Journal of Multivariate Analysis 166 (2018) 62–77

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Analysis of ordinal longitudinal data under nonignorable missingness and misreporting: An application to Alzheimer's disease study

Subrata Rana [a], Surupa Roy [b], Kalyan Das [c,*], for the Alzheimer's Disease Neuroimaging Initiative [1]

[a] Department of Statistics, University of Calcutta, Kolkata 700019, India
[b] Department of Statistics, St. Xavier's College, Kolkata 700016, India
[c] Department of Mathematics, IIT Bombay, Powai, Mumbai 400076, India

ARTICLE INFO

Article history:
Received 22 June 2017
Available online 26 February 2018

AMS subject classifications:
62-07
62F99

Keywords:
Bivariate binary model
MCNREM
Miscategorization
Missing data
Selection model

ABSTRACT

In many epidemiological and clinical studies, observations on individuals are recorded longitudinally on a Likert-type scale. In the process of recording, or due to some other causes, a proportion of outcomes and time-dependent covariates may be missing in one or more follow-up visits (non monotone missing). Even when the number of patients with intermittent missing data is small, exclusion of those patients from the study seems unsatisfactory. This apart, often due to misreporting, miscategorization of response can occur that results in potentially invalid inference when no correction is made. We propose a joint mixed model that corrects the likelihood function to account for missing response and/or covariates and adjusts the likelihood to tackle miscategorization of response. Under this extreme complex but useful setup, we seek to estimate the parameters of the proposed model that accounts for baseline and/or time dependent covariates. Monte Carlo expectation–maximization (MCEM) is a convenient approach for estimating the parameters in the model. A simulation study was carried out to assess the approach. We also analyzed Alzheimer's Disease Neuroimaging Initiative (ADNI) data where some responses and covariates are missing and some responses are possibly miscategorized. Our investigation reveals that apolipo-protein plays a significant role in Alzheimer's disease progression. This was not visible in earlier analyses of ADNI data.

1. Introduction

In many research problems related to biological, social and medical sciences, multivariate data arise from repeated measurements on a sample of subjects over time. To analyze such longitudinal data, one must consider the relationship between the serial observations made on a given unit and hence it is inappropriate to use a general multivariate model

* Corresponding author.
E-mail address: kalyan@math.iitb.ac.in (K. Das).

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

https://doi.org/10.1016/j.jmva.2018.02.004
0047-259X/

for studying covariate effects. A two-stage random-effect model [12] or a generalized linear model [17] can be used in this situation. Often, the response of interest is measured on an ordinal scale with more than two categories.

Methods for longitudinal ordinal data analysis have been actively pursued in the past; see, e.g., [13,15,20]. A marginalized latent variable model could be used to analyze data for this situation. A marginalized model separates the systematic variation in the data from random variation. Basically it uses two models in conjugation. One is the mean or regression model for the ordinal responses and the other is the dependence model. A cumulative logit model with proportional odds assumption is commonly used as the mean model. The covariates used in the regression model may include both qualitative and quantitative variables, some of which are measured at the baseline visit while others vary over time. In the dependence model, additional variables are introduced as random components to structure the relation between repeated measurements. Such a modeling approach was used by Heagerty [8] in the context of binary responses. Lee and Daniels [13] extended the marginalized latent variable model to accommodate ordinal responses.

The modeling of longitudinal responses through latent variables becomes complex when data on the response and some covariates go missing. A simple solution is to ignore the missing observations and perform a complete-case analysis. However this leads to inefficient inference, especially when the missing mechanism is non-ignorable. Following Little and Rubin [19], there are three types of missing data processes. Data are said to be missing completely at random (MCAR) if the missing data process does not depend on missing or observed data, and the process is said to be missing at random (MAR) if the missing data process depends on the observed data only.

In this paper we consider a non-ignorable missing data mechanism (MNAR) in which the missing data process depends on both observed and unobserved data. The MNAR pattern of missing data has been considered by Ibrahim and Lipsitz [10] in the context of generalized linear models when covariates are missing while Troxel et al. [30] and Ibrahim et al. [9] have considered missing responses. Most of the work on missing data focuses on either missing response or covariate. Stubbendick and Ibrahim [28,29] adopted a maximum likelihood approach for estimating the model parameters in a longitudinal study when both response and covariates are missing, but in their case the response belonged to the exponential family. Chen et al. [2] have considered missing response and covariate in the context of longitudinal binary data when the missing data mechanism is MAR. Here we have considered non-ignorable missingness in response as well as in a covariate.

Missing data analysis in a longitudinal set up is usually carried out under the assumption that the levels of the ordinal response are correctly classified. In medical or social sciences, however, the true level of the response is often not identified correctly. The reason for this misclassification may be misreporting by a subject or faulty diagnostic tests. For example, in a disease progression longitudinal study, there may be misreporting about a patient's disease severity at subsequent visits if the ordinal responses are collected and maintained by semi-experts. Also in the job characteristic data [25], employees were asked to respond to different aspects of the job which was measured on a five-point ordinal scale ranging from strongly agree to strongly disagree. The employees' response is supposed to be misclassified.

Miscategorization of categorical data has been considered by many researchers [4,26]. Espeland and Hui [7], Buonaccorsi [1], Pepe [24] considered a double sampling procedure to obtain the estimates of misclassification rates for discrete data, binary data and continuous data, respectively. However, only few studies are available in the literature on ordinal categorical data. Eickhoff and Amemiya [6] considered a known monotone misclassification pattern in either direction for polytomous outcome variable. Poon and Wang [25] discussed the use of a surrogate variable while modeling multivariate ordinal responses. Chen et al. [3] considered a generalized estimating equation approach while dealing with error prone ordinal responses and covariates.

In this paper we develop a flexible model to account for missing response and covariate which is also adjusted for misclassification of the observed ordinal response. The study, which is carried out under a longitudinal set up, incorporates the dynamic nature of the missingness pattern. However, it is assumed that the misclassification pattern remains the same over the visits.

The paper is organized as follows. In Section 2, we propose a flexible model and discuss the identifiability issues related with the model parameters. Section 3 describes the estimation methodology implemented via a Monte Carlo Newton–Raphson Expectation Maximization method. A simulation study is reported in Section 4 to assess the approach. In Section 5, Alzheimer's Disease Neuroimaging Initiative (ADNI) data are analyzed and finally, we conclude with some observations in Section 6.

2. Model formulation

2.1. The response process

Consider a study involving N subjects in which subject $i \in \{1, \ldots, N\}$ is assessed on $n_i \leq T$ occasions. Let $Y_i = (y_{i1}, \ldots, y_{in_i})^\top$ be a vector of L-category ordinal responses for the ith subject. Let also $Z_i = (z_{i1}, \ldots, z_{iq})^\top$ be a vector of baseline covariates and $\widetilde{X}_{i1}, \ldots, \widetilde{X}_{ip}$ be p time-varying covariate vectors, where $\widetilde{X}_{ij} = (x_{i1j}, \ldots, x_{itj}, \ldots, x_{in_ij})^\top$ for each $j \in \{1, \ldots, p\}$. We denote the $n_i \times p$ matrix of p-time varying covariates for the ith subject by $X_i = (\widetilde{X}_{i1}, \ldots, \widetilde{X}_{ip})$. The tth row of the matrix X_i contains the subject-specific time-dependent covariates which we denote by $X_{it} = (x_{it1}, \ldots, x_{itp})^\top$.

To start with, we assume that the responses and the covariates corresponding to each subject are completely observable on all occasions. We consider a marginalized latent variable model for the regression set up of the longitudinal ordinal response. Two separate models are considered in conjugation: the mean (or regression) model and the dependence model.

The systematic variation in the data is modeled through the mean model, whereas the dependence model takes into account the random variation in the data and separates it from the systematic variation. A cumulative logit model with proportional odds assumption is considered for the mean model, viz.

$$\text{logit}(P^M_{it\ell}) = \ln\left\{\frac{\Pr(y_{it} \le \ell \mid Z_i, X_{it})}{1-\Pr(y_{it} \le \ell \mid Z_i, X_{it})}\right\} = \beta_{0\ell} + Z_i^\top \beta_z + X_{it}^\top \beta_x, \tag{1}$$

where β_z and β_x are $q \times 1$ and $p \times 1$ vectors of regression coefficients corresponding to the baseline covariates and time-dependent covariates, respectively, and for each $\ell \in \{1, \ldots, L-1\}$, $\beta_{0\ell}$ is the intercept in the ℓth logit model. The monotonic relationship $\beta_{01} \le \cdots \le \beta_{0L-1}$ is assumed to hold. We denote the regression parameters of interest by $\beta = (\beta_{0\ell}^\top, \beta_z^\top, \beta_x^\top)^\top$. To capture the serial dependence among repeated observations (repeated over time), a dependence model is used, viz.

$$\text{logit}(P^C_{it\ell}) = \ln\frac{\Pr(y_{it} \le \ell \mid u_{it})}{1-\Pr(y_{it} \le \ell \mid u_{it})} = \Delta_{it\ell} + u_{it}. \tag{2}$$

In (2) the correlation among observations is induced via unobserved latent variables u_{it}. Here we assume that conditional on u_{it} the responses are independent, i.e.,

$$f(y_{i1}, \ldots, y_{in_i} | Z_i, X_{it}, u_{it}) = \prod_{t=1}^{n_i} f(y_{it} | u_{it}, Z_i, X_{it}).$$

The parameter $\Delta_{it\ell}$ represents a function of the marginal means such that the dependence model in (2) is consistent with the mean model in (1) and it can be derived from the relation

$$P^M_{it\ell} = \int P^C_{it\ell} f(u_{it}) du_{it}, \tag{3}$$

where f is a generic notation used to denote the probability density function of the random component u_{it} characterized by the relevant elements of Σ_i. As the intercepts in the mean model (1) are monotonic, $\Delta_{it\ell}$ is also monotonic in ℓ, i.e., $\Delta_{it1} \le \cdots \le \Delta_{itL-1}$ [14,31]. From (3), $\Delta_{it\ell}$ can be obtained in terms of $\beta_{0\ell}$, β_z, β_x and Σ_i. We assume that $\mathbf{U}_i = (u_{i1}, \ldots, u_{in_i})^\top$ is Gaussian with mean zero and variance–covariance Σ_i. We further structure Σ_i as $\Sigma_i = \sigma_i^2 \Sigma^*$ [14], where σ_i^2 is modeled as a function of baseline covariates as $\sigma_i^2 = e^{Z_i^\top \gamma}$ and

$$\Sigma^* = \begin{bmatrix} 1 & e^{-\lambda} & e^{-2\lambda} & \cdots & e^{-(n_i-1)\lambda} \\ e^{-\lambda} & 1 & e^{-\lambda} & \cdots & e^{-(n_i-2)\lambda} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ e^{-(n_i-1)\lambda} & e^{-(n_i-2)\lambda} & e^{-(n_i-3)\lambda} & \cdots & 1 \end{bmatrix}, \tag{4}$$

where γ and λ are unknown parameters that characterize the distribution of the random components. The integral in (3) cannot be expressed in a closed form and hence a numerical technique is commonly used to evaluate it. However in the specific case of normally distributed random components and a probit link function, the integral in (3) can be evaluated analytically [31] and $\Delta_{it\ell}$ takes a neat form given by

$$\Delta_{it\ell} = (\beta_{0\ell} + Z_i^\top \beta_z + X_{it}^\top \beta_x)\sqrt{1+\sigma_i^2}. \tag{5}$$

Even in the case of the logit link, an approximation works just as well. The idea is to approximate the logistic by the probit. For $c \cong 1.70$ it is well known that $H(\nu) = \Phi(\nu/c)$, where H and Φ are the cumulative distribution functions of a logistic and Gaussian distribution, respectively; see, e.g., [11,16,22]. Thus in this case an approximate expression for $\Delta_{it\ell}$ can be worked out as

$$\Delta_{it\ell} = (\beta_{0\ell} + Z_i^\top \beta_z + X_{it}^\top \beta_x)\sqrt{1+\sigma_i^2/c^2}. \tag{6}$$

The proof of (5) and (6) is given in Appendix A.1.

Suppressing the parameters, given the random components $\mathbf{U}_i$, the distribution of the multivariate ordinal response is given by

$$f(Y_i \mid X_i, Z_i, \mathbf{U}_i) = \prod_{t=1}^{n_i}\prod_{\ell=1}^{L}\left(P^C_{it\ell} - P^C_{it\overline{\ell-1}}\right)^{\mathbf{1}_{y_{it}}(c_{\ell t})}, \tag{7}$$

where $\mathbf{1}_{y_{it}}(c_{\ell t})$ takes the value 1 or 0 according as $y_{it} = c_{\ell t}$ or not. By integrating out the random components $\mathbf{U}_i$, we can write the joint distribution of the observed responses as

$$f(Y_i \mid X_i, Z_i) = \int_{\mathbf{U}_i}\prod_{t=1}^{n_i}\prod_{\ell=1}^{L}\left(P^C_{it\ell} - P^C_{it\overline{\ell-1}}\right)^{\mathbf{1}_{y_{it}}(c_{\ell t})} f(\mathbf{U}_i) d\mathbf{U}_i. \tag{8}$$

2.2. The missing data process

The model in Eq. (8) is based on the assumption that the responses and all covariates are observable. We now consider the situation where the response and a single time-varying covariate of an individual are missing on some occasions. We denote the observed and the missing responses for the ith individual by $Y_i^{(o)}$ and $Y_i^{(m)}$, respectively.

Without loss of generality we assume that the time-dependent covariate $\widetilde{X}_{i1}$ is missing on some occasions while observations on other time varying covariates $\widetilde{X}_{i2}, \ldots, \widetilde{X}_{ip}$ and the baseline covariates Z_i are completely available for each assessment time and for all study participants. Let us denote the observed covariate values (observed over time) corresponding to $\widetilde{X}_{i1}$ as $\widetilde{X}_{i1}^{(o)}$ and all the missing values (missing over time) as $\widetilde{X}_{i1}^{(m)}$. Further, we denote the matrix of all observed time-dependent covariates and the observed part of $\widetilde{X}_{i1}$ by $X_i^{(o)}$, i.e.,

$$X_i^{(o)} = (\widetilde{X}_{i1}^{(o)}, \widetilde{X}_{i2}, \ldots, \widetilde{X}_{ip}).$$

To indicate the availability of data, we introduce two indicators R_{it}^y and R_{it}^x, where R_{it}^y takes the value 1 or 0 according as y_{it} is observed or not. Similarly R_{it}^x takes the value 1 if x_{it1} is observed and 0 otherwise.

Now suppressing the parameters, the joint distribution of the multivariate ordinal response and the missing data indicators can be decomposed as

$$\begin{aligned} f(Y_i^{(o)}, R_i^y, R_i^x \mid X_i^{(o)}, Z_i) &= \int_{\mathbf{U}_i} \int_{\widetilde{X}_{i1}^{(m)}} \sum_{Y_i^{(m)}} f(R_i^y, R_i^x \mid Y_i^{(o)}, Y_i^{(m)}, X_i^{(o)}, \widetilde{X}_{i1}^{(m)}, Z_i) \\ &\qquad \times f(Y_i^{(o)}, Y_i^{(m)} \mid X_i^{(o)}, \widetilde{X}_{i1}^{(m)}, Z_i, \mathbf{U}_i) f(\widetilde{X}_{i1}^{(m)} \mid \widetilde{X}_{i1}^{(o)}) f(\mathbf{U}_i) d\mathbf{U}_i d\widetilde{X}_{i1}^{(m)} \\ &= \int_{\mathbf{U}_i} \int_{\widetilde{X}_{i1}^{(m)}} \sum_{Y_i^{(m)}} (I_1 \times I_2 \times I_3 \times I_4)\, d\mathbf{U}_i d\widetilde{X}_{i1}^{(m)}, \end{aligned} \tag{9}$$

where I_2 is given in Eq. (7), I_3 is the conditional density function of $\widetilde{X}_{i1}^{(m)}$ given $X_i^{(o)}$, viz. $f(\widetilde{X}_{i1}^{(m)} \mid X_i^{(o)}, \tau)$, and I_4 is the density function of the random component $\mathbf{U}_i$.

To model the joint distribution of the missing data indicators (I_1), we first indicate the histories of missing data processes by

$$\bar{R}_{it}^y = \left\{R_{i1}^y, \ldots, R_{it-1}^y\right\} \quad \text{and} \quad \bar{R}_{it}^x = \left\{R_{i1}^x, \ldots, R_{it-1}^x\right\},$$

where $t \leq n_i$. In longitudinal studies, the joint probability $\Pr(R_i^y = r_i^y, R_i^x = r_i^x \mid Y_i, X_i, Z_i)$ is not modeled directly. Instead, modeling is done by conditioning on the histories of missing data indicators by using probabilities of the type $\Pr(R_{it}^y = r_{it}^y, R_{it}^x = r_{it}^x \mid \bar{R}_{it}^y, \bar{R}_{it}^x, Y_i, X_i, Z_i)$. Such conditional modeling would reflect the dynamic nature of the observation process over time. The joint model can then be written as

$$\Pr(R_i^y = r_i^y, R_i^x = r_i^x \mid Y_i, X_i, Z_i) = \prod_{t=2}^{n_i} \Pr(R_{it}^y = r_{it}^y, R_{it}^x = r_{it}^x \mid \bar{R}_{it}^y, \bar{R}_{it}^x, Y_i, X_i, Z_i) \Pr(R_{i1}^y = r_{i1}^y, R_{i1}^x = r_{i1}^x \mid Y_i, X_i, Z_i). \tag{10}$$

For modeling the joint conditional probabilities for the pair (R_{it}^y, R_{it}^x) appearing in (10), we adopt marginal logistic regression models for the mean functions while the association between the missing data indicators in response and covariates is introduced through the conditional log odds ratio (ψ_{it}). We thus write

$$\begin{aligned} \mu_{it}^y &= \Pr(R_{it}^y = 1 \mid \bar{R}_{it}^y, \bar{R}_{it}^x, Y_i, X_i, Z_i) = \frac{\exp(\nu_{1it}^\top \alpha_y)}{1 + \exp(\nu_{1it}^\top \alpha_y)}, \\ \mu_{it}^x &= \Pr(R_{it}^x = 1 \mid \bar{R}_{it}^y, \bar{R}_{it}^x, Y_i, X_i, Z_i) = \frac{\exp(\nu_{2it}^\top \alpha_x)}{1 + \exp(\nu_{2it}^\top \alpha_x)}, \\ \psi_{it} &= \frac{\Pr(R_{it}^y = 1, R_{it}^x = 1 \mid \bar{R}_{it}^y, \bar{R}_{it}^x, Y_i, X_i, Z_i) \Pr(R_{it}^y = 0, R_{it}^x = 0 \mid \bar{R}_{it}^y, \bar{R}_{it}^x, Y_i, X_i, Z_i)}{\Pr(R_{it}^y = 1, R_{it}^x = 0 \mid \bar{R}_{it}^y, \bar{R}_{it}^x, Y_i, X_i, Z_i) \Pr(R_{it}^y = 0, R_{it}^x = 1 \mid \bar{R}_{it}^y, \bar{R}_{it}^x, Y_i, X_i, Z_i)}, \end{aligned}$$

where ν_{1it} and ν_{2it} are functions of responses and covariates. We keep the treatment general by making the conditional probabilities of missing data indicators to depend on the observed as well as unobserved responses and covariates. The missing data mechanism is thus considered to be MNAR. We denote the parameters of the missing data model by $\alpha = (\alpha_y^\top, \alpha_x^\top)^\top$. In case $\psi_{it} = 1$, the missing data indicators for the response and covariates at the tth assessment time are conditionally independent.

Let $p_{it}^{xy} = \Pr(R_{it}^y = 1, R_{it}^x = 1 \mid \bar{R}_{it}^y, \bar{R}_{it}^x, Y_i, X_i, Z_i)$ denote the probability for the pair (y_{it}, x_{it1}) to be observed, conditional on the histories of the indicator variables and the vector of responses and covariates. Following [18], p_{it}^{xy} can be expressed as

$$p_{it}^{xy} = \begin{cases} \dfrac{a_{it} - \{a_{it}^2 - 4\psi_{it}(\psi_{it} - 1)\mu_{it}^y \mu_{it}^x\}^{1/2}}{2(\psi_{it} - 1)} & \text{if } \psi_{it} \neq 1, \\ \mu_{it}^y \mu_{it}^x & \text{if } \psi_{it} = 1, \end{cases}$$

where $a_{it} = 1-(1-\psi_{it})(\mu^y_{it}+\mu^x_{it})$. Finally, the joint probability of observing the response and covariate at tth assessment time denoted by (π^{xy}_{11it}) is given by

$$\pi^{xy}_{11it} = \Pr(R^y_{it}=1, R^x_{it}=1 \mid Y_i, X_i, Z_i) = \sum_{\tilde{R}^y_{it}}\sum_{\tilde{R}^x_{it}} \Big\{ p^{xy}_{it} \times \prod_{j=2}^{t-1} (p^{xy}_{ij})^{r^y_{ij} r^x_{ij}} (\mu^y_{ij} - p^{xy}_{ij})^{r^y_{ij}(1-r^x_{ij})} (\mu^x_{ij} - p^{xy}_{ij})^{(1-r^y_{ij})r^x_{ij}} \times (1-\mu^y_{ij}-\mu^x_{ij}+p^{xy}_{ij})^{(1-r^y_{ij})(1-r^x_{ij})} \Big\}.$$

Here we assume that the covariate and response are observable for all study participants in the first visit, i.e., $\pi^{xy}_{11i1}=1$ for all $i \in \{1,\ldots,n\}$. The probabilities $\pi_{10it} = \Pr(R^y_{it}=1, R^x_{it}=0 \mid Y_i, X_i, Z_i)$, $\pi_{01it} = \Pr(R^y_{it}=0, R^x_{it}=1 \mid Y_i, X_i, Z_i)$, $\pi_{00it} = \Pr(R^y_{it}=0, R^x_{it}=0 \mid Y_i, X_i, Z_i)$ are given in Appendix A.2.

Thus, we can finally write

$$I_1 = \prod_{t=1}^{n_i} (\pi_{11it})^{r^y_{it} r^x_{it}} (\pi_{10it})^{r^y_{it}(1-r^x_{it})} (\pi_{01it})^{(1-r^y_{it})r^x_{it}} (\pi_{00it})^{(1-r^y_{it})(1-r^x_{it})}. \tag{11}$$

2.3. The misclassification process

In epidemiological studies, a serious source of error is the misclassification of the ordinal responses. A categorical variable is said to be subject to misclassification if the recorded category differs from the true category. Two types of misclassification models are available in literature — one is the classification error model and other is the reclassification model. In the classification error model, the observed responses are modeled as functions of true responses whereas in the reclassification model, the distribution of the true category given the observed category is specified. In this paper, the classification error model is considered. When the response is subject to misclassification, a surrogate version $\widetilde{y}_{it}$ is obtained instead of the true response y_{it}. Let $\widetilde{y}_{i1}=\widetilde{c}_{k1},\ldots,\widetilde{y}_{in_i}=\widetilde{c}_{kn_i}$ denote the multivariate ordinal data on the surrogate response. Further suppose

$$\Pr(\widetilde{y}_{i1}=\widetilde{c}_{k1},\ldots,\widetilde{y}_{in_i}=\widetilde{c}_{kn_i} \mid y_{i1}=c_{k1},\ldots,y_{in_i}=c_{kn_i}) = \prod_{t=1}^{n_i} \Pr(\widetilde{y}_{it}=\widetilde{c}_{kt} \mid y_{it}=c_{kt}) = \prod_{t=1}^{n_i} \epsilon_{itc_{kt}\widetilde{c}_{kt}}. \tag{12}$$

The above conditional independence assumption over the different occasions is meaningful in many practical situations. Let $\delta_{it} = ((\epsilon_{it\widetilde{r}\widetilde{r}}))$, where $r, \widetilde{r} \in \{0,\ldots,L-1\}$, denote the $L\times L$ matrix of misclassification rates. Clearly the correct classification is obtained when $r=\widetilde{r}$. Thus corresponding to each (i,t), there are $L\times(L-1)$ misclassification rates. We capture all the misclassification rates in a parameter δ. Writing $\widetilde{Y}_i = (\widetilde{y}_{i1},\ldots,\widetilde{y}_{in_i})^\top$, suppressing the parameters, the distribution of the surrogate response is given by

$$f(\widetilde{Y}_i \mid X_i, Z_i) = \int f(\mathbf{U}_i) \sum_{Y_i} \big\{ f(\widetilde{Y}_i \mid Y_i) f(Y_i \mid X_i, Z_i, \mathbf{U}_i) \big\} d\mathbf{U}_i,$$

where the first term within the curly brackets is the density function for the misclassification process derived from (12).

2.4. The joint model

We consider the modeling of longitudinal ordinal response in the presence of a non-ignorable missingness mechanism in covariate and responses along with the fact that the observed responses are imperfect versions of the true state of responses. To propose the parameter vector $\theta = (\beta^\top, \gamma, \lambda, \alpha^\top, \delta, \tau)^\top$, we propose a joint model which corrects the likelihood function in case of missing response and/or covariate and response miscategorization. Suppressing the parameters, we write the joint distribution of the surrogate response and missing data indicators given the observed covariates as follows:

$$\begin{aligned} f(\widetilde{Y}_i^{(o)}, R^y_i, R^x_i \mid X_i^{(o)}, Z_i) &= \sum_{Y_i^{(o)}} \sum_{Y_i^{(m)}} \int_{\mathbf{U}_i} \int_{\widetilde{X}_{i1}^{(m)}} f(\widetilde{Y}_i^{(o)}, R^y_i, R^x_i, Y_i^{(o)}, Y_i^{(m)}, \mathbf{U}_i, \widetilde{X}_{i1}^{(m)} \mid X_i^{(o)}, Z_i) d\widetilde{X}_{i1}^{(m)}\, d\mathbf{U}_i \\ &= \sum_{Y_i^{(o)}} \sum_{Y_i^{(m)}} \int_{\mathbf{U}_i} \int_{\widetilde{X}_{i1}^{(m)}} f(R^y_i, R^x_i \mid Y_i^{(o)}, Y_i^{(m)}, X_i^{(o)}, \widetilde{X}_{i1}^{(m)}, Z_i) f(Y_i^{(o)}, Y_i^{(m)} \mid \mathbf{U}_i, X_i^{(o)}, \widetilde{X}_{i1}^{(m)}, Z_i) \\ &\qquad f(\widetilde{X}_{i1}^{(m)} \mid X_i^{(o)}) f(\mathbf{U}_i) f(\widetilde{Y}_i^{(o)} \mid Y_i^{(o)})\, d\widetilde{X}_{i1}^{(m)}\, d\mathbf{U}_i \\ &= \sum_{Y_i^{(o)}} \sum_{Y_i^{(m)}} \int_{\mathbf{U}_i} \int_{\widetilde{X}_{i1}^{(m)}} (I_1 \times I_2 \times I_3 \times I_4 \times I_5)\, d\widetilde{X}_{i1}^{(m)}\, d\mathbf{U}_i, \end{aligned} \tag{13}$$

where I_1 is given in (11) and I_2–I_4 are as in Eq. (9) and I_5 comes from Eq. (12).

2.5. Identifiability and extra data

Identifiability is a major practical issue for misclassification problems as well as for non-ignorable missing data processes. While simultaneously estimating the regression parameters, the misclassification rates and the parameters of the missing data model, often a very large sample is needed to achieve convergence of an algorithm. In most situations, extra data in the form of multiple measurements, instrumental variables or validation studies are needed to resolve the identifiability issue.

In this paper, we use validation data to estimate the parameters of the misclassification model. We suppose that out of the cohort of N individuals who are followed up over time, n_t individuals appear at the tth visit. The response and time-specific covariate values of the absentees therefore automatically go missing. We make the practical assumption that all the subjects appearing at the tth visit will have their response variable measured, though some of their visit-specific covariate values might not be recorded. We select a random sample of n_{vt} subjects for whom the response variable is re-measured using special efforts, the values of which are considered as the gold standard. In doing so, the missing covariate values if any, of those individuals are also recorded. Thus these n_{vt} individuals have their gold standard response and all visit-specific covariate values recorded in addition to their surrogate response values. The remaining $n_{nvt} = n_t - n_{vt}$ individuals who form the non-validation group may or may not have their covariate values recorded in addition to their recorded surrogate response values. The non-validation group size is usually much larger than the validation group size, since cost of data collection per unit in the validation sample is much higher than unit cost in the non-validation sample.

3. Estimation: MCEM approach

For estimation of the parameters arising in the model given by (13), we adopt a Monte Carlo EM approach [21]. Let S_{vt} and S_{nvt} denote the validation and non-validation group at the tth visit, respectively. An individual qualifies to be in S_{vt} if he/she appears for the tth visit and subsequently gets selected in the random draw, while all other participants who may not be present or who may be present but not selected in the random draw will form the non-validation group for the tth visit. Let us denote the observed data for the validation group and non-validation group by $D_{ov} = \{Y_v, \widetilde{Y}_v, Z_v, X_v\}$ and $D_{onv} = \{\widetilde{Y}^{(o)}_{nv}, Z_{nv}, X^{(o)}_{nv}, R^y_{nv}, R^x_{nv}\}$, respectively. Suppressing the parameters the likelihood function is built up as follows:

$$\ell(\theta) = \sum_t \sum_{i\in S_{vt}} \ln f(y_{itv}, \widetilde{y}_{itv}, z_{iv}, X_{itv}) + \sum_t \sum_{i\in S_{nvt}} \ln f(\widetilde{y}_{itnv}, R^y_{itnv}, R^x_{itnv}, z_{inv}, X^{(o)}_{itnv}) = \ell_{ov}(\theta) + \ell_{onv}(\theta), \tag{14}$$

where $\ell_{ov}(\theta)$ and $\ell_{onv}(\theta)$ are the log-likelihood functions corresponding to the validation and non-validation data sets, respectively. These are further decomposed as

$$\begin{aligned}
\ell_{ov}(\theta) &= \sum_t \sum_{i\in S_{vt}} \ln f(\widetilde{y}_{itv} \mid y_{itv}) + \sum_t \sum_{i\in S_{vt}} \ln \int_{u_{itv}} f(y_{itv} \mid z_{iv}, X_{itv}, u_{itv}) f(u_{itv}) du_{itv} \\
&= \ell_{ov1}(\delta) + \ell_{ov2}(\beta, \gamma, \lambda), \\
\ell_{onv}(\theta) &= \sum_t \sum_{i\in S_{nvt}} \ln f(\widetilde{y}^{(o)}_{itnv}, R^y_{itnv}, R^x_{itnv}, z_{inv}, X^{(o)}_{itnv}) \\
&= \sum_t \sum_{i\in S_{nvt}} \ln \Bigg\{ \sum_{y^{(o)}_{itnv}} f(\widetilde{y}^{(o)}_{itnv} \mid y^{(o)}_{itnv}) \int_{u_{itnv}} \int_{\widetilde{x}^{(m)}_{it1}} \sum_{y^{(m)}_{itnv}} f(y^{(o)}_{itnv}, y^{(m)}_{itnv} \mid z_{inv}, X^{(o)}_{itnv}, x^{(m)}_{it1nv}, u_{itv}) \\
&\qquad f(R^y_{itnv}, R^x_{itnv} \mid y^{(o)}_{itnv}, y^{(m)}_{itnv}, X^{(o)}_{itnv}, x^{(m)}_{it1nv}) f(x^{(m)}_{it1nv} \mid X^{(o)}_{itnv}) f(u_{itnv}) dx^{(m)}_{it1nv} du_{itnv} \Bigg\}.
\end{aligned}$$

Maximizing $\ell(\theta)$ given in (14) is a challenging task because of the summation and multidimensional integrals appearing inside the logarithms. To circumvent this difficulty, we adopt a Monte Carlo based EM approach which derives the maximum likelihood estimates depending on the complete data log-likelihood function. We denote the complete data corresponding to the validation set by $D_{cv} = \{Y_v, \widetilde{Y}_v, Z_v, X_v, U_v\}$ and that corresponding to the non-validation set by $D_{cnv} = \{\widetilde{Y}^{(o)}_{nv}, Y^{(o)}_{nv}, Y^{(m)}_{nv}, Z_{nv}, X^{(o)}_{nv}, X^{(m)}_{nv}, R^y_{nv}, R^x_{nv}, U_{nv}\}$. Thus the complete data log-likelihood function can be decomposed as

$$\ell_c(\theta) = \ell_{cv}(\theta) + \ell_{cnv}(\theta),$$

where $\ell_{cv}(\theta)$ denotes the likelihood contribution from all subjects belonging to the validation set which can further be decomposed as

$$\begin{aligned}
\ell_{cv}(\theta) &= \sum_t \sum_{i\in S_{vt}} \ln f(\widetilde{y}_{itv} \mid y_{itv}; \delta) + \sum_t \sum_{i\in S_{vt}} \ln f(y_{itv} \mid z_{iv}, X_{itv}, u_{itv}; \beta, \gamma, \lambda) + \sum_t \sum_{i\in S_{vt}} \ln f(u_{itv}; \gamma, \lambda) \\
&= \ell_{cv1}(\delta) + \ell_{cv2}(\beta) + \ell_{cv3}(\gamma, \lambda),
\end{aligned}$$

while $\ell_{cnv}(\theta)$ denotes the likelihood contribution from all subjects belonging to the non-validation set which can be further partitioned as

$$\begin{aligned}\ell_{cnv}(\theta) &= \sum_t \sum_{i\in S_{nvt}} \ln f(\widetilde{y}^{(o)}_{itnv} \mid y^{(o)}_{itnv};\delta) + \sum_t \sum_{i\in S_{nvt}} \ln f(y^{(o)}_{itnv}, y^{(m)}_{itnv} \mid z_{inv}, X^{(o)}_{itnv}, x^{(m)}_{it\,1nv}, u_{itnv};\beta)\\ &\quad + \sum_t \sum_{i\in S_{nvt}} \ln f(R^y_{itnv}, R^x_{itnv} \mid y^{(o)}_{itnv}, y^{(m)}_{itnv}, X^{(o)}_{itnv}, x^{(m)}_{it\,1nv}, z_{inv};\alpha) + \sum_t \sum_{i\in S_{nvt}} \ln f(x^{(m)}_{it\,1nv} \mid X^{(o)}_{itnv};\tau)\\ &\quad + \sum_t \sum_{i\in S_{nvt}} \ln f(u_{itnv};\gamma,\lambda)\\ &= \ell_{cnv1}(\delta) + \ell_{cnv2}(\beta) + \ell_{cnv3}(\alpha) + \ell_{cnv4}(\tau) + \ell_{cnv5}(\gamma,\lambda).\end{aligned}$$

The estimation is carried out in two steps. In the first step, the misclassification rates are estimated from the validation sample. In the next step, the structural parameters along with the parameters of missing data model are estimated using both validation and non-validation data after plugging in the estimates of misclassification rates. The score functions are given below:

$$S_v(\delta) = \frac{\partial \ell_{cv}(\theta)}{\partial\delta} = \frac{\partial \ell_{cv1}(\theta)}{\partial\delta},$$

$$S(\alpha) = \mathrm{E}_{Y^{(o)},Y^{(m)},X^{(m)}}\left(\frac{\partial \ell_{cnv3}(\alpha)}{\partial\alpha} \,\middle|\, R^y, R^x, \widetilde{Y}^{(o)}, X^{(o)}, Z\right), \tag{15}$$

$$S(\beta) = \mathrm{E}_U\left(\frac{\partial \ell_{cv2}(\beta)}{\partial\beta} \,\middle|\, Y^{(o)}, X^{(o)}, Z\right) + \mathrm{E}_{U,Y^{(o)},Y^{(m)},X^{(m)}}\left(\frac{\partial \ell_{cnv2}(\beta)}{\partial\beta} \,\middle|\, \widetilde{Y}^{(o)}, X^{(o)}, Z, R^y, R^x\right), \tag{16}$$

$$S(\gamma) = \mathrm{E}_U\left(\frac{\partial \ell_{cv3}(\gamma,\lambda)}{\partial\gamma} \,\middle|\, Y^{(o)}, X^{(o)}, Z\right) + \mathrm{E}_U\left(\frac{\partial \ell_{cnv5}(\gamma,\lambda)}{\partial\gamma} \,\middle|\, \widetilde{Y}^{(o)}, X^{(o)}, Z, R^y, R^x\right),$$

$$S(\lambda) = \mathrm{E}_U\left(\frac{\partial \ell_{cv3}(\gamma,\lambda)}{\partial\lambda} \,\middle|\, Y^{(o)}, X^{(o)}, Z\right) + \mathrm{E}_U\left(\frac{\partial \ell_{cnv5}(\gamma,\lambda)}{\partial\lambda} \,\middle|\, \widetilde{Y}^{(o)}, X^{(o)}, Z, R^y, R^x\right), \tag{17}$$

$$S(\tau) = \mathrm{E}_{X_1^{(m)}}\left(\frac{\partial \ell_{cnv4}(\tau)}{\partial\tau} \,\middle|\, \widetilde{Y}^{(o)}, X^{(o)}, Z\right). \tag{18}$$

The expectations in (15)–(18) cannot be evaluated analytically because of multidimensional integrals or summations. To circumvent this difficulty, Metropolis–Hastings algorithm (MH) is used to produce random draws from conditional distributions. The details are given in Appendix A.3. The expectations are then approximated using Monte Carlo sums and score equations are solved iteratively by one-step Newton–Raphson method. For the parameter vector θ, the updated estimate at the $(t+1)$th step is given by

$$\theta^{(t+1)} = \theta^{(t)} - {\Psi^{(t)}}^{-1} S(\theta^{(t)}),$$

where $S(\theta^{(t)})$ is the score function at the parameter estimate $\theta^{(t)}$ and $\Psi^{(t)} = \frac{\partial S(\theta)}{\partial\theta}|_{\theta^{(t)}}$.

4. Simulation study

We consider a simulation setting with $N = 400$ study participants each assessed on $T = 6$ occasions. We generate the longitudinal ordinal response having $L = 4$ categories using the mean model and conditional model as described in Section 2.1. Here we briefly describe the different steps of data generation.

Step 1: For each $i \in \{1,\ldots,N\}$, the base line covariate z_{1i} is generated from multinomial distribution with the probability vector (0.4, 0.3, 0.3). For each $i \in \{1,\ldots,N\}$ and $t \in \{1,\ldots,T\}$, the time varying covariate x_{it} is generated from $\mathcal{N}(0,1)$. The random component $\mathbf{U}_i = (u_{i1},\ldots,u_{iT})^\top$ is simulated from T-variate Gaussian distribution with zero mean vector and dispersion matrix $\Sigma_i = \sigma_i^2\Sigma^*$, where $\sigma_i^2 = \exp(z_{2i}\gamma)$, with $\gamma = 0.4$ and z_{2i} is a binary covariate with success probability 0.5 which is not included in the response model. Moreover Σ^* is as given in Eq. (4) with $\lambda = 0.2$.

Step 2: For each $i \in \{1,\ldots,N\}$, $t \in \{1,\ldots,T\}$ and $\ell \in \{0,\ldots,L-1\}$, we compute $P^M_{it\ell}$ from the equation

$$P^M_{it\ell} = \frac{\exp(\beta_{0\ell} + \beta_1 z_{1i} + \beta_2 x_{it})}{1 + \exp(\beta_{0\ell} + \beta_1 z_{1i} + \beta_2 x_{it})},$$

where $\beta_0 = (-1, 0, 1)^\top$, $\beta_1 = -0.5$, and $\beta_2 = 0.5$.

Step 3: We compute $\Delta_{it\ell}$ from Eq. (6). For $c = 1.70$, $\Delta_{it\ell} = (\beta_{0\ell} + \beta_1 z_{1i} + \beta_2 x_{it})\sqrt{1 + \sigma_i^2/c^2}$, and hence we compute the conditional probabilities given by

$$P^C_{it\ell} = \frac{\exp(\Delta_{it\ell} + u_{it})}{1 + \exp(\Delta_{it\ell} + u_{it})}.$$

Step 4: For each $i \in \{1, \ldots, N\}$, $t \in \{1, \ldots, T\}$ and $\ell \in \{0, \ldots, L-1\}$ we compute $\widetilde{P}_{it\ell} = P^C_{it\ell} - P^C_{itl-1}$ and generate $(\eta_{it1}, \ldots, \eta_{it\ell})^\top$ from a multinomial distribution with parameters $(\widetilde{P}_{it1}, \ldots, \widetilde{P}_{it\ell})$ and $\eta_{it1} + \cdots + \eta_{it\ell} = 1$. Now if $\eta_{it\ell} = 1$ we declare $y_{it} = \ell$.

Step 5: For the misclassification process we make the simplifying assumption that the misclassification rates are independent of covariates or random components. We consider two schemes for the misclassification process, viz.

Scheme I.

$$P = \begin{pmatrix} 1-3\epsilon & \epsilon & \epsilon & \epsilon \\ \epsilon & 1-3\epsilon & \epsilon & \epsilon \\ \epsilon & \epsilon & 1-3\epsilon & \epsilon \\ \epsilon & \epsilon & \epsilon & 1-3\epsilon \end{pmatrix}.$$

Scheme II.

$$P = \begin{pmatrix} 1-\epsilon-\epsilon^2-\epsilon^3 & \epsilon & \epsilon^2 & \epsilon^3 \\ \epsilon & 1-2\epsilon-\epsilon^2 & \epsilon & \epsilon^2 \\ \epsilon^2 & \epsilon & 1-2\epsilon-\epsilon^2 & \epsilon \\ \epsilon^3 & \epsilon^2 & \epsilon & 1-\epsilon-\epsilon^2-\epsilon^3 \end{pmatrix}.$$

Scheme I implies that chances of misclassification are equally likely in all the cells while Scheme II implies that chances of misclassification decays as the observed category moves away from the true category. Different choices of ϵ were considered for both the misclassification schemes. Here we only report the results for $\epsilon = 0.1$ for Scheme I and $\epsilon = 0.2$ for Scheme II.

Step 6: We generate the missing data indicators R^y_{it}, R^x_{it} using the marginal models and the conditional odds ratio. Specifically we choose

$$\mu^y_{it} = \Pr(R^y_{it} = 1 \mid R^y_{it-1}, y_{it-1}, y_{it}) = \frac{\exp(\alpha_{y_0} + \alpha_{y_1} r^y_{\overline{it-1}} y_{it-1} + \alpha_{y_2} y_{it})}{1 + \exp(\alpha_{y_0} + \alpha_{y_1} r^y_{\overline{it-1}} y_{it-1} + \alpha_{y_2} y_{it})},$$

$$\mu^x_{it} = \Pr(R^x_{it} = 1 \mid R^x_{it-1}, x_{it-1}, x_{it}) = \frac{\exp(\alpha_{x_0} + \alpha_{x_1} r^x_{\overline{it-1}} x_{it-1} + \alpha_{x_2} x_{it})}{1 + \exp(\alpha_{x_0} + \alpha_{x_1} r^x_{\overline{it-1}} x_{it-1} + \alpha_{x_2} x_{it})},$$

where, $\alpha_{y_0} = 1.5, \alpha_{y_1} = 0.2, \alpha_{y_2} = 0.2, \alpha_{x_0} = 2, \alpha_{x_1} = 0.2, \alpha_{x_2} = 0.2$. Such a configuration results in an overall 10% missing response and 12% missing covariate. For all practical purposes we choose $\psi_{it} = \psi$ and carry out the analysis for four different choices of ψ namely 2, 4, 7 and 10. However, only the results for $\psi = 2$ are reported in Table 1.

Step 7: On the basis of the data generated in Steps 1–6, three different likelihoods are fitted. These are:

(i) The naive model (M_1): A complete case (CC) analysis is carried out by ignoring the miscategorization in response, thereby treating the surrogate response as the true values. The likelihood function in this case is given by

$$\ell_N = \prod_{i=1}^{N} \int \prod_{t \in S_i} \prod_{\ell=1}^{L} (P^C_{it\ell} - P^C_{itl-1})^{I_{\widetilde{y}_{it}}(c_{\ell t})} f(\mathbf{U}_i) d\mathbf{U}_i,$$

where S_i contains the pairs of all observable response and covariates for the ith study participant.

(ii) The missing data model (M_2): Here the likelihood is built up after incorporating the missing data mechanism, but ignoring the misclassification in response. The likelihood contribution of the ith individual is specified in Eq. (9) with $Y_i^{(o)}$ simply replaced by $\widetilde{Y}_i^{(o)}$.

(iii) The proposed model (M_3): Here the likelihood is built up after incorporating the missing data mechanism and adjusting for miscategorization as well. The likelihood contribution of the ith individual is given by Eq. (13).

Estimation is done using Monte Carlo Metropolis–Hastings Newton–Raphson (MCMHNR) method, where the data on random components are generated by Metropolis–Hastings algorithm. For simplicity and to save time, the MH sample size was chosen to be 500. The number of iterations needed in the Newton–Raphson method within Metropolis–Hastings algorithm was prefixed to be 30. The simulation was repeated $R = 100$ times. The validation sample at each visit comprises randomly selected 25% of the total subjects who have appeared in that particular visit.

For a generic parameter ξ, the parameter estimate is given by $\hat{\xi} = (\hat{\xi}_1 + \cdots + \hat{\xi}_R)/R$, where $\hat{\xi}_t$ is the estimate at the tth simulation, with $t \in \{1, \ldots, R\}$. The accuracy of the estimates is assessed by the bias and the precision of the estimator is assessed by MSE, viz.

$$\text{Bias}(\hat{\xi}) = \frac{1}{R} \sum_{t=1}^{R} \left| \hat{\xi}_t - \xi \right|, \quad \text{MSE}(\hat{\xi}) = \frac{1}{R} \sum_{t=1}^{R} (\hat{\xi}_t - \xi)^2.$$

Table 1
Table showing Bias and MSE of the estimates for the three models corresponding to two different misclassification patterns.

		Misclassification Scheme I			Misclassification Scheme II		
		M_1	M_2	M_3	M_1	M_2	M_3
β_{01}	Bias	0.04113	0.01477	0.00750	0.12728	0.08478	0.03175
	MSE	0.01484	0.00383	0.01259	0.03045	0.01112	0.01383
β_{02}	Bias	0.03277	0.02471	0.01444	0.07385	0.04638	0.01113
	MSE	0.01149	0.00383	0.01011	0.01954	0.00553	0.01150
β_{03}	Bias	0.02783	0.03634	0.00272	0.01534	0.01591	0.06571
	MSE	0.01279	0.00465	0.01318	0.01588	0.00429	0.01925
β_1	Bias	0.18555	0.17557	0.05569	0.07493	0.06316	0.02206
	MSE	0.04177	0.03311	0.00765	0.01197	0.00585	0.00666
β_2	Bias	0.20157	0.18707	0.00497	0.07522	0.07441	0.01113
	MSE	0.04405	0.03633	0.00380	0.00943	0.00718	0.00334
γ	Bias	0.04965	0.01903	0.00997	0.01269	0.00718	0.01521
	MSE	0.00680	0.00110	0.00124	0.00698	0.00108	0.00154
λ	Bias	0.00382	0.00152	0.00103	0.00114	0.00071	0.00122
	MSE	0.00005	0.00001	0.00001	0.00006	0.00001	0.00001
ϵ	Bias	–	–	0.00073	–	–	0.00072
	MSE	–	–	0.00004	–	–	0.00009
α_{y_0}	Bias	–	0.52626	0.53326	–	0.44331	0.48541
	MSE	–	0.29204	0.29451	–	0.21740	0.24482
α_{y_1}	Bias	–	0.08043	0.13189	–	0.02468	0.13227
	MSE	–	0.01025	0.01922	–	0.00559	0.01889
α_{y_2}	Bias	–	0.22898	0.19319	–	0.22525	0.16937
	MSE	–	0.05336	0.03785	–	0.05225	0.02921
α_{x_0}	Bias	–	0.00148	0.00961	–	0.01103	0.00903
	MSE	–	0.00423	0.00412	–	0.00440	0.00579
α_{x_1}	Bias	–	0.00693	0.01152	–	0.00367	0.00134
	MSE	–	0.00720	0.00506	–	0.00710	0.00318
α_{x_2}	Bias	–	0.10903	0.12333	–	0.08172	0.06073
	MSE	–	0.01279	0.01706	–	0.00806	0.00443

The bias and the MSE of the estimators under the mis-specified models M_1–M_2 and the proposed model M_3 are reported in Table 1 for the two misclassification schemes.

The result from Table 1 reveals that on the whole bias and MSE of the parameter estimates are smaller in the proposed model M_3 compared to the mis-specified models M_1 and M_2. Especially the parameter of interest, namely the covariate effects, are estimated with high accuracy and precision in the proposed model compared to the naive models. However, some isolated instances of slight poor performance in M_3 are observed. Model M_2, i.e., the model incorporating the missing data mechanism, is always better than the complete case model (M_1). For instance in misclassification pattern 2, the estimate of the intercept gives the bias and MSE as 0.06571 and 0.01925, respectively, under the corrected model M_3; these figures are higher than their respective counterparts in both models M_1 and M_2. The performance of M_2 and M_3 with respect to missing data model parameters is also almost at par, though in some instances model M_2 outperforms the proposed joint model M_3. This is natural since the proposed model is the most complicated one and involves the maximum number of parameters compared to the naive models. It is also observed that the parameters arising in the distribution of random component are not much affected by model misspecification. The misclassification rate is accurately estimated under both the misclassification schemes.

Table 2 presents the parameter estimates, simulated standard errors S.E. (sim) and coverage probabilities CP (Sim) for the regression parameter of interest for the models M_1–M_3 corresponding to both the misclassification patterns. The results reveal that M_3 recovers the parameter estimates well at the cost of increased standard error. This is natural since M_3 involves a larger number of parameters. Model M_1 performs poorly with respect to standard errors as well as coverage. The coverage probabilities under the correct model M_3 are always close to the nominal levels. Figs. 1–2 represent the convergence graphs of the parameters of interest for the proposed model under Misclassification Scheme I and Misclassification Scheme II, respectively.

5. Data study

Alzheimer's disease Neuroimaging Initiative (ADNI) is an ongoing longitudinal long-term non-treatment study in which more than 800 participants aged 55 to 90 were recruited from across more than 50 sites in the US and Canada. It includes approximately 200 patients diagnosed with early progression of Alzheimer's disease (AD). AD is the most common form of dementia which worsens with age. Usually it is diagnosed in people over 65 years of age but early onset of the disease can occur much earlier. The data used in this study were obtained from the ADNI database (http://adni.loni.ucla.edu). We have used the data from ADNI 1 for our investigation.

AD is characterized by a slow deterioration in cognitive and functional ability assessed by various clinical, biomedical, imaging and genetic biomarkers. The clinical dementia rating scale sum of boxes (CDRSB) score is a valuable tool for outcome

Table 2
Table showing parameter estimate, S.E. (sim) and coverage probability CP (sim) of the regression parameters for the three models under two different misclassification patterns.

		Misclassification Scheme I			Misclassification Scheme II		
		M_1	M_2	M_3	M_1	M_2	M_3
	Est	−1.04113	−1.01477	−1.00750	−1.12728	−1.08478	−0.96825
β_{01}	S.E. (Sim)	0.11522	0.06040	0.11253	0.11997	0.06300	0.11384
	CP (Sim)	93	96	95	84	75	94
	Est	−0.03277	−0.02471	−0.01444	−0.07385	−0.04638	−0.01113
β_{02}	S.E. (Sim)	0.10259	0.05704	0.10000	0.11929	0.05841	0.10725
	CP (Sim)	95	94	94	93	88	98
	Est	0.97217	0.96366	0.99728	1.01534	1.01591	0.93429
β_{03}	S.E. (Sim)	0.11016	0.05795	0.11533	0.12572	0.06383	0.12285
	CP (Sim)	95	92	93	92	94	88
	Est	−0.31445	−0.32443	−0.55569	−0.42507	−0.43684	−0.52206
β_1	S.E. (Sim)	0.08610	0.04808	0.06780	0.08014	0.04338	0.07899
	CP (Sim)	41	7	88	83	70	93
	Est	0.29843	0.31293	0.49503	0.42478	0.42559	0.48887
β_2	S.E. (Sim)	0.05879	0.03670	0.06172	0.06173	0.04080	0.05700
	CP (Sim)	6	0	96	77	51	98

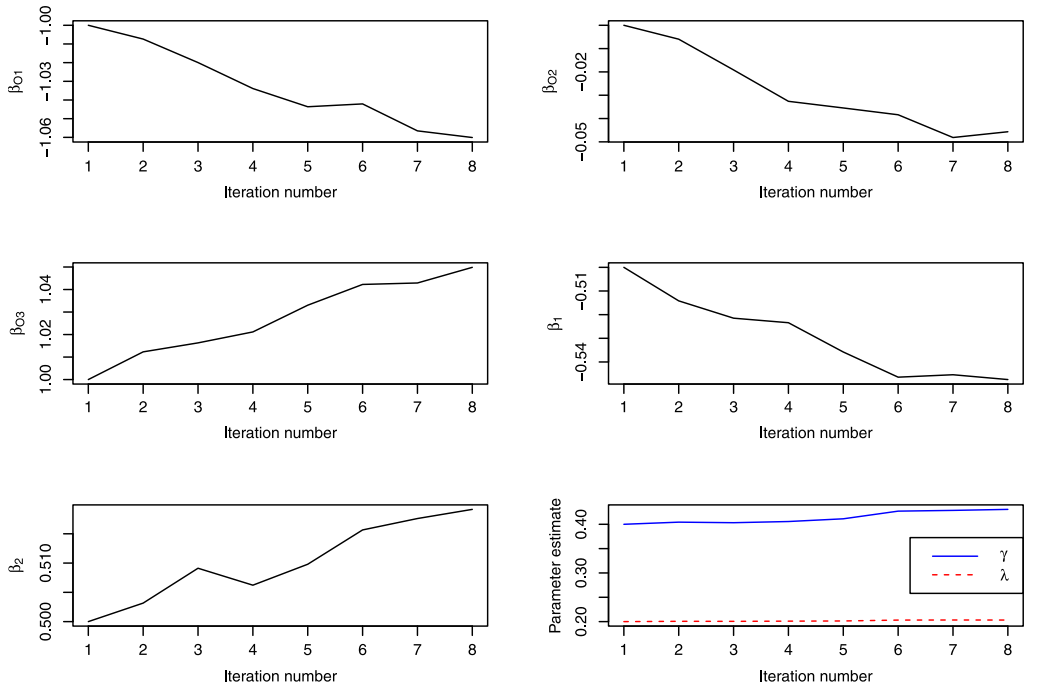

Fig. 1. Convergence graph under misclassification Scheme I.

indicator to assess both cognitive and functional impairment. Although AD develops differently for every individual, there are many common symptoms. Samtani et al. [27] modeled the baseline disease severity as a function of influential covariates.

Following the previous studies, we chose gender, apolipoprotein (APOE) $\epsilon 4$ genotype and hippocampal volume (HIPV) as the potential covariates. Among these, HIPV is considered as an important covariate in AD [5]. The hippocampus is a region of the brain that is associated primarily with memory. It is located in the inner (medial) region of the temporal lobe, forming part of the limbic system, which is particularly important in producing emotion. Individuals who suffer damage to the hippocampus experience significant memory loss or amnesia. This condition is marked by an inability to create new long-term memories. HIPV is affected by head size, age and sex, and it significantly drops in the elderly.

Data from the first six visits of ADNI 1 are considered to illustrate the methodology developed in this paper. As our missing data modeling demands that response and covariates should not be missing on the first visit, we omitted those study participants whose baseline CDRSB score and/or HIPV data were missing. We thus carried out the analysis with 376 individuals for whom CDRSB scores and/or HIPV data is missing for some occasions, while complete data on the remaining covariates are available. The study reveals 30% missing data on the covariate while approximately 14% data on the response are missing.

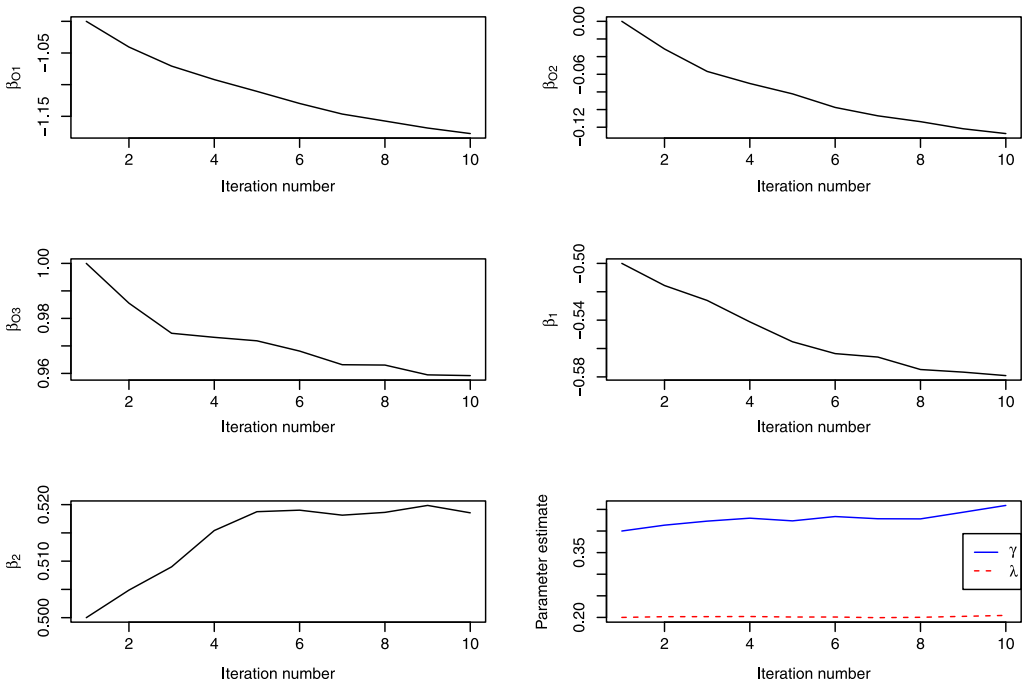

Fig. 2. Convergence graph under misclassification Scheme II.

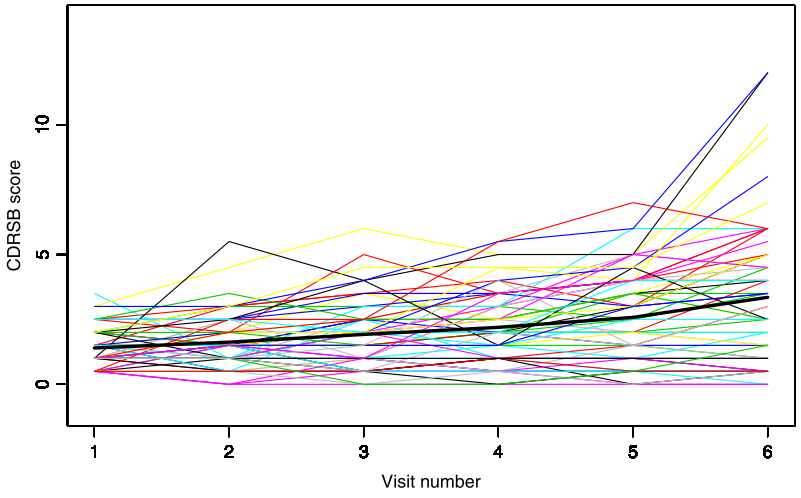

Fig. 3. Longitudinal plot of CDRSB scores.

For the available data, CDRSB scores of the study participants are plotted over the six visits in Fig. 3. The thick line gives the mean CDRSB scores which reveal an increasing trend of the disease status over time. Fig. 4 displays a bar plot of the average hippocampal volume which shows a decreasing pattern over time as is expected of patients with progression of AD.

Motivated by the work of Sid O'Bryant et al. [23], the ordinalized CDRSB scores have been categorized as 0–3 with respect to the disease status NIL, LMCI, early AD and severe AD, respectively. At the baseline visit, the CDRSB scores of all patients range between 0 and 9 indicating the absence of severe AD patients initially. In addition we work with the transformed time dependent covariate $(HIPV - M_{HIPV})/SD_{HIPV}$, where M_{HIPV} and SD_{HIPV} are the mean and standard deviations of HIPV data which were equal to 6615 and 1172, respectively. The baseline covariate (APOE) $\epsilon 4$ genotype used in the study is a nominal variable with three categories. The random component $\mathbf{U}_i = (u_{i1}, \ldots, u_{i6})^\top$ is assumed to follow a multivariate Gaussian distribution with zero mean vector and dispersion matrix $\Sigma_i = \sigma_i^2 \Sigma^*$, where Σ^* is as given in Eq. (4) and $\sigma_i^2 = \exp(Z_i\gamma)$, where $Z_i = 1$, if the ith individual is a male and 0 otherwise. The nuisance parameters arising in the distribution of the random component are estimated along with the regression parameters. For the readers' convenience, we provide a variable list in Table 3.

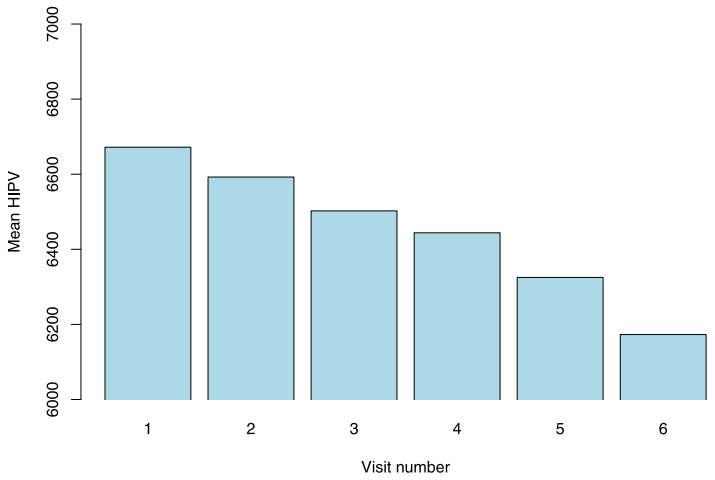

Fig. 4. Bar plot of mean hippocampal volume for 6 visits.

Table 3
Variables used in ADNI study.

Type	Variables	Description
Response (Ordinal)	CDRSB	Clinical dementia rating scale sum of boxes
Covariate (Baseline & binary)	Gender	1: Male and 0: Female
Covariate (Baseline & nominal)	APOE4	Apolipo protein genotype taking values 0, 1, 2
Covariate (Time-dependent & continuous)	HIPV	Hippocampal volume

Table 4
Misclassification matrix.

		Ordinalized CDRSB score			
		0	1	2	3
Baseline disease status	NIL	0.99537	0.00463	0	0
	LMCI	0.02767	0.78063	0.1917	0
	Early AD	0	0.02062	0.97938	0
	Severe AD	0	0	0	1

Table 5
The estimates, standard error of estimates and the t-statistic of the analysis of ADNI data.

Covariates	Complete case analysis (M_1)			Missing data analysis (M_2)			Proposed joint model (M_3)		
	Estimate	S.E.	t statistic	Estimate	S.E.	t statistic	Estimate	S.E.	t statistic
Int1	−3.0825	0.2396	−12.8659	−2.5685	0.0902	−28.5963	−1.9066	0.0724	−26.3300
Int2	2.5143	0.2010	12.5088	1.7783	0.0719	24.7434	1.4160	0.0663	21.3652
Int3	4.4876	0.4197	10.6912	3.6189	0.1146	31.5815	3.7319	0.1177	31.7126
APOE4	−0.2054	0.1618	−1.2691	−0.3896	0.0605	−6.4413	−0.2538	0.0560	−4.5331
HIPV	0.6895	0.1308	5.2725	0.5441	0.0411	13.2402	0.7433	0.0385	19.3228

Considering the data provided by ADNI as the gold standard, we observe that there are some mismatches in categorization due to discretization of the CDRSB scores. However, misclassification is prominent in the nearby categories only. We further assume that the misclassification pattern at baseline visit is carried forward in all other future visits and severe AD patients are not misclassified. The misclassification matrix is given in Table 4.

The analysis is carried out for the three models M_1–M_3 as discussed in Section 4, with known misclassification rates as given in Table 3. The ADNI data further reveal a strong association between the missing data indicators of HIPV and CDRSB scores. Driven by the data, the analysis is carried out with a known value of the association parameter chosen as $\Psi = 7$.

Table 5 reports the estimated values of the covariate effects along with their standard errors and Student t statistics for models M_1–M_3. The results indicate that the standard errors in the proposed model are lesser than that in the naive model. This is in contradiction to the simulation study findings. However, this is expected since in the simulation study, the number of nuisance parameters was larger compared to the present case, where misclassification rates and the association parameter of the missing data model are treated as known constants, chosen from the data themselves. The factor "Apolipo protein", though insignificant in the naive model, is highly significant in models M_2 and M_3. This illustrates the utility of missing data

adjusted models and the proposed joint model in lieu of naive analysis. The HIPV though significant in all three models gives extremely low p-values in the corrected models M_2 and M_3, compared to the naive model M_1.

6. Concluding remarks

The primary intention in this article was to develop a model that helps us analyze a marginal cumulative logit proportional model with missingness and miscategorization duly accounted for. Certainly the likelihood becomes complex. We consider an MCNREM approach to estimate the parameters appearing in the model. The computing time indeed depends on the dimension of the parameters. In fact as the missingness pattern is MNAR, the likelihood part involving model parameters cannot be factored out and hence we need to estimate all structural and nuisance parameters together. One possibility may be through MCMC (Markov Chain Monte Carlo) that can deal with this kind of high-dimensional situation.

Semiparametric studies are common in this kind of longitudinal framework and it would be of interest to see how the longitudinal effect can be well approximated by a non-linear function of time. Furthermore, some of the covariates (in this study may be HIPV) are often measured erroneously. Obviously, this then affects the inference in the study. It seems worthwhile to investigate situations where some of the covariates under study are subject to measurement error. In longitudinal studies, we often come across individuals who drop out from the study. Incidentally, we did not consider this here. Intermittent missing and dropout along with the miscategorization are common in medical studies. We are currently looking into this problem.

Acknowledgments

We would like to express our sincere thanks to the Reviewers, the Associate Editor and the Editor-in-Chief for their valuable suggestions and comments. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research provide funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Appendix A

A.1. Proof of Eqs. (5) and (6)

In the specific case of normally distributed random component and probit link function,

$$P^M_{it\ell} = \int P^C_{it\ell} f(u_{it}) du_{it} \Rightarrow \Phi(\beta_{0\ell} + z_i^\top \beta_z + x_{it}^\top \beta_x) = \int \Phi(\Delta_{it\ell} + u_{it}) f(u_{it}) du_{it}.$$

Now,

$$\int \Phi(\Delta_{it\ell} + u_{it}) f(u_{it}) du_{it} = \Pr\{z_{it\ell} \le \Delta_{it\ell} + u_{it} \mid u_{it} \sim \mathcal{N}(0, \sigma^2_{it})\} = \Pr\{\tau_{it\ell} \le \Delta_{it\ell} \mid u_{it} \sim \mathcal{N}(0, \sigma^2_{it})\},$$

where $\tau_{it\ell} = z_{it\ell} - u_{it}$, $\tau_{it\ell} \sim \mathcal{N}(0, 1 + \sigma^2_{it})$. Therefore,

$$\int \Phi(\Delta_{it\ell} + u_{it}) f(u_{it}) du_{it} = \Phi\left(\Delta_{it\ell} / \sqrt{1 + \sigma^2_{it}}\right).$$

Using the approximating relation between logit and probit, we get,

$$\frac{\exp(\beta_{0\ell} + z_i^\top \beta_z + x_{it}^\top \beta_x)}{1 + \exp(\beta_{0\ell} + z_i^\top \beta_z + x_{it}^\top \beta_x)} = \int \frac{\exp(\Delta_{it\ell} + u_{it})}{1 + \exp(\Delta_{it\ell} + u_{it})} f(u_{it}; \Sigma_i) du_{it}$$

$$\Rightarrow \Phi\left(\frac{\beta_{0\ell} + z_i^\top \beta_z + x_{it}^\top \beta_x}{c}\right) = \Phi\left(\frac{\Delta_{it\ell}/c}{\sqrt{1 + \sigma^2_{it}/c^2}}\right) \Rightarrow \Delta_{it\ell} = (\beta_{0\ell} + z_i^\top \beta_z + x_{it}^\top \beta_x)\sqrt{1 + \sigma^2_{it}/c^2}.$$

A.2. Detailed expression of π_{10}, π_{01} and π_{00}

Suppose, $p^{r_3 r_4}_{r_1 r_2 it} = \Pr(R^y_{it} = r_1, R^x_{it} = r_2 \mid R^y_{it-1} = r_3, R^x_{it-1} = r_4, Y_i, Z_i, X_i)$ for all $r_1, r_2, r_3, r_4 \in \{0, 1\}$. Then the unconditional probabilities unconditional on previous histories become

$$\begin{aligned}
\pi_{11it} &= p^{11}_{11it}.\pi_{11it-1} + p^{10}_{11it}.\pi_{10it-1} + p^{01}_{11it}.\pi_{01it-1} + p^{00}_{11it}.\pi_{00it-1},\\
\pi_{10it} &= p^{11}_{10it}.\pi_{10it-1} + p^{10}_{10it}.\pi_{10it-1} + p^{01}_{10it}.\pi_{01it-1} + p^{00}_{10it}.\pi_{00it-1},\\
\pi_{01it} &= p^{11}_{01it}.\pi_{11it-1} + p^{10}_{01it}.\pi_{10it-1} + p^{01}_{01it}.\pi_{01it-1} + p^{00}_{01it}.\pi_{00it-1},\\
\pi_{00it} &= p^{11}_{00it}.\pi_{11it-1} + p^{10}_{00it}.\pi_{10it-1} + p^{01}_{00it}.\pi_{01it-1} + p^{00}_{00it}.\pi_{00it-1}.
\end{aligned}$$

A.3. Metropolis–Hastings algorithm

The details of the data generation scheme using Metropolis–Hastings (MH) algorithm in Section 3 are given below.

Generating $X^{(m)(k)}$ from the conditional distribution of $X^{(m)} \mid D_{onv}, U^{(k-1)}_{nv}, Y^{(o)(k-1)}_{nv}, Y^{(m)(k-1)}_{nv}; \theta^{(t)}$: Suppressing the parameters, the conditional distribution of $X^{(m)(k)}$ is given by

$$\begin{aligned}
f(X^{(m)} \mid D_{onv}, U^{(k-1)}_{nv}, Y^{(o)(k-1)}_{nv}, Y^{(m)(k-1)}_{nv}) &\propto f(X^{(m)}, \widetilde{Y}^{(o)}_{nv}, Z_{nv}, X^{(o)}_{nv}, R^y_{nv}, R^x_{nv}, U^{(k-1)}_{nv}, Y^{(o)(k-1)}_{nv}, Y^{(m)(k-1)}_{nv})\\
&\propto f(R^y_{nv}, R^x_{nv} \mid Y^{(o)(k-1)}_{nv}, Y^{(m)(k-1)}_{nv}, X^{(o)}_{nv}, X^{(m)}, Z_{nv}) \times f(\widetilde{Y}^{(o)}_{nv} \mid Y^{(o)(k-1)}_{nv})\\
&\quad \times f(Y^{(o)(k-1)}_{nv}, Y^{(m)(k-1)}_{nv} \mid X^{(o)}_{nv}, X^{(m)}, Z_{nv}, U^{(k-1)}_{nv}) \times f(X^{(m)} \mid X^{(o)}_{nv}) \times f(U^{(k-1)}_{nv}).
\end{aligned}$$

Let us choose the proposal density for $X^{(m)}$ as $h(X^{(m)})$. Let $X^{(m)}(k-1)$ and $X^{(m)}(k)$ respectively denote the previous draw and the new values from the proposal distribution. The probability of accepting $X^{(m)}(k)$ as a potential observation from the proposal density at the kth step is given by

$$A^k_X = \min\left[\frac{f\{X^{(m)}(k) \mid D_{onv}, U^{(k-1)}_{nv}, Y^{(o)(k-1)}_{nv}, Y^{(m)(k-1)}_{nv}; \theta^{(t)}\} \times h\{X^{(m)}(k-1)\}}{f\{X^{(m)}(k-1) \mid D_{onv}, U^{(k-1)}_{nv}, Y^{(o)(k-1)}_{nv}, Y^{(m)(k-1)}_{nv}; \theta^{(t)}\} \times h\{X^{(m)}(k)\}}, 1\right].$$

In case we choose $h = f$, the acceptance function takes a neat form and is given by

$$A^k_X = \min\left[\frac{f\{R^y_{nv}, R^x_{nv} \mid Y^{(o)(k-1)}_{nv}, Y^{(m)(k-1)}_{nv}, X^{(o)}_{nv}, X^{(m)}(k), Z_{nv}\} f\{Y^{(o)(k-1)}_{nv}, Y^{(m)(k-1)}_{nv} \mid X^{(o)}_{nv}, X^{(m)}(k), Z_{nv}, U^{(k-1)}_{nv}\}}{f\{R^y_{nv}, R^x_{nv} \mid Y^{(o)(k-1)}_{nv}, Y^{(m)(k-1)}_{nv}, X^{(o)}_{nv}, X^{(m)}(k-1), Z_{nv}\} f\{Y^{(o)(k-1)}_{nv}, Y^{(m)(k-1)}_{nv} \mid X^{(o)}_{nv}, X^{(m)}(k-1), Z_{nv}, U^{(k-1)}_{nv}\}}, 1\right].$$

Generating $U^{(k)}$ from the conditional distribution of $U \mid D_{ov}; \theta^{(t)}$: Suppressing the parameters, the conditional distribution of $U|D_{ov}$ is given by

$$f(U \mid D_{ov}) \propto f(\widetilde{Y}_v, Y_v, Z_v, X_v, U) \propto f(\widetilde{Y}_v \mid Y_v) \times f(Y_v \mid X_v, Z, U) \times f(U).$$

Choosing the proposal density of U to be the same as $f(U)$, the probability of accepting a potential new value $U(k)$ as opposed to the previous value $U(k-1)$ is given by

$$A^k_U = \min\left[\frac{f\{U(k) \mid D_{ov}\} \times h\{U(k-1)\}}{f\{U(k-1) \mid D_{ov}\} \times h\{U(k)\}}, 1\right],$$

which on simplification gives

$$A^k_U = \min\left[\frac{f\{Y_v \mid X_v, Z, U(k)\}}{f\{Y_v \mid X_v, Z, U(k-1)\}}, 1\right].$$

Generating $U^{(k)}$ from the conditional distribution of $U \mid D_{onv}, X^{(m)(k)}_{nv}, Y^{(o)(k-1)}_{nv}, Y^{(m)(k-1)}_{nv}; \theta^{(t)}$: Suppressing the parameters, the conditional distribution of $U \mid D_{onv}, X^{(m)(k)}_{nv}, Y^{(o)(k-1)}_{nv}, Y^{(m)(k-1)}_{nv}$ is given by

$$\begin{aligned}
f(U \mid D_{onv}, X^{(m)(k)}_{nv}, Y^{(o)(k-1)}_{nv}, Y^{(m)(k-1)}_{nv}) &\propto f(U, \widetilde{Y}^{(o)}_{nv}, X^{(m)(k)}_{nv}, X^{(o)}_{nv}, R^y_{nv}, R^x_{nv}, Y^{(o)(k-1)}_{nv}, Y^{(m)(k-1)}_{nv}, Z_{nv})\\
&\propto f(R^y_{nv}, R^x_{nv} \mid Y^{(o)(k-1)}_{nv}, Y^{(m)(k-1)}_{nv}, X^{(o)}_{nv}, X^{(m)(k)}_{nv}, Z_{nv}) \times f(\widetilde{Y}^{(o)}_{nv} \mid Y^{(o)(k-1)}_{nv})\\
&\quad \times f(Y^{(o)(k-1)}_{nv}, Y^{(m)(k-1)}_{nv} \mid X^{(o)}_{nv}, X^{(m)(k)}_{nv}, Z_{nv}, U) \times f(X^{(m)(k)}_{nv} \mid X^{(o)}_{nv}) \times f(U).
\end{aligned}$$

Choosing the proposal density of U to be the same as $f(U)$, the probability of accepting a potential new value $U(k)$ as opposed to the previous value $U(k-1)$ is given by

$$A^k_U = \min\left[\frac{f\{U(k) \mid D_{onv}, X^{(m)(k)}_{nv}, Y^{(o)(k-1)}_{nv}, Y^{(m)(k-1)}_{nv}\} \times h\{U(k-1)\}}{f\{U(k-1) \mid D_{onv}, X^{(m)(k)}_{nv}, Y^{(o)(k-1)}_{nv}, Y^{(m)(k-1)}_{nv}\} \times h\{U(k)\}}, 1\right],$$

which on simplification gives

$$A_U^k = \min\left[\frac{f\{Y_{nv}^{(o)(k-1)}, Y_{nv}^{(m)(k-1)} \mid X_{nv}^{(o)}, X_{nv}^{(m)(k)}, Z_{nv}, U(k)\}}{f\{Y_{nv}^{(o)(k-1)}, Y_{nv}^{(m)(k-1)} \mid X_{nv}^{(o)}, X_{nv}^{(m)(k)}, Z_{nv}, U(k-1)\}}, 1\right].$$

Generating $Y^{(o)(k)}$ from the conditional distribution of $Y^{(o)} \mid D_{onv}, X_{nv}^{(m)(k)}, Y_{nv}^{(m)(k-1)}, U_{nv}^{(k)}; \theta^{(t)}$: Suppressing the parameters, the conditional distribution of $Y^{(o)} \mid D_{onv}, X_{nv}^{(m)(k)}, Y_{nv}^{(m)(k-1)}, U_{nv}^{(k)}$ is given by

$$\begin{aligned} f(Y^{(o)} \mid D_{onv}, X_{nv}^{(m)(k)}, Y_{nv}^{(m)(k-1)}, U_{nv}^{(k)}) &\propto f(Y^{(o)}, \widetilde{Y}_{nv}^{(o)}, X_{nv}^{(m)(k)}, X_{nv}^{(o)}, R_{nv}^{y}, R_{nv}^{x}, Y_{nv}^{(m)(k-1)}, U_{nv}^{(k)}, Z_{nv}) \\ &\propto f(R_{nv}^{y}, R_{nv}^{x} \mid Y^{(o)}, Y_{nv}^{(m)(k-1)}, X_{nv}^{(o)}, X_{nv}^{(m)(k)}, Z_{nv}) \times f(\widetilde{Y}_{nv}^{(o)} \mid Y^{(o)}) \\ &\quad \times f(Y^{(o)}, Y_{nv}^{(m)(k-1)} \mid X_{nv}^{(o)}, X_{nv}^{(m)(k)}, Z_{nv}, U_{nv}^{(k)}) \times f(X_{nv}^{(m)(k)} \mid X_{nv}^{(o)}) \times f(U_{nv}^{(k)}). \end{aligned}$$

Following similar lines as above, the acceptance function takes the form

$$A_{Y^{(o)k}} = \min\left[\frac{f\{Y^{(o)}(k) \mid D_{onv}, X_{nv}^{(m)(k)}, Y_{nv}^{(m)(k-1)}, U_{nv}^{(k)}\} \times h\{Y^{(o)}(k-1)\}}{f\{Y^{(o)}(k-1) \mid D_{onv}, X_{nv}^{(m)(k)}, Y_{nv}^{(m)(k-1)}, U_{nv}^{(k)}\} \times h\{Y^{(o)}(k)\}}, 1\right].$$

In case we choose $h(Y) = f(Y \mid X_{nv}^{(o)}, X_{nv}^{(m)(k)}, Z_{nv}, U_{nv}^{(k)})$, the acceptance function takes the form

$$A_{Y^{(o)k}} = \min\left[\frac{f\{R_{nv}^{y}, R_{nv}^{x} \mid Y^{(o)}(k), Y_{nv}^{(m)(k-1)}, X_{nv}^{(o)}, X_{nv}^{(m)(k)}, Z_{nv}\} f\{\widetilde{Y}_{nv}^{(o)} \mid Y^{(o)}(k)\}}{f\{R_{nv}^{y}, R_{nv}^{x} \mid Y^{(o)}(k-1), Y_{nv}^{(m)(k-1)}, X_{nv}^{(o)}, X_{nv}^{(m)(k)}, Z_{nv}\} f\{\widetilde{Y}_{nv}^{(o)} \mid Y^{(o)}(k-1)\}}, 1\right].$$

Generating $Y^{(m)(k)}$ from the conditional distribution of $Y^{(m)} \mid D_{onv}, X_{nv}^{(m)(k)}, Y_{nv}^{(o)(k)}, U_{nv}^{(k)}; \theta^{(t)}$: Suppressing the parameters, the conditional distribution of $Y^{(m)} \mid D_{onv}, X_{nv}^{(m)(k)}, Y_{nv}^{(o)(k)}, U_{nv}^{(k)}$ is given by

$$\begin{aligned} f(Y^{(m)} \mid D_{onv}, X_{nv}^{(m)(k)}, Y_{nv}^{(o)(k)}, U_{nv}^{(k)}) &\propto f(Y^{(m)}, \widetilde{Y}_{nv}^{(o)}, X_{nv}^{(m)(k)}, X_{nv}^{(o)}, R_{nv}^{y}, R_{nv}^{x}, Y_{nv}^{(o)(k)}, U_{nv}^{(k)}, Z_{nv}) \\ &\propto f(R_{nv}^{y}, R_{nv}^{x} \mid Y^{(m)}, Y_{nv}^{(o)(k)}, X_{nv}^{(o)}, X_{nv}^{(m)(k)}, Z_{nv}) \times f(\widetilde{Y}_{nv}^{(o)} \mid Y_{nv}^{(o)(k)}) \\ &\quad \times f(Y^{(m)}, Y_{nv}^{(o)(k)} \mid X_{nv}^{(o)}, X_{nv}^{(m)(k)}, Z_{nv}, U_{nv}^{(k)}) \times f(X_{nv}^{(m)(k)} \mid X_{nv}^{(o)}) \times f(U_{nv}^{(k)}). \end{aligned}$$

Following similar lines as above, the acceptance function takes the form

$$A_{Y^{(m)k}} = \min\left[\frac{f\{Y^{(m)}(k) \mid D_{onv}, X_{nv}^{(m)(k)}, Y_{nv}^{(o)(k)}, U_{nv}^{(k)}\} \times h\{Y^{(m)}(k-1)\}}{f\{Y^{(m)}(k-1) \mid D_{onv}, X_{nv}^{(m)(k)}, Y_{nv}^{(o)(k)}, U_{nv}^{(k)}\} \times h\{Y^{(m)}(k)\}}, 1\right].$$

In case we choose $h(Y) = f(Y \mid X_{nv}^{(o)}, X_{nv}^{(m)(k)}, Z_{nv}, U_{nv}^{(k)})$, the acceptance function takes the form

$$A_{Y^{(m)k}} = \min\left\{\frac{f\{R_{nv}^{y}, R_{nv}^{x} \mid Y^{(m)}(k), Y_{nv}^{(o)(k)}, X_{nv}^{(o)}, X_{nv}^{(m)(k)}, Z_{nv}\}}{f\{R_{nv}^{y}, R_{nv}^{x} \mid Y^{(m)}(k-1), Y_{nv}^{(o)(k)}, X_{nv}^{(o)}, X_{nv}^{(m)(k)}, Z_{nv}\}}, 1\right\}.$$

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2018.02.004.

References

[1] J.P. Buonaccorsi, Double sampling for exact values in the normal discriminant model with application to binary regression, Comm. Statist. Theory Methods 19 (1990) 4569–4586.
[2] B. Chen, G.Y. Yi, R.J. Cook, Weighted generalized estimating functions for longitudinal response and covariate data that are missing at random, J. Amer. Statist. Assoc. 105 (2010) 336–353.
[3] Z. Chen, G.Y. Yi, C. Wu, Marginal analysis of longitudinal ordinal data with misclassification in both response and covariates, Biometrical J. 56 (2014) 69–85.
[4] R.P. Chiacchierini, J.C. Arnold, A two-sample test for independence in 2 × 2 contingency tables with both margins subject to misclassification, J. Amer. Statist. Assoc. 72 (1977) 170–174.
[5] I. Delor, J.-E. Charoin, R. Gieschke, S. Retout, P. Jacqmin, Modeling Alzheimer's disease progression using disease onset time and disease trajectory concepts applied to CDR-SOB scores from ADNI, Pharmacomet. Syst. Pharmacol. 2 (2013) 1–10.
[6] J.C. Eickhoff, Y. Amemiya, Latent variable models for misclassified polytomous outcome variables, British J. Math. Statist. Psych. 58 (2005) 359–375.
[7] M.A. Espeland, S.L. Hui, A general approach to analyzing epidemiologic data that contain misclassification errors, Biometrics 43 (1987) 1001–1012.
[8] P.J. Heagerty, Marginally specified logistic-normal models for longitudinal binary data, Biometrics 55 (1999) 688–698.
[9] J.G. Ibrahim, M.H. Chen, S.R. Lipsitz, Weighted estimating equations with nonignorably missing response data, Biometrika 88 (2001) 551–564.
[10] J.G. Ibrahim, S.R. Lipsitz, Missing covariates in generalized linear models when the missing data mechanism is non-ignorable, J. R. Stat. Soc. Ser. B Stat. Methodol. 61 (1999) 173–190.

[11] N.L. Johnson, S. Kotz, Distribution in Statistics, Wiley, New York, 1970.
[12] N.M. Laird, J.H. Ware, Random-effects models for longitudinal data, Biometrics 38 (1982) 963–974.
[13] K. Lee, M.J. Daniels, A class of Markov models for longitudinal ordinal data, Biometrics 63 (2007) 1060–1067.
[14] K. Lee, M.J. Daniels, Marginalized models for longitudinal ordinal data with application to quality of life studies, Stat. Med. 27 (2008) 4359–4380.
[15] E. Lesaffre, G. Molenberghs, L. Dewulf, Effect of dropouts in a longitudinal study: An application of a repeated ordinal model, Stat. Med. 15 (1996) 1123–1141.
[16] K.Y. Liang, X.H. Liu, Estimating equations in generalized linear models with measurement error, in: Estimating Functions, 1991, pp. 47–63.
[17] K.Y. Liang, S.L. Zeger, Longitudinal data analysis using generalized linear models, Biometrika 73 (1986) 13–22.
[18] S.R. Lipsitz, N.M. Laird, D.P. Harrington, Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association, Biometrika 78 (1991) 153–160.
[19] R.J.A. Little, D.B. Rubin, Statistical Analysis with Missing Data, Wiley, New York, 1987.
[20] L.C. Liu, D. Hedeker, A mixed effects regression model for longitudinal multivariate ordinal data, Biometrics 62 (2006) 261–268.
[21] C.E. McCulloch, Maximum likelihood algorithms for generalized linear mixed models, J. Amer. Statist. Assoc. 92 (1997) 162–170.
[22] J.F. Monahan, L.A. Stefanski, Normal scale mixture approximations to $F^*(x)$ and computation of the logistic-normal integral, in: N. Balakrishnan (Ed.), Handbook of the Logistic Distribution, CRC Press, London, 1992, pp. 529–540.
[23] S.E. O'Bryant, S.C. Waring, C.M. Cullum, J. Hall, L. Lacritz, P.J. Massman, P.J. Lupo, J.S. Reisch, R. Doody, Texas Alzheimer's Research Consortium, Staging dementia using clinical dementia rating scale sum of boxes scores, Arch. Neurol. 65 (2008) 1091–1095.
[24] M.S. Pepe, Inference using surrogate outcome data and a validation sample, Biometrika 79 (1992) 355–365.
[25] W.Y. Poon, H.B. Wang, Bayesian analysis of multivariate probit models with surrogate outcome data, Psychometrika 75 (2010) 498–520.
[26] S.J. Press, Estimating from misclassified data, J. Amer. Statist. Assoc. 63 (1968) 123–133.
[27] M.N. Samtani, M. Farnum, V. Lobanov, E. Yang, N. Raghavan, A. Dibernardo, V. Narayan, Alzheimer's Disease Neuroimaging Initiative, An improved model for disease progression in patients from the Alzheimer's Disease Neuroimaging Initiative, J. Clin. Pharmacol. 52 (2012) 629–644.
[28] A.L. Stubbendick, J.G. Ibrahim, Maximum likelihood methods for nonignorable missing responses and covariates in random effects models, Biometrics 59 (2003) 1140–1150.
[29] A.L. Stubbendick, J.G. Ibrahim, Likelihood-based inference with nonignorable missing responses and covariates in models for discrete longitudinal data, Statist. Sinica 16 (2006) 1143–1167.
[30] A.B. Troxel, S.R. Lipsitz, T.A. Brennan, Weighted estimating equations with nonignorably missing response data, Biometrics 53 (1997) 857–869.
[31] N. Vahabi, A. Kazemnejad, S. Datta, A joint overdispersed marginalized random-effects model for analyzing two or more longitudinal ordinal responses, Stat. Methods Med. Res. (2017). http://dx.doi.org/10.1177/0962280217714616.