathbf{v}_h' = \mathbf{I}_{23}(\boldsymbol{\theta}).
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} E_{i-1}\{g(\mathbf{w}_i'\boldsymbol{\theta})\mathbf{z}_i\mathbf{z}_i'\} &= \frac{1}{n}\sum_{i=1}^{n} E\{g(\mathbf{w}_i'\boldsymbol{\theta})\mathbf{z}_i\mathbf{z}_i'|\mathcal{F}_{i-1}, C(\mathbf{z_{i-1}}), \mathbf{z}_i\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{h=1}^{m} I[\mathbf{z}_i = \mathbf{v}_h]\{\lambda_{Ah}(\boldsymbol{\theta})\pi_{iA}(\mathbf{z}_i) + \lambda_{Bh}(\boldsymbol{\theta})[1-\pi_{iA}(\mathbf{z}_i)])\}\mathbf{v}_h\mathbf{v}_h' \\
&= \sum_{h=1}^{m}\frac{n_h}{n}\left\{\lambda_{Ah}(\boldsymbol{\theta})\left[\frac{1}{n_h}\sum_{i\in J_n^{(h)}}\pi_{iA}(\mathbf{v}_h)\right] + \lambda_{Bh}(\boldsymbol{\theta})\left[1-\frac{1}{n_h}\sum_{i\in J_n^{(h)}}\pi_{iA}(\mathbf{v}_h)\right]\right\}\mathbf{v}_h\mathbf{v}_h' \\
&\xrightarrow{a.s.} \sum_{h=1}^{m}\rho_h\{\lambda_{Ah}(\boldsymbol{\theta})\pi_A(\mathbf{v}_h) + \lambda_{Bh}(\boldsymbol{\theta})[1-\pi_A(\mathbf{v}_h)]\}\mathbf{v}_h\mathbf{v}_h' = \mathbf{I}_{33}(\boldsymbol{\theta}).
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} E_{i-1}\{g(\mathbf{w}_i'\boldsymbol{\theta})x_{iA}\mathbf{z}_i\mathbf{z}_i'\} &= \frac{1}{n}\sum_{i=1}^{n} E\{g(\mathbf{w}_i'\boldsymbol{\theta})x_{iA}\mathbf{z}_i\mathbf{z}_i'|\mathcal{F}_{i-1}, C(\mathbf{z_{i-1}}), \mathbf{z}_i\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{h=1}^{m} I[\mathbf{z}_i = \mathbf{v}_h]\lambda_{Ah}(\boldsymbol{\theta})\pi_{iA}(\mathbf{z}_i)\mathbf{v}_h\mathbf{v}_h' \\
&= \sum_{h=1}^{m}\frac{n_h}{n}\lambda_{Ah}(\boldsymbol{\theta})\left[\frac{1}{n_h}\sum_{i\in J_n^{(h)}}\pi_{iA}(\mathbf{v}_h)\right]\mathbf{v}_h\mathbf{v}_h' \\
&\xrightarrow{a.s.} \sum_{h=1}^{m}\rho_h\lambda_{Ah}(\boldsymbol{\theta})\pi_A(\mathbf{v}_h)\mathbf{v}_h\mathbf{v}_h' = \mathbf{I}_{34}(\boldsymbol{\theta}).
\end{aligned}
$$

Therefore, $(1/n)\sum_{i=1}^{n} E_{i-1}\{\mathbf{H}_i(\boldsymbol{\theta})\} \xrightarrow{a.s.} \mathbf{I}(\boldsymbol{\theta})$ as $n \to \infty$. □

The next Lemma states that the average observed Fisher information matrix converges almost surely to the Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$ under some conditions.

Lemma 3.4.3. If the Assumption 3.4.1 and $\frac{n_h}{n} \xrightarrow{a.s} \rho_h$ are satisfied, then the average observed Fisher information matrix, $(1/n)\mathbf{F}_n(\boldsymbol{\theta})$, converges almost surely to $\mathbf{I}(\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \boldsymbol{\Omega}(\boldsymbol{\theta_0})$, n_h is the number of patients in $\mathbf{v}_h$ group, and $\mathbf{I}(\boldsymbol{\theta})$ is a Fisher information matrix.

Proof. Define $N_{Ah}(n) = \sum_{i \in J_n^{(h)}} x_{iA}$ as the number of patients assigned to treatment A in $\mathbf{v}_h$ group and $N_{Bh}(n) = n_h - N_{Ah}(n)$. Then, the following results hold.

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}_i'\boldsymbol{\theta})x_{iA} &= \frac{1}{n}\sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta})N_{Ah}(n) \\
&= \sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta})\frac{n_h}{n}\frac{N_{Ah}(n)}{n_h} \\
&\overset{\text{a.s.}}{\sim} \sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta})\frac{n_h}{n}\left[\frac{1}{n_h}\sum_{i \in J_n^{(h)}}\pi_{iA}(\mathbf{v}_h)\right] \\
&\xrightarrow{a.s.} \sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta})\rho_h\pi_A(\mathbf{v}_h) = I_{11}(\boldsymbol{\theta}) \\
\frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}_i'\boldsymbol{\theta})x_{iA}\mathbf{z}_i' &= \frac{1}{n}\sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta})N_{Ah}(n)\mathbf{v}_h' \\
&= \sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta})\frac{n_h}{n}\frac{N_{Ah}(n)}{n_h}\mathbf{v}_h' \\
&\overset{\text{a.s.}}{\sim} \sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta})\frac{n_h}{n}\left[\frac{1}{n_h}\sum_{i \in J_n^{(h)}}\pi_{iA}(\mathbf{v}_h)\right]\mathbf{v}_h' \\
&\xrightarrow{a.s.} \sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta})\rho_h\pi_A(\mathbf{v}_h)\mathbf{v}_h' = \mathbf{I}_{13}(\boldsymbol{\theta})
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}_i'\boldsymbol{\theta}) &= \frac{1}{n}\sum_{h=1}^{m}\{\lambda_{Ah}(\boldsymbol{\theta})N_{Ah}(h) + \lambda_{Bh}(\boldsymbol{\theta})\left[n_h - N_{Ah}(h)\right]\} \\
&= \sum_{h=1}^{m}\frac{n_h}{n}\left\{\lambda_{Ah}(\boldsymbol{\theta})\frac{N_{Ah}(h)}{n_h} + \lambda_{Bh}(\boldsymbol{\theta})\left[1 - \frac{N_{Ah}(h)}{n_h}\right]\right\} \\
&\overset{\text{a.s.}}{\sim} \sum_{h=1}^{m}\frac{n_h}{n}\left\{\lambda_{Ah}(\boldsymbol{\theta})\left[\frac{1}{n_h}\sum_{i\in J_n^{(h)}}\pi_{iA}(\mathbf{v}_h)\right] + \lambda_{Bh}(\boldsymbol{\theta})\left[1 - \frac{1}{n_h}\sum_{i\in J_n^{(h)}}\pi_{iA}(\mathbf{v}_h)\right]\right\} \\
&\xrightarrow{a.s.} \sum_{h=1}^{m}\rho_h\{\lambda_{Ah}(\boldsymbol{\theta})\pi_A(\mathbf{v}_h) + \lambda_{Bh}(\boldsymbol{\theta})[1-\pi_A(\mathbf{v}_h)]\} = I_{22}(\boldsymbol{\theta}) \\
\frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}_i'\boldsymbol{\theta})\mathbf{z}_i' &= \frac{1}{n}\sum_{h=1}^{m}\{\lambda_{Ah}(\boldsymbol{\theta})N_{Ah}(n) + \lambda_{Bh}(\boldsymbol{\theta})\left[n_h - N_{Ah}(n)\right]\}\mathbf{v}_h' \\
&= \sum_{h=1}^{m}\frac{n_h}{n}\left\{\lambda_{Ah}(\boldsymbol{\theta})\frac{N_{Ah}(n)}{n_h} + \lambda_{Bh}(\boldsymbol{\theta})\left[1 - \frac{N_{Ah}(n)}{n_h}\right]\right\}\mathbf{v}_h' \\
&\overset{\text{a.s.}}{\sim} \sum_{h=1}^{m}\frac{n_h}{n}\left\{\lambda_{Ah}(\boldsymbol{\theta})\left[\frac{1}{n_h}\sum_{i\in J_n^{(h)}}\pi_{iA}(\mathbf{v}_h)\right] + \lambda_{Bh}(\boldsymbol{\theta})\left[1 - \frac{1}{n_h}\sum_{i\in J_n^{(h)}}\pi_{iA}(\mathbf{v}_h)\right]\right\}\mathbf{v}_h' \\
&\xrightarrow{a.s.} \sum_{h=1}^{m}\rho_h\{\lambda_{Ah}(\boldsymbol{\theta})\pi_A(\mathbf{v}_h) + \lambda_{Bh}(\boldsymbol{\theta})[1-\pi_A(\mathbf{v}_h)]\}\mathbf{v}_h' = \mathbf{I}_{23}(\boldsymbol{\theta}) \\
\frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}_i'\boldsymbol{\theta})\mathbf{z}_i\mathbf{z}_i' &= \frac{1}{n}\sum_{h=1}^{m}\{\lambda_{Ah}(\boldsymbol{\theta})N_{Ah}(n) + \lambda_{Bh}(\boldsymbol{\theta})\left[n_h - N_{Ah}(n)\right]\}\mathbf{v}_h\mathbf{v}_h' \\
&= \sum_{h=1}^{m}\frac{n_h}{n}\left\{\lambda_{Ah}(\boldsymbol{\theta})\frac{N_{Ah}(n)}{n_h} + \lambda_{Bh}(\boldsymbol{\theta})\left[1 - \frac{N_{Ah}(n)}{n_h}\right]\right\}\mathbf{v}_h\mathbf{v}_h' \\
&\overset{\text{a.s.}}{\sim} \sum_{h=1}^{m}\frac{n_h}{n}\left\{\lambda_{Ah}(\boldsymbol{\theta})\left[\frac{1}{n_h}\sum_{i\in J_n^{(h)}}\pi_{iA}(\mathbf{v}_h)\right] + \lambda_{Bh}(\boldsymbol{\theta})\left[1 - \frac{1}{n_h}\sum_{i\in J_n^{(h)}}\pi_{iA}(\mathbf{v}_h)\right]\right\}\mathbf{v}_h\mathbf{v}_h' \\
&\xrightarrow{a.s.} \sum_{h=1}^{m}\rho_h\{\lambda_{Ah}(\boldsymbol{\theta})\pi_A(\mathbf{v}_h) + \lambda_{Bh}(\boldsymbol{\theta})[1-\pi_A(\mathbf{v}_h)]\}\mathbf{v}_h\mathbf{v}_h' = \mathbf{I}_{33}(\boldsymbol{\theta}) \\
\frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}_i'\boldsymbol{\theta})x_{iA}\mathbf{z}_i\mathbf{z}_i' &= \frac{1}{n}\sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta})N_{Ah}(n)\mathbf{v}_h\mathbf{v}_h' \\
&= \sum_{h=1}^{m}\frac{n_h}{n}\lambda_{Ah}(\boldsymbol{\theta})\frac{N_{Ah}(n)}{n_h}\mathbf{v}_h\mathbf{v}_h' \\
&\overset{\text{a.s.}}{\sim} \sum_{h=1}^{m}\frac{n_h}{n}\lambda_{Ah}(\boldsymbol{\theta})\left[\frac{1}{n_h}\sum_{i\in J_n^{(h)}}\pi_{iA}(\mathbf{v}_h)\right]\mathbf{v}_h\mathbf{v}_h' \\
&\xrightarrow{a.s.} \sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta})\pi_A(\mathbf{v}_h)\rho_h\mathbf{v}_h\mathbf{v}_h' = \mathbf{I}_{34}(\boldsymbol{\theta})
\end{aligned}
$$

□

In Lemma 3.4.3 we show that the average observed Fisher information, $(1/n)\mathbf{F}_n(\boldsymbol{\theta})$, converges almost surely and pointwise to $\mathbf{I}(\boldsymbol{\theta})$ for each $\boldsymbol{\theta} \in \boldsymbol{\Omega}(\boldsymbol{\theta_0})$. The next Lemma 3.4.4 states that $(1/n)\mathbf{F}_n(\boldsymbol{\theta})$ converges almost surely and uniformly on a neighbourhood of $\boldsymbol{\theta}_0$ under the same conditions in Lemma 3.4.3.

Lemma 3.4.4. If the Assumption 3.4.1, and $\frac{n_h}{n} \xrightarrow{a.s.} \rho_h$ are satisfied with $m < \infty$ and $q < \infty$, then we can find that there exists an open ball $\mathbf{G} \subseteq \boldsymbol{\Omega}(\boldsymbol{\theta_0})$, $\theta_0 \in G$ such that the average observed Fisher information $(1/n)\mathbf{F}_n(\boldsymbol{\theta})$ uniformly and almost surely converges to $\mathbf{I}(\boldsymbol{\theta})$ on $\mathbf{G}$, where n_h is the number of patients in $\mathbf{v}_h$ group.

Proof. From Lemma 3.4.3 we have that $(1/n)\mathbf{F}_n(\boldsymbol{\theta})$ almost surely converges on $\boldsymbol{\Omega}(\boldsymbol{\theta_0})$. First we consider the element $(1,1) := (1/n)\sum_{i=1}^{n} g(\mathbf{w}_i'\boldsymbol{\theta})x_{iA} = \sum_{h=1}^{m} \lambda_{Ah}(\boldsymbol{\theta})\{N_{Ah}(n)/n\}$ of the matrix $(1/n)\mathbf{F}_n(\boldsymbol{\theta})$. Let $\epsilon > 0$, for $h = 1, 2, \cdots, m$.

From Lemma 3.4.3 we have $\{N_{Ah}(n)/n\} \xrightarrow{a.s.} \rho_h \pi_A(\mathbf{v}_h)$ as $n \to \infty$ under Assumption (3.4.1) and $\frac{n_h}{n} \xrightarrow{a.s} \rho_h$. The function $\lambda_{Ah}(\boldsymbol{\theta}) : \boldsymbol{\Omega}(\boldsymbol{\theta_0}) \to \Re$ is a bounded function. Therefore, there exists $N_{11h} \in \mathbb{N}$ such that $|\{N_{Ah}(n)/n\} - \rho_h \pi_A(\mathbf{v}_h)| < \epsilon/1.5 \sum_{h=1}^{m} \lambda_{Ah}(\boldsymbol{\theta}_0)$ for all $n \geq N_{11h}$ a.s. Furthermore, $\lambda_{Ah}(\boldsymbol{\theta})$ is a continuous function, so $\lim_{\boldsymbol{\theta} \to \boldsymbol{\theta}_0} \lambda_{Ah}(\boldsymbol{\theta}) = \lambda_{Ah}(\boldsymbol{\theta}_0)$. Therefore, there exists $\delta_{11h} > 0$ such that $\lambda_{Ah}(\boldsymbol{\theta}) < 1.5\lambda_{Ah}(\boldsymbol{\theta}_0)$ for all $\boldsymbol{\theta} \in \mathbf{G}_{\delta_{11h}}$, where $\mathbf{G}_{\delta_{11h}} = \{\boldsymbol{\theta} : \| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \| < \delta_{11h}\}$. We then obtain

$$\lambda_{Ah}(\boldsymbol{\theta}) \left| \frac{N_{Ah}(n)}{n} - \rho_h \pi_A(\mathbf{v}_h) \right| < [\epsilon \lambda_{Ah}(\boldsymbol{\theta}_0)] / \sum_{h=1}^{m} \lambda_{Ah}(\boldsymbol{\theta}_0). \tag{3.12}$$

Consider

$$\begin{aligned}\left|\sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta})\frac{N_{Ah}(n)}{n}-\sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta})\rho_h\pi_A(\mathbf{v}_h)\right| &= \left|\sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta})\left\{\frac{N_{Ah}(n)}{n}-\rho_h\pi_A(\mathbf{v}_h)\right\}\right| \\ &\leq \sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta})\left|\frac{N_{Ah}(n)}{n}-\rho_h\pi_A(\mathbf{v}_h)\right| \\ &< \epsilon. \qquad (3.13)\end{aligned}$$

Therefore, $\left|\sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta})\frac{N_{Ah}(n)}{n}-\sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta})p_h\pi_A(\mathbf{v}_h)\right| < \epsilon$ for all $n \geq N_{11}$ and for all $\boldsymbol{\theta} \in \mathbf{G}_{11}$, where $N_{11} = \max_{1\leq h\leq m} N_{11h}$, $\delta_{11} = \min_{1\leq h\leq m}\delta_{11h}$, and $\mathbf{G}_{11} = \{\boldsymbol{\theta} :\| \boldsymbol{\theta}-\boldsymbol{\theta}_0 \|< \delta_{11}\}$. It follows that $\frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}_i'\boldsymbol{\theta})x_{iA}$ uniformly and almost surely converges on $\mathbf{G}_{11}$. Similarly, we can show that there exists $N_{ij} = \max_{1\leq h\leq m} N_{ijh}$ and $\delta_{ij} = \min_{1\leq h\leq m}\delta_{ijh}$ such that the $(i,j)th$ element of the matrix $(1/n)\mathbf{F}_n(\boldsymbol{\theta})$ uniformly and almost surely converges on $\mathbf{G}_{ij}$, where $\mathbf{G}_{ij} = \{\boldsymbol{\theta} :\| \boldsymbol{\theta}-\boldsymbol{\theta}_0 \|< \delta_{ij}\}$. Since q is finite, there exists $N = \max_{1\leq i,j\leq q} N_{ij}$ and $\delta = \min_{1\leq i,j\leq q}\delta_{ij}$ such that $(1/n)\mathbf{F}_n(\boldsymbol{\theta})$ uniformly and almost surely converges to $\mathbf{I}(\boldsymbol{\theta})$ on $\mathbf{G}$, where $\mathbf{G} = \{\boldsymbol{\theta} :\| \boldsymbol{\theta}-\boldsymbol{\theta}_0 \|< \delta\}$. □

Lemma 3.4.5. If the Assumption 3.4.1 is satisfied with $m < \infty$ and $q < \infty$, then $(1/n)\mu_{in}$ uniformly and almost surely converges to $\lambda_i(\boldsymbol{\theta})$ as $n \to \infty$ for $i = 1, 2, \cdots, q$, where $\mu_{1n} \leq \mu_{2n} \cdots \leq \mu_{qn}$ and $\lambda_1(\boldsymbol{\theta}) \leq \lambda_2(\boldsymbol{\theta}) \cdots \leq \lambda_q(\boldsymbol{\theta})$ are eigenvalues of the symmetric matrices $\mathbf{F}_n(\boldsymbol{\theta})$ and $\mathbf{I}(\boldsymbol{\theta})$ respectively, $\mathbf{F}_n(\boldsymbol{\theta})$ is an observed Fisher information matrix, and $\mathbf{I}(\boldsymbol{\theta})$ is a Fisher information matrix.

Proof. According to Lemma 2.1.19 (**Hoffman-Wielandt**) in Anderson, Alice and Ofer (2010), we have

$$\sum_{i=1}^{q}|(1/n)\mu_{in}-\lambda_i(\boldsymbol{\theta})|^2 \leq \text{trace}\{(1/n)\mathbf{F}_n(\boldsymbol{\theta})-\mathbf{I}(\boldsymbol{\theta})\}^2. \qquad (3.14)$$

Let the $(j,k)^{th}$ element of the matrices $(1/n)\mathbf{F}_n(\boldsymbol{\theta})$ and $\mathbf{I}(\boldsymbol{\theta})$ be defined by $(1/n)\mathbf{F}_{jkn}(\boldsymbol{\theta})$

and $\mathbf{I}_{jk}(\boldsymbol{\theta})$, respectively, where $j, k = 1, 2, \cdots, q$. Let $\epsilon > 0$. From Lemma 3.4.4, there exists $N \in \mathbb{N}$ and $\delta > 0$ such that $(1/n)\mathbf{F}_n(\boldsymbol{\theta})$ uniformly and almost surely converges to $\mathbf{I}(\boldsymbol{\theta})$ on $\mathbf{G}$, where $\mathbf{G} = \{\boldsymbol{\theta} :\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \| < \delta\}$. Therefore, there exists $N \in \mathbb{N}$ and $\delta > 0$ such that $(1/n)\mathbf{F}_{jkn}(\boldsymbol{\theta})$ uniformly and almost surely converges to $\mathbf{I}_{jk}(\boldsymbol{\theta})$ on $\mathbf{G}$, where $\mathbf{G} = \{\boldsymbol{\theta} :\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \| < \delta\}$ for all $1 \leq j, k \leq q$.

Then, there exists $N \in \mathbb{N}$ and $\delta > 0$ such that $|(1/n)\mathbf{F}_{jkn}(\boldsymbol{\theta}) - \mathbf{I}_{jk}(\boldsymbol{\theta})|^2 < \epsilon^2/q^2$ on $\mathbf{G}$ for all $\boldsymbol{\theta} \in \mathbf{G} = \{\boldsymbol{\theta} :\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \| < \delta\}$, all $n \geq N$, where $1 \leq j, k \leq q$. Furthermore,

$$\text{trace}\{(1/n)\mathbf{F}_n(\boldsymbol{\theta}) - \mathbf{I}(\boldsymbol{\theta})\}^2 = \sum_{j=1}^{q}\sum_{k=1}^{q} |(1/n)\mathbf{F}_{jkn}(\boldsymbol{\theta}) - \mathbf{I}_{jk}(\boldsymbol{\theta})|^2 < \epsilon^2 \tag{3.15}$$

Because $(1/n)\mathbf{F}_n(\boldsymbol{\theta}) - \mathbf{I}(\boldsymbol{\theta})$ is a symmetric matrix. From (3.14) and (3.15) we obtain $\sum_{i=1}^{q} |(1/n)\mu_{in} - \lambda_i(\boldsymbol{\theta})|^2 < \epsilon^2$. Therefore, there exists $N \in \mathbb{N}$ and $\delta > 0$ such that $|(1/n)\mu_{in} - \lambda_i(\boldsymbol{\theta})| < \epsilon$ for all $n \geq N$, all $\boldsymbol{\theta} \in \mathbf{G} = \{\boldsymbol{\theta} :\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \| < \delta\}$ with $1 \leq i \leq q$. Hence the Lemma holds. □

Assumption 3.4.2. $\mathbf{I}(\boldsymbol{\theta_0})$ *is a positive definite matrix, where* $\boldsymbol{\theta_0}$ *is the* q *dimensional true vector of parameters, and* $\mathbf{I}(\boldsymbol{\theta_0})$ *is a Fisher information matrix.*

Lemma 3.4.6. If the Assumptions 3.4.1 and 3.4.2 are satisfied with $m < \infty$, and $q < \infty$, then the following results hold

(i) There exists an open ball $\mathbf{G} \subseteq \boldsymbol{\Omega}(\boldsymbol{\theta_0})$ and $N \in \mathbb{N}$ such that the Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$ is a positive definite matrix for all $\boldsymbol{\theta} \in \mathbf{G}$ and the observed Fisher information matrix $\mathbf{F}_n(\boldsymbol{\theta})$ is positive definite for all $n \geq N$ and for all $\boldsymbol{\theta} \in \mathbf{G}$.

(ii) $\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s} \boldsymbol{\theta_0}$ as $n \to \infty$, where $\hat{\boldsymbol{\theta}}_n$ is the MLE of $\boldsymbol{\theta_0}$.

Proof. (i) From Lemma (3.4.5), we have that $(1/n)\lambda_{\min}\mathbf{F}_n(\boldsymbol{\theta})$ uniformly and almost surely converges to $\lambda_1(\boldsymbol{\theta})$ as $n \to \infty$, where $\lambda_{\min}\mathbf{F}_n(\boldsymbol{\theta})$ is the minimum eigenvalue of $\mathbf{F}_n(\boldsymbol{\theta})$. Therefore, there exists an open ball $\mathbf{G}_1 = \{\boldsymbol{\theta} :\| \boldsymbol{\theta}-\boldsymbol{\theta}_0 \|< \delta_1\} \subseteq \boldsymbol{\Omega}(\boldsymbol{\theta_0})$ and $N \in \mathbb{N}$ such that $\lambda_1(\boldsymbol{\theta}) - 0.5\lambda_1(\boldsymbol{\theta}_0) < (1/n)\lambda_{\min}\mathbf{F}_n(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \mathbf{G}_1$ and for all $n \geq N$.

By applying Lemma 3.4.4, $(1/n)\mathbf{F}_n(\boldsymbol{\theta})$ uniformly and almost surely converges to $\mathbf{I}(\boldsymbol{\theta})$ on $\mathbf{G}_2$. Moreover, $(1/n)\mathbf{F}_n(\boldsymbol{\theta})$ is continuous on $\mathbf{G}_2$. From the uniform convergence theorem, we have that $\mathbf{I}(\boldsymbol{\theta})$ is continuous on $\mathbf{G}_2$. Therefore, the coefficients of the polynomial for computing the eigenvalues of $\mathbf{I}(\boldsymbol{\theta})$ are continuous and finite on $\mathbf{G}_2$, since m is finite. The eigenvalues of $\mathbf{I}(\boldsymbol{\theta})$ are roots of this polynomial. Therefore, the eigenvalues of $\mathbf{I}(\boldsymbol{\theta})$ are continuous functions on $\mathbf{G}_2$ {Ortega (1932), page 45}.

It follows that, $\lambda_1(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$ on $\mathbf{G}_2$. So we have $\lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0} \lambda_1(\boldsymbol{\theta}) = \lambda_1(\boldsymbol{\theta}_0)$. Then there exists $\delta_2 > 0$ such that $0.6\lambda_1(\boldsymbol{\theta}_0) < \lambda_1(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \mathbf{G}_2 = \{\boldsymbol{\theta} :\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|< \delta_2\}$. Let $\delta = \min\{\delta_1, \delta_2\}$. Then there exists an open ball $\mathbf{G} = \{\boldsymbol{\theta} :\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|< \delta\} \subseteq \boldsymbol{\Omega}(\boldsymbol{\theta_0})$ and $N \in \mathbb{N}$ such that $0 < 0.1\lambda_1(\boldsymbol{\theta}_0) < \frac{1}{n}\lambda_{\min}\mathbf{F}_n(\boldsymbol{\theta})$ and $0 < 0.6\lambda_1(\boldsymbol{\theta}_0) < \lambda_1(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \mathbf{G}$ and for all $n \geq N$. Hence part (i) of this Lemma holds.

(ii) From part (i), we have $0.1\lambda_1(\boldsymbol{\theta}_0) < \frac{1}{n}\lambda_{\min}\mathbf{F}_n(\boldsymbol{\theta}_0)$ for all $n \geq N$. Therefore,

$$\lambda_{\min}\mathbf{F}_n(\boldsymbol{\theta_0}) \xrightarrow{a.s} \infty. \tag{3.16}$$

Furthermore, $\frac{1}{n}\lambda_{\max}\mathbf{F}_n(\boldsymbol{\theta_0}) \xrightarrow{a.s} \lambda_q(\boldsymbol{\theta_0})$, where $\lambda_q(\boldsymbol{\theta_0})$ is the maximum eigenvalue of the positive definite matrix $\mathbf{I}(\boldsymbol{\theta_0})$. Let $\epsilon > 0$ then, there exits $N_3 \in \mathbb{N}$ such that

$\lambda_q(\boldsymbol{\theta_0}) - \epsilon < \frac{1}{n}\lambda_{\max}\mathbf{F}_n(\boldsymbol{\theta}_0) < \lambda_q(\boldsymbol{\theta_0}) + \epsilon$ for all $n \geq N_3$. Choose $\epsilon = 0.5\lambda_q(\boldsymbol{\theta_0})$, then $0 < 0.5\lambda_q(\boldsymbol{\theta_0}) < \frac{1}{n}\lambda_{\max}\mathbf{F}_n(\boldsymbol{\theta}_0) < 1.5\lambda_q(\boldsymbol{\theta_0})$ for all $n \geq N_3$, and

$$\left\{\frac{1}{\mathrm{n}}\lambda_{\max}\mathbf{F}_{\mathrm{n}}(\boldsymbol{\theta}_0)\right\}^{1/4} < \{1.5\lambda_{\mathrm{q}}(\boldsymbol{\theta_0})\}^{1/4} \text{ for all n} \geq \mathrm{N}_3 \tag{3.17}$$

Let $N_4 = \max\{N, N_3\}$. Choose $\tau = 2$ and consider

$$\begin{aligned}\frac{\lambda_{\min}\mathbf{F}_n(\boldsymbol{\theta})}{[\lambda_{\max}\mathbf{F}_n(\boldsymbol{\theta_0})]^{1/(2+\tau)}} &= \frac{(1/n)\lambda_{\min}\mathbf{F}_n(\boldsymbol{\theta})}{[(1/n)\lambda_{\max}\mathbf{F}_n(\boldsymbol{\theta_0})]^{0.25}}n^{0.75} \\ &> \frac{0.1\lambda_1(\boldsymbol{\theta}_0)}{\{1.5\lambda_q(\boldsymbol{\theta_0})\}^{0.25}} \\ &= c \text{ (say) for all } n > N_4 \text{ and } \boldsymbol{\theta} \in \mathbf{G}. \end{aligned} \tag{3.18}$$

Using (3.16) and (3.18) we obtain $\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s} \boldsymbol{\theta_0}$ as $n \to \infty$ [Theorem 2 of Fahrmeir and Kaufmann (1985)].

□

Theorem 3.4.1. If the Assumptions 3.4.1 and 3.4.2 are satisfied with $m < \infty$, $q < \infty$, then we have that $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})$ is asymptotically multivariate normal in distribution with mean $\mathbf{0}$ and variance-covariance matrix $\mathbf{I}(\boldsymbol{\theta}_0)^{-1}$, where $\mathbf{I}(\boldsymbol{\theta}_0)$ is a Fisher information matrix.

Proof. We have $\mathbf{s}_n(\boldsymbol{\theta_0}) = \sum_{i=1}^{n} y_i\mathbf{w}_i - \sum_{i=1}^{n}[1 + \exp(-\mathbf{w}_i'\boldsymbol{\theta}_0)]^{-1}\mathbf{w}_i$. Then
$\mathbf{s}_{n+1}(\boldsymbol{\theta_0}) - \mathbf{s}_n(\boldsymbol{\theta_0}) = y_{n+1}\mathbf{w}_{n+1} - [1 + \exp(-\mathbf{w}_{n+1}'\boldsymbol{\theta}_0)]^{-1}\mathbf{w}_{n+1}$.
Consider
$E[\mathbf{s}_{n+1}(\boldsymbol{\theta_0}) - \mathbf{s}_n(\boldsymbol{\theta_0})|\mathcal{F}_n, C(\mathbf{z_n}), x_{(n+1)A}]$
$= E[y_{n+1}|x_{(n+1)A}]\mathbf{w}_{n+1} - [1 + \exp(-\mathbf{w}_{n+1}'\boldsymbol{\theta}_0)]^{-1}\mathbf{w}_{n+1}$
$= 0.$

Therefore $E[\mathbf{s}_{n+1}(\boldsymbol{\theta_0}) - \mathbf{s}_n(\boldsymbol{\theta_0})|\mathcal{F}_n, C(\mathbf{z_n})] = 0$.

Then $\{\mathbf{s}_n(\boldsymbol{\theta_0})\}$ is a sequence of martingale arrays. Under some regularity conditions, applying the martingale central limit theorem, $\frac{1}{\sqrt{n}}\mathbf{s}_n(\boldsymbol{\theta_0})$ follows multivariate normal distribution.

Let $\mathbf{x}_n$ be the history of treatment assignments of all n patients. That is, $\mathbf{x}_n = (x_{1A}, x_{2A}, \cdots, x_{nA})$. Then we have

$$
\begin{aligned}
&E[\mathbf{s}_n(\boldsymbol{\theta_0})|\mathbf{x}_n] \\
&= \sum_{i=1}^{n} E[y_i|x_{iA}]\mathbf{w}_i - \sum_{i=1}^{n}[1+\exp(-\mathbf{w}_i'\boldsymbol{\theta}_0)]^{-1}\mathbf{w}_i \\
&= 0.
\end{aligned}
$$

Therefore,

$$E[\mathbf{s}_n(\boldsymbol{\theta_0})] = \mathbf{0}_{q\times 1}. \tag{3.19}$$

Now,

$$
\begin{aligned}
\mathrm{Var}[\mathbf{s}_n(\boldsymbol{\theta_0})|\mathbf{x}_n] &= \mathrm{Var}\left[\sum_{i=1}^{n} y_i\mathbf{w}_i|\mathbf{x}_n\right] \\
&= \sum_{i=1}^{n}\mathrm{Var}[y_i\mathbf{w}_i|x_{iA}] \\
&= \sum_{i=1}^{n}\exp(-\mathbf{w}_i'\boldsymbol{\theta}_0)[1+\exp(-\mathbf{w}_i'\boldsymbol{\theta}_0)]^{-2}\mathbf{w}_i\mathbf{w}_i' \\
&= \mathbf{F}_n(\boldsymbol{\theta_0}),
\end{aligned}
$$

and, by Lemma (3.4.3)

$$\begin{aligned}
\frac{1}{n}\text{Var}[\mathbf{s}_n(\boldsymbol{\theta_0})] &= \frac{1}{n}E_{\mathbf{x}_n}\{\text{Var}[\mathbf{s}_n(\boldsymbol{\theta_0})]\} + \frac{1}{n}\text{Var}_{\mathbf{x}_n}\{E[\mathbf{s}_n(\boldsymbol{\theta_0})]\} \\
&= \frac{1}{n}E_{\mathbf{x}_n}\{\text{Var}[\mathbf{s}_n(\boldsymbol{\theta_0})]\} \\
&= \frac{1}{n}E_{\mathbf{x}_n}\{\mathbf{F}_n(\boldsymbol{\theta_0})\} \\
&\xrightarrow{L} \mathbf{I}(\boldsymbol{\theta}_0). \qquad (3.20)
\end{aligned}$$

Using (3.19), (3.20), and the fact that $\frac{1}{\sqrt{n}}\mathbf{s}_n(\boldsymbol{\theta_0})$ follows multivariate normal distribution we have that

$$\frac{1}{\sqrt{n}}\mathbf{s}_n(\boldsymbol{\theta_0}) \sim N_q[\mathbf{0}, \mathbf{I}(\boldsymbol{\theta}_0)]. \qquad (3.21)$$

Apply the multivariate version of the Taylor's expansion {Königsberger (2004), page 66} for the score function $\mathbf{s}_n(\boldsymbol{\theta})$ to obtain

$$\mathbf{0} = \mathbf{s}_n(\hat{\boldsymbol{\theta}}_n) = \mathbf{s}_n(\boldsymbol{\theta_0}) - \mathbf{F}_n(\boldsymbol{\theta_0})(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) + \mathbf{o}(\parallel \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0} \parallel^q), \qquad (3.22)$$

where $\mathbf{o}(\parallel \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0} \parallel^q) = [o_1(\parallel \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0} \parallel^q), o_2(\parallel \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0} \parallel^q), \cdots, o_q(\parallel \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0} \parallel^q)]'$ and the error vector, $\mathbf{o}(\parallel \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0} \parallel^q)$, in (3.22) goes to $\mathbf{0}_{q\times 1}$ faster than $\parallel \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0} \parallel^q$ goes to zero when $\hat{\boldsymbol{\theta}}_n$ is near to $\boldsymbol{\theta_0}$.

Moreover $\hat{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta_0}$ satisfying the conditions in Lemma (3.4.6). Thus, the linear approximation in (3.22) is good enough for $\mathbf{s}_n(\hat{\boldsymbol{\theta}}_n)$. Under these

conditions, we have that

$$\begin{aligned} \mathbf{F}_n(\boldsymbol{\theta_0})(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) &\approx \mathbf{s}_n(\boldsymbol{\theta_0}) \\ \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) &\approx \left[\frac{1}{n}\mathbf{F}_n(\boldsymbol{\theta_0})\right]^{-1}\left[\frac{1}{\sqrt{n}}\mathbf{s}_n(\boldsymbol{\theta_0})\right] \end{aligned} \tag{3.23}$$

since the remainder term goes faster to zero when $\hat{\boldsymbol{\theta}}_n$ closes to $\boldsymbol{\theta_0}$. Using **Lemma 3.4.3**, we have that

$$\begin{aligned} \frac{1}{n}\mathbf{F}_n(\boldsymbol{\theta_0}) &\xrightarrow{a.s.} \mathbf{I}(\boldsymbol{\theta}_0) \\ \Rightarrow \left[\frac{1}{n}\mathbf{F}_n(\boldsymbol{\theta_0})\right]^{-1} &\xrightarrow{a.s.} \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \end{aligned} \tag{3.24}$$

Therefore it follow, from (3.24), (3.21), and (3.23), we have that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) \sim N\left[\mathbf{0}, \mathbf{I}(\boldsymbol{\theta}_0)^{-1}\right]. \tag{3.25}$$

□

3.5 A Limiting Allocation for Response-Adaptive Design based on the Most Natural Mapping

In § 3.3, we discussed CARA design. The model based odds-ratio that was obtained from ML estimates and the most natural mapping were used to formulate this CARA design. However, RA designs have been well developed in literature. Thus, this RA designs have to be compared with CARA design. Moreover, we have many types of RA designs in literature. But, we selected the odds-ratio-based response-adaptive design because this design has similar formulation comparable with CARA design. In this

section, we discuss odds-ratio-based response-adaptive designs that was introduced by Basak et al. (2009). The odds ratio for comparing treatment A against treatment B is given by,

$$OR(P_{AS}, P_{BS}) = \frac{P_{AS}}{(1 - P_{AS})} \frac{(1 - P_{BS})}{P_{BS}}, \tag{3.26}$$

where P_{AS} and P_{BS} are the success probabilities for those patients assigned to treatment A and B, respectively.

The odds-ratio-based limiting allocation to treatment A using the most natural mapping is provided

$$\rho(P_{AS}, P_{BS}) = \frac{OR(P_{AS}, P_{BS})}{1 + OR(P_{AS}, P_{BS})}. \tag{3.27}$$

In our simulation study for this chapter, we use the allocation function of DBCD that is defined in (1.5) to target this limiting allocation.

3.6 Simulation Studies

In this section, we validate the theoretical results we obtained in § 3.4. Suppose, treatments A and B are to be compared among patients who have a disease, considering the covariates: gender, chronic conditions, age. The description of responses, treatment assignments, and covariates are given by

$$Y_i = \begin{cases} 1 & \text{if treatment is} \\ & \text{a success,} \\ 0 & \text{otherwise.} \end{cases} \quad \text{and } X_{iA} = \begin{cases} 1 & \text{if the patient } i \text{ is assigned} \\ & \text{to treatment } A, \\ 0 & \text{otherwise.} \end{cases}$$

$$Z_{i1} = \begin{cases} 1 & \text{Male (55\%)}, \\ 0 & \text{Female.} \end{cases} \quad \text{and } Z_{i2} = \begin{cases} 1 & \text{at least one chronic condition (60\%)}, \\ 0 & \text{otherwise.} \end{cases}$$

$$Z_{i3} = \begin{cases} 1 & 20 \leq \text{Age} \leq 50 \text{ (30\%)}, \\ 0 & \text{otherwise.} \end{cases} \quad \text{and } Z_{i4} = \begin{cases} 1 & 50 < \text{Age} \leq 65 \text{ (30\%)}, \\ 0 & \text{otherwise.} \end{cases}$$

In this simulation study, the patient responses are assumed to be instantaneous. The following model is assumed to be the true statistical model for simulation studies,

$$\begin{aligned} \text{logit}[P(Y_i = 1|x_{iA})] &= x_{iA}\beta_A + \gamma_0 + \gamma_1 z_{i1} + \gamma_2 z_{i2} + \gamma_3 z_{i3} + \gamma_4 z_{i4} + \delta_1 x_{iA} z_{i1} + \\ &\quad \delta_2 x_{iA} z_{i2} + \delta_3 x_{iA} z_{i3} + \delta_4 x_{iA} z_{i4}, \\ &= \mathbf{w}_i'\boldsymbol{\theta} \text{ for } i = 1, 2, \cdots, n \end{aligned} \tag{3.28}$$

where $\boldsymbol{\theta} = (\beta_A, \gamma_0, \boldsymbol{\gamma}', \boldsymbol{\delta}')'$ and $\mathbf{w}_i = (x_{iA}, 1, \mathbf{z}_i', x_{iA}\mathbf{z}_i')'$. Also, $P(Y_i = 1|x_{iA}) = [1 + \exp(-\mathbf{w}_i'\boldsymbol{\theta})]^{-1}$ and $P(Y_i = 0|x_{iA}) = [1 + \exp(\mathbf{w}_i'\boldsymbol{\theta})]^{-1}$.

Since we have four binary covariates, Z_1, Z_2, Z_3, Z_4, we can form 16 configuration' levels that are defined in Table 3.1.

In this simulation study, the three sets of true parameter values for the logit model in (3.28) we selected are:

(a) $\beta_{A0} = 1.25, \gamma_{00} = 0.50, \gamma_{10} = -0.18, \gamma_{20} = -0.30, \gamma_{30} = 0.25, \gamma_{40} = 0.10, \delta_{10} = 0.00, \delta_{20} = 0.00, \delta_{30} = 0.00, \delta_{40} = 0.00$

(b) $\beta_{A0} = 1.25, \gamma_{00} = 0.5, \gamma_{10} = -0.22, \gamma_{20} = -0.4, \gamma_{30} = 0.2, \gamma_{40} = 0.1, \delta_{10} =$

Table 3.1: The levels of configurations(CL) or strata

Level(h)	z_1	z_2	z_3	z_4
S1	0	0	0	0
S2	0	0	0	1
S3	0	0	1	0
S4	0	0	1	1
S5	0	1	0	0
S6	0	1	0	1
S7	0	1	1	0
S8	0	1	1	1
S9	1	0	0	0
S10	1	0	0	1
S11	1	0	1	0
S12	1	0	1	1
S13	1	1	0	0
S14	1	1	0	1
S15	1	1	1	0
S16	1	1	1	1

$0.09, \delta_{20} = -0.8, \delta_{30} = 0.06, \delta_{40} = 0.04$

(c) $\beta_{A0} = 0, \gamma_{00} = 0.25, \gamma_{10} = -0.20, \gamma_{20} = -0.35, \gamma_{30} = 0.25, \gamma_{40} = 0.15, \delta_{10} = 0.10, \delta_{20} = -1.50, \delta_{30} = 0.05, \delta_{40} = 0.05.$

The set (a) represents no treatment by covariates interactions in the true model. Thus, treatment A is the best treatment because β_{A0} is positive. However, in sets (b) and (c), the true model contains treatment by covaraite interactions. Thus, neither treatment A nor treatment B is the globally best treatment. We consider these three scenarios to conduct simulation study. The selected number of patients were 500 and 1000 to implement this simulation study through 3000 simulations.

In CARA design, the probabilities of treatment assignments depend on the current patients' covariates. On the contrary, these probabilities does not depend on the current patients' covariates in RA designs. In this simulation study, CARA design is

Table 3.2: CARA, RA, and CR procedures from 3000 simulations with 500 number of patients. The proportion of patients assigned to treatment A in stratum h ($\hat{P}_{Ah}$) for $h = 1, 2, \cdots, m$, simulated means (SM), simulated standard errors (SSE) with the model parameters $\beta_{A0} = 1.25$, $\gamma_{00} = 0.50$, $\gamma_{10} = -0.18$, $\gamma_{20} = -0.30$, $\gamma_{30} = 0.25$, $\gamma_{40} = 0.10$, $\delta_{10} = 0.00$, $\delta_{20} = 0.00$, $\delta_{30} = 0.00$, $\delta_{40} = 0.00$.

		Stratum h							
Design	Quantity	S1	S2	S3	S4	S5	S6	S7	S8
CARA	SM($\hat{P}_{Ah}$)	0.7308	0.7286	0.7260	0.7228	0.7305	0.7291	0.7220	0.7206
	SSE($\hat{P}_{Ah}$)	0.1194	0.1422	0.1572	0.1902	0.1031	0.1255	0.1366	0.1685
RA	SM($\hat{P}_{Ah}$)	0.7748	0.7734	0.7754	0.7753	0.7758	0.7768	0.7762	0.7719
	SSE($\hat{P}_{Ah}$)	0.0791	0.1024	0.1122	0.1496	0.0664	0.0836	0.0939	0.1211
CR	SM($\hat{P}_{Ah}$)	0.4964	0.5000	0.4995	0.4961	0.4981	0.4995	0.4994	0.5012
	SSE($\hat{P}_{Ah}$)	0.0790	0.1094	0.1246	0.1712	0.0650	0.0889	0.0967	0.1360
Design	Quantity	S9	S10	S11	S12	S13	S14	S15	S16
CARA	SM($\hat{P}_{Ah}$)	0.7316	0.7271	0.7238	0.7211	0.7309	0.7267	0.7246	0.7225
	SSE($\hat{P}_{Ah}$)	0.1095	0.1367	0.1426	0.1807	0.0955	0.1202	0.1257	0.1580
RA	SM($\hat{P}_{Ah}$)	0.7754	0.7736	0.7746	0.7740	0.7756	0.7754	0.7764	0.7760
	SSE($\hat{P}_{Ah}$)	0.0718	0.0902	0.1014	0.1349	0.0633	0.0775	0.0838	0.1082
CR	SM($\hat{P}_{Ah}$)	0.5012	0.4978	0.5011	0.4976	0.5006	0.5018	0.4990	0.5022
	SSE($\hat{P}_{Ah}$)	0.0720	0.0968	0.1097	0.1538	0.0580	0.0781	0.0866	0.1247

Table 3.3: CARA, RA, and CR procedures from 3000 simulations with 500 number of patients. The proportion of patients assigned to treatment A ($\hat{P}_A$), the success rates of patients ($\hat{P}_S$), simulated means (SM), and simulated standard errors (SSE) with the model parameters $\beta_{A0} = 1.25$, $\gamma_{00} = 0.50$, $\gamma_{10} = -0.18$, $\gamma_{20} = -0.30$, $\gamma_{30} = 0.25$, $\gamma_{40} = 0.10$, $\delta_{10} = 0.00$, $\delta_{20} = 0.00$, $\delta_{30} = 0.00$, $\delta_{40} = 0.00$.

Quantity	CARA	RA	CR
SM($\hat{P}_A$)	0.7279	0.7752	0.4996
SSE($\hat{P}_A$)	0.0504	0.0455	0.0222
SM($\hat{P}_S$)	0.7602	0.7718	0.7044
SSE($\hat{P}_S$)	0.0213	0.0214	0.0202

compared to CR design and RA design that is described in § 3.5. We have already mentioned that neither treatment A nor treatment B is the globally best treatment if treatment by covariate interactions exist. Thus, we compute the combined proportion of success rates of treatment A and B, say $\hat{P}_S$, to compare three designs because we would not come to any conclusion based on the results that the proportion of patients assigned to treatment A, say $\hat{P}_A$. Furthermore, we calculated the proportion of patients assigned to treatment A for stratum h, say $\hat{P}_{Ah}$, $h = 1, 2, \cdots, 16$. In fact, $\hat{P}_S$ is a measure that can be used to compare participating patients' ethics among three designs.

Table 3.4: CARA, RA, and CR procedures from 3000 simulations with 500 number of patients. Simulated means (SM), simulated standard error (SSE), estimated standard error (ESE), and coverage probability (CP) with the model parameters $\beta_{A0} = 1.25$, $\gamma_{00} = 0.50$, $\gamma_{10} = -0.18$, $\gamma_{20} = -0.30$, $\gamma_{30} = 0.25$, $\gamma_{40} = 0.10$, $\delta_{10} = 0.00$, $\delta_{20} = 0.00$, $\delta_{30} = 0.00$, $\delta_{40} = 0.00$.

Design	Quantity	$\hat{\beta}_A$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_3$	$\hat{\delta}_4$
CARA	SM	1.359	0.418	-0.183	-0.294	0.226	0.082	-0.006	-0.026	0.029	0.023
	SSE	0.561	0.449	0.397	0.416	0.443	0.446	0.496	0.521	0.556	0.548
	ESE	0.522	0.410	0.375	0.384	0.412	0.394	0.474	0.487	0.526	0.498
	CP	0.940	0.942	0.944	0.939	0.947	0.939	0.946	0.936	0.946	0.937
RA	SM	1.300	0.484	-0.181	-0.315	0.274	0.105	-0.004	-0.002	-0.014	0.004
	SSE	0.567	0.472	0.429	0.440	0.479	0.458	0.506	0.527	0.576	0.551
	ESE	0.545	0.446	0.412	0.420	0.454	0.431	0.498	0.510	0.552	0.521
	CP	0.940	0.947	0.948	0.946	0.941	0.944	0.950	0.948	0.939	0.944
CR	SM	1.287	0.512	-0.190	-0.296	0.256	0.106	0.005	-0.024	0.017	0.017
	SSE	0.495	0.292	0.267	0.276	0.296	0.283	0.445	0.458	0.511	0.470
	ESE	0.487	0.287	0.265	0.270	0.290	0.277	0.440	0.453	0.491	0.461
	CP	0.952	0.949	0.950	0.950	0.947	0.952	0.950	0.954	0.949	0.953

We also computed the simulated means (SM) to confirm the consistency of parameter estimates. The estimated standard error (ESE) is computed using the inverse of the Hessian matrix obtained at the final stage of the maximization process. Also, the simulated standard error (SSE) is calculated to validate whether ESE can be used in real data analysis. That is, if the values of SSE and ESE are close, we can use ESE

for real data analysis. The normal distribution based coverage probability (CP) is computed to verify whether the parameter estimates asymptotically follow the normal distribution.

Table 3.5: CARA, RA, and CR procedures from 3000 simulations with 500 number of patients. The proportion of patients assigned to treatment A in stratum h ($\hat{P}_{Ah}$) for $h = 1, 2, \cdots, m$, simulated means (SM), simulated standard errors (SSE) with the model parameters $\beta_{A0} = 1.25$, $\gamma_{00} = 0.5$, $\gamma_{10} = -0.22$, $\gamma_{20} = -0.4$, $\gamma_{30} = 0.2$, $\gamma_{40} = 0.1$, $\delta_{10} = 0.09$, $\delta_{20} = -0.8$, $\delta_{30} = 0.06$, $\delta_{40} = 0.04$.

		Stratum h							
Design	Quantity	S1	S2	S3	S4	S5	S6	S7	S8
CARA	SM($\hat{P}_{Ah}$)	0.7373	0.7402	0.7412	0.7414	0.5953	0.6035	0.6047	0.6123
	SSE($\hat{P}_{Ah}$)	0.1171	0.1421	0.1522	0.1872	0.1236	0.1514	0.1618	0.1920
RA	SM($\hat{P}_{Ah}$)	0.6836	0.6829	0.6844	0.6807	0.6822	0.6856	0.6848	0.6814
	SSE($\hat{P}_{Ah}$)	0.0857	0.1108	0.1240	0.1661	0.0729	0.0921	0.1020	0.1341
CR	SM($\hat{P}_{Ah}$)	0.4964	0.4999	0.4994	0.4961	0.4982	0.4995	0.4995	0.5012
	SSE($\hat{P}_{Ah}$)	0.0790	0.1094	0.1247	0.1712	0.0651	0.0889	0.0967	0.1360
Design	Quantity	S9	S10	S11	S12	S13	S14	S15	S16
CARA	SM($\hat{P}_{Ah}$)	0.7530	0.7554	0.7532	0.7564	0.6157	0.6232	0.6240	0.6351
	SSE($\hat{P}_{Ah}$)	0.1045	0.1272	0.1347	0.1717	0.1147	0.1404	0.1477	0.1767
RA	SM($\hat{P}_{Ah}$)	0.6834	0.6813	0.6835	0.6788	0.6835	0.6833	0.6841	0.6812
	SSE($\hat{P}_{Ah}$)	0.0807	0.1013	0.1129	0.1507	0.0668	0.0847	0.0928	0.1229
CR	SM($\hat{P}_{Ah}$)	0.5012	0.4978	0.5011	0.4978	0.5007	0.5018	0.4990	0.5022
	SSE($\hat{P}_{Ah}$)	0.0720	0.0968	0.1096	0.1540	0.0580	0.0781	0.0866	0.1247

The results in Tables 3.2, 3.5, 3.8, and 3.11 show that the proportion of patients assigned to treatment A are (i) approximately equal in each stratum for RA design, (ii) approximately 0.5 in each stratum for CR design, and (iii) different in values among strata for CARA design except in Table 3.2. This shows that the proportion of patients assigned to treatment A are equal in each stratum for CARA design when there is no treatment by covariate interaction in the true model [see Table 3.2].

Table 3.6: CARA, RA, and CR procedures from 3000 simulations with 500 number of patients. The proportion of patients assigned to treatment A ($\hat{P}_A$), the success rates of patients ($\hat{P}_S$), simulated means (SM), and simulated standard errors (SSE) with the model parameters $\beta_{A0} = 1.25$, $\gamma_{00} = 0.5$, $\gamma_{10} = -0.22$, $\gamma_{20} = -0.4$, $\gamma_{30} = 0.2$, $\gamma_{40} = 0.1$, $\delta_{10} = 0.09$, $\delta_{20} = -0.8$, $\delta_{30} = 0.06$, $\delta_{40} = 0.04$.

Quantity	CARA	RA	CR
SM($\hat{P}_A$)	0.6664	0.6832	0.4996
SSE($\hat{P}_A$)	0.0535	0.0484	0.0222
SM($\hat{P}_S$)	0.6767	0.6750	0.6435
SSE($\hat{P}_S$)	0.0221	0.0217	0.0214

Table 3.7: CARA, RA, and CR procedures from 3000 simulations with 500 number of patients. Simulated means (SM), simulated standard error (SSE), estimated standard error (ESE), and coverage probability (CP) with the model parameters $\beta_{A0} = 1.25$, $\gamma_{00} = 0.5$, $\gamma_{10} = -0.22$, $\gamma_{20} = -0.4$, $\gamma_{30} = 0.2$, $\gamma_{40} = 0.1$, $\delta_{10} = 0.09$, $\delta_{20} = -0.8$, $\delta_{30} = 0.06$, $\delta_{40} = 0.04$.

Design	Quantity	$\hat{\beta}_A$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_3$	$\hat{\delta}_4$
CARA	SM	1.353	0.415	-0.224	-0.370	0.173	0.082	0.101	-0.873	0.092	0.060
	SSE	0.540	0.427	0.359	0.390	0.414	0.378	0.462	0.498	0.526	0.488
	ESE	0.502	0.394	0.331	0.366	0.368	0.350	0.426	0.468	0.474	0.449
	CP	0.941	0.949	0.947	0.945	0.934	0.947	0.939	0.943	0.940	0.937
RA	SM	1.291	0.485	-0.227	-0.412	0.218	0.112	0.102	-0.823	0.058	0.036
	SSE	0.492	0.372	0.342	0.347	0.387	0.356	0.428	0.463	0.493	0.445
	ESE	0.483	0.366	0.337	0.344	0.369	0.352	0.425	0.453	0.468	0.445
	CP	0.951	0.954	0.948	0.954	0.942	0.952	0.953	0.948	0.937	0.955
CR	SM	1.274	0.515	-0.239	-0.408	0.212	0.102	0.114	-0.831	0.063	0.054
	SSE	0.478	0.290	0.262	0.274	0.288	0.281	0.402	0.448	0.458	0.424
	ESE	0.469	0.286	0.263	0.269	0.288	0.275	0.402	0.438	0.444	0.421
	CP	0.954	0.948	0.952	0.947	0.950	0.946	0.950	0.948	0.950	0.949

Table 3.8: CARA, RA, and CR procedures from 3000 simulations with 500 number of patients. The proportion of patients assigned to treatment A in stratum h ($\hat{P}_{Ah}$) for $h = 1, 2, \cdots, m$, simulated means (SM), simulated standard errors (SSE) with the model parameters $\beta_{A0} = 0$, $\gamma_{00} = 0.25$, $\gamma_{10} = -0.20$, $\gamma_{20} = -0.35$, $\gamma_{30} = 0.25$, $\gamma_{40} = 0.15$, $\delta_{10} = 0.10$, $\delta_{20} = -1.50$, $\delta_{30} = 0.05$, $\delta_{40} = 0.05$.

		Stratum h							
Design	Quantity	S1	S2	S3	S4	S5	S6	S7	S8
CARA	SM($\hat{P}_{Ah}$)	0.4962	0.5072	0.5079	0.5127	0.2108	0.2239	0.2233	0.2348
	SSE($\hat{P}_{Ah}$)	0.1513	0.1767	0.1913	0.2318	0.0948	0.1142	0.1254	0.1557
RA	SM($\hat{P}_{Ah}$)	0.3339	0.3326	0.3345	0.3314	0.3326	0.3336	0.3342	0.3311
	SSE($\hat{P}_{Ah}$)	0.0879	0.1152	0.1256	0.1704	0.0712	0.0924	0.1001	0.1356
CR	SM($\hat{P}_{Ah}$)	0.4964	0.4999	0.4994	0.4961	0.4982	0.4995	0.4995	0.5012
	SSE($\hat{P}_{Ah}$)	0.0790	0.1094	0.1247	0.1712	0.0651	0.0889	0.0967	0.1360
Design	Quantity	S9	S10	S11	S12	S13	S14	S15	S16
CARA	SM($\hat{P}_{Ah}$)	0.5214	0.5275	0.5271	0.5319	0.2244	0.2368	0.2370	0.2469
	SSE($\hat{P}_{Ah}$)	0.1416	0.1716	0.1778	0.2153	0.0955	0.1160	0.1209	0.1527
RA	SM($\hat{P}_{Ah}$)	0.3342	0.3356	0.3315	0.3317	0.3325	0.3338	0.3339	0.3308
	SSE($\hat{P}_{Ah}$)	0.0810	0.1052	0.1165	0.1522	0.0646	0.0844	0.0898	0.1253
CR	SM($\hat{P}_{Ah}$)	0.5012	0.4978	0.5011	0.4978	0.5007	0.5018	0.4990	0.5022
	SSE($\hat{P}_{Ah}$)	0.0720	0.0968	0.1096	0.1540	0.0580	0.0781	0.0866	0.1247

Table 3.9: CARA, RA, and CR procedures from 3000 simulations with 500 number of patients. The proportion of patients assigned to treatment A ($\hat{P}_A$), the success rates of patients ($\hat{P}_S$), simulated means (SM), and simulated standard errors (SSE) with the model parameters $\beta_{A0} = 0$, $\gamma_{00} = 0.25$, $\gamma_{10} = -0.20$, $\gamma_{20} = -0.35$, $\gamma_{30} = 0.25$, $\gamma_{40} = 0.15$, $\delta_{10} = 0.10$, $\delta_{20} = -1.50$, $\delta_{30} = 0.05$, $\delta_{40} = 0.05$.

Quantity	CARA	RA	CR
SM($\hat{P}_A$)	0.3417	0.3331	0.4996
SSE($\hat{P}_A$)	0.0572	0.0472	0.0222
SM($\hat{P}_S$)	0.4787	0.4577	0.4296
SSE($\hat{P}_S$)	0.0248	0.0234	0.0225

Table 3.10: CARA, RA, and CR procedures from 3000 simulations with 500 number of patients. Simulated means (SM), simulated standard error (SSE), estimated standard error (ESE), and coverage probability (CP) with the model parameters $\beta_{A0} = 0$, $\gamma_{00} = 0.25$, $\gamma_{10} = -0.20$, $\gamma_{20} = -0.35$, $\gamma_{30} = 0.25$, $\gamma_{40} = 0.15$, $\delta_{10} = 0.10$, $\delta_{20} = -1.50$, $\delta_{30} = 0.05$, $\delta_{40} = 0.05$.

Design	Quantity	$\hat{\beta}_A$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_3$	$\hat{\delta}_4$
CARA	SM	-0.017	0.213	-0.203	-0.323	0.249	0.146	0.108	1.670	0.025	0.042
	SSE	0.487	0.282	0.231	0.264	0.252	0.249	0.485	0.540	0.516	0.491
	ESE	0.437	0.269	0.228	0.252	0.249	0.239	0.429	0.492	0.468	0.448
	CP	0.935	0.946	0.949	0.944	0.947	0.946	0.928	0.946	0.939	0.938
RA	SM	-0.011	0.249	-0.205	-0.356	0.261	0.152	0.106	-1.559	0.040	0.046
	SSE	0.462	0.247	0.227	0.231	0.250	0.238	0.459	0.467	0.501	0.477
	ESE	0.452	0.243	0.226	0.230	0.246	0.235	0.442	0.447	0.479	0.461
	CP	0.948	0.943	0.951	0.951	0.946	0.953	0.944	0.948	0.947	0.943
CR	SM	0.008	0.257	-0.214	-0.361	0.265	0.156	0.112	-1.530	0.029	0.041
	SSE	0.427	0.288	0.260	0.264	0.286	0.276	0.403	0.404	0.443	0.414
	ESE	0.414	0.282	0.261	0.266	0.285	0.273	0.399	0.403	0.433	0.416
	CP	0.944	0.947	0.956	0.951	0.952	0.948	0.944	0.951	0.939	0.954

Table 3.11: CARA, RA, and CR procedures from 3000 simulations with 1000 number of patients. The proportion of patients assigned to treatment A in stratum h ($\hat{P}_{Ah}$) for $h = 1, 2, \cdots, m$, simulated means (SM), simulated standard errors (SSE) with the model parameters $\beta_{A0} = 0$, $\gamma_{00} = 0.25$, $\gamma_{10} = -0.20$, $\gamma_{20} = -0.35$, $\gamma_{30} = 0.25$, $\gamma_{40} = 0.15$, $\delta_{10} = 0.10$, $\delta_{20} = -1.50$, $\delta_{30} = 0.05$, $\delta_{40} = 0.05$.

		Stratum h							
Design	Quantity	S1	S2	S3	S4	S5	S6	S7	S8
CARA	SM($\hat{P}_{Ah}$)	0.5006	0.5036	0.5055	0.5118	0.1917	0.1997	0.1990	0.2074
	SSE($\hat{P}_{Ah}$)	0.1123	0.1382	0.1437	0.1749	0.0701	0.0864	0.0910	0.1134
RA	SM($\hat{P}_{Ah}$)	0.3330	0.3341	0.3359	0.3359	0.3348	0.3329	0.3339	0.3340
	SSE($\hat{P}_{Ah}$)	0.0630	0.0800	0.0892	0.1166	0.0499	0.0649	0.0731	0.0942
CR	SM($\hat{P}_{Ah}$)	0.4994	0.5012	0.4990	0.5004	0.4999	0.4999	0.4985	0.4972
	SSE($\hat{P}_{Ah}$)	0.0547	0.0766	0.0845	0.1179	0.0440	0.0616	0.0690	0.0962
Design	Quantity	S9	S10	S11	S12	S13	S14	S15	S16
CARA	SM($\hat{P}_{Ah}$)	0.5240	0.5270	0.5285	0.5353	0.2050	0.2123	0.2143	0.2200
	SSE($\hat{P}_{Ah}$)	0.1084	0.1285	0.1384	0.1652	0.0727	0.0881	0.0944	0.1116
RA	SM($\hat{P}_{Ah}$)	0.3346	0.3343	0.3311	0.3298	0.3338	0.3344	0.3330	0.3341
	SSE($\hat{P}_{Ah}$)	0.0586	0.0742	0.0810	0.1082	0.0464	0.0581	0.0658	0.0866
CR	SM($\hat{P}_{Ah}$)	0.5000	0.5022	0.4983	0.5014	0.5002	0.5003	0.5004	0.4980
	SSE($\hat{P}_{Ah}$)	0.0497	0.0700	0.0774	0.1054	0.0412	0.0564	0.0636	0.0871

Table 3.12: CARA, RA, and CR procedures from 3000 simulations with 1000 number of patients. The proportion of patients assigned to treatment A ($\hat{P}_A$), the success rates of patients ($\hat{P}_S$), simulated means (SM), and simulated standard errors (SSE) with the model parameters $\beta_{A0} = 0$, $\gamma_{00} = 0.25$, $\gamma_{10} = -0.20$, $\gamma_{20} = -0.35$, $\gamma_{30} = 0.25$, $\gamma_{40} = 0.15$, $\delta_{10} = 0.10$, $\delta_{20} = -1.50$, $\delta_{30} = 0.05$, $\delta_{40} = 0.05$.

Quantity	CARA	RA	CR
SM($\hat{P}_A$)	0.3289	0.3338	0.4999
SSE($\hat{P}_A$)	0.0459	0.0342	0.0161
SM($\hat{P}_S$)	0.4820	0.4578	0.4297
SSE($\hat{P}_S$)	0.0180	0.0164	0.0155

Table 3.13: CARA, RA, and CR procedures from 3000 simulations with 1000 number of patients. Simulated means (SM), simulated standard error (SSE), estimated standard error (ESE), and coverage probability (CP) with the model parameters $\beta_{A0} = 0$, $\gamma_{00} = 0.25$, $\gamma_{10} = -0.20$, $\gamma_{20} = -0.35$, $\gamma_{30} = 0.25$, $\gamma_{40} = 0.15$, $\delta_{10} = 0.10$, $\delta_{20} = -1.50$, $\delta_{30} = 0.05$, $\delta_{40} = 0.05$.

Design	Quantity	$\hat{\beta}_A$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_3$	$\hat{\delta}_4$
CARA	SM	0.002	0.229	-0.198	-0.338	0.242	0.153	0.104	-1.612	0.029	0.034
	SSE	0.322	0.192	0.160	0.180	0.172	0.167	0.316	0.389	0.342	0.338
	ESE	0.302	0.187	0.158	0.175	0.172	0.165	0.298	0.346	0.324	0.311
	CP	0.940	0.944	0.945	0.945	0.950	0.951	0.945	0.944	0.948	0.930
RA	SM	0.000	0.247	-0.197	-0.353	0.246	0.154	0.100	-1.524	0.051	0.050
	SSE	0.320	0.175	0.162	0.164	0.172	0.166	0.309	0.309	0.330	0.327
	ESE	0.312	0.170	0.158	0.161	0.172	0.165	0.305	0.308	0.330	0.318
	CP	0.948	0.946	0.943	0.946	0.951	0.955	0.950	0.951	0.948	0.944
CR	SM	0.005	0.250	-0.202	-0.352	0.250	0.156	0.099	-1.514	0.046	0.040
	SSE	0.295	0.201	0.187	0.192	0.198	0.193	0.278	0.287	0.304	0.293
	ESE	0.289	0.197	0.183	0.186	0.199	0.191	0.279	0.281	0.302	0.290
	CP	0.945	0.947	0.941	0.945	0.950	0.952	0.949	0.945	0.953	0.953

The values of $SM(\hat{P}_S)$ in Table 3.3 clearly show that RA and CARA designs increase the well-being of participating patients compared to CR designs because the values of $SM(\hat{P}_S)$ for RA, CARA, and CR designs are 77.18%, 76.02%, and 70.44%, respectively. Moreover, the values of $SM(\hat{P}_S)$ for RA design is higher than CARA design. Thus, RA design has been stably moving to achieve target allocation compared to CARA design. On the other hand, CARA design has three different stages: (i) the initial stage that is equal allocation, (ii) the second stage that attempts to detect treatment by covariate interactions, and (iii) the final stage that ethically reduces the number of patients assigned to inferior treatment. That is, after identification of treatment by covariate interactions, the success rates of participating patients are increased by CARA designs. In fact, this result confirms the statement of Ning and Huang (2010) that was mentioned in Section 1.4.1.

We now examine the scenario that true model contain treatment by covariate interactions. The results in Tables 3.6, 3.9, and 3.12 are the scenarios for which true models contain treatment by covariate interactions. These tables show that CARA and RA designs generate ethically desirable outcomes compared with CR design. Moreover, CARA design are more ethical than RA designs. In this chapter, the ethical measure is the value of $SM(\hat{P}_S)$. Also, we measure the amount of interaction from origin, which is (0,0,0,0), using the Euclidean distance denoted by EI. That is, the interaction measure is based on Euclidean distance. Therefore, EI for Table 3.6 = $\sqrt{0.09^2 + (-0.8)^2 + 0.06^2 + 0.04^2} = 0.8083$ and EI for Table 3.9 = $\sqrt{0.1^2 + (-1.5)^2 + 0.05^2 + 0.05^2} = 1.5050$. The difference of ethical measure between CARA and RA design for Table 3.6 = 67.67 - 67.50 = 0.17 and the difference of ethical measure between CARA and RA design for Table 3.9 = 47.87 - 45.77 = 2.1. These results show that, the difference in the ethical measure between CARA and RA

designs increases when the effect of treatment by covariate interactions increase.

When we investigate Table 3.9 and 3.12, the number of participating patients are the only difference between these two tables. That is, the number of participating patients in Table 3.9 is 500 and the number of participating patients in Table 3.12 is 1000. Moreover, the ethical measures for CARA, RA, and CR designs in Table 3.12 are increased by 0.33%, 0.01%, 0.01% respectively compared to Table 3.9. Thus, CARA design generate more ethically desirable outcome compared to RA and CR design after detecting the treatment by covariate interaction.

Because the parameter $\boldsymbol{\theta}$ is multi-dimensional, we use the Euclidean distance to compare the consistency and asymptotic normality among designs. We define the measure for consistency as the Euclidean distance of $\mathrm{SM}(\hat{\boldsymbol{\theta}})$ from $\boldsymbol{\theta}_0$, denoted by EC. The measure for normality is the Euclidean distance of $\mathrm{CP}(\hat{\boldsymbol{\theta}})$ from $(0.95, 0.95, \cdots, 0.95)'$ that is 10×1 dimension, denoted by EN. In Table 3.4, EC for CARA, RA, and CR design are 0.1471, 0.0618, and 0.0537 respectively. In Table 3.7, EC for CARA, RA, and CR design are 0.1633, 0.0571, and 0.0558 respectively. In the Table 3.10, EC for CARA, RA, and CR design are 3.1705, 0.0628, and 0.0475 respectively. In the Table 3.13, EC for CARA, RA, and CR design are 0.118, 0.0252, and 0.0195 respectively. According to these EC values, the ML estimator under CR design is slightly more consistent than the estimator under RA design, while the ML estimator for RA design is more consistent than CARA design. We report the values of EN to compare normality of ML estimators among designs. In Table 3.4, EN for CARA, RA, and CR design are 0.0291, 0.0202, and 0.0066 respectively. In Table 3.7, EN for CARA, RA, and CR design are 0.0286, 0.0177, and 0.0073 respectively. In Table 3.10, EN for CARA, RA, and CR design are 0.0327, 0.0133, and 0.0162 respectively. In Table 3.13, EN for

CARA, RA, and CR design are 0.0255, 0.0123, and 0.0137 respectively. Therefore, ML estimators slightly deviate from the normal distribution under CARA designs compared to RA and CR designs when the number of participating patients is 500. However, ML estimators approximately follow the normal distribution for CARA design when the number of participating patients is large. These results validate our theoretical findings.

3.7 Conclusion

In this chapter, we have established conditions for which ML estimators of parameters in GLM for adaptive designs are consistent and asymptotically follow multivariate normal distribution. One of these conditions is that the Fisher information matrix is a positive definite matrix at the true vector of parameters instead of the assumption: Fisher information and observed Fisher information matrices are positive definite matrices within a neighborhood of the vector of true parameters. We have demonstrated CARA and RA designs maximizing the well-being of participating patients in a clinical trial compared to CR design by simulation studies. Moreover, RA design generates more ethically desirable outcomes as well as efficient ML estimates than CARA design when there is no treatment by covariate interaction in the true model. If true model contains treatment by covariate interactions, then CARA design is more ethical than RA design. However, ML estimates for CARA design are less efficient compared to estimators under RA design. As we discuss in § 3.6, CARA design has three stages. Furthermore, CARA design maximizes the well-being of participating patients in the final stage after treatment by covariate interactions have been detected. Based on the above conclusions, we recommend that we can apply RA design until detecting treatment by covariate interactions. If treatment by covarite interaction is detected,

then we can apply CARA design in the final stage of a clinical trial.

Chapter 4

Investigating the Performance of Statistical Power versus Ethics between Response-Adaptive and Covariate-Adjusted Response-Adaptive Designs

4.1 Introduction

In Chapter 3, we established a set of conditions for asymptotic normality and consistency of estimators of regression parameters of logit model when we implement adaptive designs that satisfy the assumption 3.4.1. In this chapter, we establish theoretical foundation for the power computation based on Wald statistics when the model contains categorical variables with adaptive designs that satisfy assumption

3.4.1. Also, one of main objectives of a clinical trial is to test treatment effect efficiently.

This chapter is organized as follows. In § 4.2, general results for hypotheses testing are presented for adaptive designs and the model in (3.1) is considered as a true model. Similar justification of the Wald type hypotheses testing procedure is discussed in § 4.3 for adaptive designs; but the model in (4.10) is considered as the true model. In § 4.4, we investigate our theoretical results through simulation studies. At last, we provide the conclusion in § 4.5.

4.2 Hypothesis Testing: Full Model

In general, the hypotheses test for model in (3.1) are

$$H_0 : \mathbf{D}\boldsymbol{\theta}_0 = \mathbf{d}_0 \quad \text{vs} \quad H_A : \mathbf{D}\boldsymbol{\theta}_0 \neq \mathbf{d}_0 \tag{4.1}$$

where $\mathbf{D}$ is an $(d \times q)$ matrix of full row rank, $\mathbf{d}_0$ is a $(d \times 1)$ constant column vector, H_0 and H_A are null and alternative hypotheses respectively.

Theorem 4.2.1. We Assume 3.4.1 and 3.4.2 hold , with $m < \infty$, $q < \infty$. Define the Wald-type test statistic $T_W = [\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0]'[\mathbf{D}\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)^{-1}\mathbf{D}']^{-1}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0]$, where $\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)$ is the observed Fisher information matrix that is evaluated at $\hat{\boldsymbol{\theta}}_n$. Then, for fixed n,

(a) under H_0, T_W converges to the central chi-square distribution with d degrees of freedom, say, χ^2_d;

(b) under H_A, T_W is asymptotically distributed as non-central chi-square distribution with d degrees of freedom, and non-centrality parameter $\phi^{(a)}$, say, $\chi^2_d(\phi^{(a)})$;

where $\hat{\boldsymbol{\theta}}_n \in \mathbf{G}$ is the unrestricted MLE of $\boldsymbol{\theta_0}$, $n \geq N$, $\phi^{(a)} = n[\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0]'[\mathbf{D}\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{D}']^{-1}[\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0]$, $\mathbf{I}(\boldsymbol{\theta}_0)$ is the Fisher information matrix; $\mathbf{G}$, and N is defined in **Lemma 3.4.6**.

Proof. From **Lemma 3.4.6**, **Theorem 3.4.1**, and **Lemma 3.4.3**, we have the following results:

(i) $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}} \xrightarrow{a.s} \boldsymbol{\theta_0}$,

(ii) $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) \xrightarrow{d} N_q[\mathbf{0}, \mathbf{I}(\boldsymbol{\theta}_0)^{-1}]$,

(iii) $\frac{1}{n}\mathbf{F}_n(\boldsymbol{\theta_0}) \to \mathbf{I}(\boldsymbol{\theta}_0)$.

Using the result (i) and the continuous mapping theorem, the following result can be obtained

$$\frac{1}{n}\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{a.s.} \frac{1}{n}\mathbf{F}_n(\boldsymbol{\theta_0}). \tag{4.2}$$

Furthermore, applying (4.2) and the result (iii), we obtain

$$\frac{1}{n}\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{a.s.} \mathbf{I}(\boldsymbol{\theta}_0). \tag{4.3}$$

(a) Let $\hat{\boldsymbol{\theta}}_n \in \mathbf{G}$ and $n \geq N$,

we have that assumptions 3.4.1 and 3.4.2 are true, $m < \infty$, $q < \infty$; thus $\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)$ is a positive definite matrix [**Lemma 3.4.6**]. As a result, $\mathbf{D}\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)^{-1}\mathbf{D}'$ is a positive definite matrix because $\mathbf{D}$ is a matrix of full row rank. Then, $[\mathbf{D}\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)^{-1}\mathbf{D}']^{-1}$ exists.

Under H_0, T_W can be written as

$$\begin{aligned} T_W &= [\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{D}\boldsymbol{\theta}_0]'[\mathbf{D}\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)^{-1}\mathbf{D}']^{-1}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{D}\boldsymbol{\theta}_0] \\ &= [\sqrt{n}\mathbf{D}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)]'\{\mathbf{D}[(1/n)\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)]^{-1}\mathbf{D}'\}^{-1}[\sqrt{n}\mathbf{D}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)] \end{aligned}$$

$$\begin{aligned} \{\mathbf{D}[(1/n)\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)]^{-1}\mathbf{D}'\}^{-1} &\xrightarrow{a.s.} \{\mathbf{D}\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{D}'\}^{-1} \\ &= \mathbf{D}^*(\boldsymbol{\theta}_0)^{-1/2}\mathbf{D}^*(\boldsymbol{\theta}_0)^{-1/2}, \end{aligned} \tag{4.4}$$

where $\mathbf{D}^*(\boldsymbol{\theta}_0) = \mathbf{D}\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{D}'$. Since $\mathbf{D}$ is a $(d \times q)$ matrix of full row rank and $\mathbf{I}(\boldsymbol{\theta}_0)^{-1}$ is a positive definite matrix, we have that $\mathbf{D}^*(\boldsymbol{\theta}_0)$ is a positive definite matrix, [Seber and Lee (2003)]. Therefore, there exists a unique square root matrix $\mathbf{D}^*(\boldsymbol{\theta}_0)^{1/2}$ of $\mathbf{D}^*(\boldsymbol{\theta}_0)$.

Now, $\mathbf{D}$ is a $(d \times q)$ matrix and $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_\mathbf{0}) \xrightarrow{d} N_q[\mathbf{0}, \mathbf{I}(\boldsymbol{\theta}_0)^{-1}]$. Therefore, we have $\sqrt{n}\mathbf{D}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N_d[\mathbf{0}, \mathbf{D}^*(\boldsymbol{\theta}_0)]$ and $\mathbf{D}^*(\boldsymbol{\theta}_0)^{-1/2}\sqrt{n}\mathbf{D}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N_d[\mathbf{0}, \mathbf{I}_d^*]$ [Srivastava (2002)], where $\mathbf{I}_d^*$ is an identity matrix of dimension d. Therefore,

$$\begin{aligned} T_W^* &= \{\sqrt{n}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{D}\boldsymbol{\theta}_0]'\}[\mathbf{D}\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{D}']^{-1}\{\sqrt{n}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{D}\boldsymbol{\theta}_0]\} \\ &= \{\sqrt{n}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{D}\boldsymbol{\theta}_0]'\mathbf{D}^*(\boldsymbol{\theta}_0)^{-1/2}\}\{\sqrt{n}\mathbf{D}^*(\boldsymbol{\theta}_0)^{-1/2}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{D}\boldsymbol{\theta}_0]\} \\ &\xrightarrow{d} \chi^2_{(d)} \end{aligned} \tag{4.5}$$

where $\chi^2_{(d)}$ is the central chi-square distribution with d degrees of freedom. It follows from (4.4) and (4.5), that T_W asymptotically follows the central chi-square distribution with d degrees of freedom.

(b) Under H_A

$$\begin{aligned} T_W &= [\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0]'[\mathbf{D}\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)^{-1}\mathbf{D}']^{-1}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0] \\ &= \sqrt{n}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0]'[\mathbf{D}\{(1/n)\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)\}^{-1}\mathbf{D}']^{-1}\sqrt{n}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0]. \end{aligned}$$

We consider

$$\begin{aligned} \sqrt{n}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0] &= \sqrt{n}[\mathbf{D}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + (\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0)] \\ &\xrightarrow{d} N_d[\sqrt{n}(\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0), \mathbf{D}^*(\boldsymbol{\theta}_0)]. \end{aligned}$$

Thus,

$$\sqrt{n}\mathbf{D}^*(\boldsymbol{\theta}_0)^{-1/2}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0] \xrightarrow{d} N_d[\sqrt{n}\mathbf{D}^*(\boldsymbol{\theta}_0)^{-1/2}(\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0), \mathbf{I}_d^*].$$

Therefore, following Anderson (1966)

$$\begin{aligned} T_W^* &= \{\sqrt{n}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0]'\}[\mathbf{D}\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{D}']^{-1}\{\sqrt{n}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0]\} \\ &= \{\sqrt{n}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0]'\mathbf{D}^*(\boldsymbol{\theta}_0)^{-1/2}\}\{\sqrt{n}\mathbf{D}^*(\boldsymbol{\theta}_0)^{-1/2}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0]\} \\ &\xrightarrow{d} \chi^2_{(d)}(\phi^{(a)}), \end{aligned} \tag{4.6}$$

where $\phi^{(a)} = n[\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0]'[\mathbf{D}\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{D}']^{-1}[\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0]$. From (4.4) and (4.6), we have that T_W asymptotically follows the non-central chi-square distribution with d degrees of freedom, given n, and non-centrality parameter $\phi^{(a)}$.

□

4.2.1 Testing Interaction Effects

To test the interaction effects in (3.1), we test the following hypothesis:

$$H_{0I} : \boldsymbol{\delta}_0 = \mathbf{0}_{p\times 1} \quad \text{vs} \quad H_{AI} : \boldsymbol{\delta}_0 \neq \mathbf{0}_{p\times 1}. \tag{4.7}$$

Choose $\mathbf{D} = \begin{pmatrix} \mathbf{0}_{p\times(q-p)} & \mathbf{I}^*_{p\times p} \end{pmatrix}$, where $\mathbf{I}^*_{p\times p}$ is an identity matrix of dimension p and $d_0 = \mathbf{0}_{p\times 1}$, then the two hypotheses in (4.7) and (4.1) are equivalent. Define $\frac{1}{n}\mathbf{F}_n(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Delta}^{[n]}_{11}(\boldsymbol{\theta}) & \boldsymbol{\Delta}^{[n]}_{12}(\boldsymbol{\theta}) \\ \boldsymbol{\Delta}^{[n]}_{21}(\boldsymbol{\theta}) & \boldsymbol{\Delta}^{[n]}_{22}(\boldsymbol{\theta}) \end{pmatrix}$, where

$$\boldsymbol{\Delta}^{[n]}_{11}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}_i'\boldsymbol{\theta}) \begin{pmatrix} x_{iA} & x_{iA} & x_{iA}\mathbf{z}_i' \\ x_{iA} & 1 & \mathbf{z}_i' \\ x_{iA}\mathbf{z}_i & \mathbf{z}_i & \mathbf{z}_i\mathbf{z}_i' \end{pmatrix}, \boldsymbol{\Delta}^{[n]}_{21}(\boldsymbol{\theta}) = \boldsymbol{\Delta}^{[n]}_{12}(\boldsymbol{\theta})',$$

$$\boldsymbol{\Delta}^{[n]}_{12}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}_i'\boldsymbol{\theta}) \begin{pmatrix} x_{iA}\mathbf{z}_i' \\ x_{iA}\mathbf{z}_i' \\ x_{iA}\mathbf{z}_i\mathbf{z}_i' \end{pmatrix}, \text{ and } \boldsymbol{\Delta}^{[n]}_{22}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}_i'\boldsymbol{\theta}) x_{iA}\mathbf{z}_i\mathbf{z}_i'. \text{ From The-}$$

orem 4.2.1, when H_{0I} is true, we have that

$$\begin{aligned} T_{WI} &= n[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0]'\{\mathbf{D}[(1/n)\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)]^{-1}\mathbf{D}'\}^{-1}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0] \\ &= n\hat{\boldsymbol{\delta}}_n'[S_{\boldsymbol{\Delta}^{[n]}_{11}}(\hat{\boldsymbol{\theta}}_n)]\hat{\boldsymbol{\delta}}_n \sim \chi^2_p, \end{aligned} \tag{4.8}$$

where $S_{\boldsymbol{\Delta}^{[n]}_{11}}(\hat{\boldsymbol{\theta}}_n) = \boldsymbol{\Delta}^{[n]}_{22}(\hat{\boldsymbol{\theta}}_n) - \boldsymbol{\Delta}^{[n]}_{21}(\hat{\boldsymbol{\theta}}_n)[\boldsymbol{\Delta}^{[n]}_{11}(\hat{\boldsymbol{\theta}}_n)]^{-1}\boldsymbol{\Delta}^{[n]}_{12}(\hat{\boldsymbol{\theta}}_n)$.

4.2.2 Power of the Test for Interaction Effects

According to Theorem 4.2.1, when H_{AI} is true, T_{WI} is asymptotically distributed as non-central chi-square distribution with p degrees of freedom, given n, and the

non-centrality parameter ϕ, where

$$\begin{aligned} \phi &= n[\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0]'\{\mathbf{D}[(1/n)\mathbf{F}_n(\boldsymbol{\theta_0})]^{-1}\mathbf{D}'\}^{-1}[\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0] \\ &= n\boldsymbol{\delta}_0'[S_{\boldsymbol{\Delta}_{11}^{[n]}}(\boldsymbol{\theta}_0)]\boldsymbol{\delta}_0, \end{aligned} \tag{4.9}$$

with $S_{\boldsymbol{\Delta}_{11}^{[n]}}(\boldsymbol{\theta}_0) = \boldsymbol{\Delta}_{22}^{[n]}(\boldsymbol{\theta}_0) - \boldsymbol{\Delta}_{21}^{[n]}(\boldsymbol{\theta}_0)[\boldsymbol{\Delta}_{11}^{[n]}(\boldsymbol{\theta}_0)]^{-1}\boldsymbol{\Delta}_{12}^{[n]}(\boldsymbol{\theta}_0)$. When we conduct the hypothesis test for interaction effect in (4.7), we assume that the vector of parameters, $\boldsymbol{\theta}_{0I} = (\beta_{A0}, \gamma_{00}, \boldsymbol{\gamma_0}')'$, in model (3.1) are nuisance parameters. In practice the experimenter does not know the values of the nuisance parameters needed to compute power for a real data. Also to compute the value of the non-centrality parameter ϕ in (4.9), $\mathbf{F}_n(\boldsymbol{\theta_0})$ is replaced by $\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)$ [Demidenko (2007)].

4.3 Testing Hypotheses using the Wald-Type Statistic: Reduced Model

After conducting the test for interaction effects in (4.7), if we conclude that there is no evidence for interaction effects, one can drop the interaction terms from the full model in (3.1) when conducting statistical inference for the main treatment effect. Hereinafter, this sub model is called the reduced model. The reduced model is given by

$$\begin{aligned} \text{logit}[P(Y_i = 1|x_{iA}, \mathbf{z}_i)] &= x_{iA}\beta_{AR} + \gamma_{0R} + \mathbf{z}_i'\boldsymbol{\gamma}_R, \\ &= \mathbf{w}_{iR}'\boldsymbol{\theta}_R, \text{ for } i = 1, 2, \cdots, n \end{aligned} \tag{4.10}$$

where $\boldsymbol{\gamma}_R = (\gamma_{1R}, \gamma_{2R}, \cdots, \gamma_{pR})'$ are the main effects of covariates, β_{AR} is the effect of treatment A compared to treatment B, γ_{0R} is the intercept term, $\mathbf{w}_{iR} = (x_{iA}, 1, \mathbf{z}_i')'$, and $\boldsymbol{\theta}_R = (\beta_{AR}, \gamma_{0R}, \boldsymbol{\gamma}_R')'$. In fact, the model in (4.10) is a sub-model of the model in 3.1. In this case, we can define the design matrix $\mathbf{X}_R$ as

$$\mathbf{X}_R = \begin{pmatrix} \mathbf{w}_{1R}' \\ \mathbf{w}_{2R}' \\ \cdots \\ \cdots \\ \mathbf{w}_{nR}' \end{pmatrix}.$$

In this subsection, we assume that the model in (4.10) is the true model and the true vector of parameters is $\boldsymbol{\theta}_{0R}$, where $\boldsymbol{\theta}_{0R} = (\beta_{A0R}, \gamma_{00R}, \boldsymbol{\gamma}_{0R}')'$. We will investigate the influence of adaptation on the statistical power in testing for main effect.

Assumption 4.3.1. *$\boldsymbol{I}(\boldsymbol{\theta_{0R}})$ is a positive definite matrix, where $\boldsymbol{\theta_{0R}}$ is a $(q-p)$ dimensional true vector of parameters, and $\boldsymbol{I}(\boldsymbol{\theta_{0R}})$ is a Fisher information matrix.*

4.3.1 Testing for Main Effect of Treatment

The hypothesis of interest in testing for the main effect of treatment for the model in (4.10) becomes

$$H_{0TR} : \beta_{A0R} = 0 \qquad H_{ATR} : \beta_{A0R} \neq 0. \tag{4.11}$$

Now, let $\mathbf{D}_R = \begin{pmatrix} 1 & \mathbf{0}_{1\times(q-p-1)} \end{pmatrix}$, $d_{0R} = 0$, and define $\frac{1}{n}\mathbf{F}_n(\boldsymbol{\theta}_R) = \begin{pmatrix} \Delta_{11t}^{[n]}(\boldsymbol{\theta}_R) & \boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_R) \\ \boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_R) & \boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_R) \end{pmatrix}$, where

$\Delta_{11t}^{[n]}(\boldsymbol{\theta}_R) = \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}'_{iR}\boldsymbol{\theta}_R)x_{iA}$, $\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_R) = \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}'_{iR}\boldsymbol{\theta}_R)\begin{pmatrix} 1 & \mathbf{z}'_i \\ \mathbf{z}_i & \mathbf{z}_i\mathbf{z}'_i \end{pmatrix}$, $\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_R) = \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}'_{iR}\boldsymbol{\theta}_R)\begin{pmatrix} x_{iA} & x_{iA}\mathbf{z}'_i \end{pmatrix}$, $\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_R) = \boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_R)'$. Under H_{0TR}, we have that the test statistics given by

$$\begin{aligned} T_{WR} &= n[\mathbf{D}_R\hat{\boldsymbol{\theta}}_{nR}]'\{\mathbf{D}_R[(1/n)\mathbf{F}_n(\hat{\boldsymbol{\theta}}_{nR})]^{-1}\mathbf{D}'_R\}^{-1}[\mathbf{D}_R\hat{\boldsymbol{\theta}}_{nR}] \\ &= n[\hat{\beta}_{AnR}]^2[S_{\boldsymbol{\Delta}_{22t}^{[n]}}(\hat{\boldsymbol{\theta}}_{nR})] \xrightarrow{d} \chi_1^2, \end{aligned} \tag{4.12}$$

follows the chi-squared distribution with 1 degree of freedom, where $S_{\boldsymbol{\Delta}_{22t}^{[n]}}(\hat{\boldsymbol{\theta}}_{nR}) = \Delta_{11t}^{[n]}(\hat{\boldsymbol{\theta}}_{nR}) - \boldsymbol{\Delta}_{12t}^{[n]}(\hat{\boldsymbol{\theta}}_{nR})[\boldsymbol{\Delta}_{22t}^{[n]}(\hat{\boldsymbol{\theta}}_{nR})]^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\hat{\boldsymbol{\theta}}_{nR})$.

4.3.2 Statistical Power Computation for Hypothesis Testing of Main Effect

It is clear that, under H_{ATR}, T_{WR} is asymptotically distributed as non-central chi-square with 1 degree of freedom and non-centrality parameter ϕ, given by

$$\begin{aligned} \phi &= n[\mathbf{D}_R\boldsymbol{\theta}_{0R} - \mathbf{d}_{0R}]'\{\mathbf{D}_R[(1/n)\mathbf{F}_n(\boldsymbol{\theta}_{\mathbf{0R}})]^{-1}\mathbf{D}'_R\}^{-1}[\mathbf{D}_R\boldsymbol{\theta}_{0R} - \mathbf{d}_{0R}] \\ &= n[\beta_{A0R}]^2[S_{\boldsymbol{\Delta}_{22t}^{[n]}}(\boldsymbol{\theta}_{0R})], \end{aligned} \tag{4.13}$$

where $S_{\boldsymbol{\Delta}_{22t}^{[n]}}(\boldsymbol{\theta}_{0R}) = \Delta_{11t}^{[n]}(\boldsymbol{\theta}_{0R}) - \boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R})[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R})]^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R})$. In this section, the true treatment effect, β_{A0R}, is the main parameter of interest. We refer to the other parameters, $\boldsymbol{\theta}_{0NR} = (\gamma_{00R}, \boldsymbol{\gamma}_{\mathbf{0R}}')'$, in the model (4.10) as nuisance parameters. The nuisance parameters, $\boldsymbol{\theta}_{0NR}$ are however required for power computation in testing the hypothesis in (4.11). Demidenko (2007) notes that to compute the value of the non-centrality parameter ϕ in (4.13), $\mathbf{F}_n(\boldsymbol{\theta}_{\mathbf{0R}})$ has to be replaced by $\mathbf{F}_n(\hat{\boldsymbol{\theta}}_{nR})$.

As defined earlier, $N_{Ah}(n)$ is the number of patients assigned to treatment A for a given n_h number of patients in stratum h, $h = 1, 2, \cdots, m$. Define the sample proportion of patients (say $\hat{p}_{Ah}$) assigned to treatment A for a given n_h number of patients in the stratum as $\hat{p}_{Ah} = \dfrac{N_{Ah}(n)}{n_h}$, for $h = 1, 2, \cdots, m$. Furthermore, let $\hat{\mathbf{p}}_A = (\hat{p}_{A1}, \hat{p}_{A2}, \cdots, \hat{p}_{Am})'$ be the vector of sample proportions of patients assigned to treatment A for a given n and $\boldsymbol{\pi_A} = (\pi_A(\mathbf{v}_1), \pi_A(\mathbf{v}_2), \cdots, \pi_A(\mathbf{v}_m))'$ be the vector of target proportions of patients assigned to treatment A over patients' strata. Then, for a given n, we investigate the non-centrality parameter ϕ

$$\frac{\phi}{n} \;=\; [\beta_{A0R}]^2[S_{\boldsymbol{\Delta}^{[n]}_{22t}}(\boldsymbol{\theta}_{0R})], \tag{4.14}$$

where $S_{\boldsymbol{\Delta}^{[n]}_{22t}}(\boldsymbol{\theta}_{0R}) = \Delta^{[n]}_{11t}(\boldsymbol{\theta}_{0R}) - \boldsymbol{\Delta}^{[n]}_{12t}(\boldsymbol{\theta}_{0R})[\boldsymbol{\Delta}^{[n]}_{22t}(\boldsymbol{\theta}_{0R})]^{-1}\boldsymbol{\Delta}^{[n]}_{21t}(\boldsymbol{\theta}_{0R})$,

$$\Delta^{[n]}_{11t}(\boldsymbol{\theta}_{0R}) = \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}'_{iR}\boldsymbol{\theta}_{0R})x_{iA}, \;\; \boldsymbol{\Delta}^{[n]}_{22t}(\boldsymbol{\theta}_{0R}) = \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}'_{iR}\boldsymbol{\theta}_{0R}) \begin{pmatrix} 1 & \mathbf{z}'_i \\ \mathbf{z}_i & \mathbf{z}_i\mathbf{z}'_i \end{pmatrix},$$

$$\boldsymbol{\Delta}^{[n]}_{12t}(\boldsymbol{\theta}_{0R}) = \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}'_{iR}\boldsymbol{\theta}_{0R}) \begin{pmatrix} x_{iA} & x_{iA}\mathbf{z}'_i \end{pmatrix}, \;\; \boldsymbol{\Delta}^{[n]}_{21t}(\boldsymbol{\theta}_{0R}) = \boldsymbol{\Delta}^{[n]}_{12t}(\boldsymbol{\theta}_{0R})'.$$

Theorem 4.3.1. *Consider the non-centrality parameter ϕ defined in (4.14) and follow the notation that was introduced in this section. Then,*

1. *ϕ is a function of $\hat{\boldsymbol{p}}_A$, where $\hat{\boldsymbol{p}}_A = (\hat{p}_{A1}, \hat{p}_{A2}, \cdots, \hat{p}_{Am})'$ is the vector of sample proportions of patients assigned to Treatment A for given n. Let this function be $\phi(\hat{\boldsymbol{p}}_A)$.*

2. *$\phi(\hat{\boldsymbol{p}}_A)$ can be expressed the following quantities: the target allocation proportion, the bias of the randomization procedure from the target, and the variability induced by the randomization process.*

Proof. From (4.14), $\dfrac{\phi}{n} = [\beta_{A0R}]^2[S_{\boldsymbol{\Delta}^{[n]}_{22t}}(\boldsymbol{\theta}_{0R})]$.

1. First, we express the components of $\frac{\phi}{n}$, namely, $\Delta_{11t}^{[n]}(\boldsymbol{\theta}_{0R})$, $\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R})$, $\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R})$, and $\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R})$ as function of $\hat{p}_{Ah}$ for $h = 1, 2, \cdots, m$.

$$
\begin{aligned}
\Delta_{11t}^{[n]}(\boldsymbol{\theta}_{0R}) &= \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}_{iR}'\boldsymbol{\theta}_{0R}) x_{iA} \\
&= \sum_{h=1}^{m} \lambda_{AhR}(\boldsymbol{\theta}_{0R}) \frac{n_h}{n} \frac{N_{Ah}(n)}{n_h} \\
&= \sum_{h=1}^{m} \lambda_{AhR}(\boldsymbol{\theta}_{0R}) \frac{n_h}{n} \hat{p}_{Ah} \\
\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R}) &= \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}_{iR}'\boldsymbol{\theta}_{0R}) \begin{pmatrix} x_{iA} & x_{iA}\mathbf{z}_i' \end{pmatrix} \\
&= \sum_{h=1}^{m} \lambda_{AhR}(\boldsymbol{\theta}_{0R}) \frac{n_h}{n} \frac{N_{Ah}(n)}{n_h} \begin{pmatrix} 1 & \mathbf{v}_h' \end{pmatrix} \\
&= \sum_{h=1}^{m} \lambda_{AhR}(\boldsymbol{\theta}_{0R}) \frac{n_h}{n} \hat{p}_{Ah} \begin{pmatrix} 1 & \mathbf{v}_h' \end{pmatrix} \\
\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R}) &= \sum_{h=1}^{m} \lambda_{AhR}(\boldsymbol{\theta}_{0R}) \frac{n_h}{n} \hat{p}_{Ah} \begin{pmatrix} 1 & \mathbf{v}_h' \end{pmatrix}' \\
\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}) &= \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}_{iR}'\boldsymbol{\theta}_{0R}) \begin{pmatrix} 1 & \mathbf{z}_i' \\ \mathbf{z}_i & \mathbf{z}_i\mathbf{z}_i' \end{pmatrix} \\
&= \sum_{h=1}^{m} \Upsilon(\boldsymbol{\theta}_{0R}, \hat{p}_{Ah}) \begin{pmatrix} 1 & \mathbf{v}_h' \\ \mathbf{v}_h & \mathbf{v}_h\mathbf{v}_h' \end{pmatrix}
\end{aligned}
$$

where $\Upsilon(\boldsymbol{\theta}_{0R}, \hat{p}_{Ah}) = \frac{n_h}{n}\left\{\lambda_{AhR}(\boldsymbol{\theta}_{0R})\hat{p}_{Ah} + \lambda_{BhR}(\boldsymbol{\theta}_{0R})\left[1 - \hat{p}_{Ah}\right]\right\}$,
$\lambda_{AhR}(\boldsymbol{\theta}_R) = \exp(-\beta_{AR} - \gamma_{0R} - \mathbf{v}_h'\boldsymbol{\gamma}_R)[1 + \exp(-\beta_{AR} - \gamma_{0R} - \mathbf{v}_h'\boldsymbol{\gamma}_R)]^{-2}$, and
$\lambda_{BhR}(\boldsymbol{\theta}_R) = \exp(-\gamma_{0R} - \mathbf{v}_h'\boldsymbol{\gamma}_R)[1 + \exp(-\gamma_{0R} - \mathbf{v}_h'\boldsymbol{\gamma}_R)]^{-2}$. Therefore, the non-centrality parameter ϕ is a function of $\hat{\mathbf{p}}_A$.

2. We apply the multivariate version of Taylor's expansion to $\phi(\hat{\mathbf{p}}_A)$ in a neighborhood centered around $\boldsymbol{\pi_A}$. As a result,

$$\begin{aligned}
\frac{1}{n}\phi(\hat{\mathbf{p}}_A) &= \frac{1}{n}\phi(\boldsymbol{\pi_A}) + \frac{1}{n}\phi^{(1)}(\boldsymbol{\pi_A})[\hat{\mathbf{p}}_A - \boldsymbol{\pi_A}] + \frac{1}{2}[\hat{\mathbf{p}}_A - \boldsymbol{\pi_A}]'\frac{1}{n}\phi^{(2)}(\boldsymbol{\pi_A})[\hat{\mathbf{p}}_A - \boldsymbol{\pi_A}] \\
&+ \mathbf{o}(\parallel \hat{\mathbf{p}}_A - \boldsymbol{\pi_A} \parallel^m), \text{ where} \qquad (4.15) \\
\frac{1}{n}\phi^{(1)}(\boldsymbol{\pi_A}) &= \frac{1}{n}\left(\frac{\partial\phi(\hat{\mathbf{p}}_A)}{\partial\hat{\mathbf{p}}_A} \right)_{\hat{\mathbf{p}}_A=\boldsymbol{\pi_A}} \\
&= \frac{1}{n}\left(\frac{\partial\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A1}} \quad \frac{\partial\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A2}} \quad \cdots \quad \frac{\partial\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}} \quad \cdots \quad \frac{\partial\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Am}} \right)'_{\hat{\mathbf{p}}_A=\boldsymbol{\pi_A}}
\end{aligned}$$

where, for $h = 1, 2, \cdots, m$

$$\begin{aligned}
\frac{1}{n}\frac{\partial\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}} &= [\beta_{A0R}]^2\left[\frac{\partial S_{\boldsymbol{\Delta}^{[n]}_{22t}}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}}\right] \\
&= [\beta_{A0R}]^2\left[\frac{\partial\Delta^{[n]}_{11t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}} - \frac{\partial\boldsymbol{\Delta}^{[n]}_{12t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)[\boldsymbol{\Delta}^{[n]}_{22t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1}\boldsymbol{\Delta}^{[n]}_{21t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}}\right] \\
&= [\beta_{A0R}]^2\left[\frac{\partial\Delta^{[n]}_{11t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}} - \frac{\partial\boldsymbol{\Delta}^{[n]}_{12t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}}[\boldsymbol{\Delta}^{[n]}_{22t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1}\boldsymbol{\Delta}^{[n]}_{21t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)\right] \\
&- [\beta_{A0R}]^2\left[\boldsymbol{\Delta}^{[n]}_{12t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)\frac{\partial[\boldsymbol{\Delta}^{[n]}_{22t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1}}{\partial\hat{p}_{Ah}}\boldsymbol{\Delta}^{[n]}_{21t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)\right] \\
&- [\beta_{A0R}]^2\left[\boldsymbol{\Delta}^{[n]}_{12t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)[\boldsymbol{\Delta}^{[n]}_{22t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1}\frac{\partial\boldsymbol{\Delta}^{[n]}_{21t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}}\right]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\Delta^{[n]}_{11t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}} &= \lambda_{AhR}(\boldsymbol{\theta}_{0R})\frac{n_h}{n} \\
\frac{\partial\boldsymbol{\Delta}^{[n]}_{12t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}} &= \lambda_{AhR}(\boldsymbol{\theta}_{0R})\frac{n_h}{n}\left(1 \quad \mathbf{v}'_h \right) \\
\frac{\partial\boldsymbol{\Delta}^{[n]}_{21t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}} &= \lambda_{AhR}(\boldsymbol{\theta}_{0R})\frac{n_h}{n}\left(1 \quad \mathbf{v}'_h \right)'
\end{aligned}$$

$$\begin{aligned}\frac{\partial[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}}{\partial\hat{p}_{Ah}} &= -[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\left[\frac{\partial[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]}{\partial\hat{p}_{Ah}}\right][\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1} \\ &= -[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{Dh}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1},\end{aligned}$$

where $\mathbf{M}_{Dh} = \left[\frac{n_h}{n}\{\lambda_{AhR}(\boldsymbol{\theta}_{0R}) - \lambda_{BhR}(\boldsymbol{\theta}_{0R})\}\begin{pmatrix} 1 & \mathbf{v}_h' \\ \mathbf{v}_h & \mathbf{v}_h\mathbf{v}_h' \end{pmatrix}\right]$.

To evaluate the term $\frac{1}{n}\phi^{(2)}(\boldsymbol{\pi}_{\boldsymbol{A}})$ in (4.20), we require the following second derivatives:

(i) $\frac{\partial^2\Delta_{11t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}\partial\hat{p}_{Ah^*}} = 0$ for $h^* = h$ or $h^* \neq h$.

(ii) $\frac{\partial^2\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}\partial\hat{p}_{Ah^*}} = \mathbf{0}_{1\times(p+1)}$ for $h^* = h$ or $h^* \neq h$.

(iii) $\frac{\partial^2\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}\partial\hat{p}_{Ah^*}} = \mathbf{0}_{(p+1)\times1}$ for $h^* = h$ or $h^* \neq h$.

(iv) $\frac{\partial^2[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}}{\partial\hat{p}_{Ah}\partial\hat{p}_{Ah}} = 2[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{Dh}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{Dh}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}$

(v) $\frac{\partial^2[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}}{\partial\hat{p}_{Ah}\partial\hat{p}_{Ah^*}} = [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{Dh^*}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{Dh}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}$
$+ [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{Dh}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{Dh^*}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}$

$$\begin{aligned}
\frac{1}{n}\phi^{(2)}(\boldsymbol{\pi_A}) &= \frac{1}{n}\left(\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{\mathbf{p}}_A\partial\hat{\mathbf{p}}_A'} \right)_{\hat{\mathbf{p}}_A=\boldsymbol{\pi_A}} \\
&= \frac{1}{n}\begin{pmatrix}
\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A1}\partial\hat{p}_{A1}} & \frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A1}\partial\hat{p}_{A2}} & \cdots & \frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A1}\partial\hat{p}_{Ah^*}} & \cdots & \frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A1}\partial\hat{p}_{Am}} \\
\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A2}\partial\hat{p}_{A1}} & \frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A2}\partial\hat{p}_{A2}} & \cdots & \frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A2}\partial\hat{p}_{Ah^*}} & \cdots & \frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A2}\partial\hat{p}_{Am}} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}\partial\hat{p}_{A1}} & \frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}\partial\hat{p}_{A2}} & \cdots & \frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}\partial\hat{p}_{Ah^*}} & \cdots & \frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}\partial\hat{p}_{Am}} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Am}\partial\hat{p}_{A1}} & \frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Am}\partial\hat{p}_{A2}} & \cdots & \frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Am}\partial\hat{p}_{Ah^*}} & \cdots & \frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Am}\partial\hat{p}_{Am}}
\end{pmatrix}_{\hat{\mathbf{p}}_A=\boldsymbol{\pi_A}} \qquad (4.16)
\end{aligned}$$

Hence the theorem holds.

□

4.3.3 Concaveness of the Non-centrality Parameter

We begin by noting that the theoretical non-centrality parameter ϕ. Also, we consider only the one covariate Z_1. This parameter ϕ is a function of the sample proportion of patients assigned to treatment A. we assume the true model

$$\text{logit}[P(Y_i = 1|x_{iA})] = \mathbf{w}_i'\boldsymbol{\theta_{0R}} \text{ for } i = 1, 2, \cdots, n \qquad (4.17)$$

where $\boldsymbol{\theta}_{0R} = (\beta_{A0R}, \gamma_{00R}, \gamma_{10R})'$, $\mathbf{w}_i = (x_{iA}, 1, z_{i1})'$. In this section, we show that the non-centrality parameter of the distribution of the test statistic is concave, the hypothesis test for testing treatment effect

$$H_{0T} : \beta_{A0R} = 0 \qquad H_A : \beta_{A0R} \neq 0, \qquad (4.18)$$

where β_{A0R} is the effect of treatment A (the new treatment) compared to the effect of treatment B (an existing treatment).

As defined earlier, $N_{Ah}(n)$ is the number of patients assigned to treatment A for a given n_h number of patients in stratum h, $h = 1, 2$. Here, the sample proportion of patients (say $\hat{p}_{Ah}$) assigned to treatment A for a given n_h number of patients in the stratum is $\hat{p}_{Ah} = \dfrac{N_{Ah}(n)}{n_h}$, for $h = 1, 2$. Also, let $\hat{\mathbf{p}}_A = (\hat{p}_{A1}, \hat{p}_{A2})'$ be the vector of sample proportions of patients assigned to treatment A for given n and $\boldsymbol{\pi_A} = (\pi_A(\mathrm{v}_1), \pi_A(\mathrm{v}_2))'$, the vector of target proportions of patients assigned to treatment A over patients' strata. From (4.14), the non-centrality parameter ϕ for a given n is

$$\frac{1}{n}\phi(\hat{\mathbf{p}}_A) \;=\; [\beta_{A0R}]^2[S_{\boldsymbol{\Delta}^{[n]}_{22t}}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)], \tag{4.19}$$

where $S_{\boldsymbol{\Delta}^{[n]}_{22t}}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) = \Delta^{[n]}_{11t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) - \boldsymbol{\Delta}^{[n]}_{12t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)[\boldsymbol{\Delta}^{[n]}_{22t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1}\boldsymbol{\Delta}^{[n]}_{21t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)$.

Theorem 4.3.2. *Consider the non-centrality parameter defined in (4.19). If we use the notations in § 4.1, then $\phi(\hat{\boldsymbol{p}}_A)$ is a concave function.*

Proof. Using the proof of Theorem 4.3.1, we have

$$\begin{aligned}
\Delta^{[n]}_{11t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) &= \sum_{h=1}^{2} \lambda_{Ah}(\boldsymbol{\theta}_{0R})\frac{n_h}{n}\hat{p}_{Ah}, \\
\boldsymbol{\Delta}^{[n]}_{12t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) &= \sum_{h=1}^{2} \lambda_{Ah}(\boldsymbol{\theta}_{0R})\frac{n_h}{n}\hat{p}_{Ah}\begin{pmatrix} 1 & \mathrm{v}_h \end{pmatrix}, \\
\boldsymbol{\Delta}^{[n]}_{21t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) &= \sum_{h=1}^{2} \lambda_{Ah}(\boldsymbol{\theta}_{0R})\frac{n_h}{n}\hat{p}_{Ah}\begin{pmatrix} 1 & \mathrm{v}_h \end{pmatrix}', \\
\boldsymbol{\Delta}^{[n]}_{22t}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) &= \sum_{h=1}^{2} \Upsilon_h(\boldsymbol{\theta}_{0R})\begin{pmatrix} 1 & \mathrm{v}_h \\ \mathrm{v}_h & \mathrm{v}_h \end{pmatrix},
\end{aligned}$$

where $\Upsilon_h(\boldsymbol{\theta}_{0R}) = \dfrac{n_h}{n}\left\{\lambda_{Ah}(\boldsymbol{\theta}_{0R})\hat{p}_{Ah} + \lambda_{Bh}(\boldsymbol{\theta}_{0R})\left[1-\hat{p}_{Ah}\right]\right\}$,
$\lambda_{Ah}(\boldsymbol{\theta}_{0R}) = \exp(-\beta_{A0R} - \gamma_{00R} - \mathrm{v}_h\gamma_{10R})[1+\exp(-\beta_{A0R} - \gamma_{00R} - \mathrm{v}_h\gamma_{10R})]^{-2}$,
$\lambda_{Bh}(\boldsymbol{\theta}_{0R}) = \exp(-\gamma_{00R} - \mathrm{v}_h\gamma_{10R})[1+\exp(-\gamma_{00R} - \mathrm{v}_h\gamma_{10R})]^{-2}$ for $h = 1, 2$; $\mathrm{v}_1 = 1$ and $\mathrm{v}_2 = 0$.

Applying the multivariate version of Taylor's expansion to $\phi(\hat{\mathbf{p}}_A)$ in a neighborhood centered around $\boldsymbol{\pi_A}$ we obtain,

$$\frac{1}{n}\phi(\hat{\mathbf{p}}_A) = \frac{1}{n}\phi(\boldsymbol{\pi_A}) + \frac{1}{n}\phi^{(1)}(\boldsymbol{\pi_A})[\hat{\mathbf{p}}_A - \boldsymbol{\pi_A}] - \frac{1}{2}[\hat{\mathbf{p}}_A - \boldsymbol{\pi_A}]'[-\frac{1}{n}\phi^{(2)}(\boldsymbol{\pi_A})][\hat{\mathbf{p}}_A - \boldsymbol{\pi_A}]$$

Now,

$$\begin{aligned}\frac{1}{n}\phi^{(1)}(\boldsymbol{\pi_A}) &= \frac{1}{n}\left(\frac{\partial\phi(\hat{\mathbf{p}}_A)}{\partial\hat{\mathbf{p}}_A}\right)_{\hat{\mathbf{p}}_A=\boldsymbol{\pi_A}} \\ &= \frac{1}{n}\left(\frac{\partial\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A1}} \quad \frac{\partial\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A2}}\right)'_{\hat{\mathbf{p}}_A=\boldsymbol{\pi_A}}\end{aligned}$$

with

$$\begin{aligned}\frac{1}{n}\frac{\partial\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}} &= [\beta_{A0R}]^2\left[\frac{\partial S_{\boldsymbol{\Delta}_{22t}^{[n]}}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}}\right] \\ &= [\beta_{A0R}]^2\left[\frac{\partial\Delta_{11t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}} - \frac{\partial\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}}\right] \\ &= [\beta_{A0R}]^2\left[\frac{\partial\Delta_{11t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}} - \frac{\partial\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)\right] \\ &- [\beta_{A0R}]^2\left[\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)\frac{\partial[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}}{\partial\hat{p}_{Ah}}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)\right] \\ &- [\beta_{A0R}]^2\left[\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\frac{\partial\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)}{\partial\hat{p}_{Ah}}\right]\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial \Delta_{11t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)}{\partial \hat{p}_{Ah}} &= \lambda_{Ah}(\boldsymbol{\theta}_{0R})\frac{n_h}{n} \\
\frac{\partial \boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)}{\partial \hat{p}_{Ah}} &= \lambda_{Ah}(\boldsymbol{\theta}_{0R})\frac{n_h}{n}\begin{pmatrix} 1 & \mathrm{v}_h \end{pmatrix} \\
\frac{\partial \boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)}{\partial \hat{p}_{Ah}} &= \lambda_{Ah}(\boldsymbol{\theta}_{0R})\frac{n_h}{n}\begin{pmatrix} 1 & \mathrm{v}_h \end{pmatrix}'
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1}}{\partial \hat{p}_{Ah}} &= -[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1}\left[\frac{\partial [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]}{\partial \hat{p}_{Ah}}\right][\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \\
&= -[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{Dh}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1},
\end{aligned}$$

$$\mathbf{M}_{Dh} = \left[\frac{n_h}{n}\left\{\lambda_{Ah}(\boldsymbol{\theta}_{0R}) - \lambda_{Bh}(\boldsymbol{\theta}_{0R})\right\}\begin{pmatrix} 1 & \mathrm{v}_h \\ \mathrm{v}_h & \mathrm{v}_h \end{pmatrix}\right].$$

We note that here, $h = 1, 2$, and $\mathrm{v}_1 = 1$, $\mathrm{v}_2 = 0$. Also,

$$\begin{aligned}
\frac{1}{n}\phi^{(2)}(\boldsymbol{\pi_A}) &= \frac{1}{n}\left(\frac{\partial^2 \phi(\hat{\mathbf{p}}_A)}{\partial \hat{\mathbf{p}}_A \partial \hat{\mathbf{p}}_A'}\right)_{\hat{\mathbf{p}}_A = \boldsymbol{\pi_A}} \\
&= \frac{1}{n}\begin{pmatrix} \dfrac{\partial^2 \phi(\hat{\mathbf{p}}_A)}{\partial \hat{p}_{A1}\partial \hat{p}_{A1}} & \dfrac{\partial^2 \phi(\hat{\mathbf{p}}_A)}{\partial \hat{p}_{A1}\partial \hat{p}_{A2}} \\ \dfrac{\partial^2 \phi(\hat{\mathbf{p}}_A)}{\partial \hat{p}_{A2}\partial \hat{p}_{A1}} & \dfrac{\partial^2 \phi(\hat{\mathbf{p}}_A)}{\partial \hat{p}_{A2}\partial \hat{p}_{A2}} \end{pmatrix}_{\hat{\mathbf{p}}_A = \boldsymbol{\pi_A}} . \qquad (4.20)
\end{aligned}$$

To evaluate the term $\frac{1}{n}\phi^{(2)}(\boldsymbol{\pi_A})$ in 4.20, we require the following second derivatives:

(i) $\dfrac{\partial^2 \Delta_{11t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)}{\partial \hat{p}_{Ah}\partial \hat{p}_{Ah^*}} = 0$ for $h^* = h$ or $h^* \neq h$,

(ii) $\dfrac{\partial^2 \boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)}{\partial \hat{p}_{Ah}\partial \hat{p}_{Ah^*}} = \mathbf{0}_{1\times 2}$ for $h^* = h$ or $h^* \neq h$,

(iii) $\dfrac{\partial^2 \boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)}{\partial \hat{p}_{Ah} \partial \hat{p}_{Ah^*}} = \mathbf{0}_{2\times 1}$ for $h^* = h$ or $h^* \neq h$, and

(iv) $\dfrac{\partial^2 [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1}}{\partial \hat{p}_{Ah} \partial \hat{p}_{Ah^*}} = 2[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \mathbf{M}_{Dh^*} [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \mathbf{M}_{Dh} [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1}$

for $h^* = h$ or $h^* \neq h$.

The second derivatives in (4.20) are then, given by

$$
\begin{aligned}
&\frac{1}{n} \frac{\partial^2 \phi(\hat{\mathbf{p}}_A)}{\partial \hat{p}_{Ah} \partial \hat{p}_{Ah^*}} \\
&= 2[\beta_{A0R}]^2 \left[\lambda_{AhR}(\boldsymbol{\theta}_{0R}) \frac{n_h}{n} \begin{pmatrix} 1 & \mathrm{v}_h \end{pmatrix} [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \mathbf{M}_{Dh^*} [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) \right] \\
&- 2[\beta_{A0R}]^2 \left[\lambda_{AhR}(\boldsymbol{\theta}_{0R}) \frac{n_h}{n} \begin{pmatrix} 1 & \mathrm{v}_h \end{pmatrix} [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \lambda_{Ah^*R}(\boldsymbol{\theta}_{0R}) \frac{n_{h^*}}{n} \begin{pmatrix} 1 & \mathrm{v}_{h^*} \end{pmatrix}' \right] \\
&+ 2[\beta_{A0R}]^2 \left[\lambda_{Ah^*R}(\boldsymbol{\theta}_{0R}) \frac{n_{h^*}}{n} \begin{pmatrix} 1 & \mathrm{v}_{h^*} \end{pmatrix} [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \mathbf{M}_{Dh} [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) \right] \\
&- 2[\beta_{A0R}]^2 \left[\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \mathbf{M}_{Dh^*} [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \mathbf{M}_{Dh} [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \right] \\
&\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A),
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{n} \frac{\partial^2 \phi(\hat{\mathbf{p}}_A)}{\partial \hat{p}_{Ah} \partial \hat{p}_{Ah}} \\
&= 2[\beta_{A0R}]^2 \left[\lambda_{Ah}(\boldsymbol{\theta}_{0R}) \frac{n_h}{n} \begin{pmatrix} 1 & \mathrm{v}_h \end{pmatrix} [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \mathbf{M}_{Dh} [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) \right] \\
&- 2[\beta_{A0R}]^2 \left[\lambda_{Ah}(\boldsymbol{\theta}_{0R}) \frac{n_h}{n} \begin{pmatrix} 1 & \mathrm{v}_h \end{pmatrix} [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \lambda_{Ah}(\boldsymbol{\theta}_{0R}) \frac{n_h}{n} \begin{pmatrix} 1 & \mathrm{v}_h \end{pmatrix}' \right] \\
&+ 2[\beta_{A0R}]^2 \left[\lambda_{Ah}(\boldsymbol{\theta}_{0R}) \frac{n_h}{n} \begin{pmatrix} 1 & \mathrm{v}_h \end{pmatrix} [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \mathbf{M}_{Dh} [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) \right] \\
&- 2[\beta_{A0R}]^2 \left[\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \mathbf{M}_{Dh} [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \mathbf{M}_{Dh} [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \right] \\
&\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A).
\end{aligned}
$$

Using the fact that $\mathrm{v}_1 = 1$, $\mathrm{v}_2 = 0$, we obtain the following results

$$\begin{aligned}
\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) &= \sum_{h=1}^{2} \Upsilon_h(\boldsymbol{\theta}_{0R}) \begin{pmatrix} 1 & \mathrm{v}_h \\ \mathrm{v}_h & \mathrm{v}_h \end{pmatrix} \\
&= \Upsilon_1(\boldsymbol{\theta}_{0R}) \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \Upsilon_2(\boldsymbol{\theta}_{0R}) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} \Upsilon_1(\boldsymbol{\theta}_{0R}) + \Upsilon_2(\boldsymbol{\theta}_{0R}) & \Upsilon_1(\boldsymbol{\theta}_{0R}) \\ \Upsilon_1(\boldsymbol{\theta}_{0R}) & \Upsilon_1(\boldsymbol{\theta}_{0R}) \end{pmatrix},
\end{aligned}$$

with

$$\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)^{-1} = \frac{1}{\Upsilon_1(\boldsymbol{\theta}_{0R})\Upsilon_2(\boldsymbol{\theta}_{0R})} \begin{pmatrix} \Upsilon_1(\boldsymbol{\theta}_{0R}) & -\Upsilon_1(\boldsymbol{\theta}_{0R}) \\ -\Upsilon_1(\boldsymbol{\theta}_{0R}) & \Upsilon_1(\boldsymbol{\theta}_{0R}) + \Upsilon_2(\boldsymbol{\theta}_{0R}) \end{pmatrix}.$$

Recall that

$$\begin{aligned}
\mathbf{M}_{Dh} &= \left[\frac{n_h}{n} \left\{ \lambda_{Ah}(\boldsymbol{\theta}_{0R}) - \lambda_{Bh}(\boldsymbol{\theta}_{0R}) \right\} \begin{pmatrix} 1 & \mathrm{v}_h \\ \mathrm{v}_h & \mathrm{v}_h \end{pmatrix} \right] \\
&= \mu_h(\boldsymbol{\theta}_{0R}) \begin{pmatrix} 1 & \mathrm{v}_h \\ \mathrm{v}_h & \mathrm{v}_h \end{pmatrix},
\end{aligned}$$

where $\mu_h(\boldsymbol{\theta}_{0R}) = \dfrac{n_h}{n} \left\{ \lambda_{Ah}(\boldsymbol{\theta}_{0R}) - \lambda_{Bh}(\boldsymbol{\theta}_{0R}) \right\}$, and

$$\begin{aligned}
\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) &= \sum_{h=1}^{2} \lambda_{Ah}(\boldsymbol{\theta}_{0R}) \frac{n_h}{n} \hat{p}_{Ah} \begin{pmatrix} 1 & \mathrm{v}_h \end{pmatrix} \\
&= \sum_{h=1}^{2} C_h(\boldsymbol{\theta}_{0R}) \begin{pmatrix} 1 & \mathrm{v}_h \end{pmatrix} \\
&= \begin{pmatrix} C_1(\boldsymbol{\theta}_{0R}) + C_2(\boldsymbol{\theta}_{0R}) & C_1(\boldsymbol{\theta}_{0R}) \end{pmatrix}_{1\times 2},
\end{aligned}$$

where $C_h(\boldsymbol{\theta}_{0R}) = D_h(\boldsymbol{\theta}_{0R})\hat{p}_{Ah}$ and $D_h(\boldsymbol{\theta}_{0R}) = \lambda_{Ah}(\boldsymbol{\theta}_{0R})\dfrac{n_h}{n}$. It follows that,

$$
\begin{aligned}
&[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D1}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \\
&= \frac{\mu_1(\boldsymbol{\theta}_{0R})}{[\Upsilon_1(\boldsymbol{\theta}_{0R})\Upsilon_2(\boldsymbol{\theta}_{0R})]} \begin{pmatrix} \Upsilon_1(\boldsymbol{\theta}_{0R}) & -\Upsilon_1(\boldsymbol{\theta}_{0R}) \\ -\Upsilon_1(\boldsymbol{\theta}_{0R}) & \Upsilon_1(\boldsymbol{\theta}_{0R}) + \Upsilon_2(\boldsymbol{\theta}_{0R}) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \\
&= \frac{\mu_1(\boldsymbol{\theta}_{0R})}{[\Upsilon_1(\boldsymbol{\theta}_{0R})]^2\Upsilon_2(\boldsymbol{\theta}_{0R})} \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \Upsilon_1(\boldsymbol{\theta}_{0R}) & -\Upsilon_1(\boldsymbol{\theta}_{0R}) \\ -\Upsilon_1(\boldsymbol{\theta}_{0R}) & \Upsilon_1(\boldsymbol{\theta}_{0R}) + \Upsilon_2(\boldsymbol{\theta}_{0R}) \end{pmatrix} \\
&= \frac{\mu_1(\boldsymbol{\theta}_{0R})}{[\Upsilon_1(\boldsymbol{\theta}_{0R})]^2} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix};
\end{aligned}
$$

$$
\begin{aligned}
&[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D2}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \\
&= \frac{\mu_2(\boldsymbol{\theta}_{0R})}{[\Upsilon_1(\boldsymbol{\theta}_{0R})\Upsilon_2(\boldsymbol{\theta}_{0R})]} \begin{pmatrix} \Upsilon_1(\boldsymbol{\theta}_{0R}) & -\Upsilon_1(\boldsymbol{\theta}_{0R}) \\ -\Upsilon_1(\boldsymbol{\theta}_{0R}) & \Upsilon_1(\boldsymbol{\theta}_{0R}) + \Upsilon_2(\boldsymbol{\theta}_{0R}) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} [\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1} \\
&= \frac{\mu_2(\boldsymbol{\theta}_{0R})}{\Upsilon_1(\boldsymbol{\theta}_{0R})[\Upsilon_2(\boldsymbol{\theta}_{0R})]^2} \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \Upsilon_1(\boldsymbol{\theta}_{0R}) & -\Upsilon_1(\boldsymbol{\theta}_{0R}) \\ -\Upsilon_1(\boldsymbol{\theta}_{0R}) & \Upsilon_1(\boldsymbol{\theta}_{0R}) + \Upsilon_2(\boldsymbol{\theta}_{0R}) \end{pmatrix} \\
&= \frac{\mu_2(\boldsymbol{\theta}_{0R})}{[\Upsilon_2(\boldsymbol{\theta}_{0R})]^2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix};
\end{aligned}
$$

$$
\begin{aligned}
&[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D1}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) \\
&= \frac{\mu_1(\boldsymbol{\theta}_{0R})}{[\Upsilon_1(\boldsymbol{\theta}_{0R})]^2} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} C_1(\boldsymbol{\theta}_{0R}) + C_2(\boldsymbol{\theta}_{0R}) \\ C_1(\boldsymbol{\theta}_{0R}) \end{pmatrix} \qquad = \frac{\mu_1(\boldsymbol{\theta}_{0R})C_1(\boldsymbol{\theta}_{0R})}{[\Upsilon_1(\boldsymbol{\theta}_{0R})]^2} \begin{pmatrix} 0 \\ 1 \end{pmatrix};
\end{aligned}
$$

$$
\begin{aligned}
&[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D2}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) \\
&= \frac{\mu_2(\boldsymbol{\theta}_{0R})}{[\Upsilon_2(\boldsymbol{\theta}_{0R})]^2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} C_1(\boldsymbol{\theta}_{0R}) + C_2(\boldsymbol{\theta}_{0R}) \\ C_1(\boldsymbol{\theta}_{0R}) \end{pmatrix} \\
&= \frac{\mu_2(\boldsymbol{\theta}_{0R})C_2(\boldsymbol{\theta}_{0R})}{[\Upsilon_2(\boldsymbol{\theta}_{0R})]^2}\begin{pmatrix} 1 \\ -1 \end{pmatrix};
\end{aligned}
$$

$$
\begin{aligned}
&\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D1}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D1}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) \\
&= \frac{[\mu_1(\boldsymbol{\theta}_{0R})]^2}{[\Upsilon_1(\boldsymbol{\theta}_{0R})]^2}\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)]^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) \\
&= \frac{[\mu_1(\boldsymbol{\theta}_{0R})]^2}{[\Upsilon_1(\boldsymbol{\theta}_{0R})]^3}\frac{1}{\Upsilon_2(\boldsymbol{\theta}_{0R})}\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A)\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} \Upsilon_1(\boldsymbol{\theta}_{0R}) & -\Upsilon_1(\boldsymbol{\theta}_{0R}) \\ -\Upsilon_1(\boldsymbol{\theta}_{0R}) & \Upsilon_1(\boldsymbol{\theta}_{0R}) + \Upsilon_2(\boldsymbol{\theta}_{0R}) \end{pmatrix} \\
&\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) \\
&= \frac{[\mu_1(\boldsymbol{\theta}_{0R})]^2}{[\Upsilon_1(\boldsymbol{\theta}_{0R})]^3}\begin{pmatrix} C_1(\boldsymbol{\theta}_{0R}) + C_2(\boldsymbol{\theta}_{0R}) & C_1(\boldsymbol{\theta}_{0R}) \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) \\
&= \frac{[\mu_1(\boldsymbol{\theta}_{0R})]^2}{[\Upsilon_1(\boldsymbol{\theta}_{0R})]^3}C_1(\boldsymbol{\theta}_{0R})\begin{pmatrix} 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R}, \hat{\mathbf{p}}_A) \\
&= \frac{[\mu_1(\boldsymbol{\theta}_{0R})]^2}{[\Upsilon_1(\boldsymbol{\theta}_{0R})]^3}C_1(\boldsymbol{\theta}_{0R})\begin{pmatrix} 0 & 1 \end{pmatrix}\begin{pmatrix} C_1(\boldsymbol{\theta}_{0R}) + C_2(\boldsymbol{\theta}_{0R}) \\ C_1(\boldsymbol{\theta}_{0R}) \end{pmatrix} \\
&= \frac{[\mu_1(\boldsymbol{\theta}_{0R})]^2}{[\Upsilon_1(\boldsymbol{\theta}_{0R})]^3}[C_1(\boldsymbol{\theta}_{0R})]^2.
\end{aligned}
$$

The elements of (4.20) can be further simplified as follows:

$$\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D2}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D2}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)$$

$$= \frac{[\mu_2(\boldsymbol{\theta}_{0R})]^2}{[\Upsilon_2(\boldsymbol{\theta}_{0R})]^2}\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)$$

$$= \frac{[\mu_2(\boldsymbol{\theta}_{0R})]^2}{[\Upsilon_1(\boldsymbol{\theta}_{0R})][\Upsilon_2(\boldsymbol{\theta}_{0R})]^3}\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)\begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} \Upsilon_1(\boldsymbol{\theta}_{0R}) & -\Upsilon_1(\boldsymbol{\theta}_{0R}) \\ -\Upsilon_1(\boldsymbol{\theta}_{0R}) & \Upsilon_1(\boldsymbol{\theta}_{0R})+\Upsilon_2(\boldsymbol{\theta}_{0R}) \end{pmatrix}$$

$$\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)$$

$$= \frac{[\mu_2(\boldsymbol{\theta}_{0R})]^2}{[\Upsilon_2(\boldsymbol{\theta}_{0R})]^3}\begin{pmatrix} C_1(\boldsymbol{\theta}_{0R})+C_2(\boldsymbol{\theta}_{0R}) & C_1(\boldsymbol{\theta}_{0R}) \end{pmatrix}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)$$

$$= \frac{[\mu_2(\boldsymbol{\theta}_{0R})]^2}{[\Upsilon_2(\boldsymbol{\theta}_{0R})]^3}C_2(\boldsymbol{\theta}_{0R})\begin{pmatrix} 1 & -1 \end{pmatrix}\begin{pmatrix} C_1(\boldsymbol{\theta}_{0R})+C_2(\boldsymbol{\theta}_{0R}) \\ C_1(\boldsymbol{\theta}_{0R}) \end{pmatrix}$$

$$= \frac{[\mu_2(\boldsymbol{\theta}_{0R})]^2}{[\Upsilon_2(\boldsymbol{\theta}_{0R})]^3}[C_2(\boldsymbol{\theta}_{0R})]^2;$$

$$\mathbf{M}_{D2}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D1}$$

$$= \frac{\mu_2(\boldsymbol{\theta}_{0R})\mu_1(\boldsymbol{\theta}_{0R})}{\Upsilon_1(\boldsymbol{\theta}_{0R})\Upsilon_2(\boldsymbol{\theta}_{0R})}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \Upsilon_1(\boldsymbol{\theta}_{0R}) & -\Upsilon_1(\boldsymbol{\theta}_{0R}) \\ -\Upsilon_1(\boldsymbol{\theta}_{0R}) & \Upsilon_1(\boldsymbol{\theta}_{0R})+\Upsilon_2(\boldsymbol{\theta}_{0R}) \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$= \frac{\mu_2(\boldsymbol{\theta}_{0R})\mu_1(\boldsymbol{\theta}_{0R})}{\Upsilon_2(\boldsymbol{\theta}_{0R})}\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix};$$

$$
\begin{aligned}
&\frac{1}{n}\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A1}\partial\hat{p}_{A1}} \\
&= 2[\beta_{A0R}]^2\left\{D_1(\boldsymbol{\theta}_{0R})\begin{pmatrix}1 & 1\end{pmatrix}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D1}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)\right\} \\
&- 2[\beta_{A0R}]^2\left\{[D_1(\boldsymbol{\theta}_{0R})]^2\begin{pmatrix}1 & 1\end{pmatrix}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\begin{pmatrix}1 & 1\end{pmatrix}'\right\} \\
&+ 2[\beta_{A0R}]^2\left[D_1(\boldsymbol{\theta}_{0R})\begin{pmatrix}1 & 1\end{pmatrix}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D1}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)\right] \\
&- 2[\beta_{A0R}]^2\left[\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D1}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D1}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\right] \\
&\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A) \\
&= 2[\beta_{A0R}]^2 D_1(\boldsymbol{\theta}_{0R})\begin{pmatrix}1 & 1\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix}\frac{\mu_1(\boldsymbol{\theta}_{0R})C_1(\boldsymbol{\theta}_{0R})}{[\Upsilon_1(\boldsymbol{\theta}_{0R})]^2} \\
&- 2[\beta_{A0R}]^2[D_1(\boldsymbol{\theta}_{0R})]^2\frac{1}{\Upsilon_1(\boldsymbol{\theta}_{0R})\Upsilon_2(\boldsymbol{\theta}_{0R})}\begin{pmatrix}1 & 1\end{pmatrix}\begin{pmatrix}\Upsilon_1(\boldsymbol{\theta}_{0R}) & -\Upsilon_1(\boldsymbol{\theta}_{0R})\\ -\Upsilon_1(\boldsymbol{\theta}_{0R}) & \Upsilon_1(\boldsymbol{\theta}_{0R})+\Upsilon_2(\boldsymbol{\theta}_{0R})\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix} \\
&+ 2[\beta_{A0R}]^2 D_1(\boldsymbol{\theta}_{0R})\begin{pmatrix}1 & 1\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix}\frac{\mu_1(\boldsymbol{\theta}_{0R})C_1(\boldsymbol{\theta}_{0R})}{[\Upsilon_1(\boldsymbol{\theta}_{0R})]^2} \\
&- 2[\beta_{A0R}]^2\frac{[\mu_1(\boldsymbol{\theta}_{0R})]^2}{[\Upsilon_1(\boldsymbol{\theta}_{0R})]^3}[C_1(\boldsymbol{\theta}_{0R})]^2 \\
&= -\frac{2[\beta_{A0R}]^2}{\Upsilon_1(\boldsymbol{\theta}_{0R})}\left\{[D_1(\boldsymbol{\theta}_{0R})]^2 - 2D_1(\boldsymbol{\theta}_{0R})\frac{\mu_1(\boldsymbol{\theta}_{0R})C_1(\boldsymbol{\theta}_{0R})}{\Upsilon_1(\boldsymbol{\theta}_{0R})} + \frac{[\mu_1(\boldsymbol{\theta}_{0R})]^2}{[\Upsilon_1(\boldsymbol{\theta}_{0R})]^2}[C_1(\boldsymbol{\theta}_{0R})]^2\right\} \\
&= -\frac{2[\beta_{A0R}]^2}{\Upsilon_1(\boldsymbol{\theta}_{0R})}\left\{D_1(\boldsymbol{\theta}_{0R}) - \frac{\mu_1(\boldsymbol{\theta}_{0R})}{\Upsilon_1(\boldsymbol{\theta}_{0R})}C_1(\boldsymbol{\theta}_{0R})\right\}^2;
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{n}\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A2}\partial\hat{p}_{A2}} \\
&= 2[\beta_{A0R}]^2\left\{D_2(\boldsymbol{\theta}_{0R})\begin{pmatrix}1 & 0\end{pmatrix}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D2}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)\right\} \\
&- 2[\beta_{A0R}]^2\left\{[D_2(\boldsymbol{\theta}_{0R})]^2\begin{pmatrix}1 & 0\end{pmatrix}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\begin{pmatrix}1 & 0\end{pmatrix}'\right\} \\
&+ 2[\beta_{A0R}]^2\left[D_2(\boldsymbol{\theta}_{0R})\begin{pmatrix}1 & 0\end{pmatrix}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D2}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)\right] \\
&- 2[\beta_{A0R}]^2\left[\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D2}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D2}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\right] \\
&\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A) \\
&= 2[\beta_{A0R}]^2\left\{\frac{D_2(\boldsymbol{\theta}_{0R})\mu_2(\boldsymbol{\theta}_{0R})}{[\Upsilon_2(\boldsymbol{\theta}_{0R})]^2}\begin{pmatrix}1 & 0\end{pmatrix}\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)\right\} \\
&- 2[\beta_{A0R}]^2\left\{\frac{[D_2(\boldsymbol{\theta}_{0R})]^2}{\Upsilon_1(\boldsymbol{\theta}_{0R})\Upsilon_2(\boldsymbol{\theta}_{0R})}\begin{pmatrix}1 & 0\end{pmatrix}\begin{pmatrix}\Upsilon_1(\boldsymbol{\theta}_{0R}) & -\Upsilon_1(\boldsymbol{\theta}_{0R})\\ -\Upsilon_1(\boldsymbol{\theta}_{0R}) & \Upsilon_1(\boldsymbol{\theta}_{0R})+\Upsilon_2(\boldsymbol{\theta}_{0R})\end{pmatrix}\begin{pmatrix}1 & 0\end{pmatrix}'\right\} \\
&+ 2[\beta_{A0R}]^2\left[D_2(\boldsymbol{\theta}_{0R})\frac{\mu_2(\boldsymbol{\theta}_{0R})}{[\Upsilon_2(\boldsymbol{\theta}_{0R})]^2}\begin{pmatrix}1 & 0\end{pmatrix}\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)\right] \\
&- 2[\beta_{A0R}]^2\frac{[\mu_2(\boldsymbol{\theta}_{0R})]^2}{[\Upsilon_2(\boldsymbol{\theta}_{0R})]^3}[C_2(\boldsymbol{\theta}_{0R})]^2 \\
&= 2[\beta_{A0R}]^2\left\{\frac{D_2(\boldsymbol{\theta}_{0R})\mu_2(\boldsymbol{\theta}_{0R})}{[\Upsilon_2(\boldsymbol{\theta}_{0R})]^2}\begin{pmatrix}1 & -1\end{pmatrix}\begin{pmatrix}C_1(\boldsymbol{\theta}_{0R})+C_2(\boldsymbol{\theta}_{0R})\\ C_1(\boldsymbol{\theta}_{0R})\end{pmatrix}\right\} \\
&- 2[\beta_{A0R}]^2\left\{\frac{[D_2(\boldsymbol{\theta}_{0R})]^2}{\Upsilon_2(\boldsymbol{\theta}_{0R})}\begin{pmatrix}1 & -1\end{pmatrix}\begin{pmatrix}1\\ 0\end{pmatrix}\right\} \\
&+ 2[\beta_{A0R}]^2\left[D_2(\boldsymbol{\theta}_{0R})\frac{\mu_2(\boldsymbol{\theta}_{0R})}{[\Upsilon_2(\boldsymbol{\theta}_{0R})]^2}\begin{pmatrix}1 & -1\end{pmatrix}\begin{pmatrix}C_1(\boldsymbol{\theta}_{0R})+C_2(\boldsymbol{\theta}_{0R})\\ C_1(\boldsymbol{\theta}_{0R})\end{pmatrix}\right] \\
&- 2[\beta_{A0R}]^2\frac{[\mu_2(\boldsymbol{\theta}_{0R})]^2}{[\Upsilon_2(\boldsymbol{\theta}_{0R})]^3}[C_2(\boldsymbol{\theta}_{0R})]^2;
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{n}\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A2}\partial\hat{p}_{A2}} \\
&= -\frac{2[\beta_{A0R}]^2}{\Upsilon_2(\boldsymbol{\theta}_{0R})}\left\{[D_2(\boldsymbol{\theta}_{0R})]^2 - 2C_2(\boldsymbol{\theta}_{0R})D_2(\boldsymbol{\theta}_{0R})\mu_2(\boldsymbol{\theta}_{0R})\frac{1}{\Upsilon_2(\boldsymbol{\theta}_{0R})} + \frac{[\mu_2(\boldsymbol{\theta}_{0R})]^2}{[\Upsilon_2(\boldsymbol{\theta}_{0R})]^2}[C_2(\boldsymbol{\theta}_{0R})]^2\right\} \\
&= -\frac{2[\beta_{A0R}]^2}{\Upsilon_2(\boldsymbol{\theta}_{0R})}\left\{D_2(\boldsymbol{\theta}_{0R}) - \frac{\mu_2(\boldsymbol{\theta}_{0R})}{\Upsilon_2(\boldsymbol{\theta}_{0R})}C_2(\boldsymbol{\theta}_{0R})\right\}^2;
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{n}\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A1}\partial\hat{p}_{A2}} \\
&= 2[\beta_{A0R}]^2\left[\lambda_{A1}(\boldsymbol{\theta}_{0R})\frac{n_1}{n}\begin{pmatrix}1 & 1\end{pmatrix}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D2}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)\right] \\
&- 2[\beta_{A0R}]^2\left[\lambda_{A1}(\boldsymbol{\theta}_{0R})\frac{n_1}{n}\begin{pmatrix}1 & 1\end{pmatrix}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\lambda_{A2}(\boldsymbol{\theta}_{0R})\frac{n_2}{n}\begin{pmatrix}1 & 0\end{pmatrix}'\right] \\
&+ 2[\beta_{A0R}]^2\left[\lambda_{A2}(\boldsymbol{\theta}_{0R})\frac{n_2}{n}\begin{pmatrix}1 & 0\end{pmatrix}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D1}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)\right] \\
&- 2[\beta_{A0R}]^2\left[\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D2}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{D1}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\right]
\end{aligned}
$$

$$
\begin{aligned}
&\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A) \\
&= 2[\beta_{A0R}]^2\left[D_1(\boldsymbol{\theta}_{0R})\frac{\mu_2(\boldsymbol{\theta}_{0R})C_2(\boldsymbol{\theta}_{0R})}{[\Upsilon_2(\boldsymbol{\theta}_{0R})]^2}\begin{pmatrix}1 & 1\end{pmatrix}\begin{pmatrix}1\\-1\end{pmatrix}\right] \\
&- 2[\beta_{A0R}]^2\left[\frac{D_1(\boldsymbol{\theta}_{0R})D_2(\boldsymbol{\theta}_{0R})}{\Upsilon_1(\boldsymbol{\theta}_{0R})\Upsilon_2(\boldsymbol{\theta}_{0R})}\begin{pmatrix}1 & 1\end{pmatrix}\begin{pmatrix}\Upsilon_1(\boldsymbol{\theta}_{0R}) & -\Upsilon_1(\boldsymbol{\theta}_{0R})\\-\Upsilon_1(\boldsymbol{\theta}_{0R}) & \Upsilon_1(\boldsymbol{\theta}_{0R})+\Upsilon_2(\boldsymbol{\theta}_{0R})\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}\right] \\
&+ 2[\beta_{A0R}]^2\left[D_2(\boldsymbol{\theta}_{0R})\frac{\mu_1(\boldsymbol{\theta}_{0R})C_1(\boldsymbol{\theta}_{0R})}{[\Upsilon_1(\boldsymbol{\theta}_{0R})]^2}\begin{pmatrix}1 & 0\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix}\right] \\
&- 2[\beta_{A0R}]^2\left[\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\begin{pmatrix}0 & 0\\0 & 0\end{pmatrix}[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)]^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_{0R},\hat{\mathbf{p}}_A)\right]
\end{aligned}
$$

$$\frac{1}{n}\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A1}\partial\hat{p}_{A2}}$$
$$= -2[\beta_{A0R}]^2\left[\frac{D_1(\boldsymbol{\theta}_{0R})D_2(\boldsymbol{\theta}_{0R})}{\Upsilon_1(\boldsymbol{\theta}_{0R})}\begin{pmatrix} 0 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right]$$
$$= 0.$$

Since

$$\frac{1}{n}\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A1}\partial\hat{p}_{A1}}$$
$$= -\frac{2[\beta_{A0R}]^2}{\Upsilon_1(\boldsymbol{\theta}_{0R})}\left\{D_1(\boldsymbol{\theta}_{0R}) - \frac{\mu_1(\boldsymbol{\theta}_{0R})}{\Upsilon_1(\boldsymbol{\theta}_{0R})}C_1(\boldsymbol{\theta}_{0R})\right\}^2 < 0,$$
$$\frac{1}{n}\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A2}\partial\hat{p}_{A2}}$$
$$= -\frac{2[\beta_{A0R}]^2}{\Upsilon_2(\boldsymbol{\theta}_{0R})}\left\{D_2(\boldsymbol{\theta}_{0R}) - \frac{\mu_2(\boldsymbol{\theta}_{0R})}{\Upsilon_2(\boldsymbol{\theta}_{0R})}C_2(\boldsymbol{\theta}_{0R})\right\}^2 < 0, \text{and}$$
$$\frac{1}{n}\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A1}\partial\hat{p}_{A2}} = 0; \frac{1}{n}\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A2}\partial\hat{p}_{A1}} = 0.$$

The determinant of $\frac{1}{n}\phi^{(2)}(\boldsymbol{\pi_A})$ in (4.20) is positive and $\frac{1}{n}\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A1}\partial\hat{p}_{A1}}$ is negative. That is, $\frac{1}{n}\phi^{(2)}(\hat{\mathbf{p}}_A)$ is a negative definite matrix. Therefore, $\phi(\hat{\mathbf{p}}_A)$ is a concave function. □

For an example, we choose the values for true parameters: $\beta_A = 1.5$, $\gamma_0 = 0.6$, and $\gamma_1 = -0.4$ to draw the three dimensional graph of $\frac{1}{n}\phi(\boldsymbol{\pi_A})$ versus $\boldsymbol{\pi_A} = (\pi_A(\mathrm{v}_1), \pi_A(\mathrm{v}_2))'$ when the number of participating patients is 500. This graph is in the Figure 4.1. This graph confirms the concaveness of the non-centrality parameter that is a function of the proportion of treatment assignment.

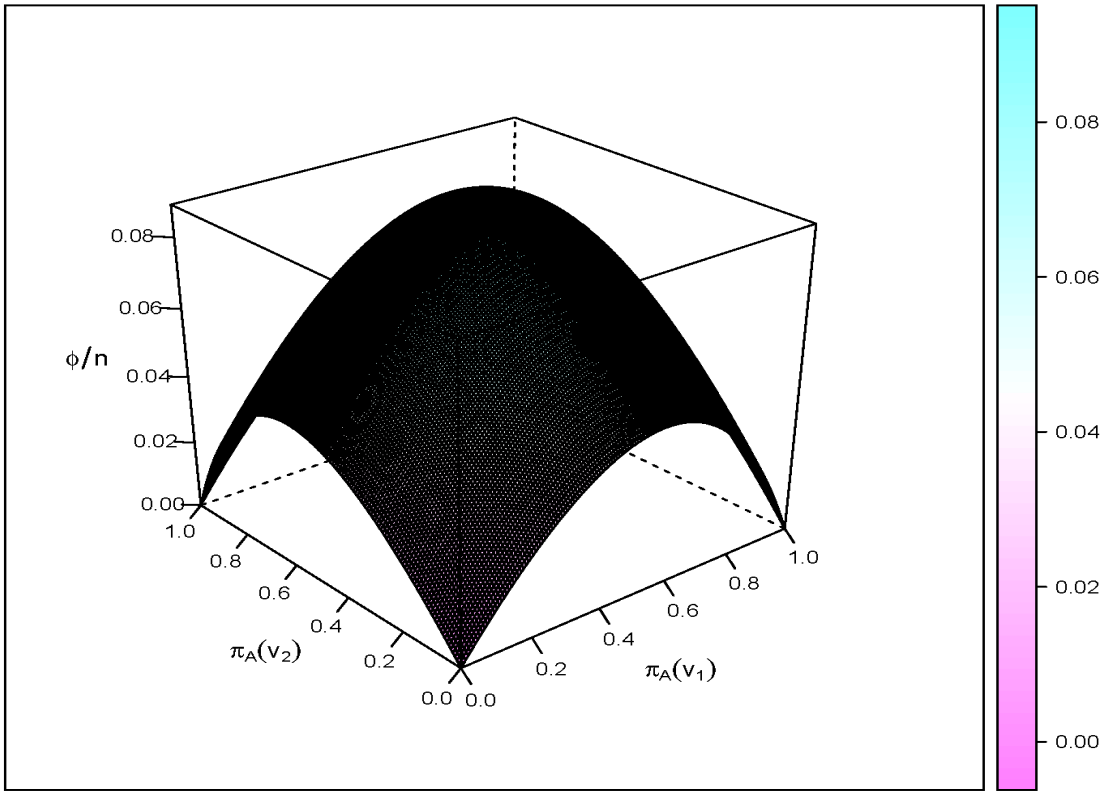

Figure 4.1: The graph of concaveness

4.4 Simulation Studies

In this section, we validate our theoretical results through simulation studies. We follow the same description and choice of true parameters used in § 3.6 for our simulation studies. Moreover, we verify whether the test statistic has valid the Type I error and discuss the computation of statistical power to apply for real data. The statistical power and Type I error rates are compared among three designs that were discussed in Chapter 3.

To verify the Type I error rates, data was generated under the null hypothesis H_{0I}. We then estimate the size of the test, α^*, by computing the proportion of rejections of H_{0I} for a fixed value α^*. We consider a test of the hypothesis for treatment by covariate interactions in the logit model defined in (3.1).

$$H_{0I} : \boldsymbol{\delta}_0 = \mathbf{0}_{4\times 1} \qquad H_{AI} : \boldsymbol{\delta}_0 \neq \mathbf{0}_{4\times 1} \tag{4.21}$$

where $\boldsymbol{\delta}_0 = (\delta_{10}, \delta_{20}, \delta_{30}, \delta_{40})'$. We use the Wald test that was derived in Section 4.2.1. Thus, the Wald test statistic for testing the hypothesis in (4.21) is

$$T_{WI} \;\; = \;\; n\hat{\boldsymbol{\delta}}_n'[S_{\boldsymbol{\Delta}_{11}^{[n]}}(\hat{\boldsymbol{\theta}}_n)]\hat{\boldsymbol{\delta}}_n \sim \chi^2_{4,\alpha}. \tag{4.22}$$

The rejection region for the test in (4.21) are

$$\{T_{WI} : T_{WI} < \chi^2_{4,\alpha^*/2} \text{ or } T_{WI} > \chi^2_{4,(1-\alpha^*/2)}\}, \tag{4.23}$$

where $P[T_{WI} < \chi^2_{4,\alpha^*/2}] = \alpha^*/2$ and $P[T_{WI} < \chi^2_{4,1-\alpha^*/2}] = 1 - \alpha^*/2$. We consider two scenarios to confirm whether Type I error is controllable for testing the hypothesis in

4.21. These results are outlined in Table 4.1 and 4.2.

Table 4.1: CARA, RA, and CR procedures from 10000 simulations with 500 number of patients. Size of the test (α^*) and estimated size of the test ($\hat{\alpha}^*$) with model parameters $\beta_{A0} = 1.00$, $\gamma_{00} = 0.50$, $\gamma_{10} = -0.18$, $\gamma_{20} = -0.30$, $\gamma_{30} = 0.25$, $\gamma_{40} = 0.10$, $\delta_{10} = 0.00$, $\delta_{20} = 0.00$, $\delta_{30} = 0.00$, $\delta_{40} = 0.00$.

Design	α^*	$\hat{\alpha}^*$	$\hat{\alpha}^* - \alpha^*$
CARA	0.10	0.1102	0.0102
	0.05	0.0517	0.0017
	0.01	0.0095	-0.0005
RA	0.10	0.0936	-0.0064
	0.05	0.0462	-0.0038
	0.01	0.0097	-0.0003
CR	0.10	0.0946	-0.0054
	0.05	0.0471	-0.0029
	0.01	0.0100	0.0000

Table 4.2: CARA, RA, and CR procedures from 10000 simulations with 500 number of patients. Size of the test (α^*) and estimated size of the test ($\hat{\alpha}^*$) with model parameters $\beta_{A0} = 0.50$, $\gamma_{00} = 0.50$, $\gamma_{10} = -0.18$, $\gamma_{20} = -0.30$, $\gamma_{30} = 0.25$, $\gamma_{40} = 0.10$, $\delta_{10} = 0.00$, $\delta_{20} = 0.00$, $\delta_{30} = 0.00$, $\delta_{40} = 0.00$.

Design	α^*	$\hat{\alpha}^*$	$\hat{\alpha}^* - \alpha^*$
CARA	0.10	0.1059	0.0059
	0.05	0.0540	0.004
	0.01	0.0116	0.0016
RA	0.10	0.0935	-0.0065
	0.05	0.0481	-0.0019
	0.01	0.0085	-0.0015
CR	0.10	0.0966	-0.0034
	0.05	0.0464	-0.0036
	0.01	0.0088	-0.0012

In CARA designs, the estimated size of errors are slightly higher than the true size of error; reverse results are generated by RA and CR designs. But, these three designs control Type I error because there is no significant deviation between the actual size

of test and estimated size of test.

Table 4.3: CARA, RA, and CR procedures from 5000 simulations with 500 number of patients and 0.05 size of the test. Power computation testing interaction with model parameters $\beta_{A0} = 0.50$, $\gamma_{00} = 0.25$, $\gamma_{10} = -0.20$, $\gamma_{20} = -0.40$, $\gamma_{30} = 0.35$, $\gamma_{40} = 0.20$, $\boldsymbol{\delta}_0 = (\delta_{10}, \delta_{20}, \delta_{30}, \delta_{40})'$.

$\boldsymbol{\delta}_0$	Design	Simulated Power	Conventional Power	Error	Average Design Variability	Overall Success Rates
$(0.10, -1.50, 0.00, 0.00)'$	CARA	0.7902	0.7186	0.0716	0.1559	0.5112
	RA	0.8068	0.7843	0.0225	0.1119	0.4851
	CR	0.8262	0.8103	0.0159	0.1038	0.4796
$(0.10, -1.50, 0.35, 0.00)'$	CARA	0.7984	0.7271	0.0713	0.1564	0.5193
	RA	0.8112	0.7928	0.0184	0.1127	0.4929
	CR	0.8284	0.8113	0.0171	0.1038	0.4903
$(0.30, -1.20, 0.35, 0.20)'$	CARA	0.6240	0.5643	0.0597	0.1569	0.5466
	RA	0.6318	0.6220	0.0098	0.1135	0.5280
	CR	0.6426	0.6231	0.0195	0.1038	0.5284
$(0.20, -2.00, 0.07, 0.05)'$	CARA	0.9608	0.8666	0.0942	0.1471	0.5152
	RA	0.9708	0.9481	0.0227	0.1108	0.4742
	CR	0.9774	0.9709	0.0065	0.1038	0.4621

Table 4.4: CARA, RA, and CR procedures from 5000 simulations with 1000 number of patients and 0.05 size of the test. Power computation testing interaction with model parameters $\beta_{A0} = 0.50$, $\gamma_{00} = 0.25$, $\gamma_{10} = -0.20$, $\gamma_{20} = -0.40$, $\gamma_{30} = 0.35$, $\gamma_{40} = 0.20$, $\boldsymbol{\delta}_0 = (\delta_{10}, \delta_{20}, \delta_{30}, \delta_{40})'$.

$\boldsymbol{\delta}_0$	Design	Simulated Power	Conventional Power	Error	Average Design Variability	Overall Success Rates
$(0.10, -1.50, 0.00, 0.00)'$	CARA	0.9916	0.9720	0.0196	0.1170	0.5131
	RA	0.9912	0.9891	0.0021	0.0784	0.4849
	CR	0.9944	0.9923	0.0021	0.0724	0.4791
$(0.10, -1.50, 0.35, 0.00)'$	CARA	0.9930	0.9756	0.0174	0.1173	0.5214
	RA	0.9960	0.9902	0.0058	0.0788	0.4933
	CR	0.9948	0.9923	0.0025	0.0722	0.4901
$(0.30, -1.20, 0.35, 0.20)'$	CARA	0.9450	0.9120	0.0330	0.1170	0.5478
	RA	0.9550	0.9442	0.0107	0.0795	0.5279
	CR	0.9540	0.9438	0.0102	0.0724	0.5279
$(0.20, -2.00, 0.07, 0.05)'$	CARA	0.9998	0.9893	0.0105	0.1103	0.5190
	RA	0.9996	0.9998	-0.0002	0.0774	0.4742
	CR	1.000	1.0000	0.0000	0.0724	0.4616

We justify the power computation method discussed in § 4.2.2 by simulation study. So, data was generated under the alternative hypothesis in (4.21). Simulated power is calculated by the proportion of rejections using the rejection region in (4.23). On the

other hand, we use the power computation method in § 4.2.2 to calculate statistical power to each simulation. Then, we calculate the average of these calculated powers and this power is called the conventional power. According to Demidenko (2007), there is no exact method to compute the statistical power for test in (4.21) because of nuisance parameters. The null model approach was used to estimate nuisance parameters in earlier literatures [see Whittemore (1981) and Self and Mauritsen (1988)]. So, estimation of nuisance parameters has not ended up in research of statistical power computation, particularly for a small sample and continuous covariates. In this simulation study, we are interested in comparing the statistical powers among three designs: CARA, RA, and CR designs. Because the power computation in § 4.2.2 is an approximate method, thus we compute error that is the difference between the simulated power and the conventional power.

The simulated powers in Tables 4.3 and 4.4 show that CR designs have more statistical powers than RA and CARA designs. Furthermore, RA designs are more efficient than CARA designs. When we examine the errors among three designs, CARA designs create more errors than RA and CR designs; also RA designs generate more errors compared to CR designs because CARA designs create more treatment imbalances over the covariate profiles [see Tables 3.5, 3.8, and 3.11]. However, ethical measures [Overall Success Rates] demonstrate reverse conclusions. Therefore, CARA designs are more ethical compared to RA and CR designs if true model contains treatment by covariate interactions. When we increase the number of patients from 500 to 1000, then (i) simulated powers increase, and (ii) errors decrease [see Tables 4.3 and 4.4].

We define the design variability for stratum h is SSE($\hat{P}_{Ah}$) for $h = 1, 2, \cdots$. Then, the

average design variability is calculated using the formula $\sum_{h=1}^{m} \hat{P}_{Ah}/m$. The results in Tables 4.3 and 4.4 show that the simulated power increases when the average design variability decreases. Moreover, CARA designs have more design variabilities than RA and CRD designs.

4.5 Conclusion

In this chapter, we have justified using the Wald test for adaptive designs when the sample size is large. We have discussed the power computation method for real data, and this method is an approximation to compute statistical power. Moreover, we have verified that the power calculation based on this approach generates values close to exact power when the sample size increases. Thus, we can use this power computation method to calculate power when we do sequential analysis in a clinical trial. We have proved that the statistical power depends on adaptive designs through the non-centrality parameter. In fact, we have demonstrated that this non-centrality parameter is a function of proportions of patients assigned to treatments over the covariate profiles. Furthermore, we have proved that this function is concave when we assume the logit model as an ideal model with binary covariates. Also, we have shown this non-centrality parameter depends on these quantities: the target allocation proportion, the bias of the randomization procedure from the target, and the variability induced by the randomization process.

Although CARA designs have less efficiency compare to RA and CR designs, CARA designs generate more ethically desirable outcomes than RA and CR designs if the exact model contains treatment by covariates interactions. The power of hypothesis

testing increases when design variability decreases. Further research requires extending Theorem 4.3.2 for more than one categorical covariates.

Chapter 5

Discussion and Conclusions

5.1 An overview for the contribution of this thesis

Treatment assignment methods play a significant role in an efficient statistical inference and the ethics of participating patients in clinical trials. Moreover, the efficient statistical inference is essential for the well-being of future patients. Adaptive designs are used to achieve these efficiencies and ethics goals. It is in interest to develop adaptive design methods to maximize the well-being of participating patients. Furthermore, Response-Adaptive (RA) and Covariate-Adjusted Response-Adaptive (CARA) designs are used to increase the well-being of participating patients. Also, the RA designs have been well established with the assumption of simple homogeneous parametric structure. The limited number of researchers have developed under the assumption of non-homogeneous parametric structure for RA designs. Due to increasing discoveries of biomarkers and identification of the observed diversity among patients, personalized medicine has interested to extend human life expectancy. In fact, these factors which are biomarkers and the observed heterogeneity among patients are covariates. Meanwhile, these covariates are used to achieve the ethics goals

for participating patients. CARA designs were developed to get benefits using covariates in randomization to maximize the well-being of participating patients.

We have investigated the performance of RA designs when a generalized linear mixed model (GLMM) is an ideal model. We have considered a logit model to examine the ethics of participating patients. We have shown that the estimators of model parameters are consistent and follow asymptotically normal distribution when the number of patients is assumed to be constant at each medical center. Also, we have introduced the new searching method to estimate of model parameters based on the influence function approach and derived a closed form expression for the asymptotic variance of the regression parameter estimators. Moreover, we have demonstrated that this searching method works better than Hessian matrix approach. We have verified that RA designs generate ethically desirable outcome compared to CR design by conducting simulation studies.

We have examined the performance of RA, CARA, and CR designs, for which responses come from GLM, measuring these quantities: (i) efficiency of statistical inference, and (ii) ethics of participating patients. Thus, we have considered the logit model with categorical covariates to investigate these designs. Furthermore, we have proved that the ML estimators of model parameters are consistent and follow asymptotically multivariate Gaussian distribution for adaptive designs. According to the present literature, when we study the large sample behavior of ML estimators for model parameters, **a regularity assumption: the Fisher information and observed Fisher information matrices are positive definite within a neighborhood near to vector of ideal parameters** is necessary to examine the asymptotic properties of these estimators. In fact, the boundary of the area is not exactly defined,

and this assumption is a strong assumption. Thus, we have proved that an assumption: the Fisher information matrix at the ideal vector of parameters is a positive definite matrix is sufficient to investigate the asymptotic properties of ML estimators.

We have demonstrated that RA design generates ethically desirable outcomes as well as more statistical efficiency compared to CARA design if there is no treatment by covariate interactions in an ideal model. Also, when a perfect model contains treatment by covariate interactions, CARA design is more ethical than RA design; however, RA design has more statistical power than CARA design.

We have justified that the Wald-type of test can be asymptotically applied for a general class of adaptive designs. Moreover, the power computation method has been discussed for adaptive designs when a logit model is an exact model. Also, we have verified that this power calculation method generates exact statistical power for a large number of participating patients based on simulation results. We have shown that the choice of adaptive designs affects the statistical power of hypothesis testing. Moreover, we have theoretically shown that the statistical power decreases with design variabilities of adaptive designs for which a covariate is in a logit model. Moreover, the simulation results have confirmed this behavior between statistical power and design variability for more than one covariates. Thus, our simulation results validate the feasibility of logical proof for which the statistical power decreases when design variability increases for more than one covariate. Therefore, we conclude this chapter by discussing some future works including the behavior of the statistical power and design variability.

5.2 Future Works

Based on contributions of this thesis, we discuss some future works to extend this research.

We have assumed that treatment assignment and center effects only influence response to investigate the performance of RA designs in a multi-center clinical trial in Chapter 2. Treatment effect can be efficiently estimated when we include covariates of patients in a model. The investigation of large sample properties of ML estimators for model parameters we have developed using the influence function method can be extended to generalized linear mixed models for the exponential family including covariates. We will examine the performance of the iteration method that we have introduced based on the influence function method and iteration method based on the Hessian matrix for GLM.

Many adaptive designs satisfy the Assumption 3.4.1 in Chapter 3. For instance, we will provide a logical proof for which the response adaptive (RA) and the covariate adaptive (CA) designs based the minimization method satisfy the Assumption 3.4.1 in Chapter 3. As we mentioned in § 5.1, we will theoretically show that the statistical power increases when the design variability decreases for more than one categorical covariates. CA designs are used in a clinical trial to improve the efficient statistical inference. In fact, the equal allocation has less efficiency compared to CA design. Thus, the design variability of CA design is smaller than CR design. We will demonstrate this conclusion using simulation study.

Bibliography

Agresti, A. and Hartzel, J. (2000). Strategies for comparing treatments on a binary response with multi-centre data, *Statistics in Medicine* **19**: 1115 – 1139.

Ai, C. and Norton, E. C. (2003). Interaction terms in logit and probit models, *Economics Letters* **80**: 123 – 129.

Albert, A. (1984). On the existence of maximum likelihood estimates in logistic regression models, *Biometrika* **71**(1): 1 – 10.

Anderson, G. W., Alice, G. and Ofer, Z. (2010). *An introduction to random matrices*, Cambridge University Press.

Anderson, T. W. (1966). *An introduction to multivariate statistical analysis*, John Wiley & Sons, Inc.

Bailey, R. A. (1987). Restricted randomization: A practical example, *Journal of the American Statistical Association* **82**(399): 712 – 719.

Baldi Antognini, A. and Zagoraiou, M. (2012). Multi-objective optimal designs in comparative clinical trials with covariates: the reinforced doubly adaptive biased coin design, *The Annals of Statistics* **40**(3): 1315 – 1345.

Basak, G., Biswas, A. and Volkov, S. (2009). An urn model for odds-ratio-based

response-adaptive phase iii clinical trials for two or more treatments, *Journal of Biopharmaceutical Statistics* **19**: 838 – 856.

Beitler, P. J. and Landis, J. R. (1985). A mixed effects model for categorical data, *Biometrics* **41**: 991 – 1000.

Brass, H. and Petras, K. (2011). *Quadrature theory : the theory of numerical integration on a compact interval, Mathematical Surveys and Monograghs*, Vol. 178, American Mathematical Society, Providence.

Broström, G. and Holmberg, H. (2001). Generalized linear models with clustered data: Fixed and random effects models, *Computational Statistics and Data Analysis* **55**: 3123 – 3134.

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multi-Model Inference*, Springer.

Castelloe, J. M. (2000). Sample size computations and power analysis with the sas system, *Proceedings of the Twenty-fifth Annual SAS Users Group International Conference*, SAS Institute Inc., Cary, NC.

Chow, SC. and Chang, M. (2008). Adaptive design methods in clinical trials - a review, *Orphanet Journal of Rare Diseases* **3:11**.

Csörgö, M. (1968). On the strong law of large numbers and the central limit theorem for martingales, *Transactions of the American Mathematical Society* **131**: 259 – 275.

Demidenko, E. (2007). Sample size determination for logistic regression revisited, *Statistics in Medicine* **26**(18): 3385 – 3397.

Dunn, P. (1997). James lind (1716-94) of edinburgh and the treatment of scurvy, *Archive of Disease in Childhood Fetal and Neonatal Edition* **76**(1): F64 – F65.

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

Eisele, J. and Woodroofe, M. (1995). Central limit theorems for doubly adaptive biased coin designs, *The Annals of Statistics* **23**(1): 234 – 254.

Fahrmeir, L. and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models, *The Annals of Statistics* **13**(1): 342 – 368.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models*, Springer-Verlag, New York.

FDA (2010). *Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics*.

FDA (2015). *Draft Guidance for Industry and Food and Drug Administration Staff: Adaptive Designs for Medical Device Clinical Studies*.

Fortin, M. (2013). Population-averaged predictions with generalized linear mixed-effects models in forestry: an estimator based on gauss-hermite quadrature, *Canadian Journal of Forest Research* **43**(2): 129 – 138.

Friedman, L. M., Furberg, C. D., DeMets, D. L., Reboussin, D. M. and Granger, C. B. (2015). *Fundamentals of Clinical Trials*, Springer.

Grizzle, J. E. (1987). Letter to the editor, *Controlled Clinical Trials* **8**: 392 – 393.

Hart, R. G., Halperin, J. L., Pearce, L. A., Anderson, D. C., Kronmal, R. A., McBride, R., Nasco, E., Sherman, D. G., Talbert, R. L. and Marler, J. R. (2003). Lessons from the stroke prevention in atrial fibrillation trials, *Annals of Internal Medicine* **138**: 831 – 838.

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

Hu, F. and Rosenberger, W. F. (2003). Optimality, variability, power: Evaluating response-adaptive randomization procedures for treatment comparisons, *Journal of the American Statistical Association* **98**(463): 671 – 678.

Hu, F., Rosenberger, W. F. and Zhang, LX. (2006). Asymptotically best response-adaptive randomization procedures, *Journal of Statistical Planning and Inference* **136**: 1911 – 1922.

Hu, F. and Zhang, LX. (2004). Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials, *The Annals of Statistics* **32**(1): 268 – 301.

Hu, J., Zhu, H. and Hu, F. (2015). A unified family of covariate-adjusted response-adaptive designs based on efficiency and ethics, *Journal of the American Statistical Association* **110**(509): 357 – 367.

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

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

Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications*, Springer New York, New York, NY.

Kahan, B. C. (2014). Accounting for centre-effects in multicentre trials with a binary outcome - when, why, and how?, *Medical Research Methodology* **14**: 20.

Königsberger, K. (2004). *Analysis 2*, Springer-Lehrbuch.

Lin, J., Lin, L. and Sankoh, S. (2016). A bayesian response-adaptive covariate-adjusted randomization design for clinical trials, *Journal of Biometrics & Biostatistics* **7**(287).

Lin, Y. and Su, Z. (2012). Balancing continuous and categorical baseline covariates in sequential clinical trials using the area between empirical cumulative distribution functions, *Statistics in Medicine* **31**: 1961 – 1971.

Lindsey, J. K. (1997). *Applying generalized linear models*, Springer Texts in Statistics.

Lyles, R. H., Lin, H. M. and Williamson, J. M. (2007). A practical approach to computing power for generalized linear models with nominal, count, or ordinal responses, *Statistics in Medicine* **26**: 1632 – 1648.

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.

McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models*, London ; New York : Chapman and Hall.

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*, London; New York : Chapman and Hall.

McCulloch, C. E., Searle, S. R. and Neuhaus, J. M. (2008). *Generalized, Linear, and Mixed Models*, Hoboken, N.J. : Wiley.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models, *Journal of the Royal Statistical Society. Series A (General)* **135**(3): 370 – 384.

Ning, J. and Huang, X. (2010). Response-adaptive randomization for clinical trials with adjustment for covariate imbalance, *Statistics in Medicine* **29**(17): 1761 – 1768.

Ortega, J. M. (1932). *Numerical analysis : a second course*, 1990 edn, Society for Industrial and Applied Mathematics.

Oyet, A. J. (1997). *Robust Designs for Wavelet Approximations of Nonlinear Models*, PhD thesis, University of Alberta.

Pocock, S. J. and Simon, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial, *Biometrics* **31**: 103 – 115.

Rosenberger, W. F., Flournoy, N. and Durham, S. D. (1997). Asymptotic normality of maximum likelihood estimators from multiparameter response-driven designs, *Journal of Statistical Planning and Inference* **60**: 69 – 76.

Rosenberger, W. F. and Hu, M. (2002). On the use of generalized linear models following a sequential design, *Statistics and Probability Letters* **56**: 155 – 161.

Rosenberger, W. F. and Lachin, J. M. (2002). *Randomization in clinical trials : theory and practice*, New York : Wiley.

Rosenberger, W. F. and Lachin, J. M. (2016). *Randomization in Clinical Trials: Theory and Practice*, second edn, Wiley.

Rosenberger, W. F., Stallard, N., Ivanova, A., Harper, C. N. and Ricks, M. L. (2001). Optimal adaptive designs for binary response trials, *Biometrics* **57**(3): 909 – 913.

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

Rosenberger, W. F., Vidyashankar, A. N. and Agarwal, D. K. (2001). Covariate-adjusted response-adaptive designs for binary response, *Journal of biopharmaceutical statistics* **11**(4): 227 – 236.

Rosenberger, W., Sverdlov, O. and Hu, F. (2012). Adaptive randomization for clinical trials, *Journal of Biopharmaceutical Statistics* **22**(4): 719 – 736.

Santner, T. J. and Duffy, D. E. (1986). A note on a. albert and j. a. anderson's conditions for the existence of maxista likelihood estimates in logistic regression models, *Biometrika* **73**(3): 755 – 758.

Seber, G. A. F. and Lee, A. J. (2003). *Linear regression analysis*, Hoboken, NJ : John Wiley.

Self, S. G., Mauritsen, R. H. and Ohara, J. (1992). Power calculations for likelihood ratio tests in generalized linear models, *Biometrics* **48**: 31 – 39.

Self, S. and Mauritsen, R. H. (1988). Power/sample size calculations for generalized linear models, *Biometrics* **44**: 79 – 86.

Selvaratnam, S., Oyet, A., Yi, Y. and Gadag, V. (2017). Estimation of a generalized linear mixed model for response-adaptive designs in multi-center clinical trials, *The Canadian Journal of Statistics* **45**(3): 310 – 325.

Shao, J. and Yu, X. (2013). Validity of tests under covariate adaptive biased coin randomization and generalized linear models, *Biometrics* **69**(4): 960 – 969.

Shen, L. Z. (1995). On optimal b-robust influence functions in semiparametric models, *The Annals of Statistics* **23**(1): 968 – 989.

Shieh, G. (2005). On power and sample size calculations for wald tests in generalized linear models, *Journal of Statistical Planning and Inference* **128**(1): 43 – 59.

Shieh, G. A. (2000). Comparison of two approaches for power and sample size calculations in logistic regression models, *Communications in Statistics - Simulation and Computation* **29**: 763 – 791.

Silvapulle, M. J. (1981). On the existence of maximum likelihood estimators for the binomial response models, *Journal of the Royal Statistical Society. Series B (Methodological)* **43**(3): 310 – 313.

Smith, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations, *Biometrika* **12**: 1 – 85.

Srivastava, M. S. (2002). *Methods of multivariate statistics*, Toronto, Ont. : Chichester : Wiley.

Sutradhar, B. C. (2003). An overview on regression models for discrete longitudinal responses, *Statistical Science* **18**(3): 377 – 393.

Sverdlov, O. (ed.) (2016). *Modern Adaptive Randomized Clinical Trials: Statistical and Practical Aspects*, Taylor & Francis Group.

Stroke Prevention in Atrial Fibrillation Investigators (1990). Design of a multicenter randomized trial for the stroke prevention in atrial fibrillation study, *Stroke* **21**: 538 – 545.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples, *Biometrika* **25**: 285 – 294.

Tuerlinckx, F., Rijmen, F., Verbeke, G. and Boeck, P. (2006). Statistical inference in generalized linear mixed models: A review, *British Journal of Mathematical and Statistical Psychology* **59**(2): 225 – 255.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the gauss-newton method, *Biometrika* **61**: 439 – 447.

Wei, L. J. (1978). An application of an urn model to the design of sequential controlled clinical trials, *Journal of the American Statistical Association* **73**: 559 – 563.

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

Whittemore, A. S. (1981). Sample size for logistic regression with small response probability, *Journal of the American Statistical Association* **76**: 27 – 32.

Yaseri, M., Zeraati, H., Mohammad, K., Soheilian, M., Ramezani, A., Eslani, M. and Peyman, G. (2014). Intravitreal bevacizumab injection alone or combined with triamcinolone versus macular photocoagulation in bilateral diabetic macular edema; application of bivariate generalized linear mixed model with asymmetric random effects in a subgroup of a clinical trial, *Journal of ophthalmic & vision research* **9**(4): 453 – 460.

Yates, F. (1948). Contribution to the discussion of "the validity of comparative experiments," by F. J. Anscombe, *Journal of the Royal Statistical Society, Series A* **111**(3): 204 – 205.

Yi, Y. (2013). Exact statistical power for response adaptive designs, *Computational Statistics and Data Analysis* **58**: 201 – 209.

Yi, Y. and Wang, X. (2007). Goodness-of-fit test for response adaptive clinical trials, *Statistics and Probability Letters* **77**: 1014 – 1020.

Yi, Y. and Wang, X. (2009). Response adaptive designs with a variance-penalized criterion, *Biometrical Journal* **51**(5): 763 – 773.

Yi, Y. and Wang, X. (2011). Comparison of wald, score, and likelihood ratio tests for

response adaptive designs, *Journal of Statistical Theory and Applications* **10**(4): 553 – 569.

Youden, W. J. (1972). Randomization and experimentation, *Technometrics* **14**(1): 13 – 22.

Yuan, Y. Huang, X. and Liu, S. (2011). A bayesian response-adaptive covariate-balanced randomization design with application to a leukemia clinical trial, *Statistics in Medicine* **30**: 1218 – 1229.

Zelen, M. (1969). Play the winner rule and the controlled clinical trial, *Journal of the American Statistical Association* **64**(325): 131 – 146.

Zhang, C. and Oyet, A. J. (2014). Second order longitudinal dynamic models with covariates: estimation and forecasting, *Metrika* **77**: 837 – 859.

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

Zhu, H. (2015). Covariate-adjusted response adaptive designs incorporating covariates with and without treatment interactions, *The Canadian Journal of Statistics* **43**(4): 534 – 553.