... Vol 32: 93–113, 2005

Marginal Regression for Binary Longitudinal Data in Adaptive Clinical Trials

BRAJENDRA C. SUTRADHAR

Department of Mathematics and Statistics, Memorial University of Newfoundland

ATANU BISWAS

Indian Statistical Institute

WASIMUL BARI

Department of Mathematics and Statistics, Memorial University of Newfoundland

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 the 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 estimates for the treatment effect as well as the effects of the other covariates in such a binary longitudinal set up. In this paper, 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 approach for the consistent and efficient estimation of the regression parameters including the treatment effects.

Key words: consistency, design misspecification effect, efficiency, longitudinal adaptive design, repeated binary responses, robust correlation structure

1. Introduction

For clinical trials on humans, it is highly desirable that one uses certain data-dependent treatment allocation rules that exploit accumulating information to assign individuals to treatments so that more study subjects are assigned to the better treatment. For example, consider a clinical study on asthma prevention. Suppose that there are two competitive treatments A and B available to treat the asthma patients. In this clinical study, it is important to treat a new patient with the better treatment. For this, although the overall effect of particular treatments are unknown, the physician must use the existing information about the effectiveness of these competitive treatments before assigning a new patient to the so-called better treatment. Suppose that the clinical study consists of K patients, and δ is a covariate representing the treatment selection, so that δ_i denote the selection of the treatment A or B for the ith ($i = 1, \ldots, K$) patient. It is then clear that this treatment covariate (δ) can neither be fixed nor completely random. The levels of such a covariate, i.e. the values of δ_i ($i = 1, \ldots, K$), are rather determined by using a suitable sequential adaptive design. For this, a special sequential adaptive design was first constructed by Zelen (1969) based on the so-called play-the-winner (PW) rule. Later, as a modification of Zelen's (PW) rule, Wei & Durham (1978) and Wei (1979) introduced the idea of the randomized play-the-winner (RPW) rule to construct better adaptive designs. For the cases where treatments are applied to obtain binary responses, Wei *et al.* (1990) and Smythe & Rosenberger (1995) studied the RPW rule-based sequential adaptive designs in allocating the patients to the better treatment. Note

however that the constructions of the sequential adaptive designs in the above studies were confined to the situations where responses are considered to be the effect of the treatment only. But there are cases in practice where the response of an individual patient may also be affected by other covariates (prognostic factors) on top of the treatment covariate. For example, in the asthma prevention study, the response of an incoming patient may be affected by the treatment as well as by certain relevant prognostic factors such as age, chronic conditions and smoking habit of the patient. Recently, some authors have modified the RPW rule to include the contributions of this type of prognostic factors in constructing the adaptive designs for the treatment covariate. For example, Bandyopadhyay & Biswas (1999, 2001) have included certain suitable prognostic factors in constructing the adaptive designs for the binary and normal responses.

Note that the construction of the adaptive designs in all of the above work was confined to the non-longitudinal (cross-sectional) set up. That is, once the treatment was assigned to an individual, the individual was expected to have only one response. In practice, there are however clinical trial experiments where it may be useful to register the study subjects sequentially and collect responses from each study subject sequentially over time. For example, in the asthma prevention study, once an individual enters into the study in a sequence of time, the individual may be examined once a week over a period of 4 weeks for the detection of 'asthma' status. Here the 'yes' or 'no' status of 'asthma' of an individual at a given week is a binary response. In this study, it is important to construct a longitudinal adaptive design by using the available repeated binary responses and covariate information such as age, chronic conditions and smoking habit, 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.

In this paper, we propose a simple longitudinal adaptive design such 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 the RPW rule in the non-longitudinal set up. The proposed design is described in section 2. In the same section, we also study the performance of the proposed design in allocating study subjects to a better treatment, through a simulation study. Furthermore, 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 section 3, following Sutradhar & Das (1999) (see also Jowaheer & Sutradhar 2002) we introduce a general autocorrelation structure for the repeated binary responses and take these correlations as well as the longitudinal adaptive design weights (to be discussed in section 2) into account for consistent and efficient estimation of the regression parameters of the model. More specifically, we use a weighted generalized quasi-likelihood (WGQL) approach for a consistent and efficient estimation. In section 4, we extend the simulation study conducted in section 2 to examine the performance of the proposed WGQL estimation approach as well as to study the misspecification effects of the longitudinal adaptive designs. It is found that the longitudinal design weights play an important role in consistent estimation. A simulation-based coverage probability for the treatment effect is also reported in the same section. Some concluding remarks are given in section 5.

2. Binary longitudinal model for clinical trial data

Let the ith ($i = 1, \ldots, K$) patient enter into the clinical trial at the time point i, giving T consecutive binary responses, namely $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$. Thus, the whole clinical trial will be completed at time point $K + T - 1$. Next suppose that $x_{it}^* = (\delta_i, x_{it}')'$ 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. For example, for the asthma prevention study mentioned in section 1, δ_i refers to the selection of the treatment for the ith patient and x_{it} denotes the prognostic factors such as age, chronic conditions and smoking habit of the ith patient, recorded at time point t ($t = 1, \ldots, T$). In all, there will be $N = KT$ binary longitudinal responses in the clinical trial. For the asthma prevention problem, these responses are the 'yes' or 'no' status of the asthma patients recorded over T repeated periods for each of the K patients. Note that as the ith patient enters the system at 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 described as the response of the ith patient at 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 system 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 continues with the selected treatment for the complete duration of T periods.

Let $\beta = (\beta_1, \beta_2, \ldots, \beta_u, \ldots, \beta_p)'$ denote the effect of the p-dimensional covariate vector $x_{it}^* = (\delta_i, x_{it}')'$, where β_1, in particular, represents the effect of the treatment, and $\beta_2, \ldots, \beta_u, \ldots, \beta_p$ denote the effect of the prognostic factors. It is of primary interest to estimate the β vector based on all covariate information and the responses available at the end of the study. Note that in the present set up, the treatment covariate δ_i ($i = 1, \ldots, K$) is chosen such that more study subjects are assigned to the better treatment. It is then clear that the consistent and efficient estimation of β, specially the estimation of treatment effect β_1, will depend on the selection probabilities of δ_i ($i = 1, \ldots, K$). Further, note that the efficient estimation of β will also depend on the longitudinal correlation structure of the repeated binary responses $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$ for all $i = 1, \ldots, K$. Following Sutradhar & Das (1999), a general longitudinal correlation structure suitable for the binary responses will be introduced in section 3 for the purpose of efficient estimation. The computation of the selection probabilities of δ_i ($i = 1, \ldots, K$) is discussed in the following section. To be specific, we introduce an appropriate longitudinal adaptive design for the computation of these probabilities in section 2.1. In section 2.2, we study the asymptotic as well as the small sample performances of the proposed longitudinal adaptive design.

2.1. Construction of the longitudinal adaptive design

To construct the longitudinal adaptive design one needs to derive the formulas for the selection probabilities of δ_i ($i = 1, \ldots, K$), where δ_i is the treatment indicator for the ith patient. Note that the selection probability for δ_i for the ith patient will be computed depending on the longitudinal outcomes of all $i - 1$ patients and their covariate information. For simplicity, the history of responses for the past $i - 1$ patients will be denoted by y_{Hi-1}. Further suppose that A is the better treatment between A and B, and

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

with

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

where w_i ($i = 1, \ldots, K$) is referred to as the design weight. Now to construct the required design we derive formulas for w_i ($i = 1, \ldots, K$) as follows.

Note that as at the start we have no reason to believe that any particular treatment is better than the other, for the first patient we choose $w_1 = 0.5$ and obtain δ_1 so that $P(\delta_1 = 1) = w_1$. Next, 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 vectors x_{rt}. 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 and the $(i-2)$th patient has yielded two responses and so on. Furthermore, as a patient is kept in the study for T repeated responses, it becomes convenient to write the formulas for w_i for two cases, first for the case when $2 \leq i \leq T$ and then for $i > T$. The formulas for w_i for these two cases will be computed using a simple longitudinal play-the-winner (SLPW) rule as a generalization of the existing randomized play-the-winner (RPW) rule (Wei & Durham, 1978). As in the RPW rule, the proposed SLPW rule will be illustrated as an urn design as described below.

As w_i is the probability of selection of the better treatment for the ith patient to be computed based on the history y_{Hi-1} [that is, y_{rt} and the prognostic factors x_{rt} $(r = 1, \ldots, i-1; 1 \leq t \leq \min(T, i-r))$], it is convenient to compute w_i by including two types of balls in the urn, the first type being the indicator for the selection of the better treatment A and the second type for the other treatment. The two types of balls are added to the urn as follows:

(a) The urn will have α balls of each type initially.
(b) For a suitable τ value and for available past responses y_{rt}, $y_{rt}\tau$ balls of the same kind by which the rth $(r = 1, \ldots, i-1)$ patient was treated and $(1 - y_{rt})\tau$ balls of the opposite kind are added, at the treatment selection stage for the ith patient.
(c) For a suitable quantity u_{rt} defined such that a larger value of u_{rt} implies the prognostic factor based on a less serious condition of the rth $(r = 1, \ldots, i-1)$ past patient, $G - u_{rt}$ balls of the same kind by which the rth patient was treated and u_{rt} balls of the opposite kind are added, at the treatment selection stage for the ith patient, where $[0, G]$ is the domain of u_{rt}.

As described in detail in the appendix, the above scheme produces the selection probabilities w_i $(i = 2, \ldots, K)$ for the cases $2 \leq i \leq T$ and $i > T$ as follows:

Case 1 $(2 \leq i \leq T)$. For this case

$$w_i = \Pr(\delta_i = 1 | y_{Hi-1}) = \frac{n^*_{i-1,A}(y_{Hi-1})}{n^*_{i-1}}, \tag{1}$$

where

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

is the total number of balls in the urn at the selection stage for the ith patient, and

$$n^*_{i-1,A}(y_{Hi-1}) = \alpha + \sum_{r=1}^{i-1}\sum_{t=1}^{i-r}[\delta_r\{(G - u_{rt}) + y_{rt}\tau\} + (1 - \delta_r)\{u_{rt} + (1 - y_{rt})\tau\}], \tag{3}$$

is the number of balls of the first type that supports the selection of the treatment A.

Case 2 $(i > T)$. For this case

$$w_i = \Pr(\delta_i = 1 | y_{Hi-1}) = \frac{\tilde{n}_{i-1,A}(y_{Hi-1})}{\tilde{n}_{i-1}}, \tag{4}$$

where

$$\tilde{n}_{i-1} = 2\alpha + \sum_{r=1}^{i-T}\sum_{t=1}^{T}(G+\tau) + \sum_{r=i-T+1}^{i-1}\sum_{t=1}^{i-r}(G+\tau) \tag{5}$$

and

$$\begin{aligned}\tilde{n}_{i-1,A}(y_{Hi-1}) = \alpha &+ \sum_{r=1}^{i-T}\sum_{t=1}^{T}[\delta_r\{(G-u_{rt})+y_{rt}\tau\} + (1-\delta_r)\{u_{rt}+(1-y_{rt})\tau\}] \\ &+ \sum_{r=i-T+1}^{i-1}\sum_{t=1}^{i-r}[\delta_r\{(G-u_{rt})+y_{rt}\tau\} + (1-\delta_r)\{u_{rt}+(1-y_{rt})\tau\}],\end{aligned} \tag{6}$$

are similar to those of n^*_{i-1} in (2) and $n^*_{i-1,A}(y_{Hi-1})$ in (3), respectively.

2.2. Performance of the proposed adaptive design

2.2.1. Limiting behaviour of design weights w_i

Note that it follows from (4) that $w_{i+1}/w_i \to 1$ as $i \to \infty$. 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 subsequence $w_{k(i)}$ that is convergent. Suppose that it converges to ω. Then from the above limiting result, we have

$$\frac{w_{k(i)+1}}{w_{k(i)}} \to 1$$

as $i \to \infty$, implying that for some $\epsilon > 0$,

$$\omega(1-\epsilon) \leq \text{liminf } w_{k(i)+1} \leq \text{limsup } w_{k(i)+1} \leq \omega(1+\epsilon),$$

and hence

$$\text{limsup } w_{k(i)+1} - \text{liminf } w_{k(i)+1} \leq 2\omega\epsilon.$$

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

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

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

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

$$(4)\ \frac{1}{iT}\sum_{r=1}^{i-T}\sum_{j=1}^{T}p_{rj1}w_r \to \pi_1\omega^*, \quad (5)\ \frac{1}{iT}\sum_{r=1}^{i-T}\sum_{j=1}^{T}p_{rj2}w_r \to \pi_2\omega^*, \quad (6)\ \frac{1}{iT}\sum_{r=1}^{i-T}\sum_{j=1}^{T}u_{rj}w_r \to u^*\omega^*.$$

Using the above limiting results from (1) to (6) in equation (4), one obtains

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

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 the allocation of 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. Note that $\omega^* > 0.5$ indicates that more study subjects will be assigned to the better treatment A.

2.2.2. Allocation performance of the proposed design: a simulation study

In the last section, we 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 $K = 100$ and 200, where K is the total number of patients involved in the clinical trial. The performance of the proposed design will be examined through a simulation study. For this, we first provide a simulation design as follows.

2.2.2.1. Simulation design. We choose $T = 4$, where T denotes the number of repeated responses collected from each of the K individuals. We choose $p = 4$ covariates, namely one treatment covariate and three prognostic factors. As before, the treatment covariate is denoted by δ_i and the other three 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 weights

$$w_i = \Pr(\delta_i = 1 | y_{H_{i-1}}),$$

constructed in section 2.1. The three prognostic factors are however chosen as follows. We consider the 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 $p = 0.5$. 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 rating for chronic disease has $x_{it2} = 0$ for all $t = 1, \ldots, T$. If the ith patient however enters the trial with a high rating for chronic disease, then $x_{it2} = 1$ for all $t = 1, \ldots, T$.

To generate the other two prognostic factors, namely, x_{it3} and x_{it4}, we now consider the age variable and generate an age between 20 and 80 from a uniform distribution. Next we create six age groups, namely 21–30, 31–40, ..., 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 comes within 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 three age groups, namely, 21–40, 41–60, and 61–80, which are 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

Table 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 values of $K =$* 100, 200

K	τ	ρ	Mean	Standard Error
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
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

It is clear from Table 1 that, irrespective of correlation values, the proposed design allocated more individuals to the better treatment A. For example, for $K = 100$, $\tau = 2.0$ and $\rho = 0.9$, 59 individuals of 100 were assigned to treatment A. Thus relatively more 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; this is about 59 per cent. Note that the allocation improves for larger τ. For example, for the same $K = 200$ and $\rho = 0.9$, the number of individuals allocated 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.

2.2.3. Expected design weights under binary models

Recall that the adaptive design weights w_i are given by (1) for $2 \leq i \leq T$ and by (4) for $i > T$, T being small in the present longitudinal set up. By (4) these design weights satisfy the limits shown in section 2.2.1. Further, it was shown in section 2.2.1 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 the estimation of β, the effects of covariates. For this, for all $i = 1, \ldots, K$ (with $K = 100$ or 200) we will compare the w_i computed based on the sample binary responses as in the last section with its expected value w_{i0}, where w_{i0} is computed as

$$w_{i0} = E_{\delta_1} E_{\delta_2|\delta_1} \cdots E_{\delta_i|\delta_1,\ldots,\delta_{i-1}}(\delta_i). \tag{7}$$

As $E_{\delta_i|\delta_1,\ldots,\delta_{i-1}}(\delta_i) = \Pr(\delta_i = 1|\delta_{i-1}, \ldots, \delta_1) = w_i$, where w_is are defined in (1) and (4), it follows that for $r = 1, \ldots, i-1$,

$$E(\delta_r Y_{rt}) = E_{\delta_1} E_{\delta_2|\delta_1} \cdots E_{\delta_r|\delta_1,\ldots,\delta_{r-1}} E(\delta_r Y_{rt}|\delta_r,\ldots,\delta_1)$$
$$= E_{\delta_1} E_{\delta_2|\delta_1} \cdots E_{\delta_r|\delta_1,\ldots,\delta_{r-1}} (\delta_r p^*_{rt}), \tag{8}$$

where $p^*_{rt} = E(Y_{rt}|\delta_r,\ldots,\delta_1) = \exp(x^{*'}_{rt}\beta)/(1+\exp(x^{*'}_{rt}\beta))$ 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 (8) then reduces to

$$E(\delta_r Y_{rt}) = w_{r0} p_{rt1}, \tag{9}$$

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

$$E(1-\delta_r)(1-Y_{rt}) = (1-w_{r0})(1-p_{rt2}), \tag{10}$$

where $p_{rt2} = \exp(z^{*'}_{rt}\beta)/(1+\exp(z^{*'}_{rt}\beta))$. Now by applying (9) and (10) to (7), it follows from (1) that for $2 \le i \le T$, the unconditional expectation of w_i is given as

$$w_{i0} = \frac{\left[\alpha + \sum_{r=1}^{i-1}\sum_{t=1}^{i-r}[\{(G-u_{rt})+p_{rt1}\tau\}w_{r0} + \{u_{rt}+(1-p_{rt2})\tau\}(1-w_{r0})]\right]}{[2\alpha+(1/2)i(i-1)(G+\tau)]}. \tag{11}$$

Similarly, it follows from (4) that for $i > T$, the unconditional expectation of w_i is given by

$$w_{i0} = \left\{2\alpha+(G+\tau)T\left(\frac{i-(T+1)}{2}\right)\right\}^{-1}\left[\alpha + \sum_{r=1}^{i-T}\sum_{t=1}^{T}\{(G-u_{rt}+p_{rt1}\tau)w_{r0}\right.$$
$$+ (u_{rt}+(1-p_{rt2})\tau)(1-w_{r0})\}$$
$$\left. + \sum_{r=i-T+1}^{i-1}\sum_{t=1}^{i-r}\{((G-u_{rt})+p_{rt1}\tau)w_{r0} + (u_{rt}+(1-p_{rt2})\tau)(1-w_{r0})\}\right]. \tag{12}$$

Note that for $i = 1,\ldots,K$, w_{i0} in (11) and (12) are the unconditional expectation of w_i under the present binary model. Further note that although β is unknown, it remains the same throughout the experiment. In the next section, we will consider the estimation of this unknown parameter β. In this section, we 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 a manner similar to that in section 2.2.2. Here, the simulations of the binary responses depend on the β and ρ parameters of the correlated binary model. Unlike w_i, w_{i0} is 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 section 2.2.2, we now compute the w_i and w_{i0} values for all $i = 1,\ldots,K$, with $K = 100$ and 200. The graphs for w_i and w_{i0} are shown in Fig. 1 for $K = 100$. Note that although the graphs for w_i and w_{i0} are plotted for $K = 200$, they are not included here in order to save space. In Fig. 1, we show the graphs for $\tau = 2.0$ and for two values of $\rho = 0.5$ and 0.9.

Note 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 based on x^*_{rt}. For convenience of numerical computations, as in section 2.2.2.1, we generated the prognostic factors x_{rt2}, x_{rt3} and x_{rt4} following certain suitable probability models. This leads to two different sets of prognostic factors as well as u_{rt} functions for the two choices $K = 100$ and 200. Consequently, for given values of β, τ, α and G, Figs 1 and 2 exhibit two similar but slightly different graphs for w_{i0} for $K = 100$ and 200. As opposed to w_{i0}, the value of w_i changes depending on the past binary responses y_{rt} ($r = 1,\ldots,i-1$; $1 \le t \le \min(T, i-r)$) which are likely to be different under different simulations, also they are different because they

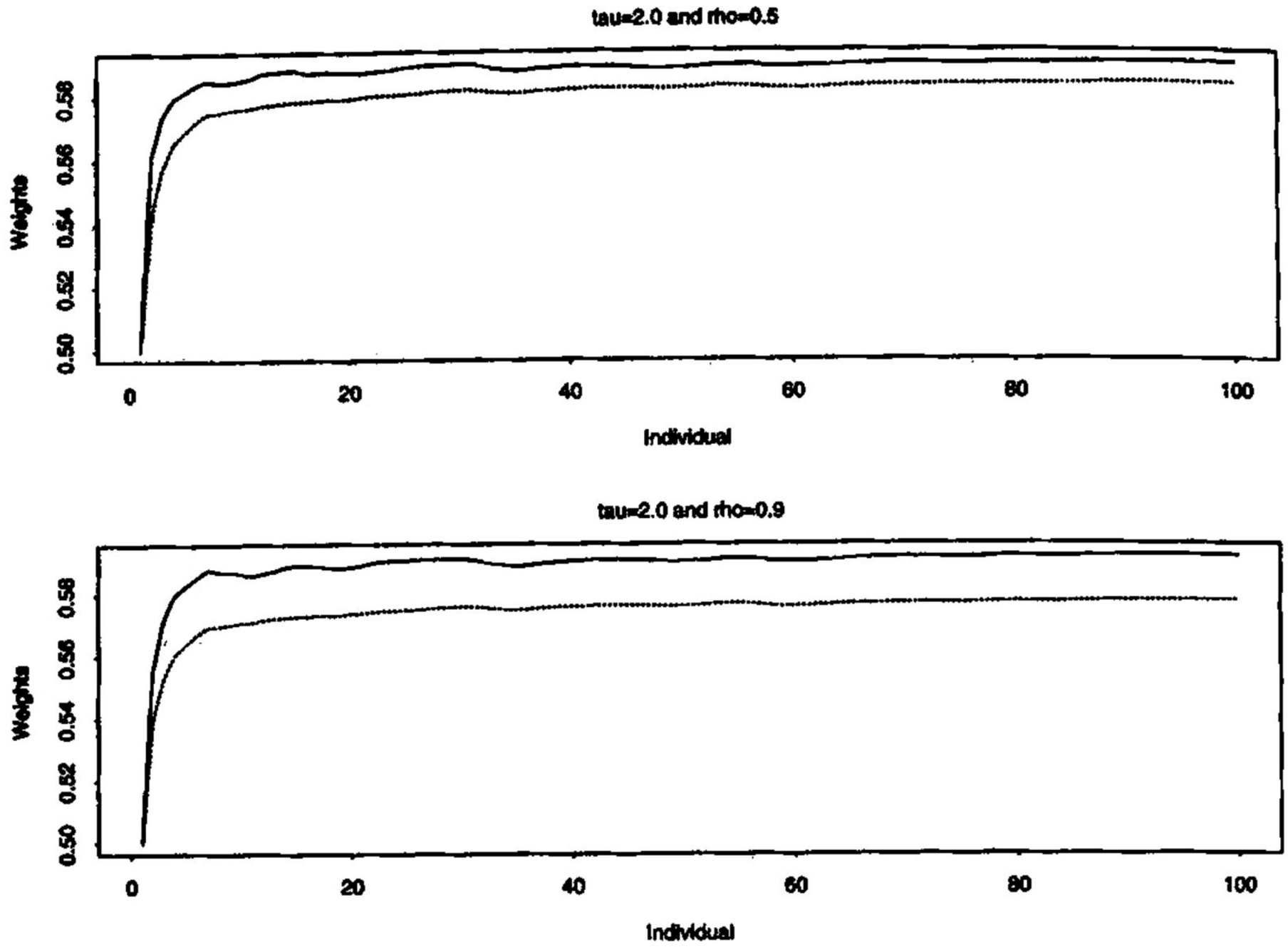

Fig. 1. Adaptive design weights (w_i) and expected weights (w_{i0}) for $K = 100$: w_i: ——; w_{i0}: ....., for τ(tau) = 2.0 and selected values of ρ(rho).

are generated with different values of longitudinal correlations, namely $\rho = 0.9$ and 0.5. For a given i, the averages of w_i over 1000 simulations are shown in Fig. 1 for $K = 100$. 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 was found to be quite satisfactory for large i, specially for the values of i closed to large K, such as for $100 \leq i \leq K$, where $K = 200$, although we have not shown these graphs here. Note that this convergence occurs irrespective of the choices of the values of τ and ρ, although the convergence is faster for larger $\tau = 4.0$ and smaller $\rho = 0.5$ as compared with smaller $\tau = 2.0$ and larger $\rho = 0.9$. In order to save space, we have also not shown the graphs for the cases with $\tau = 4.0$.

3. WGQL approaches for parameter estimation including the treatment effect

Recall that in section 2, we proposed an adaptive longitudinal design that assigns the ith ($i = 1, \ldots, K$) individual to the treatment A (between A and B) with probability w_i given by (1) for $2 \leq i \leq T$ and by (4) for $i > T$. By taking A as the better of the treatments A and B, we have also examined the performance of the proposed design by a simulation study and it was 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 in knowing the regression parameter vector β which invisibly contributes to generate binary responses $y_{rt}(r = 1, \ldots, i-1; 1 \leq t \leq \min(T, i-r))$ which is 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 and efficiently. 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 & Zeger (1986) have bypassed 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 & Das (1999). Further, additional problems mount up if the individuals enter a study in a sequence and one or more covariates, such as the treatment, for the incoming individuals are determined based on the outcomes of past individuals.

In this paper, unlike Liang & Zeger (1986), we model the correlations of the repeated binary responses following Sutradhar & Das (1999). This prompts the use of the true mean and covariance structure-based generalized quasi-likelihood (GQL) approach for the estimation of β in a non-adaptive longitudinal set up. As in the present adaptive longitudinal set up, it is important to take the design weights into account; we incorporate these design weights in the GQL approach and refer to this modified approach as the WGQL approach. This WGQL approach is explained in the following subsection.

3.1. 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 the ith individual is assigned to treatment A with probability $w_i = \Pr(\delta i = 1 | y_H)$ given by (1) for $2 \leq i \leq T$ and by (4) for $i > T$. Here, y_{it} is the tth response of the ith individual. Further note that as w_i depends on the responses from the past $i - 1$ patients, the unconditional expectation of y_{it} may be computed as

$$E(Y_{it}) = E_{\delta_1} E_{\delta_2|\delta_1} \cdots E_{\delta_i|\delta_1,\ldots,\delta_{i-1}} E(Y_{it}|\delta_i, \delta_{i-1}, \ldots, \delta_1). \tag{13}$$

It then follows from (7) to (10) that

$$E(Y_{it}) = w_{i0} p_{it1} + (1 - w_{i0}) p_{it2} = p_{it}, \tag{14}$$

where w_{i0} is given by (1) for $2 \leq i \leq T$ and by (4) for $i > T$, and p_{it1} and p_{it2} are given as

$$p_{it1} = \frac{\exp(z_{it}'\beta)}{[1 + \exp(z_{it}'\beta)]} \quad \text{and} \quad p_{it2} = \frac{\exp(z_{it}^{*\prime}\beta)}{[1 + \exp(z_{it}^{*\prime}\beta)]}, \tag{15}$$

respectively, with $z_{it}' = (\delta_i, x_{it2}, \ldots, x_{itp})|_{\delta_i = 1}$ and $z_{it}^{*\prime} = (\delta_i, x_{it2}, \ldots, x_{itp})|_{\delta_i = 0}$. Let $p_i = (p_{i1}, \ldots, p_{it}, \ldots, p_{iT})'$ where p_{it} is given by (14) for all $i = 1, \ldots, K$, so that

$$E(Y_i) = E(Y_{i1}, \ldots, Y_{iT})' = p_i. \tag{16}$$

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

$$\text{corr}(Y_{it}, Y_{iv}|\delta_i, \ldots, \delta_1) = \rho_{|(t-v)|}, \tag{17}$$

where $\rho_{|t-v|}$ denotes the lag $|t - v|$ autocorrelation. Note that the autocorrelation structure considered in (17) is general as it accommodates the Gaussian AR(1), MA(1) and exchangeable type autocorrelation structures as special cases. It then follows that the unconditional covariance between Y_{it} and Y_{iv} is given by

$$\begin{aligned} \text{cov}(Y_{it}, Y_{iv}) &= E_{\delta_1} E_{\delta_2|\delta_1} \cdots E_{\delta_i|\delta_1,\ldots,\delta_{i-1}} [\text{cov}(Y_{it}, Y_{iv})|\delta_i, \delta_{i-1}, \ldots, \delta_1)] \\ &\quad + cov_{\delta_i,\ldots,\delta_1} [E(Y_{it}|\delta_i, \delta_{i-1}, \ldots, \delta_1), E(Y_{iv}|\delta_i, \delta_{i-1}, \ldots, \delta_1)] \\ &= E_{\delta_1} E_{\delta_2|\delta_1} \cdots E_{\delta_i|\delta_1,\ldots,\delta_{i-1}} [\rho_{|t-v|} \{p_{it}^* q_{it}^* p_{iv}^* q_{iv}^*\}^{1/2}] + \text{cov}_{\delta_i,\ldots,\delta_1} [p_{it}^*, p_{iv}^*], \end{aligned} \tag{18}$$

where $E(Y_{it}|\delta_i, \ldots, \delta_1) = p_{it}^* = \exp(x_{it}^{*\prime}\beta)/(1 + \exp(x_{it}^{*\prime}\beta))$ and $\text{var}(Y_{it}|\delta_i, \ldots, \delta_1) = p_{it}^* q_{it}^*$ by (8) with $q_{it}^* = 1 - p_{it}^*$. After some algebra, equation (18) reduces to

$$\operatorname{cov}(Y_{it}, Y_{iv}) = \rho_{|t-v|}\left[w_{i0}\{p_{it1}q_{it1}p_{iv1}q_{iv1}\}^{1/2} + (1-w_{i0})\{p_{it2}q_{it2}p_{iv2}q_{iv2}\}^{1/2}\right]$$
$$+ w_{i0}\{p_{it1}p_{iv1}\} + (1-w_{i0})\{p_{it2}p_{iv2}\} - p_{it}p_{iv} = \sigma_{itv}. \tag{19}$$

When $t = v$, the covariance σ_{itv} in (19) reduces to the variance of y_{it} given by

$$\operatorname{var}(Y_{it}) = \sigma_{itt} = p_{it}q_{it}, \tag{20}$$

where $q_{it} = 1 - p_{it}$ with p_{it} as defined in (14). Let Σ_i denote the covariance matrix of y_i, which may be expressed as

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

for $t, v = 1, \ldots, T$, where σ_{itv} are given by (19) and (20).

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

$$\sum_{i=1}^{K}\left(\frac{\partial p_i'}{\partial \beta}\right)\Sigma_i^{-1}(y_i - p_i) = 0, \tag{21}$$

(Wedderburn, 1979; McCullagh, 1983), where p_i is the $T \times 1$ vector given by (16) and $\partial p_i'/\partial\beta$ 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 is a function of w_{i0}, β and ρ, where w_{i0} again depends on β. In addition, p_i vector is a function of w_{i0} which contains β. Now in solving (21) for β, in the spirit of the GEE approach (see Liang & Zeger, 1986) we re-express the QL estimating equation (21) as

$$\sum_{i=1}^{K}\left(\frac{\partial p_i'(w_{i0})}{\partial \beta}\right)\Sigma_i^{-1}(w_{i0}, \hat{\rho})(y_i - p_i(w_{i0})) = 0, \tag{22}$$

and we refer to this as the WGQL estimating equation for β, where $\hat{\rho}$ is a consistent estimator for the longitudinal correlation parameter ρ. Note that for the case when longitudinal data are subject to non-response, Robins *et al.* (1995) (see also Robins & Rotnitzky, 1995) have modified the 'working' correlation-based GEE of Liang & Zeger (1986), which they have referred to as the WGEE approach. The WGEE used in Robins *et al.* (1995) is, however, similar but quite different from our WGQL estimating equation (22). This is because the design weights in the present set up are quite different from the weights used to represent missing values. Furthermore, in (22) we use the true covariance structure, whereas Robins *et al.* (1995) used the 'working' covariance structure following Liang & Zeger (1986).

Now to solve (22) for β, one may consider the following three scenarios: first, for some initial β, w_{i0} is known in the spirit of GEE; secondly, w_{i0} is unknown but it can be replaced by the adaptive design weight w_i as $E(w_i) = w_{i0}$; thirdly, w_{i0} is an unknown function of β. The estimator of β as the solution of (22) under the above three scenarios will be denoted by $\hat{\beta}_{\mathrm{WGQL1}}$, $\hat{\beta}_{\mathrm{WGQL2}}$ and $\hat{\beta}_{\mathrm{WGQL3}}$ respectively. These solutions may be obtained by using iterative equations

$$\hat{\beta}_{(m+1)_{GQL}} = \hat{\beta}_{(m)_{GQL}} + \left[\sum_{i=1}^{K}\left(\frac{\partial p_i'(w_{i0})}{\partial \beta}\right)\Sigma_i^{-1}(w_{i0}, \hat{\rho})\left(\frac{\partial p_i(w_{i0})}{\partial \beta'}\right)\right]_m^{-1}$$
$$\times \left[\sum_{i=1}^{K}\left(\frac{\partial p_i'(w_{i0})}{\partial \beta}\right)\Sigma_i^{-1}(w_{i0}, \hat{\rho})(y_i - p_i(w_{i0}))\right]_m, \tag{23}$$

where $\hat{\beta}_{(m)_{GQL}}$ is the value of β at the mth iteration and $[\cdot]_m$ denotes that the expression within the brackets is evaluated at $\hat{\beta}_{(m)_{GQL}}$. Note that to compute $\hat{\beta}_{\mathrm{WGQL1}}$ and $\hat{\beta}_{\mathrm{WGQL2}}$, the first derivative vector $\partial p_i'(w_{i0})/\partial\beta$ has the formulas

$$\frac{\partial p_i'(w_{i0})}{\partial \beta} = w_{i0}\left(\frac{\partial p_{i1}'}{\partial \beta}\right) + (1 - w_{i0})\left(\frac{\partial p_{i2}'}{\partial \beta}\right) = w_{i0} Z_i' A_{i1} + (1 - w_{i0}) Z_i^{*'} A_{i2} = B_i, \tag{24}$$

and

$$\frac{\partial p_i'(w_{i0})}{\partial \beta} = \frac{\partial p_i'(w_{i0})}{\partial \beta}\Big|_{w_{i0}=w_i} = w_i Z_i' A_{i1} + (1 - w_i) Z_i^{*'} A_{i2} = C_i, \tag{25}$$

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 (24) and (25), $p_{i1} = (p_{i11}, \ldots, p_{it1}, \ldots, p_{iT1})'$ and $p_{i2} = (p_{i12}, \ldots, p_{it2}, \ldots, p_{iT2})'$, with $p_{it1} = \exp(z_{it}'\beta)/(1 + \exp(z_{it}'\beta))$ and $p_{it2} = \exp(z_{it}^{*'}\beta)/(1 + \exp(z_{it}^{*'}\beta))$.

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

$$\frac{\partial p_i'(w_{i0})}{\partial \beta} = w_{i0}\left(\frac{\partial p_{i1}'}{\partial \beta}\right) + (1 - w_{i0})\left(\frac{\partial p_{i2}'}{\partial \beta}\right) + \left(\frac{\partial w_{i0}}{\partial \beta}\right)(p_{i1} - p_{i2})' = D_i, \tag{26}$$

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

$$\frac{\partial w_{i0}}{\partial \beta} = \frac{\left[\sum_{r=1}^{i-1}\sum_{t=1}^{i-r}\{(p_{rt1}q_{rt1}z_{rt}\tau w_{r0}) - (p_{rt2}q_{rt2}z_{rt}^*\tau(1 - w_{r0}))\}\right]}{[2\alpha + (1/2)i(i-1)(G+\tau)]}, \tag{27}$$

and for $i > T$

$$\begin{aligned}\frac{\partial w_{i0}}{\partial \beta} = &\{2\alpha + (G+\tau)T(i - (T+1)/2)\}^{-1}\left[\sum_{r=1}^{i-T}\sum_{t=1}^{T}\{(p_{rt1}q_{rt1}z_{rt}\tau w_{r0}) - (p_{rt2}q_{rt2}z_{rt}^*\tau(1 - w_{r0}))\}\right.\\ &\left. + \sum_{r=i-T+1}^{i-1}\sum_{t=1}^{i-r}\{(p_{rt1}q_{rt1}z_{rt}\tau w_{r0}) - (p_{rt2}q_{rt2}z_{rt}^*\tau(1 - w_{r0}))\}\right].\end{aligned} \tag{28}$$

This completes the construction of the estimating equation given in (22) under the above three scenarios.

Note that the estimating equations for β require the knowledge of $\hat{\rho} = (\hat{\rho}_1, \ldots, \hat{\rho}_l, \ldots, \hat{\rho}_{T-1})$ where $\hat{\rho}_l (l = 1, \ldots, T-1)$ may be obtained consistently as in section 3.2 by using the so-called method of moments. Next, under some regularity conditions, it may be shown (Liang & Zeger, 1986) that for large K, $\hat{\beta}_{\text{WGQL1}}$ and $\hat{\beta}_{\text{WGQL3}}$ have asymptotically p-dimensional normal distribution with mean β and covariance matrices given as

$$\text{var}(\hat{\beta}_{WGQL1}) = \left[\sum_{i=1}^{K}\left(\frac{\partial p_i'(w_{i0})}{\partial \beta}\right)\Sigma_i^{-1}(w_{i0}, \hat{\rho})\left(\frac{\partial p_i(w_{i0})}{\partial \beta'}\right)\right]^{-1} = \left[\sum_{i=1}^{K} B_i \Sigma_i^{-1}(w_{i0}, \hat{\rho}) B_i'\right]^{-1}, \tag{29}$$

and

$$\text{var}(\hat{\beta}_{\text{WGQL3}}) = \left[\sum_{i=1}^{K} D_i \Sigma_i^{-1}(w_{i0}, \hat{\rho}) D_i'\right]^{-1}, \tag{30}$$

respectively, where B_i and D_i are given in (24) and (26), respectively. By similar arguments, one can show that $\hat{\beta}_{\text{WGQL2}}$ also has an asymptotically normal distribution with β mean vector and a suitable covariance matrix that can be consistently estimated by

$$\text{var}(\hat{\beta}_{\text{WGQL2}}) = \left[\sum_{i=1}^{K} C_i \Sigma_i^{-1}(w_i, \hat{\rho}) C_i'\right]^{-1}, \tag{31}$$

where $\Sigma_i(w_i)$, for example, is obtained from $\Sigma_i(w_{i0})$ by replacing w_{i0} by its data-based estimate w_i, and C_i is given by (25).