

# Modeling repeated binary data with nonignorable missingness

Cherry T. Nepomuceno<sup>1</sup>

## ABSTRACT

This paper presents a model for repeated binary response variables with nonignorable missing data mechanism. Marginal expectations of responses are related to a set of time-independent and time-dependent covariates by a logit link function and a logit model was used for the missing data mechanism. When ignorability was assumed, the model simply followed a joint distribution of responses that satisfies conditional independence. Considering normal priors, a procedure of implementing Metropolis-Hastings algorithm on a simple case is illustrated.

**KEYWORDS:** repeated measures, binary data, logistic regression, nonignorable missing data mechanism, Metropolis-Hastings algorithm

## 1. INTRODUCTION

The analysis of longitudinal data or repeated measures data is commonly confronted with problems on missing observations. Knowledge, or absence of knowledge, of the mechanisms that led to certain values being missing is a key element in choosing an appropriate analysis and in interpreting the results. In repeated measures design, incomplete data often arise as a result of drop-out in which sequences of measurements on some units terminate prematurely. Kenward and Diggle (1994) cited, in their study, classification of dropout process by Little and Rubin (1987) as completely random dropout (CRD) wherein the dropout and measurement processes are independent; random dropout (RD), the dropout process depends on the observed measurements, i.e., those preceding dropout; and informative dropout (ID), the dropout process depends on the unobserved measurements, i.e., those that would have been observed if the unit had not dropped out.

The dropout process can affect the inferences about the measurement process. It can be ignored in the RD case provided that the required inferences concern the measurement process for notionally complete sequences, and those inferences are likelihood based (Diggle and Kenward, 1994). Completely random dropout can be dealt with by using standard methods of analyses. In fact, analyses restricted to subjects with complete data yield valid estimates but there may be substantial loss of efficiency in discarding all the information on individuals with incomplete data (Fitzmaurice et al., 1996b). Dropout that is informative is said to be nonignorable since the dropout mechanism cannot be ignored when estimating parameters for the data. With nonignorable mechanism, inferences for the data parameters generally depend on the posited missingness mechanism that implies increased sensitivity of inferences to reasonable model specifications. In Diggle and Kenward (1994), Rubin pointed out in his discussion of the paper that without distributional assumptions or supplemental information it is impossible to 'test' or find evidence for or against nonignorability (informative dropout).

There are some ways of handling missing values. The first approach is simply to drop cases with missing values on any variables and to undertake complete case analyses.

---

<sup>1</sup> Mathematics Department, West Visayas State University, Iloilo City

This may lead to bias if values are missing in a systematic way and generally does not make fully efficient use of all the data. The second approach is to fill in the missing values with imputed values. This may reduce the bias of the complete case analyses but may lead to additional bias in multivariate analyses if the imputation method fails to control for all multivariate relationships. The third approach is to use model based adjustment methods that will generally be consistent and fully efficient provided that the assumed model holds (Skinner and Coker, 1996). These procedures assume ignorable missing data mechanism, which means that the likelihood can be factored into a part that contains the parameters of interest and a part that contains parameters modeling the missing data mechanism. Missing at random (MAR) and missing completely at random (MCAR) are ignorability conditions which guarantee that some inferences may be made without recourse to complicated missing-data modeling.

When missing data mechanism is nonignorable, it is not possible to factor from the likelihood a part with only parameters for the missing data mechanism (Little and Rubin, 1987). Nonignorability may be classified as: models where the missing data mechanism,  $R$  given  $Y = (Y_{\text{obs}}, Y_{\text{mis}})$  depends on  $Y_{\text{mis}}$  but does not contain unknown parameter  $\psi$  associated with  $R$ ; and models where the missing data mechanism is nonignorable and unknown with lack of knowledge reflected on the unknown parameter  $\psi$ . The term ignorable indicate that it is not necessary to specify a model for dropout in a likelihood-based analysis of the measurement process i.e. the causes of dropout can be ignored since maximum likelihood estimates of the parameters for the process under investigation are the same whether or not they are jointly estimated with the parameters for the dropout process (Fitzmaurice, et al., 1996).

Some of the studies that involved nonignorable missing data models considered continuous longitudinal response data assumed to be multivariate normal with logistic model for the dropout mechanism and used ML to estimate parameters (Diggle and Kenward, 1994); while for a similar case, Troxel, et al. (1998) formed pseudolikelihood for its estimation procedure. For binary longitudinal data subject to informative dropout Fitzmaurice, et al., (1996a) and Fitzmaurice et al. (1995) used likelihood-based analysis; while Zeger, et al. (1985) used working likelihoods to approximate the actual likelihood that lead to consistent estimates under weak assumptions. Pulkstenis, et al. (1998) addressed the problem of accounting for informative dropout in the form of rescue medication when comparing pain relievers with respect to longitudinal binary pain relief outcomes. Follmann and Wu (1995) assumed separate models for the primary response and the missingness and are linked by common random parameters. Resultant approximation is a mixed generalized linear model with possibly heterogeneous random effects. Likewise, Liang and Zeger (1986) extended GLM to the analysis of longitudinal data by introducing a class of estimating equations that give consistent estimates of the regression parameters and of their variance under mild assumptions about the time dependence.

This study is motivated by a problem of determining disease state of an individual based on some diagnostic procedure with the presence of covariates. Some illnesses are diagnosed after results from a series of tests and information on some symptoms are available. For example, diabetes can be properly diagnosed after a series of sugar test results are available such as fasting blood sugar, post prandial sugar, and glucose tolerance tests. Dengue can be diagnosed after a series of serologic tests.

Consider a study on tuberculosis (TB) prevalence. It is a two-stage study that used chest X-ray on the first stage and sputum tests on the second stage. The chest X-ray initially identified individuals from randomly chosen clusters who are TB positive and those who are not. TB positive individuals identified by such procedure were included in the second stage and were subjected to a series of sputum tests. Being positive in one of the sputum tests finally classified an individual as TB positive. Since we are dealing with repeated measures, the problem of missing observations often arises. Covariates such as sex, age, employment status, family size, coughing frequency, presence of profuse sweating at night, presence of TB patient within the household, and other physical and environmental characteristics are considered.

Suppose that  $N$  individuals are included in the first stage and there are  $n$  individuals identified by chest X-ray to be TB positive (i.e.  $(N-n)$  individuals are TB negative hence are not included in the second stage). Out of  $n$  individuals eligible for sputum testing, only  $m_t$ ,  $m_t \leq m_{t-1} \leq n$ , submitted themselves for the  $t$ th sputum test,  $t = 1, 2, \dots, T$ ;  $m_0 = n$ . It is assumed that the covariates are completely observed.

Let  $X_{i0}$  be the disease state at the first stage (chest X-ray result),  $i=1, \dots, N$ ;  $X_{it}$  be the disease state of the  $i$ th individual at the  $t$ th sputum test,  $i = 1, \dots, m_t$ ;  $t = 1, \dots, T$ ; i.e.

$$X_{it} = \begin{cases} 1 & \text{if individual } i \text{ is positive at the } t\text{th sputum test} \\ 0 & \text{otherwise.} \end{cases}$$

and  $Z_i(t)$  be a  $q_t$  vector of covariate associated with the  $t$ th sputum test on the  $i$ th individual,  $i = 1, \dots, n$ ;  $t = 1, \dots, T$ .

For some individuals, some  $X_i$ 's are missing. Data can be arranged so that it follows a monotone pattern. The following cases for covariates to be considered are: (a) all covariates are time-independent; and (b) combination of time-independent and time-dependent covariates.

The objective of this paper is to present a model for repeated binary response variables with nonignorable missing data mechanism. For estimation procedures, Metropolis-Hastings algorithm is used. This paper is organized as follows: Section 2 presents a model for repeated binary data with nonignorable missingness relating its expectations to time-dependent and time-independent covariates using logit link function; Section 3 presents the Metropolis-Hastings algorithm to estimate parameters for a simple case of univariate response; and the summary and discussions are presented in Section 4.

## 2. REPEATED BINARY DATA WITH NONIGNORABLE MISSING DATA

Suppose  $N$  individuals were observed in the first stage. Using a certain evaluation procedure to determine disease state of each individual, let

$$X_{i0} = \begin{cases} 1 & \text{if } i\text{th individual is positive of the disease of interest} \\ 0 & \text{otherwise} \end{cases} \quad \text{for } i = 1, \dots, N.$$

Individuals diagnosed as positive in the first stage will be included in the second stage, say, there are  $n$  of them. They will be subjected to another series of tests to determine their final disease state.

Suppose also that the second stage of observations involves  $T$  repeated measurements on each identified subject. At this point, it is convenient to introduce some notations that will be used throughout this paper. For  $i = 1, \dots, n$ , let

$\mathbf{X}_i' = (X_{i1}, X_{i2}, \dots, X_{iT})$  be a  $T$  vector of binary random variables on the  $i$ th subject possibly incompletely observed;

$\mathbf{Z}_i' = (\mathbf{Z}_i(1)', \mathbf{Z}_i(2)', \dots, \mathbf{Z}_i(T)')$  be a  $q$  vector of covariates,  $q = \sum_{t=1}^T q_t$ , where  $\mathbf{Z}_i(t)'$  is  $q_t$  vector of covariates associated with subject  $i$ , obtained at the  $t^{\text{th}}$  occasion;

$\beta_i' = (\beta_i(1)', \beta_i(2)', \dots, \beta_i(T)')$  be  $q$  vector of covariate effects on the response;

$\mathbf{R}_i' = (R_{i1}, R_{i2}, \dots, R_{iT})$  be a vector of  $T$  missing data indicator i.e.,  $R_{it}$  is 1 if  $X_{it}$  is observed and  $R_{it}$  is 0 if  $X_{it}$  is missing,  $t = 1, \dots, T$ ;

$\gamma_i' = (\gamma_i(1)', \gamma_i(2)', \dots, \gamma_i(T)')$  be a  $q$  vector of covariate effects on the missing data mechanism;

$\alpha$  be a vector of  $T$  unknown parameters relating response variables to the missing data mechanism;

$\delta$  be a vector of  ${}_T C_2$  parameters relating the missing data mechanism with its past values where  ${}_T C_2$  indicates a combination of  $T$  taken 2 at a time;

$\lambda$  be a vector of  ${}_T C_2$  parameters relating the missing data mechanism with the past response values;

$\rho$  be a  $2^T - T - 1$  unknown vector of bivariate and higher order relations of the response variables;

$\mathbf{Y} = (Y_i)$  be the final disease state of the individual on the second stage i.e.,

$$Y_i = \begin{cases} 1 & \text{if at least one } X_{it} = 1, t = 1, \dots, T; \\ 0 & \text{o.w.} \end{cases}$$

Assume that observations on distinct individuals are independent, i.e. given  $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n)'$ ,  $(\mathbf{X}_i, \mathbf{R}_i, \beta_i)$  and  $(\mathbf{X}_j, \mathbf{R}_j, \beta_j)$  are independent for  $i \neq j$ . Consider monotone missing data patterns, i.e.  $P[R_{it} = 1 | R_{is} = 0, \forall s < t] = 0$ .

We can determine the prevalence rate

$$E(Y_i) = P(Y_i = 1) = P\left[\bigcup_{t=1}^T \{X_{it} = 1\}\right].$$

Considering the covariates, we are particularly interested in the expression

$$E(Y | \mathbf{X}, \mathbf{Z}, \mathbf{R}) = P(Y = 1 | \mathbf{X}, \mathbf{Z}, \mathbf{R}) = P\left[\bigcup_{t=1}^T \{(X_{it} = 1) | \mathbf{Z}, \mathbf{R}\}\right] \quad (1)$$

Now for  $t = 1$ , and for  $i = 1, \dots, n$ , let  $X_{i1} \sim \text{Ber}(\mu_1)$  so that  $E(X_{i1}) = P(X_{i1} = 1) = \mu_1$  and  $\text{Var}(X_{i1}) = \mu_1(1 - \mu_1)$ . From the probability mass function of  $X_{i1}$ ,  $i = 1, \dots, n$ ,

$$p(x_{i1}) = \mu_1^{x_{i1}} (1 - \mu_1)^{1-x_{i1}} \quad (2)$$

then assuming a logistic link

$$\theta_{i1} = \ln [\mu_1 / (1 - \mu_1)] = \beta(1)' Z_i(1) \quad (3)$$

we then have

$$P[X_{i1} = x_{i1} | Z_i(1), \beta(1)] = \exp \{ x_{i1} \beta(1)' Z_i(1) - \ln (1 + \exp (\beta(1)' Z_i(1))) \} \quad (4)$$

which implies that  $X_i | Z(1) \sim \text{Ber} \left\{ \frac{\exp[\beta(1)' Z(1)]}{1 + \exp[\beta(1)' Z(1)]} \right\}$ .

Let  $R_{i1} \sim \text{Ber}(\psi_1)$ ,  $i = 1, \dots, n$  Assuming logistic regression,

$$\pi_{i1} = \ln [\psi_1 / (1 - \psi_1)] = \phi(1)' W_i(1) \quad (5)$$

where  $W_i(1) = (X_{i1}, Z_i(1))'$ ,  $(1+q_1) \times 1$  vector of the first response and the covariates at the first occasion; and  $\phi(1) = (\alpha_1, \gamma(1))'$ ,  $(1+q_1) \times 1$  vector of unknown parameters associated with the variables  $(X_{i1}, Z_i(1))'$ .

Let  $W_i = (W_i(1)', W_i(2)', \dots, W_i(T)') = (X_i, Z_i)$  be a  $1 \times (T+q)$  vector of responses and covariates, where  $q = \sum_{t=1}^T q_t$ .

Clearly,

$$P[R_{i1} = r_{i1} | W_i(1), \phi(1)] = \exp \{ r_{i1} \phi(1)' W_i(1) - \ln [1 + \exp (\phi(1)' W_i(1))] \}$$

so that

$$P[R_{i1} = r_{i1} | X_{i1}, Z_i(1), \alpha_1, \gamma(1)] = \exp \{ r_{i1} \alpha_1 X_{i1} + r_{i1} \gamma(1)' Z_i(1) - \ln [1 + \exp (\alpha_1 X_{i1} + \gamma(1)' Z_i(1))] \} \quad (6)$$

Now,

$$\begin{aligned} P \left[ \bigcup_{t=1}^T \{ X_t = 1 | Z, R, \beta, \alpha, \gamma, \delta, \lambda, \rho \} \right] &= \sum_{t=1}^T P[X_t = 1 | Z, R, \beta, \alpha, \gamma] \\ &\quad - \sum_{t>s} P[X_t = 1, X_s = 1 | Z, R, \beta, \alpha, \gamma, \rho_{st}] \\ &\quad + \sum_{t>s>r} P[X_t = 1, X_s = 1, X_r = 1 | Z, R, \beta, \alpha, \gamma, \delta, \lambda, \rho] - \dots \\ &\quad + (-1)^{T-1} P[X_1 = 1, X_2 = 1, \dots, X_T = 1 | Z, R, \beta, \alpha, \gamma, \delta, \lambda, \rho] \end{aligned} \quad (7)$$

Note that

$$P[X | Z, R, \beta, \alpha, \gamma] = P[X, R | Z, \beta, \alpha, \gamma] / \sum_x P[X, R | Z, \beta, \alpha, \gamma] \quad (8)$$

One way of formulating nonignorable missing data models is to write the joint distribution of  $X$  and  $R$  in the form suggested by Little and Rubin (1987)

$$P[X, R | Z, \beta, \alpha, \gamma] = P[X | Z, \beta] P[R | X, Z, \alpha, \gamma] \quad (9)$$

From (4) and (6)

$$P\{X = x, R = r | Z, \beta, \alpha, \gamma\} = P\{X = x | Z, \beta\} P\{R = r | X = x, Z, \alpha, \gamma\} \\ = \exp \{ x \beta' Z - \ln(1 + \exp(\beta' Z)) + r \alpha x + r \gamma' Z - \ln(1 + \exp(\alpha x + \gamma' Z)) \} \quad (10)$$

From (8) and (10),

$$P\{X = x | Z, R = r, \beta, \alpha, \gamma\} = \frac{\exp \{ x (\alpha r + \beta' Z) + \ln(1 + \exp(\gamma' Z)) - \ln(1 + \exp(\alpha x + \gamma' Z)) \}}{1 + \exp \{ (\alpha r + \beta' Z) + \ln(1 + \exp(\gamma' Z)) - \ln(1 + \exp(\alpha + \gamma' Z)) \}} \quad (11)$$

Note that

$$\sum_{(x=0,1)} P\{X = x | Z, R = r, \beta, \alpha, \gamma\} = 1 \quad (12)$$

$$E\{X | R = r, Z, \beta, \alpha, \gamma\} = P\{X = 1 | R = r, Z, \beta, \alpha, \gamma\} \\ = \frac{\exp \{ \alpha r + \beta' Z + \ln(1 + \exp(\gamma' Z)) - \ln(1 + \exp(\alpha + \gamma' Z)) \}}{1 + \exp \{ \alpha r + \beta' Z + \ln(1 + \exp(\gamma' Z)) - \ln(1 + \exp(\alpha + \gamma' Z)) \}} \quad (13)$$

$$E\{X^2 | R = r, Z, \beta, \alpha, \gamma\} = 1^2 P\{X = 1 | R = r, Z, \beta, \alpha, \gamma\} + 0^2 P\{X = 0 | R = r, Z, \beta, \alpha, \gamma\} \\ = E\{X | R = r, Z, \beta, \alpha, \gamma\} \quad (14)$$

Hence

$$\text{Var}\{X | R = r, Z, \beta, \alpha, \gamma\} = E\{X | R = r, Z, \beta, \alpha, \gamma\} \{1 - E\{X | R = r, Z, \beta, \alpha, \gamma\}\} \\ = \frac{\exp \{ \alpha r + \beta' Z + \ln(1 + \exp(\gamma' Z)) - \ln(1 + \exp(\alpha + \gamma' Z)) \}}{[1 + \exp \{ \alpha r + \beta' Z + \ln(1 + \exp(\gamma' Z)) - \ln(1 + \exp(\alpha + \gamma' Z)) \}]^2} \quad (15)$$

When  $\alpha = 0$ , i.e. the missing data mechanism does not depend on the response variable (MCAR condition) then  $P\{X = x | Z, R = r, \beta, \alpha = 0, \gamma\} = \exp [x \beta' Z] / \{1 + \exp [\beta' Z]\} = P\{X = x | Z, \beta\}$ . Consequently,  $X \sim \text{Ber} \{ \exp [\beta' Z] / \{1 + \exp [\beta' Z]\} \}$

To extend the above procedure in determining the bivariate joint conditional distribution of  $(X_1, X_2)$  given  $[Z(1), Z(2), \beta(1), \beta(2), R = (r_1, r_2), \alpha_1, \alpha_2, \gamma(1), \gamma(2), \rho_{12}, \delta_{12}, \lambda_{12}]$ , we have

$$P\{X_1, X_2 | Z_2 = (Z(1), Z(2)), R^{(2)} = (R_1, R_2), \beta_2 = (\beta(1), \beta(2)), \alpha^{(2)} = (\alpha_1, \alpha_2), \gamma_2 = (\gamma(1), \gamma(2)), \rho_{12}, \delta_{12}, \lambda_{12}\} \\ = \frac{P\{X_1, X_2, R_1, R_2 | Z_2 = (Z(1), Z(2)), \beta_2, \alpha^{(2)}, \gamma_2, \rho_{12}, \delta_{12}, \lambda_{12}\}}{\sum_{(X_1, X_2)} P\{X_1, X_2, R_1, R_2 | Z_2 = (Z(1), Z(2)), \beta_2, \alpha^{(2)}, \gamma_2, \rho_{12}, \delta_{12}, \lambda_{12}\}} \quad (16)$$

$$\begin{aligned} P[X_1, X_2, R_1, R_2 | Z_2 = (Z(1), Z(2)), \beta_2, \alpha^{(2)}, \gamma_2, \rho_{12}, \delta_{12}, \lambda_{12}] \\ = P[X_1, X_2 | Z_2, \beta_2, \rho_{12}] P[R_1, R_2 | X_1, X_2, Z_2, \alpha^{(2)}, \gamma_2, \delta_{12}, \lambda_{12}] \end{aligned} \quad (17)$$

$$\begin{aligned} P[X_1, X_2 | Z_2 = (Z(1), Z(2)), \beta_2 = (\beta(1), \beta(2)), \rho_{12}] \\ = P[X_2 | X_1, Z(2), \beta(2), \rho_{12}] P[X_1 | Z(1), \beta(1)] \end{aligned} \quad (18)$$

Assuming logistic regression

$$\begin{aligned} P[X_2 = x_2 | X_1 = x_1, Z(2), \beta(2), \rho_{12}] \\ = \exp \{x_2 [\beta(2)' Z(2) + x_1 \rho_{12}] - \ln [1 + \exp(\beta(2)' Z(2) + x_1 \rho_{12})]\} \end{aligned} \quad (19)$$

From (18),

$$\begin{aligned} P[X_1 = x_1, X_2 = x_2 | Z_2 = (Z(1), Z(2)), \beta_2 = (\beta(1), \beta(2)), \rho_{12}] \\ = \exp \{x_2 [\beta(2)' Z(2) + x_1 \rho_{12}] - \ln [1 + \exp(\beta(2)' Z(2) + x_1 \rho_{12})]\} \exp \{x_1 [\beta(1)' Z(1)] - \\ \ln [1 + \exp(\beta(1)' Z(1))]\} \\ = \exp \{x_1 \beta(1)' Z(1) + x_2 \beta(2)' Z(2) + x_1 x_2 \rho_{12} - \ln [1 + \exp(\beta(1)' Z(1))] - \\ \ln [1 + \exp(\beta(2)' Z(2) + x_1 \rho_{12})]\} \end{aligned} \quad (20)$$

Assume that nonresponse is independent of the future values but is dependent on the present and past values. Hence, we model the probability distribution as

$$\begin{aligned} P[R_1 = r_1, R_2 = r_2 | X_1 = x_1, X_2 = x_2, Z_2, \alpha^{(2)}, \gamma_2, \delta_{12}, \lambda_{12}] \\ = \exp \{r_1 [x_1 \alpha_1 + \gamma(1)' Z(1)] + r_2 [x_2 \alpha_2 + \gamma(2)' Z(2) + r_1 \delta_{12} + x_1 \lambda_{12}] - \\ \ln [1 + \exp(x_1 \alpha_1 + \gamma(1)' Z(1))] - \\ \ln [1 + \exp [x_2 \alpha_2 + \gamma(2)' Z(2) + r_1 \delta_{12} + x_1 \lambda_{12}]]\} \end{aligned} \quad (21)$$

From (20) and (21), (17) becomes

$$\begin{aligned} P[X_1 = x_1, X_2 = x_2, R_1 = r_1, R_2 = r_2 | Z_2, \beta_2, \alpha^{(2)}, \gamma_2, \rho_{12}, \delta_{12}, \lambda_{12}] \\ = \exp \{x_1 [\beta(1)' Z(1) + r_1 \alpha_1 + r_2 \lambda_{12}] + x_2 [\beta(2)' Z(2) + r_2 \alpha_2] + \\ x_1 x_2 \rho_{12} + r_1 \gamma(1)' Z(1) + r_2 \gamma(2)' Z(2) + \\ r_1 r_2 \delta_{12} - \ln [1 + \exp(\beta(1)' Z(1))] - \ln [1 + \exp(\beta(2)' Z(2) + x_1 \rho_{12})] - \\ \ln [1 + \exp(x_1 \alpha_1 + \gamma(1)' Z(1))] - \ln [1 + \exp(\alpha_2 x_2 + \gamma(2)' Z(2) + \\ r_1 \delta_{12} + x_1 \lambda_{12})]\} \end{aligned} \quad (22)$$

Taking the sum of (22) over all possible values of  $(X_1, X_2)$  then (16) becomes

$$P[X_1 = x_1, X_2 = x_2 | Z(1), Z(2), \beta(1), \beta(2), R_1 = r_1, R_2 = r_2, \alpha_1, \alpha_2, \gamma(1), \gamma(2), \rho_{12}, \delta_{12}, \lambda_{12}] = \frac{\exp\{x_1\beta(1)'Z(1) + x_2\beta(2)'Z(2) + x_1x_2\rho_{12} + x_1r_1\alpha_1 + r_2\alpha_2x_2 + r_2x_1\lambda_{12}\}}{K_2 [1 + \exp(\beta(2)'Z(2) + x_1\rho_{12})][1 + \exp(\alpha_1x_1 + \gamma(1)'Z(1))][1 + \exp(\alpha_2x_2 + \gamma(2)'Z(2) + r_1\delta_{12} + x_1\lambda_{12})]} \quad (23)$$

where

$$K_2 = [1 + \exp(\beta(2)'Z(2))]^{-1} [1 + \exp(\gamma(1)'Z(1))]^{-1} \left\{ \frac{1}{[1 + \exp(\gamma(2)'Z(2) + r_1\delta_{12})] + \frac{\exp\{\beta(2)'Z(2) + r_2\alpha_2\}}{[1 + \exp(\alpha_2 + \gamma(2)'Z(2) + r_1\delta_{12})]}} \right\} + \frac{\exp\{\beta(1)'Z(1) + r_1\alpha_1 + r_2\lambda_{12}\}}{[1 + \exp(\beta(2)'Z(2) + \rho_{12})][1 + \exp(\alpha_1 + \gamma(1)'Z(1))]} \left\{ \frac{1}{[1 + \exp(\lambda_{12} + \gamma(2)'Z(2) + r_1\delta_{12})] + \frac{\exp\{\beta(2)'Z(2) + r_2\alpha_2 + \rho_{12}\}}{[1 + \exp(\alpha_2 + \lambda_{12} + \gamma(2)'Z(2) + r_1\delta_{12})]}} \right\}$$

Now, evaluate

$$E[X_1, X_2 | Z(1), Z(2), \beta(1), \beta(2), R_1 = r_1, R_2 = r_2, \alpha_1, \alpha_2, \gamma(1), \gamma(2), \rho_{12}, \delta_{12}, \lambda_{12}] = E[X_1, X_2 | R^{(2)}, Z_2, \Omega_2].$$

It can be shown that

$$E[X_1, X_2 | R^{(2)}, Z_2, \Omega_2] = \sum_{(x_1, x_2)} (x_1, x_2) P[X_1 = x_1, X_2 = x_2 | R^{(2)}, Z_2, \Omega_2] = (\mu_{12}, \mu_{22})$$

where

$$\mu_{12} = \sum_{x_2=\{0,1\}} P[X_1 = 1, X_2 | R^{(2)}, Z_2, \Omega_2] = \left( \frac{K_2^{-1} \exp\{\beta(1)'Z(1) + r_1\alpha_1 + r_2\lambda_{12}\}}{[1 + \exp(\beta(2)'Z(2) + \rho_{12})][1 + \exp(\alpha_1 + \gamma(1)'Z(1))]} \left\{ \frac{1}{[1 + \exp(\lambda_{12} + \gamma(2)'Z(2) + r_1\delta_{12})] + \frac{\exp\{\beta(2)'Z(2) + r_2\alpha_2 + \rho_{12}\}}{[1 + \exp(\alpha_2 + \lambda_{12} + \gamma(2)'Z(2) + r_1\delta_{12})]}} \right\} \right)$$

and

$$\begin{aligned} \mu_{22} &= \sum_{x_1=\{0,1\}} P[X_1, X_2 = 1 | R^{(2)}, Z_2, \Omega_2] \\ &= K_2^{-1} \exp\{\beta(2)'Z(2) + \alpha_2r_2\} \\ &\quad \left\{ \frac{1}{[1 + \exp(\beta(2)'Z(2))][1 + \exp(\gamma(1)'Z(1))][1 + \exp(\alpha_2 + \gamma(2)'Z(2) + r_1\delta_{12})]} + \frac{\exp\{\beta(1)'Z(1) + r_1\alpha_1 + r_2\lambda_{12} + \rho_{12}\}}{[1 + \exp(\beta(2)'Z(2) + \rho_{12})][1 + \exp(\gamma(1)'Z(1) + \alpha_1)][1 + \exp(\alpha_2 + \gamma(2)'Z(2) + r_1\delta_{12} + \lambda_{12})]} \right\} \end{aligned}$$



Also

$$\begin{aligned} & E[X_1, X_2 | \mathbf{R}^{(2)}, \mathbf{Z}_2, \Omega_2] [X_1, X_2 | \mathbf{R}^{(2)}, \mathbf{Z}_2, \Omega_2] \\ &= \sum_{(x_1, x_2)} \begin{bmatrix} x_1^2 & x_1 x_2 \\ x_1 x_2 & x_2^2 \end{bmatrix} P[X_1 = x_1, X_2 = x_2 | \mathbf{R}^{(2)}, \mathbf{Z}_2, \Omega_2] \\ &= \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix} \end{aligned}$$

where

$$\psi_{11} = \mu_{12}$$

$$\psi_{22} = \mu_{22}$$

$$\begin{aligned} \psi_{12} = \psi_{21} &= P[X_1 = 1, X_2 = 1 | \mathbf{R}^{(2)}, \mathbf{Z}_2, \Omega_2] \\ &= \frac{\exp\{\beta(1)' \mathbf{Z}(1) + \beta(2)' \mathbf{Z}(2) + \rho_{12} + r_1 \alpha_1 + r_2 \alpha_2 + r_2 \lambda_{12}\}}{K_2 [1 + \exp\{\beta(2)' \mathbf{Z}(2) + \rho_{12}\}] [1 + \exp\{\alpha_1 + \gamma(1)' \mathbf{Z}(1)\}] [1 + \exp\{\alpha_2 + \gamma(2)' \mathbf{Z}(2) + r_1 \delta_{12} + \lambda_{12}\}]} \end{aligned}$$

Hence

$$\text{Cov}[X_1, X_2 | \mathbf{R}^{(2)}, \mathbf{Z}_2, \Omega_2] = \begin{bmatrix} \mu_{12}(1 - \mu_{12}) & \psi_{12} - \mu_{12}\mu_{22} \\ \psi_{12} - \mu_{12}\mu_{22} & \mu_{22}(1 - \mu_{22}) \end{bmatrix}$$

When  $X_1$  and  $X_2$  are not related i.e.  $\rho_{12} = 0$  and the missing data mechanism is ignorable i.e.,  $\alpha_1 = \alpha_2 = \lambda_{12} = \delta_{12} = 0$  then (23) becomes

$$\begin{aligned} & P[X_1 = x_1, X_2 = x_2 | \mathbf{Z}(1), \mathbf{Z}(2), \beta(1), \beta(2), R_1 = r_1, R_2 = r_2, \alpha_1, \alpha_2, \gamma(1), \gamma(2), \rho_{12}, \delta_{12}, \lambda_{12}] = \\ &= \frac{\exp\{x_1 \beta(1)' \mathbf{Z}(1)\} \exp\{x_2 \beta(2)' \mathbf{Z}(2)\}}{[1 + \exp\{\beta(1)' \mathbf{Z}(1)\}] [1 + \exp\{\beta(2)' \mathbf{Z}(2)\}]} \\ &= P[X_1 = x_1 | \mathbf{Z}(1), \beta(1)] P[X_2 = x_2 | \mathbf{Z}(2), \beta(2)] \end{aligned} \quad (24)$$

When the missing data mechanism is ignorable i.e.  $\alpha_1 = \alpha_2 = \lambda_{12} = \delta_{12} = 0$ , equation (23) becomes

$$\begin{aligned} & P[X_1 = x_1, X_2 = x_2 | \mathbf{Z}(1), \mathbf{Z}(2), \beta(1), \beta(2), R_1 = r_1, R_2 = r_2, \alpha_1, \alpha_2, \gamma(1), \gamma(2), \rho_{12}, \delta_{12}, \lambda_{12}] = \\ &= \frac{\exp\{x_1 \beta(1)' \mathbf{Z}(1) + x_2 \beta(2)' \mathbf{Z}(2) + x_1 x_2 \rho_{12}\}}{[1 + \exp\{\beta(1)' \mathbf{Z}(1)\}] [1 + \exp\{\beta(2)' \mathbf{Z}(2) + x_1 \rho_{12}\}]} \\ &= P[X_1 = x_1 | \mathbf{Z}(1), \beta(1)] P[X_2 = x_2 | X_1 = x_1, \mathbf{Z}(2), \beta(2)] \end{aligned} \quad (25)$$

Using the same scheme, the joint distribution of  $\mathbf{X}^{(3)} = (X_1, X_2, X_3)$  given  $\mathbf{Z}_3, \beta_3, \alpha^{(3)}, \gamma_3, \rho^{(3)}, \delta^{(3)}, \lambda^{(3)}$  and  $\mathbf{R}^{(3)}$  can be obtained. The above procedure may likewise be extended to obtain

the joint distribution of the T dimensional repeated response measures  $\mathbf{X}^{(T)} = (X_1, \dots, X_T)$  given  $[\mathbf{Z}_T, \beta_T, \alpha^{(T)}, \gamma_T, \rho^{(T)}, \delta^{(T)}, \lambda^{(T)}, \mathbf{R}^{(T)}]$  where

$$\begin{aligned} \mathbf{Z}_T &= (\mathbf{Z}(1), \dots, \mathbf{Z}(T)), \\ \beta_T &= (\beta(1), \dots, \beta(T)), \\ \alpha^{(T)} &= (\alpha_1, \dots, \alpha_T), \\ \gamma_T &= (\gamma(1), \dots, \gamma(T)), \\ \rho^{(T)} &= (\rho_{12}, \rho_{13}, \dots, \rho_{12\dots T}), \\ \delta^{(T)} &= (\delta_{12}, \delta_{13}, \dots, \delta_{T-1,T}), \\ \lambda^{(T)} &= (\lambda_{12}, \lambda_{13}, \dots, \lambda_{T-1,T}), \\ \mathbf{R}^{(T)} &= (\mathbf{R}_1, \dots, \mathbf{R}_T), \end{aligned}$$

Also, let

$$\begin{aligned} \theta_t &= \beta(t)' \mathbf{Z}(t); \\ \pi_t &= \alpha_t x_t + \gamma(t)' \mathbf{Z}(t), \end{aligned}$$

$$\begin{aligned} \theta^{(t)} &= (\theta_1, \dots, \theta_t), \\ \pi^{(t)} &= (\pi_1, \dots, \pi_t) \text{ for } t = 1, \dots, T. \end{aligned}$$

$$\begin{aligned} P[\mathbf{X}^{(t)}, \mathbf{R}^{(t)} \mid \mathbf{Z}_t, \beta_t, \rho^{(t)}, \alpha^{(t)}, \gamma_t, \delta^{(t)}, \lambda^{(t)}] &= \\ P[\mathbf{X}^{(t)} \mid \mathbf{Z}_t, \beta_t, \rho^{(t)}] P[\mathbf{R}^{(t)} \mid \mathbf{X}^{(t)}, \mathbf{Z}_t, \alpha^{(t)}, \gamma_t, \delta^{(t)}, \lambda^{(t)}] & \quad (26) \end{aligned}$$

$$\begin{aligned} P[\mathbf{X}^{(t)} \mid \mathbf{Z}_t, \beta_t, \rho^{(t)}] &= \\ \frac{\exp\{\sum_{j=1}^t x_j \theta_j + U^{(t)}, \rho^{(t)}\}}{\prod_{j=1}^t \left( 1 + \exp\left( \theta_j + \sum_{l=1}^{j-1} x_l \rho_{lj} + \sum_{l=1}^{j-2} \sum_{m=l+1}^{j-1} x_l x_m \rho_{lmj} + \dots + x_1 \dots x_{j-1} \rho_{12\dots j} \right) \right)} & \quad (27) \end{aligned}$$

where  $\rho_{ij\dots l} = 0, \forall i \geq j \geq \dots \geq l, l = 2, \dots, t$ .

$$\begin{aligned} P[\mathbf{R}^{(t)} = \mathbf{r}^{(t)} \mid \mathbf{X}^{(t)} = \mathbf{x}^{(t)}, \mathbf{Z}_t, \alpha^{(t)}, \gamma_t, \delta^{(t)}, \lambda^{(t)}] &= \\ \frac{\exp\left[ \sum_{j=1}^t r_j \pi_j + \sum_{l=1}^{t-1} \sum_{m=l+1}^t (r_l r_m \delta_{lm} + x_l r_m \lambda_{lm}) \right]}{\prod_{j=1}^t \left( 1 + \exp\left( \pi_j + \sum_{l=1}^{j-1} r_l \delta_{lj} + \sum_{l=1}^{j-1} x_l \lambda_{lj} \right) \right)} & \quad (28) \end{aligned}$$

$$P[\mathbf{X}^{(t)} | \mathbf{Z}_t, \mathbf{R}^{(t)}, \boldsymbol{\beta}_t, \boldsymbol{\rho}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\gamma}_t, \boldsymbol{\delta}^{(t)}, \boldsymbol{\lambda}^{(t)}] = \frac{P[\mathbf{X}^{(t)}, \mathbf{R}^{(t)} | \mathbf{Z}_t, \boldsymbol{\beta}_t, \boldsymbol{\rho}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\gamma}_t, \boldsymbol{\delta}^{(t)}, \boldsymbol{\lambda}^{(t)}]}{\sum_{\mathbf{X}^{(t)}} P[\mathbf{X}^{(t)}, \mathbf{R}^{(t)} | \mathbf{Z}_t, \boldsymbol{\beta}_t, \boldsymbol{\rho}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\gamma}_t, \boldsymbol{\delta}^{(t)}, \boldsymbol{\lambda}^{(t)}]} \quad (29)$$

$$P[\mathbf{X}^{(t)}, \mathbf{R}^{(t)} | \mathbf{Z}_t, \boldsymbol{\beta}_t, \boldsymbol{\rho}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\gamma}_t, \boldsymbol{\delta}^{(t)}, \boldsymbol{\lambda}^{(t)}] = \frac{\exp\left[\sum_{j=1}^t (x_j \theta_j + r_j \tau_j) + \mathbf{U}^{(t)'} \boldsymbol{\rho}^{(t)} + \sum_{l=1}^{t-1} \sum_{m=l+1}^t (r_l r_m \delta_{lm} + x_l r_m \lambda_{lm})\right]}{\prod_{j=1}^t \left(1 + \exp\left(\theta_j + \sum_{l=1}^{j-1} x_l \rho_{lj} + \sum_{l=1}^{j-2} \sum_{m=l+1}^{j-1} x_l x_m \rho_{lmj} + \dots + x_1 \dots x_{j-1} \rho_{12..j}\right)\right) \left(1 + \exp\left(\pi_j + \sum_{l=1}^{j-1} r_l \delta_{lj} + \sum_{l=1}^{j-1} x_l \lambda_{lj}\right)\right)} \quad (30)$$

$$= \frac{\exp\left[\sum_{j=1}^t x_j (\theta_j + \alpha_j \tau_j) + \mathbf{U}^{(t)'} \boldsymbol{\rho}^{(t)} + \sum_{l=1}^{t-1} \sum_{m=l+1}^t (x_l r_m \lambda_{lm})\right]}{\prod_{j=1}^t \left(1 + \exp\left(\theta_j + \sum_{l=1}^{j-1} x_l \rho_{lj} + \sum_{l=1}^{j-2} \sum_{m=l+1}^{j-1} x_l x_m \rho_{lmj} + \dots + x_1 \dots x_{j-1} \rho_{12..j}\right)\right) \left(1 + \exp\left(\pi_j + \sum_{l=1}^{j-1} r_l \delta_{lj} + \sum_{l=1}^{j-1} x_l \lambda_{lj}\right)\right)} \quad (31)$$

$$= \frac{\sum_{\mathbf{X}^{(t)}} \prod_{j=1}^t \left(1 + \exp\left(\theta_j + \sum_{l=1}^{j-1} x_l \rho_{lj} + \sum_{l=1}^{j-2} \sum_{m=l+1}^{j-1} x_l x_m \rho_{lmj} + \dots + x_1 \dots x_{j-1} \rho_{12..j}\right)\right) \left(1 + \exp\left(\pi_j + \sum_{l=1}^{j-1} r_l \delta_{lj} + \sum_{l=1}^{j-1} x_l \lambda_{lj}\right)\right)}{\sum_{\mathbf{X}^{(t)}} \prod_{j=1}^t \left(1 + \exp\left(\theta_j + \sum_{l=1}^{j-1} x_l \rho_{lj} + \sum_{l=1}^{j-2} \sum_{m=l+1}^{j-1} x_l x_m \rho_{lmj} + \dots + x_1 \dots x_{j-1} \rho_{12..j}\right)\right) \left(1 + \exp\left(\pi_j + \sum_{l=1}^{j-1} r_l \delta_{lj} + \sum_{l=1}^{j-1} x_l \lambda_{lj}\right)\right)}$$

$$\sum_{\mathbf{X}^{(t)}} P[\mathbf{X}^{(t)} | \mathbf{Z}_t, \mathbf{R}^{(t)}, \boldsymbol{\beta}_t, \boldsymbol{\rho}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\gamma}_t, \boldsymbol{\delta}^{(t)}, \boldsymbol{\lambda}^{(t)}] = 1. \quad (32)$$

$$E[\mathbf{X}^{(t)} | \mathbf{Z}_t, \mathbf{R}^{(t)}, \boldsymbol{\beta}_t, \boldsymbol{\rho}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\gamma}_t, \boldsymbol{\delta}^{(t)}, \boldsymbol{\lambda}^{(t)}] = E[\mathbf{X}^{(t)} | \mathbf{Z}_t, \mathbf{R}^{(t)}, \boldsymbol{\Omega}_t] = (\mu_{1t}, \mu_{2t}, \dots, \mu_{tt}) \quad (33)$$

Let  $\mathbf{X}_{-j}^{(t)}$  be a  $t-1$  vector of responses without the  $j$ th response; and  $\mathbf{X}_{-ij}^{(t)}$  be a  $t-2$  vector of responses without the  $i$ th and  $j$ th responses,  $i, j=1, \dots, t$ . Then

$$\mu_{jt} = \sum_{\mathbf{X}_{-j}^{(t)}} P[X_j = 1, \mathbf{X}_{-j}^{(t)} | \mathbf{R}^{(t)}, \mathbf{Z}_t, \boldsymbol{\Omega}_t] \quad \text{for } j=1, \dots, t \quad (34)$$

where the summation is over all possible values of the vector  $\mathbf{X}_{-j}^{(t)}$ .

$$\text{Cov}[\mathbf{X}^{(t)} | \mathbf{Z}_t, \mathbf{R}^{(t)}, \boldsymbol{\Omega}_t] = \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1t} \\ \vdots & & \vdots \\ \sigma_{t1} & \dots & \sigma_{tt} \end{bmatrix} \quad (35)$$

where

$$\sigma_{ij} = E[X_i, X_j | \mathbf{Z}_t, \mathbf{R}^{(t)}, \boldsymbol{\Omega}_t] - \mu_{it} \mu_{jt}, \quad i, j=1, \dots, t \quad (36)$$

and

$$E[X_i, X_j | Z_t, R^{(t)}, \Omega_t] = \sum_{(x_{-j}^{(t)})} P[X_i = 1, X_j = 1, X_{-j}^{(t)} | Z_t, R^{(t)}, \Omega_t] \quad (37)$$

again, the summation is over all possible values of the vector  $X_{-j}^{(t)}$ .

Assuming  $\rho^{(t)} = \mathbf{0}$  and ignorability, i.e.  $\alpha^{(t)} = \mathbf{0}, \gamma_t = \mathbf{0}, \delta^{(t)} = \mathbf{0}, \lambda^{(t)} = \mathbf{0}$  then

$$P[X^{(t)} | Z_t, R^{(t)}, \beta_t, \rho^{(t)} = \mathbf{0}, \alpha^{(t)} = \mathbf{0}, \gamma_t, \delta^{(t)} = \mathbf{0}, \lambda^{(t)} = \mathbf{0}] = \frac{\exp\{\sum_{j=1}^t x_j \theta_j\}}{\prod_{j=1}^t (1 + \exp(\theta_j))} = \prod_{j=1}^t P[X_j = x_j | Z(j), \beta(j)] \quad (38)$$

since  $\theta_j = \beta(j)'Z(j), j = 1, \dots, t$ .

Assuming CRD,

$$P[X^{(t)} | Z_t, R^{(t)}, \beta_t, \rho^{(t)}, \alpha^{(t)} = \mathbf{0}, \gamma_t, \delta^{(t)} = \mathbf{0}, \lambda^{(t)} = \mathbf{0}] = \prod_{j=1}^t P[X_j = x_j | X_1, \dots, X_{j-1}, Z_t, R^{(t)}, \beta_t, \rho^{(t)}] \quad (39)$$

Now for fixed T, substitute the derived marginal and joint distributions of the repeated responses given the covariates and the missing data mechanism to equation (7) to obtain the prevalence rate of the disease of interest or  $E(Y|X, Z, R)$ .

### 3. MARKOV CHAIN MONTE CARLO SAMPLING

To estimate the parameters in the model presented in the previous section, Markov chain Monte Carlo(MCMC) methods are proposed. For the general procedures of implementing MCMC methods, refer to Gilks et al., (1996). In the simple case of a univariate response, the procedure is illustrated below.

Here  $t = 1$ , so we may write  $X_1 = X, R_1 = R, Z(1) = Z, \beta(1) = \beta, \alpha_1 = \alpha, \gamma(1) = \gamma, q_1 = q$ .

$$\text{From (4), } X|Z, \beta \sim \text{Ber} \left\{ \frac{\exp[\beta'Z]}{1 + \exp[\beta'Z]} \right\}.$$

Now, assuming that  $\beta | \mathbf{b}, \Sigma \sim N_q(\mathbf{b}, \Sigma), \mathbf{b} \sim N_q(\mathbf{0}, \mathbf{I}), \Sigma \sim \text{Inv Wishart}(v, S^{-1})$  where  $v$  and  $S$  are known and  $S$  is symmetric, positive definite, the posterior distribution is

$$p[\beta | X, Z, \mathbf{b}, \Sigma] = p[X | Z, \beta] p[\beta | \mathbf{b}, \Sigma] p[\mathbf{b}] p[\Sigma] \quad (40)$$

Thus, we have,

$$\begin{aligned} & \frac{\exp[\mathbf{x}\boldsymbol{\beta}'\mathbf{Z}]}{1 + \exp[\boldsymbol{\beta}'\mathbf{Z}]} (2\pi)^{-q/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{b})' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta} - \mathbf{b})\right\} (2\pi)^{-q/2} \exp\left\{-\frac{1}{2} \sum_{j=1}^q b_j^2\right\} \\ & \cdot \left[ 2^{vq/2} \pi^{q(q-1)/4} \prod_{j=1}^q \left[ \frac{v+1-j}{2} \right]^{-1} \right] |\mathbf{S}|^{v/2} |\boldsymbol{\Sigma}|^{-(v+q+1)/2} e^{-\frac{1}{2}\mathbf{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1})} \\ & \propto \frac{\exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{b})' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta} - \mathbf{b}) + \mathbf{x}\boldsymbol{\beta}'\mathbf{Z}\right\}}{[1 + \exp(\boldsymbol{\beta}'\mathbf{Z})]} \end{aligned} \quad (41)$$

Since the observations are independent,

$$p[\boldsymbol{\beta} | \mathbf{X}, \mathbf{Z}, \mathbf{b}, \boldsymbol{\Sigma}] \propto \frac{\exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{b})' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta} - \mathbf{b}) + \sum_{i=1}^n x_i \boldsymbol{\beta}' \mathbf{Z}_i\right\}}{\prod_{i=1}^n [1 + \exp(\boldsymbol{\beta}' \mathbf{Z}_i)]} \quad (42)$$

The expression in (42) may thus be written as  $g(\boldsymbol{\beta})$ .

Consider the case where  $q = 3$ , i.e.  $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ . Since the probability function  $g(\boldsymbol{\beta})$  in (42) is not of closed form, find an envelope function  $h(\boldsymbol{\beta}) \ni h(\boldsymbol{\beta}) \geq g(\boldsymbol{\beta}) \forall \boldsymbol{\beta} \in \mathcal{R}^q$ , and sampling from  $h(\boldsymbol{\beta})$  is possible. Consider  $h(\cdot)$  to be Normal( $\cdot, \cdot$ ), i.e.

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \sim N \left[ \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{00} & \sigma_{01} & \sigma_{02} \\ \sigma_{10} & \sigma_{11} & \sigma_{12} \\ \sigma_{20} & \sigma_{21} & \sigma_{22} \end{bmatrix} \right]$$

where  $\boldsymbol{\Sigma}$  is symmetric and positive definite, hence  $\boldsymbol{\Sigma}^{-1}$  exists and is also symmetric.

Let

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \sigma^{00} & \sigma^{01} & \sigma^{02} \\ \sigma^{10} & \sigma^{11} & \sigma^{12} \\ \sigma^{20} & \sigma^{21} & \sigma^{22} \end{bmatrix} \text{ and } \mathbf{Z} = \begin{bmatrix} 1 \\ \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{bmatrix}$$

Now by assumption

$$\begin{aligned} h(\boldsymbol{\beta}) &= (2\pi)^{-3/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}[(\boldsymbol{\beta} - \mathbf{b})' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta} - \mathbf{b})] + \sum_{i=1}^n \boldsymbol{\beta}' \mathbf{Z}_i x_i\right\} \\ h(\boldsymbol{\beta}) &= (2\pi)^{-3/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2} \left[ \left( \boldsymbol{\beta} - \left( \mathbf{b} + \sum_{i=1}^n x_i \mathbf{Z}_i \boldsymbol{\Sigma} \right) \right)' \boldsymbol{\Sigma}^{-1} \left( \boldsymbol{\beta} - \left( \mathbf{b} + \sum_{i=1}^n x_i \mathbf{Z}_i \boldsymbol{\Sigma} \right) \right) \right] \right\} \\ & \quad \cdot \exp\left\{ \mathbf{b}' \sum_{i=1}^n x_i \mathbf{Z}_i + \frac{1}{2} \left[ \sum_{i=1}^n x_i \mathbf{Z}_i \right]' \boldsymbol{\Sigma} \left[ \sum_{i=1}^n x_i \mathbf{Z}_i \right] \right\} \end{aligned}$$

Define

$$h(\beta_j | \beta_{-j}) = \frac{h(\beta)}{\int h(\beta) d\beta_j}$$

where  $\beta_{-j} \equiv$  vector of  $\beta$  excluding the  $j$ th component.

Applying Metropolis-Hastings algorithm assuming that only  $n_1$  responses were observed out of  $n$ ,  $n_1 \leq n$ , the following iterations are performed:

a) Given  $x_{\text{mis}} = (x_{n_1+1}, \dots, x_n)$ , generate  $\beta^* = (\beta_1^*, \beta_2^*, \dots, \beta_q^*)$  as follows:

Use  $\hat{\beta}_{\text{MLE}}$  based on  $P[X|Z, \beta]$  with observed responses only, as the initial value

$\beta^{(0)} = (\beta_0^{(0)}, \beta_1^{(0)}, \beta_2^{(0)})$  i.e.

$$P[X|Z, \beta] = \prod_{i=1}^{n_1} \left[ \frac{\exp\{x_i \beta' Z_i\}}{1 + \exp\{\beta' Z_i\}} \right] = \frac{\exp\left\{ \beta' \sum_{i=1}^{n_1} x_i Z_i \right\}}{\prod_{i=1}^{n_1} [1 + \exp\{\beta' Z_i\}]}$$

$$\exp\left\{ \beta' \sum_{i=1}^{n_1} x_i Z_i - \sum_{i=1}^{n_1} \ln[1 + \exp\{Z_i' \beta\}] \right\}$$

Then equating  $\frac{\partial P[X|Z, \beta]}{\partial \beta}$  to 0, we obtain

$$\sum_{i=1}^{n_1} x_i Z_i = \sum_{i=1}^{n_1} Z_i \left( \frac{\exp\{Z_i' \hat{\beta}\}}{1 + \exp\{Z_i' \hat{\beta}\}} \right)$$

Use Newton-Raphson iterative procedure to approximate  $\hat{\beta} = \beta^{(0)} = (\beta_0^{(0)}, \beta_1^{(0)}, \beta_2^{(0)})$

a1] Now sample  $\beta_0^{(1)}$  from

$$h(\beta_0 | \beta_1^{(0)}, \beta_2^{(0)}) = \frac{h(\beta_0, \beta_1^{(0)}, \beta_2^{(0)})}{\int h(\beta_0, \beta_1^{(0)}, \beta_2^{(0)}) d\beta_0}$$

Sample  $U$  from  $U(0, 1)$ .

$$\text{If } U \leq \min \left[ 1, \frac{g(\beta_0 | \beta_1^{(0)}, \beta_2^{(0)}) h(\beta_0^{(0)} | \beta_1^{(0)}, \beta_2^{(0)})}{g(\beta_0^{(0)} | \beta_1^{(0)}, \beta_2^{(0)}) h(\beta_0 | \beta_1^{(0)}, \beta_2^{(0)})} \right]$$

accept  $\beta_0^{(1)}$ , else  $\beta_0^{(1)} = \beta_0^{(0)}$ .

Consider the conditional distribution

$$h(\beta_0 | \beta_1, \beta_2) = \frac{h(\beta_0, \beta_1, \beta_2)}{\int h(\beta_0, \beta_1, \beta_2) d\beta_0} =$$

$$\frac{\exp\left\{-\frac{1}{2}\left[\left(\beta - \left(\mathbf{b} + \sum_{i=1}^n x_i \mathbf{Z}_i \Sigma\right)\right)' \Sigma^{-1} \left(\beta - \left(\mathbf{b} + \sum_{i=1}^n x_i \mathbf{Z}_i \Sigma\right)\right)\right]\right\}}{\int \exp\left\{-\frac{1}{2}\left[\left(\beta - \left(\mathbf{b} + \sum_{i=1}^n x_i \mathbf{Z}_i \Sigma\right)\right)' \Sigma^{-1} \left(\beta - \left(\mathbf{b} + \sum_{i=1}^n x_i \mathbf{Z}_i \Sigma\right)\right)\right]\right\} d\beta_0}$$

It can be shown that

$$\beta_0 | \beta_1, \beta_2 \sim$$

$$N_1\left\{b_0 + \sum_{i=1}^n x_i \sum_{j=0}^2 Z_{ij} \sigma_{j0} - \frac{\sigma_{01}}{\sigma_{00}} \left[\beta_1 - \left(b_1 + \sum_{i=1}^n x_i \sum_{j=0}^2 Z_{ij} \sigma_{j1}\right)\right] - \frac{\sigma_{02}}{\sigma_{00}} \left[\beta_2 - \left(b_2 + \sum_{i=1}^n x_i \sum_{j=0}^2 Z_{ij} \sigma_{j2}\right)\right]\right\} \frac{1}{\sigma_{00}}$$

Let the acceptance/rejection criterion be denoted by

$$F = \frac{g(\beta_0 | \beta_1^{(0)}, \beta_2^{(0)}) h(\beta_0^{(0)} | \beta_1^{(0)}, \beta_2^{(0)})}{g(\beta_0^{(0)} | \beta_1^{(0)}, \beta_2^{(0)}) h(\beta_0 | \beta_1^{(0)}, \beta_2^{(0)})}$$

$$= \frac{\frac{g(\beta_0, \beta_1^{(0)}, \beta_2^{(0)})}{\int g(\beta_0, \beta_1^{(0)}, \beta_2^{(0)}) d\beta_0} \frac{h(\beta_0^{(0)}, \beta_1^{(0)}, \beta_2^{(0)})}{\int h(\beta_0^{(0)}, \beta_1^{(0)}, \beta_2^{(0)}) d\beta_0^{(0)}}}{\frac{g(\beta_0^{(0)}, \beta_1^{(0)}, \beta_2^{(0)})}{\int g(\beta_0^{(0)}, \beta_1^{(0)}, \beta_2^{(0)}) d\beta_0^{(0)}} \frac{h(\beta_0, \beta_1^{(0)}, \beta_2^{(0)})}{\int h(\beta_0, \beta_1^{(0)}, \beta_2^{(0)}) d\beta_0}}$$

It can be shown that

$$F = \prod_{i=1}^n \left[ \frac{1 + \exp(\beta_0^{(0)} + \beta_1^{(0)} Z_{i1} + \beta_2^{(0)} Z_{i2})}{1 + \exp(\beta_0 + \beta_1^{(0)} Z_{i1} + \beta_2^{(0)} Z_{i2})} \right] \quad (43)$$

This implies that the sampling acceptance/rejection criterion does not depend on the response variable.

$$\text{Then, accept } \beta_0^{(1)} \text{ if } U \leq \min \left[ 1, \prod_{i=1}^n \left[ \frac{1 + \exp(\beta_0^{(0)} + \beta_1^{(0)} Z_{i1} + \beta_2^{(0)} Z_{i2})}{1 + \exp(\beta_0 + \beta_1^{(0)} Z_{i1} + \beta_2^{(0)} Z_{i2})} \right] \right], \text{ otherwise } \beta_0^{(1)}$$

$$= \beta_0^{(0)}.$$

a2] Next, sample  $\beta_1^{(1)}$  from

$$h(\beta_1 | \beta_0^{(1)}, \beta_2^{(0)}) = \frac{h(\beta_0^{(1)}, \beta_1, \beta_2^{(0)})}{\int h(\beta_0^{(1)}, \beta_1, \beta_2^{(0)}) d\beta_1}$$

Again, it can be shown that

$$\beta_1 | \beta_0, \beta_2 \sim$$

$$N_1 \left\{ \left[ b_1 + \sum_{i=1}^n x_i \sum_{j=0}^2 Z_{ij} \sigma_{j1} - \frac{\sigma_{01}}{\sigma_{11}} \left[ \beta_0 - \left( b_0 + \sum_{i=1}^n x_i \sum_{j=0}^2 Z_{ij} \sigma_{j0} \right) \right] - \frac{\sigma_{12}}{\sigma_{11}} \left[ \beta_2 - \left( b_2 + \sum_{i=1}^n x_i \sum_{j=0}^2 Z_{ij} \sigma_{j2} \right) \right] \right], \frac{1}{\sigma_{11}} \right\}$$

Sample U from U(0, 1).

$$\text{If } U \leq \min \left[ 1, \frac{g(\beta_1 | \beta_0^{(1)}, \beta_2^{(0)}) h(\beta_1^{(0)} | \beta_0^{(1)}, \beta_2^{(0)})}{g(\beta_1^{(0)} | \beta_0^{(1)}, \beta_2^{(0)}) h(\beta_1 | \beta_0^{(1)}, \beta_2^{(0)})} \right]$$

accept  $\beta_1^{(1)}$ , else  $\beta_1^{(1)} = \beta_1^{(0)}$ .

$$\text{Similar to (43) } \frac{g(\beta_1 | \beta_0^{(1)}, \beta_2^{(0)}) h(\beta_1^{(0)} | \beta_0^{(1)}, \beta_2^{(0)})}{g(\beta_1^{(0)} | \beta_0^{(1)}, \beta_2^{(0)}) h(\beta_1 | \beta_0^{(1)}, \beta_2^{(0)})} =$$

$$\prod_{i=1}^n \left[ \frac{1 + \exp(\beta_0^{(1)} + \beta_1^{(0)} Z_{i1} + \beta_2^{(0)} Z_{i2})}{1 + \exp(\beta_0^{(1)} + \beta_1 Z_{i1} + \beta_2^{(0)} Z_{i2})} \right].$$

Hence accept  $\beta_1^{(1)}$  if

$$U \leq \prod_{i=1}^n \left[ \frac{1 + \exp(\beta_0^{(1)} + \beta_1^{(0)} Z_{i1} + \beta_2^{(0)} Z_{i2})}{1 + \exp(\beta_0^{(1)} + \beta_1 Z_{i1} + \beta_2^{(0)} Z_{i2})} \right] \leq 1, \text{ o.w. } \beta_1^{(1)} = \beta_1^{(0)}$$

a3] Next, sample  $\beta_2^{(1)}$  from

$$h(\beta_2 | \beta_0^{(1)}, \beta_1^{(1)}) = \frac{h(\beta_0^{(1)}, \beta_1^{(1)}, \beta_2)}{\int h(\beta_0^{(1)}, \beta_1^{(1)}, \beta_2) d\beta_2}$$

As in the previous cases, it can be shown that

$$\beta_2 | \beta_0, \beta_1 \sim$$

$$N_1 \left\{ \left[ b_2 + \sum_{i=1}^n x_i \sum_{j=0}^2 Z_{ij} \sigma_{j2} - \frac{\sigma_{02}}{\sigma_{22}} \left[ \beta_0 - \left( b_0 + \sum_{i=1}^n x_i \sum_{j=0}^2 Z_{ij} \sigma_{j0} \right) \right] - \frac{\sigma_{12}}{\sigma_{22}} \left[ \beta_1 - \left( b_1 + \sum_{i=1}^n x_i \sum_{j=0}^2 Z_{ij} \sigma_{j1} \right) \right] \right], \frac{1}{\sigma_{22}} \right\}$$

Sample U from U(0, 1).

$$\text{If } U \leq \min \left[ 1, \frac{g(\beta_2 | \beta_0^{(1)}, \beta_1^{(1)}) h(\beta_2^{(0)} | \beta_0^{(1)}, \beta_1^{(1)})}{g(\beta_2^{(0)} | \beta_0^{(1)}, \beta_1^{(1)}) h(\beta_2 | \beta_0^{(1)}, \beta_1^{(1)})} \right]$$

accept  $\beta_2^{(1)}$ , else  $\beta_2^{(1)} = \beta_2^{(0)}$ .



Again, similar to (43)

$$\frac{g(\beta_2 | \beta_0^{(i)}, \beta_1^{(i)}) h(\beta_2^{(0)} | \beta_0^{(i)}, \beta_1^{(i)})}{g(\beta_2^{(0)} | \beta_0^{(i)}, \beta_1^{(i)}) h(\beta_2 | \beta_0^{(i)}, \beta_1^{(i)})} = \prod_{i=1}^n \left[ \frac{1 + \exp(\beta_0^{(i)} + \beta_1^{(i)} Z_{i1} + \beta_2^{(0)} Z_{i2})}{1 + \exp(\beta_0^{(i)} + \beta_1^{(i)} Z_{i1} + \beta_2 Z_{i2})} \right]$$

Hence accept  $\beta_2^{(1)}$  if

$$U \leq \prod_{i=1}^n \left[ \frac{1 + \exp(\beta_0^{(i)} + \beta_1^{(i)} Z_{i1} + \beta_2^{(0)} Z_{i2})}{1 + \exp(\beta_0^{(i)} + \beta_1^{(i)} Z_{i1} + \beta_2 Z_{i2})} \right] \leq 1$$

otherwise set  $\beta_2^{(1)} = \beta_2^{(0)}$ .

Continuing in this fashion, we arrive at a Markov Chain Monte Carlo sample of  $\beta$  whose estimated posterior mean we denote as  $\beta^{(m_1)} = (\beta_0^{(m_1)}, \beta_1^{(m_1)}, \beta_2^{(m_1)})$ .

b) Now given  $\beta^{(m_1)}$ , generate  $\mathbf{x}_{\text{mis}}^{(1)} = (x_{n_1+1}, \dots, x_n)'$  using Gibbs sampling from

$$P(X | \beta^{(m_1)}, Z) \equiv \text{Ber} \left\{ \frac{\exp[\beta^{(m_1)'} Z]}{1 + \exp[\beta^{(m_1)'} Z]} \right\}.$$

Assuming  $\{x_i, i = n_1 + 1, n_1 + 2, \dots, n\}$  are independent and identically distributed, we sample  $x_i$  from

$$P(X | \beta^{(m_1)}, Z) = \frac{\exp\{x_i (\beta_0^{(m_1)} + \beta_1^{(m_1)} Z_{i1} + \beta_2^{(m_1)} Z_{i2})\}}{1 + \exp(\beta_0^{(m_1)} + \beta_1^{(m_1)} Z_{i1} + \beta_2^{(m_1)} Z_{i2})} \text{ for } i = n_1 + 1, n_1 + 2, \dots, n.$$

Starting with the new values of  $\mathbf{x}_{\text{mis}}^{(1)} = (x_{n_1+1}, \dots, x_n)'$ , iterate (a) and (b) until  $\beta^*$  and  $\mathbf{x}_{\text{mis}}^*$  converge, say to  $\hat{\beta}_{MH}$  and  $\hat{\mathbf{x}}_{\text{mis}, MH}$ , respectively.

Now from (6)

$$P[R = r | X = x, Z, \alpha, \gamma] = \frac{\exp\{r\alpha x + r\gamma' Z\}}{1 + \exp\{r\alpha x + r\gamma' Z\}}$$

Let

$$\begin{aligned} (\alpha, \gamma') | \mathbf{c}, \Sigma_\gamma &\sim N_{q+1}(\mathbf{c}, \Sigma_\gamma) \\ \mathbf{c} &\sim N_{q+1}(\mathbf{0}, \mathbf{I}) \\ \Sigma_\gamma &\sim \text{Inv Wishart}(w, W) \end{aligned}$$

where  $w$  and  $W$  are known and  $W$  is sym, positive definite.

Similar Metropolis-Hastings procedure as above is applied to obtain estimates of  $(\alpha, \gamma')$ . Iterative procedures are continuously applied to the models derived in the previous section.

#### 4. SUMMARY AND DISCUSSIONS

This paper presents a scheme for modeling repeated binary responses with a nonignorable missing data mechanism. The models used a simple relationship between dropout, responses and their past values integrating covariate information. In the univariate case, the conditional distribution of the response given the covariates and missingness showed a perturbed Bernoulli distribution; and its conditional expectation and variance follow the form of the Bernoulli distribution. For the bivariate case, the form of the conditional distribution of the responses given covariates, missing data mechanism and their past values has yet to be compared with the standard probability distributions. When ignorability was assumed though, the joint bivariate model can be expressed as the product of the marginals given previous response. The same result extended to the case of  $t$  repeated binary measures.

In the case of univariate response, a sequence of Metropolis-Hastings algorithm and Gibbs sampler is illustrated to estimate the parameters of the models. These MCMC methods are sequentially combined to obtain estimates of the parameters and the missing data. It is observed from the models that the number of parameters increases considerably as the number of repeated observations increases.

#### ACKNOWLEDGEMENTS

The author would like to thank Dr. Joselito C. Magadia, Dr. Ana L. Tabunda, Dr. Jesus N. Sarol Jr., Dr. Jose Ramon G. Albert, Dr. Emiel B. Barrios and the referee for their helpful comments and suggestions. Conversations with professors and friends at the School of Statistics, UP Diliman have been encouraging and informative. Funding for this study is provided by PCASTRD-DOST.

#### References

- DIGGLE, P. AND KENWARD, M. 1994 *Informative drop-out in longitudinal data analysis*. Applied Statistics 43, 49-93.
- FITZMAURICE, G., CLIFFORD, P. AND HEATH, A. 1996a *Logistic regression models for binary panel data with attrition*. Journal of the Royal Statistical Society Ser. A 159, 249-263.
- FITZMAURICE, G., LAIRD, N. AND ZAHNER, G. 1996b *Multivariate logistic models for incomplete binary responses*. Journal of the American Statistical Association. 91, 99-108.
- FITZMAURICE, G., MOLENBERGS, G. AND LIPSITZ, S. 1995 *Regression models for longitudinal binary responses with informative drop-outs*. Journal of the Royal Statistical Society Ser. B 58, 691-704.
- FOLLMANN, D. AND WU, M. 1995 *An approximate generalized linear model with random effects for informative missing data*. Biometrics 51, 151-168.
- GILKS, W.R., RICHARDSON, S., AND SPIEGELHALTER, D.G. (eds.) 1996 *Markov chain Monte Carlo in Practice*. London: Chapman and Hall.
- LIANG, K. AND ZEGER, S. 1986 *Longitudinal data analysis using generalized linear models*. Biometrika 73, 13-22.
- LIPSITZ, S.R., FITZMAURICE, G.M., SLEEPER, L., AND ZHAO, L.P. 1995 *Estimation methods for the joint distribution of repeated binary observations*. Biometrics 51, 562-570.

LITTLE, R. AND RUBIN, D. 1987 *Statistical analysis with missing data*. New York: John Wiley.

SKINNER, C. AND COKER, O. 1996 *Regression analysis for complex survey data with missing values covariate*. Journal of the Royal Statistical Society Ser. A 159, 265-274.

TROXEL, A. B., LIPSITZ, S. R. AND HARRINGTON, D. P. 1998 *Marginal models for the analysis of longitudinal measurements with nonignorable non-monotone missing data*. Biometrika 85, 661-672.

ZEGER, S. LIANG, K. AND SELF, S. 1985 *The analysis of binary longitudinal data with time-dependent covariates*. Biometrika 72, 31-38.