

A Correlated Binary Model for Ignorable Missing Data: Application to Rheumatoid Arthritis Clinical Data

Francis Erebholo^{1*}, Victor Apprey³, Paul Bezandry², John Kwagyan^{3,4}

¹ *Department of Mathematics, Hampton University*

² *Department of Mathematics, Howard University*

³ *National Human Genome Center, Howard University*

⁴ *Department of Community and Family Medicine, Howard University*

Abstract: Incomplete data are common phenomenon in research that adopts the longitudinal design approach. If incomplete observations are present in the longitudinal data structure, ignoring it could lead to bias in statistical inference and interpretation. We adopt the disposition model and extend it to the analysis of longitudinal binary outcomes in the presence of monotone incomplete data. The response variable is modeled using a conditional logistic regression model. The nonresponse mechanism is assumed ignorable and developed as a combination of Markov's transition and logistic regression model. MLE method is used for parameter estimation. Application of our approach to rheumatoid arthritis clinical trials is presented.

Key words: Binary data; disposition model; dropout mechanism; ignorable missingness; maximum likelihood estimation.

1. Introduction

An Correlated data are very common in clinical and social science research and include nested and clustered data. Data are correlated because of common attributes that are shared among the members of the group, or among several measures of a member over time. Longitudinal and repeated data are specific cases of correlated data. Longitudinal data refer to data that are collected by repeatedly observing the same subject over a period of time. Repeated measures, which include longitudinal data, refer to data on subjects measured repeatedly over a period of time, or under different conditions, or both. In epidemiology, a study on members of the same family will be correlated on different covariate than non-family members. This means that the probability that a family member has an outcome is not necessarily the same as that of an individual randomly selected from the population.

Likelihood based approaches for analyzing correlated binary data are limited. Bonney (1998) introduced the term disposition to represent the conditional probability of the outcome of one

* Corresponding author.

member of a cluster given another member has the attribute. The development of the disposition model starts with random effects formulation and then introduces a theory for constructing likelihoods utilizing moment series representations. The disposition model was further investigated by Kwagyan (2001) through an alternative formulation from a finite mixture modeling perspective.

The attractive feature of reproducibility of the disposition model (Bonney, 1998, 2003) makes it desirable to naturally extend it to capture the types of correlation or dependence that arises in longitudinal data.

The problem of incomplete responses is a common occurrence in longitudinal data. This happens when one or more of the subject measurement from a unit is not taking, lost or otherwise unavailable. Missingness could be related to the outcome of interest. When it is unrelated to the outcome of interest, the effect is weak and analysis of the parameters of interest is less complicated. However, when it is related to the outcome of interest, the impact of the missing data is great, and the analysis, which is complicated, should be carried out with care to avoid a potential bias of inference on the parameters of interest. This in particular is the case when individuals with missing data differ significantly in important ways from those with complete data structure (Molenberghs et. al., 2015).

When a missing data is related to the history of the observed responses, it is known as missing at random (MAR), when it is related to the current unobserved response, it is known as missing not at random (MNAR) (Little and Rubin, 2002). When the missingness is MAR, estimates will be valid and fully efficient when the likelihood and missing data model is correctly specified (Molenberghs et. al., 2002, Diggle and Kenward, 1994). However, when the missingness is MNAR, statisticians are faced with difficulties when the parameters of interest are to be estimated. The pattern of missingness could be monotone and non-monotone. If an individual, or a subject miss an appointment for an observation and he or she is never observed again, the pattern of missingness is said to be monotone otherwise it is non-monotone.

Since missing data are common and still a challenging problem in longitudinal study design, several approaches for dealing with them has been suggested. Some of these methods include the Likelihood-based and Bayesian method, Multiple Imputation and the Weighted Equations. For a study with missing data, the validness, soundness and efficacy of any method of analysis will require the tenability of some assumptions regarding the reasons the missing values occurred to avoid a bias and misleading inference about the parameters of interest (Molenberghs et. al., 2015).

In Section 2, we introduced the joint distribution of the incomplete data by combining the model of disposition and the dropout model and present the corresponding likelihood function. In Section 3, we present and discuss the result of the application of our approach to the rheumatoid arthritis clinical trial. Conclusions are presented in Section 4.

2. Joint Distribution for Incomplete Data

In this section, we introduce the disposition model and adopt it to develop a model in the presence of incomplete data. A joint distribution for the incomplete data will be constructed and

models for different dropout mechanism will be developed. Consider a sample of N clusters, each of size $n_i, i = 1, \dots, N$ and $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$ denote the vector of binary outcomes for the i^{th} cluster. Let δ_{ik} denote the conditional probability of $Y_{ik} = 1$ given that $Y_{ik'} = 1$ that is,

$$\delta_{ik} = Pr(Y_{ik} = 1 | Y_{ik'} = 1), k \neq k'; k, k' = 1, 2, \dots, n$$

Let us further assume that a pair of observed response within the same group satisfies the following relation

$$\frac{Pr(Y_{ik} = 1 | Y_{ik'} = 1)}{Pr(Y_{ik} = 1)Pr(Y_{ik'} = 1)} = \frac{1}{\alpha_i}, \alpha_i > 0, k \neq k'; k, k' = 1, 2, \dots, n$$

where α_i , called the relative disposition is common for all pairs of observation and it measures the within-group aggregation (correlation): $\alpha_i = 1$ implies independence or no aggregation. With this, Bonney (1998, 2003) has shown that the joint distribution of the i^{th} cluster is given as

$$P(Y_{i1}, \dots, Y_{in_i}) = (1 - \alpha_i) \prod_{k=1}^{n_i} (1 - y_{ik}) + \alpha_i \prod_{k=1}^{n_i} \delta_{ik} (1 - \delta_{ik})^{(1-y_{ik})} \quad (1)$$

Let us temporarily drop the subscript i representing the i^{th} unit when discussing a model of a cluster for simplicity of notation and let $\mathbf{Y}^* = (Y_1^*, \dots, Y_n^*)$ denote the complete vector of intended sequence of measurement on an experimental unit, and $\mathbf{t} = (t_1, \dots, t_n)$ the set of times that corresponds to the intended measurement. The joint probability distribution of \mathbf{Y}^* is

$$P(\mathbf{Y}^*; \alpha, \delta) = (1 - \alpha) \prod_{k=1}^n (1 - y_k^*) + \alpha \prod_{k=1}^n \delta_k (1 - \delta_k)^{(1-y_k^*)}$$

Let $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ denote the vector of complete observed sequences of binary observation for the i^{th} unit.

The assumption for the dropout process is that if an experimental unit is still in the study at time t_k ($2 \leq k \leq n$), the sequence of measurement $Y_j; j = 1, 2, \dots, k$ associated with it follows the same joint distribution as that of the corresponding intended sequence ($Y_j^*; j = 1, 2, \dots, k$).

We define the preceding outcome Y_j as:

$$Y_j = \begin{cases} 2Y_j^* - 1; & \text{for } j = 1, \dots, (D - 1) (Y_j^* \text{ is observed}) \\ 0 & \text{for } j \geq D, (Y_j^* \text{ is missing}) \end{cases},$$

where D is a random variable such that $2 \leq D \leq n$ denotes the dropout time, and $D = n + 1$ indicates no dropout.

For each k , let $\mathbf{H}_k = (Y_1, \dots, Y_{k-1})$ denote the observed history up to time t_{k-1} , and y_k^* , the value that would have been observed at time t_k , if there was no dropout in the unit. Similar to Diggle and Kenward (1994) selection model with non-ignorable dropout, we assume that the probability of dropout at time d depends on the history of the measurement process up to, and including the time of dropout t_d .

That is,

$$Pr(D = d | \text{History}) = p_d(\mathbf{H}_d, y_d^*; \boldsymbol{\phi}),$$

where $\boldsymbol{\phi} = (\phi_0, \dots, \phi_{2+p})$ is a vector of unknown parameters. With this, the following patterns of dropout process are identified: *Dropout Completely At Random (DCAR)*. Dropout is completely at random if the dropout process is independent of \mathbf{H}_d , and y_d^* . That is,

$$Pr(D = d | \text{History}) = p_d(d; \boldsymbol{\phi})$$

Dropout At Random (DAR). Dropout is at random if the dropout process depends on \mathbf{H}_d , and not y_d^* . That is,

$$Pr(D = d | \text{History}) = p_d(\mathbf{H}_d; \boldsymbol{\phi})$$

Informative or Dropout Not At Random (DNAR). This is when the dropout process depends on y_d^* .

We adopt the regressive logistic models of Bonney (1986, 1987, 1998) to model the dropout process and define the logit as,

$$\begin{aligned} \theta_k = \text{logit}[p_k(\mathbf{H}_k, \mathbf{y}; \boldsymbol{\phi})] &= \phi_0 + \phi_1 y_k + \sum_{j=2}^k \phi_j y_{k+1-j} \\ &+ \phi_{k+1} X_{k1} + \dots + \phi_{k+p} X_{kp} \end{aligned} \tag{2}$$

where $\mathbf{X} = (X_{k1}, \dots, X_{kp})^T$ is the p individual-specific covariates.

Our motivation of this choice is that the probability of dropout at time t_d is a direct consequence of the past outcomes, the present outcome, and possible set of covariates. Interested readers should see Bonney (1986, 1987) for full-parameterized forms with the specification of the dependence structure of the model.

Following Diggle and Kenward (1994), the joint distribution for an incomplete sequence with dropout at the t_d^{th} time point is:

$$P(\mathbf{Y}) = P^*(y_1, \dots, y_{d-1}) \left[\prod_{k=2}^{d-1} 1 - p_k(\mathbf{H}_k, y_k) \right] Pr(Y_d = 0 | \mathbf{H}_d, Y_{d-1} \neq 0) \tag{3}$$

Thus, the full log-likelihood for the i^{th} cluster for $\boldsymbol{\Theta}$ based on the data $(\mathbf{y}_i; i = 1, \dots, N)$ is given as

$$\ell(\boldsymbol{\Theta}) = \sum_{i=1}^N \log \left\{ P^*(y_1, \dots, y_{d-1}; \alpha, \delta) \left[\prod_{k=2}^{d-1} 1 - p_k(\mathbf{H}_k, y_k; \boldsymbol{\phi}) \right] Pr(Y_d = 0 | \mathbf{H}_d, Y_{d-1} \neq 0; \alpha, \delta, \boldsymbol{\phi}) \right\}$$

which is partitioned as:

$$\ell(\boldsymbol{\Theta}) = \ell_1(\alpha, \delta) + \ell_2(\boldsymbol{\phi}) + \ell_3(\alpha, \delta, \boldsymbol{\phi}) \tag{4}$$

where

$$\ell_1(\alpha, \delta) = \sum_{i=1}^N \log \left\{ (1 - \alpha_i) \prod_{k=1}^{d_i-1} (1 - y_{ik}) + \alpha \prod_{k=1}^{d_i-1} \delta_{ik} (1 - \delta_{ik})^{(1-y_{ik})} \right\}$$

is the log-likelihood for the observed response,

$$\ell_2(\boldsymbol{\phi}) = \sum_{i=1}^N \sum_{k=1}^{d_i-1} \log[1 - p_k(\mathbf{H}_k, y_k; \boldsymbol{\phi})]$$

and

$$\ell_3(\alpha, \delta, \boldsymbol{\phi}) = \sum_{i \leq N; d_i \leq n} \log\{\Pr(D = d_i | \mathbf{y}_i)\}$$

together corresponds to the log-likelihood function for the dropout process,

$$\Pr(D = d_i | \mathbf{y}_i) = \begin{cases} \sum_y p_d(\mathbf{H}_{d_i}, \mathbf{y}; \boldsymbol{\phi}) P_d^*(\mathbf{y} | \mathbf{H}_{d_i}, \alpha, \delta) & \text{for } d_i < n \\ 1 & \text{for } d_i = n + 1 \end{cases} \quad (5)$$

and

$P_k^*(\mathbf{y} | \mathbf{H}_k; \alpha, \delta)$ denote the conditional probability distribution function of \mathbf{Y}_k^* given \mathbf{H}_k

For non-ignorable dropout $\ell_3(\alpha, \delta, \boldsymbol{\phi})$ contains information on $(\alpha, \delta, \boldsymbol{\phi})$ as such, cannot be ignored.

By applying the DAR (ignorability) condition to Eqn. (4), the reduced log-likelihood function is

$$\ell(\boldsymbol{\theta}) = \ell_1(\alpha, \delta) + \ell_2(\boldsymbol{\phi}) \quad (6)$$

where $\ell_3(\alpha, \delta, \boldsymbol{\phi}) = \ell_3(\boldsymbol{\phi})$ depends only on $\boldsymbol{\phi}$ and is absorbed by $\ell_2(\boldsymbol{\phi})$

3. Applications to Rheumatoid Arthritis Clinical Trial Data

In this section we use data consisting of 200 subjects from the rheumatoid arthritis clinical data reported in Bombardier et. al., (1986) to illustrate different ways we can fit the disposition model when the data are incomplete. Because closed form of solutions to the score function do not exist, estimation of the parameters will be done using MULTIMAX (Kwagyan, 2001, Bonney, 2003, Kwagyan et. al., 2003) for maximization likelihood estimation.

Patients in this study have at most five unequally spaced binary self-assessment measurements of arthritis, where self-assessment equals 0 if "poor" and 1 if "good". An initial self-assessment measurement of all patients was recorded at the first time point (month 1), after which follow-up self-assessment measurements were taken monthly up to 5 months (k=2,3,4,5).

Patients were randomized to one of the two treatments: placebo or auranofin at the second self-assessment time. After randomization, patients remained in the treatment groups for the entire 5 months study period. The covariates used in this study are age in years, sex (1= male,

0=female), and treatment (1=Auranofin, 0=Placebo). Details about eligibility criteria and the result of the study are reported elsewhere (Bombardier et. al., 1986).

Of the 200 patients, 33(16.5%) subjects had some of their response missing. Missingness is assumed to be monotone in the sense that if an observation is missing at time t_k it is missing for time $t_j, j > k$. The primary objective of the study was to determine the effects of treatment to positive self-assessment, adjusted for age and gender. Of importance to us is the effect of the dropout process to positive self-assessment. If we work with the assumption that subjects who showed no improvement or positive response to the treatment are likely to dropout of the study, then we cannot rule out a DAR mechanism as such, we constrain $\phi_2 = 0$.

To test this assumption, three different models: a complete case where the analysis is based on those subject who did not have missingness in their observation, and two incomplete models that incorporates the dropout will be fitted. We consider the case when the regression parameters in the response model and dropout process are the same and when they are different.

The dropout probability is model by:

$$\theta = \text{logit}[p_k] = \phi_0 + \phi_1 y_{k-1} + \phi_3 \text{SEX} + \phi_4 \text{AGE} + \phi_5 \text{TRT}_k \quad (7)$$

while the logit of the individual disposition and the relative disposition are modeled by:

$$\begin{cases} \text{logit}[\delta_k] = \gamma_0 + \lambda_0 + \beta_1 \text{SEX} + \beta_2 \text{AGE} + \beta_3 \text{TRT}_k \\ \alpha = \frac{1 + e^{-(\gamma_0 + \lambda_0)}}{1 + e^{-\gamma_0}} \end{cases} \quad (8)$$

where γ_0 is the parameter measuring the within cluster or group dependence and λ_0 is the intercept or the mean effect.

3.1 Results of Analysis

Three different analyses are carried out to investigate the impact of the dropout process in the estimation of the response variables.

Complete Case: This analysis is done by deleting all the subjects with missing values from the data set and estimate the parameters using only the dataset from those subjects without missing values also called the completers using the disposition model given by Eqn. (8).

Incomplete Model DAR: For this model, the parameter for the current response $\phi_2=0$ is constrained while assuming the covariate parameters for the dropout model and the model of disposition are the same. This is done because of the need to ascertain the significance or non-significance of the missingness.

$$\text{logit}[\delta_k] = \gamma_0 + \lambda_0 + \beta_1 \text{SEX} + \beta_2 \text{AGE} + \beta_3 \text{TRT}_k$$

$$\text{logit}[p_k] = \phi_0 + \phi_1 y_{k-1} + \beta_1 \text{SEX} + \beta_2 \text{AGE} + \beta_3 \text{TRT}_k$$

$$\alpha = \frac{1 + e^{-(\gamma_0 + \lambda_0)}}{1 + e^{-\gamma_0}}$$

Incomplete Model DAR II: In this model, we want to know if the covariates had any effect or significance on the dropout process. To do this, we work with the same DAR assumption and chose different parameters for the covariates in the dropout and the model of disposition respectively. The models are:

$$\text{logit}[\delta_k] = \gamma_0 + \lambda_0 + \beta_1 \text{SEX} + \beta_2 \text{AGE} + \beta_3 \text{TRT}_k$$

$$\text{logit}[p_k] = \phi_0 + \phi_1 y_{k-1} + \phi_3 \text{SEX} + \phi_4 \text{AGE} + \phi_5 \text{TRT}_k$$

$$\alpha = \frac{1 + e^{-(\gamma_0 + \lambda_0)}}{1 + e^{-\gamma_0}}$$

Table 1 shows results of the fitted models.

Complete Case: We observe that the parameter γ_0 measuring the within cluster dependence was not statistically significant, the sex (β_1) of the subjects and the treatment (β_3) received were statistically significant to the way the subjects perceived the positive self assessment of their arthritis status. The result suggests that treatment with auranofin tends to increase the odds of a positive self-assessment by $e^{1.021} \approx 2.80$. There was no age effect to a positive self-assessment of the subject as the age (β_2) of the subjects was not statistically significant. This result is similar to the result obtained by Fitzmaurice and Lipsitz, (1995) who analyzed a subset of the data under the assumption of missing completely at random using the generalized estimation equations.

Table 1: Parameter estimates and standard errors of the Complete Case and Incomplete data.

	Complete Case	DAR I	DAR II
Disposition parameters			
λ_0	0.4902 (0.3939)	0.3787 (0.3659)	0.2642 (0.3647)
γ_0	0.19746 (0.1687)	0.1475 (0.0563)*	0.1182 (0.052)*
SEX (β_1)	0.3609 (0.1449)*	0.3811 (0.1723)*	0.3921 (0.1569)*
AGE (β_2)	-0.0065 (0.0071)	-0.0014 (0.0074)	-0.0032 (0.0066)
TRT (β_3)	0.4840 (0.1554)*	0.4840 (0.1554)*	0.8341 (0.1434)*
Dropout parameters			
y_{k-1} (ϕ_1)	-	-0.8908 (0.4009)*	-0.9846 (0.3981)*
SEX (ϕ_3)	-	-	0.5568 (166.24)
AGE (ϕ_4)	-	-	-0.00254 (17.96)
TRT (ϕ_5)	-	-	-0.0076 (-)
log Lik. value	-541.35	-529.15	-534.8
-2 log Lik.	1082.7	1058.3	1069.6
AIC	1092.7	1066.3	1085.6

Incomplete Models: The parameter γ_0 measuring the within cluster dependence is statistically significant for DAR I, and DAR II. This means that the clusters are correlated. This

was expected since the observation is repeated in each experiment with only one subject in each cluster. Also, there was no age effect to the positive self-assessment of the subject as the age (β_2) of the subjects was not statistically significant. Further, there were gender and treatment effects as the gender (β_1) and treatment (β_3) parameters in the response model were statistically significant for DAR II models and I. The result suggests that treatment with auranofin has the tendency to increase the odds of a positive self-assessment by $e^{0.4840} \approx 1.62$ and $e^{0.8341} \approx 2.30$ for DAR I and DAR II models respectively.

The dropout parameter ϕ_1 measuring the response status at the previous time point was statistically significant for model DAR I when the covariates of the dropout model and the response model was the same meaning that we cannot rule out a DAR dropout mechanism. A similar result was obtained when the parameters of the covariates in the dropout and response model were different in DAR II. However, the covariate parameters ϕ_3, ϕ_4 , and ϕ_5 for sex, age and treatment were not statistically significant. This means that the dropout is depended on the outcome of the previous visit (past history) rather than the treatment, gender or age of the subject.

Negative estimates of the dropout parameter ϕ_1 imply that the dropout time is more likely for those subjects who did not show any positive improvement in their self-assessment at the previous visit. The result suggests that holding other covariates constant, patients who did not show any positive improvement in their self-assessment at the previous visit are likely to have $e^{-0.8909} \approx 0.41$ and $e^{-0.9846} \approx 0.37$ times odds of continuing the study than their counterparts who experienced a positive self assessment of arthritis for DAR II and I respectively. The DAR I model was the best fitted according to Akaike's AIC, and the Likelihood Ratio Test. So we can conclude, for this example that the dropout process is random.

Finally, since the purpose of this study was to determine the effects of treatment and the impact of the dropout process to positive self-assessment, adjusted for age and gender, a complete case analysis that cannot capture the dropout process cannot be relied upon to produce an unbiased estimate even though the result in the disposition parameters across the three models are similar. For example, the complete case analysis showed there is no dependence or correlation within the clusters, whereas, there is correlation within the clusters as captured by the incomplete DAR models.

It is not uncommon for the dropout process to only depend on the observed history. If this is the case, then incomplete DAR I model should be adopted. However, it is possible that the reason for the dropout is related to the observed history of the patient and other covariates. To analyze data that falls within this framework, the incomplete DAR II model should be used.

4. Concluding Remarks

In discussing an example to illustrate the applications of the disposition model to longitudinal binary response when the data is incomplete, we have provided two different models that can be fitted when the dropout mechanism is assumed to be at random. We considered the case when the regression parameters in the response model and dropout model are the same and when they

are different. The choice of a model, for any given datasets, should be guided by the purpose of the analysis and assumption of the dropout process.

Acknowledgment

The Authors are grateful to the anonymous reviewer for helpful comments and suggestions on earlier drafts of the paper.

This work is supported by: NIH/NCAT grant UL1TR000101 and NIH/NIMHHD grant G12MD007597.

Appendix: Estimation and Inference

In this section we determine the maximum likelihood estimates MLE of the parameters of Eqn. (6). Since we are dealing with binary outcomes, we will adopt the logit transformation and model the parameters (α, δ) in terms of certain covariate as:

$$\alpha_i(\mathbf{\Lambda}, \mathbf{\Gamma}) = \frac{1 + e^{-\{M(\mathbf{Z}_i) + D(\mathbf{Z}_i)\}}}{1 + e^{-M(\mathbf{Z}_i)}}$$

$$\delta_{ik}(\mathbf{\Lambda}, \mathbf{\Gamma}, \mathbf{\beta}) = \frac{1}{1 + e^{-\{M(\mathbf{Z}_i) + D(\mathbf{Z}_i) + W(\mathbf{X}_{ik})\}}}$$

where $M(\mathbf{Z}_i)$ denotes the mean effects, $D(\mathbf{Z}_i)$ represents the within group dependence and $W(\mathbf{X}_{ik})$ is a function that describes the individual covariates effect, which are parameterized as

$$M(\mathbf{Z}_i) = \lambda_0 + \lambda_1 Z_{i1} + \dots + \lambda_q Z_{iq}$$

$$D(\mathbf{Z}_i) = \gamma_0 + \gamma_1 Z_{i1} + \dots + \gamma_q Z_{iq}$$

$$W(\mathbf{X}_{ik}) = \beta_1 X_{ik1} + \dots + \beta_p X_{ikp}.$$

The set of parameters to be determined in the model is therefore.

$$\mathbf{\Phi} = (\mathbf{\Lambda}, \mathbf{\Gamma}, \mathbf{\beta}) = \{\gamma_0, \dots, \gamma_q, \lambda_0, \dots, \lambda_q, \beta_1, \dots, \beta_p\}$$

For the ignorable model with reduced log-likelihood function given by Eqn. (6), the score function of the log-likelihood function is.

$$\frac{\partial \ell(\mathbf{\Theta})}{\partial \mathbf{\Theta}} = \frac{\partial \ell_1(\mathbf{\Phi})}{\partial \mathbf{\Theta}} + \frac{\partial \ell_2(\mathbf{\phi})}{\partial \mathbf{\Theta}}$$

where $\mathbf{\Theta} = (\mathbf{\Phi}, \mathbf{\phi})^T$ is the vector parameter of the model.

Since $\ell_2(\mathbf{\phi})$ does not depend on $\mathbf{\Phi}$,

$$\frac{\partial \ell(\boldsymbol{\Theta})}{\partial \boldsymbol{\Phi}} = \frac{\partial \ell_1(\boldsymbol{\Phi})}{\partial \boldsymbol{\Phi}}$$

Let.

$$\tau_{ik} = M(\mathbf{Z}_i) + D(\mathbf{Z}_i) + W(\mathbf{X}_{ik})$$

Then

$$\tau_{ik}^{(1)} = \frac{\partial \tau_{ik}}{\partial \boldsymbol{\Phi}} = \begin{bmatrix} \mathbf{Z}_i \\ \mathbf{Z}_i \\ \mathbf{X}_{ik} \end{bmatrix}$$

and further define the following formulas for the i^{th} unit

$$\tau_{ik}^{(2)} = \frac{\partial \tau_{ik}^{(1)}}{\partial \boldsymbol{\Phi}}$$

$$L_{0i}(\boldsymbol{\Phi}|\mathbf{y}_i) = \prod_{k=1}^{d_i-1} \delta_{ik}(1 - \delta_{ik})^{(1-y_{ik})}$$

$$\ell_{0i}(\boldsymbol{\Phi}|\mathbf{y}_i) = \log L_{0i}(\boldsymbol{\Phi}|\mathbf{y}_i) = \sum_{k=1}^{d_i-1} \{y_{ik} \log \delta_{ik} + (1 - y_{ik}) \log(1 - \delta_{ik})\}$$

$$\mathbf{S}_{0i}(\boldsymbol{\Phi}|\mathbf{y}_i) = \frac{\partial}{\partial \boldsymbol{\Phi}} \log L_{0i}(\boldsymbol{\Phi}|\mathbf{y}_i) = \sum_{k=1}^{d_i-1} (y_{ik} - \delta_{ik}) \tau_{ik}^{(1)}$$

$$\mathbf{H}_{0i}(\boldsymbol{\Phi}|\mathbf{y}_i) = - \sum_{k=1}^{d_i-1} \{\delta_{ik}(1 - \delta_{ik}) \tau_{ik}^{(1)} \tau_{ik}^{(1)T}\}$$

$$I_{0i}(\boldsymbol{\Phi}|\mathbf{y}_i) = -E\{\mathbf{H}_{0i}(\boldsymbol{\Phi}|\mathbf{y}_i)\} = \sum_{k=1}^{d_i-1} \{\delta_{ik}(1 - \delta_{ik}) \tau_{ik}^{(1)} \tau_{ik}^{(1)T}\}$$

Then

$$L_i(\boldsymbol{\Phi}|\mathbf{y}_i) = (1 - \alpha_i) \prod_{k=1}^{d_i-1} (1 - y_{ik}) + \alpha_i L_{0i}(\boldsymbol{\Phi}|\mathbf{y}_i)$$

the log-likelihood term is given by

$$\ell_1(\Phi|\mathbf{y}_i) = \log L_i(\Phi|\mathbf{y}_i) = \log \left\{ (1 - \alpha_i) \prod_{k=1}^{d_i-1} (1 - y_{ik}) + \alpha_i L_{0i}(\Phi|\mathbf{y}_i) \right\}$$

The Score Vector with respect to Φ is obtained from

$$\mathbf{S}(\Phi|\mathbf{y}_i) = \frac{\partial}{\partial \Phi} \ell_1(\Phi|\mathbf{y}_i) = A_i(\Phi) \alpha_i^* B_i(\Phi) + \mathbf{S}_{0i}(\Phi|\mathbf{y}_i) \quad (\text{A1})$$

where

$$A_i = \frac{\alpha_i [L_{0i}(\Phi|\mathbf{y}_i) - \prod_{k=1}^{d_i-1} (1 - y_{ik})]}{L_i}$$

$$B_i = \frac{\alpha_i L_{0i}(\Phi|\mathbf{y}_i)}{L_i}$$

$$\alpha_i^* = \frac{\partial}{\partial \Phi} \log \alpha_i$$

Thus, the Score vector with respect to the parameters Φ is given by

$$\mathbf{S}(\beta) = \frac{\partial}{\partial \beta} \ell_1(\Phi|\mathbf{y}_i) = A_i \alpha_i^*(\beta) + B_i \mathbf{S}_{0i}(\beta)$$

$$\mathbf{S}(\Gamma) = \frac{\partial}{\partial \Gamma} \ell_1(\Phi|\mathbf{y}_i) = A_i \alpha_i^*(\Gamma) + B_i \mathbf{S}_{0i}(\Gamma)$$

$$\mathbf{S}(\Lambda) = \frac{\partial}{\partial \Lambda} \ell_1(\Phi|\mathbf{y}_i) = A_i \alpha_i^*(\Lambda) + B_i \mathbf{S}_{0i}(\Lambda)$$

The Hessian terms are obtained by taking the second partial derivatives of Eqn. (A1) which are written in compact form as

$$\mathbf{H}(\Phi) = \sum_{i=1}^N \left\{ \frac{\prod_{k=1}^{d_i-1} (1 - y_{ik}) A_i \alpha_i^* \alpha_i^{*T}}{L_i} + \frac{\prod_{k=1}^{d_i-1} (1 - y_{ik}) B_i}{L_i} [\alpha_i^* \mathbf{S}_{0i}^T(\Phi) + \mathbf{S}_{0i}(\Phi) \alpha_i^{*T}] \right. \\ \left. + \frac{\prod_{k=1}^{d_i-1} (1 - y_{ik}) A_i \alpha_i^* \alpha_i^{*T}}{L_i} B_i \mathbf{S}_{0i}(\Phi) \mathbf{S}_{0i}^T(\Phi) + B_i \mathbf{H}_{0i}(\Phi) + A_i \frac{\partial}{\partial \Phi} \alpha_i^* \right\}$$

So the Hessian terms are:

$$\mathbf{H}_{\beta\beta} = \sum_{i=1}^N \left\{ \frac{\prod_{k=1}^{d_i-1} (1 - y_{ik}) A_i \alpha_i^* \alpha_i^{*T}}{L_i} B_i \mathbf{S}_{0i}(\boldsymbol{\beta}) \mathbf{S}_{0i}^T(\boldsymbol{\beta}) + B_i \mathbf{H}_{0i}(\boldsymbol{\beta}) \right\}$$

$$\begin{aligned} \mathbf{H}_{\Gamma\Gamma} = \sum_{i=1}^N & \left\{ \frac{\prod_{k=1}^{d_i-1} (1 - y_{ik}) A_i \alpha_i^*(\boldsymbol{\Gamma}) \alpha_i^{*T}(\boldsymbol{\Gamma})}{L_i} + \frac{\prod_{k=1}^{d_i-1} (1 - y_{ik}) B_i}{L_i} [\alpha_i^*(\boldsymbol{\Gamma}) \mathbf{S}_{0i}^T(\boldsymbol{\Gamma}) + \mathbf{S}_{0i}(\boldsymbol{\Gamma}) \alpha_i^{*T}(\boldsymbol{\Gamma})] \right. \\ & \left. + \frac{\prod_{k=1}^{d_i-1} (1 - y_{ik}) A_i \alpha_i^* \alpha_i^{*T}}{L_i} B_i \mathbf{S}_{0i}(\boldsymbol{\Gamma}) \mathbf{S}_{0i}^T(\boldsymbol{\Gamma}) + B_i \mathbf{H}_{0i}(\boldsymbol{\Gamma}) + A_i \frac{\partial}{\partial \boldsymbol{\Gamma}} \alpha_i^*(\boldsymbol{\Gamma}) \right\} \end{aligned}$$

$$\begin{aligned} \mathbf{H}_{\beta\Gamma} = \sum_{i=1}^N & \left\{ \frac{\mathbf{1}}{L_i} \left\{ L_{0i} \alpha_i^{*T}(\boldsymbol{\Gamma}) \mathbf{S}_{0i}^T(\boldsymbol{\beta}) + \alpha_i(\boldsymbol{\Omega}) \mathbf{S}_{0i}^T(\boldsymbol{\Gamma}) \mathbf{S}_{0i}(\boldsymbol{\beta}) + \alpha_i(\boldsymbol{\Omega}) L_{0i}(\boldsymbol{\Phi}) \sum_{k=1}^{d_i-1} \{\delta_{ik} (1 - \delta_{ik}) \mathbf{X}_{ik} \mathbf{Z}_i^T\} \right\} \right. \\ & \left. - \frac{\alpha_i(\boldsymbol{\Omega}) L_{0i}(\boldsymbol{\Phi}) \mathbf{S}_{0i}(\boldsymbol{\beta})}{L_i^2} \left\{ \alpha_i(\boldsymbol{\Omega}) [L_{0i}(\boldsymbol{\Phi}) - \prod_{k=1}^{d_i-1} (1 - y_{ik})] \alpha_i^*(\boldsymbol{\Gamma}) + \alpha_i(\boldsymbol{\Omega}) L_{0i}(\boldsymbol{\Phi}) \mathbf{S}_{0i}^T(\boldsymbol{\Gamma}) \right\} \right\} \end{aligned}$$

$$\begin{aligned} \mathbf{H}_{\beta\Lambda} = \sum_{i=1}^N & \left\{ \frac{\mathbf{1}}{L_i} \left\{ L_{0i} \alpha_i^{*T}(\boldsymbol{\Lambda}) \mathbf{S}_{0i}^T(\boldsymbol{\beta}) + \alpha_i(\boldsymbol{\Omega}) \mathbf{S}_{0i}^T(\boldsymbol{\Lambda}) \mathbf{S}_{0i}(\boldsymbol{\beta}) + \alpha_i(\boldsymbol{\Omega}) L_{0i}(\boldsymbol{\Phi}) \sum_{k=1}^{d_i-1} \{\delta_{ik} (1 - \delta_{ik}) \mathbf{X}_{ik} \mathbf{Z}_i^T\} \right\} \right. \\ & \left. - \frac{\alpha_i(\boldsymbol{\Omega}) L_{0i}(\boldsymbol{\Phi}) \mathbf{S}_{0i}(\boldsymbol{\beta})}{L_i^2} \left\{ \alpha_i(\boldsymbol{\Omega}) [L_{0i}(\boldsymbol{\Phi}) - \prod_{k=1}^{d_i-1} (1 - y_{ik})] \alpha_i^*(\boldsymbol{\Lambda}) + \alpha_i(\boldsymbol{\Omega}) L_{0i}(\boldsymbol{\Phi}) \mathbf{S}_{0i}^T(\boldsymbol{\Lambda}) \right\} \right\} \end{aligned}$$

$$\begin{aligned} \mathbf{H}_{\Lambda\Lambda} = \sum_{i=1}^N & \left\{ \frac{\prod_{k=1}^{d_i-1} (1 - y_{ik})}{L_i} A_i \alpha_i^*(\boldsymbol{\Lambda}) \alpha_i^{*T}(\boldsymbol{\Lambda}) + \frac{\prod_{k=1}^{d_i-1} (1 - y_{ik})}{L_i} B_i [\alpha_i^*(\boldsymbol{\Lambda}) \mathbf{S}_{0i}^T(\boldsymbol{\Lambda}) + \mathbf{S}_{0i}(\boldsymbol{\Lambda}) \alpha_i^{*T}(\boldsymbol{\Lambda})] \right. \\ & \left. + \frac{(1 - \alpha_i) \prod_{k=1}^{d_i-1} (1 - y_{ik})}{L_i} B_i \mathbf{S}_{0i}(\boldsymbol{\Lambda}) \mathbf{S}_{0i}^T(\boldsymbol{\Lambda}) + B_i \mathbf{H}_{0i}(\boldsymbol{\Lambda}) + A_i \frac{\partial}{\partial \boldsymbol{\Lambda}} \alpha_i^*(\boldsymbol{\Lambda}) \right\}. \end{aligned}$$

Finally

$\mathbf{H}_{\Gamma\Lambda} = \mathbf{H}_{\Lambda\Gamma}^T$ is the term

$$\begin{aligned} \mathbf{H}_{\Gamma\Lambda} = & \sum_{i=1}^N \frac{1}{L_i} \{ \delta_{0i}(1 - \delta_{0i}) \alpha_i [L_{0i} - \prod_{k=1}^{d_i-1} (1 - y_{ik})] \} \mathbf{Z}_i \mathbf{Z}_i^T + [L_{0i} - \prod_{k=1}^{d_i-1} (1 - y_{ik})] \alpha_i^*(\Gamma) \alpha_i^{*T}(\Lambda) \\ & + \alpha_i \alpha_i^*(\Gamma) \mathbf{S}_{0i}(\Lambda) + L_{0i} \alpha_i^{*T}(\Lambda) \mathbf{S}_{0i}(\Gamma) + \alpha_i \mathbf{S}_{0i}^T(\Lambda) \mathbf{S}_{0i}(\Gamma) \\ & + \alpha_i L_{0i} \left(- \sum_{k=1}^{d_i-1} \delta_{ik} (1 - \delta_{ik}) \mathbf{Z}_i \mathbf{Z}_i^T \right) \\ & - \frac{\alpha_i \alpha_i^*(\Gamma) [L_{0i} - \prod_{k=1}^{d_i-1} (1 - y_{ik})] + \alpha_i L_{0i} \mathbf{S}_{0i}(\Gamma)}{L_i^2} \{ \alpha_i [L_{0i} - \prod_{k=1}^{d_i-1} (1 - y_{ik})] \alpha_i^{*T}(\Lambda) \\ & + \alpha_i L_{0i} \mathbf{S}_{0i}(\Lambda) \} \end{aligned}$$

In addition,

$$\begin{aligned} \mathbf{H}_{\phi\Gamma} &= \mathbf{H}_{\Gamma\phi}^T = \mathbf{0} \\ \mathbf{H}_{\phi\Lambda} &= \mathbf{H}_{\Lambda\phi}^T = \mathbf{0} \end{aligned}$$

Since $\ell_1(\Theta)$ does not depend on ϕ ,

$$\frac{\partial \ell(\Theta)}{\partial \phi} = \frac{\partial \ell_2(\phi)}{\partial \phi}$$

The score vector $\mathbf{S}(\phi | \mathbf{y}_i)$ is given by

$$\mathbf{S}(\phi | \mathbf{y}_i) = \frac{\partial \ell_2(\phi)}{\partial \phi} = - \sum_{i=1}^N \sum_{k=1}^{d_i-1} (y_{ik}, (\mathbf{X}_{ik})^T) p_k$$

where

$$(\mathbf{X}_{ik})^T = (X_{ik1}, \dots, X_{ikp})$$

and

$$p_k = p_k(\mathbf{H}_k; \phi)$$

The score equation is obtained by solving the equation

$$\mathbf{S}(\phi) = \mathbf{0}$$

The Hessian term is

$$\mathbf{H}_{\phi\phi} = - \sum_{i=1}^N \sum_{k=1}^{d_i-1} (y_{ik}, (\mathbf{X}_{ik})^T)^T \mathbf{U}_k (y_{ik}, (\mathbf{X}_{ik})^T)$$

where

$$\mathbf{U}_k = \text{diag}(p_k(1 - p_k)).$$

References

- [1] Albert, P.S., and Follman, D.A. (2003). A random Effects Transition Model for Longitudinal Binary Data with Informative Missingness. *Statistica Neerlandica*, 57,100-111.
- [2] Bombardier, C., Ware, J.H., and Russell, I.J. (1986). Auranofin Therapy and Quality of Life in Patients with Rheumatoid Arthritis. *Am. J. Med.* 81, 565-578.
- [3] Bonney, G.E. (1984). On the Statistical Determination of Major Gene Mechanism in Continuous Human Traits: Regressive Models. *American Journal of Medical Genetics*, 18, 731-749.
- [4] Bonney, G.E. (1986). Regressive Logistics Models for Familial Diseases and other Binary Traits. *Biometrics*, 42, 611-625.
- [5] Bonney, G.E. (1987). Logistic Regression for Dependent Binary Observations. *Biometrics*, 43, 951-973.
- [6] Bonney, G.E. (1998). Regression Analysis of Disposition to Correlated Outcomes. *Dept. of Biostatistics Technical Report No. 1998-001*. Fox Chase Cancer Center, Philadelphia, PA. USA.
- [7] Bonney, G.E. (2003). Disposition to a Correlated Binary Outcomes and its Regression Analysis. *Journal of Statistics, Biometry and Genetics*. 1, 1-30.
- [8] Diggle, P.J., Liang, K.Y., and Zeger, S.L. (1994). *Analysis of Longitudinal Data*. Clarendon Press, Oxford, Great Britain.
- [9] Diggle, P.J., Heagerty, P., Liang, K.Y. and Zeger, S.L. (2002). *Analysis of Longitudinal Data*. 2 ed. Clarendon Press, Oxford, Great Britain.
- [10] Diggle, P.J. and Kenward, M.G. (1994). Informative Dropout in Longitudinal Data Analysis (with discussion). *Journal of Royal Statistical Society B*, 43, 49-93.
- [11] Diggle, P.J. (2005). Dealing with Missing Values in Longitudinal Studies. *Statistical Analysis of Medical Data*, 203-228.
- [12] Diggle, P.J. (1998). Diggle-Kenward Model for Dropouts. In *Encyclopedia of Biostatistics*. 2 ed. Armitage, P. and Colton, T., 1160-1, New York:Wiley.

- [13] Erebholo, F.O. (2015). Application of the Disposition Model to the Analysis of Longitudinal Binary Outcome in the Presence of Incomplete Data. Ph.D dissertation submitted to the College of Art and Sciences Graduate School, Howard University, Washington DC. USA.
- [14] Fitzmaurice, G.M. and Laird, N.M. (1993). A likelihood-based Method for Analyzing Longitudinal Binary Responses. *Biometrika*, 80, 141-151.
- [15] Fitzmaurice, G.M., Laird, N.M., and Lipsitz, S.R. (1994). Analyzing Incomplete Binary Responses: A Likelihood-based Approach. *Biometrics*, 50, 601-612
- [16] Fitzmaurice, G.M. and Lipsitz, S.R. (1995). A Model for Binary Time Series data with Serial Odds Ratio Patterns. *Journal of the Royal Statistical Society, series C*. 44 (1), 51-61.
- [17] Fitzmaurice, G.M., Molenberghs, G., and Lipsitz, S.R. (1995). Regression Models for Longitudinal Binary responses with Informative Drop-outs. *Journal of Royal Statistical Society. Series B*, 57, 691-704.
- [18] Fitzmaurice, G.M., Heath, A.F. and Clifford, P. (1996). Logistic Regression Models for Binary Panel with Attrition. *Journal of Royal Statistical Society. A*, 159, 249-263.
- [19] Kurland, B.F. and Heagerty, P.J. (2004). Marginalized Transition Models for Longitudinal Binary Data with Ignorable and Non-ignorable Dropout. *Statistics in Medicine*, 23, 2673-2695.
- [20] Kwagyan, J. (2001). Further Investigations of the Disposition Model for Correlated Binary Outcomes. Ph.D dissertation, The Temple University Graduate School, USA
- [21] Kwagyan, J. Apprey, V., and Bonney, G.E. (2003) Maximum Likelihood Inference in the Models of Disposition for Correlated Binary Outcomes, *J. Statistics, Biometry and Genetics*, 1, 31-48.
- [22] Kwagyan, J. Apprey, V., and Bonney, G.E. (2003) Gram-Schmidt Parameter Orthogonalization on Maximum Likelihood. *J. Statistics, Biometry and Genetics*, 1, 49-62..
- [23] Kwagyan, J. Apprey, V., and Bonney, G.E. (2003) Fitting the Disposition Model via EM Algorithm. *J. Statistics, Biometry and Genetics*, 1, 63-79.
- [24] Laird, N.M. and Ware, J.H. (1982). Random-effects Models for Longitudinal Data. *Biometrics*, 38, 963-974.
- [25] Liang, K.Y. and Zeger, S.L (1986). Longitudinal Data Analysis using Generalized Linear Models. *Biometrika*, 73, 13-22.

-
- [26] Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. John Wiley and Sons, New York, USA.
- [27] Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. 2 ed. John Wiley and Sons, New York, USA.
- [28] Little, R.J.A. (1995). Modeling the Dropout Mechanism in Repeated Measure Studies. *J. of the American Statistical Association*, 90, 1112-1121.
- [29] Molenberghs, G., Fitzmaurice, G., Kenward, M.G., Tsiatis, A., and Verbeke, G. (2015). *Handbook of Missing Data Methodology*. Chapman and Hall/CRC. New York.
- [30] Molenberghs, G., and Kenward, M.G. (2005). *Missing Data in Clinical Studies*. John Wiley and Sons, Chichester, England.
- [31] Molenberghs, G., and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer, New York, USA.
- [32] Newey, W.K. (1993). Efficient Estimation of Models with Conditional Moment Restrictions. *Handbook of Statistics*, 11, 419-454.
- [33] Rubin, R.D. (1976). Inference and Missing Data. *Biometrika*, 63, 581-592.
- [34] Yi, G.Y. and Thompson, M.E. (2005) Marginal and Association Regression Models for Longitudinal Binary Data with Drop-outs: A Likelihood-based Approach. *The Canadian Journal of Statistics*, 33, 3-20.
- [35] Zeger, C.L., and Liang, K.Y. (1986). Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics*, 42, 121-130.

Received March 15, 2015; accepted November 19, 2015.

Francis Erebholo
Department of Mathematics
Hampton University,
Hampton, VA, USA
francis.erbholo@hamptonu.edu

Victor Apprey
National Human Genome Center
Howard University
Washington DC, USA
vapprey@howard.edu

Paul Bezandry
Department of Mathematics
Howard University,
Washington DC, USA
pbezandry@howard.edu

John Kwagyan
Department of Community and Family Medicine
College of Medicine
Howard University Hospital
Washington DC, USA
jkwagyan@howard.edu