



Analysis of Incomplete Longitudinal Binary Data-A Combined Markov's Transition and Logistic Model for Non-ignorable Missingness

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

¹Department of Mathematics

Hampton University

Hampton, Virginia 23668 USA

Email: francis.erbholo@hamptonu.edu

²Department of Mathematics

Howard University, Washington DC 20059 USA

³National Human Genome Center

Howard University, Washington DC 20059 USA

⁴Department of Community and Family Medicine

Howard University, Washington DC 20059 USA

* Corresponding Author

Received: November 27, 2015; Accepted: May 24, 2016

Abstract

The problem of incomplete data is a common phenomenon in research that involves the longitudinal design approach. We investigate and develop a likelihood-based approach for incomplete longitudinal binary data using the disposition model when the missing value mechanism is non-ignorable. We combined Markov's transition and a logistic regression model to build the dropout process and model the response using conditional logistic regression model. By holding the missingness parameter that is weakly identified constant, we analyzed their effects through a sensitivity analysis as the estimation of parameters in MLE for non-ignorable missing data is not generally plausible. An application of our approach to Schizophrenia clinical trial is presented.

Keywords: Binary data; Disposition model; Dropout mechanism; Maximum Likelihood Estimation; Non-ignorable missingness

MSC 2010 No.: 62F03, 62J02, 62H12, 62H20

1. Introduction

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.

Incomplete or missing data are common occurrences in longitudinal studies because many subjects are not available to be measured at all points. A subject may miss an appointment for a measurement and is never measured again resulting in a monotone missing data pattern. Further, a subject can be missing at one follow-up time and be available to be measured at one of the next, resulting in non-monotone missing data pattern. These kind of missing data, if not handled or accounted for properly could lead to a bias when inferences on one or more covariates on a response variable of interest are made.

Missing data could be related to, or unrelated to the outcome of interest. When it is unrelated to the outcome of interest, the effect is weak and analyses of the parameters of interest are less complicated. However, when it is related to the outcome of interest, the impact of the missing data is great, and the analyses, which are 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 response, 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 are correctly specified (Yi et al., 2005, 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 attractive feature of reproducibility of the disposition model (Bonney, 1998, 2003) makes it desirable to naturally extend it to capture the type of correlation or dependence that arises in longitudinal data. The original development of the disposition model starts with random effects formulation and then introduces a theory for constructing likelihoods utilizing moment series representations. Kwagyan (2001) further investigated the disposition model through an alternative formulation from a finite mixture modeling perspective. Erebholo (2015) and Erebholo et al. (2016) adopt the disposition model, and extend it to the analysis of longitudinal binary outcomes in the presence of monotone incomplete data under the dropout at random mechanism.

The paper is organized as follows. In Section 2, we introduce 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 PANSS Schizophrenia data. Section 4 is focused on the

sensitivity analysis of the dropout not at random model. We end with concluding remarks in Section 5.

2. The 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. We will construct a joint distribution for the incomplete data and develop models for different dropout mechanism.

2.1. The Models of Disposition

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 with size $n_i \times 1$. Let δ_{ik} denote the conditional probability of $Y_{ik} = 1$ given that $Y_{ik'} = 1$. That is,

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

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

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

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, $0 < \alpha_i < 1$ implies positive aggregation, and $\alpha_i > 1$ implies negative aggregation. With this, Bonney (1998, 2003) has shown that the joint distribution of the i^{th} cluster is given as

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

In general, α_i and δ_{ik} are modeled as

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

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

where $M(\mathbf{Z}_i)$ represents the mean effect, $D(\mathbf{Z}_i)$ represents the within group dependence, and $W(\mathbf{X}_{ik})$ is the adjustment due to individual-specific covariates and are parameterized as

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

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

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

and

$$(\mathbf{\Lambda}, \mathbf{\Gamma}, \mathbf{\beta}) = \{ \gamma_0, \gamma_1, \dots, \gamma_q, \lambda_0, \lambda_1, \dots, \lambda_q, \beta_1, \dots, \beta_p \}$$

are the unknown parameters.

2.2. Modeling the Incomplete Data

Let $\mathbf{Y}^* = (Y_1^*, \dots, Y_n^*)$ denote the complete vector of intended sequence of measurement on an experimental unit, and $\mathbf{t}_i = (t_{i1}, \dots, t_{in_i})$ the set of times that corresponds to the intended measurement. Then the joint probability distribution of \mathbf{Y}^* is

$$P(\mathbf{Y}^* ; \alpha, \delta) = (1 - \alpha_i) \prod_{k=1}^{n_i} (1 - y_{ik}^*) + \alpha_i \prod_{k=1}^{n_i} \delta_{ik} (1 - \delta_{ik})^{(1-y_{ik}^*)}.$$

Let $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$ 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 ($Y_{ij}: j = 1, 2, \dots, k$) associated with it follows the same joint distribution as that of the corresponding intended sequence ($Y_{ij}^*: j = 1, 2, \dots, k$).

We define the preceding outcome Y_j as:

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

where D_i is a random variable.

For each k , let $\mathbf{H}_{ik} = (Y_{i1}, \dots, Y_{ik-1})$ denote the observed history up to time t_{ik-1} , and y_{ik}^* , the value that would have been observed at time t_{ik} , if there was no dropout in the unit. Analogous to Diggle and Kenward (1994) selection model with non-ignorable dropout, we assume that the probability of dropout at time d_i is assumed to depends on the history of the measurement process up to, and including the time of dropout t_{d_i} . That is,

$$\Pr(D_i = d_i | \text{History}) = p_d(\mathbf{H}_{d_i}, y_{d_i}^*; \boldsymbol{\phi}),$$

where $\boldsymbol{\phi} = (\phi_0, \dots, \phi_{2+p})$ is a vector of unknown parameters. With this, we identify the following patterns of dropout process:

Dropout Completely At Random (DCAR). Dropout is completely at random when the dropout process is independent of \mathbf{H}_{d_i} , and $y_{d_i}^*$. That is,

$$\Pr(D_i = d_i | \text{History}) = p_d(d_i; \boldsymbol{\phi}).$$

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

$$\Pr(D_i = d_i | \text{History}) = p_d(\mathbf{H}_{d_i}; \boldsymbol{\phi}).$$

Dropout Not At Random (DNAR). This is when the dropout process depends on $y_{d_i}^*$. That is,

$$\Pr(D_i = d_i | \text{History}) = p_d(\mathbf{H}_{d_i}, y_{d_i}^*; \boldsymbol{\phi}).$$

We adopt the regressive logistic models of Bonney (1986, 1987, 1998) to model the dropout process $p_k(\mathbf{H}_{ik}, y_i; \boldsymbol{\phi})$ and define the logit as

$$\begin{aligned} \theta_{ik} &= \text{logit}[p_k(\mathbf{H}_{ik}, y_i; \boldsymbol{\phi})], \\ &= \phi_0 + \phi_1 y_{d_i} + \sum_{j=2}^k \phi_j y_{ik+1-j} + \phi_{k+1} X_{ik1} + \dots + \phi_{k+p} X_{ikp}, \end{aligned} \quad (2)$$

where $\mathbf{X}_{ik} = (X_{ik1}, \dots, X_{ikp})^T$ is the p individual-specific covariates.

The reason for this choice is that the probability of dropout at time t_{d_i} is a direct consequence of the past outcomes, the present outcome, and possible set of covariates.

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

$$P(\mathbf{Y}_i) = P^*(y_{i1}, \dots, y_{d_i-1}) \left[\prod_{k=2}^{d_i-1} 1 - p_k(\mathbf{H}_{ik}, y_{ik}) \right] \Pr(Y_{d_i} = 0 | \mathbf{H}_{d_i}, Y_{d_i-1} \neq 0). \quad (3)$$

Hence, 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

$$\ell(\boldsymbol{\Theta}) = \sum_{I=1}^N \log \left\{ P^*(y_{i1}, \dots, y_{d_i-1}; \alpha, \delta) \left[\prod_{k=2}^{d_i-1} 1 - p_k(\mathbf{H}_{ik}, y_{ik}; \boldsymbol{\phi}) \right] \times \Pr(Y_{d_i} = 0 | \mathbf{H}_{d_i}, Y_{d_i-1} \neq 0; \alpha, \delta, \boldsymbol{\phi}) \right\},$$

and is partitioned as:

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

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,

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

and

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

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

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

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

Let us temporarily drop the subscript i for ease of notation and without the loss of generality. Following from Equation (2),

$$\begin{aligned} \theta_k &= \phi_0 + \phi_1 y_k + \phi_2 y_{k-1} + \phi_3 X_{k1} + \cdots + \phi_{2+p} X_{kp}, \\ \theta_d^* &= \phi_0 + \phi_1 y_d + \phi_2 y_{d-1} + \phi_3 X_{d1} + \cdots + \phi_{2+p} X_{dp}. \end{aligned}$$

Using Equation (5), Equation (3) becomes

$$\begin{aligned} P(\mathbf{y}) &= P^*(y_1, \dots, y_{d-1}) \left[\prod_{k=2}^{d-1} 1 - p_k(\mathbf{H}_k, y_k) \right] \\ &\quad \times \sum_{\mathbf{y}^{(m)}} p_d(\mathbf{H}_d, \mathbf{y}_d; \boldsymbol{\phi}) P_d^*(\mathbf{y}^{(m)} | \mathbf{H}_d; \alpha, \delta). \end{aligned} \quad (6)$$

We adopt the Markov transition model-to-model $P^*(Y_d = y_d | \mathbf{H}_d; \alpha, \delta)$.

Let

$$\pi_k = \Pr(Y_k = 1 | H_k)$$

and

$$\xi_k = \eta Y_{k-1}$$

be the first order Markov chain, where η is the dependence parameter; that is, the odds that compare the participants who did not drop out of the study at the current measure with the participants who dropped out of the study at the previous measure keeping all other covariates constant.

We now use the logit and model the function as

$$\xi_k = \log \left\{ \frac{\Pr(Y_k = 1 | H_k)}{1 - \Pr(Y_k = 1 | H_k)} \right\}.$$

So that

$$\pi_k = \frac{e^{\xi_k}}{1 + e^{\xi_k}}.$$

Hence, the conditional distribution of the current observation given the history is given by

$$P^*(y_k | \mathbf{H}_k; \eta) = \frac{e^{\xi_k y_k}}{1 - e^{\xi_k}}.$$

So Equation (6) becomes

$$P(\mathbf{y}) = P^*(y_1, \dots, y_{d-1}) \left[\prod_{k=2}^{d-1} 1 - p_k(\mathbf{H}_k, y_k; \boldsymbol{\phi}) \right] \sum_{y^{(m)}} p_d(\mathbf{H}_d, y_d; \boldsymbol{\phi}) P_d^*(y^{(m)} | \mathbf{H}_d; \eta). \quad (7)$$

By the definition of non-ignorable dropout, we can see that $\Pr(D = d | \mathbf{H}_d) = p_d(\mathbf{H}_d, y_d; \boldsymbol{\phi})$ solely depends on $y^{(m)} = y_k^*$ -the current unobserved response through y_d .

Also, the factor $\sum_{y^{(m)}} p_d(\mathbf{H}_d, y_d; \boldsymbol{\phi}) P_d^*(y^{(m)} | \mathbf{H}_d; \eta)$ in Equation (7) represents the conditional expectation of $p_d(\mathbf{H}_d, y; \boldsymbol{\phi})$ under the distribution of $P_d^*(y^{(m)} | \mathbf{H}_d; \eta) = P_d^*(y_d | \mathbf{H}_d; \eta)$ and is evaluated as

$$E[p_d(\mathbf{H}_d, y_d; \boldsymbol{\phi})] = \sum_{y_d=0,1} p_d(\mathbf{H}_d, y_d; \boldsymbol{\phi}) P_d^*(Y_d | \mathbf{H}_d; \eta),$$

$$E[p_d(\mathbf{H}_d, y_d; \boldsymbol{\phi})] = \pi_d q_{d1} + (1 - \pi_d) q_{d0}, \quad (8)$$

where

$$q_{d1} = p_d(\mathbf{H}_d, y_d = 1; \boldsymbol{\phi}) = \frac{e^{\theta^* | y_d=1}}{1 - e^{\theta^* | y_d=1}},$$

$$q_{d0} = 1 - q_{d1},$$

$$\pi_d = \frac{e^{\xi_d}}{1 - e^{\xi_d}},$$

$$\theta^* | y_d=1 = \theta_d^* = \phi_0 + \phi_1 + \phi_2 y_{d-1} + \phi_3 X_{k1} + \dots + \phi_{2+p} X_{kp}.$$

A compact form of the distribution of the incomplete data is obtained by substituting Equation (8) into Equation (7) as

$$P(\mathbf{y}) = P^*(y_1, \dots, y_{d-1}) \left\{ \prod_{k=2}^{d-1} 1 - p_k(\mathbf{H}_k, y_k) \right\} \{ \pi_d q_{d1} + (1 - \pi_d) q_{d0} \}.$$

Thus, the full log-likelihood for the i^{th} unit with an incomplete measurement sequence is

$$\ell(\boldsymbol{\Theta}) = \ell_1(\alpha, \delta) + \ell_2(\boldsymbol{\phi}) + \ell_3(\boldsymbol{\phi}, \eta), \quad (9)$$

where

$$\begin{aligned}\ell_1(\alpha, \delta) &= \sum_{i=1}^N \log\{(1 - \alpha_i) \prod_{k=1}^{d_i-1} (1 - y_{ik}) + \alpha_i \prod_{k=1}^{d_i-1} \delta_{ik} (1 - \delta_{ik})^{(1-y_{ik})}\}, \\ \ell_2(\boldsymbol{\phi}) &= -\sum_{i=1}^N \sum_{k=2}^{d_i-1} \log[1 + e^{\theta_{ik}}], \\ \ell_3(\boldsymbol{\phi}, \boldsymbol{\eta}) &= \sum_{i \leq N; d_i \leq n_i} \log\{\pi_{d_i} q_{d1} + (1 - \pi_{d_i}) q_{d0}\}.\end{aligned}$$

Since closed form solution of the score function does not exist, numerical techniques will be used to obtain the estimates of the parameters of interest.

3. Application to PANSS Clinical Study

In this section we use data from the PANSS Schizophrenia data to illustrate different ways we can fit the disposition model when the data is incomplete. Estimation of the parameters will be done using MULTIMAX (Kwagyan, 2001, Bonney, 2003, Kwagyan et al. 2003) for maximization likelihood estimation.

These data were analyzed by Kurland (2002), Kurland and Heagerty (2004) using marginalized transition model. The Positive and Negative Syndrome Scale (PANSS) schizophrenia study (Chouinard et al., 1993; Marder and Meibach, 1994; Kurland, 2002) is a longitudinal clinical trial with monotone pattern of missingness (or dropout). Data consisted of 519 participants that were randomly placed into six different treatment groups: Placebo, Haloperidol 20mg/day, Risperidone at 2mg, 6mg, 10mg and 16mg/day over a period of 8 weeks.

The treatment covariates are: PLAC (1=placebo, 0 otherwise), RISP[2mg] (1 = risp (2), 0 otherwise), RISP[6mg] (1 = risp (6), 0 otherwise), RISP[10mg] (1 = risp (10), 0 otherwise), RISP[16mg] (1 = risp. (16), 0 otherwise) and HALO (1=haloperidol, 0 otherwise). In our study, we considered placebo, haloperidol, low dose of risperidone (2mg & 6mg), and high dose of risperidone (10 mg & 16 mg) over the 5 post-baseline scores. The treatment covariates used were: PLAC (1 = placebo, 0 otherwise), RISP[Low] (1= risp (2, 6), 0 other- wise), RISP[High] (1=risp (10, 16), 0 otherwise), and HALO (1=haloperidol, 0 otherwise).

Following Chouinard et al. (1993) and Marder and Meibach (1994), we used binary outcome, which was dichotomized as clinically significant improvement in symptoms of subject at time k , at a 20% reduction compared to baseline according to PANSS. Of 519 patients, 275 (53%) had some of their responses missing. We deleted and excluded 13 observations from the data because they did not have any measurement at the baseline and post baseline time, while the entries of two of the participants with non-monotone data structure were deleted to make their data monotone. In so doing, we had a total of 506 participants with 2531 measured response.

The primary research question is to know how patients respond to haloperidone, and risperidone in the treatment of schizophrenia. In addition, we seek to understand the effects of the dropout process in the treatment of schizophrenia. In the analysis, we considered the case when the regression parameters in the response and dropout models are the same and when they are different.

The dropout probability is modeled as

$$\theta = \text{logit}[p_k] = \phi_0 + \phi_1 y_{k-1} + \phi_2 y_k + \phi_3 HALO + \phi_4 RISP(L) + \phi_5 RISP(H). \quad (10)$$

The logit of the individual disposition and the relative disposition are modeled as

$$\begin{cases} \text{logit}[\delta_k] = \gamma_0 + \lambda_0 + \beta_1 HALO + \beta_2 RISP(L) + \beta_3 RISP(H), \\ \alpha = \frac{1+e^{-(\gamma_0+\lambda_0)}}{1+e^{-\gamma_0}}, \end{cases} \quad (11)$$

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

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

Complete Case: In this analysis, we delete all the subjects with missing values from the data set, and then estimate the parameters using only the data set from those subjects without missing values using the disposition model given by Equation (11).

Incomplete DAR I Model: For this analysis, the parameter for the current response ϕ_2 is constrained (i.e., $\phi_2 = 0$), 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.

$$\begin{aligned} \text{logit}[\delta_k] &= \gamma_0 + \lambda_0 + \beta_1 HALO + \beta_2 RISP(L) + \beta_3 RISP(H), \\ \text{logit}[p_k] &= \phi_1 y_{k-1} + \beta_1 HALO + \beta_2 RISP(L) + \beta_3 RISP(H), \\ \alpha &= \frac{1+e^{-(\gamma_0+\lambda_0)}}{1+e^{-\gamma_0}}. \end{aligned}$$

Incomplete DAR II Model: This analysis seeks to answer the question of the significance effect of the covariates on the dropout process. To do this, we work with the same DAR assumption and choose different parameters for the covariates in the dropout and the model of disposition respectively.

$$\begin{aligned} \text{logit}[\delta_k] &= \gamma_0 + \lambda_0 + \beta_1 HALO + \beta_2 RISP(L) + \beta_3 RISP(H), \\ \text{logit}[p_k] &= \phi_1 y_{k-1} + \phi_3 HALO + \phi_4 RISP(L) + \phi_5 RISP(H), \\ \alpha &= \frac{1+e^{-(\gamma_0+\lambda_0)}}{1+e^{-\gamma_0}}. \end{aligned}$$

Incomplete DNAR Model: In this analysis, the current response parameter is not constrained i.e., $\phi_2 \neq 0$, although ϕ_0 and ϕ_1 may be constrained.

$$\begin{aligned} \text{logit}[\delta_k] &= \gamma_0 + \lambda_0 + \beta_1 \text{HALO} + \beta_2 \text{RISP(L)} + \beta_3 \text{RISP(H)}, \\ \text{logit}[p_k] &= \phi_1 y_{k-1} + \phi_2 y_k + \beta_1 \text{HALO} + \beta_2 \text{RISP(L)} + \beta_3 \text{RISP(H)}, \\ \text{logit}[\xi_k] &= \eta y_{d-1}, \\ \alpha &= \frac{1+e^{-(\gamma_0+\lambda_0)}}{1+e^{-\gamma_0}}. \end{aligned}$$

Table 1 shows results of the fitted models.

Complete Case: When fitted, we observed that the parameter γ_0 , measuring the within cluster dependence was statistically significant. In addition, there was no haloperidone treatment effect since it was not statistically significant. However, the low and high doses of risperidone were statistically significant in the treatment of schizophrenia. It is estimated that the patients taking the high and low doses of risperidone have $e^{0.7255} \approx 2.066$ and $e^{0.7706} \approx 2.161$ times higher odds to improve in the treatment of schizophrenia. In other words, treatment with both low and high doses of risperidone tends to increase the odds of a schizophrenia treatment.

Incomplete DAR I and DAR II Models: The parameter γ_0 , measuring the dependence within the cluster was statistically significant. This implies there is a strong correlation within the clusters. This was expected since the observation is repeated in each experiment with only one subject in each cluster. The parameter β_1 for the treatment of haloperidone was not statistically significant.

Table 1: "Parameter estimates and standard error for CC and DAR models"

| | Complete Case | DAR I | DAR II |
|------------------------|-------------------|-------------------|-------------------|
| Parameters | Est. (Std. error) | Est.(Std. error) | Est. (Std. error) |
| Disposition parameters | | | |
| λ_0 | -1.3422 (0.1475)* | -1.4636 (0.1482)* | -1.3863 (0.1494)* |
| γ_0 | 0.5413 (0.1011)* | 0.8349 (0.0861)* | 0.7346 (0.0663)* |
| HALO (β_1) | 0.3849 (0.1945) | 0.4041 (0.2938) | 0.4290 (0.2360) |
| RISP(L) (β_2) | 0.7255 (0.1680)* | 0.7956 (0.1688)* | 0.8386 (0.1709)* |
| RISP(H) (β_3) | 0.7706 (0.1685)* | 0.8058 (0.1700)* | 0.8326 (0.1718)* |
| Dropout parameters | | | |
| $y_{k-1}(\phi_1)$ | - | 1.2811(0.1995)* | 2.944 (0.4588)* |
| HALO (ϕ_3) | - | - | - |
| RISP(L) (ϕ_4) | - | - | -18.98 (826.30) |
| RISP(H) (ϕ_5) | - | - | -18.91 (772.39) |
| loglik. Value | -1193.4 | -1184.1 | -1147.4 |
| -2 loglik. | 2386.8 | 2368.2 | 2294 |
| AIC | 2396.8 | 2380.2 | 2310 |

Note: * means significant and y_{k-1} is schi. status at previous time point

However, the parameters β_2 and β_3 measuring the low and high doses of risperidone were statistically significant for both DAR models. This suggests that patients taking both the low and high doses of risperidone have $e^{0.7956} \approx 2.216$ and $e^{0.8058} \approx 2.239$ times higher odds to show clinical improvement in the treatment of schizophrenia for DAR I, and $e^{0.8386} \approx 2.313$ and $e^{0.8326} \approx 2.3$ times higher odds to show clinical improvement in the treatment of schizophrenia

for DAR II model.

The dropout parameter ϕ_1 measuring the previous response in both models was statistically significant. Positive estimate of the dropout parameter indicates that patients who showed positive clinical improvement in the treatment of schizophrenia are more likely to continue the study. This suggests that patients who showed clinical improvement in the previous response measurement have $e^{1.2811} \approx 3.6$ and $e^{2.944} \approx 18.99$ times higher odds of continuing the study for models DAR I and DAR II respectively. The covariate parameters ϕ 's in DAR II model were not statistically significant.

Although it seems both DAR models are good fits for the data, we cannot conclude just yet that the dropout mechanism is random without investigating the effect of the current response to the dropout process. To do this, we will investigate the DNAR model by incorporating the parameter ϕ_1 , measuring the effect of the response at the previous visit into the model.

Incomplete DNAR Model: An initial analysis of the DNAR model (results not published) revealed that the parameter for the current response ϕ_2 is weakly identified and as such, it was not significant even though the parameter ϕ_1 measuring the previous response was. This is not surprising as most DNAR parameters are not only weakly identified, but also their estimation will become sensitive to the assumptions of the distribution. In situations like this, a sensitivity analysis on the DNAR model will be performed.

4. Sensitivity Analysis for DNAR

In the spirit of Kurland (2002), we fix the parameter for the current response for values between $[-1.5, 1.5]$ and conduct a sensitivity analysis using DNAR to know the effect of the dropout process in the treatment of schizophrenia. For example, fixing $\phi_2 = 0.5$, the odds of a patient to remain in the study when he or she experiences a significant clinical improvement is $e^{0.5} \approx 1.65$ times the odds when the patient did not experience a significant clinical improvement.

In the same way, if $\phi_2 = -0.5$, the odds of a patient to remain in the study when he or she did not experience a significant clinical improvement is $e^{-0.5} \approx 0.61$ times the odds when the patient experiences a significant clinical improvement. For the DNAR model, a bound was found for the current response parameter ϕ_2 while estimation was carried out at selected points. Parameter estimates and model-based standard errors for the sensitivity analysis are presented. Two different analyses are fitted (the independence and dependence) based on the output of some preliminary analyses.

Table 2 and Table 3 below show the parameter estimates and standard error of the independence case ($\gamma_0 = 0$) and dependence case ($\gamma_0 \neq 0$).

Independence Case ($\gamma_0 = 0$): The optimal solution for the analysis was obtained when $\phi_2 = 1.0$ in Table 2. Both low and high doses of risperidone were statistically significant. This suggests that treatment with both low and high doses of risperidone tends to increase the odds of a schizophrenia treatment by $e^{1.1193} \approx 3.06$ and $e^{1.0733} \approx 2.92$ respectively. In addition, the parameter γ_0 , which measures the correlation within the groups was statistically significant, while the Markov parameter η , was not. Now, $\phi_2 = 1.0$ and ϕ_1 statistically significant with

positive estimates suggests that patients who demonstrate a positive clinical improvement at the previous visit are estimated to have $e^{1.0} \approx 2.71$ times higher odds of remaining in the study.

Dependence Case ($\gamma_0 \neq 0$): From Table 3, the optimal solution for this analysis was obtained when $\phi_2 = 0.8$. The parameter γ_0 measuring correlation within the groups (cluster) was statistically significant while η was not. The low dose and high dose of risperidone were statistically significant. With this, it is estimated that patients taking the low dose of haloperidone have a $e^{1.2525} \approx 3.5$ higher odds to experience significant improvement in their treatment of schizophrenia while those taking a higher dose of risperidone have $e^{1.0605} \approx 2.9$ higher odds to experience significant improvement in treatment of schizophrenia. In other words, treatment with both low and high doses of risperidone tends to increase the odds of a schizophrenia treatment by $e^{1.2525} \approx 3.5$ and $e^{1.0605} \approx 2.9$ respectively.

Now, $\phi_2 = 0.8$, and ϕ_1 statistically significant with positive estimates imply that patients who demonstrate a positive clinical improvement at the previous visit are estimated to have $e^{0.8} \approx 2.23$ times higher odds of remaining in the study than their counterparts who did not show any significant improvement. Finally, a comparison of DNAR with the complete case according to Akaike's Information Criteria (AIC) showed that the DNAR model is a better fit.

5. Conclusion

To study a procedure for fitting, and analyzing the model of disposition in the presence of incomplete or missing data, we adopted the selection model of Diggle and Kenward (1994) for binary response and extended it to model the joint distribution of the incomplete data reported in Erebholo et al. (2016) under the ignorable dropout condition. For the non-ignorable mechanism, we developed a combined Markov's transition and a logistic regression model to build the dropout process while modeling the response using conditional logistic regression.

In discussing an example to illustrate this application, we considered the case when the regression parameters in the response model and dropout model are the same and when they are different. The ignorable and non-ignorable models are fitted. When the dropout mechanism is not ignorable, we hold the dropout parameters that are weakly identified constant and analyzed their effects through a sensitivity analysis.

The choice for a model, for any given data sets, should be guided by the purpose of the analysis and assumptions of the dropout process. For example, 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 fall within this framework, the incomplete DAR II model should be used. To analyze data for non-ignorable dropout analysis, when the DNAR parameter is weakly identified (as was in our example) a sensitivity analysis is recommended to know the effect of the current response to the dropout process.

Finally, for this example, both DAR and DNAR models are good fits and as such, choosing a specific model to adopt for an analysis could be very difficult to justify. Because of this, there is

a need to be very careful in deciding on a model to adopt.

Acknowledgements

The authors are grateful to the anonymous reviewers and the Editor-in-Chief Professor Aliakbar M. Haghghi for helpful comments and suggestions on earlier draft of the paper. This work is supported by: National Institute of Health/National Center for Advancing Translational Sciences Grant UL1TR000101 and National Institute of Health Grant G12MD007597.

REFERENCES

- Bonney, G.E. (1986). Regressive Logistics Models for Familial Diseases and other Binary Traits. *Biometrics*, Vol.42, 611-625.
- Bonney, G.E. (1987). Logistic Regression for Dependent Binary Observations. *Biometrics*, Vol. 43, 951-973.
- 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.
- Bonney, G.E. (2003). Disposition to a Correlated Binary Outcome and its Regression Analysis. *Journal of Statistics, Biometry and Genetics*, Vol.1, 1-30.
- Chouinard, G., Jones, B., Remington, G., Bloom, D., Addington, D., MacEWAN, G. W., Labelle, A., Beauclair, L., and Arnott, W. (1993). A Canadian Multicenter Placebo-Controlled Study of Fixed Doses of Risperidone and Haloperidol in the Treatment of Chronic Schizophrenic Patients. *Journal of clinical psychopharmacology*, Vol. 13, 25-40.
- Diggle, P.J. and Kenward, M.G. (1994). Informative Dropout in Longitudinal Data Analysis (with discussion). *Journal of Royal Statistical Society B*, Vol. 43, 49-93.
- Diggle, P.J., Heagerty, P., Liang, K.Y, and Zeger, S.L. (2002). *Analysis of Longitudinal Data*. 2 Ed. Clarendon Press, Oxford, Great Britain.
- 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.
- Erebholo, F.O, Bezandry, P., Apprey, V. and Kwagyan, J. (2016) A Correlated Binary Model for Ignorable Missing Data: Application to Rheumatoid Arthritis Clinical Data. *Journal of Data Science*. Vol. 14.
- Fitzmaurice, G.M. and Laird, N.M. (1993). A Likelihood-based Method for Analyzing Longitudinal Binary Responses. *Biometrika*, Vol. 80, 141-151.
- Fitzmaurice, G.M., Laird, N.M., and Lipsitz, S.R. (1994). Analyzing Incomplete Binary Responses: A likelihood-based Approach. *Biometrics*, Vol. 50, 601-612.
- Kurland, B.F. (2002). *Analysis of Binary Longitudinal Data with Dropout and Death*. Ph.D. dissertation, University of Washington, Seattle, WA.
- Kurland, B.F. and Heagerty, P.J. (2004). Marginalized Transition Models for Longitudinal Binary Data with Ignorable and Non-ignorable Dropout. *Statistics in Medicine*, Vol. 23, 2673-2695.
- Kwagyan, J. (2001). *Further Investigations of the Disposition Model for Correlated Binary*

Outcomes. Ph.D. dissertation, The Temple University Graduate School, USA.

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*, Vol. 1, 31-48.

Little, R.J.A., and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. 2 Ed. John Wiley and Sons, New York, USA.

Marder, S.R., and Meibach, R.C.(1994). Risperidone in the Treatment of Schizophrenia. *American Journal of Psychiatry*, Vol. 151, 825-35.

Molenberghs, G., Fitzmaurice, G., Kenward, M.G., Tsiatis, A., and Verbeke, G. (2015). *Handbook of Missing Data Methodology*. Chapman and Hall/CRC, New York.

Rubin, R.D. (1976). Inference and Missing Data. *Biometrika*, Vol. 63, 581-592.

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*, Vol. 33, 3-20.

Table 2: Parameter estimates and standard error of the sensitivity analysis I for DNAR model with γ_0 fixed.

| Parameters | $\Phi_2=-1.5$ Est. (Std. error) | $\Phi_2=-1.2$ Est. (Std. error) | $\Phi_2=-1.0$ Est. (Std. error) | $\Phi_2=-0.5$ Est. (Std. error) | $\Phi_2=0$ Est. (Std. error) | $\Phi_2=0.5$ Est. (Std. error) | $\Phi_2=0.8$ Est. (Std. error) | $\Phi_2=1.0$ Est. (Std. error) | $\Phi_2=1.5$ Est. (Std. error) |
|-------------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|---------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
| λ_0 | -0.9583 (0.1324)* | -1.3367(0.1025)* | -1.4694(0.1496)* | -1.5465(0.1194)* | -1.5482 (0.1093)* | -1.5359(0.1031)* | -1.4721(0.1065)* | -1.5212(0.1006)* | -19.83(352.05) |
| HALO (β_1) | 0.1245 (0.1102) | 0.5924 (0.3621) | 0.2107 (0.2016) | 0.1229 (0.2024) | 0.4439 (0.3428) | 0.1792 (0.1419) | 0.7819 (0.5210) | 0.7479 (0.8156) | 0.7829 (0.6419) |
| RISP(L) (β_2) | 0.3194 (0.1915)* | 0.8331(0.2636)* | 1.0448(0.1970)* | 1.1485(0.1516)* | -1.1519 (0.1369)* | 1.1451 (0.1281)* | 1.0513(0.1571)* | 1.1193(0.1247)* | 0.7622(0.1209)* |
| RISP(H) (β_3) | 0.3517 (0.1833)* | 0.8429(0.2502)* | 1.0118(0.1925)* | 1.1081(0.1520)* | 1.1091 (0.1387)* | 1.0973 (0.1307)* | 1.0163 (0.1371)* | 1.0733(0.1272)* | 0.7349(0.1232)* |
| $\gamma_{k-1} (\phi_1)$ | -0.2675 (0.1278)* | -0.1514(0.0413)* | -0.0766(0.0297)* | 0.05115(0.0163)* | 0.1678 (-) | 0.2196 (-) | 1.1550 (-) | 0.4585 (0.2147)* | 0.2493 (0.0231)* |
| η | 0.0748 (0.0386) | 0.1481(0.0443)* | 0.1696(0.0304)* | 0.1736(0.0066)* | 0.1678 (-) | 0.1120 (-) | 0.4944 (-) | 0.1686 (0.1014) | 2.3357 (39.117) |
| log likelihood | -1212.14 | -1208 | -1203.3 | -1191.5 | -1182.4 | -1275.2 | -1274.6 | -1171.5 | -1178.4 |
| -2 loglkd | 2424.3 | 2416 | 2406.6 | 2383 | 2364.8 | 2550.4 | 2549.2 | 2343 | 2348.8 |
| AIC | 2444.3 | 2436 | 2426.4 | 2403 | 2384.8 | 2570.4 | 2569.2 | 2363 | 2368.8 |

Note: * means significant and γ_{k-1} is schizophrenia status at previous time point

Table 3: Parameter estimates and standard error of the sensitivity analysis II for DNAR model with γ_0

| Parameters | $\phi_2=-1.0$ Est. (Std. error) | $\phi_2=-0.5$ Est. (Std. error) | $\phi_2=0.5$ Est. (Std. error) | $\phi_2=0.8$ Est. (Std. error) | $\phi_2=1.0$ Est. (Std. error) | $\phi_2=1.5$ Est. (Std. error) |
|------------------------------|------------------------------------|------------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
| λ_0 | -1.4463 (0.2669)* | -1.5188 (0.1319)* | -1.4801 (0.1183)* | -1.4661(0.1135)* | -2.7455 (0.1400)* | -1.3590 (0.0509)* |
| γ_0 | 0.6934 (0.1121)* | 0.6681 (0.0715)* | 0.6811 (0.07015)* | 0.0696 (0.0145)* | -0.2615 (0.0014)* | -0.9417 (0.0897)* |
| HALO (β_1) | 0.1831 (0.1324) | 1.0017(0.8025) | 1.4463(0.8669) | 0.5188(0.3319) | 0.5482 (0.3109) | 0.4801(0.2818) |
| RISP(L) (β_2) | 1.2403 (0.4528)* | 1.3500 (0.1965)* | 1.2776 (0.1733)* | 1.2525 (0.1643)* | 0.8325 (0.1191)* | 0.9043 (0.0547)* |
| RISP(H) (β_3) | 1.0067 (0.4528)* | 0.1139 (0.2050) | 1.0823 (0.1784)* | 1.0605 (0.1689)* | 0.7959 (0.1216)* | 0.8618 (0.0623)* |
| $\gamma_{k-1} (\phi_1)$ | -0.0658 (0.0659) | 0.0628 (0.0201)* | 0.5136 (0.2025)* | 0.4883 (0.2301)* | 0.2090 (0.0662)* | -2.3966 (-) |
| Markov dependence (η) | 0.1792 (0.0616)* | 0.1846 (0.0111)* | 0.3103 (0.1201)* | 0.2270 (0.1204) | 4.7514 (238.6) | -0.7300 (-) |
| log likelihood. value | -1176.2 | -1164.2 | -1148.2 | -1145.9 | -1198.2 | -1213.7 |
| -2log likelihood. | 2352.4 | 2328.4 | 2296.4 | 2291.8 | 2396.4 | 2427.4 |
| AIC | 2374.4 | 2350.4 | 2318.4 | 2313.8 | 2418.4 | 2449.4 |

Note: * means significant and γ_{k-1} is schizophrenia status at previous time point.