COVARIATE-ADJUSTED RESPONSE-ADAPTIVE DESIGN WITH LONGITUDINAL RESPONSES

YISHU WANG

Covariate-adjusted response-adaptive design with longitudinal responses

by

© Yishu Wang

A thesis submitted to the

School of Graduate Studies

in partial fulfilment of the

requirements for the degree of

Master of Science

Department of Mathematics and Statistics

Memorial University of Newfoundland

June 2012

St. John's Newfoundland

Abstract

Response-adaptive designs have been extensively studied and applied in clinical trials. However, few research is for longitudinal data and less is known on limiting allocation properties when response adaptive designs are used in a longitudinal setting. Zhang et. al. (2007) proposed a general covariate-adjusted response-adaptive (CARA) design for non-longitudinal clinical trials and explored its asymptotic properties. The objectives of this research are to extend the general CARA design to clinical trials with longitudinal responses, and to study the asymptotic properties of the parameter estimators and the allocation proportion. The explicit expressions for the limiting allocation proportions of the extended design are obtained. The generalized estimating equations and martingale theories are used to develop the asymptotic properties of regression parameters and allocation proportions. This research is also the first study on covariate-adjusted response-adaptive design with longitudinal clinical responses and with more than two competing treatments.

Acknowledgements

There are many people who helped me advance my research work to its current state.

Firstly, I would like to express my deep gratitude to my supervisor Dr. Yanqing Yi for her guidance and supervision throughout my Master's study. I am grateful to the financial support from Dr. Yi's research grant funded by the IRIF start-up from the Government of Newfoundland and Labrador, and appreciated for her encouragement during these years not only in study but also in many aspects of my everyday life.

Many thanks go to Dr. Brajendra Sutradhar, Dr. Zhaozhi Fan, Dr. Concepcion Loredo-Osti, and Dr. Asokan Variyath, who taught me courses and opened the gate for me to the world of Statistics. I would also like to give my sincere appreciation to Dr. Jie Xiao, who offered his generous help during my darkest time when working on my thesis. Not only his vast knowledge in mathematics helped me solving the mathematical problems, his patience and dedication in research also inspired me in my future study.

I also want to thank all my friends who have always been there for me, in good times and in bad times. With your wisdom, love and support, I had passed and will pass through all the difficulties in my life.

Last but not least, I would like to give my thanks to my parents. Without their love, unconditional support and sincere belief in me, I will never be a person like I am right now. Love you forever Mom and Dad.

Contents

Chapter 1

Introduction

This chapter introduces the background of response-adaptive designs of clinical trials, and gives a brief review of some important works that have been done in this field.

1.1 Introduction to Response-Adaptive Designs of Clinical Trials

"A clinical trial is basically an experiment designed to evaluate the beneficial and adverse effects of a new medical treatment or intervention." (Rosenberger and Lachin, 2002). The focus of this research is on clinical trials that aim at comparing two or more treatments, where a group of new therapy treated subjects may be compared to a group of conventional therapy treated subjects or a group of subjects who receive a placebo control.

Randomization is a fundamental principle that has been widely applied throughout the design, conduct and analysis of clinical trials. The use of randomization

serves several important purposes in a clinical study. It prevents allocating patients to treatments from personal bias, and provides a firm basis for most of the statistical methodology used to assess results of the trials. Most importantly, it promotes the comparability among the treatment groups by distributing both the known and unknown variables in a random way, which also ensures the interpretation of an observed group difference largely unambiguous (Brian and Andrew, 1999).

However, randomized clinical trials also create much controversy. The most challenging issue is the ethics of randomization. Response-adaptive designs were developed as one of the solutions to this ethical issue, and have been widely studied. The following sections will further discuss the ethics of randomization and the development of response-adaptive design.

1.1.1 Ethical Issues in Randomized Clinical Trials

Ethical issues generally exist throughout the design, conduct and analyze of clinical trials, while randomization is often identified as the central ethical issue in clinical trials which relates directly to the statistical aspects of design and analysis. Brian and Andrew (1999) illustrated the potential risk of employing randomization to assign participants in clinical trials through the example of using zidovudine (AZT) to treat AIDS. When the trials of AZT as a therapy for AIDS "were first announced, there was a large, vocal lobby against testing the drug in a controlled clinical trial where necessarily some patients would receive an 'inferior treatment'. Later, however, when the severity of some side effects was identified and the long term effectiveness of the drug in doubt, an equally vocal lobby called for AZT treatment to be aban-

doned. Expanding networks of 'support groups' makes these problems increasingly likely."(Brian and Andrew, 1999).

This example shows the uncertainty of treatment effects in clinical trials, and the concern of adverse impacts on patients who participate in the study, which reveal the major problems exist in randomly assigning patients to treatments in a study. When using probability as a method of assigning treatments to patients in clinical trials for randomization purpose, it gives patients a chance of being assigned to a potentially effective treatment, but also a chance not to receive a treatment which might potentially be very beneficial or miss the best time for receiving treatments. In addition, it exposes patients to the risk of receiving a treatment which may turn out to be highly toxic. This concern is referred to as "individual ethics" (Schwartz and Lellouch, 1971), the goal of which is for individual patients to receive the best possible treatment.

Despite not bringing individuals their best interest, randomized clinical trials have been embraced by most of today's scientists. Since when taking into consideration of "collective ethics" where to find the best treatment for the entire patient group is of concern (Schwartz and Lellouch, 1971), the fact that randomization can provide a solid foundation for inference of treatment effects, thus leading to a scientific comparison of treatment effects and the advancement of public health is very important to our society.

To achieve the balance between "individual ethics" and "collective ethics" is naturally challenging, and is still the subject of many researchers. As one of the attempts to solve this problem, information of accumulated responses in a clinical trial is used to skew the assignment of future patients in favour of the treatment that so far has

a "better" performance. Response-adaptive designs were thus proposed.

1.1.2 Development of Response-Adaptive Designs

Response-adaptive randomization is developed as an effort to mitigate the conflict between "individual ethics" and "collective ethics". In a response-adaptive randomized trial, "the probability of being assigned to a treatment is changed throughout the trial according to data which have already accrued about the treatment effect." (Rosenberger and Lachin, 2002). The goal is to assign more patients to the potentially "better" treatment for the benefit of "individual ethics", while the allocation procedure is still fully randomized for the inference of treatment effect which will benefit the "collective ethics". The early work on response-adaptive designs may be traced back to the play-the-winner rule by Zelen (1969). The work considered clinical trials with two treatments. The incoming patient's assignment depends on the response of the last patient in the study. If it was a success, the incoming patient is assigned to the same treatment, otherwise, to the opposite treatment. However, there design introduces selection biases. To deal with this issue, different allocation procedures have been proposed by many scholars, among which two major approaches have been broadly studied.

The first approach is based on an intuitive rule to adapt the allocation probabilities when assigning new patient to a treatment, which is not designed to target some specific allocation proportions. It is completely nonparametric but is not considered as optimal in a formal sense. Most designs using this approach are realized in the context of various urn models (Hu and Rosenberger, 2006; Rosenberger and Lachin, 2002).

Wei and Durham (1978) proposed the randomized play-the-winner rule based on Zelen's research (1969). The two competing treatments A and B are each represented by α balls in an urn. A ball is drawn and replaced. Patient is assigned to the treatment represented by the ball. A success will result in β,($\beta > 0$) balls representing the same treatment as the patient was assigned to added to the urn, a failure will result in β,($\beta > 0$) balls representing the opposite treatment being added to the urn. Their design skews the probability of treatment assignment to favor the potentially "better" treatment, rather than switching deterministically between treatments as in Zelen's (1969) play-the-winner rule. Athreya and Karlin (1968) proposed the generalized Friedman's urn model which is a natural design for clinical trials with K ($K \geq 2$) treatments. The allocation of incoming patient depends on a random variable whose distribution is related to the previous treatment assignment. Wei (1979) further explored the generalized Friedman's urn model and proposed a simple allocation rule based on the original work. Some other popular designs are the birth and death urn (Ivanova, Rosenberger, Durham, et al. , 2000), and the drop-the-loser rule (Ivanova, 2002) in which a tenary urn model is used and the variability of allocation proportions was proved to attain the lower bound of the variance of allocation proportions. (Hu et al. 2006).

The second approach is based on parametric models and designed to target the allocation proportion before the start of a trial. The target allocation proportion is a function of parameters which represent the treatment effects. Responses of sequentially accrued patients are used to update estimates of unknown parameters in most designs. This approach is often refereed to as sequential estimation procedures (Hu and Rosenberger, 2006). To obtain an initial estimate of unknown parameters, some

data must be available to compute the estimates. In practice, a sequential estimating trial usually begins with a certain number of patients assigned to each treatment before the proposed procedure begins. Melfi and Page (1995), and Melfi, Page and Geraldes (2001) explored properties of the sequential maximum likelihood procedure targeting the Neyman allocation defined by the ratio of the standard deviations of two competing treatments. Jennison and Turnbull (2000) proposed a group sequential adaptive design to targeting a predefined optimal allocation ratio for the difference of normal means. Eisele (1994), and Eisele and Woodroofe (1995) proposed a doubly-adaptive biased coin design to achieve any desired allocation proportion when comparing two treatments. The doubly-adaptive biased coin design yields a large family of sequential estimation procedures. Hu and Zhang (2004) generalized the design to multi-treatment cases, and derived the strong consistency and asymptotic normality of the design under some widely satisfied conditions.

1.1.3 Covariate-Adjusted Response-Adaptive (CARA) Designs

Although response-adaptive randomization successfully utilize accumulated information in responses in a study to assign more patients to a "better" treatment, as the complexity of modern clinical trials grows, adaptive designs based solely on patients responses to treatments are inadequate to address unique covariate structures underlying each patients's prognosis. The covariate information associated with each patient may have strong influences on comprehensive evaluation of the effectiveness of treatments. For example, the effectiveness of asthma treatment may depend on

whether the patient is a smoker or non-smoker and which age group the patient belongs to. The covariate-adjusted response-adaptive (CARA) designs are thus introduced.

Rosenberger et al. (2001) firstly used a logistic model to incorporate covariates into the allocation scheme. Even though they did not give any theoretical justifications and asymptotic properties, their simulation study indicated that their approach, together with the inclusion of the covariates, significantly reduced the percentage of treatment failures. Bandyopadhyay et al. (2007), developed a two-stage allocation rule for binary responses incorporating covariates. They showed several exact and limiting properties for the proportion of allocation and treatment failures in their work. Zhang et al.(2007) laid out a framework for a general CARA design, which can be applied to clinical trials to compare K treatments ($K \geq 2$) and are suitable for both discrete and continuous responses. Asymptotic properties of this general CARA design under certain widely satisfied conditions have also been studied.

As have been discussed, extensive research have been conducted towards response-adaptive designs. However, the methodology discussed above are only suitable for clinical trials in which responses are only observed once. When the responses of each patient are repeatedly recorded through a certain period of time, more complexity are added. The following discussion is on response-adaptive designs with longitudinal responses.

1.1.4 Response-Adaptive Designs with Longitudinal Responses

The responses of patients may often be observed repeatedly over different monitoring times which make the data of responses longitudinal in nature. For example, a randomized clinical trial was conducted for treating drug addiction where two treatments (Buprenorphine and Methadone) were compared for their ability to reduce opiate use among a group of 162 addicts (Johnson R E. et al., 1992). The outcome of this trial is a vector of repeated binary responses of whether an individual failed a urine test at each of 3 visits per week (on Monday, Wednesday and Friday) over a 17-week period. The reasons for collecting longitudinal data in clinical trials are to obtain a more precise estimate of the outcome and hence the treatment effect or to evaluate the treatment effect over time (Albert P., 1999). Therefore, if repeated responses are used in response-adaptive design to update the allocation probability of the incoming patient, intuitively it would be more efficient to assign more patients to the potentially "better" treatment. However, this setup is usually complicated due to the correlation within the longitudinal responses of each patient and the dependency results from the adaptation of treatment allocation. The literature in this case is also scanty.

Biswas and Dewanji (2004) developed the longitudinal randomized play-the-winner (LRPW) rule, which is an urn design that extends the randomized play-the-winner (RPW) rule to accommodate longitudinal binary responses. Biswas and Dewanji (2004) applied this design to investigate the effect of pulse electro-magnetic field (PEMF) for the treatment of patients with rheumatoid arthritis in the study conducted in the Indian Statistical Institute, Kolkata. The design successfully assigns

more patients to the better performing treatment. However, they did not consider the available covariates such as gender or age which may have significantly influenced the treatment effects and the allocation procedure of the design was restricted to clinical studies with only two competing treatments. Sutradhar, Biswas and Bari (2005) introduced a binary response-based longitudinal adaptive design. They utilized a similar allocation rule as the LRPW, but proposed a weighted generalized quasi-likelihood (WGQL) approach for the consistent and efficient estimation of the regression parameters including the treatment effects. Subsequently, Sutradhar and Jowaheer (2006) applied WGQL approach to analyze longitudinal count data accrued in clinical trials. However, these designs still did not go beyond the LRPW rule of Biswas and Dewanji (2004), and the current patient's covariate were not considered in the allocation design. The extension of allocation procedure to clinical trials with more than two competing treatments had not been studied.

Biswas et al. (2010) considered clinical trials with two competing treatments and proposed a covariate-adjusted longitudinal response-adaptive (CALRA) design using the log-odds ratio within the Bayesian framework. Their main allocation scheme is as follows. First, a total number of patients in the trial N is fixed before the study begins. To begin the adaptive process, $2m$ ($m \in \mathbb{N}^+$) patients were assigned equally to the two competing treatments using a restricted randomization procedure at the first stage. The second stage involves the remaining $(N-2m)$ patients, during which patients with covariate vector X is assigned to the default treatment with probability $p = p(\mathbf{X})$. The allocation rule is a function of the covariate of the incoming patient, and is based on all the available data up to that time point. To determine $p(X)$ appropriately, they define an utility function based on the likelihood function

of the available information. The correlation within the responses of each subject are assumed to follow an AR(1) structure. By maximizing the utility function, the allocation function of the incoming patient $p(X)$ is determined. The CALRA design is the first optimal design in clinical trials with longitudinal responses. It also managed to incorporate covariate information of the current patient into a longitudinal response setup and obtained its statistical properties through some optimality criterion.

However, this design can only be applied to clinical trials with two competing treatments, and it only designed for clinical trials with binary responses. The condition on using this design is too restrictive to generate to more general cases or other types of responses. The CALRA design also does not have an explicit expression for the variability of the design allocation and the limiting allocation proportion. Only a numeric study and an example on the effect of pulsed electro-magnetic field (PEMF) for the treatment of rheumatoid arthritis were provided to illustrate the effectiveness of the CALRA design.

1.2 Objective of This Research

The objectives of this research are to extend the application of the general CARA design to clinical trials with longitudinal responses and to study the asymptotic properties as well as to give explicit expressions for the limiting allocation proportions of the extended design. The generalized estimating equations and martingale theory are used to develop the asymptotic properties of regression parameters and allocation proportions. This research is also the first study on a covariate-adjusted response-adaptive design with longitudinal responses and more than two competing treatments.

Chapter 2

General CARA Design with Longitudinal Responses

This chapter is devoted to extend the application of the general CARA design (Zhang et al., 2007) to clinical trials with longitudinal responses. In section 2.1, we describe the framework and asymptotic properties of the general CARA design. In section 2.2, we give the framework and the general data setup in the longitudinal clinical trials. In section 2.3, the proposed allocation procedure for longitudinal clinical trials is introduced.

2.1 Introduction to the General CARA Design

The general framework of the CARA design is introduced by Zhang et al.(2007) and is reproduced in this section.

Zhang et al.(2007) considered a clinical trial with K treatments. Let $\mathbf{X_i}$ be the random treatment assignment for the i^{th} patient ($i = 1, 2, \cdots$), where $\mathbf{X_i} =$

$(x_{i1} \cdots x_{ik} \cdots x_{iK})$. If a patient is assigned to treatment k $(k = 1, 2, \ldots, K)$, then $x_{ik} = 1$ while $x_{ij} = 0, (j \neq k, j = 1, 2, \ldots, K)$. Let N_n^k be the number of subjects assigned to the k^{th} treatment among the first n patients. Write $\mathbf{N}_n = (N_n^1, \ldots, N_n^k)$, then $\mathbf{N}_n = \sum_{i=1}^{n} \mathbf{X}_i$. The observed response of a patient assigned to the k^{th} treatment is denoted as Y_k, the covariate vector of whom is denoted as $\boldsymbol{\xi}$. The response and covariate vector are supposed to satisfy

$$E[Y_k|\boldsymbol{\xi}] = p_k(\boldsymbol{\theta_k}, \boldsymbol{\xi}) \quad \boldsymbol{\theta_k} \in \boldsymbol{\Theta_k} \tag{2.1}$$

where $p_k(\cdot, \cdot)$, $k = 1, 2, \ldots, K$, are known functions, and θ_k are unknown parameters. The sequence $\{(Y_{m,1}, \ldots, Y_{m,K}, \boldsymbol{\xi}_m), m = 1, 2, \ldots\}$ is assumed to be independent and identically distributed as that of $\{(Y_1, \ldots, Y_K, \boldsymbol{\xi})\}$.

To start, a restricted randomization is used to assign m_0 $(m_0 \in \mathbb{N}^+)$ patients to each treatment. Assume that $m(m \geq Km_0)$ patients have already been assigned to treatments. Let $\hat{\boldsymbol{\theta}}_{\boldsymbol{m}}$ be the estimate of $\boldsymbol{\theta}$ at that stage. When the $(m+1)^{th}$ patient is ready to enter the study, the covariate of whom $\boldsymbol{\xi}_{m+1}$ is recorded and the patient will be assigned to treatment k with probability

$$\psi_k = P(X_{m+1,k} = 1|\mathcal{F}_m, \boldsymbol{\xi}_{m+1}) = \pi_k(\hat{\theta}_m, \boldsymbol{\xi}_{m+1}) \quad k = 1, \ldots, K \tag{2.2}$$

where $\mathcal{F}_m = \sigma(\mathbf{X}_1, \ldots, \mathbf{X}_m, \mathbf{Y}_1, \ldots, \mathbf{Y}_m, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m)$ is the sigma field of the history and $\pi_k(\cdot, \cdot)$ are some given functions that satisfy $0 < \pi_k(\cdot, \cdot) < 1$ for each k, and $\sum_{k=1}^{K} \pi_k(\cdot, \cdot) = 1$. The vector function $\boldsymbol{\pi(\cdot, \cdot)} = (\pi_1(\cdot, \cdot), \ldots, \pi_K(\cdot, \cdot))$ is called an allocation function.

Let

$$g_k(\boldsymbol{\beta^*}) = E(\pi_k(\boldsymbol{\beta^*}, \boldsymbol{\xi})), \quad \nu_k = g_k(\boldsymbol{\theta}) = E(\pi_k(\boldsymbol{\theta}, \boldsymbol{\xi})) \tag{2.3}$$

Then write

$$\boldsymbol{g}(\boldsymbol{\theta}) = (g_1(\boldsymbol{\theta}), \ldots, g_K(\boldsymbol{\theta})) \tag{2.4}$$

$$\boldsymbol{\nu} = (\nu_1, \ldots, \nu_K) \tag{2.5}$$

Under the assumption of the continuity of the allocation function $\boldsymbol{\pi(\theta, \xi)}$, Zhang et al.(2007) derived the following results. If for each treatment k ($k = 1, \ldots, K$), the parameter estimators satisfy the following equation

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{nk}} - \boldsymbol{\theta_k} = \frac{1}{n}\sum_{m=1}^{n} X_{m,k}\mathbf{h}_k(Y_{m,k}, \boldsymbol{\xi}_m)(1 + o(1)) + o(n^{-1/2}) \quad a.s., \tag{2.6}$$

where $\mathbf{h}_k$ are K functions with $E(\mathbf{h}_k(Y_k, \boldsymbol{\xi})|\boldsymbol{\xi}) = 0$, and $E\|\mathbf{h}_k(Y_k, \xi)\|^2 \leq \infty$, then the allocation proportion for the k^{th} treatment has a limiting value of ν_k. For patients with certain specific covariates $\boldsymbol{\xi}^*$, the limiting allocation proportion to treatment k is $\pi_k(\boldsymbol{\beta}, \boldsymbol{\xi}^*)$. The convergence rate for the allocation proportions and the estimators of parameters are $O\left(\sqrt{\frac{loglogn}{n}}\right)$. The allocation proportions and the estimators of parameters both have asymptotically multivariate normal distributions.

Zhang et al.(2007) examined a large applicable case where generalized linear models are used in the estimation of regression parameters for the exponential family. They defined

$$\mathbf{I}_k = \mathbf{I}_k(\boldsymbol{\theta}) = E(\pi_k(\boldsymbol{\theta}, \boldsymbol{\xi})I_k(\boldsymbol{\theta_k}|\boldsymbol{\xi})), \quad k = 1, \ldots, K \tag{2.7}$$

where $\mathbf{I}_k(\boldsymbol{\theta}|\boldsymbol{\xi}) = -E[\frac{\partial^2 log f_k(y_k|\boldsymbol{\xi}, \boldsymbol{\theta_k})}{\partial \boldsymbol{\theta_k^2}}|_{\boldsymbol{\xi}}] = a_k''(\mu_k)\mathbf{h}_k(\boldsymbol{\xi}\boldsymbol{\theta_k^T})\boldsymbol{\xi}^{\boldsymbol{T}}\boldsymbol{\xi}$ is the conditional Fisher information matrix. If $\mathbf{I}_k$ are nonsingular, then under certain condition, the asymptotic variance of the design is $\mathbf{V}_k = \mathbf{I}_k^{-1}$.

To extend their work to the longitudinal clinical trials, we need to set up a framework for longitudinal data in addition to their original notations.

2.2 Longitudinal Response Setup

Let $\mathbf{X_i} = (x_{i1} \cdots x_{ik} \cdots x_{iK})$ be the treatment assignment for the i^{th} patient. Let $\mathbf{Y}_i^k = (y_{i1}^k \cdots y_{iT_i}^k)'$ be a $T_i \times 1$ vector which denotes the responses of the i^{th} patient allocated to the k^{th} treatment collected over T_i different time points.

Assume there are p different covariates considered in the evaluation of patients' response, let $\boldsymbol{\xi_i} = (\boldsymbol{\xi_{i1}} \ldots \boldsymbol{\xi_{it}} \ldots \boldsymbol{\xi_{iT_i}})'$ be the $T_i \times p$ matrix that denotes the covariates of the i^{th} patient corresponding to the T_i different collecting time points.

Without lose of generality, we assume each patient's responses and covariates are repeatedly examined and recorded for the same fixed times denoted as T. i.e., $T_i = T, (i = 1, 2, \ldots)$. The time interval between all the patients' examination are equally spaced, and we treated it as a unit time. This assumption is reasonable when dealing with real life problems. Since most of the time, we would prefer combining the same amount of information from each individual to draw conclusions about treatment effects in clinical trials. Unbalanced information from different individual may result in biases.

Patients are assumed to sequentially enter into the study. Patients' first covariates are observed every time right before they entering the trial, and their following covariates are observed at the same time point as their repeated responses being observed. Patients' responses are also assumed to be able to be collected immediately after they are assigned to the treatment and after they are repeatedly examined over different time points.

Responses from different patients are assumed to be independent. However, the responses collected over different time points from the same patient are usually de-

pendent. The correlations between responses of each individual introduce much complication to the statistical inference for the designs. Varying from different problems we are dealing with, the covariates of each patient at different time point may change through time or may stay the same as collecting time point changing. Similarly, We assume that the covariates from different patients are independent, and the covariates collected from each patient have an independent and identical distribution as a known random variable $\boldsymbol{\xi} = (\boldsymbol{\xi_1}, \ldots, \boldsymbol{\xi_T})$.

Suppose given $\boldsymbol{\xi}_t, (t = 1, 2, \ldots, T)$, that the response of a trial to the treatment k $(k = 1, 2, \ldots, K)$ at measure time point t y_t^k has a marginal density in an exponential family with the form

$$f(y_t^k|\boldsymbol{\xi_t}) = \exp\{[y_t^k\theta_t^k - a_k(\theta_t^k)]\phi^k + b_k(y_t^k, \phi^k)\} \tag{2.8}$$

where $\theta_t^k = h(\eta_t^k)$ with $\eta_t^k = \boldsymbol{\xi_t'}\boldsymbol{\beta}^k$, and $\boldsymbol{\beta}^k \in \boldsymbol{\Theta}^k(k = 1, 2\cdots, K)$ are the $p \times 1$ vector of unknown parameters. $\boldsymbol{\Theta}^k \subset \mathbb{R}^p$ is the parameter space of $\boldsymbol{\beta}^k$. Write $\boldsymbol{\beta} = (\boldsymbol{\beta^{1\prime}} \cdots \boldsymbol{\beta^{k\prime}} \cdots \boldsymbol{\beta^{K\prime}})$ and $\boldsymbol{\Theta} = \boldsymbol{\Theta^1} \times \cdots \times \boldsymbol{\Theta^K}$. Let $\phi^k(k = 1, 2, \cdots, K)$ be the possibly unknown scale parameters.

2.3 Allocation Procedure of the General CARA Design with Longitudinal Responses

We consider clinical trials where patients sequentially enter the study. For the allocation of patients, we modify the scheme used in the general CARA design to adapt to the longitudinal data setting.

First, assign i_0 patients to each treatment by using a restricted randomization

and collect T times responses and covariates of all patients. Assume that $i(i \geq Ki_0)$ patients have already been assigned to treatments. By the time the $(i+1)^{th}$ patient is ready to be assigned to a treatment, the covariates of the $(i+1)^{th}$ patient are observed for the first time and denoted as $\boldsymbol{\xi_{i+1,1}}$. Until then, the response and covariates of the i^{th} patient are only observed once, which are recorded as $\mathbf{Y}_i^k = (y_{i1}^k), (k = 1, 2, \ldots, K)$ and $\boldsymbol{\xi_i} = (\boldsymbol{\xi_{i1}})'$. The responses and covariates of the $(i-1)^{th}$ patient only have two repeatedly measured records which are denoted as $\mathbf{Y}_{i-1}^k = (y_{i-1,1}^k, y_{i-1,2}^k), (k = 1, 2, \ldots, K)$ and $\boldsymbol{\xi_{i-1}} = \boldsymbol{(\xi_{i-1,1}, \xi_{i-1,2})'}$ respectively. In general, the responses and covariates of the $(i-t+1)^{th}$ patient by the time the $(i+1)^{th}$ patient entering the trial have t $(t = 1, 2 \ldots, T-1 \wedge (i-Ki_0))$ repeatedly measured records denoted as $\mathbf{Y}_{i-t+1}^k = (y_{i-t+1,1}^k, \ldots, y_{i-t+1,t+1}^k), (k = 1, 2, \ldots, K)$ and $\boldsymbol{\xi_{i-t+1}} = \boldsymbol{(\xi_{i-t+1,1}, \ldots, \xi_{i-t+1,t+1})'}$ for responses and covariates respectively. If $(i - i_0) \geq T$, then the first $(i - T + 1)$ patients would have their T times repeatedly observed responses and covariates.

After the $(i + 1)^{th}$ patient is assigned to a treatment, the response of whom will be observed for the first time and recorded as $\mathbf{Y}_{i+1}^k = (y_{i+1,1}^k)$. The responses and covariates of the previous $(i-t+1)$ $(1 \leq t \leq T-1 \wedge (i-i_0))$ patients will be observed and recorded again.

Usually, T is small but the total number of patients n in a trial is large, i.e., when n is large, only T patients do not have completely recorded responses and covariates before the i^{th} patient entering the trial, and this number is small compared to the number of patients with fully T records. Therefore, for notation simplicity, we still use $\mathbf{Y}_{i-t+1}^k$ and $\boldsymbol{\xi_{i-t+1}}$ to denote the responses and covariates of the $(i - t + 1)^{th}$ patient. The difference is when $T - 1 < t < i$, $\mathbf{Y}_{i-t+1}^k$ is a $T \times 1$ vector and $\boldsymbol{\xi_{i-t+1}}$ is a $T \times p$ matrix, whereas when $1 \leq t \leq T - 1$, $\mathbf{Y}_{i-t}^k$ is a $t \times 1$ vector and $\boldsymbol{\xi_{i-t}}$ is a $t \times p$

matrix.

Let $\mathcal{F}_i = \sigma(\boldsymbol{Y_1}, \boldsymbol{X_1}, \boldsymbol{\xi_1}, \ldots \boldsymbol{Y_i}, \boldsymbol{X_i}, \boldsymbol{\xi_i})$ be the sigma field generated from the entire history of all previous patients before the $(i+1)^{th}$ patient enters the trial. Let $\hat{\boldsymbol{\beta}}_i^k$ be an estimate of $\boldsymbol{\beta}^k (k = 1, 2, \ldots, K)$, which is based on the observed responses and their corresponding covariates among those previous i patients who were assigned to treatment k. Then $\hat{\boldsymbol{\beta}}_i = (\hat{\boldsymbol{\beta}}_i^{\boldsymbol{1\prime}}, \ldots, \hat{\boldsymbol{\beta}}_i^{\boldsymbol{K\prime}})$ is an estimate of $\boldsymbol{\beta} = (\boldsymbol{\beta}^{\boldsymbol{1\prime}}, \ldots, \boldsymbol{\beta}^{\boldsymbol{K\prime}})$ before the $(i+1)^{th}$ patient entering the study. Suppose that the $(i+1)^{th}$ patient is assigned to treatment k with probability

$$P(X_{i+1,k} = 1 | \mathcal{F}_i, \boldsymbol{\xi_{i+1,1}}) = \pi_k(\hat{\boldsymbol{\beta}}_{\boldsymbol{i}}, \boldsymbol{\xi_{i+1,1}}), \qquad k = 1, 2, \ldots, K \tag{2.9}$$

where $\pi_k(\cdot, \cdot)$ are some given functions that satisfy $0 < \pi_k(\cdot, \cdot) < 1$ for each k, and $\sum_{k=1}^{K} \pi_k(\cdot, \cdot) = 1$. We use the same notation as that in Zhang et al. (2007), and refer the vector function $\boldsymbol{\pi(\cdot, \cdot)} = (\pi_1(\cdot, \cdot), \ldots, \pi_K(\cdot, \cdot))$ as a allocation function. Let $g_k(\boldsymbol{\beta^*}) = E(\pi_k(\boldsymbol{\beta^*}, \boldsymbol{\xi_1}))$. Then

$$\begin{aligned} P(X_{i+1,k} = 1 | \mathcal{F}_i) &= E_{\boldsymbol{\xi}}(P(X_{i+1,k} = 1 | \mathcal{F}_i, \boldsymbol{\xi_{i+1,1}})) \\ &= g_k(\hat{\boldsymbol{\beta}}_{\boldsymbol{i}}), \qquad k = 1, \ldots, K \end{aligned} \tag{2.10}$$

Define

$$\lambda_k = g_k(\boldsymbol{\beta}) = E_{\boldsymbol{\xi}}(\pi_k(\boldsymbol{\beta}, \boldsymbol{\xi_1})) \quad k = 1, \ldots, K \tag{2.11}$$

and

$$\boldsymbol{\Lambda} = (\lambda_1, \ldots, \lambda_K) \tag{2.12}$$

When $\boldsymbol{\beta}$ are the real parameters from the distribution of responses and the distribution of the covariates are known, $\boldsymbol{\Lambda}$ is the allocation proportion among the K treatments based on the scheme described above. As discussed in Zhang et al.(2007), different choices of $\boldsymbol{\pi(\cdot, \cdot)}$ can generate different classes of designs.

Patients enter the trial sequentially and are allocated to treatments sequentially. The assignment of treatment to the $(i+1)^{th}$ patient depends only on the previously collected information $\mathcal{F}_i$ and his/her covariates $\boldsymbol{\xi_{i+1,1}}$. Therefore, the above equation (1.4) can also be written as

$$P(X_{i+1,k} = 1|\mathcal{F}_i, \boldsymbol{\xi_{i+1}}) = P(X_{i+1,k} = 1|\mathcal{F}_i, \boldsymbol{\xi_{i+1,1}}) = \pi_k(\hat{\boldsymbol{\beta}}_{\boldsymbol{i}}, \boldsymbol{\xi_{i+1,1}}), \qquad k = 1, 2, \ldots, K \tag{2.13}$$

The described allocation procedure above will be performed every time when a new patient entering the trial and their responses as well as the corresponding covariates will be measured and recorded each time. We follow this allocation method until the last patient enters into the trial and all the patients' responses and covarites over T time points are recorded.

Chapter 3

Asymptotic Properties

This chapter first introduces different methods for longitudinal data analysis. A detailed introduction is given for the generalized estimating equations which was used to analyse the longitudinal data in our design. The estimation of regression parameters is followed. Then the asymptotic properties of the estimators of parameters and the allocation proportions of our proposed designed are studied. A lemma and a theorem are proposed, the proofs to which are given respectively. At last, a comparison between our design and the general CARA design proposed by Zhang et al. (2007) is discussed. A comparison between our design and other designs with longitudinal responses is also discussed in this chapter.

3.1 Longitudinal Data Analysis

The data obtained from a longitudinal study are characterized by the fact that repeated observations for a subject tend to be correlated. When the outcomes are continuous, some common methods are mixed linear model (Laird and Ware, 1982;

Ware, 1985) and the general linear mixed-effects model (Verbeke and Molenberghs, 2000). These models sometimes are referred to as subject-specific (SS) models (Zeger, Liang and Albert, 1988), since the focus of which is usually on the response for an individual rather than for the population. Whereas in population studies, such as in clinical trials, where the difference in population-averaged response between several treatments is more of concern than the change in an individual's response, marginal models are usually used instead of a full likelihood procedure (Zeger, Liang and Albert, 1988).

Liang and Zeger (1986), and Zeger and Liang (1986) developed generalized estimating equations(GEE) procedure, which are essentially extended generalized linear models for the situation of correlated data. With the possible application to continuous data, GEE is most commonly used for discrete measurement sequences. The method combines estimating equations for the regression parameters with moment-based estimation for the correlation parameters based on the "working" correlation assumption. The model requires only the correct specification of the univariate marginal distributions provided the primary concern is on the regression parameters, not on the correlation structure (Liang and Zeger, 1986). It is assumed that the correlation matrix R, thus R_i, depends on a vector of association parameters denoted as $\boldsymbol{\alpha}$. The "working" correlation matrix is assumed to be of the same structure for all subjects which represents the average dependence among the repeated observations across subjects.

There are some theoretical considerations for the problems that may occur in the GEE estimation procedure. Crowder (1995), Sutradhar and Das (1999) argued that the strong difference between the "working" correlation and the true underlying

structure might result in efficiency loss of the estimator, and in some special cases there might not be consistent estimator for the "working" correlation. However, since the "working" correlations are only treated like nuisance parameters and used as devices to support estimation of the regression parameters, they should not be made a part of formal inference (Molenberghs and Verbeke, 2005). The GEE method yields consistent estimates of the regression coefficients and their standard errors, even with misspecification of the correlational structure. The loss of efficiency due to the misspecification of the correlation structure can be lessened as the number of subjects gets large (Molenberghs and Verbeke, 2005). Consistent variance estimates are also available under the weak assumption that a weighted average of the estimated correlation matrices converges to a fixed matrix (Molenberghs and Verbeke, 2005).

When GEE is deemed unsatisfactory in the cases when the correlation structure is of interest, there are some extension of GEE methods one can turn to, such as second-order extensions of these estimating equations, which are usually referred as GEE2 (Zhao and Prentice (1990)), and alternating logistic regressions (Carey, Zeger, and Diggle (1993)).

Since the regression parameters are the primary concern in this paper, we will use GEE method to estimate the longitudinal data collected in the clinical trial. The following section will give a detailed introduction to the GEE models.

3.1.1 Generalized Estimating Equations

Let $\mathbf{Y}_i = (y_{i1}, \ldots, y_{iT_i})^T$ be the $T_i \times 1$ vector of outcomes and $\boldsymbol{\xi}_i = (\boldsymbol{\xi}_{i1}, \ldots, \boldsymbol{\xi}_{iT_i})$ be the $p \times T_i$ matrix of covariate values for the i^{th} subject $(i = 1, \ldots, n)$.

The marginal density of y_{it} is assumed to be in an exponential family having the density

$$f(y_{it}|\boldsymbol{\xi_{it}}) = \exp\{[y_{it}\theta_{it} - a(\theta_{it})]\phi + b(y_{it}, \phi)\} \tag{3.1}$$

where $\theta_{it} = h(\eta_{it})$, $\eta_{it} = \boldsymbol{\xi'_{it}}\boldsymbol{\beta}$. As such, the first two moments of y_{it} are given by

$$E(y_{it}) = a'(\theta_{it}), \qquad Var(y_{it}) = a''(\theta_{it})\phi \tag{3.2}$$

Let $\boldsymbol{R_i(\alpha)}$ be a $T_i \times T_i$ symmetric matrix which fulfills the requirement of being a correlation matrix, and $\boldsymbol{\alpha}$ be an $s \times 1$ vector which fully characterizes $\boldsymbol{R_i(\alpha)}$. $\boldsymbol{R_i(\alpha)}$ is called a "working" correlation matrix.

Define

$$\boldsymbol{V_i} = \boldsymbol{A_i^{\frac{1}{2}} R_i(\alpha) A_i^{\frac{1}{2}}}\phi. \tag{3.3}$$

$\boldsymbol{V_i}$ will be the covariance matrix of y_i, $\boldsymbol{cov(Y_i)}$, if $\boldsymbol{R_i(\alpha)}$ is indeed the true correlation matrix for the $\mathbf{Y}_i$'s. The general estimating equations are defined to be

$$\sum_{i=1}^{n} \boldsymbol{D_i^T V_i^{-1} S_i} = \mathbf{0} \tag{3.4}$$

where $\mathbf{D}_i = \partial(\mathbf{a}'(\theta_i))/\partial\boldsymbol{\beta} = \boldsymbol{A_i \Delta_i X_i}$, $\boldsymbol{\Delta}_i = diag(\frac{d\theta_{it}}{d\eta_{it}})$. Equation can be reexpressed as a function of $\boldsymbol{\beta}$ alone by first replacing $\boldsymbol{\alpha}$ by $\hat{\boldsymbol{\alpha}}(\boldsymbol{Y}, \boldsymbol{\beta}, \phi)$, a $n^{\frac{1}{2}}$-consistent estimator of $\boldsymbol{\alpha}$ when $\boldsymbol{\beta}$ and ϕ are known. In addition, we replace ϕ by $\phi(\boldsymbol{Y}, \boldsymbol{\beta})$, a $n^{\frac{1}{2}}$- consistent estimator of ϕ when $\boldsymbol{\beta}$ is known. Then the estimating equation has the form

$$\sum_{i=1}^{n} \boldsymbol{U_i}(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta}))) = \mathbf{0} \tag{3.5}$$

and $\hat{\boldsymbol{\beta}}_n$ is the solution of the above equation.

Solving the GEE involves iterating between a modified Fisher scoring for estimating $\boldsymbol{\beta}$ and moment estimation for estimating $\boldsymbol{\alpha}$ and ϕ as a function of $\boldsymbol{\beta}$. Essentially, it involves the following steps as suggested by Liang and Zeger (1986).

1. Compute initial estimates for $\boldsymbol{\beta}$, say $\hat{\boldsymbol{\beta}}_0$, assuming independence between repeated responses first, or, in other words, using conventional linear regression.

2. Compute the Pearson residuals from the equation below based on the current value for $\boldsymbol{\beta}$.

$$\hat{r}_{i,t} = (y_{it} - a'(\hat{\theta_{it}}))/\sqrt{a''(\hat{\theta_{it}})} \tag{3.6}$$

The specific estiamator of $\boldsymbol{\alpha}$ depends on the choice of $\boldsymbol{R_i(\alpha)}$. The general approach is to estimate $\boldsymbol{\alpha}$ by a simple function of

$$\hat{R}_{uv} = \sum_{i=1}^{n} \hat{r_{iu}} \hat{r_{iv}}/(Q-p) \tag{3.7}$$

The value of ϕ can be estimated by

$$\hat{\phi}^{-1} = \sum_{i=1}^{n} \sum_{t=1}^{T_i} \hat{r_{it}}^2/(Q-p) \tag{3.8}$$

where $Q = \sum_{i=1}^{n} T_i$.

3. Based on the above estimation, $\boldsymbol{R_i(\alpha)}$ can be computed, as well as V_i from

$$\boldsymbol{V_i} = \boldsymbol{A_i^{\frac{1}{2}} R_i(\alpha) A_i^{\frac{1}{2}}} \phi \tag{3.9}$$

4. Then, given the current estimate of $\boldsymbol{\beta}$ after m iterations, say $\hat{\boldsymbol{\beta}}_m$, update the estimate for $\boldsymbol{\beta}$ by

$$\hat{\boldsymbol{\beta}}_{m+1} = \hat{\boldsymbol{\beta}}_m - [\sum_{i=1}^{n} \frac{\boldsymbol{\partial \mu_i}}{\boldsymbol{\partial \beta}} \boldsymbol{V_i^{-1}} (\frac{\boldsymbol{\partial \mu_i}}{\boldsymbol{\partial \beta^k}})']^{-1} \times [\sum_{i=1}^{n} \frac{\boldsymbol{\partial \mu_i}}{\boldsymbol{\partial \beta^k}} \boldsymbol{V_i^{-1}} (\boldsymbol{y_i - \mu_i})] \tag{3.10}$$

Repeat step 2, 3, 4 until convergence.

3.2 Estimation of Regression Parameters

The GEE approach is generalized to the response-adaptive designs with longitudinal responses in this section.

Binary data and count data are two most common types of responses obtained from clinical trials. In our design, we will mainly focus on these two types of response. We use the longitudinal data set up as described in section 2.2. The exponential family with the marginal density function as in equation (3.1) will accommodate these two important discrete distributions, binary and poisson, when ϕ^k are constants. We assume $\phi^k = 1$, $k = (1, 2, \ldots, K)$ in the following discussion. The results can be generalized to situations where ϕ^k, $k = (1, 2, \ldots, K)$ are unknown. The mean and variance of the response at time point t are thus given by

$$E(y_t^k|\boldsymbol{\xi_t}, \boldsymbol{\beta}^k) = a_k'(\theta_t^k) \qquad , var(y_t^k|\boldsymbol{\xi_t}, \boldsymbol{\beta}^k) = a_k''(\theta_t^k). \tag{3.11}$$

Let

$$\begin{aligned}
\boldsymbol{\mu_i^k} &= E(\boldsymbol{Y_i^k}|\boldsymbol{\xi_i}, \boldsymbol{\beta^k}) \\
&= (\mu_{i1}^k, \cdots, \mu_{it}^k, \cdots \mu_{iT}^k) \\
&= (a_k'(\theta_{i1}^k), \cdots a_k'(\theta_{it}^k), \cdots a_k'(\theta_{iT}^k)) \\
\mathbf{A_i^k} &= diag(\sigma_{i1}^k, \cdots, \sigma_{it}^k, \cdots, \sigma_{iT}^k) \\
&= diag(a_k''(\theta_{i1}^k), \cdots, a_k''(\theta_{it}^k), \cdots, a_k''(\theta_{iT}^k)).
\end{aligned} \tag{3.12}$$

For patients being allocated to the k^{th} treatment ($k = 1, 2, \ldots, K$), we define a likelihood function for $\boldsymbol{\beta}^k$ denoted as $l(\boldsymbol{Y_i^k}, \boldsymbol{\mu_i^k}|\boldsymbol{\beta^k})$ based on i^{th} patient information in a similar way as the quasi-likelihood function (Wedderburn, 1974). The likelihood

function satisfies the following condition:

$$\frac{\partial l(\boldsymbol{Y_i^k}, \boldsymbol{\mu_i^k}|\beta^{\boldsymbol{k}})}{\partial \boldsymbol{\mu_i^k}} = \boldsymbol{V_i}^{-1}(\boldsymbol{\alpha^k})(\boldsymbol{Y_i^k} - \boldsymbol{\mu_i^k}) \tag{3.13}$$

where

$$\boldsymbol{V_i}(\boldsymbol{\alpha^k}) = (\boldsymbol{A_i^k})^{\frac{1}{2}} \boldsymbol{R_i}(\boldsymbol{\alpha^k})(\boldsymbol{A_i^k})^{\frac{1}{2}}, \tag{3.14}$$

$\boldsymbol{R_i}(\boldsymbol{\alpha^k})$ is the "working" correlation for the k^{th} treatment, and $\boldsymbol{V_i}(\boldsymbol{\alpha^k})$ is the variance-covariance matrix of the responses from the i^{th} patient towards the k^{th} treatment which is fully characterized by an unknown $s \times 1$ vector parameter $\boldsymbol{\alpha}$. According to the properties of quasi-likihood function and its resemblence with the log-liklihood, we define the likihood function for $\boldsymbol{\beta}$ based on i^{th} patient information as:

$$L(\boldsymbol{Y_i}, \boldsymbol{\mu_i}|\beta) = \sum_{k=1}^{K} x_{ik} l(\boldsymbol{Y_i^k}, \boldsymbol{\mu_i^k}|\boldsymbol{\beta^k}) \tag{3.15}$$

Suppose the n^{th} patient just entered the trial. For each k,the estimating equation for $\boldsymbol{\beta^k}$ is

$$\begin{aligned}
\sum_{i=1}^{n} \frac{\partial L(\boldsymbol{Y_i}, \boldsymbol{\mu_i}|\boldsymbol{\beta})}{\partial \boldsymbol{\beta^k}} &= \sum_{i=1}^{n} x_{ik} \frac{\partial l(\boldsymbol{Y_i^k}, \boldsymbol{\mu_i^k}|\boldsymbol{\beta^k})}{\partial \boldsymbol{\beta^k}} \\
&= \sum_{i=1}^{n} x_{ik} \frac{\partial \boldsymbol{\mu_i^{k'}}}{\partial \boldsymbol{\beta^k}} \frac{\partial l(\boldsymbol{Y_i^k}, \boldsymbol{\mu_i^k}|\boldsymbol{\beta^k})}{\partial \boldsymbol{\mu_i^k}} \\
&= \sum_{i=1}^{n} x_{ik} \frac{\partial \boldsymbol{\mu_i^{k'}}}{\partial \boldsymbol{\beta^k}} \boldsymbol{V_i}^{-1}(\boldsymbol{\alpha^k})(\boldsymbol{Y_i^k} - \boldsymbol{\mu_i^k}) \\
&= \mathbf{0}
\end{aligned} \tag{3.16}$$

That is, the GEE estimating equation for $\boldsymbol{\beta^k}$ is

$$\sum_{i=1}^{n} x_{ik} \frac{\boldsymbol{\partial \mu_i^{k\prime}}}{\boldsymbol{\partial \beta^k}} ((\boldsymbol{A_i^k})^{\frac{1}{2}} \boldsymbol{R_i}(\boldsymbol{\alpha^k})(\boldsymbol{A_i^k})^{\frac{1}{2}})^{-1}(\boldsymbol{Y_i^k} - \boldsymbol{\mu_i^k}) = \mathbf{0} \tag{3.17}$$

One thing needs to be noticed is that the information of the $(n-t)^{th}$ $(t = 1, 2 \ldots, T)$ patient is not complete. However, since consistency and efficiency of the estimators will not be affected by the unbalanced responses (Liang and Zeger, 1986), GEE estimation procedure can still be applied.

3.3 Asymptotic Properties of the Estimators and Allocation Proportions

This section examines the asymptotic properties of the GEE estimators and the allocation proportion. To start, we firstly need to give some new notations and some assumptions.

Suppose there are already n patients entered the study, let

$$\mathcal{G}_i = \sigma(\mathcal{F}_i, \boldsymbol{\xi_i}, \boldsymbol{\xi_{i-1}}, \ldots, \boldsymbol{\xi_{i-T+2}}) \tag{3.18}$$

for $(i \leq n)$, then $\mathcal{G}_i$ is the $\sigma-$field generated by all the available information of previous i patients when there are already $n(n \geq i)$ patients entered the study. It includes the repeated responses and covariates of the $(i+1)^{th}$ patient which is not available in $\mathcal{F}_i$, when the patient first entered the study. Then, $\mathcal{F}_i \subseteq \mathcal{G}_i$ is trivial.

The allocation functions $\boldsymbol{\pi}(\cdot, \cdot)$ need to satisfy certain conditions similar as those assumed in Zhang et al. (2007), which are as follows:

Condition 1. *Assume that the parameter space $\boldsymbol{\Theta}_k$ is a bounded domain in $\mathbb{R}^p$, and that the true value $\boldsymbol{\beta}^k$ is an interior point of $\boldsymbol{\Theta}_k$ $(k = 1, 2, \ldots, K)$. Furthermore,*

1. *For each fixed* $\boldsymbol{\xi}$, $\pi_k(\boldsymbol{\beta^*}, \boldsymbol{\xi}) > 0$ *is a continuous function of* $\boldsymbol{\beta^*}$, $k = 1, 2, \ldots, K$.

2. *For each* $k = 1, \ldots, K$, $\pi_k(\boldsymbol{\beta^*}, \boldsymbol{\xi})$ *is differentiable with respect to* $\boldsymbol{\beta^*}$ *under the expectation, and there is a* $\delta > 0$ *such that*

$$g_k(\boldsymbol{\beta^*}) = g_k(\boldsymbol{\beta}) + (\boldsymbol{\beta^*} - \boldsymbol{\beta})(\frac{\partial g_k}{\partial \boldsymbol{\beta^*}}|_{\boldsymbol{\beta}})^T + o(\|\boldsymbol{\beta^*} - \boldsymbol{\beta}\|^{1+\delta}) \tag{3.19}$$

, *where* $\frac{\partial g_k}{\partial \boldsymbol{\beta^*}} = (\frac{\partial g_k}{\partial \beta^*_{1,1}}, \ldots, \frac{\partial g_k}{\partial \beta^*_{K,p}})$

Apply similar notations as used in GEE (Liang and Zeger, 1986). Let

$$\boldsymbol{D_i^k} = \frac{\boldsymbol{\partial \mu_i}}{\boldsymbol{\partial \beta^k}} = \boldsymbol{A_i^k \xi_i'} \quad \boldsymbol{S_i^k} = \boldsymbol{Y_i^k} - \boldsymbol{\mu_i^k} \tag{3.20}$$

Then we first introduce the following lemma.

Lemma 1. *Let*

$$\mathbf{B}_n^k = \frac{1}{n} \sum_{i=1}^{n} E(\pi_k(\hat{\boldsymbol{\beta}}_{i-1}, \boldsymbol{\xi_1})(\boldsymbol{D^k})' \boldsymbol{V^{-1}} \boldsymbol{D^k}) \tag{3.21}$$

and

$$\mathbf{B}^k = E(\pi_k(\boldsymbol{\beta}, \boldsymbol{\xi_1})(\boldsymbol{D^k})' \boldsymbol{V^{-1}} \boldsymbol{D^k}), \tag{3.22}$$

where $\boldsymbol{D^k} = \boldsymbol{A^k}(\boldsymbol{\beta})\boldsymbol{\xi'}$.

If the allocation functions satisfy condition 1., and $\hat{\boldsymbol{\beta}}_{\boldsymbol{i}}$, $i \in N^+$ *is a consistent estimator of the parameter* $\boldsymbol{\beta}$, *then*

$$\mathbf{B}_n^k \to \mathbf{B}^k \qquad a.s. \qquad as \ n \to \infty \tag{3.23}$$

Proof. Let

$$f_n(\boldsymbol{\xi}) = \frac{1}{n} \sum_{i=1}^{n} \pi_k(\hat{\boldsymbol{\beta}}_{i-1}, \boldsymbol{\xi_1})(\boldsymbol{D^k})' \boldsymbol{V^{-1}} \boldsymbol{D^k} \tag{3.24}$$

and

$$f(\boldsymbol{\xi}) = \pi_k(\boldsymbol{\beta}, \boldsymbol{\xi_1})(\boldsymbol{D^k})' \boldsymbol{V^{-1}} \boldsymbol{D^k}. \tag{3.25}$$

Under the assumptions in the lemma, for each fixed $\boldsymbol{\xi}$, $\pi_k(\boldsymbol{\beta^*}, \boldsymbol{\xi}) > 0$ is a continuous function of $\boldsymbol{\beta^*}$, $k = 1, 2, \ldots, K$, and

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{i}} \to \boldsymbol{\beta} \quad a.s. \qquad as \ i \to \infty. \tag{3.26}$$

We have, for each fixed $\boldsymbol{\xi}$,

$$f_n(\boldsymbol{\xi}) \to f(\boldsymbol{\xi}) \quad a.s. \qquad as \ n \to \infty. \tag{3.27}$$

Also, since for all n,

$$|f_n(\boldsymbol{\xi})| = |\frac{1}{n}\sum_{i=1}^{n} \pi_k(\hat{\boldsymbol{\beta}}_{\boldsymbol{i-1}}, \boldsymbol{\xi_1})(\boldsymbol{D^k})'\boldsymbol{V^{-1}}\boldsymbol{D^k}| \leq (D^k)'V^{-1}D^k \tag{3.28}$$

According to the Lebesgue's dominated convergence theorem,

$$E(f_n(\boldsymbol{\xi})) \to E(f(\boldsymbol{\xi})) \quad a.s. \quad as \ n \to \infty. \tag{3.29}$$

That is,

$$\begin{aligned} \mathbf{B}_n^k &= E(\frac{1}{n}\sum_{i=1}^{n} \pi_k(\hat{\boldsymbol{\beta}}_{\boldsymbol{i-1}}, \boldsymbol{\xi_1})\boldsymbol{D'}\boldsymbol{V^{-1}}\boldsymbol{D}) \\ &\to E(\pi_k(\boldsymbol{\beta}, \boldsymbol{\xi_1})(\boldsymbol{D^k})'\boldsymbol{V^{-1}}\boldsymbol{D^k}) \\ &= \mathbf{B}^k \quad a.s. \quad as \ n \to \infty \end{aligned} \tag{3.30}$$

□

Additional assumptions about the estimating equations are also needed.

Condition 2. *For $k = 1, \ldots, K$, assume the estimating equations satisfy the following conditions:*

1. *$\hat{\boldsymbol{\alpha}}^k$ converges to some limit as $n \to \infty$, and $\hat{\boldsymbol{\alpha}}^k$ are $n^{\frac{1}{2}}$-consistent estimators for $\boldsymbol{\alpha}^k$ given $\boldsymbol{\beta}^k$,*

2. $\mathbf{B}^k$, $\mathbf{B}_n^k$ *are nonsingular for all* $n \in \mathbb{N}^+$.

3. $\sum_{i=1}^{n} Var(x_{i,k}(\boldsymbol{B^k})^{-1}\boldsymbol{D_i'V_i^{-1}S_i}|\mathcal{G}_{i-1}) \to \infty$ *as* $n \to \infty$

4. $\sum_{i=1}^{n} \frac{1}{i^2} Var(x_{i,k}\boldsymbol{D_i'V_i^{-1}D_i}|\mathcal{G}_{i-1}) < \infty$ *as* $n \to \infty$

5. $\|(\boldsymbol{B^k})^{-1}\boldsymbol{D_n'V_n^{-1}S_n}\|_2 < \infty$ *for all* $n \in \mathbb{N}^+$

Theorem 1. *Under mild regularity conditions, and Condition(1) and Condition(2), we have for* $k = (1, \ldots, K)$*:*

$$P(x_{n,k} = 1) \to \lambda_k, \quad P(x_{n,k} = 1|\mathcal{F}_{n-1}, \boldsymbol{\xi}_{n,1} = \boldsymbol{\xi^*}) \to \pi_k(\boldsymbol{\beta}, \boldsymbol{\xi^*}) \quad a.s \tag{3.31}$$

and

$$\frac{\boldsymbol{N_n}}{\boldsymbol{n}} - \boldsymbol{\Lambda} = O\left(\sqrt{\frac{loglogn}{n}}\right) \quad \hat{\boldsymbol{\beta}}_{\boldsymbol{n}} - \boldsymbol{\beta} = O\left(\sqrt{\frac{loglogn}{n}}\right) \quad a.s. \tag{3.32}$$

Further, let

$$\boldsymbol{Var}^k = (\boldsymbol{B}^k)^{-1}E[\pi_k(\boldsymbol{\beta}, \boldsymbol{\xi_1})(\boldsymbol{D^k})'\boldsymbol{V}^{-1}\boldsymbol{Cov}(\boldsymbol{Y}|\boldsymbol{\xi_1})\boldsymbol{V}^{-1}\boldsymbol{D^k}](\boldsymbol{B}^k)^{-1}$$

$$\boldsymbol{Var} = \boldsymbol{diag}(\boldsymbol{Var^1}, \ldots, \boldsymbol{Var^K})$$

$$\boldsymbol{\Sigma} = \boldsymbol{diag}(\boldsymbol{\Lambda}) - \boldsymbol{\Lambda}^T\boldsymbol{\Lambda} + 2\sum_{k=1}^{K}(\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\beta^k}})\boldsymbol{Var}^k(\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\beta^k}})^T$$

Then

$$\sqrt{n}\left(\frac{N_n}{n} - \Lambda\right) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\Sigma}), \quad \sqrt{n}\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} - \boldsymbol{\beta}\right) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{Var}) \tag{3.33}$$

Proof. We complete the proof in three steps. First, we prove the result 3.32.

Let

$$\boldsymbol{U}_i(\boldsymbol{\beta^k}, \boldsymbol{\alpha}) = x_{ik}(\boldsymbol{D_i^k})'\boldsymbol{V}_i(\boldsymbol{\alpha}^k)^{-1}\boldsymbol{S_i^k}. \tag{3.34}$$

For notation simplicity, we drop the superscript k in the estimating equation (3.6) from all the term except for $\boldsymbol{\beta^k}$ and x_{ik} for the moment, and use x_{ik}, $\boldsymbol{\beta^k}$ only to represent that it is the k^{th} treatment that we are dealing with. Then, the estimating equation for $\boldsymbol{\beta^k}$ can be written as,

$$\begin{aligned}\mathbf{0} &= \sum_{i=1}^{n} x_{ik} \frac{\boldsymbol{\partial \mu_i'}}{\boldsymbol{\partial \beta^k}} (\boldsymbol{A_i^{\frac{1}{2}} R_i(\alpha) A_i^{\frac{1}{2}}})^{-1} (\boldsymbol{Y_i - \mu_i}) \\ &= \sum_{i=1}^{n} x_{ik} \boldsymbol{D_i' V_i(\alpha)^{-1} S_i} \\ &= \sum_{i=1}^{n} \boldsymbol{U_i(\beta^k, \alpha)}\end{aligned} \tag{3.35}$$

.

Let $\hat{\boldsymbol{\beta}}_n^k$ denote the solution to the estimation equation (3.35), under the first term in Condition 2, when there are n patients in the trial. There exists $\varepsilon > 0$, such that

$$\begin{aligned}\mathbf{0} &= \sum_{i=1}^{n} \boldsymbol{U_i}(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}^{\boldsymbol{k}}, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}^{\boldsymbol{k}})) \\ &= \sum_{i=1}^{n} \boldsymbol{U_i}(\boldsymbol{\beta^k}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta^k})) + \sum_{i=1}^{n} \frac{\boldsymbol{dU_i}(\boldsymbol{\beta^k}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta^k}))}{\boldsymbol{d\beta^k}} (\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}^{\boldsymbol{k}} - \boldsymbol{\beta^k}) + o(\|\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}^{\boldsymbol{k}} - \boldsymbol{\beta^k}\|^{1+\varepsilon})\end{aligned} \tag{3.36}$$

.

Thus

$$\frac{\sum_{i=1}^{n} \boldsymbol{U_i}(\boldsymbol{\beta^k}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta^k}))}{n^{\frac{1}{2}}} + \frac{\sum_{i=1}^{n} \frac{\boldsymbol{dU_i}(\boldsymbol{\beta^k}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta^k}))}{\boldsymbol{d\beta^k}}}{n} n^{\frac{1}{2}} (\hat{\boldsymbol{\beta}}^{\boldsymbol{k}} - \boldsymbol{\beta^k}) + \frac{o(\|\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}^{\boldsymbol{k}} - \boldsymbol{\beta^k}\|^{1+\varepsilon})}{n^{\frac{1}{2}}} = \mathbf{0}. \tag{3.37}$$

Since

$$\frac{d\boldsymbol{U_i}(\boldsymbol{\beta^k}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta^k}))}{d\boldsymbol{\beta^k}} = \frac{\partial \boldsymbol{U_i}(\boldsymbol{\beta^k}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta^k}))}{\partial \boldsymbol{\beta^k}} + \frac{\boldsymbol{\partial U_i}(\boldsymbol{\beta^k}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta^k}))}{\boldsymbol{\partial \hat{\alpha}}} \frac{\boldsymbol{\partial \hat{\alpha}}(\boldsymbol{\beta^k})}{\boldsymbol{\partial \beta^k}} \tag{3.38}$$

and $\frac{\partial \boldsymbol{U_i}(\boldsymbol{\beta^k}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta^k}))}{\partial \hat{\boldsymbol{\alpha}}}$ are linear combinations of $\boldsymbol{y_i} - \boldsymbol{\mu_i}$, the expectation of $\boldsymbol{y_i} - \boldsymbol{\mu_i}$ given $\boldsymbol{\xi}_i$ is $\mathbf{0}$.

We have,

$$\sum_{i=1}^{n} \frac{\partial \boldsymbol{U_i}(\boldsymbol{\beta^k}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta^k}))}{\partial \hat{\boldsymbol{\alpha}}} = o_p(n) \qquad \frac{\boldsymbol{\partial \hat{\alpha}(\beta^k)}}{\boldsymbol{\partial \beta^k}} = O_p(1) \tag{3.39}$$

$$\frac{\partial \boldsymbol{U_i}(\boldsymbol{\beta^k}, \hat{\alpha}(\boldsymbol{\beta^k}))}{\partial \boldsymbol{\beta^k}} = -x_{ik}\mathbf{D_i'V_i^{-1}D_i} \tag{3.40}$$

Then according to equations (3.38), (3.39), (3.40)

$$\frac{\sum_{i=1}^{n} \frac{d\boldsymbol{U_i}(\boldsymbol{\beta}, \hat{\alpha}(\boldsymbol{\beta}))}{d\boldsymbol{\beta}}}{n} = -\frac{\sum_{i=1}^{n} x_{ik}\mathbf{D_i'V_i^{-1}D_i}}{n} + o_p(1) \tag{3.41}$$

Fix $\boldsymbol{\beta^k}$, use Taylor expansion to expand function $\frac{\sum_{i=1}^{n} \boldsymbol{U_i}(\boldsymbol{\beta^k}, \hat{\alpha}(\boldsymbol{\beta^k}))}{n^{\frac{1}{2}}}$ at α,

$$\frac{\sum_{i=1}^{n} \boldsymbol{U_i}(\boldsymbol{\beta^k}, \hat{\alpha}(\boldsymbol{\beta^k}))}{n^{\frac{1}{2}}} = \frac{\sum_{i=1}^{n} \boldsymbol{U_i}(\boldsymbol{\beta^k}, \alpha)}{n^{\frac{1}{2}}} + \frac{\sum_{i=1}^{n} \frac{\partial \boldsymbol{U_i}(\boldsymbol{\beta^k}, \hat{\alpha}(\boldsymbol{\beta^k}))}{\partial \alpha}}{n} n^{\frac{1}{2}}(\hat{\alpha} - \alpha) + o_p(n^{-\frac{1}{2}}) \tag{3.42}$$

for a $\sqrt{n}$-consistent estimator of α, $\sqrt{n}(\hat{\alpha} - \alpha) = O_p(1)$.

Therefore,

$$\frac{\sum_{i=1}^{n} \boldsymbol{U_i}(\boldsymbol{\beta^k}, \hat{\alpha}(\boldsymbol{\beta^k}))}{n^{\frac{1}{2}}} = \frac{\sum_{i=1}^{n} \boldsymbol{U_i}(\boldsymbol{\beta^k}, \alpha)}{n^{\frac{1}{2}}} + o_p(1) \tag{3.43}$$

According to equations (3.37), (3.41), (3.43)

$$\frac{\sum_{i=1}^{n} x_{ik}\mathbf{D_i'V_i^{-1}D_i}}{n} n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}^{\boldsymbol{k}} - \boldsymbol{\beta^k}) = \frac{\sum_{i=1}^{n} \boldsymbol{U_i}(\boldsymbol{\beta^k}, \alpha)}{n^{\frac{1}{2}}} + o(n^{-\frac{1}{2}}) \tag{3.44}$$

Write

$$M_n = \sum_{i=1}^{n} [x_{ik}\mathbf{D_i'V_i^{-1}D_i} - E(x_{ik}\mathbf{D_i'V_i^{-1}D_i}|\mathcal{G}_{i-1})]. \tag{3.45}$$

For $m \leq n$, we have

$$\begin{aligned}
E(M_n|\mathcal{G}_m) &= E(\sum_{i=1}^{n}[x_{ik}\mathbf{D_i'V_i^{-1}D_i} - E(x_{ik}\mathbf{D_i'V_i^{-1}D_i}|\mathcal{G}_{i-1})]|\mathcal{G}_m) \\
&= \sum_{i=1}^{n} E(x_{i,k}\boldsymbol{D_i'V_i^{-1}D_i}|\mathcal{G}_m) - \sum_{i=1}^{n} E[E(x_{i,k}\boldsymbol{D_i'V_i^{-1}D_i}|\mathcal{G}_{i-1})|\mathcal{G}_m] \\
&= \sum_{i=m+1}^{n} E(x_{ik}\boldsymbol{D_i'V_i^{-1}D_i}|\mathcal{G}_m) + \sum_{i=1}^{m} x_{ik}\mathbf{D_i'V_i^{-1}D_i} \\
&- \sum_{i=m+1}^{n} E(x_{ik}\boldsymbol{D_i'V_i^{-1}D_i}|\mathcal{G}_m) - \sum_{i=1}^{m} E(x_{ik}\mathbf{D_i'V_i^{-1}D_i}|\mathcal{G}_{i-1})] \\
&= \sum_{i=1}^{m}[x_{ik}\mathbf{D_i'V_i^{-1}D_i} - E(x_{ik}\mathbf{D_i'V_i^{-1}D_i}|\mathcal{G}_{i-1})] \\
&= M_m.
\end{aligned} \tag{3.46}$$

According to the definition, M_n is a martingale, and

$$\begin{aligned}
\triangle M_i &= M_i - M_{i-1} \\
&= x_{ik}\boldsymbol{D_i'V_i^{-1}D_i} - E(x_{ik}\boldsymbol{D_i'V_i^{-1}D_i}|\mathcal{G}_{i-1})
\end{aligned} \tag{3.47}$$

is a martingale difference.

Under the fourth term in Condition 2,

$$\sum_{n=1}^{\infty}\frac{1}{n^2}E(\triangle M_n \triangle M_n'|\mathcal{G}_{n-1}) = \sum_{n=1}^{\infty}\frac{1}{n^2}Var(x_{i,k}\boldsymbol{D_i'V_i^{-1}D_i}|\mathcal{G}_{n-1}) < \infty. \tag{3.48}$$

According to the law of large numbers for martingales, $\frac{1}{n}M_n \to 0$ with probability 1.

From lemma 1, $\mathbf{B}_n^k$ converges almost surely to $\mathbf{B}^k$. Thus we have

$$\begin{aligned}
\sum_{i=1}^{n} x_{ik} \boldsymbol{D_i'} \boldsymbol{V_i^{-1}} \boldsymbol{D_i} &= \sum_{i=1}^{n} E(x_{ik} \boldsymbol{D_i'} \boldsymbol{V_i^{-1}} \boldsymbol{D_i} | \mathcal{G}_{i-1}) + o(n) \\
&= \sum_{i=1}^{n} E(E(x_{ik} \boldsymbol{D_i'} \boldsymbol{V_i^{-1}} \boldsymbol{D_i} | \mathcal{G}_{i-1}, \boldsymbol{\xi_i})) + o(n) \\
&= \sum_{i=1}^{n} E(\pi_k(\hat{\boldsymbol{\beta}}_{\boldsymbol{i-1}}, \boldsymbol{\xi_1}) \boldsymbol{D'} \boldsymbol{V^{-1}} \boldsymbol{D}) + o(n) \\
&= n\mathbf{B}_n^k + o(n) \\
&= n\mathbf{B}^k + o(n) \qquad a.s
\end{aligned} \tag{3.49}$$

Under the fifth term in Condition 2, matrix $\mathbf{B}_n^k$ and $\mathbf{B}^k$ are nonsingular. Substitute equation (3.49) into (3.44),

$$\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\beta}}^{\boldsymbol{k}} - \boldsymbol{\beta}^{\boldsymbol{k}}) &= (\frac{1}{n} \sum_{i=1}^{n} E(\pi_k(\hat{\boldsymbol{\beta}}_{\boldsymbol{i-1}}^{\boldsymbol{k}}, \boldsymbol{\xi}_{i,1}) \boldsymbol{D_i'} \boldsymbol{V_i^{-1}} \boldsymbol{D_i}))^{-1} \frac{\sum_{i=1}^{n} \boldsymbol{U_i}(\boldsymbol{\beta^k}, \boldsymbol{\alpha})}{n^{\frac{1}{2}}} + o(n^{-\frac{1}{2}}) \\
&= (\mathbf{B_n^k})^{-1} \frac{\sum_{i=1}^{n} \boldsymbol{U_i}(\boldsymbol{\beta^k}, \boldsymbol{\alpha})}{n^{\frac{1}{2}}} + o(n^{-\frac{1}{2}}) \\
&= \frac{1}{\sqrt{n}} (\mathbf{B}^k)^{-1} \sum_{i=1}^{n} x_{ik} \boldsymbol{D_i'} \boldsymbol{V_i^{-1}} \boldsymbol{S_i} + o(n^{-\frac{1}{2}}) \qquad a.s
\end{aligned} \tag{3.50}$$

Let

$$\boldsymbol{Q_{n,k}} = (\mathbf{B}^k)^{-1} \sum_{i=1}^{n} x_{i,k} \boldsymbol{D_i'} \boldsymbol{V_i^{-1}} \boldsymbol{S_i}, \tag{3.51}$$

$$\triangle \boldsymbol{Q_{n,k}} = x_{n,k} (\mathbf{B}^k)^{-1} \mathbf{D}_n' \mathbf{V}_n^{-1} \mathbf{S}_n. \tag{3.52}$$

Then, from equation (3.50)

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{\boldsymbol{k}} - \boldsymbol{\beta}^{\boldsymbol{k}} &= \frac{1}{n}(\mathrm{B}^k)^{-1}\sum_{i=1}^{n} x_{ik}\boldsymbol{D}_{\boldsymbol{i}}'\boldsymbol{V}_{\boldsymbol{i}}^{-\mathbf{1}}\boldsymbol{S}_{\boldsymbol{i}} + o(n^{-\frac{1}{2}}) \\ &= \frac{1}{n}\boldsymbol{Q}_{\boldsymbol{n},\boldsymbol{k}} + o(n^{-\frac{1}{2}}) \qquad\qquad a.s.\end{aligned} \tag{3.53}$$

Therefore,

$$\begin{aligned}E(\triangle\, \boldsymbol{Q}_{\boldsymbol{n},\boldsymbol{k}}|\mathcal{G}_{n-1}) &= E(x_{n,k}(\boldsymbol{B}^{\boldsymbol{k}})^{-\mathbf{1}}\boldsymbol{D}_{\boldsymbol{n}}'\boldsymbol{V}_{\boldsymbol{n}}^{-\mathbf{1}}\boldsymbol{S}_{\boldsymbol{n}}|\mathcal{G}_{n-1}) \\ &= E(E(x_{n,k}(\boldsymbol{B}^{\boldsymbol{k}})^{-\mathbf{1}}\boldsymbol{D}_{\boldsymbol{n}}'\boldsymbol{V}_{\boldsymbol{n}}^{-\mathbf{1}}\boldsymbol{S}_{\boldsymbol{n}}|\mathcal{G}_{n-1},\xi_n)) \\ &= E(\pi_{n,k}E((\boldsymbol{B}^{\boldsymbol{k}})^{-\mathbf{1}}\boldsymbol{D}_{\boldsymbol{n}}'\boldsymbol{V}_{\boldsymbol{n}}^{-\mathbf{1}}\boldsymbol{S}_{\boldsymbol{n}}|\mathcal{G}_{n-1},\xi_n)) \\ &= 0\end{aligned} \tag{3.54}$$

For $m \leq n$, it holds that

$$\begin{aligned}E(\boldsymbol{Q}_{\boldsymbol{n}}|\mathcal{G}_m) &= E(\sum_{i=m+1}^{n} \triangle\, \boldsymbol{Q}_{\boldsymbol{i}}|\mathcal{G}_m) + Q_m \\ &= \boldsymbol{Q}_{\boldsymbol{m}}.\end{aligned} \tag{3.55}$$

That is, $\boldsymbol{Q}_{\boldsymbol{n}}$ is a martingale, and $\{\triangle\, \boldsymbol{Q}_{\boldsymbol{n},\boldsymbol{k}}, \mathcal{G}_n, n \geq 1\}$ is a martingale difference sequence. Under the fifth term in Condition 2,

$$\begin{aligned}\|\triangle\, \boldsymbol{Q}_{\boldsymbol{n},\boldsymbol{k}}\|_2 &= \|x_{i,k}(\boldsymbol{B}^{\boldsymbol{k}})^{-\mathbf{1}}\boldsymbol{D}_{\boldsymbol{n}}'\boldsymbol{V}_{\boldsymbol{n}}^{-\mathbf{1}}\boldsymbol{S}_{\boldsymbol{n}}\|_2 \\ &= \sqrt{E(x_{n,k}^2(\boldsymbol{B}^{\boldsymbol{k}})^{-\mathbf{1}}\boldsymbol{D}_{\boldsymbol{n}}'\boldsymbol{V}_{\boldsymbol{n}}^{-\mathbf{1}}\boldsymbol{S}_{\boldsymbol{n}}\boldsymbol{S}_{\boldsymbol{n}}'\boldsymbol{V}_{\boldsymbol{n}}^{-\mathbf{1}}\boldsymbol{D}_{\boldsymbol{n}}(\boldsymbol{B}^{\boldsymbol{k}})^{-\mathbf{1}})} \\ &\leq \|(\boldsymbol{B}^{\boldsymbol{k}})^{-\mathbf{1}}\boldsymbol{D}_{\boldsymbol{n}}'\boldsymbol{V}_{\boldsymbol{n}}^{-\mathbf{1}}\boldsymbol{S}_{\boldsymbol{n}}\|_2 \\ &< \infty\end{aligned} \tag{3.56}$$

Therefore,

$$\begin{aligned}\|\boldsymbol{Q_{n,k}}\|_2 &= \sqrt{E(\boldsymbol{Q_{n,k}Q'_{n,k}})} \\ &= \sqrt{\sum_{m=1}^{n} E(\triangle\, Q_{m,k} \triangle\, Q'_{m,k})} \\ &= \sqrt{\sum_{m=1}^{n}\sum_{j=1}^{K} E(\triangle\, Q^2_{n,j})} \\ &= O(\sqrt{n})\end{aligned} \tag{3.57}$$

Under the third term in condition 2, $\sum_{i=1}^{n} Var(x_{i,k}\boldsymbol{D'_i V_i^{-1} S_i}|\mathcal{G}_{i-1}) \to \infty$ as $n \to \infty$, according to the iterated law of logarithm for martingales,

$$\boldsymbol{Q_{n,k}} = O(\sqrt{nloglogn}) \qquad a.s \tag{3.58}$$

Therefore, from (3.53)

$$\hat{\boldsymbol{\beta}}^{\boldsymbol{k}} - \boldsymbol{\beta}^{\boldsymbol{k}} = O\left(\sqrt{\frac{loglogn}{n}}\right) \qquad a.s \tag{3.59}$$

Let $I_{n,k} = \sum_{i=1}^{n}(x_{i,k} - E(x_{i,k}|\mathcal{F}_{i-1}))$. For $j \leq n$,

$$
\begin{aligned}
E(I_{n,k}|\mathcal{F}_j) &= E(\sum_{i=1}^{n}(x_{i,k} - E(x_{i,k}|\mathcal{F}_{i-1}))|\mathcal{F}_j) \\
&= \sum_{i=1}^{n} E(x_{i,k}|\mathcal{F}_j) - \sum_{i=1}^{n} E(E(x_{i,k}|\mathcal{F}_{i-1})|\mathcal{F}_j) \\
&= \sum_{i=1}^{j} x_{i,k} + \sum_{i=j+1}^{n} E(x_{i,k}|\mathcal{F}_j) - (\sum_{i=1}^{j} E(x_{i,k}|\mathcal{F}_{i-1}) + \sum_{i=j+1}^{n} E(x_{i,k}|\mathcal{F}_j)) \\
&= \sum_{i=1}^{j}(x_{i,k} - E(x_{i,k}|\mathcal{F}_{i-1})) \\
&= I_{j,k}
\end{aligned}
\tag{3.60}
$$

Then $I_{n,k}$ is a martingale. We use $\triangle\ I_{n,k} = I_{n,k} - I_{n-1,k} = x_{n,k} - E(x_{n,k}|\mathcal{F}_{n-1})$ to denote the martingale difference, and $\boldsymbol{I_n} = (I_{n,1}, \ldots, I_{n,K})$. Since $\| \triangle\ I_{n,k}\| \leqq 1$, $\boldsymbol{I_n} = \sum_{m=1}^{n} \triangle\ \boldsymbol{I_m}$, we have

$$
\begin{aligned}
\|\boldsymbol{I_n}\|_2 &= \sqrt{E(\boldsymbol{I_n}\boldsymbol{I_n'})} \\
&= \sqrt{E((\sum_{m=1}^{n} \triangle\ \mathbf{I}_m)(\sum_{m=1}^{n} \triangle\ \mathbf{I}_m)')} \\
&= \sqrt{E(\sum_{m=1}^{n} \triangle\ \mathbf{I}_m \triangle\ \mathbf{I}_m' + 2\sum_{m=1}^{n}\sum_{j=m}^{n} \triangle\ I_m \triangle\ I_j')} \\
&= \sqrt{\sum_{m=1}^{n} E(\triangle\ \mathbf{I}_m \triangle\ \mathbf{I}_m')} \\
&= \sqrt{\sum_{m=1}^{n}\sum_{j=1}^{K} E(\triangle\ I_{m,k}^2)} \\
&\leq \sqrt{Kn}.
\end{aligned}
\tag{3.61}
$$

Therefore, $\|\boldsymbol{I_n}\|_2 = O(\sqrt{n})$.

According to the law of the iterated logarithm for martingales,

$$I_n = O(\sqrt{nloglogn}) \qquad a.s \tag{3.62}$$

For each $k = 1, 2 \ldots, K$, notice that $x_{i+1,k}$ can be written as

$$x_{i+1,k} = x_{i+1,k} - E(x_{i+1,k}|\mathcal{F}_i) + g_k(\hat{\boldsymbol{\beta}}_{\boldsymbol{i}}), \tag{3.63}$$

Then

$$\begin{aligned} N_n^k &= x_{1,k} + \sum_{i=2}^{n} x_{i,k} \\ &= E(x_{1,k}|\mathcal{F}_0) + (x_{1,k} - E(x_{1,k}|\mathcal{F}_0)) + \sum_{i=1}^{n-1}[x_{i+1,k} - E(x_{i+1,k}|\mathcal{F}_i) + g_k(\hat{\boldsymbol{\beta}}_{\boldsymbol{i}})] \\ &= E(x_{1,k}|\mathcal{F}_0) + \sum_{i=1}^{n}(x_{i,k} - E(x_{i,k}|\mathcal{F}_{i-1})) + \sum_{i=1}^{n-1} g_k(\hat{\boldsymbol{\beta}}_{\boldsymbol{i}}) \\ &= E(x_{1,k}|\mathcal{F}_0) + I_{n,k} + \sum_{i=1}^{n-1} g_k(\hat{\boldsymbol{\beta}}_{\boldsymbol{i}}). \end{aligned} \tag{3.64}$$

Therefore

$$\begin{aligned} N_n^k - n\lambda^k &= I_{n,k} + \sum_{i=1}^{n-1} g_k(\hat{\boldsymbol{\beta}}_{\boldsymbol{i-1}}) - (n-1)g_k(\boldsymbol{\beta}) \\ &= I_{n,k} + \sum_{i=1}^{n-1}(g_k(\hat{\boldsymbol{\beta}}_{\boldsymbol{i-1}}) - g_k(\boldsymbol{\beta})) \\ &= I_{n,k} + \sum_{i=1}^{n-1}[(\hat{\boldsymbol{\beta}}_{\boldsymbol{i}} - \boldsymbol{\beta})(\frac{\partial g_k}{\partial \boldsymbol{\beta}})^T + o(\|\hat{\boldsymbol{\beta}}_{\boldsymbol{i}} - \boldsymbol{\beta}\|^{1+\delta})] \\ &= I_{n,k} + \sum_{i=1}^{n-1}\sum_{j=1}^{K}(\hat{\boldsymbol{\beta}}_{\boldsymbol{i}}^{\boldsymbol{j}} - \boldsymbol{\beta}^{\boldsymbol{j}})^T(\frac{\partial g_k}{\partial \boldsymbol{\beta}^{\boldsymbol{j}}})^T + \sum_{i=1}^{n-1} o(\|\hat{\boldsymbol{\beta}}_{\boldsymbol{i}} - \boldsymbol{\beta}\|^{1+\delta}) \end{aligned} \tag{3.65}$$

Substitute (3.53) into the above equation (3.65),

$$\begin{aligned} N_n^k - n\lambda^k &= I_{n,k} + \sum_{i=1}^{n-1}\sum_{j=1}^{K}(\frac{\boldsymbol{Q}_{i,j}}{i} + o(i^{-\frac{1}{2}}))^T(\frac{\partial g_k}{\partial \boldsymbol{\beta^j}})^T + \sum_{j=1}^{n-1} o(\|\hat{\boldsymbol{\beta}}_{\boldsymbol{j}} - \boldsymbol{\beta}\|^{1+\delta}) \\ &= I_{n,k} + \sum_{i=1}^{n}\sum_{j=1}^{K}(\frac{\boldsymbol{Q}_{i,j}}{i})^T(\frac{\partial g_k}{\partial \boldsymbol{\beta^j}})^T + o(n^{\frac{1}{2}}) \qquad a.s. \end{aligned} \tag{3.66}$$

That is,

$$\begin{aligned} \boldsymbol{N_n} - \boldsymbol{n\Lambda} &= \boldsymbol{I_n} + \sum_{i=1}^{n}\sum_{j=1}^{K}(\frac{\boldsymbol{Q}_{i,j}}{i})^T(\frac{\boldsymbol{\partial g}}{\partial \boldsymbol{\beta^j}})^T + o(n^{\frac{1}{2}}) \\ &= \boldsymbol{I_n} + \sum_{i=1}^{n}\sum_{j=1}^{K}(\triangle\, \boldsymbol{Q}_{i,j})^T(\frac{\boldsymbol{\partial g}}{\partial \boldsymbol{\beta^j}})^T\sum_{m=i}^{n}\frac{1}{m} + o(n^{\frac{1}{2}}) \qquad a.s. \end{aligned} \tag{3.67}$$

From equations (3.67) and (3.62), we have

$$\begin{aligned} \boldsymbol{N_n} - \boldsymbol{n\Lambda} &= O(\sqrt{nloglogn}) + \sum_{m=1}^{n}\sum_{j=1}^{K}\frac{O(\sqrt{mloglogm})}{m} \\ &= O(\sqrt{nloglogn}) \quad a.s \end{aligned} \tag{3.68}$$

Therefore,

$$\frac{\boldsymbol{N_n}}{\boldsymbol{n}} - \boldsymbol{\Lambda} = O(\sqrt{\frac{loglogn}{n}}) \tag{3.69}$$

From equations (3.59) and (3.69), we have proved equation (3.32).

According to condition 1, the continuity of the allocation function π_k, and the equation 3.32, equation 3.31 can be proved.

Next, to prove the asymptotic normality of the estimator of regression parameters and the allocation proportion, we first notice the equation (3.50). According to the central limit theorem, as $n \to \infty$, $n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}^{\boldsymbol{k}} - \boldsymbol{\beta}^{\boldsymbol{k}})$ has an asymptotic Normal

distribution with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{Var^k}$ where

$$\begin{aligned}
\boldsymbol{Var^k} &= \lim_{n\to\infty} Cov(\frac{\sum_{i=1}^{n}(\boldsymbol{B^k})^{-1}\boldsymbol{U_i}(\boldsymbol{\beta^k},\boldsymbol{\alpha})}{\sqrt{n}}) \\
&= \lim_{n\to\infty}\frac{1}{n}Cov(\sum_{i=1}^{n} x_{ik}(\mathrm{B^k})^{-1}\boldsymbol{D_i'V_i(\alpha)^{-1}S_i}) \\
&= \lim_{n\to\infty}\frac{1}{n}\{E[(\sum_{i=1}^{n} x_{ik}(\boldsymbol{B^k})^{-1}\boldsymbol{D_i'V_i(\alpha)^{-1}S_i})(\sum_{i=1}^{n} x_{ik}(\boldsymbol{B^k})^{-1}\boldsymbol{D_i'V_i(\alpha)^{-1}S_i})'] \\
&- E(\sum_{i=1}^{n} x_{ik}(\boldsymbol{B^k})^{-1}\boldsymbol{D_i'V_i(\alpha)^{-1}S_i})E(\sum_{i=1}^{n} x_{ik}(\boldsymbol{B^k})^{-1}\boldsymbol{D_i'V_i(\alpha)^{-1}S_i})'\} \\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} E[E(x_{i,k}^2(\boldsymbol{B^k})^{-1}\boldsymbol{D_i'V_i^{-1}S_iS_i'V_i^{-1}D_i}(\boldsymbol{B^k})^{-1})|\boldsymbol{\mathcal{G}_{i-1},\xi_i}] \\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} E[(\boldsymbol{B^k})^{-1}\boldsymbol{D_i'V_i^{-1}}E(x_{ik}^2|\boldsymbol{\mathcal{G}_{i-1},\xi_i})E(\boldsymbol{S_iS_i'}|\boldsymbol{\mathcal{G}_{i-1},\xi_i})\boldsymbol{V_i^{-1}D_i}(\boldsymbol{B^k})^{-1}] \\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} E[\boldsymbol{\pi_k}(\boldsymbol{\hat{\beta}_{i-1}^k,\xi_i})(\boldsymbol{B^k})^{-1}\boldsymbol{D_i'V_i^{-1}}Var(\boldsymbol{Y_i}|\boldsymbol{\xi_i})\mathrm{V_i^{-1}D_i}(\mathrm{B^k})^{-1}] \\
&= (\boldsymbol{B^k})^{-1}E[\pi_k(\boldsymbol{\beta,\xi_1})(\boldsymbol{D^k})'\boldsymbol{V^{-1}Cov(Y|\xi_1)V^{-1}D^k}](\boldsymbol{B^k})^{-1}
\end{aligned} \tag{3.70}$$

If $\boldsymbol{R_i(\alpha)}$ is indeed the true correlation matrix for the $\boldsymbol{Y_i}'s$, then $\boldsymbol{V_i(\alpha)}$ will be equal to $Cov(\boldsymbol{Y_i}|\boldsymbol{\xi_i})$. Therefore

$$\begin{aligned}
\mathbf{Var^k} &= (\mathrm{B^k})^{-1}E(\pi_k(\boldsymbol{\beta,\xi_1})(\boldsymbol{D^k})'\boldsymbol{V^{-1}D^k})(\mathrm{B^k})^{-1} \\
&= E(\pi_k(\boldsymbol{\beta,\xi_1})(\boldsymbol{D^k})'\boldsymbol{V^{-1}D^k})^{-1}.
\end{aligned} \tag{3.71}$$

Finally, we prove $\sqrt{n}\left(\frac{N_n}{n}-\Lambda\right)\xrightarrow{D} N(\mathbf{0},\boldsymbol{\Sigma})$. Since

$$\begin{aligned}
E[(\triangle\ \boldsymbol{I_n})'\ \triangle\ \boldsymbol{I_n}|\mathcal{G}_{n-1}] &= \boldsymbol{diag}(\boldsymbol{g}(\hat{\boldsymbol{\beta}}_{n-1})) - (\boldsymbol{g}(\hat{\boldsymbol{\beta}}_{n-1})'\boldsymbol{g}(\hat{\boldsymbol{\beta}}_{n-1})) \\
&\to \boldsymbol{diag}(\boldsymbol{\Lambda}) - \boldsymbol{\Lambda'\Lambda}
\end{aligned} \tag{3.72}$$

and for all k, j,

$$
\begin{aligned}
& E(\triangle \boldsymbol{I_{m,k}})' \triangle \boldsymbol{Q_{m,j}}|\boldsymbol{\mathcal{G}_{m-1}}) \\
& = E[(x_{m,k} - E(x_{m,k}|\mathcal{F}_{m-1}))x_{m,j}\boldsymbol{(B^j)^{-1}D'_m V_m^{-1}(Y_m - \mu_m)}|\mathcal{G}_{m-1}] \\
& = \mathbf{0},
\end{aligned}
\tag{3.73}
$$

We have

$$
\begin{aligned}
& E(\triangle \boldsymbol{Q_{n,k}}(\triangle \boldsymbol{Q_{n,k}})^{\boldsymbol{T}}|\boldsymbol{\mathcal{G}_{n-1}}) \\
& = E[\pi_k(\hat{\boldsymbol{\beta}}_{\boldsymbol{n-1}}, \boldsymbol{\xi_{i1}})\boldsymbol{(B^k)^{-1}D'_n V_n^{-1}}] \\
& \to E[\pi_k(\boldsymbol{\beta}, \boldsymbol{\xi_1})\boldsymbol{(B^k)^{-1}(D^k)'V^{-1}Cov(Y|\xi)V^{-1}D^k(B^k)^{-1}}]
\end{aligned}
\tag{3.74}
$$

and

$$
E(\triangle \boldsymbol{Q_{m,j}} \triangle \boldsymbol{Q_{m,k}}|\boldsymbol{\mathcal{G}_{m-1}}) = 0. \tag{3.75}
$$

Therefore,

$$
\begin{aligned}
& Var(\sum_{i=1}^{n}\sum_{j=1}^{K}(\triangle \boldsymbol{Q_{i,j}})(\frac{\boldsymbol{\partial g}}{\partial \boldsymbol{\beta^j}})^T \sum_{m=i}^{n}\frac{1}{m}) \\
& = \textstyle\sum_{i=1}^{n-1}\sum_{l=1}^{n}\sum_{j=1}^{K}\frac{l\wedge i}{li}(\frac{\boldsymbol{\partial g}}{\partial \boldsymbol{\beta^j}})Var(\boldsymbol{(B^j)^{-1}U_i(\beta^j, \alpha)})(\frac{\boldsymbol{\partial g}}{\partial \boldsymbol{\beta^j}})^T \\
& \to 2n(\textstyle\sum_{j=1}^{K}(\frac{\boldsymbol{\partial g}}{\partial \boldsymbol{\beta^j}})Var^j(\frac{\boldsymbol{\partial g}}{\partial \boldsymbol{\beta^j}})^T))
\end{aligned}
\tag{3.76}
$$

$$
\begin{aligned}
Var(\boldsymbol{I_n}) &= \sum_{i=1}^{n}\boldsymbol{diag(g(\hat{\beta}_i))} - \boldsymbol{g(\hat{\beta}_i)'g(\hat{\beta}_i)} \\
&\to n(\boldsymbol{diag(\Lambda)} - \boldsymbol{\Lambda'\Lambda})
\end{aligned}
\tag{3.77}
$$

Let $\Sigma = \boldsymbol{diag(\Lambda)} - \boldsymbol{\Lambda'\Lambda} + 2\sum_{k=1}^{K}(\frac{\boldsymbol{\partial g}}{\partial \boldsymbol{\beta^k}})Var^k(\frac{\boldsymbol{\partial g}}{\partial \boldsymbol{\beta^k}})^T$, then

$$
\begin{aligned}
Var(\boldsymbol{N_n} - \boldsymbol{n\Lambda}) &= n(\boldsymbol{diag(\Lambda)} - \boldsymbol{\Lambda'\Lambda} + 2\sum_{k=1}^{K}(\frac{\boldsymbol{\partial g}}{\partial \boldsymbol{\beta^k}})Var^k(\frac{\boldsymbol{\partial g}}{\partial \boldsymbol{\beta^k}})^T) + O(n) \\
&= n\Sigma + O(n)
\end{aligned}
\tag{3.78}
$$

Since $\boldsymbol{N_n - n\nu}$ are linear combinations of zero mean martingales with bounded variance, by the central limit theorem for martingales, we have

$$\sqrt{n}(\frac{\boldsymbol{N_n}}{\boldsymbol{n}} - \boldsymbol{\Lambda}) = \boldsymbol{n^{-\frac{1}{2}}(N_n - n\Lambda)} \xrightarrow{\mathcal{D}} \boldsymbol{N(0, \Sigma)}. \tag{3.79}$$

Equation (3.33) thus is proved. □

3.4 Comparison Between the Extended General CARA and the General CARA Design

We utilized a widly applicable estimation approach to estimate the longitudinal data accrued in clinical studies and extended the general CARA design (L-X Zhang et al., 2007) to adapt to the trials with longitudinal responses. Because of the longitudinal setting, both the conduct of the design and the analysis of the accrued data are different from the original design.

Firstly, our extended design becomes much more complicated due to the repeated examined data. When a new patient enters the trial, the complete information of the previous $T-1$ patients have not been obtained yet, and they have to be observed and recorded at the same time as the new patient being observed. In addition, the probability of a new patient being assigned to each treatment depends on the estimated value of the parameter $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}^{\boldsymbol{1\prime}}, \ldots, \hat{\boldsymbol{\beta}}^{\boldsymbol{K\prime}})$. In the CARA design with one time response, only one parameter among the K of them is updated from the information of previously treated patients. While in design with longitudinal responses, there can be more than one parameter updated from the information of the previous several patients whose responses and covariates are measured repeatedly. With re-

peated measurements, more information are available for the adaptation of treatment allocations.

Secondly, the result of the general CARA design can only be applied to clinical studies with one-time responses. This thesis extended the result of Zhang et al. to the longitudinal clinical trials. After careful study of some widely applicable methods for the longitudinal data analysis, we choose the generalized estimating equations (Liang and Zeger, 1986) to analyze the longitudinal data accrued in the clinical trials, and successfully derived the asymptotic properties for our design. When there are only one response from each patient,

$$\mathbf{B}^k = E(\pi_k(\boldsymbol{\beta}, \boldsymbol{\xi_1})(\boldsymbol{D^k})'\boldsymbol{V^{-1}}\boldsymbol{D^k}) \tag{3.80}$$

will reduce to

$$\mathbf{B}^k = E(\pi_k(\boldsymbol{\beta}, \boldsymbol{\xi_1})\boldsymbol{a''}(\boldsymbol{\theta^k})\boldsymbol{\xi}^T\boldsymbol{\xi}) \tag{3.81}$$

and

$$\boldsymbol{Var}^k = (\boldsymbol{B}^k)^{-1}E[\pi_k(\boldsymbol{\beta}, \boldsymbol{\xi_1})(\boldsymbol{D^k})'\boldsymbol{V^{-1}}\boldsymbol{Cov}(\boldsymbol{Y}|\boldsymbol{\xi_1})\boldsymbol{V^{-1}}\boldsymbol{D^k}](\boldsymbol{B}^k)^{-1} \tag{3.82}$$

will reduce to

$$\boldsymbol{Var}^k = (E(\pi_k(\boldsymbol{\beta}, \boldsymbol{\xi_1})\boldsymbol{a''}(\boldsymbol{\theta^k})\boldsymbol{\xi}^T\boldsymbol{\xi}))^{-1} \tag{3.83}$$

which is the case when the generalized linear models are used in the sequential estimation of Zhang et al.'s general CARA design.

3.5 Comparison Between the Extended General CARA and the other Designs with Longitudinal Responses

As discussed in section 1.1.4, the CALRA design (Biswas et al., 2010) is the only optimal design, to our knowledge, for longitudinal clinical trials that consider the covariates of both the previous patients and the current patient. Compared to the CALRA design, our extended general CARA design can target any desired allocation proportion whereas the CALRA design can only target a certain optimal allocation proportion decided by the utility function they defined. In addition, the extended CARA design can be applied to clinical trials with more than two competing treatments and any responses which follow a distribution in the exponential family. To the contrast, the CALRA design is only suitable for clinical trials with two competing treatments and binary responses.

Chapter 4

Conclusions and Future Work

We have conducted a review on the development of response-adaptive randomization, and have discussed the necessity of response-adaptive design with longitudinal responses. The contribution of this research is that we proposed a general CARA design for longitudinal clinical trials which extends the framework of the CARA design by Zhang et al.(2007). Our design considers the covariate information of both the previous patients in the study and the current patient who is ready for treatment assignment. The design is also the first response-adaptive designs that can be applied to longitudinal clinical trials with more than two competing treatments, and can target different desired allocation proportions. We have also explored the asymptotic properties of estimators of parameters and proved the asymptotic normality of the regression parameters and the allocation proportion. The explicit form of the variability of the allocation proportion is also obtained.

There are still problems left for further research. The allocation function which is directly related to the variability of the allocation proportion only has a general form

in our design. More studies are needed to find allocation functions which will result in smaller variability for the design, or to look into optimality criteria for an optimal allocation function that gives the smallest variability. Also, the design we proposed requires that the responses of each patient are complete or missing completely at random (MCAR) (Cornfield, 1959). The collecting time point for longitudinal responses are assumed to be equally spaced as well. However, in practice, often some patients drop off the trial before their complete responses are collected, or there may be delayed responses from some patients, or the repeated measurement of each patient are unrealistic to be equally spaced. Little is known about response-adaptive randomization with longitudinal responses when missing data with certain pattern present, and when the repeated measurement are not equally spaced. Better methods in analyzing longitudinal data collected in clinical trials are needed. These will be the direction of future research.

Bibliography

Bibliography

[1] Athreya k B, Karlin S. Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. Annals of Mathematical Statistical Statistics (1968), Vol 39 pp 1801-1817.

[2] Albert P S. Longitudianl data analysis (repeated measures) in clinical trials. Statistics in Medicine. (1999), 18:1707-1732.

[3] Brian S. Everitt and Andrew Pickles. Statistical aspects of the design and analysis of clinical trials. World Scientific Pub Co Inc; Revised edition (March 30, 2004).

[4] Biswas A, Dewanji A. Inference for a RPW-Type clinical trial with repeated monitoring for the treatment of rheumatoid arthritis. Biometrical Journal (2004), Vol 46, pp769-779.

[5] Biswas A, Dewanji A. A randomized longitudianl play-the-winner design for repeated binary data. Australian and New Zealand Journal of Statistics.(2004), Vol.46, pp 675-684.

[6] Biswas A, Park E, Bhattacharya R. Covariate-adjusted response-adaptive designs for longitudinal treatment responses: PEMF trial revisited. Statistical Methods in Medical Research 2010 0(0) 1-14

[7] Bandyopadhyay U, Biswas A and Bhattacharya R. A covariate adjusted two-stage allocation design for binary responses in randomized clinical trials. Statistics in Medicine (2007). 26: 4386-4399.

[8] Blackwell D, Hodges J L. Design for the control of selection bias. Annals of Mathematical Statistics (1957), Vol. 28 pp 449-460.

[9] Cornfield J. Principles of research. American Journal of Mental Deficiency, Vol. 64, (1959), 240-252.

[10] Crowder M. On the use of a working correlation matrix in using generalized linear models for repeated measures. Biometrika (1995) 82, 407-410.

[11] Carey, V., Zeger, S.L., and Diggle, P. Modelling multivariate binary data with alternating logistic regressions. Biometrika (1993), 517-526

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

[13] Eisele J R. The doubly adaptive biased coin design for sequential clinical trials. Journal of Statistical Planning and Inference (1994), Vol. 38, 249-262.

[14] Eisele J R, Woodroofe M. Central limit theorems for doubly adaptive biased coin designs. Annals of Statistics (1995), Vol. 23, 234-254.

[15] Fisher E. On the law of the iterated logarithm for martingales. The Annals of Probability, Vol.20, No. 2(April, 1992), 675-680

[16] Hu F, Zhang L X. Asymptotic properties of doubly adaptive biased coin designs for multi-treatment clinical trials. Annals of Statistics (2004), Vol. 32, 268-301.

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

[18] Hu F, Rosenberger W F. The Theory of response-adaptive randomization in clinical trials. New York: Wiley.(2006)

[19] Ivanova A. A play-the-winner-type urn design with reduced variability. Metrika (2003) 58:1-13

[20] Ivanova A, Rosenberger W F, Durham S D, Flournoy N. A birth and death urn for randomized clinical trials: asymptotic methods. The Indian Journal of Statistics, Special Issue on Biostatistics (2000), Volume 62, Series B, Pt. 1, 104-118

[21] Johnson R E., Jale J H. and Fudala P J. A controlled trial of buprenorphine treatment for opiate dependence, Journal of the American Medical Association (1992), 267, 2750-2755.

[22] Liang Y, Carriere K C. Multiple-objective response-adaptive repeated measurement designs for clinical trials. Journal of Statistical Planning and Inference, 139(2009) 1134-1145

[23] Liang K Y, Zeger S L. Longitudianl data analysis using generalized linear models, Biometrika(1986) . 73.1. 13-22

[24] Lachin J M. Properties of simple randomization in clinical trials. Controlled Clinical Trials (1988), Vol 9, 312-326.

[25] Legedza A T R, ScD, Ibrahim J G. Longitudianl design for phase I clinical trails using the continual reassessment method. Controlled Clinical Trials (2000) 21:574-588

[26] Miklos Csorgo. On the strong law of large numbers and the central limit theorem for martingales (1967). http://www.ams.org/journals/tran/1968-131-01/s0002-9947-1968-0221562-x.pdf.

[27] McCullagh P. Quasi-Likelihood functions. The Annals of Statistics. Vol. 11, No. 1(Mar., 1983), 59-67

[28] Melfi V F., Page Connie. Variability in adaptive designs for estimation of success probabilities. In New Developments and Applications in Experimental Design (Flournoy N., Rosenberger W F., and Wong W K.).(1995) Institute of Mathematical Statistics, Hayward 106-114.

[29] Melfi V F., Page Connie., Geraldes M. An adaptive randomized design with application to estimation. Canadian Journal of Statistics. (2001) Vol. 29. 107-116.

[30] Melfi V F., Page Connie. Estimation after adaptive allocation. Journal of Statistical Planning and Inference, 87 (2000) 353-363

[31] Molenberghs G., and Verbeke G. (2005) Models for discrete longitudinal data. New York: Springer.

[32] Robbins H. (1952). Some aspects of the sequential design of experiments. Bulletin of the American Mathematics Society 58, 527-535.

[33] Rosenberger, W. F. and Lachin, J. L. Randomization in clinical trials: Theory and Practice. New York: Wiley (2002).

[34] Rosenberger WF, Vidyashankar AN and Agarwal DK. Covariate-adjusted response-adaptive designs for binary response. Journal of Biopharmaceutical Statistics (2001). 11: 227-236.

[35] Sutradhar B C, and Das K. On the efficiency of regression estimators in generalized linear models for longitudinal data. Biometrika (1999) 86, 459-465.

[36] Sutradhar B C. Inferences in generalized linear longitudinal mixed models, The Canadian Journal of Statistics Vol.38, No. 2,2010, Pages 174-196

[37] Sutradhar B C, Jowaheer V. Analyzing longitudinal count data from adaptive clinical trials: A weighted generalized quasi-likelihood approach. Journal of Statistical Computation and Simulation, Vol. 76, No.12 (Dec., 2006), 1079-1093

[38] Sutradhar B C, Biswas A, Bari W. Marginal regression for binary longitudinal data in adaptive clinical trials, Scandinavian Journal of Statistics (2005), Vol. 32 93-113.

[39] Sun R, Cheung S H, Zhang L X. A generalized drop-the-loser rule for multi-treatment clinical trials. Journal of Statistical Planning and Inference, 137(2007) 2011-2023

[40] Schwartz, D. and Lellouch, J. (1967). Explanatory and pragmatic attitudes in therapeutic trials, J. Chron. Disorders 20: 637?648.

[41] Wedderburn R.W.M. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. Biometrika(1974),61,3,439

[42] Wei L J, Smythe R T, Lin D Y, Park T S. Statistical inference with data-dependent treatment allocation rules. Journal of the American Statistical Association, Vol. 85, No.409(Mar., 1990), 156-162

[43] Wei L J, Durham S. The Randomized play-the-winner rule in medical trials. Journal of the American Statistical Association, Vol. 73, No.364 (Dec., 1978), 840-843

[44] Wald A. differentiation under the expectation sign in the fundamental identity of sequential analysis. The Annals of Mathematical Statistics, Vol. 17, No.4(Dec., 1946), 493-497

[45] William F. Rosenberger, John M. Lachin. Randomization in clinical trials: Theory and Practice. Wiley, New York, 2002.

[46] Wei L J., Durham S D.(1978). The randomized play-the-winner rule in medical trials. Journal of the American Statistical Association (1978), Vol 73, pp840-843.

[47] Wei L J. The generalized polya's urn design for sequential medical trials. Annals of Statistics.(1979) 7, 291-196.

[48] Zhang L X, Hu F, Cheung S H, et al. Asymptotic properties of covariate-adjusted adaptive designs, Annals of Statistics, 2007, 35:1166-1182

[49] Zhang L X, Hu F. A new family of covariate-adjusted response adaptive designs and their properties, Applied Mathematical Journal of Chinese University 2009, 24(1) 1-13

[50] Zeger S L, Kung-Yee Liang, Albert P S. Models for longitudianl data: A generalized estimating equation approach. Biometrics, Vol. 44, No. 4(Dec. 1988), 1049-1060

[51] Zeger S L, Kung-Yee Liang, Longitudinal data analysis for discrete and continuous outcomes. Biometrics (1986), 121-130.

[52] Zhao, L. P. and Prentice, R. L. Correlated binary regression using a quadratic exponential model. Biometrika (1990) 77, 642-648.

[53] Zelen, M. (1969). Play the winner rule and the controlled clinical trial. Journal of the American Statistical Association. 64 131-146.

Chapter 5

Appendix

5.1 Martingale Theory

This section introduces the definition of martingales and some of its useful properties.

5.1.1 General Definition and Properties of Martingales

Definition 1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space: Ω is a set, $\mathcal{F}$ a σ-field of subsets of Ω, and P a probability measure defined on $\mathcal{F}$. Let I be any interval of the form $(a,b), [a,b), (a,b]$ or $[a,b]$ of the ordered set $\{-\infty, \ldots, -1, 0, 1, \ldots, \infty\}$. Let $\{\mathcal{F}_n, n \in I\}$ be an increasing sequence of σ-fields of $\mathcal{F}$ sets. Suppose that $\{Z_n, n \in I\}$ is a sequence of random variables on Ω satisfying*

1. *Z_n is measurable with respect to $\mathcal{F}_n$*
2. *$E|Z_n| < \infty$*
3. *$E(Z_n|\mathcal{F}_m) = Z_m$ a.s for all $m < n$, $m, n \in I$*

Then, the sequence $\{Z_n, n \in I\}$ *is said to be a martingale with respect to* $\{\mathcal{F}_n, n \in I\}$. *We write that* $\{Z_n, \mathcal{F}_n, n \in I\}$ *is a martingale.*

Definition 2. *A sequence* $(Y_n, \mathcal{G}_n, n \geqslant 1)$ *defined on* $(\Omega, \mathcal{F}, P)$ *is called a martignale difference sequence if* $\mathcal{G}_n \subset \mathcal{F}$ *are increasing* σ*-fields with* Y_n *being* $\mathcal{G}_n$ *measurable and* $E(Y_n|\mathcal{G}_n) = 0$ *for all* $n \geqslant 2$.

5.1.2 Useful Theorems of Martingales

Theorem 2. *(Law of Large Numbers for Martingales):*
Let $\{\triangle\ \boldsymbol{S_i}\}$ *denotes a margtingale difference sequence, then* $\boldsymbol{S_n} = \sum_{i=1}^{n} \triangle\ \boldsymbol{S_i}$ *will be a margtingale. If*

$$\sum_{i=1}^{\infty} \boldsymbol{i^{-2}} \boldsymbol{E}(\triangle\ \boldsymbol{S_i}\ \triangle\ \boldsymbol{S_i'}) < \boldsymbol{\infty}, \tag{5.1}$$

then

$$\frac{1}{n}\boldsymbol{S_n} \to 0 \quad a.s. \tag{5.2}$$

For each $j = 1, \ldots, n$, let S_{nk} be a martingale with respect to nested sigma-algebras $\mathcal{F}_{nk}$. Let $X_{nk} = S_{nk} - S_{n,k-1}$, $S_{n0} = 0$, denote the martignale differences. Then $\{S_{nk}, \mathcal{F}_{nk}\}$, for $k = 1, \ldots, n$, $n \geq 1$ is a double sequence of triangular arrays, called a martingale array (Hall and Heyde(1980)). The central limit theorem for martingale arrays $\{S_{nk}, \mathcal{F}_{nk}\}$ states as below.

Theorem 3. *(Billingsley's Central Limit Theorem):*
Let $S_n = \sum_{j=1}^{n} X_j$ *be a zero mean martingale with respect to* $\mathcal{F}_n$. *Assume the following two conditions:*

$$\varinjlim n^{-1} \sum_{j=1}^{n} E(X_j^2|\mathcal{F}_{j-1}) = \eta^2 \tag{5.3}$$

and

$$\lim_{\longrightarrow} n^{-1-\delta/2} \sum_{j=1}^{n} E(|X_j|^{2+\delta} | \mathcal{F}_{j-1}) = 0 \quad a.s \tag{5.4}$$

for positive constants δ *and* η^2. *Then*

$$n^{-\frac{1}{2}} \sum_{j=1}^{n} X_j \xrightarrow{\mathcal{D}} N(0, \eta^2) \quad as \quad n \to \infty \tag{5.5}$$

Theorem 4. *(Law of the Iterated Logarithm for Martingales):*

Let $(Z_n, \mathcal{F}_n, n \geqslant 1)$ *be a martingale defined on a probability space* $(\Omega, \mathcal{F}, P)$ *with* $E(Z_1) = 0$. *Let* $Y_n = Z_n - Z_{n-1}$, *for* $n \geqslant 1$, $Z_0 = 0$, $\mathcal{F}_0 = (\phi, \Omega)$, $s_n^2 = \sum_{i=1}^{n} E(Y_i^2 | \mathcal{F}_{i-1})$, *and* $u_n = \sqrt{2loglogs_n^2}$. *If* $s_n^2 \to \infty$ *and*

$$|Y_n| \leq K_n s_n / u_n \quad for \ \ n \geq 1 \tag{5.6}$$

where K_n *are* $\mathcal{F}_{n-1}$ *measurable and* $K_n \to 0$, *then* $\limsup Z_n/(s_n u_n) = 1$

If $|Y_n| \leq K$ for some constant K and $s_n^2 \to \infty$, then the law of the iterated logarithm holds easily by setting $K_n = K u_n / s_n$.