ANALYZING BINARY LONGITUDINAL DATA
IN ADAPTIVE CLINICAL TRIALS

WASIMUL BARI

Analyzing Binary Longitudinal Data in Adaptive Clinical Trials

by

A thesis submitted to the School of Graduate Studies
in partial fulfillment of the requirement for the Degree of
Master of Science in Statistics

Department of Mathematics and Statistics
Memorial University of Newfoundland

July, 2003

St. John's   **Newfoundland**   **Canada**

Abstract

In an adaptive clinical trial research, it is common to use certain data dependent design weights to assign individuals to treatments so that more study subjects are assigned to the better treatment. These design weights must also be used for consistent estimation of the treatment effects as well as the effects of other prognostic factors. In practice, there are however situations where it may be necessary to collect binary responses repeatedly from an individual over a period of time and to obtain consistent and efficient estimates for the treatment effects as well as the effects of the other covariates. In this thesis, we introduce a binary response based longitudinal adaptive design for the allocation of individuals to a better treatment, and propose a weighted generalized quasi-likelihood (WGQL) approach for the consistent and efficient estimation of the regression parameters, including the treatment effects. We also introduce a binary longitudinal adaptive mixed model assuming that given the treatment effects and the unobservable individual random effect, repeated responses of an individual are longitudinally correlated. An extended WGQL approach is also used to obtain consistent and efficient estimators for the regression parameters and the variance component of individual random effects.

Acknowledgments

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation of the Problem

A clinical trial is basically an experimental design to evaluate the effects of a new medical treatment or intervention. For clinical trials on human beings, it is highly desirable that one uses certain data-dependent treatment allocation rules which exploit accumulating information to assign individuals to treatments so that more study subjects are assigned to the better treatment. More clearly, even though the main objective of the clinical experiment is to identify a better treatment, the experiment is designed in such a way so that more study subjects are likely to be assigned to the better treatment during the process for ethical reasons (cf. Armitage (1975), Anscombe (1963), Colton (1963), and Cornfield, Halprin, and Greenhouse (1969)). This type of experimental design may be referred to as a sequential adaptive design, where the allocation of treatment to an incoming subject is defined by what has already been learned at earlier stages. There exists a vast literature on the development of such sequential adaptive designs. For example, Zelen (1969) introduced the play-the-winner(PW) rule, which prescribes that a success with a given treatment generates a future trial with the same treatment, while a failure generates a trial with the alternative treatment. If the patients respond to the treatments without much delay, the PW rule specifies that after each success we continue to use the

same treatment and after each failure we switch to the other treatment. Zelen (1969) called this rule the modified play-the-winner (MPW) rule. The MPW rule tends to assign more patients to the better treatment, but it is deterministic and may bias the trial in various ways. The MPW rule maximizes the selection biases because the experimenter knows the next assignment for sure. Simon, Weiss, and Hoel (1975) proposed a non-deterministic plan based on the likelihood function, but it is rather complicated for practical use.

As a modification of Zelen's (1969) PW rule, Wei and Durham (1978) and Wei (1979) proposed a randomized play-the-winner (RPW) rule (see Chapter 2 for details) which has advantages that it is not deterministic and is less vulnerable to experimental bias, and it is also easily implemented in a real trial. One disadvantage of the RPW scheme is that it does not include the balance over the covariates or prognostic factors which may affect the response of the patient to the treatment. In conducting a clinical trial, it is desirable that the trial should be balanced, not only with respect to the overall assignment of patients to treatments but also with respect to the various prognostic factors, such as age, sex, and major indicators of clinical condition. Biased coin schemes which do force balance over both treatments and prognostic factors are given by Pocock and Simon (1975) and Efron (1980). The properties of the design have been elucidated by numerical studies and they are now being increasingly used in clinical trials. However, the designs suffer from the disadvantage that they rely on arbitrary functions to achieve the desired balance. The procedures thus lack a firm theoretical framework. Begg and Igelewicz (1980) introduced an alternative approach in the presence of the prognostic factors, which uses optimum design theory to suggest a deterministic design criterion, which is then modified for computational convenience. Atkinson (1982) suggested an optimum design theory to provide a procedure of the biased coin type for an arbitrary number of treatments in the presence and absence of prognostic factors. This has the theoretical advantage of obviating dependence on a series of arbitrary functions. Smith (1984) showed how martingle methods may be used to study the procedures of Wei (1978) and a number of generalizations of these

methods.

Later on, by using different inference procedures, many authors, for example, Wei, Smythe, Lin, and Park (1990) and Smythe and Rosenberger (1995) studied the effects of the treatment on the binary responses, where the treatment is assigned by using the RPW rule based adaptive designs on the binary responses. Recently, some authors have modified the RPW rule to accommodate the contributions of other possible ordinal covariates (prognostic factors) in constructing the adaptive designs for the assignment of the treatment to the incoming patients and for the examination of the treatment effect on discrete or continuous responses. For example, Bandyopadhyay and Biswas (1999) and Bandyopadhyay and Biswas (2001) have accommodated the suitable prognostic factors in constructing the adaptive designs for binary and normal responses, respectively. Remark that the construction of the adaptive designs in all of the above works is however confined to the non-longitudinal (cross-sectional) set up. That is, once the treatment was assigned to an individual, the individual was expected to provide only one response.

In practice, there are however clinical trial experiments where it may be useful to register the study subjects sequentially over time as in the above studies, but collect repeated binary responses from each study subject. While the responses under an adaptive design may be normal, binary, or count, in the present thesis, we consider the adaptive design for longitudinal binary responses because of its wide range of applications. For example, in a psychological study, once an individual enters into the study sequentially, the individual may be asked to report daily over a period of 7 days on the presence of 'anxiety'. Here the 'yes' or 'no' status of 'anxiety' of an individual on a given day is a binary response. To address such problems, in this thesis, our motivation is to construct a longitudinal adaptive design by using the available repeated binary responses and covariate information such as age and education level, for the purpose of assigning more study subjects to a better treatment. Here it is also of interest to compute the treatment effect as well as the effects of the other covariates based on all covariate information and the responses available at the

end of the study. We do this in the spirit of Liang and Zeger (1986), Sutradhar and Das (1999) and Jowaheer and Sutradhar (2002), where they proposed generalized estimating equations approach to obtain the consistent and efficient estimators of the regression parameters in the class of generalized linear models under a classical longitudinal set up. Furthermore, the longitudinal correlation computed from the repeated binary data will be utilized to obtain consistent as well as efficient estimates of regression effects.

1.2   Objective of the Thesis

As the analysis of binary longitudinal data in adaptive clinical trials is not adequately addressed in the literature, in this thesis, we propose a simple longitudinal adaptive design so that more study subjects may be assigned to the better treatment. The construction of such a longitudinal adaptive binary design may be considered as an extension of the existing adaptive design in the non-longitudinal set up constructed on the idea of the RPW rule. The existing adaptive designs in the non-longitudinal (cross-sectional) set up are discussed in brief in Chapter 2.

In Chapter 3, we deal with the longitudinal adaptive clinical trial studies. More specifically, in Chapter 3, we construct the so-called longitudinal adaptive designs so that more study subjects are assigned to the better treatment. The performance of this type of design is examined through a simulation study. To accommodate the longitudinal correlations, we follow Sutradhar and Das (1999)(see also Jowaheer and Sutradhar (2002)) and use a general auto-correlation structure for the repeated binary responses and take these correlations as well as the longitudinal adaptive design weights into account for consistent and efficient estimation of the regression parameters of the models through a weighted generalized quasi-likelihood(WGQL) approach. We also examine the performance of the WGQL estimation approach, mainly for the estimation of the treatment effect. In the same chapter, the misspecification effects of the longitudinal adaptive designs are also examined through a simulation study.

Note that it has been assumed in Chapter 3 that the repeated responses of an individual patient are likely to be correlated following a general autocorrelation structure. But, as there may be situations in practice where repeated responses from an individual may also be affected by an individual random effect, in Chapter 4 we consider a more general correlation structure which is constructed by accommodating the individual random effect as well as the repetition of the data. The adaptive weights are constructed accordingly. Next we discuss the WGQL approach to estimate the dispersion parameter of the individual random effect and the regression parameters involved in the mixed model, consistently and efficiently. Similar to Chapter 3, the performance of the WGQL approach is also studied by a simulation. Longitudinal correlations are still estimated by the moment method as in Chapter 3.

We conclude the thesis in Chapter 5 with some remarks on the usefulness of the longitudinal adaptive designs that we constructed in Chapters 3 and 4. In the same chapter, we also provide some remarks on the possibilities of future research in this area.

Chapter 2

On Existing Adaptive Clinical Trials in Cross-Sectional Set up

Adaptive clinical trials have attracted the attention of practitioners mainly because of ethical reasons, as this type of design allows the incoming patients to receive the better treatment with greater likelihood. More specifically, in any sequential medical experiment on a cohort of human beings, there is an ethical imperative to provide the best possible medical care for individual patient. This ethical imperative may not be satisfied if a 50-50 randomized allocation scheme is used. Adaptive designs have long been proposed to remedy this situation.

Adaptive designs that have been developed over the past few decades are based on single response of treatment from each study subject. These responses may be either discrete or continuous. Moreover, covariates or prognostic factors that may influence the response of an individual are not considered in all adaptive clinical trials so far developed. In the following sections, we briefly describe the existing adaptive clinical trials with single discrete or continuous response from an individual patient in the presence or absence of prognostic factors.

2.1 Adaptive Clinical Trials with Discrete Response Variables

In many clinical trials, discrete responses are collected from the patients involved in the experiment. Consequently, most of the available works on adaptive clinical trials are based on discrete responses. This type of adaptive design can be described through the generalized Pólya urn (GPU). The play-the-winner (PW) rule for dichotomous responses in clinical trials introduced by Zelen (1969) is a special case of the GPU. Later on, as a modification of Zelen's rule, Wei and Durham (1978) and Wei (1979) proposed the idea of a randomized play-the-winner (RPW) rule. Further works in this direction are due to Wei, Smythe, and Mehta (1989), Wei (1988), and Begg (1990). In all the above works and almost all other works available in the literature on clinical trials, it is assumed that the entering patients are homogeneous. But, in practice, there may be many prognostic factors like age, sex, blood pressure, heart beat, blood sugar etc. which usually make the patients involved heterogeneous. The construction of an adaptive design in the presence of prognostic factors requires much more attention as compared to the construction of the adaptive clinical design in the absence of prognostic factors. The treatment allocation problem in the presence of prognostic factors was considered by Begg and Iglewicz (1980) and Atkinson (1982), among others. Begg and Iglewicz (1980) proposed a treatment allocation procedure based on the minimization of a function that is an easily evaluated approximation to the variance of the treatment effect in a linear model relating the outcome variable to the chosen prognostic factors and selected interactions, whereas Atkinson (1982) used optimum design theory to provide a procedure of the biased coin type for an arbitrary number of treatments in the presence of prognostic factors. Bandyopadhyay and Biswas (1999) proposed an adaptive RPW (ARPW) rule to incorporate the presence of prognostic factors. They considered both the cases where prognostic factors are non-stochastic and stochastic. In the following subsection, we describe some of the existing important adaptive designs for discrete response variables in the absence of

prognostic factors. In the next subsection, we describe similar existing designs in the presence of prognostic factors.

2.1.1 Adaptive Clinical Trials with Discrete Response Variables in the Absence of Prognostic Factors

2.1.1.1 Generalized Pólya Urn (GPU)

The randomized adaptive designs are constructed to assign the incoming patients to a better treatment with greater likelihood. A large family of randomized adaptive designs is developed from the generalized Pólya urn (GPU) model (originally designated by Athreya and Karlin, 1968, as the "generalized Friedman's urn"). For simplicity, consider a clinical experiment with two treatments. To develop the adaptive design for this experiment, Athreya and Karlin (1968) considered a GPU model which can be described as follows. Suppose that an urn contains two types of particles with initial number as $n_0 = (n_{01}, n_{02})$ particles where n_{0m} denotes the initial number of particles of type m, $m = 1, 2$. A particle is drawn or split at random from the urn. Its type is observed and the particle is put back to the urn. When a particle of type m ($m = 1, 2$) is drawn, it is said that a type m split occurs. Following the type m split, suppose that $R_{mm'}$ particles of type m', for $m' = 1, 2$, are added to the urn, or generated. Thus, if type m is drawn at the first draw (i=1), then $n_1 = (n_{11}, n_{12})$ would represent the composition of the urn after the first draw. Here, $n_{1m'} = n_{0m'} + R_{mm'}$, for $m' = 1, 2$. In the most general sense, $R_{mm'}$ can be random and can be some function of the responses of patients. A particle must always be generated at each stage, in addition to the replacement so that $\Pr\{R_{mm'} = 0, m' = 1, 2\}$ is assumed to be zero.

Allocation Probability

Let us define the indicator variable δ_r for the rth draw as follows:

$$\delta_r = \begin{cases} 1, & \text{if type 1 particle is drawn} \\ \\ 0, & \text{if type 2 particle is drawn} \end{cases}$$

After i splits and generations, the urn composition is given by the vector $n_i = (n_{i1}, n_{i2})$ where n_{im}, for $m = 1, 2$, represents the number of particles in the urn of type m after i splits given by

$$n_{im} = n_{0m} + \sum_{r=1}^{i}[\delta_r R_{1m} + (1 - \delta_r)R_{2m}].$$

Thus the proportion of type m splits after K splits is $\frac{n_{Km}}{\sum_{m=1}^{2} n_{Km}}$. For given all the previous draws $(\delta_1, \ldots, \delta_K)$, let w_{K+1} be the conditional probability that type 1 particle is drawn at the $(K + 1)st$ draw. Then

$$w_{K+1} = \frac{n_{K1}}{\sum_{m=1}^{2} n_{Km}}. \tag{2.1}$$

Note that in terms of assigning an incoming patient, this w_{K+1} would represent the probability of assigning the $(K + 1)st$ patient to the type 1 treatment, whereas $(1 - w_{K+1})$ would represent the probability of assigning this patient to the type 2 treatment.

Athreya and Karlin (1968) showed that

$$\frac{n_{Km}}{\sum_{m=1}^{2} n_{Km}} \to \nu_m, \tag{2.2}$$

almost surely as $K \to \infty$, where ν_m is the mth $(m = 1, 2)$ element of the left eigen vector $\nu = (\nu_1, \nu_2)$ with $\sum_{m=1}^{2} \nu_m = 1$ (see, e.g. Gantmacher, 1959) such that the eigen vector ν is constructed corresponding to the maximal eigen value, say ρ, of the matrix $E = [E(R_{mm'})]$.

Allocation Performance

Once the w_i's are constructed, all δ's become known. Then $\frac{\sum_{r=1}^{K} \delta_r}{K}$ indicates the proportion of patients who receive the type 1 treatment and $\frac{\sum_{r=1}^{K}(1-\delta_r)}{K}$ indicates the

proportion of patients who receive the type 2 treatment. It can be shown that

$$\frac{\sum_{r=1}^{K} \delta_r}{K} \to \nu_1 \quad \text{and} \quad \frac{\sum_{r=1}^{K}(1 - \delta_r)}{K} \to \nu_2,$$

almost surely as $K \to \infty$, where ν_m $(m = 1, 2)$ is defined as before (cf. Athreya and Karlin (1968)).

The above two proportions can be re-expressed by using a single formula. For this, let X_i be a categorical variable such that $X_i \equiv m$ if the ith split is type m, $m = 1, 2$ and let $I_{im} = 1$ if $X_i = m$ and $I_{im} = 0$ otherwise. Then the proportion of type m splits after K splits is $\frac{\delta_{Sm}(K)}{K} \equiv \frac{\sum_{i=1}^{K} I_{im}}{K}$. Thus

$$\frac{\delta_{Sm}(K)}{K} \to \nu_m, \tag{2.3}$$

almost surely as $K \to \infty$, where ν_m is defined as before.

GPU Rule Applied to Binary Case

Consider a clinical experiment where binary responses are collected from the subjects. Let $Y_i = 1$ if the response of the ith subject is a "success", 0 otherwise. The total proportion of successes with the type 1 treatment will be then $\frac{\sum_{r=1}^{K} \delta_r}{K}$ and with the type 2 treatment it will be $\frac{\sum_{r=1}^{K}(1 - \delta_r)}{K}$. These two proportions can be expressed through a single formula by using the notation I_{im}. It then follows that the total proportion of successes with treatment m, for $m = 1, 2$, in the trial is $\frac{\sum_{i=1}^{K} Y_i I_{im}}{K}$.

Note that one may be interested to estimate p_m, where $p_m = \Pr\{Y_i = 1 | X_i = m\}$, for $i = 1, \ldots, K$ and $m = 1, 2$. It can be obtained by the maximum likelihood estimation approach (cf. Rosenberger and Sriram (1997)). The maximum likelihood estimator of p_m is given by

$$\hat{p}_m = \frac{\sum_{i=1}^{K} Y_i I_{im}}{\sum_{i=1}^{K} I_{im}}, \tag{2.4}$$

which is the observed proportion of successes on treatment m. Rosenberger and Sriram (1997) have shown that $\hat{p}_m$ is strongly consistent for p_m. For $p = (p_1, p_2)'$, Rosenberger, Flournoy, and Durham (1997) have shown that the vector $K^{\frac{1}{2}}(\hat{p} - p)$ is asymptotically normal with mean vector 0, and diagonal covariance matrix, $diag[\frac{p_1 q_1}{\nu_1}, \frac{p_2 q_2}{\nu_2}]$,

where $q_m = 1 - p_m$. By Slutsky's theorem, it further follows that $[\delta_{Sm}(K)]^{\frac{1}{2}}(\hat{p}_m - p_m)$ has asymptotically a univariate normal distribution with 0 mean and variance $p_m q_m$. In the same manner as for the independent case, one may make asymptotic inference on the p_m's using the usual contrasts and the χ^2 statistic.

2.1.1.2 Randomized Play-the-Winner (RPW) Rule as a GPU Model: A Binary Case

The randomized play-the-winner (RPW) rule is an adaptive design introduced by Wei and Durham (1978), motivated as an extension to Zelen's (1969) play-the-winner rule. Wei (1979) first noted that the RPW rule can be formulated as a GPU model, which can be described for the case with two treatments as follows:

The GPU rule is described in the previous sub-section § 2.1.1.1. Let A and B denote two treatments and the response of each patient to treatment is dichotomous, either a success or a failure. To construct the RPW rule as a special case of GPU, one can start with α particles of each type, i.e. $n_0 = (\alpha, \alpha)$ in the urn. When a patient is available for an assignment, a particle is drawn at random and replaced. If it is type m, then treatment m is assigned to this patient, where $m \equiv A, B$. When the response of a previous patient to treatment m is available, we change the structure of the urn based on the following rule: if this response is a success, then additional β_0 particles of type m and additional α_0 particles of type m' are put in the urn; if this response is failure, then additional α_0 particles of type m and β_0 particles of type m' are put in the urn, where $\beta_0 \geq \alpha_0 \geq 0$, $m, m' \equiv A, B$, and $m' \neq m$. It is to be noted that after each response, exactly $\alpha_0 + \beta_0$ additional particles are added to the urn. This rule is denoted by $RPW(\alpha, \alpha_0, \beta_0)$. It is also applicable when responses are delayed.

Allocation Probability

Based on the above $RPW(\alpha, \alpha_0, \beta_0)$, we now show how to compute the probability for an incoming subject to be assigned under treatment A or B. For this purpose

define indicator variables (δ_i, Y_i) as:

$$
\delta_i = \begin{cases} 1, & \text{if the } i\text{th patient is assigned to } A \\ \\ 0, & \text{if the } i\text{th patient is assigned to } B \end{cases}
$$

and

$$
Y_i = \begin{cases} 1, & \text{if the } i\text{th patient's response is success} \\ \\ 0, & \text{if the } i\text{th patient's response is failure} \end{cases}
$$

Let S_{Km} and F_{Km} be the number of successes and failures with treatment m after K assignments, respectively, where $m \equiv A, B$. Then

$$
S_{KA} = \sum_{i=1}^{K} \delta_i Y_i, \quad F_{KA} = \sum_{i=1}^{K} \delta_i (1 - Y_i), \quad \text{and}
$$

$$
S_{KB} = \sum_{i=1}^{K} (1 - \delta_i) Y_i, \quad F_{KB} = \sum_{i=1}^{K} (1 - \delta_i)(1 - Y_i).
$$

Further let n_{KA} and n_{KB} be the number of particles of types A and B in the urn, respectively, after K responses. Then

$$
n_{KA} = \alpha + \beta_0 [S_{KA} + F_{KB}] + \alpha_0 [S_{KB} + F_{KA}],
$$

$$(2.5)$$

$$
n_{KB} = \alpha + \beta_0 [S_{KB} + F_{KA}] + \alpha_0 [S_{KA} + F_{KB}].
$$

Hence, after K responses the total number of particles in the urn will be n_K where $n_K = n_{KA} + n_{KB} = 2\alpha + (\alpha_0 + \beta_0)K$. Given all the previous assignments $(\delta_1, \ldots, \delta_K)$ and responses $(y_1, \ldots, y_K)$, let w_{K+1} be the conditional probability of assigning the $(K+1)st$ patient to the treatment A. Then

$$
w_{K+1} = \frac{n_{KA}}{n_K}. \tag{2.6}
$$

Allocation Performance

Let the probability of a single trial success for treatment m is p_m, where $0 < p_m < 1$ and $m \equiv A, B$, i.e. $p_A = \Pr\{success|A\}$, $p_B = \Pr\{success|B\}$, $q_A = 1 - p_A$, and $q_B = 1 - p_B$. The random variables n_{KA} form a stochastic process with transition probabilities

$$\Pr[n_{(K+1)A} = n_{KA} + \beta_0 | n_{KA}] = [p_A n_{KA} + q_B n_{KB}]/n_K,$$

(2.7)

$$\Pr[n_{(K+1)A} = n_{KA} + \alpha_0 | n_{KA}] = [q_A n_{KA} + p_B n_{KB}]/n_K.$$

From (2.7) we have the following recursive relations for expectations

$$E[n_{(K+1)A}] = [1 + (p_A - q_B)(\beta_0 - \alpha_0)/n_K] E[n_{KA}]$$

$$+ \alpha_0 p_B + \beta_0 q_B.$$

(2.8)

Recall from (2.3) that $\delta_{SA}(K)$ denotes the total number of patients assigned to treatment A after K assignments. It then follows (Wei and Durham (1978)) that by (2.5) we have

$$E[\delta_{SA}(K)] = [E(n_{KA}) - \alpha - K(\beta_0 q_B + \alpha_0 p_B)]/$$

$$[(p_A - p_B)(\beta_0 - \alpha_0)],$$

(2.9)

where $p_A \neq p_B$ and $\beta_0 \neq \alpha_0$.

Without any loss of generality, one can assume that $p_A \geq p_B$. By (2.8) and lemma 6.6 of Freedman (1965), it can be shown that

$$\lim_{K \to \infty} \frac{E[n_{KA}]}{K} = \frac{(\alpha_0 p_B + \beta_0 q_B)}{[1 - (p_A - p_B)(\beta_0 - \alpha_0)/(\alpha_0 + \beta_0)]}.$$

(2.10)

It then follows from (2.9) that

$$\lim_{K \to \infty} \frac{E[\delta_{SA}(K)]}{n} = \frac{\alpha_0 p_B + \beta_0 q_B}{[\alpha_0(p_A + p_B) + \beta_0(q_A + q_B)]}.$$

(2.11)

It is to be noted that (2.11) is increasing in β_0/α_0 and tend to $q_B/(q_A + q_B)$ as $\beta_0/\alpha_0 \to \infty$. Therefore, if β_0 is large with respect to α_0, we force the trial to put more patients on the better treatment. It is also interesting to note that for the rule $RPW(\alpha, 0, \beta_0)$, (2.11) becomes $q_B/(q_A + q_B)$, which is the asymptotic proportion of patients treated by A when the play-the-winner (PW) rule is utilized. The PW rule prescribes that a success with a given treatment generates a future trial with the same treatment, while a failure generates a trial with the alternative treatment. It follows that, for a large trial, the $RPW(\alpha, 0, \beta_0)$ is as good as the modified play-the-winner (MPW) rule for placing more patients on the better treatment, where MPW specifies that after each success we continue to use the same treatment and after each failure we switch to the other treatment.

2.1.2 Adaptive Clinical Trials with Discrete Response Variables in the Presence of Prognostic Factors

Note that a selected treatment may work differently on a subject if there is a possibility that the response of a subject may also be affected by certain covariates (prognostic factors). One disadvantage of the RPW rule described in § 2.1.1.2 is that it does not include the prognostic factors that may affect the response of a patient. Pocock and Simon (1975), Efron (1980), Begg and Iglewicz (1980), Atkinson (1982), and Bandyopadhyay and Biswas (1999) considered prognostic factors in designing the adaptive clinical trials. Bandyopadhyay and Biswas (1999) proposed an adaptive randomized play-the-winner (ARPW) scheme in the presence of prognostic factors with a goal of allocating more patients to the better treatment in the course of sampling. In the following sub-section, we discuss the ARPW sampling scheme with non-stochastic prognostic factors. Furthermore, as prognostic factors may also be random, we consider this situation in § 2.1.2.2.

2.1.2.1 ARPW Rule with Non-Stochastic Prognostic Factors

Construction of ARPW Rule

Assume that there is only one prognostic factor C, which is non-stochastic, and the corresponding variable is either discrete or can be easily transformed to a discrete random variable with $(G+1)$ ordered grades $0, 1, \ldots, G$, defined with consulting a clinician. Grade 0 is for the least favorable and grade G for the most favorable condition. Clearly, the response of the *ith* patient depends not only on the treatment (A or B) by which the patient is treated, but also on the grade $u_i \in \{0, 1, \ldots, G\}$ of the *ith* patient. Using this prognostic factor C with its $(G+1)$ grades and treatments A and B, ARPW can be described through an urn model as follows:

One can start with two types of balls A and B such that there will be α balls of each type. An entering patient, say the *ith* patient, of grade u_i can be treated by any one of the two treatments by drawing a ball with replacement. If success occurs an additional $(G - u_i + \beta_0)\tilde{\beta}$ balls of the same kind and $(u_i + \alpha_0)\tilde{\beta}$ of the opposite kind are added in the urn. On the other hand, if a failure occurs, we add an additional $(G - u_i + \alpha_0)\tilde{\beta}$ balls of the same kind and $(u_i + \beta_0)\tilde{\beta}$ balls of the opposite kind in the urn. Thus for every entering patient, $(G + \alpha_0 + \beta_0)\tilde{\beta}$ balls are added in the total, $G\tilde{\beta}$ for the grade and $(\alpha_0 + \beta_0)\tilde{\beta}$ for a success or failure response. For $G = 0$, $u_i = 0$, and $\tilde{\beta} = 1$, this ARPW procedure reduces to $RPW(\alpha, \alpha_0, \beta_0)$.

Bandyopadhyay and Biswas (1999) designed this ARPW rule by assuming $\alpha_0 = 0$ and $\beta_0 = \tau$. For given $(\alpha, \tilde{\beta}, \tau)$, they denoted their procedure as $ARPW(\alpha, \tilde{\beta}, \tau)$.

Allocation Probability

One may be interested to compute the probability of assigning an incoming patient to a particular treatment. For this assignment we now explain the $ARPW(\alpha, \tilde{\beta}, \tau)$ rule due to Bandyopadhyay and Biswas (1999). Suppose that there is a sequential chain of patient's entrance up to a maximum of K patients. Corresponding to the *ith* $(i = 1, \ldots, K)$ entering patient with grade u_i, let us define a pair of indicator

variables $\{\delta_i, Y_i\}$ as follows:

$$\delta_i = \begin{cases} 1, & \text{if the } i\text{th patient is assigned to } A \\ \\ 0, & \text{if the } i\text{th patient is assigned to } B \end{cases}$$

and

$$Y_i = \begin{cases} 1, & \text{if the } i\text{th patient response is success} \\ \\ 0, & \text{if the } i\text{th patient response is failure} \end{cases}$$

Suppose that at the treatment selection stage for the $(i + 1)$st patient, the total number of balls in the urn is n_i, where $n_i = 2\alpha + i(G + \tau)\tilde{\beta}$. Also suppose that among the n_i balls, the number of balls of type A is n_{iA} which is given by

$$
\begin{aligned}
n_{iA} &= \alpha + \sum_{j=1}^{i} \delta_j Y_j (G - u_j + \tau)\tilde{\beta} + \sum_{j=1}^{i} \delta_j (1 - Y_j)(G - u_j)\tilde{\beta} \\
&\quad + \sum_{j=1}^{i} (1 - \delta_j) Y_j u_j \tilde{\beta} + \sum_{j=1}^{i} (1 - \delta_j)(1 - Y_j)(u_j + \tau)\tilde{\beta} \\
&= \alpha + \beta \left[2\tau \sum_{j=i}^{i} \delta_j Y_j + \sum_{j=1}^{i} (u_j + \tau) \right. \\
&\qquad \left. - \sum_{j=1}^{i} (\tau + 2u_j - G)\delta_j - \tau \sum_{j=1}^{i} Y_j \right].
\end{aligned}
\tag{2.12}
$$

Let w_{i+1} be the conditional probability of $\delta_{i+1} = 1$ given all the previous assignments $\{\delta_1, \ldots, \delta_i\}$, and all the previous responses $\{y_1, \ldots, y_i\}$. Then it follows that

$$w_{i+1} = \Pr(\delta_{i+1} = 1 | \delta_1, \ldots, \delta_i; y_1, \ldots, y_i) = \frac{n_{iA}}{n_i}, \quad i \geq 1. \tag{2.13}$$

Unconditional Allocation Probability

Assume that

$$\Pr(Y_i = 1 | \delta_i = h, u_i) = p_{2-h} a^{G - u_i}, \quad h = 0, 1, \tag{2.14}$$

where $a \in (0, 1)$, called the prognostic factor index, is either known from past experience or can be estimated from past data and $p_1, p_2 \in (0, 1)$, the success probabilities

of treatment A and B respectively at grade G, are unknown. However, it is noted that under equivalence of treatment effects (i.e. when $p_1 = p_2 = p$), the δ_i's are identically distributed as Bernoulli with success probability $\frac{1}{2}$ and Y_i's are independently distributed as $\Pr(Y_i = 1) = 1 - \Pr(Y_i = 0) = pa^{G-u_i}$, and the δ_i's are independent of Y_i's.

From the urn model, it is clear that $w_1 = \frac{1}{2}$. Now from (2.13), the distribution of δ_i $(i = 1, \ldots, K)$ is given by

$$\Pr(\delta_i = 1) = \frac{1}{2}, \quad \text{for } i = 1, \tag{2.15}$$

and for $i \geq 1$,

$$\Pr(\delta_{i+1} = 1) = \frac{1}{2} - d_{i+1}, \tag{2.16}$$

where, by the method of induction,

$$d_{i+1} = \frac{\tilde{\beta}}{2\alpha + i(G+\tau)\tilde{\beta}} + (p_2 - p_1)\sum_{j=1}^{i} a^{G-u_j}\left(\frac{1}{2} + d_j\right)$$

$$+ \frac{\tilde{\beta}}{2\alpha + i(G+\tau)\tilde{\beta}}\sum_{j=1}^{i}[2\tau p_1 a^{G-u_j} - (\tau + 2u_j - G)]d_j, \tag{2.17}$$

(cf. Bandyopadhyay and Biswas (1999)).

Inference Based on ARPW Rule

Suppose that one is interested in any one of the following decisions:

$$H_1 : A > B, \quad H_2 : B > A, \tag{2.18}$$

where $' > '$ means better than. Under H_1, we have $p_1 > p_2$. For $a \in (0,1)$ as in (2.14), let after K allocations,

$$T_{AK} = \sum_{i=1}^{K} a^{u_i} Y_i \delta_i \quad \text{and} \quad T_{BK} = \sum_{i=1}^{K} a^{u_i} Y_i (1 - \delta_i),$$

denote the total number of successes (in the presence of the prognostic factor) with treatment A and B respectively. Further, let

$$\delta_{SA}(K) = \sum_{i=1}^{K} \delta_i = \text{Total number of allocations with treatment } A, \quad \text{and}$$

$$\delta_{SB}(K) = \sum_{i=1}^{K}(1 - \delta_i) = \text{Total number of allocations with treatment } B.$$

Then, for $m \equiv A, B$,

$$g_{mK} = \frac{T_{mK}}{\delta_{Sm}(K)}. \tag{2.19}$$

denotes the proportion of successes due to treatment m.

For a particular treatment $m \equiv A, B$, T_{mK} not only accounts for the total number of successes, but also the grades from which the successes have occurred as a^{u_i} is inversely proportional to the success probability at grade u_i. Based on (2.19), one may then develop the decision rules as follows:

Rule 1: The terminal decision rule: This rule is to accept H_1 if $g_{AK} > g_{BK}$

and H_2 if $g_{AK} < g_{BK}$. If $g_{AK} = g_{BK}$, accept H_1 with probability

$\frac{1}{2}$.

Rule 2: Termination rule with early stopping: Suppose that for $m \equiv A, B$, and after allocation of k $(k = 1, \ldots, K)$ patients, one writes

$$P_{mk}(\nu) = \frac{T_{mk} + \nu}{\delta_{Sm}(k) + \nu}, \quad \text{and} \quad Q_{mk}(\nu) = \frac{T_{mk}}{\delta_{Sm}(k) + K - k - \nu},$$

where $\nu = 0, 1, \ldots, K - k$, indicates the number of patients treated by treatment m. In case $\delta_{Sm}(k) = 0$, consider $P_{mk}(0) = Q_{mk}(K - k) = 0$. Here, $P_{mk}(\nu)$ represents a possible value of g_{mK} where among the future $(K - k)$ incoming patients (after the kth one) exactly ν patients each of grade 0 will be treated by treatment m and for all of them the result will be success. Similarly, $Q_{mk}(\nu)$ is a possible value of g_{mK} where among $(K - k)$ remaining patients $(K - k - \nu)$ patients will be treated by treatment m and for each of them the result will be failure. We then stop sampling and accept A or B at the kth stage if

$$\min_{\nu}(Q_{Ak}(\nu) - P_{Bk}(\nu)) > 0 \quad \text{or} \quad \min_{\nu}(Q_{Bk}(\nu) - P_{Ak}(\nu)) > 0,$$

respectively. So far it is assumed the prognostic factor to be non-stochastic. The case when it is stochastic is discussed in the following section.

2.1.2.2 ARPW Rule with Stochastic Prognostic Factors

Suppose that the variable U corresponding to the prognostic factor has distribution function $F(u), u = 0, 1, \ldots, G$. Recall from (2.13) that

$$w_{i+1} = \Pr(\delta_{i+1} = 1 | \delta_1, \ldots, \delta_i; y_1, \ldots, y_i) = \frac{n_{iA}}{n_i}, \quad i \geq 1.$$

We now re-express this probability in the presence of the stochastic prognostic factor as follows:

Based on the $ARPW(\alpha, \tilde{\beta}, \tau)$, at the $(i+1)$st selection stage, one may write

$$\Pr(\delta_{i+1} = 1 | \delta_1, \ldots, \delta_i; y_1, \ldots, y_i; u_1, \ldots, u_i) = \frac{n_{iA}}{n_i}. \tag{2.20}$$

It then follows from (2.20) that given all the previous assignments $(\delta_1, \ldots, \delta_i)$ and responses $(y_1, \ldots, y_i)$, the allocation probability for the $(i+1)$st patient is

$$w_{i+1} = \Pr(\delta_{i+1} = 1 | \delta_1, \ldots, \delta_i; y_1, \ldots, y_i) = \frac{E_U(n_{iA})}{n_i}, \tag{2.21}$$

where $E_U(n_{iA})$ denotes the expected value of n_{iA} over the distribution of U.

To compute this allocation probability, Bandyopadhyay and Biswas (1999) defined $\psi_l(a) = E[a^{G-U} U^l]$ (provided it exists) and $P_U(w)$, the probability generating function (p.g.f) of U. Then the marginal distribution of the δ_i's can be obtained from (2.15)-(2.17) by replacing $a^{G-u_j} u_j^l$ and a^{G-u_j} with $\psi_l(a)$ and $a^G P_U(a^{-1})$, respectively, at every stage. Subsequent analysis can be carried out in a similar way. If we consider the simplest case where $G = 1$, then U follows a Bernoulli distribution with probability of success p. In this case we have $E(a^{G-U}) = 1 - q + qa$ and $E(a^{G-U} U^l) = q$ for each l, where $q = 1 - p$.

All the ARPW rules with discrete response are discussed above by considering only one prognostic factor. If there are more than one prognostic factor, we can proceed in the following direction. Suppose there are S prognostic factors $C_1, \ldots, C_S$ with grades $0, 1, \ldots, G_\ell$ for the ℓth factor. First, consider $G + 1 = \prod_{\ell=1}^{S}(G_\ell + 1)$ factor combinations. We can arrange these $(G + 1)$ combinations according to the favorable conditions $0, 1, \ldots, G$ and carry out the same procedure discussed above. If G is

moderately large, the revised grading may be difficult as it involves the combination of different grades. In that case for an entering patient with grade $u_{\ell j}$ for the factor $C_\ell, \ell = 1, \ldots, S$, we have

$$\Pr(Y_j = 1 | \delta_j = h, u_{\ell j}) = p_{2-h} \prod_{\ell=1}^{S} a_\ell^{G_\ell - u_{\ell j}}, h = 0, 1, \qquad (2.22)$$

where we have ideas about the prognostic factor indices $a_1, \ldots, a_S$ from past experience. Then the same procedure can be carried out. However, it requires more modeling and knowledge about parameters.

2.2 Adaptive Clinical Trials with Continuous Response Variables

Most of the available works on adaptive clinical trials in the literature are based on a discrete response variable. The urn models that we have discussed so far are applicable for binary or polychotomous responses. One may however collect continuous responses in some clinical trials. For example, it may be necessary to see the effect of treatments and other covariates on the blood pressure of patients. In this example, blood pressure is a continuous response variable. Rosenberger (1993) introduced a response-adaptive design for continuous responses, in the spirit of the RPW rule. Bandyopadhyay and Biswas (2001) considered the case of two competing treatments with continuous responses. We provide a brief discussion on these approaches below.

2.2.1 Adaptive Clinical Trials with Continuous Response Variables in the Absence of Prognostic Factors

Rosenberger (1993) developed a biased coin randomization scheme for continuous responses based on a liner rank statistic without considering prognostic factors. According to his scheme, at each stage of the trial, the next treatment assignment is generated from a rank-type statistic, giving the higher probability of assignment to

the treatment that is "winning" at that stage. Assume that there are two treatments, say A and B. Define for the *ith* patient $(i = 1, \ldots, K)$

$$\delta_i = \begin{cases} 1, & \text{if the patient is assigned to treatment } A \\ 0, & \text{if the patient is assigned to treatment } B \end{cases}$$

For $1 \leq i \leq j \leq K$, suppose that the outcomes of j patients are available. Next, let r_{ij} be the rank for the *ith* patient among j patients. Also, let a_{ij} be the score functions corresponding to r_{ij}, where $\sum_{i=1}^{j} a_{ij} = 0$, $j = 1, \ldots, K$. Define $a_{ij}^{+} = a_{ij} I(a_{ij} > 0)$, where I is an indicator function. Then, given all the previous assignments $(\delta_1, \ldots, \delta_{i-1})$ and responses $(y_1, \ldots, y_{i-1})$, the probability that the *ith* patient is allocated to treatment A is w_i given by

$$\begin{aligned} w_i &= \Pr(\delta_i = 1 | \delta_1, \ldots, \delta_{i-1}, y_1, \ldots, y_{i-1}) \\ &= \frac{1}{2} \left\{ 1 + \frac{\sum_{j=1}^{i-1} a_{j,i-1}(\delta_i - \frac{1}{2})}{\sum_{j=1}^{i-1} a_{j,i-1}^{+}} \right\}. \end{aligned} \tag{2.23}$$

Remark that each patient is randomized with a probability that is a function of the current value of the rank statistic. The better the responses of previous patients on treatment A, relative to those on B, the larger will be the probability for the next patient to be assigned to the treatment A. One can use a permutation test based on the rank scores to test the hypotheses of the treatment effect. The form of the test statistic is given in Rosenberger(1993) and simulations indicate that it is asymptotically standard normal.

2.2.2 Adaptive Clinical Trials with Continuous Response Variables in the Presence of Prognostic Factors

It is reasonable to assume that incoming patients to an adaptive clinical trials are heterogeneous with respect to some prognostic factors. Bandyopadhyay and Biswas (2001) considered the case of competing treatments with continuous treatment responses and proposed an allocation design provided by means of a link function that

accommodates prognostic factors. The allocation design bridges the past allocation and response histories and present allocation pattern. Based on a normal linear model they discussed the allocation design as follows.

Allocation Probability

Suppose that there are two competing treatments A and B. For the ith patient, let

$$\delta_i = \begin{cases} 1, & \text{if the patient is treated by treatment } A \\ \\ 0, & \text{if the patient is treated by treatment } B \end{cases}$$

Also let x_i be the $p \times 1$ vector representing the prognostic factors for the ith patient, which does not include the treatment covariate. Let Y_i be the continuous variable representing the response of the ith patient, treated by either A or B following the adaptive design. Assume that responses are instantaneous and normally distributed. Suppose μ_A and μ_B are population characteristics representing the treatment effects of A and B, respectively.

Initially, one may allocate the first $2k$ patients to the two treatments randomly, k to each treatment. This ensures that every treatment will have at least k allocations, and k is so chosen that estimates of parameters can be obtained from this initial sample. Now, for $i \geq 2k$, suppose that $\hat{\mu}_{Ai}$ and $\hat{\mu}_{Bi}$ are the estimates of μ_A and μ_B, respectively, on the basis of the responses $Y_1, \ldots, Y_{2k}, \ldots, Y_i$, eliminating the effects of the prognostic factors. Let us consider a suitable link function which bridges the past histories to the $(i + 1)$st allocation. This may be a suitable cumulative distribution function $G(\cdot)$ that is symmetric about 0, that is, $G(0) = \frac{1}{2}$ and $G(-x) = 1 - G(x)$. A natural choice for G is the probit link function $G(x) = \Phi(x, C_0)$, where Φ is the standard normal cumulative distribution function and the choice of the tuning constant C_0 should be handled with care.

One can allocate the $(i+1)$st patient to treatment A with probability $G(\hat{\mu}_{Ai} - \hat{\mu}_{Bi})$ and to treatment B with probability $1 - G(\hat{\mu}_{Ai} - \hat{\mu}_{Bi}) = G(\hat{\mu}_{Bi} - \hat{\mu}_{Ai})$. The allocation procedure favors the treatment doing better at that stage. This procedure is continued

up to a predetermined number K of patients. According to Bandyopadhyay and Biswas (2001), the responses of the patients are assumed to be linear models with normally distributed errors. Then the observation for the ith patient can be expressed as

$$Y_i = \delta_i \mu_A + (1 - \delta_i)\mu_B + x_i'\beta + \varepsilon_i, \quad i = 1, \ldots, K; K \geq 1, \tag{2.24}$$

where the ε_i's are assumed to be independently and identically distributed $N(0, \sigma^2)$ random variables. Then the sample means corresponding to treatments A and B up to the ith patients are

$$\overline{Y}_{Ai} = \frac{\sum_{j=1}^{i} \delta_j Y_j}{\sum_{j=1}^{i} \delta_j}, \quad \overline{Y}_{Bi} = \frac{\sum_{j=1}^{i}(1 - \delta_j)Y_j}{\sum_{j=1}^{i}(1 - \delta_j)}.$$

Then, if

$$\delta_{SA}(i) = \sum_{j=1}^{i} \delta_j, \quad \delta_{SB}(i) = \sum_{j=1}^{i}(1 - \delta_j),$$

$$S_{xx}^{(i)} = \sum_{j=1}^{i} \delta_j(x_j - \overline{x}_{Ai})(x_j - \overline{x}_{Ai})' + \sum_{j=1}^{i} \overline{\delta_j}(x_j - \overline{x}_{Bi})(x_j - \overline{x}_{Bi})',$$

$$S_{xy}^{(i)} = \sum_{j=1}^{i} Y_j x_j - \delta_{SA}(i)\overline{Y}_{Ai}\overline{x}_{Ai} - \delta_{SB}(i)\overline{Y}_{Bi}\overline{x}_{Bi},$$

$$\overline{\delta}_j = 1 - \delta_j, \quad \overline{x}_{Ai} = \frac{\sum_{j=1}^{i} \delta_j x_j}{\sum_{j=1}^{i} \delta_j}, \quad \overline{x}_{Bi} = \frac{\sum_{j=1}^{i}(1 - \delta_j)x_j}{\sum_{j=1}^{i}(1 - \delta_j)},$$

then from (2.24), it can be shown that

$$\hat{\mu}_{Ai} - \hat{\mu}_{Bi} = \overline{Y}_{Ai} - \overline{Y}_{Bi} - (\overline{x}_{Ai} - \overline{x}_{Bi})\hat{\beta}^{(i)}, \tag{2.25}$$

where

$$\hat{\beta}^{(i)} = S_{xx}^{(i)^{-1}} S_{xy}^{(i)}. \tag{2.26}$$

Given the past allocation histories $\{\delta_1, \ldots, \delta_i\}$, responses $(y_1, \ldots, y_i)$, and prognostic factors $(x_1, \ldots, x_i)$ and based on the probit link function, the conditional probability that the $(i + 1)$st patient will be treated by treatment A is w_{i+1} where

$$w_{i+1} = \Pr(\delta_{i+1} = 1|\delta_1, \ldots, \delta_i; y_1, \ldots, y_i; x_1, \ldots, x_i) = \Phi\left(\frac{\hat{\mu}_{Ai} - \hat{\mu}_{Bi}}{C_0}\right), \tag{2.27}$$

where C_0 being the tuning constant as mentioned before.

Unconditional Allocation Probability

Note that given $(\delta_1, \ldots, \delta_i)$, one may have (Bandyopadhyay and Biswas (2001))

$$\hat{\mu}_{Ai} - \hat{\mu}_{Bi} \sim N\left\{\mu_A - \mu_B, \sigma^2\left(\frac{1}{\delta_{SA}(i)} + \frac{1}{\delta_{SB}(i)} + V\right)\right\}, \qquad (2.28)$$

where

$$V = (\overline{x}_{Ai} - \overline{x}_{Bi})' S_{xx}^{(i)-1} (\overline{x}_{Ai} - \overline{x}_{Bi}).$$

Assuming the x_i's are independently and identically distributed as $N_p(\mu_x, \Sigma)$, it can be shown that

$$V = \left(\frac{\delta_{SA}(i) + \delta_{SB}(i)}{\delta_{SA}(i)\delta_{SB}(i)}\right)\frac{p}{i-p-1}W,$$

where W has the $F - distribution$, i.e. $W \sim F(p, i - p - 1)$. Thus given $(\delta_1, \ldots, \delta_i)$ and W, one can show that

$$\hat{\mu}_{Ai} - \hat{\mu}_{Bi} \sim N\left\{\mu_A - \mu_B, \sigma^2\left(\frac{1}{\delta_{SA}(i) + \delta_{SB}(i)}\right)\left(1 + \frac{p}{i-p-1}W\right)\right\}.$$

Furthermore, it is well known that, in probability, as $i \to \infty$,

$$\frac{W}{i-p-1} \to 0. \qquad (2.29)$$

Thus, if one takes expectation over $(\delta_1, \ldots, \delta_i)$ and W, the unconditional probability that the $(i + 1)$st patient will be treated by treatment A will be

$$\begin{aligned}
\Pr(\delta_{i+1} = 1) &= E\left\{\Phi\left(\frac{\hat{\mu}_{Ai} - \hat{\mu}_{Bi}}{C_0}\right)\right\} \\
&= E\left[E\left\{\Phi\left(\frac{\hat{\mu}_{Ai} - \hat{\mu}_{Bi}}{C_0}\right)|\delta_1, \ldots, \delta_i; W\right\}\right] \\
&= E\left\{\Pr\left(U < \frac{\hat{\mu}_{Ai} - \hat{\mu}_{Bi}}{C_0}\right)\right\}, \qquad (2.30)
\end{aligned}$$

where U follows a standard normal distribution. Clearly this unconditional probability reduces to

$$E\left(\Phi\left[\frac{\mu_A - \mu_B}{\left\{C_0^2 + \sigma^2\left(\frac{1}{\delta_{SA}(i)} + \frac{1}{\delta_{SB}(i)}\right)\left(1 + \frac{p}{i-p-1}W\right)\right\}^{\frac{1}{2}}}\right]\right),$$

which equals $\frac{1}{2}$ when $\mu_A = \mu_B$. Thus, the allocation pattern is balanced only when there is no difference in treatment effects. In this case, the expected number of allocations to any treatment is $\frac{K}{2}$. If treatment A is better than treatment B, that is $\mu_A > \mu_B$, the unconditional probability is skewed in favor of that better treatment.

Remark that the construction of the adaptive design in all of the above works is based on the non-longitudinal (cross-sectional) set up. That is, once the treatment was assigned to an individual, the individual was expected to provide only one response. In this thesis, we consider both longitudinal fixed and mixed models to analyze repeated binary responses in adaptive clinical trials which are discussed in details in Chapters 3 and 4, respectively.

Chapter 3

Longitudinal Fixed Model For Binary Data in Adaptive Clinical Trials

In Chapter 2, we have summarized the adaptive designs developed so far, for clinical trials in the cross-sectional set up. Note however that in practice there may be clinical trials where patients enter the studies sequentially over a period of time and it is useful to collect repeated observations from each patient. Therefore, as opposed to the cross-sectional adaptive designs, it is important to construct longitudinal adaptive designs by using available repeated responses and covariate information of patients for assigning more patients to the better treatment. In this chapter, we introduce a longitudinal model for binary responses in the adaptive clinical trial by assuming that the repeated observations of an individual are likely to be correlated following an autocorrelation structure, but the longitudinal data collected from individual patients are not affected by any unknown individual effect. Recently, Sutradhar and Biswas (2001) proposed a simple longitudinal adaptive design so that more study subjects may be assigned to the better treatment. The construction of such a longitudinal adaptive design may be considered as an extension of the existing adaptive designs based on the idea of RPW rule in the non-longitudinal set up. Following Sutradhar

and Biswas (2001), we describe a longitudinal adaptive design for binary responses in § 3.2. We also study the performance of the proposed design in allocating study subjects to a better treatment, through a simulation study in § 3.3. Here it is also of interest to compute the treatment effect as well as the effects of the other prognostic factors based on all the covariate information and the responses available at the end of the study. With regard to the estimation of the effects of the covariates including the treatment effect, one must take the longitudinal adaptive design weights as well as the correlation of the repeated binary responses into account. In § 3.4, following Sutradhar and Das (1999)(see also Jowaheer and Sutradhar (2002)) we introduce a general auto-correlation structure for the repeated binary responses and take these correlations as well as the longitudinal adaptive design weights (to be discussed in § 3.2) into account for the consistent and efficient estimation of the regression parameters of the model (see also Sutradhar and Biswas (2001)). More specifically, we use a weighted generalized quasi-likelihood (WGQL) approach for such consistent and efficient estimation. In § 3.6, we examine the performance of the proposed $WGQL$ estimation approach through a simulation study. In the same section, similar to Sutradhar and Biswas (2001), we also conduct a separate simulation study to examine the misspecification effect of the longitudinal adaptive designs. A simulation based coverage probabilities for the treatment effect are also reported in § 3.6.

3.1 Simple Random Sampling Design Based Binary Longitudinal Model

Let $y_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT})'$ be a vector of T-dimensional repeated binary responses for the ith $(i = 1, \ldots, K)$ individual. Let $x_{it}^* = (x_{it1}, \ldots, x_{itu}, \ldots, x_{itp})'$ be a vector of p covariates associated with the response y_{it}. As, in general, it is difficult to write the multivariate binary distribution for the repeated binary responses $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$,

Liang and Zeger (1986) by-passed the specification of the joint distribution and introduced a 'working' correlation structure based generalized estimating equations (GEE) approach for the estimation of β. This 'working' correlation approach has however many pitfalls which are discussed by Crowder (1995) and Sutradhar and Das (1999).

Let β be the $p \times 1$ vector of regression parameters of interest. For known functional forms $a_i(\cdot)$, let $E(Y_{it}) = \mu_{it} = a_i'(\theta_{it})$ and $var(Y_{it}) = \sigma_{it}^{*2} = \phi^{-1}a_i''(\theta_{it})$ where $\theta_{it} = x_{it}^{*\prime}\beta$, $a_i'(\theta_{it})$ and $a_i''(\theta_{it})$ are, respectively, the first and second derivatives of $a_i(\theta_{it})$ with respect to θ_{it}, and ϕ is a possibly unknown scale parameter. In many important situations, ϕ may be assumed to be known. For example, one may use $\phi = 1$ for binary and Poisson data. We therefore consider the case $\phi = 1$. In the longitudinal set up, the components of the vector y_i are repeated responses, which are likely to be correlated. In practice, this longitudinal correlation structure is unknown, which makes it difficult to estimate β. Further, let $\Sigma_i^*(\rho) = A_i^{\frac{1}{2}} C(\rho) A_i^{\frac{1}{2}}$ be the true covariance matrix of y_i $(i = 1, \ldots, K)$, where $A_i = \text{diag}[var(Y_{i1}), \ldots, var(Y_{it}), \ldots, var(Y_{iT})]$ and $C(\rho)$ is the $T \times T$ correlation matrix characterized by the ρ correlation parameters. In the longitudinal set-up, $C(\rho)$ is usually considered as an autocorrelation matrix of the form

$$C(\rho_1, \ldots, \rho_{T-1}) = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{T-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \cdots & 1 \end{bmatrix}, \tag{3.1}$$

(cf. Sutradhar and Das (1999)) where for $\ell = 1, \ldots, T - 1$, ρ_ℓ is referred to as the autocorrelation of lag ℓ. Here one is interested to study the dependence of the binary responses on the covariates, the time dependence among repeated measurements for an individual being of secondary interest. This set up covers the exponential family model with canonical link functions. For the case when covariates are fixed, the covariate effects β may be estimated by solving the generalized quasi-likelihood (GQL)

estimation equation

$$\sum_{i=1}^{K} X_i^{*'} A_i \Sigma_i^{*-1}(\hat{\rho}_1, \ldots, \hat{\rho}_{T-1})(y_i - \mu_i) = 0, \tag{3.2}$$

(Sutradhar and Das (1999)) where $X_i^{*'} = (x_{i1}^*, \ldots, x_{it}^*, \ldots, x_{iT}^*)$ is the $p \times T$ matrix, $\Sigma_i^*(\hat{\rho}_1, \ldots, \hat{\rho}_{T-1}) = A_i^{\frac{1}{2}} C(\hat{\rho}_1, \ldots, \hat{\rho}_{T-1}) A_i^{\frac{1}{2}}$ with $\hat{\rho}_\ell$, a consistent estimator of ρ_ℓ, and $\mu_i = (\mu_{i1}, \ldots, \mu_{it}, \ldots, \mu_{iT})'$ with $\mu_{it} = E(Y_{it})$. Because the responses are binary, a logistic regression model for each response is quite a natural choice. More formally, we assume that

$$\ell n\{\mu_{it}/(1 - \mu_{it})\} = x_{it}^{*'}\beta, \tag{3.3}$$

where $\mu_{it} = E(Y_{it}) = \Pr\{Y_{it} = 1 | x_{it}, \beta\}$ is the probability of 'success,' say, for the tth response of the ith individual. The GQL estimator obtained from (3.2) is consistent as well as highly efficient. The consistent estimator of the longitudinal correlation parameter ρ_ℓ can be achieved by using the method of moments, that is, by solving the estimating equation derived by equating the sample covariance to its population counterpart. To be specific, the moment estimator of ρ_ℓ is given by

$$\hat{\rho}_\ell = \frac{\sum_{i=1}^{K} \sum_{t=1}^{T-\ell} y_{it}^* y_{i(t+\ell)}^* / (T - \ell)}{\sum_{i=1}^{K} \sum_{t=1}^{T} y_{it}^{*2} / T} \quad (\ell = 1, \ldots, T-1),$$

(cf. Jowaheer and Sutradhar (2002), equation (3.8)) where $y_{it}^* = \frac{y_{it} - \mu_{it}}{[\mu_{it}(1 - \mu_{it})]^{1/2}}$.

3.2 Adaptive Design in Longitudinal Clinical Trial Set up

In the context of adaptive clinical trials, additional problems get mounted as the individuals enter the study in a sequence and one or more covariates, such as the treatment, for the incoming individuals are determined based on the outcomes of the past individuals. For example, in clinical trial studies, it is highly desirable that one uses certain data-dependent treatment allocation rules which exploit accumulating information from the past, and assign incoming individuals to treatments so that

more study subjects are assigned to the better treatment. Once the study subject is assigned to a treatment, the individual remains under the same treatment during the study period. Consequently, the repeated responses for an individual do not only depend on the design weights used to choose his or her treatment group, they become correlated among themselves as they are repeatedly collected from the same individual under a fixed treatment. These longitudinal correlations along with design weights of the individuals must be taken into account in estimating the effects of covariates, including the particular treatment effect.

The purpose of this section is to introduce a longitudinal adaptive design appropriate for repeated binary data in a clinical trial set up. This we do following Sutradhar and Biswas (2001). In § 3.3, we study the performance of the adaptive design discussed in this section. We will return to the estimation of the regression effects in § 3.4, where we introduce a weighted generalized quasi-likelihood (WGQL) approach as an extension of the GQL approach of Sutradhar and Das (1999, § 3) [see also Jowaheer and Sutradhar (2002)] to the clinical trial data in order to accommodate the design weights, for the consistent estimation of the parameters of the model.

Let the ith $(i = 1, \ldots, K)$ patient enter the clinical trial at the time point i and give T consecutive responses. Thus, the whole clinical trial will be completed at time point $K + T - 1$. Suppose $x_{it}^* = (\delta_i, x_{it}')'$ with $x_{it} = (x_{it2}, \ldots, x_{itu}, \ldots, x_{itp})'$. Here δ_i is the treatment covariate and the other $p - 1$ covariates are treated as prognostic factors. In all, there will be $N = KT$ binary longitudinal responses in the clinical trial. Note that as the ith patient enters the system at the ith time point, under the present sequential set up, the tth response of the ith patient is actually collected at time point $i + t - 1$ for $t = 1, \ldots, T$. Consequently, y_{it} may be explained as the response of the ith patient at the tth time sequence where $t = i, i + 1, \ldots, i + T - 1$. We however will explain y_{it} as the tth $(t = 1, \ldots, T)$ repeated response of the ith individual, where the ith individual enters the trial at the ith time point. Further note that the treatment covariate δ_i does not depend on t. This is because, once a patient is assigned to a treatment, the patient remains under the selected treatment

for the complete duration of T periods.

For simplicity, suppose that there are two treatments A and B, and for $i = 2, \ldots, K$, the ith patient is allotted to either of the two treatments depending on the longitudinal outcomes of all $i - 1$ patients and their covariates information. Also suppose that A is the better treatment. The first patient is allotted to A or B randomly with equal probability. Let

$$\delta_i = \begin{cases} 1, & \text{if the ith patient is assigned to treatment } A \\ 0, & \text{if the ith patient is assigned to treatment } B \end{cases}$$

with

$$\Pr.(\delta_i = 1|y_H) = w_i \text{ and } \Pr.(\delta_i = 0|y_H) = 1 - w_i,$$

where y_H indicates the history of the past $i - 1$ patients. It will be assumed that $w_1 = \frac{1}{2}$. In general, for $i = 2, \ldots, K$, the distribution of δ_i will depend on $\{\delta_1, \ldots, \delta_{i-1}\}$ and available responses y_{rt} ($r = 1, \ldots, i - 1$; $1 \leq t \leq \min(T, i - r)$) along with their corresponding covariate vector x_{rt}. We now provide the treatment allocation rule and show how to construct the design weights w_i ($i = 2, \ldots, K$) for the selection of the treatment for the ith individual.

3.2.1 Construction of the Longitudinal Adaptive Design Weights (w_i)

Our longitudinal adaptive design is motivated by the popular randomized play-the-winner (RPW) rule (Wei and Durham (1978)). As in the RPW rule, we illustrate the proposed procedure as an urn design. We start with an urn that will reflect the relative performance of the two treatments A and B at any time point and accordingly the urn proportion will determine the probability for an entering patient to get treated by one of the two treatments. As in the beginning we have no reason to believe that any particular treatment is better than the other, we take the initial urn composition in a 50:50 fashion. Thus, the urn will have two types of balls, say α balls of each type

at the outset, and the probability that the first patient will be treated by treatment A is 0.5, i.e., $\Pr(\delta_1 = 1) = w_1 = 0.5$. Note that the objective of the present longitudinal design is to treat the entering patient by one of the competing treatments in such a way that eventually the better treatment is applied to a larger number of patients. This means that the selection of the treatment for the ith patient ready to enter the trial will be based on its past performance in treating the patients from 1 to $i-1$. Here i ranges from 2 to K. Note that in the present longitudinal set up, a patient provides T consecutive responses. Furthermore, it is assumed that the responses are instantaneous in the sense that the first response (at $t=1$) of the existing $(i-1)$th $(i = 2, \ldots, K)$ patient is obtained before the entry of the ith patient. Suppose that at the selection stage of the ith patient, $\{y_{rt}\}$ denotes all available responses for $r = 1, \ldots, i-1$ and $1 \leq t \leq \min(T, i-r)$. The range of t here depends on the value of r. For example, for the selection time of the ith $(i = 2, \ldots, K)$ patient, $t = 1$ when $r = i - 1$. Similarly $t = 1, 2$ for $r = i - 2$. Also suppose that at this selection stage we take all these available responses into account and for a suitable τ value and for specific available response y_{rt}, we add $y_{rt}\tau$ balls of the same kind by which the patient was treated, and add $(1 - y_{rt})\tau$ balls of the opposite kind in the urn. Thus, at this stage, we add τ balls for each and every available response.

On top of the past responses, it may also be sensible to take into account the condition of certain covariates which, along with the treatment (A or B) were responsible to yield the past responses y_{rt}. Partition the $p - 1$ covariates into two subsets of p_1 and p_2 covariates, so that $p_1 + p_2 = p - 1$. More specifically,

$$x_{rt} = (x_{rt2}, \ldots, x_{rtp})' = (\tilde{x}'_{rt1}, \tilde{x}'_{rt2})',$$

with $\tilde{x}_{rt1} = (x_{rt2}, \ldots, x_{rt(p_1+1)})'$ and $\tilde{x}_{rt2} = (x_{rt(p_1+2)}, \ldots, x_{rt(p_1+p_2+1)})'$. Also suppose that for a suitable known function $\psi(\cdot)$,

$$u_{rt} = \psi(x_{rt(p_1+2)}, \ldots, x_{rt(p_1+p_2+1)})$$

where u_{rt} is a non-stochastic continuous quantity with the domain being $[0, G]$. Here u_{rt} measures the condition of the covariate vector $\tilde{x}_{rt2}$ corresponding to the past

response y_{rt} for $r = 1, \ldots, i-1$, $1 \le t \le \min(T, i-r)$. More specifically, suppose that a greater value of u_{rt} implies a better condition of the rth past patient and it was a more favorable condition of the patient to treat. In the same token, a smaller value of u_{rt} means that the patient was serious. Now to make sure so that this better or serious condition of the past patient does not influence the selection of the treatment for the present ith patient, we add $G - u_{rt}$ balls of the same kind by which the patient was treated and u_{rt} balls of the opposite kind in the urn. This means for every $\{u_{rt}\}$ corresponding to every past $\{y_{rt}\}$ we add G balls to the urn before the selection of the treatment is made for the ith patient.

Now one may update the urn combination in light of the past response y_{rt} and the condition of the corresponding covariates u_{rt}, for $r = 1, \ldots, i-1$, $1 \le t \le \min(T, i-r)$, by adding $\{(G - u_{rt}) + y_{rt}\tau\}$ balls of the same kind by which the rth patient was treated, and $\{u_{rt} + (1 - y_{rt})\tau\}$ balls of the opposite kind in the urn. Thus for every available response of the past patients, we add $G + \tau$ balls in the urn, G balls for the covariate dependent condition or grade and τ balls for the response. We refer to this design as simple longitudinal play-the-winner (SLPW) design. We are now ready to use this SLPW design to compute the treatment selection probability for the ith patient as follows.

Recall that the probability that the first patient is assigned to treatment A is $w_1 = \Pr(\delta_1 = 1) = 1/2$. As mentioned earlier, at this stage we have α balls of each type in the urn. Now to compute w_i for $i = 2, \ldots, K$, it is necessary in the present longitudinal set up to derive its formula for two cases, first for the case when $2 \le i \le T$ and then for $i > T$, T being the total number of repeated responses recorded from each patient.

Case 1. $2 \le i \le T$

As the selection of the ith patient is made at the ith time point, by this time, the $(i-1)$th patient has yielded one response, the $(i-2)$th patient has yielded 2 responses

and so on. Thus, at this treatment selection stage for the ith patient, there are

$$2\alpha + \sum_{r=1}^{i-1}\sum_{t=1}^{i-r}(G+\tau) = 2\alpha + \frac{1}{2}i(i-1)(G+\tau) = n_{i-1}^* \tag{3.4}$$

balls in total in the urn. Among these balls, there are

$$\alpha + \sum_{r=1}^{i-1}\sum_{t=1}^{i-r}[\delta_r\{(G-u_{rt})+y_{rt}\tau\}+(1-\delta_r)\{u_{rt}+(1-y_{rt})\tau\}] = n_{i-1,1}^*(y_H) \tag{3.5}$$

balls of first type, where y_H indicates the history of responses from the past $i-1$ patients. Consequently, given y_H, the conditional probability that $\delta_i = 1$ is given by

$$w_i = \Pr(\delta_i = 1|y_H) = \frac{n_{i-1,1}^*(y_H)}{n_{i-1}^*}, \tag{3.6}$$

by (3.4) and (3.5).

Case 2. $i > T$

Under this case, at the treatment selection stage for the ith patient, there are

$$2\alpha + \sum_{r=1}^{i-T}\sum_{t=1}^{T}(G+\tau) + \sum_{r=i-T+1}^{i-1}\sum_{t=1}^{i-r}(G+\tau) = \tilde{n}_{i-1} \tag{3.7}$$

balls in total in the urn. Among these balls, there are $\tilde{n}_{i-1,1}(y_H)$ balls of first type, where

$$\begin{aligned}
\tilde{n}_{i-1,1}(y_H) &= \alpha + \sum_{r=1}^{i-T}\sum_{t=1}^{T}[\delta_r\{(G-u_{rt})+y_{rt}\tau\}+(1-\delta_r)\{u_{rt}+(1-y_{rt})\tau\}] \\
&\quad + \sum_{r=i-T+1}^{i-1}\sum_{t=1}^{i-r}[\delta_r\{(G-u_{rt})+y_{rt}\tau\} \\
&\quad +(1-\delta_r)\{u_{rt}+(1-y_{rt})\tau\}].
\end{aligned} \tag{3.8}$$

Consequently, one can obtain the design weights w_i as

$$w_i = \Pr(\delta_i = 1|y_H) = \frac{\tilde{n}_{i-1,1}(y_H)}{\tilde{n}_{i-1}}, \tag{3.9}$$

by (3.7) and (3.8).

3.3   Performance of the Proposed Adaptive Design

3.3.1   Limiting Behavior of Design Weights w_i

Note that it follows from (3.9) that as $i \to \infty$,

$$\frac{w_{i+1}}{w_i} \to 1.$$

Again the sequence $\{w_i, i \geq 1\}$ is bounded by 0 from the left and by 1 from the right. Hence there exists a subsequent $w_{k(i)}$ which is convergent. Suppose that it converges to ω. Then from the above limiting result, we have

$$\underset{i \to \infty}{\text{limit}} \ \frac{w_{k(i)+1}}{w_{k(i)}} = 1,$$

implying for some $\epsilon > 0$,

$$\omega(1 - \epsilon) \leq \text{limit inf } w_{k(i)+1} \leq \text{limit sup } w_{k(i)+1} \leq \omega(1 + \epsilon),$$

and hence

$$\text{limit sup } w_{k(i)+1} - \text{limit inf } w_{k(i)+1} \leq 2\omega\epsilon.$$

Since ϵ is arbitrary, we conclude that $\{w_i, i \geq 1\}$ is convergent. Suppose that it converges to ω^*.

To have some feelings about ω^*, we now make an attempt to derive a closed form formula for this convergent property. Let $p_{rj}^* = E(Y_{rj} | \delta_r, \ldots, \delta_1) = \frac{\exp(x_{rj}^{*\prime}\beta)}{1 + \exp(x_{rj}^{*\prime}\beta)}$ be the conditional probability for the binary response y_{rj} given the treatment $\delta_r, \ldots, \delta_1$. Further for $\delta_r = 1$, let p_{rj}^* reduce to p_{rj1} and to p_{rj2} for $\delta_r = 0$. At this stage we assume that, as $i \to \infty$,

$$(1) \quad \frac{1}{iT} \sum_{r=1}^{i-T} \sum_{j=1}^{T} p_{rj1} \to \pi_1, \quad (2) \quad \frac{1}{iT} \sum_{r=1}^{i-T} \sum_{j=1}^{T} p_{rj2} \to \pi_2, \quad (3) \quad \frac{1}{iT} \sum_{r=1}^{i-T} \sum_{j=1}^{T} u_{rj} \to u^*.$$

Next,

$$\frac{1}{iT} \sum_{r=1}^{i-T} \sum_{j=1}^{T} p_{rj1} w_r - \pi_1 \omega^* = \frac{1}{iT} \sum_{r=1}^{i-T} \sum_{j=1}^{T} p_{rj1}(w_r - \omega^*)$$

$$+ \omega^* \left[\frac{1}{iT} \sum_{r=1}^{i-T} \sum_{j=1}^{T} p_{rj1} - \pi_1 \right] \to 0,$$

as $i \to \infty$. It then follows that

$$(4) \quad \frac{1}{iT} \sum_{r=1}^{i-T} \sum_{j=1}^{T} p_{rj1} w_r \to \pi_1 \omega^*, \quad (5) \quad \frac{1}{iT} \sum_{r=1}^{i-T} \sum_{j=1}^{T} p_{rj2} w_r \to \pi_2 \omega^*,$$

$$(6) \quad \frac{1}{iT} \sum_{r=1}^{i-T} \sum_{j=1}^{T} u_{rj} w_r \to u^* \omega^*.$$

Using the above limiting results from (1) to (6) in (3.9), one obtains

$$\omega^* = \frac{1}{G+\tau} \left[(G - u^* + \pi_1 \tau) \omega^* + (u^* + (1 - \pi_2)\tau)(1 - \omega^*) \right],$$

yielding

$$\omega^* = \frac{u^* + (1 - \pi_2)\tau}{2u^* + (2 - \pi_1 - \pi_2)\tau},$$

which is primarily a function of τ. In fact, this ω^* is the limiting value of the probability of allocation to treatment A. This can be viewed as the limiting proportion of allocation to treatment A in this adaptive allocation scheme. For example, if $u^* = 2.0$, $\pi_1 = 0.8, \pi_2 = 0.2$, and $\tau = 2.0$, ω^* reduces to 0.6. Similarly, for $u^* = 2.0$, $\pi_1 = 0.8$, $\pi_2 = 0.2$, and $\tau = 4.0$, ω^* reduces to 0.65. Remark that $\omega^* > 0.5$ indicates that more study subjects will be assigned to the better treatment A.

3.3.2 Allocation performance of the proposed design

In the last subsection, we have computed the limiting value of w_i as $i \to \infty$. As in practice, a large but limited number of patients are considered in a clinical trial study, we examine the performance of the proposed adaptive design for various K as large as $K = 200$, where K is the total number of patients involved in the clinical trial experiment. More specifically, we consider $K = 75, 100,$ and 200. The performance of the proposed design will be examined through a simulation study. To be specific, we will consider 1000 simulations and examine the distribution of $\delta_s = \sum_{i=1}^{K} \delta_i$ where $w_i = \Pr(\delta_i = 1 | y_H)$ are the design weights defined in § 3.2.1. Note that the longitudinal adaptive design proposed in § 3.2.1 is expected to assign more subjects to the better treatment. For this to happen, $\delta_s = \sum_{i=1}^{K} \delta_i$, say, has to be greater than $K/2$.

For the purpose of examining the performance of δ_s, we require to compute w_i for $i = 2, \ldots, K$ by using (3.6) or (3.9). The computation of w_i $(i = 2, \ldots, K)$ however requires knowledge of the past responses y_{rt} $(r = 1, \ldots, i-1; 1 \leq t \leq \min(T, i-r))$ as well as $x_{rt}^* = (\delta_r, x_{rt2}, \ldots, x_{rtp})'$ where δ_r is known and $x_{rt2}, \ldots, x_{rtp}$ are known prognostic factors. To ensure that A is the better treatment we choose $\beta_1 = 1.5$, for example. The regression coefficients of three other prognostic covariates are chosen to be $\beta_2 = 0.0$, $\beta_3 = 0.2$, and $\beta_4 = 0.1$, for example, and the prognostic covariates themselves are chosen as in § 3.6. Remark that under the present set up, all y_{i1} $(i = 1, \ldots, K)$ are generated following the logistic binary model given by

$$\Pr(y_{i1} = 1 | \delta_i, \ldots, \delta_1) = \frac{\exp(x_i^{*'} \beta)}{1 + \exp(x_i^{*'} \beta)}$$

$$= p_i^*, \text{ (say)},$$

assuming that $x_{it} = x_i$ for all $t = 1, \ldots, T$. However, for a given i, to generate y_{it} for $2 \leq t \leq T$, one must ensure that $y_{i1}, \ldots, y_{iT}$ satisfy the underlying longitudinal correlation structure appropriate for repeated binary data. For example, if the repeated responses follow AR(1) binary correlation structure with correlation parameter ρ (Zeger, Liang and Self (1985)), then one generates $y_{i2}, \ldots, y_{iT}$ as follows. If $y_{i1} = 0$, then generate y_{i2} with probability $p_i^*(1 - \rho)$, otherwise generate y_{i2} with probability $p_i^* + \rho(1 - p_i^*)$. Continue this to get y_{i3} depending only on y_{i2} and so on. This assures that the lag $\ell = 1, 2, \ldots, T - 1$ correlation between y_{it} and $y_{i,t+\ell}$ is ρ^ℓ. In the present simulation study, we choose $T = 4$ and $\rho = 0.3, 0.5, 0.7,$ and 0.9 to represent small as well as large correlations.

Note that the computation of w_i $(i = 1, \ldots, K)$ further requires the knowledge of α, G, τ, and the non-stochastic function u_{rt} for $(r = 1, \ldots, i-1; 1 \leq t \leq \min(T, i-r))$. As the u_{rt} functions are constructed from the prognostic covariates, we also refer to § 3.6 for their choices along with the choices for the covariates. The choice of u_{rt} functions however yielded $G = 3.0$. Next for simplicity, we choose $\alpha = 1.0$. As far as the choice of τ is concerned, we recall from the previous subsection that the design weights w_i in the limit primarily depend on τ. Consequently, we now choose two

values of τ, namely $\tau = 2.0$ and 4.0 in order to see the effect of small and large τ values on the construction of w_i (see § 3.2.1).

Note that once w_i is known, the corresponding δ_i is generated from a binary distribution with probability w_i. As mentioned earlier, to understand whether the proposed design can allocate more individuals to the better treatment, we now compute $\delta_s = \sum_{i=1}^{K} \delta_i$ under each of the 1000 simulations. The simulated mean and standard deviation of δ_s for various choices of K, τ, and ρ are shown in Table A.1.

It is clear from Table A.1 that irrespective of correlation values, the proposed design allocated more individuals to the better treatment A. For example, for $K = 75$, $\tau = 2.0$, and $\rho = 0.9$, 44 individuals out of 75 were assigned to treatment A. Thus about 59% individuals were assigned to the better treatment. Similarly for $K = 200$, $\tau = 2.0$, and $\rho = 0.9$, 117 individuals were allocated to treatment A which is about 59%. Remark that allocation gets better for larger τ. For example, for the same $K = 200$, and $\rho = 0.9$, the allocated number of individuals to treatment A is 125 for the case with $\tau = 4.0$, whereas the allocated number is 117 for $\tau = 2.0$. Thus the proposed design works well in assigning more subjects to a better treatment.

Note that we have considered $\delta_s = \sum_{i=1}^{K} \delta_i > \frac{K}{2}$ as the criterion to examine the performance of the proposed allocation scheme. Although the mean of δ_s statistic was found to reflect the goodness of the scheme, this approach however produced relatively large standard error. As a remedy it seems that $\bar{\delta}_s = \frac{\sum_{i=1}^{K} \delta_i}{K}$ would be a much more stable statistic whose mean will be greater than 0.5, but standard error will be relatively smaller.

3.3.3 Expected Design Weights Under Binary Models

Recall that the adaptive design weights w_i are given by (3.6) for $2 \leq i \leq T$ and by (3.9) for $i > T$, T being small in the present longitudinal set up. The design weights given by (3.9) satisfy $\{\frac{w_{i+1}}{w_i}\} \to 1$ as $i \to \infty$. Further it has been shown in § 3.3.1 that in the limit as $i \to \infty$, $w_i \to \omega^*$, which is primarily a function of τ. However, as w_i is a function of the past responses y_{rt} and the covariates x_{rt}

for $(r = 1, \ldots, i - 1; 1 \leq t \leq \min(T, i - r))$, it may be of interest to examine the difference between w_i and its expected value $E(w_i) = w_{i0}$, say, under the true model that generates all the past responses y_{rt}. In the present set up, we consider a correlated binary model for all y_{rt}, $(r = 1, \ldots, i - 1; 1 \leq t \leq \min(T, i - r))$. This issue of examining the difference between w_i and w_{i0} is particularly important in a situation where one would like to use w_i as an estimator of w_{i0} in any statistical analysis, such as in estimation of β, the effects of covariates. For this purpose, for all $i = 1, \ldots, K$ (with $K = 75, 100,$ or 200) we will compare the w_i computed using the sample binary responses as in the last sub-section with their expected values w_{i0} where w_{i0} is computed as

$$w_{i0} = E_{\delta_1} E_{\delta_2 | \delta_1} \ldots E_{\delta_i | \delta_1, \ldots, \delta_{i-1}} (\delta_i). \tag{3.10}$$

Since $E_{\delta_i | \delta_1, \ldots, \delta_{i-1}}(\delta_i) = \Pr(\delta_i = 1 | \delta_{i-1}, \ldots, \delta_1) = w_i$, where the w_i's are defined in (3.6) and (3.9), it then follows that for $r = 1, \ldots, i - 1$,

$$
\begin{aligned}
E(\delta_r Y_{rt}) &= E_{\delta_1} E_{\delta_2 | \delta_1} \ldots E_{\delta_r | \delta_1, \ldots, \delta_{r-1}} E(\delta_r Y_{rt} | \delta_r, \ldots, \delta_1) \\
&= E_{\delta_1} E_{\delta_2 | \delta_1} \ldots E_{\delta_r | \delta_1, \ldots, \delta_{r-1}} (\delta_r p_{rt}^*),
\end{aligned}
\tag{3.11}
$$

where $p_{rt}^* = E(Y_{rt} | \delta_r, \ldots, \delta_1) = \frac{\exp(x_{rt}^{*'} \beta)}{1 + \exp(x_{rt}^{*'} \beta)}$ with $x_{rt}^* = (\delta_r, x_{rt2}, \ldots, x_{rtp})'$. Suppose that $z_{rt} = x_{rt}^* |_{\delta_r=1}$ and $z_{rt}^* = x_{rt}^* |_{\delta_r=0}$. The expectation in (3.11) then reduces to

$$E(\delta_r Y_{rt}) = w_{r0} p_{rt1}, \tag{3.12}$$

where $p_{rt1} = \frac{\exp(z_{rt}' \beta)}{1 + \exp(z_{rt}' \beta)}$. By similar calculation, it can be shown that

$$E(1 - \delta_r)(1 - Y_{rt}) = (1 - w_{r0})(1 - p_{rt2}), \tag{3.13}$$

where $p_{rt2} = \frac{\exp(z_{rt}^{*'} \beta)}{1 + \exp(z_{rt}^{*'} \beta)}$. Now by applying (3.12) and (3.13) to (3.10), it follows from (3.6) that for $2 \leq i \leq T$, the unconditional expectation of w_i is given as

$$w_{i0} = \frac{\alpha + \sum_{r=1}^{i-1} \sum_{t=1}^{i-r} [\{(G - u_{rt}) + p_{rt1} \tau\} w_{r0} + \{u_{rt} + (1 - p_{rt2}) \tau\}(1 - w_{r0})]}{2\alpha + \frac{1}{2} i(i - 1)(G + \tau)}. \tag{3.14}$$

Similarly, it follows from (3.9) that for $i > T$, the unconditional expectation of w_i is given by

$$
\begin{aligned}
w_{i0} \;=\; & \left\{ 2\alpha + (G+\tau)T\left(i - \frac{T+1}{2}\right) \right\}^{-1} \left[\alpha + \sum_{r=1}^{i-T}\sum_{t=1}^{T}\{(G - u_{rt} + p_{rt1}\tau)w_{r0} \right. \\
& + (u_{rt} + (1 - p_{rt2})\tau)(1 - w_{r0})\} + \sum_{r=i-T+1}^{i-1}\sum_{t=1}^{i-r}\{((G - u_{rt}) + p_{rt1}\tau)w_{r0} \\
& \left. + (u_{rt} + (1 - p_{rt2})\tau)(1 - w_{r0})\} \right].
\end{aligned}
\tag{3.15}
$$

Note that for $i = 1, \ldots, K$, w_{i0} in (3.14) and (3.15) are the unconditional expectations of w_i under the present binary model. Further note that although β is unknown, it remains the same all through the experiment. In the next section, we will consider the estimation of this unknown parameter β. In this sub-section, we however compare the w_i values with their corresponding w_{i0} values for known β as well as for other given parameters such as τ and ρ. It is clear that w_i is a function of binary responses for the past $i - 1$ patients that we simulate in the manner similar to that of § 3.3.2. Here, the simulations of the binary responses depend on β and ρ parameters of the correlated binary model. As opposed to w_i, w_{i0} is however not dependent on the responses, rather it directly depends on the parameters of the underlying binary model such as β. For given β, τ, ρ, α, G, and non-stochastic function u_{rt} as given in the last § 3.3.2, we now compute the w_i and w_{i0} values for all $i = 1, \ldots, K$, with $K = 75, 100,$ and 200. The graphs for w_i and w_{i0} are shown in Figure C.1, C.2, and C.3 for $K = 75, 100,$ and 200 respectively. In each of these three figures, we show the graphs for two values of $\tau = 2.0$ and 4.0 and for two values of $\rho = 0.5$ and 0.9.

Remark that as w_{i0} is the expected value of w_i under the binary model, the value of w_{i0} changes with regard to i ($i = 1, \ldots, K$) only through the prognostic factors x_{rt}^* ($r = 1, \ldots, i-1; 1 \le t \le \min(T, i-r)$) and the non-stochastic u_{rt} functions constructed using x_{rt}^*. For convenience of numerical computations, as in § 3.6, we generated the prognostic factors $x_{rt2}, x_{rt3},$ and x_{rt4} following certain suitable probability models. This leads to three different sets of prognostic factors as well as u_{rt} functions

for three choices of $K = 75, 100,$ and 200. Consequently, for given values of β, τ, α, and G, Figures C.1 to C.3 exhibit three similar but slightly different graphs for w_{i0} for three choices of $K = 75, 100,$ and 200. As opposed to w_{i0}, the value of w_i changes depending on the past binary responses y_{rt} $(r = 1, \ldots, i - 1; 1 \leq t \leq \min(T, i - r))$ which are likely to be different under different simulations. They are also different because they are generated with different values of longitudinal correlations such as $\rho = 0.9$ and 0.5. For a given i, the average of w_i over 1000 simulations are displayed in Figures C.1-C.3 for $K = 75, 100,$ and 200 respectively. It is clear that for given values of τ and ρ, the value of w_i converges to w_{i0} for large $i \leq K$. The convergence is quite satisfactory for large i, specially for the values of i close to large K, such as for $100 \leq i \leq K$, where $K = 200$. This happens irrespective of the choices of the values of τ and ρ, although the convergence is quicker for larger $\tau = 4.0$ and smaller $\rho = 0.5$ as compared to smaller $\tau = 2.0$ and larger $\rho = 0.9$, respectively. For small values of K such as $K = 75$, there always remains a difference between w_i and w_{i0} even for large i. These differences however get smaller as the value of τ gets larger and the value of ρ gets smaller. For $K = 100$, the situation is better as compared to $K = 75$, which is expected.

3.4 WGQL Approaches For Parameter Estimation Including the Treatment Effect

Recall that in § 3.2, we have proposed an adaptive longitudinal design which assigns the ith $(i = 1, \ldots, K)$ individual to treatment A (between A and B) with probability w_i given by (3.6) for $2 \leq i \leq T$ and by (3.9) for $i > T$. By treating A as the better treatment between A and B, we have also examined the performance of the proposed design by a simulation study and found that the proposed design allocates more study subjects to the better treatment. In practice, however one may be interested to know the effects of the treatment as well as the effects of other prognostic covariates. This means that one is interested to know the regression parameter vector β which invisibly

contributes to generate binary responses y_{rt} $(r = 1, \ldots, i - 1; 1 \leq t \leq \min(T, i - r))$ necessary for the construction of w_i. Note that the longitudinal correlations of the repeated responses have to be taken into account in estimating this β parameter consistently. We provide a weighted generalized quasi-likelihood (WGQL) approach in the following sub-section for such consistent estimation of the regression effects.

3.4.1 WGQL Estimation Approach for Regression Effects

Let $y_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT})'$ be a $T \times 1$ vector of repeated binary responses for the ith $(i = 1, \ldots, K)$ individual. Note that this individual is assigned to treatment A with probability $w_i = \Pr(\delta i = 1 | y_H)$ given by (3.6) for $2 \leq i \leq T$ and by (3.9) for $i > T$. Here, y_{it} is the tth response of the ith individual. Further note that since w_i depends on the responses from the past $i - 1$ patients, the unconditional expectation of Y_{it} may be computed as

$$E(Y_{it}) = E_{\delta_1} E_{\delta_2 | \delta_1} \ldots E_{\delta_i | \delta_1, \ldots, \delta_{i-1}} E(Y_{it} | \delta_i, \delta_{i-1}, \ldots, \delta_1). \tag{3.16}$$

It then follows by (3.10)-(3.13) that

$$
\begin{aligned}
E(Y_{it}) &= E_{\delta_1} E_{\delta_2 | \delta_1} \ldots E_{\delta_i | \delta_1, \ldots, \delta_{i-1}} (p_{it}^*) \\[2mm]
&= w_{i0} p_{it1} + (1 - w_{i0}) p_{it2} \\[2mm]
&= p_{it}, \text{ say}, \tag{3.17}
\end{aligned}
$$

where w_{i0} is given by (3.14) for $2 \leq i \leq T$ and by (3.15) for $i > T$, and p_{it1} and p_{it2} are given as

$$p_{it1} = \frac{\exp(z_{it}'\beta)}{1 + \exp(z_{it}'\beta)} \text{ and } p_{it2} = \frac{\exp(z_{it}^{*'}\beta)}{1 + \exp(z_{it}^{*'}\beta)}, \tag{3.18}$$

respectively, with $z_{it}' = (\delta_i, x_{it2}, \ldots, x_{itp})|_{\delta_i = 1}$ and $z_{it}^{*'} = (\delta_i, x_{it2}, \ldots, x_{itp})|_{\delta_i = 0}$. Let $p_i = (p_{i1}, \ldots, p_{it}, \ldots, p_{iT})'$ where p_{it} is given by (3.17) for all $i = 1, \ldots, K$, so that

$$
\begin{aligned}
E(Y_i) &= E(Y_{i1}, \ldots, Y_{iT})' \\[2mm]
&= p_i. \tag{3.19}
\end{aligned}
$$

Next, we compute the unconditional covariance matrix of Y_i where $Y_i = (Y_{i1}, \ldots, Y_{it}, \ldots, Y_{iT})'$. For this purpose, following Sutradhar and Das (1999, §3) we now assume that conditional on $\delta_i, \ldots, \delta_1$, the repeated responses Y_{it} and Y_{iv} at two time points t and v $(t, v = 1, \ldots, T)$ have longitudinal correlation structure given by

$$corr(Y_{it}, Y_{iv}|\delta_i, \ldots, \delta_1) = \rho_{|(t-v)|}, \tag{3.20}$$

where $\rho_{|t-v|}$ denotes the lag $|t - v|$ auto-correlation. Note that the auto-correlation structure considered in (3.20) is general as it accommodates the Gaussian AR(1), MA(1), and exchangeable type auto-correlation structures as special cases. It then follows that the unconditional covariance between Y_{it} and Y_{iv} is given by

$$
\begin{aligned}
cov(Y_{it}, Y_{iv}) &= E_{\delta_1} E_{\delta_2|\delta_1} \ldots E_{\delta_i|\delta_1, \ldots, \delta_{i-1}} \left[cov(Y_{it}, Y_{iv})|\delta_i, \delta_{i-1}, \ldots, \delta_1) \right] \\
&\quad + cov_{\delta_i, \ldots, \delta_1} \left[E(Y_{it}|\delta_i, \delta_{i-1}, \ldots, \delta_1), E(Y_{iv}|\delta_i, \delta_{i-1}, \ldots, \delta_1) \right] \\
&= E_{\delta_1} E_{\delta_2|\delta_1} \ldots E_{\delta_i|\delta_1, \ldots, \delta_{i-1}} \left[\rho_{|t-v|} \{p_{it}^* q_{it}^* p_{iv}^* q_{iv}^*\}^{\frac{1}{2}} \right] \\
&\quad + cov_{\delta_i, \ldots, \delta_1} \left[p_{it}^*, p_{iv}^* \right], \tag{3.21}
\end{aligned}
$$

where $E(Y_{it}|\delta_i, \ldots, \delta_1) = p_{it}^* = \frac{\exp(x_{it}^{*'}\beta)}{1+\exp(x_{it}^{*'}\beta)}$ and $var(Y_{it}|\delta_i, \ldots, \delta_1) = p_{it}^* q_{it}^*$ by (3.8) with $q_{it}^* = 1 - p_{it}^*$. After some algebra, the equation (3.21) reduces to

$$
\begin{aligned}
cov(Y_{it}, Y_{iv}) &= \rho_{|t-v|} \left[w_{i0}\{p_{it1}q_{it1}p_{iv1}q_{iv1}\}^{\frac{1}{2}} + (1 - w_{i0})\{p_{it2}q_{it2}p_{iv2}q_{iv2}\}^{\frac{1}{2}} \right] \\
&\quad + w_{i0}\{p_{it1}p_{iv1}\} + (1 - w_{i0})\{p_{it2}p_{iv2}\} - p_{it}p_{iv} \\
&= \sigma_{itv}, \text{ say,} \tag{3.22}
\end{aligned}
$$

When $t = v$, the covariance σ_{itv} in (3.22) reduces to the variance of Y_{it} given by

$$var(Y_{it}) = \sigma_{itt} = p_{it}q_{it}, \tag{3.23}$$

where $q_{it} = 1 - p_{it}$ with p_{it} as defined in (3.17). Let Σ_i denote the covariance matrix of Y_i, which may be expressed as

$$\Sigma_i = cov(Y_i) = (\sigma_{itv}), \tag{3.24}$$

for $t, v = 1, \ldots, T$, where σ_{itv} are given by (3.22) and (3.23).

Next for known Σ_i, one may write the quasi-likelihood (QL) estimating equation for β as

$$\sum_{i=1}^{K} \frac{\partial p_i'}{\partial \beta} \Sigma_i^{-1} (y_i - p_i) = 0, \tag{3.25}$$

(McCullagh (1983), Wedderburn (1979)), where p_i is the $T \times 1$ vector given by (3.19) and $\frac{\partial p_i'}{\partial \beta}$ is the $p \times T$ first derivative vector of p_i' with respect to β. Note that, in practice, however Σ_i is unknown and it is a function of w_{i0}, β, and ρ, where w_{i0} again depends on β. Also p_i vector is a function of w_{i0} which contains β. Now in solving (3.25) for β, in the spirit of the GEE (generalized estimating equations) approach (see Liang and Zeger (1986)) we re-express the QL estimating equation (3.25) as

$$\sum_{i=1}^{K} \frac{\partial p_i'(w_{i0})}{\partial \beta} \Sigma_i^{-1}(w_{i0}, \hat{\rho}) \, (y_i - p_i(w_{i0})) = 0, \tag{3.26}$$

and we refer to this as the weighted generalized quasi-likelihood (WGQL) estimating equation for β, where $\hat{\rho}$ is a consistent estimator for the longitudinal correlation parameter ρ.

Note that to solve (3.26) for β, one may consider the following **three** scenarios: **first**, for some initial β, w_{i0} is known in the spirit of GEE; **second**, w_{i0} is unknown but it can be replaced with the adaptive design weight w_i as $E(w_i) = w_{i0}$; **third**, w_{i0} is an unknown function of β. The estimator of β as the solution of (3.26) under the above three scenarios will be denoted by $\hat{\beta}_{WGQL1}$, $\hat{\beta}_{WGQL2}$, and $\hat{\beta}_{WGQL3}$ respectively. These solutions may be obtained by using iterative equations

$$\hat{\beta}_{(m+1)GQL} = \hat{\beta}_{(m)GQL} + \left[\sum_{i=1}^{K} \frac{\partial p_i'(w_{i0})}{\partial \beta} \Sigma_i^{-1}(w_{i0}, \hat{\rho}) \frac{\partial p_i(w_{i0})}{\partial \beta'} \right]_m^{-1}$$

$$\times \left[\sum_{i=1}^{K} \frac{\partial p_i'(w_{i0})}{\partial \beta} \Sigma_i^{-1}(w_{i0}, \hat{\rho}) \, (y_i - p_i(w_{i0})) \right]_m, \tag{3.27}$$

where $\hat{\beta}_{(m)GQL}$ is the value of β at the mth iteration and $[\cdot]_m$ denotes that the expression within brackets is evaluated at $\hat{\beta}_{(m)GQL}$. Remark that to compute $\hat{\beta}_{WGQL1}$ and

$\hat{\beta}_{WGQL2}$, the first derivative vector $\frac{\partial p_i'(w_{i0})}{\partial \beta}$ has the formulas

$$
\begin{aligned}
\frac{\partial p_i'(w_{i0})}{\partial \beta} &= w_{i0}\frac{\partial p_{i1}'}{\partial \beta} + (1 - w_{i0})\frac{\partial p_{i2}'}{\partial \beta} \\[2mm]
&= w_{i0}Z_i'A_{i1} + (1 - w_{i0})Z_i^{*'}A_{i2} \\[2mm]
&= B_i, \text{ say,}
\end{aligned}
\tag{3.28}
$$

and

$$
\begin{aligned}
\frac{\partial p_i'(w_{i0})}{\partial \beta} &= \frac{\partial p_i'(w_{i0})}{\partial \beta}\Big|_{w_{i0}=w_i} = w_i Z_i'A_{i1} + (1 - w_i)Z_i^{*'}A_{i2} \\[2mm]
&= C_i, \text{ say,}
\end{aligned}
\tag{3.29}
$$

respectively, where $Z_i' = (z_{i1}, \ldots, z_{it}, \ldots, z_{iT})$ and $Z_i^{*'} = (z_{i1}^*, \ldots, z_{it}^*, \ldots, z_{iT}^*)$ are $p \times T$ matrices, $A_{i1} = \text{diag}[p_{i11}q_{i11}, \ldots, p_{iT1}q_{iT1}]$, and $A_{i2} = \text{diag}[p_{i12}q_{i12}, \ldots, p_{iT2}q_{iT2}]$ are $T \times T$ matrices. Moreover, in (3.28) and (3.29), $p_{i1} = (p_{i11}, \ldots, p_{it1}, \ldots, p_{iT1})'$ and $p_{i2} = (p_{i12}, \ldots, p_{it2}, \ldots, p_{iT2})'$, with $p_{it1} = \frac{\exp(z_{it}'\beta)}{1+\exp(z_{it}'\beta)}$ and $p_{it2} = \frac{\exp(z_{it}^{*'}\beta)}{1+\exp(z_{it}^{*'}\beta)}$.

To compute $\hat{\beta}_{WGQL3}$, one may simplify the first derivative vector as

$$
\begin{aligned}
\frac{\partial p_i'(w_{i0})}{\partial \beta} &= w_{i0}\frac{\partial p_{i1}'}{\partial \beta} + (1 - w_{i0})\frac{\partial p_{i2}'}{\partial \beta} + \frac{\partial w_{i0}}{\partial \beta}(p_{i1} - p_{i2})' \\[2mm]
&= D_i, \text{ say,}
\end{aligned}
\tag{3.30}
$$

where for $2 \leq i \leq T$,

$$
\frac{\partial w_{i0}}{\partial \beta} = \frac{\sum_{r=1}^{i-1}\sum_{t=1}^{i-r}\left\{(p_{rt1}q_{rt1}z_{rt}\tau w_{r0}) - (p_{rt2}q_{rt2}z_{rt}^*\tau(1-w_{r0}))\right\}}{2\alpha + \frac{1}{2}i(i-1)(G+\tau)},
\tag{3.31}
$$

and for $i > T$

$$
\begin{aligned}
\frac{\partial w_{i0}}{\partial \beta} &= \left\{2\alpha + (G+\tau)T\left(i - \frac{T+1}{2}\right)\right\}^{-1} \\[2mm]
&\quad \times \left[\sum_{r=1}^{i-T}\sum_{t=1}^{T}\left\{(p_{rt1}q_{rt1}z_{rt}\tau w_{r0}) - (p_{rt2}q_{rt2}z_{rt}^*\tau(1-w_{r0}))\right\}\right.
\end{aligned}
$$

$$+ \sum_{r=i-T+1}^{i-1} \sum_{t=1}^{i-r} \{(p_{rt1}q_{rt1}z_{rt}\tau w_{r0}) - (p_{rt2}q_{rt2}z_{rt}^{*}\tau(1 - w_{r0}))\} \Bigg].$$

(3.32)

This completes the construction of the estimating equation given in (3.26) under the above three scenarios.

Note that the estimating equations for β require knowledge of $\hat{\rho} = (\hat{\rho}_1, \ldots, \hat{\rho}_\ell, \ldots, \hat{\rho}_{T-1})$ where $\hat{\rho}_\ell$ ($\ell = 1, \ldots, T-1$) may be obtained consistently as in § 3.5 by using the so-called method of moments. Next, under some regularity conditions, it may be shown (Liang and Zeger (1986)) that for large K, $\hat{\beta}_{WGQL1}$ and $\hat{\beta}_{WGQL3}$ have asymptotically p–dimensional normal distributions with mean β and covariance matrices given as

$$var(\hat{\beta}_{WGQL1}) = \left[\sum_{i=1}^{K} \frac{\partial p_i'(w_{i0})}{\partial \beta} \Sigma_i^{-1}(w_{i0}, \hat{\rho}) \frac{\partial p_i(w_{i0})}{\partial \beta'} \right]^{-1}$$

$$= \left[\sum_{i=1}^{K} B_i \Sigma_i^{-1}(w_{i0}, \hat{\rho}) B_i' \right]^{-1},$$

(3.33)

and

$$var(\hat{\beta}_{WGQL3}) = \left[\sum_{i=1}^{K} D_i \Sigma_i^{-1}(w_{i0}, \hat{\rho}) D_i' \right]^{-1},$$

(3.34)

respectively, where B_i and D_i are given in (3.28) and (3.30) respectively. By similar arguments, one may show that $\hat{\beta}_{WGQL2}$ also has an asymptotically normal distribution with mean vector β and a suitable covariance matrix which can be consistently estimated by

$$v\hat{a}r(\hat{\beta}_{WGQL2}) = \left[\sum_{i=1}^{K} C_i \Sigma_i^{-1}(w_i, \hat{\rho}) C_i' \right]^{-1},$$

(3.35)

where $\Sigma_i(w_i)$, for example, is obtained from $\Sigma_i(w_{i0})$ by replacing w_{i0} with its data based estimate w_i, and C_i is given by (3.29).

3.5 Moment Approach For Longitudinal Correlations

Note that the estimating equation (3.26) or equivalently, the iterative equation (3.27) requires knowledge of $\rho = (\rho_1, \ldots, \rho_\ell, \ldots, \rho_{T-1})'$. For a given value of the estimate of β, we now obtain a moment estimator $\hat{\rho}$, which is consistent for ρ. To be specific, for the computation of $\hat{\beta}_{WGQL1}$ and $\hat{\beta}_{WGQL3}$, $\hat{\rho}$ may be obtained by following Jowaheer and Sutradhar (2002, p. 394). That is, by (3.22), $\hat{\rho}_\ell$, the ℓth $(\ell = |t-v| = 1, \ldots, T-1)$ component of $\hat{\rho}$ has the formula given by

$$\hat{\rho}_\ell = \frac{\frac{\sum_{i=1}^{K}\sum_{|t-v|=\ell}[(y_{it}-p_{it})(y_{iv}-p_{iv})/K(T-\ell)}{\sum_{i=1}^{K}\sum_{t=1}^{T}[y_{it}-p_{it}]^2/KT} - \frac{\sum_{i=1}^{K}\sum_{|t-v|=\ell}[w_{i0}p_{it1}p_{iv1}+(1-w_{i0})p_{it2}p_{iv2}-p_{it}p_{iv}]/K(T-\ell)}{\sum_{i=1}^{K}\sum_{t=1}^{T}[p_{it}q_{it}]/KT}}{\frac{\sum_{i=1}^{K}\sum_{|t-v|=\ell}\left[w_{i0}\{p_{it1}q_{it1}p_{iv1}q_{iv1}\}^{\frac{1}{2}}+(1-w_{i0})\{p_{it2}q_{it2}p_{iv2}q_{iv2}\}^{\frac{1}{2}}\right]/K(T-\ell)}{\sum_{i=1}^{K}\sum_{t=1}^{T}[p_{it}q_{it}]/KT}}.$$

(3.36)

Similarly, for the computation of $\hat{\beta}_{WGQL2}$, the ℓth component of $\hat{\rho}$, i.e., $\hat{\rho}_\ell$ may be obtained from (3.36) by replacing w_{i0} with the data based adaptive design weight w_i $(i = 1, \ldots, K)$.

3.6 Performance of the WGQL Estimation Approaches: A Simulation Study

Recall from § 3.2 that in a clinical trial set up, the adaptive longitudinal design is constructed such that more study subjects are assigned to the better treatment. Once the clinical trial is over, one is then ready to compute the actual effects of the covariates (prognostic factors) including the effect of the treatment. Note that in the longitudinal study, the repeated responses are collected from the same individual. Consequently, it is important to take the correlations of the repeated data into account along with the adaptive design weights in estimating the regression effects. The WGQL approach discussed in § 3.4 takes these design weights as well as longitudinal correlations into account in estimating the effects of the covariates. In this section,

we conduct a simulation study to examine the performance of the WGQL estimator of β for different choices of cluster number (small and large), where β is the $p \times 1$ vector of effects of the covariates including the treatment.

3.6.1 Simulation Design

To understand the effects of small as well as large samples, we have chosen $K = 75, 100,$ and 200, where K is the total number of individuals in the adaptive longitudinal clinical study. Next, we choose $T = 4$, where T denotes the number of repeated responses collected from each of the K individuals. As far as the covariates are concerned, we choose $p = 4$ covariates: 1 treatment covariate and 3 others as prognostic factors. Let δ_i denote the treatment covariate so that $\delta_i = 1$ indicates that the ith patient is treated by the better treatment and $\delta_i = 0$ indicates otherwise. The other 3 covariates, that is, the prognostic factors are denoted by x_{it2}, x_{it3}, and x_{it4} for the ith individual at the tth $(t = 1, \ldots, T)$ data collection time. Note that the values of δ_i for all i $(i = 1, 2, \ldots, K)$ are determined based on the adaptive longitudinal design such that

$$\Pr(\delta_i = 1 | y_H) = w_i,$$

where w_i values are computed from (3.6) and (3.9) for $2 \leq i \leq T$ and $i > T$ cases respectively and y_H is the history of the responses from the past $i - 1$ patients. The prognostic factors are however chosen as follows.

We consider chronic disease condition of an incoming patient as the first prognostic factor denoted by x_{it2}. To generate x_{it2} for all i $(i = 1, 2, \ldots, K)$, we consider c_i as a binomial variable with parameters m and p, i.e., $c_i \sim b(m, p)$, where c_i represents the number of chronic diseases for the ith patient at his or her entry time to the clinical trial. We choose, for example, $m = 5$ and $p = \frac{1}{2}$. We then consider $x_{it2} = 0$ for $c_i = 0, 1$ and $x_{it2} = 1$ for $c_i = 2, 3, 4, 5$. Thus, the ith patient with a low occurrence of chronic disease has $x_{it2} = 0$ for all $t = 1, \ldots, T$. If the ith patient however enters to the trial with a large occurrence of chronic disease, then $x_{it2} = 1$ for all $t = 1, \ldots, T$.

To generate the 2 other prognostic factors, namely, x_{it3} and x_{it4}, we now consider

an age variable and generate an age between 20 and 80 from a uniform distribution. Next we create 6 age groups, namely $21-30, 31-40, \ldots, 71-80$ and define d_i as an ordinal variable such that $d_i = 1, 2, \ldots, 6$, where, for example, $d_i = 1$ indicates that the age of the ith patient belongs to the first age group 21-30. To generate x_{it3} and x_{it4}, we consider the merging of two consecutive age groups into one age group and obtain 3 age groups, namely, 21-40, 41-60, and 61-80, which may be referred to as the young, middle, and old age groups respectively. We now define $x_{it3} = 1$ and $x_{it4} = 0$ if the ith patient belongs to the young group 21-40, and $x_{it3} = 0$ and $x_{it4} = 1$ if the ith patient belongs to the middle age group, otherwise $x_{it3} = 0$ and $x_{it4} = 0$.

In order to compute the adaptive longitudinal design weights w_i (using 3.6 and 3.9) we also require a non-stochastic continuous quantity with domain $[0, G]$, say. More specifically, we require to construct $u_{rt} = \psi(x_{rt2}, x_{rt3}, x_{rt4})$, where u_{rt} measures the condition of covariates x_{rt2}, x_{rt3}, and x_{rt4} through the ψ function. Recall from § 3.2 that ψ function has to be chosen so that a larger value of u_{rt} implies the better condition of the rth $(r = 1, \ldots, i-1)$ patient. In the simulation study, we choose

$$u_{rt} = \frac{2}{c_r + 1} + \frac{1}{d_r},$$

for all t $(1 \leq t \leq min(T, i-r))$, where c_r is an implicit function of x_{rt2}, and similarly d_r is an implicit function of x_{rt3} and x_{rt4}. Note that as $c_r = 0, 1, \ldots, 5$ and $d_r = 1, 2, \ldots, 6$, it then follows that u_{rt} lies in the range of 0 to 3 yielding $G = 3$. Next, for simplicity we consider $\alpha = 1.0$, and two values of $\tau = 2$ and 4. Remark that as the limiting value of w_i mainly depends on τ as shown in § 3.3.1, we have considered two values of τ.

As far as the generation of y_{rt} $[r = 1, \ldots, i-1; 1 \leq t \leq min\{T, i-1\}]$ is concerned, these repeated responses for the rth individual, namely, $y_{r1}, \ldots, y_{r[min\{T,i-r\}]}$ are generated from a multivariate binary distribution, as explained in § 3.3.2, with ℓth lag correlation $\rho_\ell = \rho^\ell$ for selected values of ρ, namely, $\rho = 0.3, 0.5, 0.7$, and 0.9. We are now ready to compute the adaptive longitudinal design weights w_i by using (3.6) and (3.9). In generating the design weights w_i and binary responses y_{it} for all $i = 1, \ldots, K$ and all $t = 1, \ldots, T$, we have used the treatment effect parameter

β_1 as $\beta_1 = 1.50$. The other 3 regression parameters representing the effects of the prognostic factors were chosen to be $\beta_2 = 0.0$, $\beta_3 = 0.20$, and $\beta_4 = 0.10$, respectively.

3.6.2 Estimation Performance

The purpose of the simulation study is to examine the performance of the estimators of $\beta = (\beta_1, \ldots, \beta_p)'$ and $\rho = (\rho_1, \ldots, \rho_{T-1})'$ developed in § 3.4 and § 3.5. Remark that the estimating formulas for the regression parameters were constructed in § 3.4 under 3 scenarios, and the estimation approaches were named as $WGQL_1$, $WGQL_2$, and $WGQL_3$. We now conduct 1000 simulations and report the simulated means and standard deviations of estimators of $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)'$ and $\rho = (\rho_1, \rho_2, \rho_3)'$ parameters in Tables A.2, A.3, and A.4 for $K = 75, 100$, and 200, respectively. We have also computed the estimated standard errors for the estimators of regression parameters $\beta_1, \ldots, \beta_4$ by using the asymptotic variance formula for $\hat{\beta}$ given by (3.33)-(3.35). The simulated means of these estimated standard errors are also reported in the same tables for $K = 75, 100$, and 200.

It is clear from Tables A.2, A.3 and A.4 that all three approaches, namely, $WGQL_1$, $WGQL_2$, and $WGQL_3$ perform well in estimating longitudinal correlations. For example, for $K = 100$, $\tau = 4.0$, and $\rho = 0.7$, the robust estimating formula (3.36) yields lag 1,2,and 3 correlations as 0.683, 0.460, and 0.304 respectively, whereas the true lag correlations are 0.7, 0.49, and 0.34 respectively. The simulated standard errors of these correlation estimators are reasonably small. For the above case, these simulated standard errors are 0.069, 0.117, and 0.149 respectively. With regard to the estimation of the effects of the prognostic factors for given $K = 75, 100$, and 200, all three approaches yield almost the same estimates. For example, for $K = 100$, $\tau = 4.0$, and $\rho = 0.7$, the $WGQL_1$, $WGQL_2$, and $WGQL_3$ approaches yield -0.05, 0.23, and 0.09; 0.03, 0.28, and 0.15; and 0.07, 0.21, and 0.09, respectively for β_2, β_3, and β_4, the true parameter values being 0.0, 0.2, and 0.1 respectively.

For the estimation of the treatment effect , the performances of these three approaches are not quite the same. All three approaches appear to produce downward

biases for $K = 75$. When K increases to 100 or 200, they appear to yield upward biases except for the $WGQL_2$ approach with $K = 100$. Note however that the biases for large sample cases are reasonably small. Nevertheless, we have computed the mean squared errors (MSE) of the estimator of β_1 under all three approaches to understand their overall performances. These MSE's are reported in Table A.5. It is clear that as the sample size increases the MSE gets smaller. For example, for $K = 75$, $\tau = 4.0$, and $\rho = 0.7$, the $WGQL_1$ produces MSE as 0.880, whereas for $K = 100$ and 200, this approach gives the MSE as 0.746 and 0.432 respectively. It is clear from Table A.5 that the $WGQL_1$ approach, in general, produces treatment estimates with smaller MSE followed by the $WGQL_3$ approach. Recall that the $WGQL_1$ and $WGQL_3$ approaches use adaptive weights as functions of β, whereas the weights in the $WGQL_2$ approach are free from β as these are constructed based on past data. Consequently, although the $WGQL_2$ approach trails the other two approaches with regard to the MSE for the estimation of β_1, the MSEs are however close to the MSEs of the other two approaches. Thus, from the viewpoint of practitioners, we recommend this data dependent weights based $WGQL_2$ approach for the estimation of the regression parameters including the treatment effect. In the next subsection, we examine the misspecification effect of the adaptive design under the recommended $WGQL_2$ approach.

3.6.3 Design Misspecification Effect

To examine the effect of using the non-adaptive design in estimating the regression parameters of the model, we obtain the regression estimates from the iterative equation (3.27) by using $w_i = 0.5$ for all $i = 1, \ldots, K$, even though the longitudinal binary data were generated as before based on the adaptive longitudinal design based variable weights w_i. The mean squared errors (MSEs) of the regression estimates were computed for the two cases : **case 1)** estimation was carried out based on weights following the adaptive longitudinal design; **case 2)** estimation was carried out using

$w_i = 0.5$ following the non-adaptive design with equal weights. The MSEs are reported in Table A.6. It is clear from the table that the MSEs for the estimates of β_2, β_3 and β_4 are not so different under the two choices of the design weights. This is however not true for the estimation of the treatment effect β_1. This is because the MSEs of the β_1 estimates are always larger when the estimation is carried out based on $w_i = 0.5$ as compared to the estimation based on true weights following the adaptive longitudinal design. For example, for $K = 100$, $\tau = 4.0$, and $\rho = 0.9$, although the adaptive longitudinal design based MSE of β_1 estimate is 0.884, it is 1.045 for the working design ($w_i = 0.5$) based estimate. Thus, it is clear that if the adaptive design based weights are ignored during the estimation, one would obtain an unreliable estimate for the treatment effect.

3.6.4   Confidence Interval for Treatment Effect

Recall from § 3.6.2 that the $WGQL_2$ approach was found to be practically suitable for the point estimation of the regression parameters including the treatment effect. In this subsection, we take a further look on the interval estimation of the main parameter of interest, namely, the treatment effect. For this purpose, we recall from § 3.4 that $\hat{\beta} = (\hat{\beta}_{1,WGQL_2}, \hat{\beta}_{2,WGQL_2}, \hat{\beta}_{3,WGQL_2}, \hat{\beta}_{4,WGQL_2})'$ has asymptotically a 4-dimensional normal distribution with mean β and a suitable covariance matrix that can be consistently estimated by (3.35), namely,

$$\hat{cov}(\hat{\beta}_{WGQL2}) = \left[\sum_{i=1}^{K} C_i \Sigma_i^{-1}(w_i, \hat{\rho}) C_i' \right]^{-1}. \tag{3.37}$$

Consequently, one may construct $(1 - \alpha)100\%$ confidence interval for β_1 given by

$$\hat{\beta}_{1,WGQL_2} - z_{\frac{\alpha}{2}} s.e.(\hat{\beta}_{1,WGQL_2}) \leq \beta_1 \leq \hat{\beta}_{1,WGQL_2} + z_{\frac{\alpha}{2}} s.e.(\hat{\beta}_{1,WGQL_2}), \tag{3.38}$$

where $s.e.(\hat{\beta}_{1,WGQL_2})$ is computed from the first diagonal element of the 4×4 covariance matrix in (3.37).

Next to examine the performance of the interval estimation of β_1 by (3.38), we conduct a limited Monte Carlo study with 1000 simulations. By using $|z_{\frac{\alpha}{2}}| = 1.96$,

in each simulation, we computed the lower and upper limits for β_1 given by (3.38) for true $\beta_1 = 1.5$, and calculated the proportion of simulations with true value of $\beta_1 = 1.5$ bounded by these limits. These proportions, which are known as coverage probabilities, are reported in Table A.7. For the selected values of K, τ, and ρ, it is clear from the table that the coverage probabilities lie in the range between 93% and 96%. More specifically, for $K = 100$, $\tau = 4.0$, and $\rho = 0.5$, the coverage probability was found to be 96% and for $K = 200$, $\tau = 2.0$, and $\rho = 0.9$, the coverage probability was 93%. For other cases the coverage probabilities were equal or close to 95%. This shows that $WGQL_2$ estimation approach is quite satisfactory for the estimation of the treatment effect β_1.

In this chapter, we have introduced a longitudinal adaptive design in order to assign more study subjects to a better treatment in the longitudinal clinical trial set up. The proposed design is constructed by generalizing the adaptive designs used so far in the literature in the non-longitudinal set up. We have studied various limiting properties of the proposed design weights as well as examined the performance of the proposed design in allocating incoming individuals to a better treatment through a simulation study. The design was found to work quite well in allocating more study subjects to a better treatment.

This chapter also introduced a weighted generalized quasi-likelihood (WGQL) approach for the estimation of the effects of the prognostic factors including the treatment effect. The WGQL approach exploits both longitudinal design weights and the longitudinal correlations of the binary responses yielding consistent and efficient estimates for the parameters involved. The performance of the WGQL approach was studied through a simulation study and it was found that this approach works quite well in estimating the parameters, including the treatment effect. The coverage probabilities for the treatment effects were also found to be highly satisfactory.

Chapter 4

Longitudinal Mixed Model For Binary Data in Adaptive Clinical Trials

In the previous chapter, we introduced longitudinal fixed models in adaptive clinical trials where repeated responses from an individual patient are not affected by any unknown individual effect. But, it may happen in practice that the variability and the correlations of the repeated data may not be completely explained by the model parameters considered in Chapter 3. As a remedy, in this chapter we introduce a binary longitudinal mixed model for adaptive clinical trials under the assumption that conditional on a particular individual random effect, the repeated responses of an individual follow a correlation structure as in Chapter 3. Thus, unconditionally, the mean, variance, and covariances of repeated responses will also be affected by the random effects of the individuals under the study.

Note that in the traditional simple random sampling set up (non-adaptive set up), some authors have analyzed repeated binary and count data by assuming that conditional on the random effects the repeated responses are independent. For example, we refer to Zeger (1988), Sutradhar and Das (1995), Heagerty (1999), and Davis, Dunsmuir, and Wang (2000). But as discussed by Jowaheer and Sutradhar (2002),

these random effects based models are not suitable to represent the auto-correlation structure (such as AR(1)) of the repeated data. There exist some studies (Sashegyi, Brown, and Farrell (2000), Sutradhar and Sinha (2002), Sutradhar and Farrell (2003)) where the longitudinal correlations of the repeated responses of an individual are modeled conditional on the individual random effect. Note that these studies are however confined to non-clinical studies where adaptive designs do not play any role.

Remark that as opposed to the analysis of longitudinal fixed model data collected from individuals, in the longitudinal mixed model set up, there are individual random effects on the longitudinal responses of individuals under study. Consequently, in the longitudinal mixed model set up, the correlations among the unconditional responses arise due to the variation in the random effects as well as the longitudinal correlations considered in the conditional set up. Note that the consistent and efficient estimation of the parameters of this type of longitudinal mixed model is however much more involved as compared to the estimation of the parameters under the longitudinal fixed model. Further, additional problems arise in the longitudinal adaptive clinical trial set up, where individuals enter the trials in sequence and individuals are assigned to the treatment based on the outcomes of the past individuals. The unconditional correlations (involving the variance of the random effects as well as the longitudinal conditional correlations) along with the design weights must be taken into account to estimate the parameters involved in the model.

In this chapter, we consider the longitudinal mixed model for binary responses in adaptive clinical trials where incoming patients are assigned (i.e. adaptive design is constructed) to a better treatment on the basis of certain data-dependent rules. In § 4.1, we introduce this longitudinal mixed model for binary data collected based on adaptive designs. The longitudinal mixed model, unlike the longitudinal fixed model discussed in Chapter 3, contains regression as well as the variance component of the individual random effects as main parameters, and longitudinal correlations parameters as nuisance parameters. In order to estimate the parameters, in § 4.2, we provide a weighted generalized quasi-likelihood (WGQL) approach for the estimation of the

regression vector (β) after taking the design weights and the longitudinal correlations into account. The WGQL estimator of β is consistent and highly efficient. In § 4.3, we exploit this WGQL approach and develop a unified estimating equation for the consistent and efficient estimation of the variance component of individual random effects. We also provide a consistent estimator for longitudinal correlations by the method of moments in § 4.4. We examine the performances of the WGQL estimation approach and the the method of moments approach through a simulation study in § 4.5.

4.1 Adaptive Design Based Binary Longitudinal Mixed Model

As in Chapter 3, we consider two treatments A and B so that the probability that the ith $(i = 1, \ldots, K)$ individual is assigned to the treatment A is w_i given by (3.6) for $2 \leq i \leq T$ and by (3.9) for $i > T$ where

$$w_i = \Pr[\delta_i | y_H],$$

with

$$\delta_i = \begin{cases} 1, & \text{if the } i\text{th patient is assigned to treatment } A \\ 0, & \text{if the } i\text{th patient is assigned to treatment } B, \end{cases}$$

and where y_H indicates the history for the past $i - 1$ patients. Here w_i is referred to as the design weight for the selection of a treatment for the individual. Recall from Chapter 3 that under the fixed longitudinal binary adaptive design, y_{it}, the response of the ith individual was assumed to follow the binary logistic model with

$$\Pr(y_{it} = 1 | \delta_i, \ldots, \delta_1) = p_{it}^* = \frac{exp(x_{it}^{*'}\beta)}{1 + exp(x_{it}^{*'}\beta)},$$

where the ith individual was selected for treatment A with probability w_i. In the present mixed model set up, we now assume that the ith individual has an unobservable random effect which naturally will affect the responses $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$

collected longitudinally. To incorporate this individual random effect we now consider that

$$\Pr(y_{it} = 1 | \delta_i, \ldots, \delta_1; \gamma_i^*) = \tilde{p}_{it}^* = \frac{exp(x_{it}^{*\prime}\beta + \gamma_i^*)}{1 + exp(x_{it}^{*\prime}\beta + \gamma_i^*)}, \quad (4.1)$$

where γ_i^* is the random effect of the *ith* individual. It will be assumed that $\gamma_i^* \overset{i.i.d.}{\sim} N(0, \sigma^2)$. Note that under the present set up, to initiate the construction of the estimating equations we will require the computations of the unconditional mean vector and the covariance matrix for the repeated responses of the individuals having a general autocorrelation structure conditional on the individual random effect and treatment effects. This we do in the following two sub-sections.

4.1.1 Construction of the Unconditional Mean Vector

Let $y_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT})'$ be a T-dimensional vector of repeated responses to a treatment for the *ith* $(i = 1, \ldots, K)$ individual. It then follows from (4.1) that the unconditional expectation of Y_{it} may be computed as

$$E(Y_{it}) = E_{\gamma_i} E_{\delta_1} E_{\delta_2 | \delta_1} \ldots E_{\delta_i | \delta_1, \ldots, \delta_{i-1}} E(Y_{it} | \delta_i, \ldots, \delta_1; \gamma_i), \quad (4.2)$$

where $\gamma_i = \frac{\gamma_i^*}{\sigma} \overset{i.i.d.}{\sim} N(0, 1)$. Now, by using (3.10) and (4.1), we obtain

$$
\begin{aligned}
E(Y_{it}) &= E_{\gamma_i} E_{\delta_1} E_{\delta_2 | \delta_1} \ldots E_{\delta_i | \delta_1, \ldots, \delta_{i-1}} (\tilde{p}_{it}^*) \\
\\
&= E_{\gamma_i} [w_{i0} \tilde{p}_{it1} + (1 - w_{i0}) \tilde{p}_{it2}], \quad (4.3)
\end{aligned}
$$

where

$$\tilde{p}_{it1} = \frac{exp(z_{it}'\beta + \sigma\gamma_i)}{1 + exp(z_{it}'\beta + \sigma\gamma_i)} \quad \text{and} \quad \tilde{p}_{it2} = \frac{exp(z_{it}^{*\prime}\beta + \sigma\gamma_i)}{1 + exp(z_{it}^{*\prime}\beta + \sigma\gamma_i)} \quad (4.4)$$

respectively, with $z_{it}' = (\delta_i, x_{it2}, \ldots, x_{itp})|_{\delta_i=1}$ and $z_{it}^{*\prime} = (\delta_i, x_{it2}, \ldots, x_{itp})|_{\delta_i=0}$ and w_{i0} as given in (3.14) for $2 \leq i \leq T$ and in (3.15) for $i > T$. The expectation over γ_i in (4.3) may be computed based on a simulation approach (Fahrmeir and Tutz (1994)) as

$$E(Y_{it}) = w_{i0} \frac{1}{M} \sum_{\xi=1}^{M} \frac{exp(z_{it}'\beta + \sigma\gamma_{i,\xi})}{1 + exp(z_{it}'\beta + \sigma\gamma_{i,\xi})} + (1 - w_{i0}) \frac{1}{M} \sum_{\xi=1}^{M} \frac{exp(z_{it}^{*\prime}\beta + \sigma\gamma_{i,\xi})}{1 + exp(z_{it}^{*\prime}\beta + \sigma\gamma_{i,\xi})}$$

$$= w_{i0} \frac{1}{M} \sum_{\xi=1}^{M} g_{i1}^{(t)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_{i0}) \frac{1}{M} \sum_{\xi=1}^{M} g_{i2}^{(t)}(\beta, \sigma, \gamma_{i,\xi})$$

$$= w_{i0} \frac{1}{M} \sum_{\xi=1}^{M} \tilde{p}_{it1,\xi} + (1 - w_{i0}) \frac{1}{M} \sum_{\xi=1}^{M} \tilde{p}_{it2,\xi}$$

$$= \tilde{p}_{it}, say, \tag{4.5}$$

where for $\xi = 1, \ldots, M$, $\gamma_{i,\xi}$ are M simulated values of γ_i from a standard normal distribution. Here M is usually large, say $M = 5,000$.

Now, by using the $\tilde{p}_{it} = E(Y_{it})$ from (4.5), one may construct the unconditional mean vector of Y_i as

$$\tilde{p}_i = (\tilde{p}_{i1}, \ldots, \tilde{p}_{it}, \ldots, \tilde{p}_{iT})'. \tag{4.6}$$

4.1.2 Construction of the Unconditional Variance and Co-variance

In this subsection, we develop the formulas for the variance of Y_{it} $(t = 1, \ldots, T)$ and the covariance between Y_{it} and Y_{iv} for $t \neq v$. For this purpose, we first compute the covariances as follows. For given $\delta_i, \ldots, \delta_1$ and individual random effect γ_i, the two responses Y_{it} and Y_{iv} are assumed to have correlation given by

$$corr(Y_{it}, Y_{iv} | \delta_i, \ldots, \delta_1; \gamma_i) = \rho_{|t-v|}. \tag{4.7}$$

The correlation structure defined in (4.7) is general as it accommodates the Gaussian type AR(1), MA(1), and exchangeable auto-correlation structures as special cases.

To compute the covariance between Y_{it} and Y_{iv}, that is, to compute

$$cov(Y_{it}, Y_{iv}) = E(Y_{it} Y_{iv}) - E(Y_{it}) E(Y_{iv}), \tag{4.8}$$

we simplify $E(Y_{it} Y_{iv})$ as follows:

$$E(Y_{it} Y_{iv}) = E_{\gamma_i} E_{\delta_1} E_{\delta_2 | \delta_1} \ldots E_{\delta_i | \delta_1, \ldots, \delta_{i-1}} E(Y_{it} Y_{iv} | \delta_i, \ldots, \delta_1; \gamma_i)$$

$$= E_{\gamma_i} E_{\delta_1} E_{\delta_2 | \delta_1} \ldots E_{\delta_i | \delta_1, \ldots, \delta_{i-1}} \Big[cov(Y_{it}, Y_{iv} | \delta_i, \ldots, \delta_1; \gamma_i)$$

$$+E(Y_{it}|\delta_i,\ldots,\delta_1;\gamma_i)\ E(Y_{iv}|\delta_i,\ldots,\delta_1;\gamma_i)\Big]$$

$$= E_{\gamma_i}E_{\delta_1}E_{\delta_2|\delta_1}\ldots E_{\delta_i|\delta_1,\ldots,\delta_{i-1}}\left[\rho_{|t-v|}\left(\tilde{p}_{it}^*\tilde{q}_{it}^*\tilde{p}_{iv}^*\tilde{q}_{iv}^*\right)^{1/2}+\tilde{p}_{it}^*\tilde{p}_{iv}^*\right]$$

$$(4.9)$$

Now by (4.3), we compute the expectation over $\delta_i,\ldots,\delta_1$ and obtain

$$E(Y_{it}Y_{iv}) = E_{\gamma_i}\Big[w_{i0}\rho_{|t-v|}\left(\tilde{p}_{it1}\tilde{q}_{it1}\tilde{p}_{iv1}\tilde{q}_{iv1}\right)^{1/2}+(1-w_{i0})\rho_{|t-v|}\left(\tilde{p}_{it2}\tilde{q}_{it2}\tilde{p}_{iv2}\tilde{q}_{iv2}\right)^{1/2}$$

$$+w_{i0}\tilde{p}_{it1}\tilde{p}_{iv1}+(1-w_{i0})\tilde{p}_{it2}\tilde{p}_{iv2}\Big], \qquad (4.10)$$

where $\tilde{q}_{it1}=1-\tilde{p}_{it1}$, $\tilde{q}_{it2}=1-\tilde{p}_{it2}$. Now based on the simulation approach as shown in (4.5), we can simplify the above expectation as

$$E(Y_{it}Y_{iv}) = w_{i0}\rho_{|t-v|}\frac{1}{M}\sum_{\xi=1}^{M}g_{i3}^{(tv)}(\beta,\sigma,\gamma_{i,\xi})+(1-w_{i0})\rho_{|t-v|}\frac{1}{M}\sum_{\xi=1}^{M}g_{i4}^{(tv)}(\beta,\sigma,\gamma_{i,\xi})$$

$$+w_{i0}\frac{1}{M}\sum_{\xi=1}^{M}g_{i5}^{(tv)}(\beta,\sigma,\gamma_{i,\xi})+(1-w_{i0})\frac{1}{M}\sum_{\xi=1}^{M}g_{i6}^{(tv)}(\beta,\sigma,\gamma_{i,\xi}), \qquad (4.11)$$

where

$$g_{i3}^{(tv)}(\beta,\sigma,\gamma_{i,\xi})=\left(\tilde{p}_{it1,\xi}\ \tilde{q}_{it1,\xi}\ \tilde{p}_{iv1,\xi}\ \tilde{q}_{iv1,\xi}\right)^{1/2},$$

$$g_{i4}^{(tv)}(\beta,\sigma,\gamma_{i,\xi})=\left(\tilde{p}_{it2,\xi}\ \tilde{q}_{it2,\xi}\ \tilde{p}_{iv2,\xi}\ \tilde{q}_{iv2,\xi}\right)^{1/2},$$

$$g_{i5}^{(tv)}(\beta,\sigma,\gamma_{i,\xi})=\tilde{p}_{it1,\xi}\ \tilde{p}_{iv1,\xi},$$

and

$$g_{i6}^{(tv)}(\beta,\sigma,\gamma_{i,\xi})=\tilde{p}_{it2,\xi}\ \tilde{p}_{iv2,\xi}.$$

Now by applying (4.5) and (4.11) into (4.8), we obtain the covariance as

$$cov(Y_{it},Y_{iv}) = w_{i0}\rho_{|t-v|}\frac{1}{M}\sum_{\xi=1}^{M}g_{i3}^{(tv)}(\beta,\sigma,\gamma_{i,\xi})+(1-w_{i0})\rho_{|t-v|}\frac{1}{M}\sum_{\xi=1}^{M}g_{i4}^{(tv)}(\beta,\sigma,\gamma_{i,\xi})$$

$$+w_{i0}\frac{1}{M}\sum_{\xi=1}^{M}g_{i5}^{(tv)}(\beta,\sigma,\gamma_{i,\xi})+(1-w_{i0})\frac{1}{M}\sum_{\xi=1}^{M}g_{i6}^{(tv)}(\beta,\sigma,\gamma_{i,\xi})-\tilde{p}_{it}\tilde{p}_{iv}$$

$$= \tilde{\sigma}_{itv},\ say. \qquad (4.12)$$

When $t = v$, (4.8) yields the variance of Y_{it} given by

$$
\begin{aligned}
var(Y_{it}) &= E(Y_{it}^2) - [E(Y_{it})]^2 \\[2mm]
&= E(Y_{it}) - [E(Y_{it})]^2 \\[2mm]
&= \tilde{p}_{it}\tilde{q}_{it},
\end{aligned}
\tag{4.13}
$$

since $E(Y_{it})$, by (4.5), is given as $E(Y_{it}) = \tilde{p}_{it}$, and in (4.13), $\tilde{q}_{it} = 1 - \tilde{p}_{it}$. Suppose that $\tilde{\Sigma}_i$ is the covariance matrix of Y_i such that

$$
\tilde{\Sigma}_i = cov(Y_i) = (\tilde{\sigma}_{itv}),
\tag{4.14}
$$

for $t, v = 1, \ldots, T$, where $\tilde{\sigma}_{itv}$ and $\tilde{\sigma}_{itt}$ are defined as in (4.12) and (4.13), respectively.

Note that all the elements of the $\tilde{\Sigma}_i$ matrix are computed by simulating normal random effects implying that $\tilde{\Sigma}_i$ may be referred to as the simulated unconditional covariance matrix. This $\tilde{\Sigma}_i$ matrix along with the unconditional mean vector $\tilde{p}_i$ will be exploited in the next two sections to construct the weighted generalized estimating equations for the regression parameters β and the variance component of the random effects σ^2. The longitudinal correlation parameters ρ_ℓ ($\ell = |t - v| = 1, \ldots, T - 1$) will be estimated by using the well-known method of moments. More specifically, a moment estimate of ρ_ℓ will be obtained by equating the sample autocorrelation functions to their population counterparts (cf. Jowaheer and Sutradhar (2002) and Sutradhar and Kovacevic (2000)).

4.2   WGQL Approach for Regression Effects

In practice, $\tilde{\Sigma}_i$ is unknown and it is a function of w_{i0}, β, σ, and ρ where w_{i0} is again a function of β and σ. The mean vector $\tilde{p}_i$ is also a function of β, σ, and w_{i0}. Then for known ρ and σ, the weighted generalized quasi-likelihood (WGQL) estimating equation for β may be written as

$$
\sum_{i=1}^{K} \tilde{B}_i \tilde{\Sigma}_i^{-1}(w_{i0}, \hat{\rho}, \hat{\sigma})(y_i - \tilde{p}_i(w_{i0}, \hat{\sigma})) = 0,
\tag{4.15}
$$

where $\tilde{p}_i$ and $\tilde{\Sigma}_i$ are defined as in (4.6) and (4.14) respectively, and $\tilde{B}_i = \frac{\partial \tilde{p}'_i(w_{i0},\hat{\sigma})}{\partial \beta}$ is the $p \times T$ first order derivative matrix. Remark that in Chapter 3, among three WGQL approaches, we recommended the data dependent weights based WGQL approach for practitioners for estimating the parameters, including the treatment effect. In this chapter, we follow this recommendation and use the data dependent weights based WGQL approach for estimating parameters involved in the model. In data dependent weights based WGQL approach, it is assumed that w_{i0} is unknown, but it can be replaced by w_i as $E(w_i) = w_{i0}$, where w_i is given by (3.6) for $2 \le i \le T$ and by (3.9) for $i > T$. Then estimating equation (4.15) can be re-expressed as

$$\sum_{i=1}^{K} \tilde{C}_i \tilde{\Sigma}_i^{-1}(w_i, \hat{\rho}, \hat{\sigma})(y_i - \tilde{p}_i(w_i, \hat{\sigma})) = 0, \qquad (4.16)$$

where $\tilde{C}_i = \frac{\partial \tilde{p}'_i(w_i,\hat{\sigma})}{\partial \beta}$. Note that since $\tilde{p}_i = (\tilde{p}_{i1}, \dots, \tilde{p}_{it}, \dots, \tilde{p}_{iT})'$ the derivative of $\tilde{p}'_i$ with respect to β requires the derivative of $\tilde{p}_{it}$ as defined in (4.5) with respect to β. To be specific, $\frac{\partial \tilde{p}_{it}}{\partial \beta}$ is the $p \times 1$ vector given by

$$\frac{\partial \tilde{p}_{it}}{\partial \beta} = w_i \frac{1}{M} \sum_{\xi=1}^{M} \frac{\partial}{\partial \beta} g_{i1}^{(t)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i) \frac{1}{M} \sum_{\xi=1}^{M} \frac{\partial}{\partial \beta} g_{i2}^{(t)}(\beta, \sigma, \gamma_{i,\xi}).$$

As , following (4.5),

$$g_{i1}^{(t)}(\beta, \sigma, \gamma_{i,\xi}) = \tilde{p}_{it1,\xi} = \frac{exp(z'_{it}\beta + \sigma\gamma_{i,\xi})}{1 + exp(z'_{it}\beta + \sigma\gamma_{i,\xi})},$$

and

$$g_{i2}^{(t)}(\beta, \sigma, \gamma_{i,\xi}) = \tilde{p}_{it2,\xi} = \frac{exp(z^{*\prime}_{it}\beta + \sigma\gamma_{i,\xi})}{1 + exp(z^{*\prime}_{it}\beta + \sigma\gamma_{i,\xi})},$$

one obtains

$$\frac{\partial \tilde{p}_{it}}{\partial \beta} = \frac{1}{M} \sum_{\xi=1}^{M} \left[w_i \tilde{p}_{it1,\xi} \tilde{q}_{it1,\xi} z_{it} + (1 - w_i) \tilde{p}_{it2,\xi} \tilde{q}_{it2,\xi} z^*_{it} \right], \qquad (4.17)$$

where $\tilde{q}_{it1,\xi} = 1 - \tilde{p}_{it1,\xi}$ and $\tilde{q}_{it2,\xi} = 1 - \tilde{p}_{it2,\xi}$. Also in (4.17), $z_{it} = x^*_{it}|_{\delta_i=1}$ and $z^*_{it} = x^*_{it}|_{\delta_i=0}$ are $p \times 1$ vectors of all covariates for the ith individual at time t. Consequently,

$$\frac{\partial \tilde{p}'_i}{\partial \beta} = \frac{1}{M} \sum_{\xi=1}^{M} \left[w_i Z'_i \tilde{A}_{i1,\xi} + (1 - w_i) Z^{*\prime}_i \tilde{A}_{i2,\xi} \right]$$

$$= \tilde{C}_i, \qquad (4.18)$$

where $Z_i' = (z_{i1}, \ldots, z_{it}, \ldots, z_{iT})$ and $Z_i^{*'} = (z_{i1}^*, \ldots, z_{it}^*, \ldots, z_{iT}^*)$ are $p \times T$ matrices and $\tilde{A}_{i1,\xi} = diag(\tilde{p}_{i11,\xi}\tilde{q}_{i11,\xi}, \ldots, \tilde{p}_{iT1,\xi}\tilde{q}_{iT1,\xi})$ and $\tilde{A}_{i2,\xi} = diag(\tilde{p}_{i12,\xi}\tilde{q}_{i12,\xi}, \ldots, \tilde{p}_{iT2,\xi}\tilde{q}_{iT2,\xi})$ are $T \times T$ matrices.

Next, let $\hat{\beta}_{WGQL}$ denote the weighted generalized quasi-likelihood estimator of β which is the solution of the estimating equation (4.16). The solution of (4.16), that is, $\hat{\beta}_{WGQL}$ may be obtained by using the customary Newton-Raphson iterative method. Given the value $\hat{\beta}_{WGQL}(r)$ at the rth iteration, $\hat{\beta}_{WGQL}(r+1)$ may be obtained at the $(r+1)st$ iteration as

$$\hat{\beta}_{WGQL}(r+1) = \hat{\beta}_{WGQL}(r) + \Big[\sum_{i=1}^{K} \tilde{C}_i \tilde{\Sigma}_i^{-1}(w_i, \hat{\rho}, \hat{\sigma}) \tilde{C}_i'\Big]_r^{-1}$$

$$\times \Big[\sum_{i=1}^{K} \tilde{C}_i \tilde{\Sigma}_i^{-1}(w_i, \hat{\rho}, \hat{\sigma})(y_i - \tilde{p}_i(w_i, \hat{\sigma}))\Big]_r, \qquad (4.19)$$

where $[\cdot]_r$ denotes the fact that the expression in the brackets is evaluated at $\beta = \hat{\beta}_{WGQL}(r)$.

The estimator $\hat{\beta}_{WGQL}$ is consistent and it is highly efficient for known σ^2 and ρ_ℓ ($\ell = 1, \ldots, T-1$). This is because $\hat{\beta}_{WGQL}$ is obtained by solving the estimating equation (4.16), where the weight matrix $\tilde{\Sigma}_i$ is the correct covariance matrix of the responses. If σ^2 and ρ_ℓ are unknown, which is usually the case in practice, the use of their consistent estimates in the weighted matrix in (4.16) still provides a highly efficient estimator of β. As we discuss in the next section, the efficiency of the estimator of σ^2 will however depend on the correct specification of a weight matrix (a fourth order moment matrix for longitudinal mixed model data) to be used for the construction of the estimating equation for σ^2. Note that if one could obtain the maximum likelihood estimator of β, it would have been fully efficient. It is however extremely difficult to compute the maximum likelihood estimator, as the joint density of the longitudinal responses is unknown.

Further, by arguments analogous to those given in Liang and Zeger (1986), it may be shown that $K^{\frac{1}{2}}(\hat{\beta}_{WGQL} - \beta)$ has an asymptotic normal distribution with mean

vector 0 and covariance matrix given by

$$K\Big(\sum_{i=1}^{K}\tilde{C}_i\tilde{\Sigma}_i^{-1}(w_i,\hat{\rho},\hat{\sigma})\tilde{C}_i'\Big)^{-1}. \tag{4.20}$$

Remark that the computation of the estimate of β by (4.16) requires σ^2 and ρ_ℓ to be known, where for $\ell = 1,\ldots,T-1$, the ρ_ℓ's are longitudinal correlation parameters, which are treated as nuisance. In a manner similar to that of (4.16), in the next section, we develop the weighted generalized quasi-likelihood estimating equation for σ^2. A consistent estimator for ρ_ℓ obtained by the method of moments is provided in § 4.4.

4.3   WGQL Approach for Variance Component

To develop the estimating equation for σ^2, similar to Jowaheer and Sutradhar (2002), we use the squared and distinct cross-product responses of the ith individual ($i = 1,\ldots,K$) as data. Let $f_{is} = (y_{i1}^2,\ldots,y_{it}^2,\ldots,y_{iT}^2)'$ be the T-dimensional vector of squares of the repeated observations and $f_{ip} = (y_{i1}y_{i2},\ldots,y_{it}y_{iv},\ldots,y_{iT-1}y_{iT})'$ be the $\frac{T(T-1)}{2}$-dimensional vector of distinct pair-wise-products of repeated observations for $t,v = 1,\ldots,T; t < v$. Since observations here are binary, f_{is} can be re-written as $f_{is} = (y_{i1},\ldots,y_{it},\ldots,y_{iT})'$. Further let $f_i = (f_{is}',f_{ip}')'$ be the $\frac{T(T+1)}{2}$-dimensional combined vector of squares and pair-wise products of repeated observations for the ith individual. Let $m_{is}^* = (\tilde{m}_{i1},\ldots,\tilde{m}_{it},\ldots,\tilde{m}_{iT})'$ be the mean vector of f_{is} with $\tilde{m}_{it} = E(Y_{it})$ and $m_{ip}^* = (\tilde{m}_{i12},\ldots,\tilde{m}_{itv},\ldots,\tilde{m}_{i(T-1)T})'$ be the mean vector of f_{ip} with $\tilde{m}_{itv} = E(Y_{it}Y_{iv})$. Also suppose that mean vector of combined vector f_i is m_i^* such that $m_i^* = (m_{is}^{*\,\prime}, m_{ip}^{*\,\prime})'$. Let the covariance matrix of f_i be Ω_i.

As explained in § 4.2, m_i^* is a function of β, σ, ρ, and longitudinal design weights w_{i0} defined as in (3.14) for $2 \leq i \leq T$ and in (3.15) for $i > T$. Here, like the previous section, we use the data dependent design weights because it is recommended in Chapter 3 for practical application. In other words, we use w_i instead of w_{i0}. Note that in the present longitudinal mixed model set up, it is however impossible to

compute the fourth order moments matrix Ω_i. This is because the joint distribution of repeated responses $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$ is unknown. To overcome this problem, in the longitudinal set up, Jowaheer and Sutradhar (2002), following Prentice and Zhao (1991), approximated Ω_i matrix by a normality based 'working' matrix $\Omega_i^{(N)}$, say. To compute $\Omega_i^{(N)}$ matrix in the present set up, one may follow Jowaheer and Sutradhar (2002) and compute the elements of $\Omega_i^{(N)}$ by pretending that $\tilde{\Sigma}_i$ in (4.16) is the covariance matrix of the normal vector $y_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT})'$, whereas y_i is truly the vector of binary responses. The computation of $\Omega_i^{(N)}$ matrix is however complicated. As opposed to the 'working normality' based fourth order moments matrix, we here propose a 'longitudinal independence' based working fourth order moments matrix $\Omega_i^{(I)}$, say. Note that unlike the 'working normality' based approach, one retains the true nature of the responses in tact in computing the 'working independence' $\Omega_i^{(I)}$ matrix. Moreover, the computation of the $\Omega_i^{(I)}$ matrix is much more simpler than that of the $\Omega_i^{(N)}$ matrix. Since, in the present set up, we use the 'working independence' based approach, $\Omega_i^{(I)}$ will be a function of β, σ, and the longitudinal design weights w_{i0}. Like § 4.2, we also use the data dependent design weight based WGQL approach to estimate the variance component σ^2 of individual random effects. Note that although, in the present approach, $\Omega_i^{(I)}$ will be free from ρ, the mean function m_i^* and its derivative vector $\tilde{D}_i$ (with respect to σ^2) are both functions of β, σ, and ρ. Then for known β and ρ, the weighted generalized quasi-likelihood estimating equation for σ^2 is

$$\sum_{i=1}^{K} \tilde{D}_i \Omega_i^{(I)^{-1}}(w_i, \hat{\beta})(f_i - m_i^*(w_i, \hat{\beta}, \hat{\rho})) = 0, \qquad (4.21)$$

where w_i is given by (3.6) for $2 \leq i \leq T$ and by (3.9) for $i > T$ and $\tilde{D}_i$ is $1 \times \frac{T(T+1)}{2}$ first order derivative vector $\frac{\partial m_i^{*\prime}(w_i, \hat{\beta}, \hat{\rho})}{\partial \sigma^2}$. In the following sub-sections we show how we can compute the elements of m_i^*, $\tilde{D}_i$, and $\Omega_i^{(I)}$, for $i = 1, \ldots, K$.

4.3.1 Construction of Mean Vector m_i^*

Recall that $m_i^* = (m_{is}^{*}{}', m_{ip}^{*}{}')'$ where $m_{is}^* = (\tilde{m}_{i1}, \ldots, \tilde{m}_{it}, \ldots, \tilde{m}_{iT})'$ with $\tilde{m}_{it} = E(Y_{it})$ and $m_{ip}^* = (\tilde{m}_{i12}, \ldots, \tilde{m}_{itv}, \ldots, \tilde{m}_{i(T-1)T})'$ with $\tilde{m}_{itv} = E(Y_{it}Y_{iv})$ for $i = 1, \ldots, K$; $t, v = 1, \ldots, T$; $t < v$. It then follows from (4.5) and (4.11) that

$$\tilde{m}_{it} = w_i \frac{1}{M} \sum_{\xi=1}^{M} g_{i1}^{(t)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i)\frac{1}{M} \sum_{\xi=1}^{M} g_{i2}^{(t)}(\beta, \sigma, \gamma_{i,\xi}), \qquad (4.22)$$

and

$$
\tilde{m}_{itv} = w_i \rho_{|t-v|} \frac{1}{M} \sum_{\xi=1}^{M} g_{i3}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i)\rho_{|t-v|}\frac{1}{M} \sum_{\xi=1}^{M} g_{i4}^{(tv)}(\beta, \sigma, \gamma_{i,\xi})
$$

$$
+ w_i \frac{1}{M} \sum_{\xi=1}^{M} g_{i5}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i)\frac{1}{M} \sum_{\xi=1}^{M} g_{i6}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}), \qquad (4.23)
$$

respectively, where $g_{i1}^{(t)}(\beta, \sigma, \gamma_{i,\xi})$, $g_{i2}^{(t)}(\beta, \sigma, \gamma_{i,\xi})$, $g_{i3}^{(tv)}(\beta, \sigma, \gamma_{i,\xi})$, $g_{i4}^{(tv)}(\beta, \sigma, \gamma_{i,\xi})$, $g_{i5}^{(tv)}(\beta, \sigma, \gamma_{i,\xi})$, and $g_{i6}^{(tv)}(\beta, \sigma, \gamma_{i,\xi})$ are defined in § 4.1.1 and § 4.1.2.

4.3.2 Construction of First Order Derivative Vector $\frac{\partial m_i^{*}{}'}{\partial \sigma^2}$

The first order derivative of the mean vector m_i^* ($i = 1, \ldots, K$) with respect to the variance component σ^2 can be defined as

$$
\frac{\partial m_i^{*}{}'}{\partial \sigma^2} = \left[\frac{\partial}{\partial \sigma^2} m_{is}^{*}{}', \frac{\partial}{\partial \sigma^2} m_{ip}^{*}{}' \right]
$$

$$
= \tilde{D}_i, \qquad (4.24)
$$

where $\tilde{D}_i$ is a $1 \times \frac{T(T+1)}{2}$ vector. Since $m_{is}^* = (\tilde{m}_{i1}, \ldots, \tilde{m}_{it}, \ldots, \tilde{m}_{iT})'$ and $m_{ip}^* = (\tilde{m}_{i12}, \ldots, \tilde{m}_{itv}, \ldots, \tilde{m}_{i(T-1)T})'$, to compute $\tilde{D}_i$ it is sufficient to compute the derivatives of $\tilde{m}_{it}$ and $\tilde{m}_{itv}$ with respect to σ^2 where $\tilde{m}_{it}$ and $\tilde{m}_{itv}$ are defined as in (4.22) and (4.23) respectively. By direct calculations, these derivatives are given by

$$
\frac{\partial \tilde{m}_{it}}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2}\left[w_i \frac{1}{M} \sum_{\xi=1}^{M} g_{i1}^{(t)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i)\frac{1}{M} \sum_{\xi=1}^{M} g_{i2}^{(t)}(\beta, \sigma, \gamma_{i,\xi}) \right]
$$

$$
= \frac{1}{2M\sigma}\left[w_i \sum_{\xi=1}^{M} g_{i7}^{(t)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i)\sum_{\xi=1}^{M} g_{i8}^{(t)}(\beta, \sigma, \gamma_{i,\xi}) \right], \qquad (4.25)
$$

where

$$g_{i7}^{(t)}(\beta, \sigma, \gamma_{i,\xi}) = \gamma_{i,\xi} \tilde{p}_{it1,\xi} \tilde{q}_{it1,\xi},$$

$$g_{i8}^{(t)}(\beta, \sigma, \gamma_{i,\xi}) = \gamma_{i,\xi} \tilde{p}_{it2,\xi} \tilde{q}_{it2,\xi},$$

and

$$
\begin{aligned}
\frac{\partial \tilde{m}_{itv}}{\partial \sigma^2} &= \frac{\partial}{\partial \sigma^2} \Big[w_i \rho_{|t-v|} \frac{1}{M} \sum_{\xi=1}^{M} g_{i3}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) + (1-w_i)\rho_{|t-v|} \frac{1}{M} \sum_{\xi=1}^{M} g_{i4}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) \\
&\quad + w_i \frac{1}{M} \sum_{\xi=1}^{M} g_{i5}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) + (1-w_i)\frac{1}{M} \sum_{\xi=1}^{M} g_{i6}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) \Big] \\
&= w_i \rho_{|t-v|} \frac{1}{2M\sigma} \sum_{\xi=1}^{M} g_{i9}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) + (1-w_i)\rho_{|t-v|} \frac{1}{2M\sigma} \sum_{\xi=1}^{M} g_{i10}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) \\
&\quad + w_i \frac{1}{2M\sigma} \sum_{\xi=1}^{M} g_{i11}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) + (1-w_i)\frac{1}{2M\sigma} \sum_{\xi=1}^{M} g_{i12}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}), \quad (4.26)
\end{aligned}
$$

where

$$g_{i9}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) = \gamma_{i,\xi} \Big(\tilde{p}_{it1,\xi} \tilde{q}_{it1,\xi} \tilde{p}_{iv1,\xi} \tilde{q}_{iv1,\xi} \Big)^{1/2} (1 - \tilde{p}_{it1,\xi} - \tilde{p}_{iv1,\xi}),$$

$$g_{i10}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) = \gamma_{i,\xi} \Big(\tilde{p}_{it2,\xi} \tilde{q}_{it2,\xi} \tilde{p}_{iv2,\xi} \tilde{q}_{iv2,\xi} \Big)^{1/2} (1 - \tilde{p}_{it2,\xi} - \tilde{p}_{iv2,\xi}),$$

$$g_{i11}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) = \gamma_{i,\xi} \tilde{p}_{it1,\xi} \tilde{p}_{iv1,\xi} (2 - \tilde{p}_{it1,\xi} - \tilde{p}_{iv1,\xi}),$$

and

$$g_{i12}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) = \gamma_{i,\xi} \tilde{p}_{it2,\xi} \tilde{p}_{iv2,\xi} (2 - \tilde{p}_{it2,\xi} - \tilde{p}_{iv2,\xi}),$$

respectively.

4.3.3 Construction of the 'Working' Covariance Matrix $\Omega_i^{(I)}$

Recall that Ω_i is the covariance matrix of $f_i = (f'_{is}, f'_{ip})'$ where f_{is} is the vector of repeated observations and f_{ip} is the vector of distinct pair-wise-products of the repeated observations of the *ith* individual ($i = 1, \ldots, K$). Thus, to compute Ω_i, it

is sufficient to compute three matrices, namely, $\Omega_{iss} = var(F_{is})$, $\Omega_{isp} = cov(F_{is}, F_{ip})$, and $\Omega_{ipp} = var(F_{ip})$. In other words, Ω_i can be expressed as

$$\Omega_i = \begin{pmatrix} \Omega_{iss} & \Omega_{isp} \\ & \Omega_{ipp} \end{pmatrix}. \tag{4.27}$$

It then follows that for the construction of the $\Omega_i^{(I)}$ matrix, one needs to compute $\Omega_{iss}^{(I)}$, $\Omega_{isp}^{(I)}$, and $\Omega_{ipp}^{(I)}$ matrices. This may be achieved by pretending that for given $\delta_i, \ldots, \delta_1$ and γ_i, Y_{it} and Y_{iv} are independent. This implies that

$$corr\Big\{(Y_{it}, Y_{iv})|\delta_i, \ldots, \delta_1; \gamma_i\Big\} = \rho_{|t-v|} = 0, \tag{4.28}$$

for $t, v = 1, \ldots, T$, $t \neq v$, although, in practice, $\rho_{|t-v|}$ may not be zero.

Construction of $\Omega_{iss}^{(I)}$ Matrix

Recall that $f_{is} = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT})'$ for $(i = 1, \ldots, K)$. Then it follows that $\Omega_{iss}^{(I)}$ is a $T \times T$ matrix defined as

$$\Omega_{iss}^{(I)} = \begin{pmatrix} var(Y_{i1}) & \cdots & \cdots & cov(Y_{i1}, Y_{iT}) \\ \vdots & & & \vdots \\ cov(Y_{iT}, Y_{i1}) & \cdots & \cdots & var(Y_{iT}) \end{pmatrix}.$$

For the computation of $\Omega_{iss}^{(I)}$, it is sufficient to compute $cov(Y_{it}, Y_{iv})$, $(t, v = 1, \ldots, T; t \neq v)$ and $var(Y_{it})$. More specifically, the covariance may be computed from (4.12) by evaluating it at $\rho_{|t-v|} = 0$ and $w_{i0} = w_i$. That is,

$$
\begin{aligned}
cov(Y_{it}, Y_{iv}) &= \tilde{\sigma}_{itv}|_{\rho_{|t-v|}=0, w_{i0}=w_i} \\
&= w_i \frac{1}{M} \sum_{\xi=1}^{M} g_{i5}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i)\frac{1}{M} \sum_{\xi=1}^{M} g_{i6}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) \\
&\quad -\tilde{\tilde{p}}_{it}\tilde{\tilde{p}}_{iv},
\end{aligned}
\tag{4.29}
$$

where $g_{i5}^{(tv)}(\beta, \sigma, \gamma_{i,\xi})$ and $g_{i6}^{(tv)}(\beta, \sigma, \gamma_{i,\xi})$ are defined in § 4.1.2 and $\tilde{p}_{it} = E(Y_{it}) = \tilde{p}_{it}|_{w_{i0}=w_i}$ where $\tilde{p}_{it}$ is defined as in (4.5). Again, following (4.13), $var(Y_{it})$ can be

computed as

$$var(Y_{it}) = \tilde{p}_{it}\tilde{\tilde{q}}_{it}, \tag{4.30}$$

where $\tilde{\tilde{q}}_{it} = 1 - \tilde{p}_{it}$.

Construction of $\Omega_{isp}^{(I)}$ Matrix

Note that $\Omega_{isp}^{(I)}$ is the $T \times \frac{T(T-1)}{2}$ matrix of covariances between F_{is} and F_{ip}, f_{is} being the $T \times 1$ vector of repeated observations and f_{ip} being the $\frac{T(T-1)}{2} \times 1$ vector of distinct pair-wise-product of elements of f_{is}. Then one can express $\Omega_{isp}^{(I)}$ as

$$\Omega_{isp}^{(I)} = \begin{pmatrix} cov(Y_{i1}, Y_{i1}Y_{i2}) & \cdots & \cdots & cov(Y_{i1}, Y_{iT-1}Y_{iT}) \\ \vdots & & & \vdots \\ cov(Y_{iT}, Y_{i1}Y_{i2}) & \cdots & \cdots & cov(Y_{iT}, Y_{iT-1}Y_{iT}) \end{pmatrix}.$$

To construct the elements of $\Omega_{isp}^{(I)}$, it is necessary to compute $E(Y_{it})$, $E(Y_{it}Y_{iv})$, and $E(Y_{it}Y_{iv}Y_{ir})$ for $i = 1, \ldots, K$; $t, v, r = 1, \ldots, T$. These expectations can be computed as follows. To be specific, by (4.5),

$$
\begin{aligned}
E(Y_{it}) &= \tilde{p}_{it}|_{w_{i0}=w_i} \\
&= w_i \frac{1}{M} \sum_{\xi=1}^{M} g_{i1}^{(t)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i)\frac{1}{M} \sum_{\xi=1}^{M} g_{i2}^{(t)}(\beta, \sigma, \gamma_{i,\xi}) \\
&= \tilde{p}_{it}. \tag{4.31}
\end{aligned}
$$

Next, under the independence assumption, one may compute $E(Y_{it}Y_{iv})$, for $t \neq v$, as

$$
\begin{aligned}
E(Y_{it}Y_{iv}) &= E_{\gamma_i}E_{\delta_1}E_{\delta_2|\delta_1}\ldots E_{\delta_i|\delta_1,\ldots,\delta_{i-1}}E(Y_{it}Y_{iv}|\delta_i,\ldots,\delta_1;\gamma_i) \\
&= E_{\gamma_i}E_{\delta_1}E_{\delta_2|\delta_1}\ldots E_{\delta_i|\delta_1,\ldots,\delta_{i-1}}\Big\{E(Y_{it}|\delta_i,\ldots,\delta_1;\gamma_i)E(Y_{iv}|\delta_i,\ldots,\delta_1;\gamma_i)\Big\} \\
&= E_{\gamma_i}E_{\delta_1}E_{\delta_2|\delta_1}\ldots E_{\delta_i|\delta_1,\ldots,\delta_{i-1}}\Big(\tilde{p}_{it}^*\tilde{p}_{iv}^*\Big) \\
&= E_{\gamma_i}\Big[w_i\tilde{p}_{it1}\tilde{p}_{iv1} + (1 - w_i)\tilde{p}_{it2}\tilde{p}_{iv2}\Big]
\end{aligned}
$$

$$= \quad w_i \frac{1}{M} \sum_{\xi=1}^{M} g_{i5}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) + (1-w_i)\frac{1}{M} \sum_{\xi=1}^{M} g_{i6}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}). \qquad (4.32)$$

Similarly, $E(Y_{it}Y_{iv}Y_{ir})$ for $t \neq v \neq r$ can be expressed as

$$E(Y_{it}Y_{iv}Y_{ir}) = w_i \frac{1}{M} \sum_{\xi=1}^{M} g_{i13}^{(tvr)}(\beta, \sigma, \gamma_{i,\xi}) + (1-w_i)\frac{1}{M} \sum_{\xi=1}^{M} g_{i14}^{(tvr)}(\beta, \sigma, \gamma_{i,\xi}), \qquad (4.33)$$

where

$$g_{i13}^{(tvr)}(\beta, \sigma, \gamma_{i,\xi}) = \tilde{p}_{it1,\xi}\tilde{p}_{iv1,\xi}\tilde{p}_{ir1,\xi}$$

and

$$g_{i14}^{(tvr)}(\beta, \sigma, \gamma_{i,\xi}) = \tilde{p}_{it2,\xi}\tilde{p}_{iv2,\xi}\tilde{p}_{ir2,\xi}.$$

We now turn back to write the formulas for the elements of the $\Omega_{isp}^{(I)}$ matrix by using the above expectations. One must consider the following cases for this purpose.

Case I: $t \neq v, \ v < r; \ t,v,r = 1,\ldots,T$

$$cov(Y_{it}, Y_{iv}Y_{ir}) \quad = \quad E(Y_{it}Y_{iv}Y_{ir}) - E(Y_{it})E(Y_{iv}Y_{ir})$$

$$= \quad w_i \frac{1}{M} \sum_{\xi=1}^{M} g_{i13}^{(tvr)}(\beta, \sigma, \gamma_{i,\xi}) + (1-w_i)\frac{1}{M} \sum_{\xi=1}^{M} g_{i14}^{(tvr)}(\beta, \sigma, \gamma_{i,\xi})$$

$$-\tilde{\tilde{p}}_{it}\Big[w_i \frac{1}{M} \sum_{\xi=1}^{M} g_{i5}^{(vr)}(\beta, \sigma, \gamma_{i,\xi}) + (1-w_i)\frac{1}{M} \sum_{\xi=1}^{M} g_{i6}^{(vr)}(\beta, \sigma, \gamma_{i,\xi})\Big].$$

$$(4.34)$$

Case II: $t = v, \ v < r; \ t,v,r = 1,\ldots,T$

$$cov(Y_{it}, Y_{iv}Y_{ir}) \quad = \quad E(Y_{it}Y_{iv}Y_{ir}) - E(Y_{it})E(Y_{iv}Y_{ir})$$

$$= \quad [1 - E(Y_{it})]E(Y_{it}Y_{ir})$$

$$= \quad (1 - \tilde{p}_{it})\Big[w_i \frac{1}{M} \sum_{\xi=1}^{M} g_{i5}^{(tr)}(\beta, \sigma, \gamma_{i,\xi}) + (1-w_i)\frac{1}{M} \sum_{\xi=1}^{M} g_{i6}^{(tr)}(\beta, \sigma, \gamma_{i,\xi})\Big].$$

$$(4.35)$$

Case III: $t = r$, $v < r$; $t, v, r = 1, \ldots, T$

$$
\begin{aligned}
cov(Y_{it}, Y_{iv}Y_{ir}) &= E(Y_{it}Y_{iv}Y_{ir}) - E(Y_{it})E(Y_{iv}Y_{ir}) \\[2mm]
&= [1 - E(Y_{it})]E(Y_{it}Y_{iv}) \\[2mm]
&= (1 - \tilde{p}_{it})\Big[w_i \frac{1}{M} \sum_{\xi=1}^{M} g_{i5}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i) \frac{1}{M} \sum_{\xi=1}^{M} g_{i6}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) \Big].
\end{aligned}
$$

$$(4.36)$$

Construction of $\Omega_{ipp}^{(I)}$ Matrix

Recall that $\Omega_{ipp}^{(I)}$ is the $\frac{T(T-1)}{2} \times \frac{T(T-1)}{2}$ covariance matrix of vector $F_{ip} = (Y_{i1}Y_{i2}, \ldots, Y_{it}Y_{iv},$ $\ldots, Y_{iT-1}Y_{iT})'$ which can be expressed as

$$
\Omega_{ipp}^{(I)} = \begin{pmatrix}
var(Y_{i1}Y_{i2}) & \cdots & \cdots & cov(Y_{i1}Y_{i2}, Y_{iT-1}Y_{iT}) \\
\vdots & & & \vdots \\
cov(Y_{iT-1}Y_{iT}, Y_{i1}Y_{i2}) & \cdots & \cdots & var(Y_{iT-1}Y_{iT})
\end{pmatrix}.
$$

Note that for the computation of elements of $\Omega_{ipp}^{(I)}$, we require to compute $E(Y_{it})$, $E(Y_{it}Y_{iv})$, $E(Y_{it}Y_{iv}Y_{ir})$, and $E(Y_{it}Y_{iv}Y_{ir}Y_{is})$ for $i = 1, \ldots, K$; $t, v, r, s = 1, \ldots, T$. The expectations, $E(Y_{it})$, $E(Y_{it}Y_{iv})$, and $E(Y_{it}Y_{iv}Y_{ir})$, are already shown in (4.31), (4.32), and (4.33) respectively, whereas $E(Y_{it}Y_{iv}Y_{ir}Y_{is})$ can be computed under the working independence assumption for $t \neq v \neq r \neq s$ as follows.

$$
\begin{aligned}
E(Y_{it}Y_{iv}Y_{ir}Y_{is}) &= E_{\gamma_i} E_{\delta_1} E_{\delta_2|\delta_1} \ldots E_{\delta_i|\delta_1,\ldots,\delta_{i-1}} E(Y_{it}Y_{iv}Y_{ir}Y_{is}|\delta_i,\ldots,\delta_1;\gamma_i) \\[2mm]
&= E_{\gamma_i} E_{\delta_1} E_{\delta_2|\delta_1} \ldots E_{\delta_i|\delta_1,\ldots,\delta_{i-1}} \Big(\tilde{p}_{it}^* \tilde{p}_{iv}^* \tilde{p}_{ir}^* \tilde{p}_{is}^* \Big) \\[2mm]
&= E_{\gamma_i} \Big[w_i \tilde{p}_{it1} \tilde{p}_{iv1} \tilde{p}_{ir1} \tilde{p}_{is1} + (1 - w_i) \tilde{p}_{it2} \tilde{p}_{iv2} \tilde{p}_{ir2} \tilde{p}_{is2} \Big] \\[2mm]
&= w_i \frac{1}{M} \sum_{\xi=1}^{M} g_{i15}^{(tvrs)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i) \frac{1}{M} \sum_{\xi=1}^{M} g_{i16}^{(tvrs)}(\beta, \sigma, \gamma_{i,\xi}),
\end{aligned}
$$

$$(4.37)$$

where

$$g_{i15}^{(tvrs)}(\beta, \sigma, \gamma_{i,\xi}) = \tilde{p}_{it1,\xi}\tilde{p}_{iv1,\xi}\tilde{p}_{ir1,\xi}\tilde{p}_{is1,\xi}$$

and

$$g_{i16}^{(tvrs)}(\beta, \sigma, \gamma_{i,\xi}) = \tilde{p}_{it2,\xi}\tilde{p}_{iv2,\xi}\tilde{p}_{ir2,\xi}\tilde{p}_{is2,\xi}.$$

One must consider the following cases to compute the elements of the $\Omega_{ipp}^{(I)}$ matrix.

Case I: $t < v, r < s,\ t = r, v = s;\ \ t, v, r, s = 1, \ldots, T$

$$
\begin{aligned}
var(Y_{it}Y_{iv}) &= E[Y_{it}Y_{iv}] - [E(Y_{it}Y_{iv})]^2 \\[2mm]
&= E(Y_{it}Y_{iv})[1 - E(Y_{it}Y_{iv})] \\[2mm]
&= \Big[w_i \frac{1}{M}\sum_{\xi=1}^{M} g_{i5}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i)\frac{1}{M}\sum_{\xi=1}^{M} g_{i6}^{(tv)}(\beta, \sigma, \gamma_{i,\xi})\Big] \\[2mm]
&\quad \times \Big[1 - \Big\{w_i \frac{1}{M}\sum_{\xi=1}^{M} g_{i5}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i)\frac{1}{M}\sum_{\xi=1}^{M} g_{i6}^{(tv)}(\beta, \sigma, \gamma_{i,\xi})\Big\}\Big].
\end{aligned}
$$

$$(4.38)$$

Case II: $t < v, r < s,\ t \neq r, v \neq s,\ t \neq s,\ v \neq r;\ \ t, v, r, s = 1, \ldots, T$

$$
\begin{aligned}
cov(Y_{it}Y_{iv}, Y_{ir}Y_{is}) &= E(Y_{it}Y_{iv}Y_{ir}Y_{is}) - E(Y_{it}Y_{iv})E(Y_{ir}Y_{is}) \\[2mm]
&= w_i \frac{1}{M}\sum_{\xi=1}^{M} g_{i15}^{(tvrs)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i)\frac{1}{M}\sum_{\xi=1}^{M} g_{i16}^{(tvrs)}(\beta, \sigma, \gamma_{i,\xi}) \\[2mm]
&\quad - \Big[\Big\{w_i \frac{1}{M}\sum_{\xi=1}^{M} g_{i5}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i)\frac{1}{M}\sum_{\xi=1}^{M} g_{i6}^{(tv)}(\beta, \sigma, \gamma_{i,\xi})\Big\} \\[2mm]
&\quad \times \Big\{w_i \frac{1}{M}\sum_{\xi=1}^{M} g_{i5}^{(rs)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i)\frac{1}{M}\sum_{\xi=1}^{M} g_{i6}^{(rs)}(\beta, \sigma, \gamma_{i,\xi})\Big\}\Big].
\end{aligned}
$$

$$(4.39)$$

Case III: $t < v, r < s,\ t = r, v \neq s;\quad t, v, r, s = 1, \dots, T$

$$
\begin{aligned}
cov(Y_{it}Y_{iv}, Y_{ir}Y_{is}) &= E(Y_{it}Y_{iv}Y_{is}) - E(Y_{it}Y_{iv})E(Y_{it}Y_{is}) \\[2mm]
&= w_i \frac{1}{M} \sum_{\xi=1}^{M} g_{i13}^{(tvs)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i)\frac{1}{M}\sum_{\xi=1}^{M} g_{i14}^{(tvs)}(\beta, \sigma, \gamma_{i,\xi}) \\[2mm]
&\quad - \Big[\Big\{ w_i \frac{1}{M}\sum_{\xi=1}^{M} g_{i5}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i)\frac{1}{M}\sum_{\xi=1}^{M} g_{i6}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) \Big\} \\[2mm]
&\quad \times \Big\{ w_i \frac{1}{M}\sum_{\xi=1}^{M} g_{i5}^{(ts)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i)\frac{1}{M}\sum_{\xi=1}^{M} g_{i6}^{(ts)}(\beta, \sigma, \gamma_{i,\xi}) \Big\} \Big].
\end{aligned}
$$

$$(4.40)$$

Case IV: $t < v, r < s,\ t = s, v > r;\quad t, v, r, s = 1, \dots, T$

$$
\begin{aligned}
cov(Y_{it}Y_{iv}, Y_{ir}Y_{is}) &= E(Y_{it}Y_{iv}Y_{ir}) - E(Y_{it}Y_{iv})E(Y_{it}Y_{ir}) \\[2mm]
&= w_i \frac{1}{M} \sum_{\xi=1}^{M} g_{i13}^{(tvr)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i)\frac{1}{M}\sum_{\xi=1}^{M} g_{i14}^{(tvr)}(\beta, \sigma, \gamma_{i,\xi}) \\[2mm]
&\quad - \Big[\Big\{ w_i \frac{1}{M}\sum_{\xi=1}^{M} g_{i5}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i)\frac{1}{M}\sum_{\xi=1}^{M} g_{i6}^{(tv)}(\beta, \sigma, \gamma_{i,\xi}) \Big\} \\[2mm]
&\quad \times \Big\{ w_i \frac{1}{M}\sum_{\xi=1}^{M} g_{i5}^{(tr)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i)\frac{1}{M}\sum_{\xi=1}^{M} g_{i6}^{(tr)}(\beta, \sigma, \gamma_{i,\xi}) \Big\} \Big].
\end{aligned}
$$

$$(4.41)$$

Case V: $t < v, r < s,\ v = r, t < s;\quad t, v, r, s = 1, \dots, T$

$$
\begin{aligned}
cov(Y_{it}Y_{iv}, Y_{ir}Y_{is}) &= E(Y_{it}Y_{iv}Y_{is}) - E(Y_{it}Y_{iv})E(Y_{iv}Y_{is}) \\[2mm]
&= w_i \frac{1}{M} \sum_{\xi=1}^{M} g_{i13}^{(tvs)}(\beta, \sigma, \gamma_{i,\xi}) + (1 - w_i)\frac{1}{M}\sum_{\xi=1}^{M} g_{i14}^{(tvs)}(\beta, \sigma, \gamma_{i,\xi})
\end{aligned}
$$

$$-\Big[\Big\{w_i\frac{1}{M}\sum_{\xi=1}^{M}g_{i5}^{(tv)}(\beta,\sigma,\gamma_{i,\xi})+(1-w_i)\frac{1}{M}\sum_{\xi=1}^{M}g_{i6}^{(tv)}(\beta,\sigma,\gamma_{i,\xi})\Big\}$$

$$\times\Big\{w_i\frac{1}{M}\sum_{\xi=1}^{M}g_{i5}^{(vs)}(\beta,\sigma,\gamma_{i,\xi})+(1-w_i)\frac{1}{M}\sum_{\xi=1}^{M}g_{i6}^{(vs)}(\beta,\sigma,\gamma_{i,\xi})\Big\}\Big].$$

$$(4.42)$$

Case VI: $t<v, r<s,\ v=s, t\neq r;\ \ t,v,r,s=1,\ldots,T$

$$
\begin{aligned}
cov(Y_{it}Y_{iv},Y_{ir}Y_{is}) &= E(Y_{it}Y_{iv}Y_{ir})-E(Y_{it}Y_{iv})E(Y_{iv}Y_{ir})\\[2mm]
&= w_i\frac{1}{M}\sum_{\xi=1}^{M}g_{i13}^{(tvr)}(\beta,\sigma,\gamma_{i,\xi})+(1-w_i)\frac{1}{M}\sum_{\xi=1}^{M}g_{i14}^{(tvr)}(\beta,\sigma,\gamma_{i,\xi})\\[2mm]
&\quad -\Big[\Big\{w_i\frac{1}{M}\sum_{\xi=1}^{M}g_{i5}^{(tv)}(\beta,\sigma,\gamma_{i,\xi})+(1-w_i)\frac{1}{M}\sum_{\xi=1}^{M}g_{i6}^{(tv)}(\beta,\sigma,\gamma_{i,\xi})\Big\}\\[2mm]
&\quad \times\Big\{w_i\frac{1}{M}\sum_{\xi=1}^{M}g_{i5}^{(vr)}(\beta,\sigma,\gamma_{i,\xi})+(1-w_i)\frac{1}{M}\sum_{\xi=1}^{M}g_{i6}^{(vr)}(\beta,\sigma,\gamma_{i,\xi})\Big\}\Big].
\end{aligned}
$$

$$(4.43)$$

This completes the construction of the 'working independence' based covariance matrix $\Omega_i^{(I)}$.

Let $\hat{\sigma}_{WGQL}^2$ denote the weighted generalized quasi-likelihood estimator of σ^2, which is the solution of the estimating equation (4.21). This estimator is consistent, but loses its efficiency slightly because of the use of a 'working' covariance matrix as the weight matrix in the estimating equation. Note that the degree of efficiency loss will depend on the extent of misspecification of the 'working' fourth order moments matrix to be used in the place of the true fourth order moments matrix. Remark that this problem of efficiency loss may arise in some situations only under the longitudinal mixed model (Sutradhar and Kumar (2001)), but not at all for the generalized linear mixed model. This is because when $T=1$, one can compute the exact fourth order moments matrix.

Turning back to the asymptotic distribution of the variance component estimator, one may show, by similar arguments as in Liang and Zeger (1986), that $K^{\frac{1}{2}}(\hat{\sigma}^2_{WGQL} - \sigma^2)$ has a univariate normal distribution, as $K \to \infty$, with mean zero and the variance which may be consistently estimated by

$$\Big[\sum_{i=1}^{K} \tilde{D}_i \Omega_i^{(I)^{-1}}(w_i, \hat{\beta}) \tilde{D}_i'\Big]^{-2} \Big[\sum_{i=1}^{K} \tilde{D}_i \Omega_i^{(I)^{-1}}(w_i, \hat{\beta})(f_i - \hat{m}_i^*(w_i, \hat{\beta}, \hat{\rho}))$$

$$\times (f_i - \hat{m}_i^*(w_i, \hat{\beta}, \hat{\rho}))' \Omega_i^{(I)^{-1}}(w_i, \hat{\beta}) \tilde{D}_i'\Big], \tag{4.44}$$

where $\hat{m}_i^*$ is computed by using $\hat{\beta}_{WGQL}$ and $\hat{\sigma}^2_{WGQL}$ in the formula for m_i^* given in § 4.3.1.

4.4 Moment Approach for Longitudinal Correlations

Similar to Jowaheer and Sutradhar (2002) (see also Sutradhar and Kovacevic (2000)), the longitudinal correlations ρ_ℓ ($\ell = 1, \ldots, T-1$) in the longitudinal mixed model are treated to be nuisance parameters. Nevertheless, it is evident from sections § 4.2 and § 4.3 that the iterative solutions of the estimating equations (4.16) and (4.21) for β and σ^2, respectively, require a consistent estimator for the longitudinal correlation parameter ρ_ℓ. This may be obtained using the method of moments and by solving the moment estimating equation derived by equating the unconditional sample covariance with its population counterpart given by (4.12). One may then compute $\hat{\rho}_\ell$ ($\ell = |t-v| = 1, \ldots, T-1$), a consistent estimator of ρ_ℓ, by using the formula

$$\hat{\rho}_\ell = \frac{\dfrac{\sum_{i=1}^{K}\sum_{|t-v|=\ell}[(y_{it}-\tilde{p}_{it})(y_{iv}-\tilde{p}_{iv})]/K(T-\ell)}{\sum_{i=1}^{K}\sum_{t=1}^{T}(y_{it}-\tilde{p}_{it})^2/KT} - \dfrac{\bar{S}}{\sum_{i=1}^{K}\sum_{t=1}^{T}(\tilde{p}_{it}\tilde{q}_{it})/KT}}{\dfrac{\sum_{i=1}^{K}\sum_{|t-v|=\ell}\Big[w_i\frac{1}{M}\sum_{\xi=1}^{M}g_{i3}^{(tv)}(\beta,\sigma,\gamma_{i,\xi})+(1-w_i)\frac{1}{M}\sum_{\xi=1}^{M}g_{i4}^{(tv)}(\beta,\sigma,\gamma_{i,\xi})\Big]/K(T-\ell)}{\sum_{i=1}^{K}\sum_{t=1}^{T}(\tilde{p}_{it}\tilde{q}_{it})/KT}}, \tag{4.45}$$

where

$$\bar{S} = \frac{\sum_{i=1}^{K}\sum_{|t-v|=\ell}w_i\frac{1}{M}\sum_{\xi=1}^{M}g_{i5}^{(tv)}(\beta,\sigma,\gamma_{i,\xi}) + (1-w_i)\frac{1}{M}\sum_{\xi=1}^{M}g_{i6}^{(tv)}(\beta,\sigma,\gamma_{i,\xi}) - \tilde{p}_{it}\tilde{p}_{iv}}{K(T-\ell)}.$$

This correlation estimate is used in (4.16) and (4.21) to obtain further improved estimates of β and σ^2 respectively, which are in turn used here in (4.45) to obtain a further improved estimate of ρ_ℓ. This cycle of iteration continues until convergence.

4.5 Performance of the WGQL Estimation Approach For Mixed Model: A Simulation Study

In this section, we examine the performance of the weighted generalized quasi-likelihood (WGQL) approach for the estimation of the effects of covariates, including the treatment effect and the variance component of individual random effects of the binary longitudinal mixed model under an adaptive clinical set up by conducting a simulation study. As in Chapter 3, we consider two treatments A and B assuming that treatment A is the better treatment. In § 3.2.1, we have shown how to construct the longitudinal adaptive design weights w_i $(i = 1, \ldots, K)$ so that more study subjects are allocated to the better treatment. Let δ_i be the treatment covariate. For $i = 1, \ldots, K$, the treatment covariate δ_i is chosen to be 1 if the ith patient is allocated to the better treatment A and to be 0 otherwise. As in Chapter 3, the values of δ_i are determined based on the adaptive longitudinal design such that

$$\Pr(\delta_i = 1 | y_H) = w_i,$$

where w_i values are computed from (3.6) for $2 \le i \le T$ and from (3.9) for $i > T$ respectively and y_H is the history of the response from past $i - 1$ patients. It is also noted that the treatment covariate δ_i does not depend on the time t. This is because once the patient is assigned to a treatment , the patient remains under the selected treatment for the complete duration of T periods.

For the construction of w_i for $i = 2, \ldots, K$ by using (3.6) and (3.9), we require the knowledge of past responses y_{rt} $(r = 1, \ldots, i - 1,\ 1 \le t \le min(T, i - r))$ and $x_{rt}^* = (\delta_r, x_{rt2}, \ldots, x_{rtp})'$ where δ_r is known and $x_{rt2}, \ldots, x_{rtp}$ are known prognostic factors. For our simulation study we consider $K = 100$, K being the number of individuals

in the longitudinal adaptive clinical study. Further we consider $T = 4$, the number of repeated data from each individual. Usually, under the longitudinal set up, the number of individuals K is large and the number of time periods T is small. For the sake of simplicity, we consider $p = 1$ covariate: the treatment covariate only. To ensure that treatment A is better, we choose $\beta = 1.5$, for example. As far as the variance component of the random effects is concerned, in this simulation study, we consider small as well as large values of σ^2, namely, $\sigma^2 = 0.1, 0.2, 0.5, 0.75,$ and 1.0. With regard to the correlations of the repeated data, for the present set up, we consider a binary AR(1) model with different values for the autocorrelation parameter ρ, namely, $\rho = 0.3, 0.5, 0.7, 0.8, 0.9,$ and 0.95 to represent small as well as large correlation.

Note that the computation of w_i further requires the values of α, G, τ, and the non-stochastic function of prognostic factors other than treatment covariate u_{rt} $(r = 1, \ldots, i-1, \ 1 \leq t \leq min(T, i-r))$, where G is the upper bound for the domain of u_{rt}. Since, in this simulation, we only consider treatment as a prognostic factor, here, $u_{rt} = 0$ and hence $G = 0$. For simplicity we choose $\alpha = 1.0$ and $\tau = 2.0$.

To generate the individual random effects, we generate independent $\gamma_1, \ldots, \gamma_K$ from N(0,1) for $K = 100$. Note that all y_{i1} $(i = 1, \ldots, K)$ [the first observation in the cluster under the *ith* individual] are generated by using a logistic binary model given by

$$\Pr(y_{i1} = 1 | \delta_i, \ldots, \delta_1; \gamma_i) = \frac{exp(x_{i\cdot}^{*\prime}\beta + \sigma\gamma_i)}{1 + exp(x_{i\cdot}^{*\prime}\beta + \sigma\gamma_i)} = \tilde{p}_i^*,$$

since $x_{it}^* = x_{i\cdot}^*$ for all $t = 1, \ldots, T$. In other words, $x_{it}^* = \delta_i$. To generate y_{it} for $2 \leq t \leq T$, one must ensure that $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$ satisfy the underlying longitudinal correlation structure appropriate for the binary repeated data. If the repeated binary responses follow AR(1), the correlated binary responses $(y_{i2}, \ldots, y_{it}, \ldots, y_{iT})$ for each patient i can be generated as follows. If $y_{i1} = 0$, y_{i2} will be generated with probability $\tilde{p}_i^*(1-\rho)$; if $y_{i1} = 1$, then y_{i2} will be generated with probability $\tilde{p}_i^* + \rho(1 - \tilde{p}_i^*)$. One may continue this to generate y_{i3} depending on y_{i2} and so on. This assures that the lag $\ell = 1, \ldots, T-1$ correlation between y_{it} and $y_{i(t+\ell)}$ is ρ^ℓ. In the following sub-section, we report the performance of estimators of β, σ^2, and $\rho = (\rho_1, \ldots, \rho_\ell)'$ based on 500

simulations.

4.5.1 Estimation Performance

With starting values zero for longitudinal correlations and small positive values for regression and variance component parameters, we obtained the estimates of β and σ^2 by using (4.16) and (4.21) respectively and the estimate of lag-ℓ ($\ell = 1, \ldots, T-1$) autocorrelation i.e. of ρ_ℓ, by using (4.45). The estimation procedure was repeated for 500 simulated runs as mentioned above. We report the simulated means and simulated standard errors of β, σ^2, and $\rho = (\rho_1, \rho_2, \rho_3)'$ parameters in Table B.1. We have also computed the estimated standard errors for $\hat{\beta}$ and $\hat{\sigma}^2$ by using the asymptotic variance formulas for $\hat{\beta}$ and $\hat{\sigma}^2$ given by (4.20) and (4.44) respectively. The means of these estimated standard errors are also reported in Table B.1.

It is clear from Table B.1 that the robust estimating formula (4.45) for the autocorrelations performs extremely well in estimating the autocorrelation parameters ρ_ℓ ($\ell = 1, \ldots, T-1$). As expected, these estimates appear to approximately satisfy the AR(1) relationship $\rho_\ell = \rho^\ell$. For example, for $\sigma^2 = 0.50$ and $\rho = 0.8$, the moment estimating formula (4.45) yields the lag 1, 2, and 3 correlations as 0.797, 0.636, and 0.507 respectively, whereas the true lag correlations are 0.8, 0.64, and 0.512 respectively. The simulated standard errors of these correlations are 0.044, 0.084, and 0.115 respectively, which are reasonably small.

With regard to the estimation of the regression parameter (treatment effect), the weighted generalized quasi-likelihood estimating equation (4.16) appears to perform well. This method however appears to overestimate β slightly. The simulated standard errors of this regression estimator are smaller for small values of ρ and larger for large values of ρ, irrespective of the values of σ^2 (small or large). The estimates of β are less biased for small values of σ^2 than for large values of σ^2.

To estimate the variance component σ^2, the weighted generalized quasi-likelihood estimation equation (4.21) appears to underestimate the large σ^2 and to overestimate the small σ^2. The performance of this method in estimating the parameter σ^2 is

relatively better for small values of σ^2, namely, $\sigma^2 = 0.1, 0.2$ as compared to the cases with large values of σ^2, namely, $\sigma^2 = 0.5, 0.75, 1.0$. For large values of σ^2, the amount of bias appears to decrease as the value of ρ increases. On the other hand, the amount of bias in the estimate of σ^2 is insignificant for small values of σ^2. The simulated standard errors of the estimate of the variance component σ^2 are small as expected. The simulated standard errors appear to be stable irrespective of the value of ρ (small or large).

Further, to examine the performance of the estimated standard errors of the regression estimator (treatment effect) $\hat{\beta}$ and variance estimator $\hat{\sigma}^2$, we computed the averages of the estimated standard errors calculated from (4.20) and (4.44) respectively. These values agreed closely with the simulated standard errors of the regression estimator and the variance component estimator, as shown in Table B.1. For example, for $\sigma^2 = 0.5$ and $\rho = 0.8$, the estimated standard error of the treatment effect and variance component estimators are 0.407 and 0.123 respectively. These are close to the corresponding simulated standard errors 0.366 and 0.149, respectively.

Note that the simulation results reported in Table B.1 clearly show that the WGQL approach performs well in estimating the treatment effect of the adaptive longitudinal binary mixed model. The performance of this approach in estimating the variance component is also satisfactory, even though the variance estimates are slightly biased in some cases, especially for large values of the true variance component, along with small values of the longitudinal correlations. Further note that as the data dependent design weights w_i $(i = 1, \ldots, K)$ were already shown to be important in the adaptive set up, we included these weights under the present mixed model in order to obtain consistent and efficient estimates for the treatment effect as well as the variance component of the individual random effects.

Chapter 5

Concluding Remarks

5.1 General Remarks

In clinical trial studies, the individuals are included in the sample in sequence, and the available information about the treatment and other covariates are used to assign a new individual to a better treatment. Thus, the allocation of the treatment to an individual depends on an adaptive design as opposed to the simple random sampling design. It is of interest to examine the effects of the treatment and other covariates at the end of the clinical trial study. These statistical inferences are made by exploiting the adaptive design weights properly in the estimation and testing processes. Note that there are however situations where individuals are kept on the system for a small period of time after their enrollment under a particular treatment. This makes the repeated responses of an individual longitudinally correlated. Recently, Sutradhar and Biswas (2001) have introduced a longitudinal binary adaptive design and dealt with the inferences about the treatment effects. The estimation of the treatment and other covariate effects, in this set up, requires adaptive design weights and longitudinal correlations of the responses to be known. With regard to the longitudinal correlations, they have incorporated them following Sutradhar and Das (1999). But, these authors have exploited the expected design weights (instead of the data dependent design weights) for the estimation of parameters by using the weighted

generalized quasi-likelihood (WGQL) approach. In Chapter 3, we have re-examined the WGQL estimation procedure of Sutradhar and Biswas (2001) by using the data dependent design weights in the estimation process. More specifically, in Chapter 3, we have compared the performance of the data dependent design weights based WGQL estimation process with that of Sutradhar and Biswas (2001). It was found based on a simulation study that this data dependent design weights based WGQL technique performs better than other techniques based on the limiting design weights.

Note that it may further happen that individuals under study may have unobservable random effects which along with covariates may affect their repeated responses. This random effect issue under an adaptive set up is however not addressed in the literature so far. With this in view, in Chapter 4, we have developed a longitudinal binary mixed model under an adaptive clinical trial set up assuming that given the treatment effects and individual random effect, the repeated responses of an individual follow a specific autocorrelation structure. Remark that in the adaptive longitudinal fixed model, for given design weights, regression effects and longitudinal correlations are the only parameters to be estimated, whereas, in the adaptive longitudinal mixed model, along with these parameters the variance component of the random effects of individuals is also to be estimated. In practice, obtaining consistent and efficient estimates of the parameters involved in such an adaptive longitudinal mixed model is much more complicated as compared to the estimation of parameters of the adaptive longitudinal fixed model. This is because, unlike the adaptive longitudinal fixed model, in the adaptive longitudinal mixed model, it is essential to take the individual random effects into account for the estimation.

In the non-adaptive longitudinal mixed model set up, recently Sutradhar and Farrell (2003) and Sutradhar and Sinha (2002) [see also Sashegyi, Brown, and Farrell (2000)] studied the estimation procedure for the estimation of regression effects, the variance component of the individual random effects, and the longitudinal correlations. Thus, our adaptive longitudinal mixed model, introduced in Chapter 4, may be considered as a direct generalization of the longitudinal mixed models considered

by Sutradhar and Farrell (2003) and Sutradhar and Sinha (2002). Note that, unlike these authors, we have estimated the covariate effects and variance component of individual random effects by using the adaptive design based WGQL approach. We have exploited the method of moments to estimate the nuisance conditional longitudinal correlations. To examine the performance of the adaptive design based WGQL approach to estimate the regression and variance component parameters and the method of moments to estimate the longitudinal correlations, we have conducted a limited simulation study. It was found that the moment method performs extremely well in estimating the longitudinal correlations under the adaptive longitudinal mixed model. The WGQL approach appears to slightly overestimate the treatment effect with small standard error. For small values of the variance component, the WGQL approach performs well to estimate the parameter σ^2, whereas for large values of variance component, this approach underestimates the parameter σ^2. The estimated standard errors of the estimator of σ^2 were found to be small in magnitude as expected. In conclusion, our proposed adaptive design based WGQL approach performs well in estimating all the parameters of the proposed adaptive longitudinal binary mixed model.

5.2   Proposal for Future Research

As discussed in Chapters 3 and 4, in this thesis, we have analyzed the longitudinal fixed and mixed model-based binary data obtained from adaptive clinical trials. Note that, under the proposed models, the repeated responses were collected from an individual admitted to the adaptive clinical trial. It may however be necessary to include the members of a family instead of an individual, and to collect data over a period of time. This type of multidimensional data exhibits two-way correlations: first, the individuals of a family are likely to share a common random family effect causing familial correlations among the responses of the family; secondly, the repeated responses of the individuals of the same family are likely to be longitudinally correlated due to

the repetition, causing the familial correlated data also be longitudinally correlated. The analysis of this type of data is possible by generalizing the proposed adaptive longitudinal binary mixed model for a large number of individuals to the cases with large number of families. This is however beyond the scope of this thesis.

Appendix A

Tables: Adaptive Longitudinal Binary Fixed Model

Table A.1: Simulated means and standard errors of δ_s (total number of patients receiving the better treatment) for selected values of the true correlation parameter ρ under AR(1) binary model with $\beta_1 = 1.5$, $\beta_2 = 0.0$, $\beta_3 = 0.2$, and $\beta_4 = 0.1$; and adaptive design parameters $\alpha = 1.0$, $G = 3.0$, and $\tau = 2.0, 4.0$; for different values of $K = 75, 100, 200$.

K	τ	ρ	Mean	Standard Error
75	2.0	0.3	43.638	7.024
		0.5	43.480	7.022
		0.7	43.672	7.101
		0.9	43.839	7.190
	4.0	0.3	46.595	7.339
		0.5	46.023	7.701
		0.7	46.549	7.576
		0.9	46.566	8.135
100	2.0	0.3	58.703	8.505
		0.5	58.634	8.376
		0.7	58.632	8.588
		0.9	58.890	8.745
	4.0	0.3	62.483	8.779
		0.5	62.528	8.857
		0.7	62.348	9.047
		0.9	62.825	9.745

(Table A.1 contd....)

K	τ	ρ	Mean	Standard Error
200	2.0	0.3	116.660	11.097
		0.5	116.657	11.331
		0.7	116.291	11.451
		0.9	116.887	11.485
	4.0	0.3	124.693	11.668
		0.5	124.310	12.347
		0.7	123.675	12.349
		0.9	124.839	13.004

Table A.2: Simulated means(SM), simulated standard errors (SSE), and estimated standard errors (ESE) of the $WGQL_1$, $WGQL_2$, and $WGQL_3$ estimates for the regression and correlation parameters for selected values of the true correlation parameter ρ under AR(1) binary model with $\beta_1 = 1.5$, $\beta_2 = 0.0$, $\beta_3 = 0.2$, and $\beta_4 = 0.1$; and adaptive design parameters $\alpha = 1.0$, $G = 3.0$, and $\tau = 2.0, 4.0$; for $K = 75$ subjects.

Method	τ	ρ	Statistic	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
$WGQL_1$	2.0	0.3	SM	1.421	0.030	0.272	0.156	0.263	0.042	-0.021
			SSE	0.893	0.443	0.490	0.526	0.133	0.170	0.196
			ESE	1.350	0.460	0.535	0.555			
		0.5	SM	1.407	0.017	0.297	0.155	0.469	0.203	0.078
			SSE	0.962	0.495	0.538	0.582	0.118	0.172	0.207
			ESE	1.525	0.528	0.596	0.619			
		0.7	SM	1.379	0.071	0.277	0.124	0.681	0.460	0.306
			SSE	0.952	0.530	0.589	0.635	0.082	0.140	0.585
			ESE	1.718	0.585	0.688	0.708			
		0.9	SM	1.300	0.059	0.328	0.246	0.892	0.793	0.708
			SSE	1.048	0.604	0.720	0.724	0.044	0.087	0.119
			ESE	1.940	0.676	0.795	0.802			
	4.0	0.3	SM	1.499	0.012	0.226	0.126	0.255	0.028	-0.036
			SSE	0.852	0.458	0.529	0.550	0.135	0.177	0.204
			ESE	1.253	0.482	0.557	0.566			
		0.5	SM	1.468	-0.007	0.246	0.125	0.464	0.197	0.070
			SSE	0.920	0.512	0.582	0.613	0.124	0.171	0.204
			ESE	1.412	0.549	0.632	0.639			
		0.7	Mean	1.442	0.031	0.257	0.109	0.678	0.454	0.302
			SSE	0.936	0.556	0.630	0.651	0.083	0.141	0.177
			ESE	1.590	0.616	0.714	0.729			
		0.9	SM	1.331	0.063	0.294	0.202	0.892	0.793	0.707
			SSE	1.021	0.605	0.744	0.752	0.044	0.088	0.120
			ESE	1.816	0.697	0.832	0.845			

(Table A.2 contd....)

Method	τ	ρ	Statistic	Estimates						
				$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
$WGQL_2$	2.0	0.3	SM	1.360	0.061	0.299	0.184	0.276	0.058	-0.006
			SSE	0.893	0.420	0.474	0.526	0.123	0.162	0.193
			ESE	1.305	0.443	0.511	0.496			
		0.5	SM	1.332	0.047	0.333	0.193	0.477	0.213	0.089
			SSE	0.999	0.486	0.530	0.558	0.117	0.175	0.210
			ESE	1.438	0.495	0.565	0.586			
		0.7	SM	1.299	0.091	0.331	0.177	0.686	0.469	0.321
			SSE	0.972	0.514	0.572	0.627	0.080	0.136	0.174
			ESE	1.594	0.559	0.652	0.673			
		0.9	SM	1.213	0.094	0.362	0.282	0.895	0.799	0.715
			SSE	1.004	0.589	0.688	0.719	0.040	0.079	0.108
			ESE	1.805	0.644	0.749	0.774			
	4.0	0.3	SM	1.415	0.055	0.279	0.181	0.271	0.050	-0.015
			SSE	0.891	0.424	0.498	0.518	0.127	0.166	0.196
			ESE	1.252	0.447	0.522	0.537			
		0.5	SM	1.329	0.051	0.321	0.205	0.477	0.216	0.092
			SSE	0.948	0.497	0.538	0.559	0.122	0.182	0.209
			ESE	1.283	0.502	0.573	0.582			
		0.7	SM	1.330	0.070	0.329	0.193	0.684	0.466	0.317
			SSE	0.956	0.538	0.582	0.624	0.083	0.141	0.177
			ESE	1.477	0.572	0.655	0.668			
		0.9	SM	1.229	0.095	0.367	0.270	0.895	0.798	0.714
			SSE	0.976	0.569	0.684	0.711	0.042	0.082	0.115
			ESE	1.611	0.650	0.746	0.767			

(Table A.2 contd....)

Method	τ	ρ	Statistic	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
$WGQL_3$	2.0	0.3	SM	1.452	0.017	0.265	0.152	0.257	0.034	-0.031
			SSE	0.941	0.454	0.508	0.528	0.141	0.179	0.209
			ESE	1.124	0.459	0.537	0.544			
		0.5	SM	1.431	0.009	0.289	0.142	0.464	0.196	0.070
			SSE	1.007	0.504	0.543	0.597	0.124	0.181	0.217
			ESE	1.265	0.521	0.597	0.618			
		0.7	SM	1.481	0.036	0.262	0.101	0.670	0.441	0.286
			SSE	1.110	0.568	0.624	0.676	0.099	0.164	0.202
			ESE	1.529	0.609	0.714	0.740			
		0.9	SM	1.339	0.061	0.312	0.218	0.890	0.791	0.704
			SSE	1.238	0.632	0.773	0.795	0.046	0.089	0.121
			ESE	1.771	0.693	0.851	0.863			
	4.0	0.3	SM	1.554	-0.009	0.201	0.111	0.244	0.014	-0.050
			SSE	0.933	0.489	0.558	0.573	0.145	0.193	0.216
			ESE	0.995	0.485	0.569	0.568			
		0.5	SM	1.536	-0.029	0.212	0.094	0.451	0.178	0.047
			SSE	1.045	0.538	0.622	0.652	0.137	0.196	0.236
			ESE	1.171	0.556	0.660	0.666			
		0.7	SM	1.531	-0.002	0.226	0.084	0.670	0.440	0.282
			SSE	1.068	0.603	0.665	0.699	0.095	0.161	0.204
			ESE	1.315	0.631	0.738	0.748			
		0.9	SM	1.432	0.023	0.259	0.172	0.889	0.787	0.700
			SSE	1.182	0.653	0.799	0.811	0.048	0.094	0.128
			ESE	1.543	0.713	0.861	0.865			

Table A.3: Simulated means(SM), simulated standard errors (SSE), and estimated standard errors (ESE) of the $WGQL_1$, $WGQL_2$, and $WGQL_3$ estimates for the regression and correlation parameters for selected values of the true correlation parameter ρ under AR(1) binary model with $\beta_1 = 1.5$, $\beta_2 = 0.0$, $\beta_3 = 0.2$, and $\beta_4 = 0.1$; and adaptive design parameters $\alpha = 1.0$, $G = 3.0$, and $\tau = 2.0, 4.0$; for $K = 100$ subjects.

Method	τ	ρ	Statistic	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
						Estimates				
$WGQL_1$	2.0	0.3	SM	1.541	-0.002	0.215	0.101	0.263	0.040	-0.025
			SSE	0.787	0.396	0.425	0.415	0.125	0.157	0.181
			ESE	1.051	0.426	0.430	0.418			
		0.5	SM	1.502	0.021	0.232	0.103	0.474	0.211	0.085
			SSE	0.805	0.408	0.459	0.466	0.096	0.136	0.169
			ESE	1.179	0.473	0.480	0.472			
		0.7	SM	1.523	-0.003	0.257	0.115	0.677	0.450	0.287
			SSE	0.918	0.507	0.537	0.530	0.075	0.128	0.166
			ESE	1.402	0.557	0.557	0.553			
		0.9	SM	1.428	0.061	0.262	0.124	0.894	0.798	0.713
			SSE	0.930	0.511	0.602	0.597	0.037	0.070	0.097
			ESE	1.520	0.605	0.640	0.623			
	4.0	0.3	SM	1.544	-0.017	0.211	0.099	0.266	0.042	-0.021
			SSE	0.718	0.417	0.433	0.429	0.117	0.149	0.175
			ESE	0.933	0.431	0.432	0.421			
		0.5	SM	1.539	-0.005	0.214	0.085	0.473	0.208	0.081
			SSE	0.758	0.445	0.489	0.488	0.096	0.141	0.172
			ESE	1.065	0.487	0.493	0.483			
		0.7	SM	1.580	-0.047	0.228	0.094	0.676	0.449	0.288
			SSE	0.860	0.538	0.558	0.545	0.072	0.122	0.157
			ESE	1.251	0.578	0.571	0.557			
		0.9	SM	1.501	0.020	0.233	0.109	0.893	0.796	0.710
			SSE	0.928	0.550	0.638	0.611	0.037	0.072	0.099
			ESE	1.464	0.647	0.664	0.644			

(Table A.3 contd....)

Method	τ	ρ	Statistic	Estimates						
				$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
$WGQL_2$	2.0	0.3	SM	1.465	0.036	0.251	0.128	0.275	0.057	-0.009
			SSE	0.810	0.374	0.403	0.407	0.120	0.150	0.176
			ESE	1.034	0.403	0.414	0.407			
		0.5	SM	1.471	0.052	0.249	0.121	0.478	0.217	0.092
			SSE	0.872	0.387	0.462	0.463	0.098	0.141	0.172
			ESE	1.188	0.451	0.472	0.466			
		0.7	SM	1.439	0.053	0.290	0.146	0.684	0.461	0.303
			SSE	0.920	0.465	0.506	0.509	0.070	0.120	0.151
			ESE	1.312	0.512	0.526	0.523			
		0.9	SM	1.365	0.107	0.286	0.162	0.895	0.802	0.718
			SSE	0.976	0.493	0.576	0.569	0.035	0.067	0.094
			ESE	1.497	0.574	0.609	0.590			
	4.0	0.3	SM	1.485	0.032	0.243	0.120	0.274	0.053	-0.010
			SSE	0.761	0.387	0.417	0.427	0.116	0.149	0.177
			ESE	0.920	0.406	0.419	0.414			
		0.5	SM	1.489	0.045	0.243	0.109	0.477	0.214	0.088
			SSE	0.832	0.410	0.483	0.486	0.099	0.146	0.176
			ESE	1.069	0.457	0.487	0.481			
		0.7	SM	1.478	0.026	0.282	0.147	0.683	0.460	0.304
			SSE	0.851	0.483	0.519	0.512	0.069	0.117	0.147
			ESE	1.164	0.519	0.531	0.524			
		0.9	SM	1.400	0.087	0.289	0.159	0.895	0.801	0.716
			SSE	0.935	0.509	0.606	0.585	0.037	0.070	0.099
			ESE	1.1361	0.586	0.623	0.604			

(Table A.3 contd....)

Method	τ	ρ	Statistic	Estimates						
				$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
$WGQL_3$	2.0	0.3	SM	1.550	-0.007	0.212	0.098	0.261	0.039	-0.027
			SSE	0.793	0.401	0.434	0.420	0.128	0.158	0.182
			ESE	0.855	0.421	0.432	0.418			
		0.5	SM	1.540	0.004	0.216	0.095	0.469	0.202	0.074
			SSE	0.862	0.426	0.488	0.478	0.106	0.154	0.189
			ESE	0.966	0.471	0.486	0.472			
		0.7	SM	1.578	-0.024	0.243	0.104	0.671	0.440	0.276
			SSE	0.976	0.526	0.560	0.558	0.082	0.140	0.180
			ESE	1.143	0.553	0.561	0.551			
		0.9	SM	1.501	0.030	0.235	0.112	0.892	0.794	0.707
			SSE	1.025	0.535	0.621	0.604	0.038	0.074	0.103
			ESE	1.342	0.628	0.658	0.633			
	4.0	0.3	SM	1.567	-0.027	0.202	0.092	0.262	0.037	-0.026
			SSE	0.751	0.438	0.453	0.450	0.125	0.157	0.180
			ESE	0.738	0.428	0.439	0.427			
		0.5	SM	1.557	-0.016	0.208	0.081	0.470	0.203	0.075
			SSE	0.807	0.460	0.498	0.496	0.100	0.149	0.178
			ESE	0.844	0.486	0.495	0.484			
		0.7	SM	1.635	-0.067	0.214	0.085	0.672	0.441	0.278
			SSE	0.915	0.554	0.568	0.551	0.078	0.134	0.170
			ESE	0.984	0.570	0.567	0.552			
		0.9	SM	1.580	-0.014	0.203	0.100	0.890	0.792	0.704
			SSE	1.013	0.577	0.683	0.616	0.040	0.078	0.106
			ESE	1.157	0.647	0.671	0.637			

Table A.4: Simulated means(SM), simulated standard errors, (SSE) and estimated standard errors (ESE) of the $WGQL_1$, $WGQL_2$, and $WGQL_3$ estimates for the regression and correlation parameters for selected values of the true correlation parameter ρ under AR(1) binary model with $\beta_1 = 1.5$, $\beta_2 = 0.0$, $\beta_3 = 0.2$, and $\beta_4 = 0.1$; and adaptive design parameters $\alpha = 1.0$, $G = 3.0$, and $\tau = 2.0, 4.0$; for $K = 200$ subjects.

| Method | τ | ρ | Statistic | Estimates | | | | | | |
				$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
$WGQL_1$	2.0	0.3	SM	1.575	-0.025	0.199	0.101	0.272	0.054	-0.011
			SSE	0.613	0.325	0.289	0.310	0.095	0.118	0.138
			ESE	0.743	0.320	0.290	0.304			
		0.5	SM	1.560	-0.016	0.203	0.093	0.479	0.217	0.085
			SSE	0.662	0.343	0.326	0.323	0.077	0.112	0.139
			ESE	0.837	0.358	0.325	0.339			
		0.7	SM	1.538	-0.026	0.223	0.124	0.686	0.464	0.310
			SSE	0.720	0.393	0.365	0.400	0.060	0.102	0.131
			ESE	0.939	0.405	0.364	0.382			
		0.9	SM	1.580	-0.046	0.241	0.116	0.894	0.798	0.712
			SSE	0.821	0.465	0.421	0.434	0.029	0.055	0.077
			ESE	1.135	0.479	0.420	0.446			
	4.0	0.3	SM	1.571	-0.028	0.195	0.095	0.274	0.057	-0.009
			SSE	0.556	0.337	0.289	0.314	0.089	0.114	0.133
			ESE	0.665	0.323	0.292	0.306			
		0.5	SM	1.563	-0.029	0.199	0.090	0.480	0.219	0.087
			SSE	0.610	0.365	0.331	0.331	0.076	0.108	0.134
			ESE	0.748	0.362	0.326	0.341			
		0.7	SM	1.537	-0.038	0.219	0.112	0.686	0.465	0.312
			SSE	0.656	0.402	0.374	0.406	0.055	0.093	0.118
			ESE	0.840	0.410	0.366	0.383			
		0.9	SM	1.621	-0.063	0.208	0.085	0.893	0.797	0.711
			SSE	0.748	0.466	0.431	0.456	0.028	0.053	0.075
			ESE	1.044	0.499	0.436	0.465			

(Table A.4 contd....)

Method	τ	ρ	Statistic	Estimates						
				$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
$WGQL_2$	2.0	0.3	SM	1.558	0.003	0.208	0.109	0.275	0.058	-0.008
			SSE	0.682	0.304	0.282	0.308	0.097	0.122	0.143
			ESE	0.769	0.311	0.286	0.303			
		0.5	SM	1.558	0.008	0.216	0.103	0.480	0.219	0.087
			SSE	0.736	0.320	0.322	0.320	0.080	0.116	0.142
			ESE	0.902	0.349	0.322	0.337			
		0.7	SM	1.522	0.006	0.235	0.132	0.688	0.468	0.313
			SSE	0.771	0.356	0.361	0.397	0.055	0.094	0.125
			ESE	0.973	0.390	0.358	0.377			
		0.9	SM	1.514	0.004	0.265	0.138	0.896	0.801	0.717
			SSE	0.856	0.424	0.405	0.429	0.029	0.054	0.077
			ESE	1.122	0.449	0.409	0.434			
	4.0	0.3	SM	1.561	0.000	0.207	0.101	0.275	0.058	-0.009
			SSE	0.637	0.313	0.283	0.311	0.096	0.123	0.146
			ESE	0.694	0.315	0.290	0.306			
		0.5	SM	1.569	-0.003	0.216	0.103	0.478	0.217	0.084
			SSE	0.710	0.343	0.328	0.330	0.084	0.121	0.150
			ESE	0.798	0.352	0.325	0.340			
		0.7	SM	1.525	-0.001	0.238	0.133	0.687	0.468	0.315
			SSE	0.756	0.372	0.370	0.394	0.057	0.096	0.124
			ESE	0.894	0.395	0.363	0.380			
		0.9	SM	1.585	-0.007	0.229	0.110	0.894	0.798	0.711
			SSE	0.842	0.424	0.433	0.457	0.031	0.059	0.083
			ESE	1.052	0.458	0.424	0.446			

(Table A.4 contd....)

Method	τ	ρ	Statistic	Estimates						
				$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
$WGQL_3$	2.0	0.3	SM	1.575	-0.025	0.199	0.101	0.272	0.054	-0.011
			SSE	0.612	0.326	0.289	0.310	0.095	0.118	0.138
			ESE	0.605	0.316	0.289	0.303			
		0.5	SM	1.576	-0.023	0.198	0.092	0.477	0.213	0.080
			SSE	0.683	0.353	0.330	0.325	0.084	0.121	0.148
			ESE	0.686	0.356	0.326	0.339			
		0.7	SM	1.563	-0.038	0.220	0.120	0.683	0.460	0.305
			SSE	0.750	0.407	0.367	0.403	0.064	0.110	0.137
			ESE	0.776	0.405	0.365	0.383			
		0.9	SM	1.629	-0.060	0.225	0.101	0.892	0.795	0.707
			SSE	0.872	0.475	0.449	0.468	0.032	0.060	0.085
			ESE	0.950	0.482	0.432	0.458			
	4.0	0.3	SM	1.570	-0.027	0.195	0.095	0.275	0.058	-0.008
			SSE	0.552	0.336	0.289	0.314	0.088	0.112	0.131
			ESE	0.515	0.317	0.292	0.305	0.478		
		0.5	SM	1.574	-0.035	0.195	0.089	0.478	0.216	0.083
			SSE	0.631	0.376	0.335	0.332	0.086	0.120	0.152
			ESE	0.585	0.360	0.329	0.341			
		0.7	SM	1.550	-0.044	0.215	0.110	0.685	0.464	0.311
			SSE	0.672	0.411	0.378	0.407	0.056	0.095	0.121
			ESE	0.658	0.406	0.367	0.383			
		0.9	SM	1.648	-0.074	0.202	0.075	0.892	0.795	0.708
			SSE	0.788	0.487	0.441	0.0476	0.030	0.056	0.079
			ESE	0.796	0.486	0.435	0.459			

Table A.5: Comparison of mean squared errors (MSE) for the estimators of the treatment effect(β_1) under three weighted generalized quasi-likelihood approaches, based on 1000 simulations.

K	Method	$\tau = 2.0$				$\tau = 4.0$			
		ρ				ρ			
		0.3	0.5	0.7	0.9	0.3	0.5	0.7	0.9
75	$WGQL_1$	0.803	0.934	0.921	1.137	0.726	0.847	0.880	1.071
	$WGQL_2$	0.817	1.026	0.985	1.090	0.800	0.929	0.943	1.026
	$WGQL_3$	0.888	1.019	1.232	1.560	0.874	1.093	1.142	1.401
100	$WGQL_1$	0.621	0.648	0.843	0.871	0.518	0.576	0.746	0.862
	$WGQL_2$	0.657	0.761	0.850	0.970	0.579	0.693	0.724	0.885
	$WGQL_3$	0.631	0.745	0.958	1.051	0.569	0.654	0.855	1.033
200	$WGQL_1$	0.381	0.442	0.519	0.681	0.315	0.377	0.432	0.573
	$WGQL_2$	0.468	0.545	0.594	0.732	0.410	0.509	0.572	0.716
	$WGQL_3$	0.380	0.473	0.566	0.777	0.310	0.404	0.454	0.643

Table A.6: Comparison of SLPW and randomized designs based mean squared errors (MSEs) of the estimates of the regression parameters of a binary AR(1) longitudinal model with true regression parameters $\beta_1 = 1.5, \beta_2 = 0.0, \beta_3 = 0.20, \beta_4 = 0.10$ and AR(1) correlation parameter $\rho = 0.5, 0.9$, based on two selected values of $\tau = 2.0$ and 4.0, for $K = 75, 100,$ and 200 subjects.

| | | | | | Estimates | | | |
True design	Working design	K	τ	ρ	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$w_i \geq 0.5$	$w_i \geq 0.5$	75	2.0	0.5	1.026	0.238	0.299	0.320
				0.9	1.090	0.347	0.500	0.550
			4.0	0.5	0.928	0.250	0.304	0.324
				0.9	1.026	0.333	0.496	0.534
		100	2.0	0.5	0.761	0.152	0.216	0.215
				0.9	0.970	0.254	0.339	0.328
			4.0	0.5	0.692	0.170	0.235	0.236
				0.9	0.884	0.267	0.375	0.346
		200	2.0	0.5	0.545	0.102	0.104	0.102
				0.9	0.733	0.180	0.168	0.185
			4.0	0.5	0.509	0.118	0.108	0.109
				0.9	0.716	0.180	0.188	0.209

(Table A.6 contd....)

True design	Working design	K	τ	ρ	Estimates			
					$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$w_i \geq 0.5$	$w_i = 0.5$	75	2.0	0.5	1.047	0.215	0.275	0.304
				0.9	1.220	0.321	0.447	0.485
			4.0	0.5	0.985	0.227	0.290	0.312
				0.9	1.198	0.346	0.471	1.503
		100	2.0	0.5	0.887	0.170	0.219	0.214
				0.9	1.008	0.250	0.371	0.311
			4.0	0.5	0.962	0.165	0.217	0.216
				0.9	1.045	0.251	0.382	0.329
		200	2.0	0.5	0.869	0.103	0.100	0.112
				0.9	0.954	0.152	0.146	0.156
			4.0	0.5	1.085	0.107	0.105	0.121
				0.9	1.004	0.149	0.153	0.167

Table A.7: Coverage probabilities for β_1 using $WGQL_2$ approach based on 1000 simulations.

K	τ	ρ	Coverage probability
100	2.0	0.3	0.94
		0.5	0.95
		0.7	0.95
		0.9	0.96
	4.0	0.3	0.95
		0.5	0.96
		0.7	0.95
		0.9	0.95
200	2.0	0.3	0.95
		0.5	0.94
		0.7	0.96
		0.9	0.93
	4.0	0.3	0.96
		0.5	0.96
		0.7	0.95
		0.9	0.95

Appendix B

Table: Adaptive Longitudinal Binary Mixed Model

Table B.1: Simulated means(SM), simulated standard errors (SSE), and estimated standard errors (ESE) of the WGQL estimates for the variance component of random effect, regression and correlation parameters for selected values of variance component σ^2 and the true correlation parameter ρ under AR(1) binary model with $\beta = 1.5$, and adaptive design parameters $\alpha = 1.0$, $G = 0.0$, and $\tau = 2.0$; for $K = 100$ subjects.

σ^2	ρ	Statistic	Estimates				
			$\hat{\sigma}^2$	$\hat{\beta}$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
0.10	0.30	SM	0.128	1.514	0.298	0.0764	0.010
		SSE	0.175	0.289	0.086	0.109	0.136
		ESE	0.168	0.292			
	0.50	SM	0.142	1.518	0.490	0.242	0.118
		SSE	0.147	0.341	0.084	0.123	0.153
		ESE	0.191	0.327			
	0.70	SM	0.142	1.541	0.689	0.470	0.317
		SSE	0.129	0.422	0.065	0.110	0.136
		ESE	0.108	0.381			
	0.80	SM	0.131	1.523	0.796	0.632	0.496
		SSE	0.160	0.410	0.045	0.081	0.116
		ESE	0.136	0.393			
	0.90	SM	0.112	1.556	0.899	0.807	0.720
		SSE	0.119	0.424	0.035	0.065	0.093
		ESE	0.126	0.473			
	0.95	SM	0.109	1.554	0.950	0.901	0.854
		SSE	0.155	0.448	0.023	0.046	0.068
		ESE	0.191	0.456			

(Table B.1 contd....)

σ^2	ρ	Statistic	Estimates				
			$\hat{\sigma}^2$	$\hat{\beta}$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
0.20	0.30	SM	0.192	1.524	0.296	0.081	0.015
		SSE	0.160	0.286	0.089	0.111	0.133
		ESE	0.120	0.299			
	0.50	SM	0.209	1.530	0.491	0.244	0.120
		SSE	0.165	0.355	0.082	0.120	0.149
		ESE	0.227	0.335			
	0.70	SM	0.201	1.571	0.691	0.474	0.323
		SSE	0.214	0.450	0.066	0.112	0.143
		ESE	0.235	0.399			
	0.80	SM	0.206	1.600	0.794	0.626	0.491
		SSE	0.271	0.527	0.053	0.098	0.133
		ESE	0.322	0.466			
	0.90	SM	0.191	1.614	0.898	0.804	0.718
		SSE	0.186	0.586	0.037	0.067	0.094
		ESE	0.135	0.510			
	0.95	SM	0.199	1.585	0.952	0.903	0.857
		SSE	0.240	0.534	0.023	0.049	0.070
		ESE	0.161	0.507			

(Table B.1 contd....)

σ^2	ρ	Statistic	Estimates				
			$\hat{\sigma}^2$	$\hat{\beta}$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
0.50	0.30	SM	0.392	1.574	0.297	0.083	0.023
		SSE	0.195	0.308	0.090	0.110	0.135
		ESE	0.167	0.326			
	0.50	SM	0.393	1.581	0.492	0.248	0.125
		SSE	0.169	0.359	0.078	0.110	0.134
		ESE	0.103	0.359			
	0.70	SM	0.423	1.584	0.691	0.477	0.328
		SSE	0.202	0.431	0.064	0.107	0.138
		ESE	0.120	0.405			
	0.80	SM	0.416	1.574	0.797	0.636	0.507
		SSE	0.149	0.366	0.044	0.084	0.115
		ESE	0.123	0.407			
	0.90	SM	0.435	1.625	0.895	0.801	0.718
		SSE	0.144	0.519	0.036	0.070	0.094
		ESE	0.143	0.483			
	0.95	SM	0.462	1.617	0.949	0.900	0.855
		SSE	0.147	0.569	0.025	0.050	0.071
		ESE	0.156	0.518			

(Table B.1 contd....)

σ^2	ρ	Statistic	Estimates				
			$\hat{\sigma}^2$	$\hat{\beta}$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
0.75	0.30	SM	0.570	1.606	0.298	0.088	0.025
		SSE	0.165	0.300	0.086	0.108	0.132
		ESE	0.090	0.343			
	0.50	SM	0.573	1.600	0.494	0.250	0.132
		SSE	0.187	0.348	0.080	0.109	0.132
		ESE	0.109	0.370			
	0.70	SM	0.617	1.593	0.694	0.481	0.328
		SSE	0.169	0.345	0.055	0.094	0.124
		ESE	0.124	0.397			
	0.80	SM	0.656	1.636	0.793	0.627	0.495
		SSE	0.168	0.436	0.050	0.091	0.126
		ESE	0.143	0.443			
	0.90	SM	0.680	1.588	0.898	0.806	0.727
		SSE	0.117	0.408	0.033	0.063	0.088
		ESE	0.159	0.442			
	0.95	SM	0.704	1.683	0.948	0.900	0.851
		SSE	0.149	0.566	0.025	0.050	0.073
		ESE	0.178	0.517			

(Table B.1 contd....)

σ^2	ρ	Statistic	Estimates				
			$\hat{\sigma}^2$	$\hat{\beta}$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
1.00	0.30	SM	0.784	1.627	0.301	0.091	0.029
		SSE	0.174	0.311	0.088	0.110	0.136
		ESE	0.100	0.361			
	0.50	SM	0.803	1.626	0.494	0.250	0.128
		SSE	0.142	0.354	0.079	0.110	0.136
		ESE	0.115	0.387			
	0.70	SM	0.876	1.613	0.694	0.480	0.327
		SSE	0.173	0.353	0.055	0.096	0.125
		ESE	0.137	0.412			
	0.80	SM	0.893	1.644	0.792	0.623	0.497
		SSE	0.175	0.418	0.049	0.089	0.114
		ESE	0.159	0.450			
	0.90	SM	0.918	1.665	0.894	0.799	0.715
		SSE	0.199	0.490	0.036	0.069	0.096
		ESE	0.182	0.481			
	0.95	SM	0.937	1.621	0.949	0.900	0.856
		SSE	0.205	0.436	0.024	0.048	0.068
		ESE	0.191	0.475			

Appendix C

Graphs: Adaptive Longitudinal Binary Fixed Model

Figure C.1: Adaptive design weights (w_i) and expected weights (w_{i0}) for $K = 75$: $w_i :$ ———; $w_{i0} : \ldots\ldots\ldots$, for selected τ(tau) and ρ(rho).
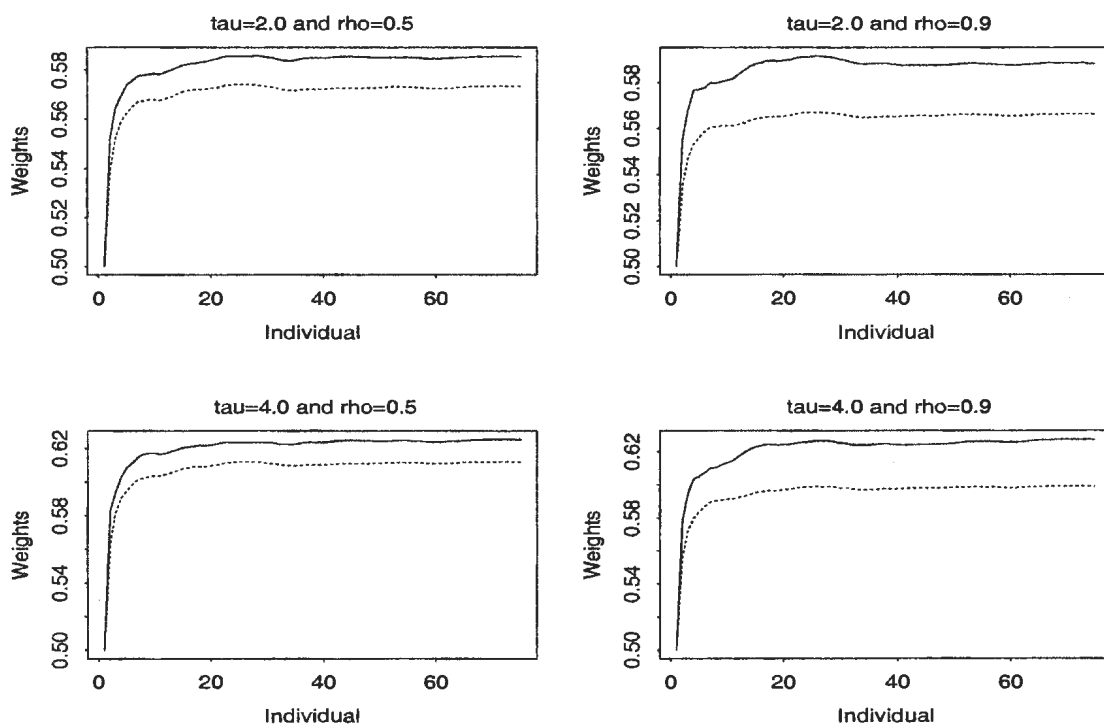
Figure C.2: Adaptive design weights (w_i) and expected weights (w_{i0}) for $K = 100$: w_i : ———; w_{i0} : ...., for selected τ(tau) and ρ(rho).
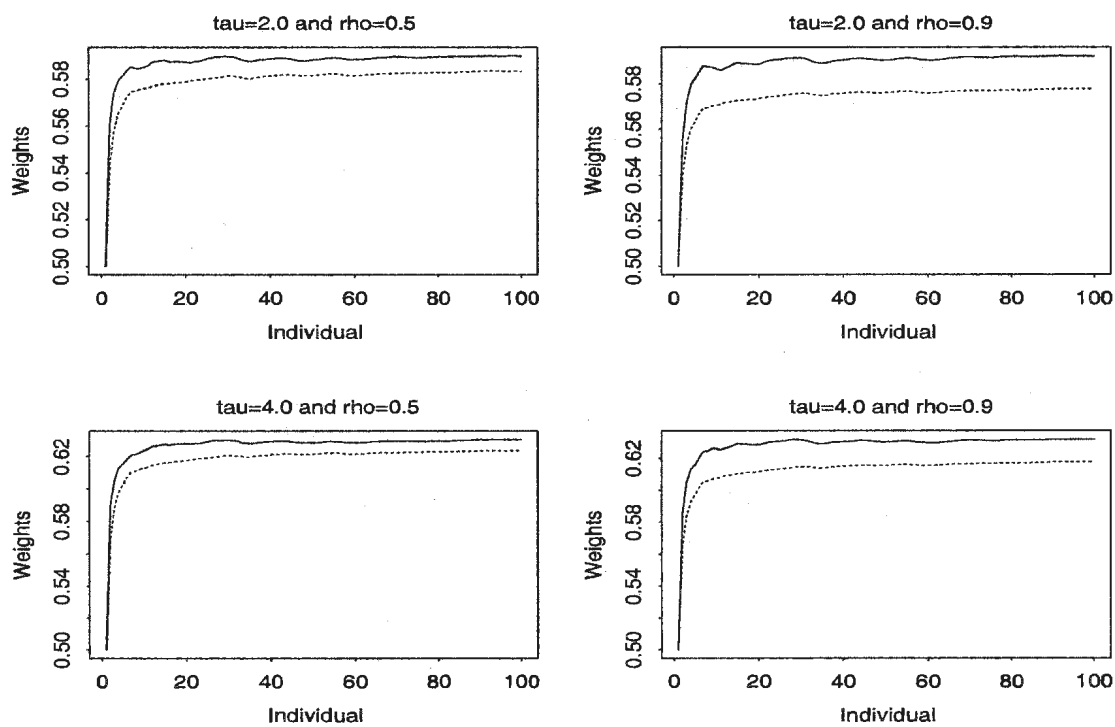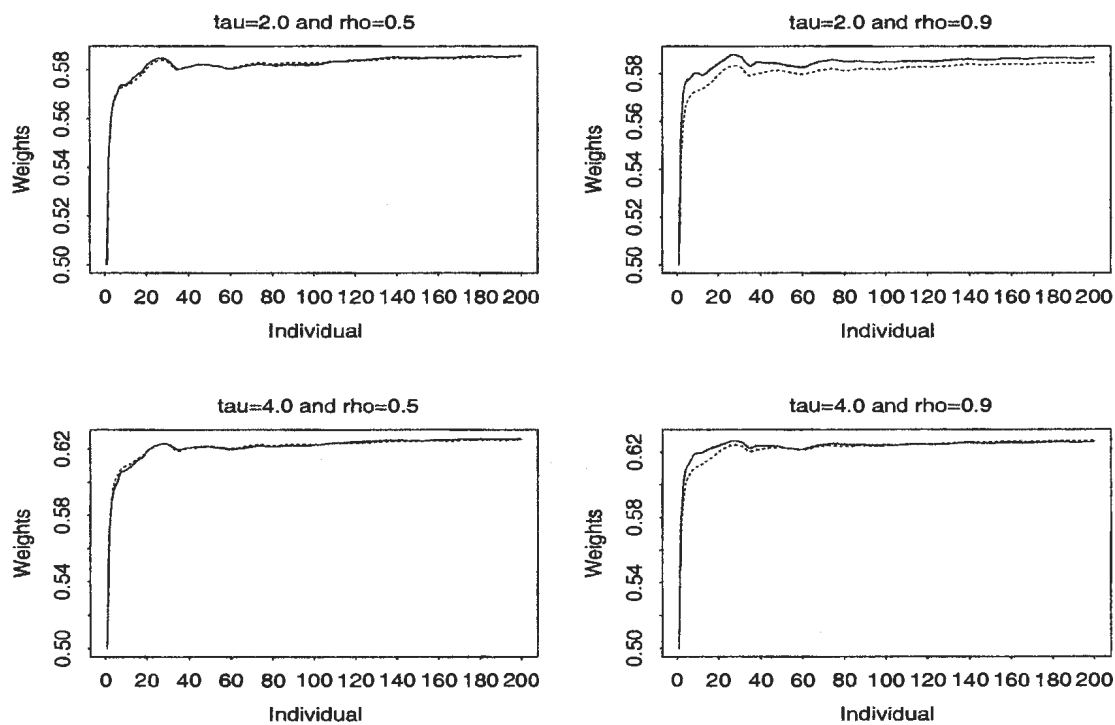
Figure C.3: Adaptive design weights (w_i) and expected weights (w_{i0}) for $K = 200$: w_i : ———; w_{i0} : . . . ., for selected τ(tau) and ρ(rho).

Bibliography

[1] ANSCOMBE, F.J. (1963). Sequential medical trials. *Journal of the American Statistical Association*, **58**, 365-383.

[2] ARMITAGE, P. (1975). *Sequential Medical Trials*. New York: John Wiley and Sons.

[3] ATHREYA, K.B. & KARLIN, S. (1968). Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Annals of Mathematical Statistics*, **39**, 1801-1817.

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

[5] BANDYOPADHYAY, U. & BISWAS, A. (1999). Allocation by randomized play-the-winner rule in the presence of prognostic factors. *Sankhya A*, **61**, 396-412.

[6] BANDYOPADHYAY, U. & BISWAS, A. (2001). Adaptive designs for normal responses with prognostic factors. *Biometrika*, **88**, 409-419.

[7] BEGG, C. B. (1990). On inference from Wei's biased coin design for clinical trials. *Biometrika*, **77**, 467-484.

[8] BEGG, C. B. & IGLEWICZ, B. (1980). A treatment allocation procedure for sequential clinical trials. *Biometrics*, **63**, 81-90.

[9] COLTON, T. (1963). A model for selecting one of the two medical treatments. *Journal of the American Statistical Association*, **58**, 388-400.

[10] CORNFIELD, J., HALPRIN, M., & GREENHOUSE, S. W. (1969). An adaptive procedure for sequential clinical trials. *Journal of the American Statistical Association*, **64**, 759-770.

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

[12] DAVIS, R. A., DUNSMUIR, W.T.M., & WANG, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika*, **87**, 491-506.

[13] EFRON, B. (1980). Randomizing and balancing a complicated sequential experiment. In *Biostatistics Case-Book*, Ed. R. G. Miller, B. Efron,, B. W. Brown, & L. E. Moses, 19-30, New York: Wiley.

[14] FAHRMEIR, L. & TUTZ, G. T. (1994). *Multivariate statistical modelling based on generalized linear models*. New York, Springer-Verlag.

[15] FREEDMAN, D. A. (1965). Bernard Friedman's urn. *Annals of Mathematical Statistics*, **36**, 956-970.

[16] GANTMACHER, F. R. (1959). *Matrix Theory 2*. Chelsea, New York.

[17] HEAGERTY, P. J. (1999). Marginally specified logistic-normal models for longitudinal binary data. *Biometrics*, **55**, 688-698.

[18] JOWAHEER,V. & SUTRADHAR, B. C. (2002). Analyzing longitudinal count data with overdispersion. *Biometrika*, **89**, 389-399.

[19] LIANG, K. Y. & ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22.

[20] McCULLAGH, P. (1983). Quasilikelihood functions. *The Annals of Statistics*, **11**, 59-67.

[21] POCOCK, S. J. & SIMON, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*, **31**, 103-115.

[22] PRENTICE, R. L. & ZHAO, L. P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics*, **47**, 825-839.

[23] ROSENBERGER, W. F. (1993). Asymptotic inference with response-adaptive treatment allocation designs. *The Annals of Statistics*, **21**, 2098-2107.

[24] ROSENBERGER, W. F., FLOURNOY, N., & DURHAM, S. D. (1997). Asymptotic normality of maximum likelihood estimators from multiparameter response-driven designs. *Journal of Statistical Planning and Inference*, **60**, 69-76.

[25] ROSENBERGER, W. F. & SRIRAM, T. N. (1997). Estimation for an adaptive allocation design. *Journal of Statistical Planning and Inference*, **59**, 309-319.

[26] SASHEGYI, A. I., BROWN, K. S., & FARRELL, P. J. (2000). Estimation in an empirical Bayes model for longitudinal data and cross-sectionally clustered binary data. *The Canadian Journal of Statistics*, **28**, 45-64.

[27] SIMON, R., WEISS, G. H., & HOEL, D.G. (1975). Sequential analysis of binomial clinical trials. *Biometrika*, **62**, 195-200.

[28] SMITH, R. L. (1984). Properties of biased coin designs in sequential trials. *The Annals of Statistics*, **12**, 1018-1034.

[29] SMYTHE, R. T. & ROSENBERGER, W. F. (1995). Play-the-winner designs, generalized Pólya urns, and Markov branching processes. In *Adaptive Designs* (IMS Lecture Notes-Monograph Series, **Vol. 25**), N. Flournoy and W. F. Rossenberger (eds.). IMS.

[30] SUTRADHAR, B. C. & BISWAS, A. (2001). Marginal regression for binary longitudinal data in adaptive clinical trials. *Technical Report*, Department of Mathematics and Statistics, Memorial University of Newfoundland, Canada.

[31] SUTRADHAR, B. C. & DAS, K. (1995). Analysing exponential family based messy data. *Communication in Statistics - Theory and Methods*, **24**, 2683-2699.

[32] SUTRADHAR, B. C. & DAS, K. (1999). On the efficiency of regression estimators in generalized linear models for longitudinal data. *Biometrika*, **86**, 459-465.

[33] SUTRADHAR, B. C. & FARRELL, P. J. (2003). Analyzing multivariate longitudinal binary data: A generalized estimating equation approach. *Submitted.*

[34] SUTRADHAR, B. C. & KOVACEVIC, M. (2000). Analysing ordinal longitudinal survey data: Generalized estimating equations approach. *Biometrika*, **87**, 837-848.

[35] SUTRADHAR, B. C. & KUMAR, P. (2001). On the efficiency of extended generalized estimating equation approaches. *Statistics and Probability Letters*, **55**, 53-61.

[36] SUTRADHAR, B. C. & SINHA, S. K. (2002). On pseudo-likelihood inference in the binary longitudinal mixed model. *Communication in Statistics - Theory and Methods*, **31**, 397-417.

[37] WEDDERBURN, R. (1979). Quasi-likelihood functions, generalized linear models, and Gauss-Newton method. *Biometrika*, **61**, 439-447.

[38] WEI, L. J. (1978). An application of an urn model to the design of sequential control clinical trials. *Journal of the American Statistical Association*, **73**, 559-563.

[39] WEI, L. J. (1979). The generalized Pólya urns for sequential medical trials. *The Annals of Statistics*, **7**, 291-296.

[40] WEI, L. J. (1988). Exact two sample permutation tests based on the play-the-winner rule. *Biometrika*, **75**, 603-606.

[41] WEI, L. J. & DURHAM, S. (1978). The randomized play-the-winner rule in medical trials. *Journal of the American Statistical Association*, **73**, 840-843.

[42] WEI, L. J., SMYTHE, R. T., LIN, D. Y., & PARK, T. S. (1990). Statistical inference with data-dependent treatment allocation rules. *Journal of the American Statistical Association*, **85**, 156-162.

[43] WEI, L. J., SMYTHE, R. T., & MEHTA, C. R. (1989). Interval estimation with restricted randomization rules. *Biometrika*, **76**, 363-368.

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

[45] ZEGER, S.L. (1988). A regression model for time series of counts. *Biometrika*, **75**, 621-629.

[46] ZEGER, S.L., LIANG, K. Y., & SELF, S. G. (1985). The analysis of binary longitudinal data with time independent covariates. *Biometrika*, **72**, 31-38.