Missingness Mechanism that Incorporated Joint Modeling of Longitudinal Data with Monotone Dropout

ALUKO O.*, MWAMBI H.

School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Private Bag X01 Scottsville 3209, Pietermaritzburg, South Africa

Abstract  We analyzed repeated measurement of continuous responses with monotone dropout. We are interested in reducing the bias associated with treatment effects, but the results’ credibility relies on the validity of the techniques applied to analyze the data, and under the conditions where the techniques gives reliable answers. Furthermore, the robustness of the trial findings are determined through the application of sensitivity analysis which verifies to which extent the results are affected by changes in techniques, values of unmeasured variables and model assumptions. Moreover, the results obtain from the missing not at random (MNAR) is the same as their counterpart in missing at random (MAR). In addition, using multiple imputation (MI) in the analysis also improves the accuracy of results.

Keywords  Sensitivity Analysis, Multiple Imputation, Selection Model, Pattern Mixture Model, Monotone Dropout

1 Introduction

The prevalence of missing data is high in real application data. The techniques for handling missing data is classified into ignorable and non-ignorable, within the likelihood-based (parametric) framework [30]. In ignorable, the model for the distribution of the response variable is the only requirement, while the latter requires a model for the full-data distribution which represents the distribution of both the response and the missing data indicators. The identifiability issues was introduced by non-ignorable missingness; by the fact that the missing data were not observed and have no data with which the distribution can be estimated. In literature, we have techniques which resolves this issue via recourse to parametric modeling assumptions [39, 5, 21, 20]. According to [3], if an issue cannot be resolved nonparametrically given an infinite sample then it is usually...
complex to resolve it parametrically. In addition, the use of parametric approaches; should not indirectly resolve identifiability issues [13].

The alternative option as part of the analysis is the adoption of non-identifiability where the full-data distribution can be factored into two components; the observed-data distribution which can be nonparametrically identified from the observed data, and the conditional distribution of the missing data given the observed data which is sometimes termed extrapolation distribution, but which is not [4, 18]. The expression of different assumptions about missing data is termed identifying restrictions. However, the assumptions that link the marginal distribution of the observed data to the extrapolation distribution is identifying restriction. The most popular identifying restriction is the missing at random (MAR) assumption, but many others exist. However, we are estimating the observed-data distribution either parametrically or nonparametrically and also identifying restriction to identify the extrapolation distribution.

An advocacy was raised by the national research council [27] for practitioners to conduct sensitivity analyses where the assumptions about the non-identified model components are to determine how much conclusions are being driven by untestable assumptions about the missingness. Considering family of such assumptions, reference was made to the estimates which index the restrictions as a sensitivity parameter. Frequentist semiparametric approach was motivated because of the issues about parametric assumptions [31, 33]. There are avoidance of assumptions by these methods about the full-data distribution, and the parametric model for the missingness given the outcome (reference as to the missing data mechanism) is specified and, it is optional for semiparametric model [13]. Different methods have been proposed using likelihood-based methods which are flexible for the semiparametric approaches and allow flexibility of sensitivity analysis [38, 14, 15]. Some of those works have bayesian approach, but also include proposals related to frequentist framework [34]. The Bayesian approach is advantageous about the unidentified components of the model to be encoded in an informative prior, and in order to have a valid inference; the requisite subject-matter expertise is incorporated into the analysis formally. In our study, we propose the flexibility and fully Bayesian framework for missing outcome variable. The use of this framework handles any type of vector-valued outcome data that has these features. In Section 2, we describe the notation and general concept of selection and pattern mixture models. Section 3 gives a description of the modeling framework use in the analysis. In Section 4, we give brief overview of the data use. The paper ends with a discussion in Section 5.

2 Notation and concepts

The modeling of missingness notation was introduced by [30, 19]. Let the response of interest be $y_{ij}$, for the $i$th study subject where $i = 1, ..., N$, designed to be measured at occasion $t_j$, where $j = 1, ..., n$. When the intention was to have $n$ observations per subject, but due to dropout some subjects may end up having less than the $n$ intended observations. However, we may assume that the $i$th individual is actually observed $n_i$ times. For subject $i$ and occasion $j$, we define $R_{ij} = 1$, if $y_{ij}$ is observed, and $0$ otherwise. The $y_{ij}$ is divided into two sub-vectors, $y^n_i$ and $y^m_i$ which represents $y_{ij}$ for $R_{ij} = 1$ and $R_{ij} = 0$ respectively. Furthermore, due to dropout the measurements for each individual is recorded up to a certain time point, after which all data are unobserved. The dropout indicator is defined as $D_i$ given by $D_i = 1 + \sum_{j=1}^{n} R_{ij}$ which represents the occasion at which the
first dropout occurs. It is essential to consider the joint modeling for the measurement process together with the dropout process, especially in the modeling of the missing data. Therefore, we assume the full data density is given by

\[ f(y_i, r_i | X_i, Z_i, \theta, \psi), \]  

(1)

where \( X_i \) and \( Z_i \) represent the design matrices for measurement process and dropout mechanism respectively, while \( \theta \) and \( \psi \) denote the parametrization vectors for the joint distribution. (1), the factorization of joint density function can be done in two ways to facilitate modeling. First, the selection and pattern mixture models are defined by the conditional factorizations of the joint distribution of \( Y \) and \( R \) which are discussed in details in [18], and briefly analyzed below. The factorization of the selection model is as follows

\[ f(y_i, r_i | X_i, Z_i, \theta, \psi) = f(y_i | X_i, Z_i, \theta)f(r_i | y_i, X_i, \psi), \]  

(2)

the first part of the factorization represents the marginal density of the measurement process, while the second part is the density of the dropout process, conditional on measurements. In addition, the alternative factorization based on the pattern mixture models [16, 17] is given by

\[ f(y_i, r_i | X_i, Z_i, \theta, \psi) = f(y_i | r_i, X_i, Z_i, \theta)f(r_i | X_i, \psi). \]  

(3)

These models excellently in [7, 9, 10, 19]. Through the selection model framework, the missing data process was developed by [30, 19]. Among the several missing data processes, they make distinctions. Based on the second part of model (2), these processes may be derived, that is

\[ f(r_i | y_i, X_i, \psi) = f(r_i | y_{i0}, y_{im}, X_i, \psi). \]  

(4)

A process is termed missing completely at random (MCAR), if the distribution of missingness process is reduced to \( f(r_i | y_i, X_i, \psi) = f(r_i, X_i, \psi) \) meaning that the process is independent of the measurements. When the dependence of the missingness probability is on the observed measurement \( y_{i0} \), but not on the unobserved measurements \( y_{im} \), which means that \( f(r_i | y_i, X_i, \psi) = f(r_i | y_{i0}, X_i, \psi) \), then the process is called missing at random(MAR). Lastly, when data are missing not at random (MNAR) or exhibit an informative process then the missingness probability depends on the unobserved measurement, \( y_{im} \) and possibly on the observed measurement, \( y_{i0} \) which means \( f(r_i | y_i, X_i, \psi) = f(r_i | y_{i0}, y_{im}, X_i, \psi) \). However, expression (4) is an informative process which is irreducible.

3 Modeling framework

In this section, we compare selection model (SM) and pattern mixture model (PMM) using marginal restrictions which is a useful form of the sensitivity analysis. In addition, each model is discussed briefly.
3.1 Selection model

As earlier mentioned, selection model factors the joint distribution into two parts: the marginal measurements and missingness models, where the former discusses the complete measurements and the latter characterizes the conditional distribution of the outcome indicators given the observed and unobserved measurements. In this model, the first specification is the measurement distribution, then followed by the probability of being observed which depends on the data. Thus, using a selection model as specified in equation (2), [5] combined a multivariate normal model for the measurement process \( f(y_i|X_i, \theta) \), and a logistic regression model for the dropout process \( f(r_i|y_i, Z_i, \psi) \).

When we assume that the missingness is due to monotone dropout, for each individual the first measurement \( y_{i1} \) is observed. However, recall that \( D_i \) has been stated to be the dropout indicator which denotes the occasion at which the dropout first occurs. Also, let \( D_i = d_i \) express the dropout time for individual \( i \), where \( D_i = n + 1 \), if the sequence of measurement is complete. Thus, the selection model expressed in equation (2) arise when the joint likelihood of the measurement and dropout processes is factorized as following

\[
f(y_i, D_i|X_i, Z_i, \theta, \psi) = f(y_i, D_i|X_i, Z_i, \theta) f(d_i|y_i, \psi).
\]

We recall that model for dropout process is based on a logistic regression for the probability of dropout at event \( j \), was present in the study at the previous event. However, this probability is represented by \( g_i(y_{ij}, h_{ij}) \), where \( h_{ij} \) denotes the history of the measurement process. Therefore, one may assume that \( g_i(y_{ij}, h_{ij}) \) satisfies the model of

\[
\text{logit}[g(h_{ij}, y_{ij})] = \text{logit}[p(D_i = j|D_i \geq j, h_{ij}, y_{ij})] = \eta(h_{ij}, y_{ij}),
\]

the linear predictor \( \eta(h_{ij}, y_{ij}) \) which depends on \( h_{ij} \) and \( y_{ij} \). Equation (5) simply expresses the modeling of the dropout mechanism by assuming \( \eta(h_{ij}, y_{ij}) \) which depends only on the current measurement and previous measurement \( y_{ij-1} \), but not on future measurements or higher order history, with corresponding regression coefficients, \( \psi_1 \) and \( \psi_2 \) [32]. Justifying the dependence on future unobserved measurement is difficult and therefore not modeled here. The observed history of subject \( i \) up to time \( t_{i,j-1} \) can be included, but we assume first order history for simplicity. The logistic dropout model could be written as

\[
\text{logit}[g(y_{i,j-1}, y_{ij})] = \text{logit}[p(D_i = j|y_{i,j-1}, y_{ij})] = \psi_0 + \psi_1 y_{i,j-1} + \psi_2 y_{ij}.
\]

The model expressed in equation (6) is MAR when \( \psi_2 \) equal to zero, since the dropout does not depend on the current measurement this model becomes MCAR, since the dropout does not depend on outcome at all, and finally when \( \psi_1 \) is equal to zero the dropout mechanism is MNAR. We use the likelihood ratio (LRT) to compare model fit under a model that assumes the incomplete data due to monotone dropout MCAR versus MAR [5]. The LRT test statistic follows a null asymptotic \( \chi^2 \) distribution. The details of this statistic derivation are in [5, 21]. The significance of the LRT is an indication that the least restrictive of the two models is preferred. However, Based on [5], the use of LRT statistic to test the MAR against MNAR is not the behaviour of the LRT statistic for the MNAR parameter \( \psi_2 \) is the information available on \( \psi_2 \) is rare and interwoven with other features of both measurement and dropout models [11]. Thus, considering the model based on Diggle and Kenward type; it is assumed to be
true, but their tests is conditional on the alternative model. The difference between the MAR and MNAR mechanism can only be made using untestable modeling assumptions, such as distributional form [22]. This assumption gives rise to the dropout in a sample which is unverifiable by the observed measurements, and any test concerning the dropout process is invalidated as stated in [25]. The justification is that the parameters of the dropout model process depend partially on dropout. Furthermore, unless one clarifies it in the posited MNAR model; the distinction (MAR/MNAR) [26]. However, the challenge of model identifiability poses a complex issue when considering the models for dropout mechanism. In addition, it is recommended to conduct sensitivity analysis of the parameters of the measurement model across models that make different assumptions about the dropout process [25].

3.2 Pattern mixture model (PMM)

3.2.1 Identifying restrictions

In this study, patterns are defined based on dropout at every time point, except baseline [2]. However, when the last observed subject has \( t \)th outcome value and one dropouts after this point, this individual belongs to pattern \( t = 1, \ldots, T \). The uniqueness of Pattern mixture model (PMM) starts from a particular decomposition of the joint distribution of the response variable together with the dropout indicator [2]. The pattern-mixture distribution of complete outcome values \( y_1, \ldots, y_T \) is written as

\[
f(y_1, \ldots, y_T) = \sum_{t=1}^{T} \alpha_t f_t(y_1, \ldots, y_T), \tag{7}
\]

the proportion of pattern \( t \) is represented by \( \alpha_t \) and \( f_t(y_1, \ldots, y_T) \) stands for \( f(y_1, \ldots, y_T | t) \). In equation (7), the expression of the whole population distribution is in terms of a mixture of the pattern populations distributions. The decomposition of these can be written as

\[
f_t(y_1, \ldots, y_T) = f_t(y_1, \ldots, y_t) f_t(y_{t+1}, \ldots, y_T|y_1, \ldots, y_t)
\]

\[= f_t(y_1, \ldots, y_t) \prod_{s=t+1}^{T} f_t(y_s|y_1, \ldots, y_{s-1}). \tag{8}\]

In equation (8), the first component is identified from the observed response values. Thus, identification of unidentified parameters is an important part of the distributions of conditional pattern that forms the basic framework of the imputation process. This complexity is overcome by identifying restrictions: the unidentified parameters of missing patterns are set equal to appropriate functions of the parameters describing the distributions of other patterns. Further descriptions on the identifying restrictions could be found in [35]. Under the complete-case missing value (CCMV), the identification is based on the completers’ pattern (i.e. borrow from subjects with complete outcome profile), which has pattern \( T \). However, this identification can be written as

\[
f_t(y_s|y_1, \ldots, y_{s-1}) = f_T(y_s|y_1, \ldots, y_{s-1}), \ s = t + 1, \ldots, T. \tag{9}\]
In neighboring-case missing values (NCMV) approach, the closest neighboring pattern is applied. In this case, an incomplete observation at a time point is borrowed from the nearby pattern for which the response values are observed at this time point, but unobserved later. Then we have

\[ f_t(y_s | y_1, \ldots, y_{s-1}) = f_s(y_s | y_1, \ldots, y_{s-1}), \quad s = t + 1, \ldots, T. \]  

(10)

Thus, identification can be based on the identified patterns as suggested in the expression

\[ f_t(y_s | y_1, \ldots, y_{s-1}) = \sum_{j=s}^{T} \omega_{sj} f_j(y_s | y_1, \ldots, y_{s-1}), \quad s = t + 1, \ldots, T. \]  

(11)

The set of positive \( \omega_{sj} \)'s is represented as \( \omega_s \). All \( \omega_s \) that adds up to 1 gives an identification scheme that is validity. Expression (11) becomes a special case of equation (9) by setting \( \omega_{sT}=1 \) and all other \( \omega_{sj}=0 \) and to the special case of equation (10) by setting \( \omega_{ss}=1 \) and all other \( \omega_{sj}=0 \). Finally, the available-case missing values (ACMV) approach offers a compromise between CCMV and NCMV as missing information is borrowed from all available pattern weighted by occurrence of each pattern. The \( \omega_s \) was determined in [23] such that equation (11) corresponds to ACMV. The coefficients are defined as

\[ \omega_{sj} = \frac{\alpha_j f_j(y_1, \ldots, y_{s-1})}{\sum_{l=s}^{T} \alpha_l f_l(y_1, \ldots, y_{s-1})}, \quad j = s, \ldots, T. \]  

(12)

The status of ACMV is unique because it is a natural counterpart of MAR in the PMM framework. In application, analysis under MAR can be the starting point for sensitivity analyses under MNAR.

### 3.3 Multiple imputation method

In this section, we explain the implementation of PMM analysis for Gaussian outcome variables following the standard three-stage MI approach described in [28].

#### 3.3.1 Pattern parameter estimation

Unique models are developed to fit the response variable within each pattern. Time is the class variable and no specification of random effects.

Let \( Y_i = (y_i, 1, \ldots, y_i, T) \) the complete outcome vector for the \( i \)th subject of pattern \( t \) and the observed part \( Y_{i,obs} = (y_i, 1, \ldots, y_i, t) \). The expression for MMRMs per pattern is written as

\[ Y_{i,obs} = X_i \beta_t + \epsilon_i, \]  

(13)

where \( \epsilon_i \sim N(0, \Sigma_t) \), \( \Sigma_t \) is unstructured, and \( \epsilon_i \)'s are independent. The matrix \( X_i \) contains the known fixed-effects covariates whereas \( \beta_t \) contains the unknown parameters. In the first stage, we obtain estimates \( \hat{\beta}_t, \hat{VAR}(\hat{\beta}_t), \hat{\Sigma}_t \) and \( \hat{VAR}(\text{col}(\hat{\Sigma}_t)) \), where \( \text{col}(\hat{\Sigma}_t) \) is the vector containing the coefficients of the diagonal and lower part of \( \hat{\Sigma}_t \) [2].
3.3.2 Imputation

The missing response values is imputed sequentially by value. Thus, on how to obtain a run of \( M \) imputed values of \( y_{i,t+1} \). Multiple imputation of \( y_{i,t+1}, \ldots, y_{i,T} \) follows the same process by considering the previous imputed values as observed ones [2]. The imputed values of \( y_{i,t+1} \) are drawn from conditional pattern distributions. The pattern of imputation selection is done through the identifying restrictions chosen. For the purpose of explanation, the first thing is to use pattern \( r (t + 1 \leq r \leq T) \). However, according to [2] \( \mu_{i,r} \) was introduced for the mean of \( Y_i \), which is \( \mu_{i,r} = X_i \beta_r \). Based on appropriate parts of \( \mu_{i,r} \) and \( \Sigma_r \), we define the distributions of \( Y_i \) components as \( Y_{i,\text{obs}} \sim N(\mu_{i,r,1}, \Sigma_{r,11}) \) and \( y_{i,t+1} \sim N(\mu_{i,r,2}, \Sigma_{r,22}) \). The notation of their covariances are \( \Sigma_{r,12} \) and \( \Sigma_{r,21} \). Furthermore, using \( 2|1 \) as notation for \( y_{i,t+1}|y_{i,1}, \ldots, y_{i,t} \) given \( y_{i,1}, \ldots, y_{i,t} \) is described as

\[
 f_r(y_{i,t+1}|y_{i,1}, \ldots, y_{i,t}) \sim N(\mu_{i,r,2|1}, \Sigma_{r,2|1}),
\]

where

\[
 \begin{align*}
 \mu_{i,r,2|1} &= \mu_{i,r,2} + \Sigma_{r,21}[\Sigma_{r,11}]^{-1}(Y_{i,\text{obs}} - \mu_{i,r,1}), \\
 \Sigma_{r,2|1} &= \Sigma_{r,22} - \Sigma_{r,21}[\Sigma_{r,11}]^{-1}\Sigma_{r,12}.
\end{align*}
\]

(14)

Through the Bayesian posterior predictive distributions; the uncertainty pertaining to the pattern. This is based on the Gaussian distributions and non-informative Jeffreys’ priors, where the values of \( \hat{\beta}_r^{(m)} \) and \( \hat{\Sigma}_r^{(m)} \), \( m = 1, \ldots, M \). When the derivation of \( \hat{\mu}_{i,r}^{(m)} \) is obtained, the imputed values of \( y_{i,t+1} \) are drawn from the conditional pattern distributions which are written as

\[
 f_r^{(m)}(y_{i,t+1}|y_{i,1}, \ldots, y_{i,t}) \sim N(\hat{\mu}_{i,r,2|1}^{(m)}, \hat{\Sigma}_r^{(M)}), m = 1, \ldots, M.
\]

(15)

In equation (11), the ACMV suggests that imputing the missing \( y_{i,t+1} \) value is based on a sum of conditional distributions of several patterns weighted by occurrence of each pattern. In accordance with the MI procedure, handling the weighted summation is done via random pattern selection over imputations. The coefficients \( \omega_{sk} \) in equation (11) is calculated for each imputation and this characterizes the pattern probabilities. In the imputation, random pattern selection is based on coefficient values \( \omega_{sk}^{(m)} \) and drawn from the uniform distribution. Therefore, select pattern \( p \) if

\[
 \Sigma_{j=s}^{p-1} \omega_{sk}^{(m)} < U^{(m)} < \Sigma_{j=s}^{P} \omega_{sk}^{(m)}.
\]

(16)

3.3.3 Pooled analysis

Fitting the complete datasets is done using modeling strategy explained in equation (13) for pattern parameter estimation. The procedure for MMRM incorporates the interaction between the full group-by-time at every time point for the fixed effects and ARH(1) error covariance matrix. The analysis model can be expressed as

\[
 Y_i = X_i \beta^* + \epsilon_i
\]

(17)
where \( \epsilon_i^* \sim N(0, \Sigma) \), \( \Sigma \) is ARH(1) structure and the \( \epsilon_i^* \)’s are independent. The unknown parameters of fixed effects is contained in the vector \( \beta^* \). The MI approach combines the inferences made on the \( M \) imputations into a single one. Let us define \( \hat{\beta}^{*(m)} \), where \( m = 1, \ldots, M \), the estimators of \( \beta^* \) by imputation and \( \hat{\beta}^* \) the pooled estimator of \( \beta \). Thus, according to Rubin’s expression
\[
\beta - \hat{\beta}^* \sim N(0, V),
\]
\( \hat{\beta}^* \) is the average of \( \hat{\beta}^{*(m)} \)’s whereas the pooled variance \( V \) is expressed as
\[
V = W + \left( \frac{M+1}{M} \right) B,
\]
the average of \( \overline{VAR}(\hat{\beta}^{*(m)}) \) is \( W \) and \( B \) is the between-imputation variance:
\[
B = \frac{1}{M-1} \sum_{m=1}^{M} (\hat{\beta}^{*(m)} - \hat{\beta}^*)(\hat{\beta}^{*(m)} - \hat{\beta}^*)^T.
\]
Under MNAR, the pooled estimator \( \hat{\beta}^* \) is consistent. The \( F \)- distributions of the tests of fixed effects are described in [12].

### 4 Description of the case study

#### 4.1 Age-related macular degeneration (ARMD) trial

The age-related macular degeneration (ARMD) data arise from a randomized multi-center clinical trial comparing an experimental (Interferon-\( \alpha \)) versus placebo for patients diagnosed with the ARMD [8]. The focus was on the efficacy of the placebo and the highest dose (6 million units daily - MIU) of the interferon. The doses were 1.5, 3.0 and 6.0 MIU.

Patients with macular degeneration progressively lose vision. In the trial, each of 266 patients’s visual acuity were assessed at baseline and at four post randomization visits i.e, visit 1 at 4 weeks, visit 2 at 12 weeks, visit 3 at 24 weeks and visit 4 at 52 weeks respectively [2]. Treatment-effect inferences on visual acuity at visit 4 is the primary focus of the statistical analysis. Each patient’s visual acuity was evaluated based on the ability to read lines of letters on standardized vision charts. The charts display lines of five letters decreasing size, which the patient must read from top (largest letters) to bottom (smallest letters). When at least four letters are read correctly on each line; it is called one “line of vision”. Furthermore, interest is also on the visual acuity which is defined as the total number of letters read correctly. Thus, another possible approach was to consider visual acuity measured by the number lines read correctly. However, the two approaches are closely linked, as each line of vision contains five letters. We present the general overview of the different missingness patterns in Table 1.

| pattern | visit 1 | visit 2 | visit 3 | visit 4 | pattern distribution |
|---------|---------|---------|---------|---------|----------------------|
| freq    | percent | freq    | percent |         | freq    | percent |
| 1       | 0       | 0       | 0       | 0       | 188     | 83.2    |
| 2       | 0       | 0       | 0       | M       | 24      | 10.6    |
| 3       | 0       | 0       | M       | M       | 8       | 3.5     |
| 4       | 0       | M       | M       | M       | 6       | 2.7     |

Notes: Missingness patterns and the frequencies with which they occur. ‘O’ denotes observed and ‘M’ denotes missing.
4.2 Fitting selection model

We fit the selection models to the ARMD data by combining the measurement model with the logistic regression for dropout model, in line with [5] using a generic function maximization routine. In order to obtain the initial values for the parameters of the measurement model, we use the MMRM of the equation (13). Fitting the model, we assume different intercepts and treatments effects for each of the four visits, with a \((4 \times 4)\) ARH(1) structured variance-covariance matrix. Considering a multivariate normal model with unconstrained visits trend under placebo and an occasion-specific treatment effect using equation (13). From this, we obtain the parameter estimates and standard errors for the eight mean model parameters. We fit this model using the SAS procedure MIXED with REPEATED statement. Thereafter, we consider the dropout model. The dropout is an independent covariates. We fit the model with an intercept, an effects for the previous response and current unobserved measurement corresponding to MCAR, MAR and MNAR respectively. It is possible analytically to have dependence on future unobserved measurements, but in avoidance of complication, we model dependence on the current unobserved measurements. The probability of ARMD data is assumed to follow the logistic regression model (it is the commonly used model for dropout process, as indicated in [25]) in equation (6). Three parameters are involved in the logistic regression model; that is, an intercept \((\psi_0)\), the effect of the measurement prior to dropout \((\psi_1)\) and effect of the measurement at the visits of dropout \((\psi_2)\). For the four visits, the model can be written as

\[
\logit[g(y_{i,j-1}, y_{ij})] = \logit[p(D_i = j|y_{i,j-1}, y_{ij})] = \psi_0 + \psi_1 y_{i,j-1} + \psi_2 y_{ij}, j = 2, 3, 4, 5.
\] (18)

It is still a major challenge to estimate the selection model for MNAR because the dropout indicators depend on the unobserved measurement. For illustration, in the stated above selection model; the dropout indicators depend in part on the unobserved measurements at the time of dropout. Difficulty in assessing the likelihood function can be handle as provided in [6] that maximizes the log-likelihood for the model using PROC IML.

| Effect                | Parameter | MCAR     | MAR      | MNAR     |
|-----------------------|-----------|----------|----------|----------|
|                       |           | Est  (s.e) | Est  (s.e) | Est  (s.e) |
| Measurement model     |           |          |          |          |
| Intercept\(_i\)      | \(\beta_{1i}\) | 54.00 (1.47) | 54.00 (1.47) | 54.00 (1.47) |
| Intercept\(_{i1}\)   | \(\beta_{21}\) | 53.23 (1.95) | 52.89 (1.61) | 52.96 (2.20) |
| Intercept\(_{i3}\)   | \(\beta_{31}\) | 49.31 (2.29) | 49.45 (1.79) | 49.05 (2.49) |
| Intercept\(_{i5}\)   | \(\beta_{41}\) | 44.67 (2.68) | 44.69 (2.35) | 44.40 (2.73) |
| Treatment effect\(_i\) | \(\beta_{12}\) | -3.11 (2.10) | -3.11 (2.10) | -3.11 (2.10) |
| Treatment effect\(_{i1}\) | \(\beta_{22}\) | -4.73 (2.69) | -4.26 (2.36) | -4.56 (2.71) |
| Treatment effect\(_{i3}\) | \(\beta_{32}\) | -3.86 (3.14) | -3.77 (2.55) | -3.79 (2.92) |
| Treatment effect\(_{i5}\) | \(\beta_{42}\) | -4.90 (2.87) | -4.72 (2.60) | -4.76 (2.89) |
| dropout model         | \(\psi_0\) | -2.79 0.17 | -1.86 0.46 | -1.81 0.46 |
|                       | \(\psi_1\) | -0.02 0.01 | 0.02 0.02  | 0.04 0.02  |
|                       | \(\psi_2\) |          |          |          |
| -2log-likelihood      |          | 3391.3 | 3389.2 | 3387.9 |
The results in Table 2 display the parameter estimates and standard errors for the fixed effects for the selection model, which includes the eight mean model parameters all into marginal measurement model as well as in the logistic dropout model. For comparison, the MCAR, MAR and MNAR produce results are close to one another, except for negligible differences as shown in the estimates and standard error. From the assumed model, the significance of the results in the dropout model is examined using the LRT test statistic which compares that the MCAR. However, there is a challenge when assessing that dropout mechanism is MNAR because neither the LRT statistic of the models that assume the dropout is MAR as against MNAR nor there is reliability in the assessment of \( \psi_2 \) in relation to the its standard error[11]. In addition, the dropout mechanism for MNAR is unverifiable [26]. In this situation, the focus is on the marginal effect because the overall treatment effect is exempted.

Here the conditions that influence dropout are briefly discussed. Based on equation (18), modeling the logistic regression for dropout in selection models are obtained. From Table 2, it is observed that the maximum likelihood estimates for \( \psi_1 \) and \( \psi_2 \) are (0.02) and (-0.04) which give different signs and also not the same in value. This outcome is not surprising, but confirms the argument raised by [24]. In addition, it is stated in [24] that when there is positive correlation between two simultaneous measurements, the dropout model may rely on the increment, that is, \( y_{ij} - y_{i,j-1} \). Furthermore, the dropout estimates from the MNAR models is stated below

\[
\text{logit}\left[p(D_i = j | y_{ij-1} - y_{ij})\right] = -1.81 - 0.02y_{i,j-1} - 0.04(y_{ij} - y_{i,j-1}).
\] (19)

However, insight obtained in the fitted model was achieved by re-parameterizing the parameters for dropout in accordance with increment and sum of the successive measurements. The dropout probabilities from the dropout model in equation (18) was re-parametrized to obtain

\[
\text{logit}\left[p(D_i = j | y_{ij-1}, y_{ij})\right] = \vartheta_0 + \vartheta_1(y_{i,j} + y_{i,j-1}) + \vartheta_2(y_{ij} - y_{i,j-1}), \quad j = 2, 3, 4, 5.
\] (20)

Thus, \( \vartheta_1 = (\psi_1 + \psi_2)/2 \) and \( \vartheta_2 = (\psi_1 - \psi_2)/2 \). In the ARMD data, the parameters stated above represent the dependence on level and increment, and these quantities are likely to be much less strongly correlated than are \( y_{ij} \) and \( y_{i,j-1} \). The MNAR model is rewritten as in equation (15)

\[
\text{logit}\left[p(D_i = j | y_{ij-1}, y_{ij})\right] = -1.81 - 0.01(y_{i,j} + y_{i,j-1}) + 0.03(y_{ij} - y_{i,j-1}),
\] (21)

which indicates that the probability of dropout increases with larger negative increments. This means that patients with reduction in the overall level of the visual acuity from the previous week may have higher probability of dropping out from the experimental trial.

4.3 Multiple imputation and sensitivity analysis of the pattern mixture models

Multiple imputation, at least in its basic form [29] requires the missingness mechanism to be MAR. Our investigation commenced with the ACMV methods because it works under MAR. However, the ACMV, CCMV and NCMV analyses were imple-
mented as follows: first, the procedure fits the linear mixed model per pattern using PROC SORT and MIXED. Thereafter, MI (SAS 9.4) is used for the imputations and all the completed datasets are analyzed using the a linear-mixed model using PROC MIXED for modeling the visual acuity with the variance matrix, and conclusively the separate analysis results are combined into single one using the MIANALYZE procedure.

In Table 3, we have the results of the three types of the identifying restrictions. It can be seen that $p$-values associated with the marginal effect assessments are all nonsignificant. On the other hand, the $p$-values associated with the intercepts are highly significant. Furthermore, it is also observed that the three restrictions ACMV, CCMV and NCMV have the same values for parameters estimates and standard errors for the first marginal effect, see the effects for intercept$_4$ and treatment$_4$. Such results are expected because the first observation is observed for all the subjects considered in the analysis. Thus, the contrast between the model building of the ACMV, CCMV and NCMV and the selection model did not account for whether the dropout process is MNAR or not, because the model assumptions are different [32]. Consequently, the parameter estimates and standard errors obtain for the treatment effects by applying CCMV are smaller than those obtain for ACMV and NCMV because many completers are available. In the case of ACMV, the PMMs use data from the different patterns to multiply impute new observations, however in NCMV, the PMMs get information from the neighboring case patterns only [32].

5 Discussion

In this study, we analyzed the incomplete longitudinal data using the SM and PMM where the response variable miss across visit. We used the models to illustrate and compare the results of the analysis. Our focus is on a scenario where the responses are continuous. The focus of the study was on the special cases of SM and PMMs using [5] model and identifying restrictions strategy [16, 17]. In selection model, logistic regression was used to model the dropout. The objective was to examine the influence dropout might have on the measurement of the data and how to handle the incomplete observations. Furthermore, the results obtain from the models were compared and also given insights to the ARMD data. In the findings, a great confidence was established because the joint applications of the models display inferences that are similar. In our study, we compared and noticed similar values in the results. A challenge always arises when dealing with dropout that are MNAR. Due to this challenge, it is pertinent to be aware that this assumption increases the dropout which is not to be known in real application. Moreover, several proposed application procedures applied to handle MNAR dropout is unverifiable. In addition, no one is sure if the process of dropout is captured accurately by any of the methods applied. Due to this, several methods were to raised by [25]. In this situation, the sensitivity analysis of the parameter estimates of several mechanism about dropout process is examined [32]. If the parameter estimates are the same under different methods; this may be an indication for ignoring the dropout process. On the other hand, if different parameter estimates are obtained under different methods; this may likely mean that dropout process is an important element for describing the data in the analysis.

The structure of the selection dropout model adopted that dropout increases with a unit change in the ARMD trial [32], this reveals that dropout is in relation to the larger negative increments than the main observation. This further shows that patients with decrease in the overall level from the previous week have a higher probability of dropping out of the trial.
In this research, we proposed an application of joint modeling on ARMD; the study is a longitudinal continuous outcome where some of the patients left the study. The dropout process is modeled jointly. Thus, we extend our model by incorporating missingness in diverse ways to estimate the effect on the results. With several application of MNAR models, is the same fit with its MAR version. This indicates that the MAR counterpart properly fits the observed data. Thus, the model formulation included only random intercepts [1].

Moreover, several modifications of the MI analysis were adopted because of the assumptions behind the posterior distribution of the dropouts. These assumptions are untestable and conducting sensitivity analysis is primal. Starting with the basic MI analysis form [29] and then extended to MNAR, under a suite of the identifying restrictions [25, 36, 37], we say that the performance of the sensitivity analysis using ARMD trial, the inferences pertaining to some parameters can be modified by some identifying restrictions. From the results, we can conclude that the ACMV, CCMV and NCMV restrictions did not differ considerably. In addition, extension can be made to models with more complex data structures. This area can be further exploited for future research.
Table 3: Multiple imputation parameter estimates (Est), standard error (s.e.) and p-values of the regression coefficients for available case missing value (ACMV), complete case missing value (CCMV) and neighboring case missing value (NCMV).

| Effect          | Parameter | ACMV          | CCMV          | NCMV          |
|-----------------|-----------|---------------|---------------|---------------|
|                 |           | Est(s.e) p-value | Est(s.e) p-value | Est(s.e) p-value |
| Intercept₁₂     | β₁₁       | 54.00(1.47) -  | 54.00(1.47) -  | 54.00(1.47) -  |
| Intercept₁₂     | β₂₁       | 53.23(1.95) <.0001 | 52.89(1.61) <.0001 | 52.96(2.20) <.0001 |
| Intercept₂₄     | β₃₁       | 49.31(2.29) <.0001 | 49.45(1.79) <.0001 | 49.05(2.49) <.0001 |
| Intercept₅₄     | β₄₁       | 44.67(2.29) <.0001 | 44.69(2.35) <.0001 | 44.40(2.73) <.0001 |
| Treatment effect₁₂ | β₁₂      | -3.11(2.10) -  | -3.11(2.10) -  | -3.11(2.10) -  |
| Treatment effect₂₄ | β₂₂     | -4.73(2.69) 0.0820 | -4.26(2.36) 0.0710 | -4.56(2.71) 0.0929 |
| Treatment effect₂₄ | β₃₂     | -3.86(3.14) 0.2243 | -3.77(2.55) 0.1402 | -3.79(2.92) 0.1955 |
| Treatment effect₅₄ | β₄₂     | -4.90(2.87) 0.0893 | -4.72(2.60) 0.0693 | -4.76(2.89) 0.1008 |
REFERENCES

[1] Anna I, Molenberghs G and Verbeke G. (2016). Mechanism for missing data incorporated in joint modelling of ordinal responses. *Journal of the Royal Statistical Society Series C Applied Statistics*. DOI: 10.1111/rssc.12201

[2] Bunouf P., Grouin Jean-Marie, Molenberghs G and Thijs H. (2015). A SAS Program Combining R Functionalities to Implement Pattern-Mixture Models. *Journal of Statistics Softwares* Vol. 68:8. doi:10.18637/jss.v068.i08.

[3] Cox, D and Xing, C. (2011). *Principles of Applied Statistics*. Cambridge University Press, first edition.

[4] Daniels MJ and Hogan, JW. (2008). *Missing data in longitudinal studies*. Chapman and Hall/CRC, first edition.

[5] Diggle, P. and Kenward, MG. (1994). Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 43:49-73.

[6] Dmitrienko, A Offen, WW Faries, D Christy Chuang-Stein, JL and Molenberghs G (2005). *Analysis of Clinical Trial Data Using the SAS System*, Cary, NC:SAS Publishing.

[7] Ekholm, A and Skinner, C. (1998). The muscatine children’s obesity data reanalysed using pattern mixture models. *Applied Statistics*, 47,251-263.

[8] Galecki A and Burzykowski T. (2013). Linear mixed-effects models using R: A step-by-step approach. Springer Text in Statistics. doi:10.1007/978-1-4614-3900-4-1, Springer Science + Business Media, New York. Page 10-11.

[9] Glynn, RJ, Laird, NM and Rubin DB. (1986). Selection modeling versus mixture modeling with nonignorable nonresponse. In: *Drawing Inferences from Self-Selected Samples*. New York: Springer-verlag.

[10] Hogan, J and Laird, N. (1997). Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine*, 16,259-272.

[11] Jansen, I., Hens, N., Molenberghs G., Aerts, S., Verbeke, G and Kenward, M.K. (2006). The nature of sensitivity in missing not at random models. *Computational Statistics and Data Analysis*, 50, 830-858.

[12] Li k, Raghunathan T and Rubin D. (1991). Large sample significance levels from multiply imputed data using moment-based statistics and an F reference distribution. *Journal of the American Statistical Association*, 86(416), 125-134. doi:10.1080/01621459.1991.10475152.

[13] Linero, AR and Daniels, MJ. (2017). A general bayesian nonparametric approach for missing outcome data. www.ani.stat.fsu.edu/ arlinero/assets/documents/GeneralBNPApproach.pdf

[14] Linero, AR and Daniels MJ. (2015). A flexible bayesian approach to monotone missing data in longitudinal studies with informative dropout with application to a schizophrenia clinical trial. *Journal of the American Statistical Association*, 110(1):45-55.
[15] Linero AR. (2017). Bayesian nonparametric analysis studies in the presence of informative missingness. *Biometrika*, 104(2):327-341.

[16] Little, RJA. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Association*, 88, 125-134.

[17] Little, RJA. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika*, 81,471-483.

[18] Little, RJ. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistics Association*, 90(431):1112-1121.

[19] Little, RJA and Rubin, DB. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.

[20] Miao, W, Ding, P and Geng, Z. (2016). Identifiability of normal and normal mixture of components. *Journal of the American Statistical Association*, 111(516):1673-1683.

[21] Molenberghs, G Kenward, MG and Lesaffre, E. (1997). The analysis of longitudinal ordinal data with nonrandom drop-out. *Biometrika*, 84(1):33-44.

[22] Molenberghs, G., Michiels, B and Kenward, M. (1998a). Pseudo-likelihood for combined selection and pattern-mixture models for missing data problems. *Biometrical Journal*, 40,5557-572.

[23] Molenberghs G, Michiels B, Kenward M and Diggle P. (1998b). Missing data mechanisms and pattern-mixture models. *Statistica Neerlandica*, 52(2), 153-161. doi:10.1111/1467-9574.00075.

[24] Molenberghs G, Verbeke G, Thijs H, Lesaffre E and Kenward M. (2001). Mastitis in diary cattle: influence analysis to assess sensitivity of the dropout process. *Computational Statistics and Data Analysis*, 37, 93-113. DOI: 10.1016/S0167-9473(00)00065-7.

[25] Molenberghs G and Kenward, M. (2007). *Missing Data in Clinical Studies*. New York: Wiley.

[26] Molenberghs G., Beunckens, C., Sotto, C and Kenward, M. (2008). Every missing not at random model has got a missing at random counterpart with equal fit. *Journal of Royal Statistics Society*, 70,371-388.

[27] National research council (2010). *The prevention and treatment of missing data in clinical trials*. The National Academies Press.

[28] Rubin D. (1987). Multiple imputation for nonresponses in surveys. John Wiley and Sons. doi:10.1002/bimj.4710310118.

[29] Rubin DB. (1978). Multiple imputations in sample surveys- a phenomenological Bayesian approach to nonresponse. In: *Imputation and Editing of Faulty or Missing Survey Data*. Washington, DC: U.S. Department of Commerce, pp. 1-23. *Biometrika*, 63, 581-92.

[30] Rubin, DB. (1976). Inference and missing data. *Biometrika*, 63:581-592.
Robins, JM., Rotnitzky, A and Zhao LP. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistics Association*, 90(429):106-121.

Satty A and Mwambi H (2013). Selecting and pattern mixture models for modelling longitudinal data with dropout: An application study. *Statistics and Operations Research Transactions*, SORT 37:131-152.

Scharfstein DO., Rotnitzky, A., and Robins, JM. (1999). Adjusting for nonignorable dropout using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94.

Scharfstein D, McDermott A, Olson W and Wiegand F. (2014). Global sensitivity analysis for repeated measures studies with informative dropout: A fully parametric approach. *Statistics in Biopharmaceutical Research*, 6(4):338-348. doi:org/10.1080/19466315.2014.966920

Thijs H, Molenberghs G, Michiels B, Verbeke G and Curran D. (2002). Strategies to Fit Pattern-Mixture Models. *Biostatistics*, 3(2), 245-265. doi.org/10.1093/biostatistics/3.2.245.

van Buuren S (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research*, June; 16(3):219-242.

Verbeke G and Molenberghs G. (2000). Linear Mixed Models for Longitudinal Data. New York: Springer.

Wang C, Daniels, MJ., Scharfstein, DO and Land S. (2010). A bayesian shrinkage model for incomplete longitudinal binary data application to the breast cancer prevention trial. *Journal of the American Statistical Association*, 105:1333-1346.

Wu, MC. and Carroll, RJ. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 45:175-188.