Published in final edited form as:

Stat Med. 2013 October 30; 32(24): 4180–4195. doi:10.1002/sim.5812.

Bayesian Modeling and Inference for Clinical Trials with Partial Retrieved Data Following Drop-out

Qingxia Chen^a, Ming-Hui Chen^b, David Ohlssen^c, and Joseph G. Ibrahim^{d,*}

^aDepartments of Biostatistics and Biomedical Informatics, Vanderbilt University, Nashville, Tennessee, 37232, U.S.A.

^bDepartment of Statistics, University of Connecticut, Storrs, Connecticut, 06269, U.S.A.

^cNovartis Pharmaceuticals Corporation, East Hanover, New Jersey, 07936, U.S.A.

^dDepartment of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, U.S.A.

Abstract

In randomized clinical trials, it is common that patients may stop taking their assigned treatments and then switch to a standard treatment (standard of care available to the patient) but not the treatments under investigation. While the availability of limited retrieved data on patients who switch to standard treatment, called off-protocol data, could be highly valuable in assessing the associated treatment effect with the experimental therapy, it leads to a complex data structure requiring the development of models that link the information of per-protocol data with the off-protocol data. In this paper, we develop a novel Bayesian method to jointly model longitudinal treatment measurements under various dropout scenarios. Specifically, we propose a multivariate normal mixed-effects model for repeated measurements from the assigned treatments and the standard treatment, a multivariate logistic regression model for those stopping the assigned treatments, logistic regression models for those starting a standard treatment off protocol, and a conditional multivariate logistic regression model for completely withdrawing from the study. We assume that withdrawing from the study is non-ignorable but intermittent missingness is assumed to be at random. Various properties of the proposed model are examined. An efficient Markov chain Monte Carlo sampling algorithm is developed. A real data set from a clinical trial is analyzed in detail via the proposed method.

Keywords

Intermittent Missingness; Logistic Regression Model; Markov Chain Monte Carlo; Missing At Random; Multivariate Mixed-effects Model

1. Introduction

In a drug development setting, research and debate on improving the handling of missing data in clinical trials has often focused on demonstrating the advantage of model-based approaches that aim to employ a clear set of assumptions, such as using a mixed model in a longitudinal analysis that implicitly assumes either missing at random assumption (MAR) or missing not at random (MNAR), over the more traditional approaches that are based on single data imputations, including last observation carried forward (LOCF) and baseline observation carried forward (BOCF) ([1, 2, 3, 4]). As the LOCF and BOCF have been applied in the primary analysis of the vast majority of confirmatory phase III trials, the arguments for those developments are well motivated. However, recent work by the National Research Council Panel on Handling Missing Data in Clinical Trials emphasizes the need to combine better methods for the prevention of missing data with well principled modeling strategies (Panel on Handling Missing Data in Clinical Trials and National Research Council [5]).

Here focus shall be placed on the third recommendation of the National Research Council that trial sponsors should continue to collect information on key outcomes on participants who discontinue their protocol-specified intervention in the course of the study. If improvements in data retrieval strategies are to be made, the likely consequence is a final analysis dataset consisting of patients with at least three types of follow-up: 1) patients with complete data; 2) patients who drop-out of the study at a certain time point and no further data is collected; 3) patients who drop-out but a certain amount of data is retrieved following withdrawal from study medication when it is ethical. Up until now, little thought has gone into appropriate statistical analysis for such “mixed data” structures.

The literature on missing data methods for longitudinal studies is enormous and far too numerous to list here. There have been many likelihood-based approaches as well as weighted estimating approaches proposed in the literature. Estimating parameters with nonignorable missing data is complex. Likelihood-based methods require specification of the joint distribution of the data and the missing data mechanism. This specification can be further classified into three types of models: selection, pattern-mixture, and shared-parameter models [13]. The selection modeling approach models the hypothetical complete data together with the missing data process conditional on the hypothetical complete data. The pattern-mixture approach models the distribution of the data conditional on the missing data pattern. The third approach, shared-parameter models, accounts for the dependence between the measurement and missingness processes by means of latent variables such as random effects ([6, 7, 8, 9]). The literature on maximum likelihood estimation with nonignorable missing longitudinal data models includes [10] and [11]. The main emphasis is on continuous and normally distributed outcomes. A substantial part of the literature also assumes monotone patterns of missingness, where sequences of measurements on some subjects simply terminate prematurely. Approaches using selection models include [12, 13, 14] among others. Approaches based on pattern-mixture models include [13, 15, 16, 17, 18]. A comprehensive review paper on missing data in longitudinal studies is given in [19] and contains many references. We mention here that although there has been a substantial literature for longitudinal data with nonignorable missing outcomes, there has been virtually

no literature at all for modeling and accommodating longitudinal “off-protocol data”, that is modeling the longitudinal process while patients are on the assigned treatment, and then also jointly modeling the longitudinal process when patients stop taking the assigned treatment, and then start taking a “standard treatment”, which is the term we shall use for the standard of care available to a patient once they stop taking the treatment assigned to them in the clinical trial. Related but different challenges can also be found in the literature of noncompliance, in which causal inference has been shown to be a useful inference tool ([20, 21]). In the noncompliance studies, it is typically assumed that the patients would switch to the other treatment arm under investigation when noncompliant, however, this is not the situation for our motivating study, in which the “standard-treatment” is different from the treatments under investigation. Modeling such “off-protocol data” poses many new statistical challenges not addressed in the literature. First, there is the issue of how to model the response and missingness mechanisms for the longitudinal component for those on the assigned treatment and those on the standard treatment. Second, there is the issue of defining an intent-to-treat analysis. Third, there is also the critical issue of model identifiability which needs to be addressed. In this paper, we address these issues and present a novel model for missing longitudinal data that accommodates a longitudinal structure for patients taking the assigned experimental treatment as well as patients taking a standard treatment after stopping the assigned treatment upon dropping out of the study.

The remainder of this paper is structured as follows. Section 2 introduces a motivating example based on a trial examining an experimental treatment for mild cognitive impairment (MCI). In Section 3 we develop a methodological framework that is later applied to the MCI trial in Section 5. Prior and posterior distributions are presented in Section 4, and in the same section, an efficient Gibbs sampling algorithm is developed. Finally, the ideas and limitations is discussed along with various possible extensions in Section 6.

2. The Mild Cognitive Impairment Study

To help exemplify and develop methods for the “mixed data structure” described in Section 1, a Phase III randomized, multicenter, double-blind, placebo-controlled study is considered, that examined the effect of the experimental treatment Exelon on subjects with mild cognitive impairment (MCI). Although MCI can present with a variety of symptoms, when memory loss is the predominant symptom it is termed “amnesic MCI” and is frequently seen as a risk factor for Alzheimer’s disease [22]. At the time of the study, the experimental treatment Exelon was already approved for the treatment of Alzheimer’s disease (AD). Therefore, it was hypothesized that treatment of subjects with MCI might delay progression of amnesic MCI to AD due to the induction of symptomatic improvements. One of the co-primary endpoints for this study is the change from baseline in the long-term effects of taking Exelon versus placebo on a cognitive function score made up of a battery of cognitive tests.

As described in the protocol, the enrolled patients had regular visits every six months until the 36th month, corresponding to visits 1 to 6. At each visit, the cognitive function score was measured and then transformed to a continuous composite z -score. In the trial, investigators

encouraged subjects who discontinued study treatment to continue coming to the clinic for regularly scheduled 6-month visits, until the planned end of the study. Therefore, to clarify the difference, we describe the score as a per-protocol outcome if it was measured when the patient was taking the assigned treatment, and as an off-protocol outcome if it was measured after the patient discontinued their assigned treatment and returned to standard of care. It is noted that other types of protocol violations could occur, here to simplify the presentation, we shall only focus on withdrawal from a randomized study treatment. The per- and off-protocol schedule is complex in this study. To illustrate, we listed some selected per- and off-protocol patterns together with one of its associated outcome measurements in the top section of Figure 1. We called the patients “Return Patients” if they ever returned to the clinic after discontinuing the assigned treatment “Non-return Patients” if they never came back to the clinic after discontinuing the assigned treatment, and “Never-stopped Patients” if they never discontinued the assigned treatment. In the bottom section of Figure 1, we presented the boxplots of the per-protocol outcome measurements for “Return Patients” (top), “Non-return Patients” (top middle), and “Never-stopped Patients” (bottom middle), and the off-protocol outcome measurements for “Return Patients” (bottom) only. No clinically meaningful difference between the two treatments was observed from the graph.

Given the longitudinal follow-up, it is interesting to examine longitudinal models that incorporate information on subjects with retrieved data, to form more appropriate assumption about subjects who dropped out with no further retrieved data. In the following section, a general methodological framework is developed for this type of modeling.

3. Methodological Development

3.1. Notation

To accommodate the longitudinal process for patients on the assigned treatment, standard treatment, and study dropout, we need to define three random variables, R , S , and L , and hence develop a regression model for each variable. Let r_i be the time point at which the i th patient stops taking the assigned treatment, and thus r_i is a categorical response variable. The possible values of r_i include $1, \dots, K, K+1$, where $r_i = K+1$ implies that the i th patient never switches the treatment and thus never drops out from the study. Let s_i be the time point at which i th patient starts the standard treatment. The possible values of s_i include $1, \dots, K, K+1$ with the constraint that $s_i \geq r_i$, where $s_i = K+1$ implies that the i th patient never starts the standard treatment. Let ℓ_i be the time point at which a patient completely withdraws from the study. As in the case of ℓ_i , the possible values of ℓ_i include $1, \dots, K, K+1$, under the restriction that $\ell_i \geq r_i$, and $\ell_i = K+1$ implies that the i th patient never drops out from the study. For example, for a patient with $r_i = s_i = 3$ and $\ell_i = 5$, he/she switched from assigned treatment to standard treatment between the second and third visits, and then dropped out and did not come to the fifth visit.

The variables r_i , s_i and ℓ_i categorize the patients into four groups:

1. patients stick to the planned treatment for the whole study period (never-stopped patients): $r_i = s_i = \ell_i = K+1$;

2. patients discontinue the study treatment but still come back to the clinic for all scheduled visits until the end of study (return patients without dropout): $r_i = s_i < K + 1$ and $\ell_i = K + 1$;
3. patients discontinue the study treatment but still come back to the clinic for some scheduled visits (return patients with dropout): $r_i = s_i < \ell_i < K + 1$; and
4. patients discontinue the study treatment and never come back to the clinic for regularly scheduled visits (never-returned patients): $r_i = \ell_i < K + 1$ and $s_i = K + 1$.

Due to the nature of clinical trials, the patients switching to the standard treatment will not resume their assigned treatments. The following two patterns hold in general regarding variables r_i , s_i and ℓ_i : (i) if $r_i = K + 1$, then $s_i = \ell_i = K + 1$; (ii) if $r_i < K + 1$ and $s_i = K + 1$, then $\ell_i = r_i$. Another pattern observed in the MCI study and required by our proposed approach, but may not hold in general is:

Assumption A.1: If $r_i < K + 1$, then $s_i = r_i$ or $s_i = K + 1$.

We called this seamless assumption. It assumes that if the patients stop taking the assigned treatment, they will either start the standard treatment right away without missing the visit or drop out from the study completely. This assumption leads to another general pattern: (iii) if $r_i < K + 1$ and $s_i < K + 1$, then $s_i = r_i$ and $s_i + 1 = \ell_i = K + 1$. The proposed method is developed under the Assumption A.1 and the patterns (i) to (iii).

We let y_{ik} denote the observed continuous primary outcome, which is the outcome measured from the assigned treatment when $k < r_i$ and the outcome measured from the standard treatment when $k = r_i$. Let z_i denote the binary treatment covariate at baseline, where $Z_i = 1$ indicates a patient is assigned to the experimental group and $z_i = 0$ indicates a patient is assigned to control, and let $\mathbf{x}_i$ be a column vector of other baseline covariates for the i th patient. We also let g_{ik} denote the missing indicator for outcome measurement such that $g_{ik} = 1$ if y_{ik} is missing and $g_{ik} = 0$ if y_{ik} is observed. When $g_{ik} = 1$ and $k < r_i$, y_{ik} from the assigned treatment is intermittently missing. Similarly, when $g_{ik} = 1$ and $s_i - k < \ell_i$, y_{ik} from the standard treatment is intermittently missing. Furthermore, we make the following assumption for notation convenience.

Assumption A.2: If $g_{ik} = 1$ and $\ell_i - k < s_i$, the missing value y_{ik} is from the standard treatment.

In other words, we assume for those never-returned patients, if the outcome measurements after dropout were observed, they were measured from standard treatment.

3.2. The Model

We use $[A|B]$ to denote the conditional distribution of A given B . We will model the observed data through a sequence of conditional distribution functions $[R_i] [Y_i|R_i] [S_i|R_i, Y_i] [L_i|R_i, Y_i, S_i]$.

We first model $[R_i]$. Since R_i , the time point at which the i th patient stops taking the assigned treatment, is a categorical response variable taking $K + 1$ distinct values, we propose a multivariate logistic regression model for R_i as follows:

$$f_R(r_i | z_i, \mathbf{x}_i, \boldsymbol{\theta}) = P(R_i = r_i | z_i, \mathbf{x}_i, \boldsymbol{\theta}) = \begin{cases} \frac{\exp\{\theta_{0r_i} + \theta_{1r_i} z_i + \boldsymbol{\theta}'_{2r_i} \mathbf{x}_i\}}{1 + \sum_{k=1}^K \exp\{\theta_{0k} + \theta_{1k} z_i + \boldsymbol{\theta}'_{2k} \mathbf{x}_i\}}, & \text{if } r_i = 1, \dots, K \\ \frac{1}{1 + \sum_{k=1}^K \exp\{\theta_{0k} + \theta_{1k} z_i + \boldsymbol{\theta}'_{2k} \mathbf{x}_i\}}. & \text{if } r_i = K + 1. \end{cases} \quad (3.1)$$

A multivariate normal mixed effects model is proposed for the observed continuous outcome y_{ik} given r_i . Because the treatment effects will be included in the both measurement types at the k th visit of time t_k , the model must be structured in a way that allows synthesis of the treatment effects, in-turn producing an overall treatment effect. Taking these factors into consideration, the model for $\mathbf{y}_i = (y_{i1}, \dots, y_{iK})$ assumes:

$$y_{ik} = \begin{cases} v_i + \gamma_{10k} + \gamma_{11k} z_i + \boldsymbol{\beta}'_1 \mathbf{x}_i + \varepsilon_{i1k}, & \text{if } k < r_i \\ v_i + \gamma_{20k} + \gamma_{21k} z_i + \boldsymbol{\beta}'_2 \mathbf{x}_i + \varepsilon_{i2k}, & \text{if } k \geq r_i \end{cases} \quad (3.2)$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the vectors of regression coefficients corresponding to the baseline covariates $\mathbf{x}_i$ and

$$v_i \stackrel{\text{i.i.d.}}{\sim} N(0, \tau^2). \quad (3.3)$$

Let $\boldsymbol{\varepsilon}_{i1} = (\varepsilon_{i11}, \dots, \varepsilon_{i1K})'$ and $\boldsymbol{\varepsilon}_{i2} = (\varepsilon_{i21}, \dots, \varepsilon_{i2K})'$ denote the vectors of error terms in (3.2). We assume $\boldsymbol{\varepsilon}_{i1}$ and $\boldsymbol{\varepsilon}_{i2}$ are independent,

$$\boldsymbol{\varepsilon}_{i1} \stackrel{\text{i.i.d.}}{\sim} N(\mathbf{0}, \Sigma_1), \text{ and } \boldsymbol{\varepsilon}_{i2} \stackrel{\text{i.i.d.}}{\sim} N(\mathbf{0}, \Sigma_2), \quad (3.4)$$

where $\Sigma_j = (\sigma_{jkk})$ is a $K \times K$ unstructured variance-covariance matrix for $j = 1, 2$. We note that in (3.2), the subject level random effects v_i , $i = 1, \dots, n$, are assumed to be common across the measurements from the assigned and standard treatments. Note that unlike the regular longitudinal study, τ , Σ_1 , and Σ_2 are estimable since τ can be estimated from patients receiving both the assigned and standard treatments, while the patients with more than one visit when they are taking the assigned treatment and more than one visit when they are taking standard treatment are used to estimate Σ_1 and Σ_2 , respectively. The inclusion of z_i in the y_{ik} model when $k < r_i$ could account for possible carry-over effects. Under the model in (3.2), given the subject random effect v_i , the outcome measurements from the assigned treatment and standard treatment are assumed to be independent and, however, they are dependent across different visits when receiving the same treatment. Let $\boldsymbol{\gamma}_j = (\gamma_{j01}, \dots, \gamma_{j0K}, \gamma_{j11}, \dots, \gamma_{j1K})'$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}'_1, \boldsymbol{\gamma}'_2)'$, $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2)'$, $\boldsymbol{\Sigma} = (\Sigma_1, \Sigma_2)$, and partition Σ_j as

$$\Sigma_j = \begin{pmatrix} \Sigma_{j,1:k,1:k} & \Sigma_{j,1:k,(k+1):K} \\ \Sigma_{j,(k+1):K,1:k} & \Sigma_{j,(k+1):K,(k+1):K} \end{pmatrix},$$

for $j = 1, 2$. Then, conditional on the subject level random effect v_i and r_i , the distribution of y_i is given by

$$f(\mathbf{y}_i | z_i, \mathbf{x}_i, r_i, v_i; \boldsymbol{\gamma}, \boldsymbol{\beta}, \Sigma) = \frac{|W_i|^{-1/2}}{(2\pi)^{K/2}} \exp \left\{ -\frac{1}{2} (\mathbf{y}_i - \boldsymbol{\mu}_i)' W_i^{-1/2} (\mathbf{y}_i - \boldsymbol{\mu}_i) \right\}, \quad (3.5)$$

where

$\boldsymbol{\mu}_i = v_i \mathbf{1} + (\gamma_{101} + \gamma_{111} z_i + \boldsymbol{\beta}'_1 \mathbf{x}_i, \dots, \gamma_{10, r_i-1} + \gamma_{11, r_i-1} z_i + \boldsymbol{\beta}'_1 \mathbf{x}_i, \gamma_{20, r_i} + \gamma_{21, r_i} z_i + \boldsymbol{\beta}'_2 \mathbf{x}_i, \dots, \gamma_{20K} + \gamma_{21K} z_i + \boldsymbol{\beta}'_2 \mathbf{x}_i)'$, and $W_i = \text{diag}(\Sigma_{1,1:r_i-1,1:r_i-1}, \Sigma_{2,r_i:K,r_i:K})$. The conditional density of y_i given r_i is given by

$$f(\mathbf{y}_i | z_i, \mathbf{x}_i, r_i; \boldsymbol{\gamma}, \boldsymbol{\beta}, \Sigma) = \int f(\mathbf{y}_i | z_i, \mathbf{x}_i, r_i, v_i; \boldsymbol{\gamma}, \boldsymbol{\beta}, \Sigma) f(v_i | \tau^2) dv_i \quad (3.6)$$

Next, conditioning on $R_i = r_i$ and y_i , we model the variable S_i , which is the time-point at which the i th patient starts the standard treatment. This conditional model for S_i given $R_i = r_i$ = assumes

$$f_{S|R}(s_i | r_i, z_i, \mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\lambda}) = P(S_i = s_i | R_i = r_i, z_i, \mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\lambda}) = \begin{cases} \text{logit}^{-1}[\lambda_{0r_i} + \lambda_{1r_i} z_i + \boldsymbol{\lambda}'_2(\mathbf{x}_i, y_{i,r_i-1})], & \text{if } r_i < K+1 \text{ and } s_i = r_i, \\ 1 - P(S_i = r_i | R_i = r_i, z_i, \mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\lambda}), & \text{if } r_i < K+1 \text{ and } s_i = K+1, \\ 1, & \text{if } r_i = K+1 \text{ and } s_i = K+1, \\ 0, & \text{if } r_i = K+1 \text{ and } s_i < K+1, \end{cases} \quad (3.7)$$

where $\text{logit}^{-1}(u) = \exp(u)/(1 + \exp(u))$, $\boldsymbol{\lambda} = (\lambda_{01}, \dots, \lambda_{0K}, \lambda_{11}, \dots, \lambda_{1K}, \boldsymbol{\lambda}'_2)'$ and $\boldsymbol{\lambda}_2$ is a vector of the regression coefficients of $(\mathbf{x}_i, y_{i,r_i-1})$. We assume S that is MAR but may depend on the observed outcome.

Conditional on $R_i = r_i$ and $S_i = s_i$, we model the random variable L_i , which is the time-point a patient completely withdraws from the study, that is, the patient stops taking the assigned treatment as well as the standard treatment if any. Note that we need to model L_i only when $r_i \leq s_i < K$. The subset of the data with $r_i \leq s_i < K$ is often sparse, especially when K is small. Conditional on $R_i = r_i$ and $S_i = s_i$, we propose a more parsimonious multivariate logistic regression model for L_i , which is stated as follows:

$$\begin{aligned} f_{L|R,S}(\ell_i | r_i, s_i, \mathbf{t}, \mathbf{w}_i, \mathbf{y}_i, \boldsymbol{\alpha}) &= P(L_i = \ell_i | R_i = r_i, S_i = s_i, \mathbf{t}, \mathbf{w}_i, \mathbf{y}_i, \boldsymbol{\alpha}) \\ &= \begin{cases} \frac{\exp(\alpha_0 + \alpha_1 t_{\ell_i} + \alpha_2 y_{i,\ell_i-1} + \boldsymbol{\alpha}'_3 \mathbf{w}_i)}{1 + \sum_{k=s_i+1}^K \exp(\alpha_0 + \alpha_1 t_k + \alpha_2 y_{i,k-1} + \boldsymbol{\alpha}'_3 \mathbf{w}_i)}, & \text{if } r_i = s_i < K \text{ and } s_i+1 \leq \ell_i < K+1, \\ \frac{1}{1 + \sum_{k=s_i+1}^K \exp(\alpha_0 + \alpha_1 t_k + \alpha_2 y_{i,k-1} + \boldsymbol{\alpha}'_3 \mathbf{w}_i)}, & \text{if } r_i = s_i < K \text{ and } \ell_i = K+1, \\ 1, & \text{if } r_i = s_i = K \text{ and } \ell_i = K+1, \\ 1, & \text{if } r_i \leq K+1, s_i = K+1, \text{ and } \ell_i = r_i, \\ 0, & \text{otherwise,} \end{cases} \quad (3.8) \end{aligned}$$

where $\mathbf{t} = (t_1, \dots, t_K)'$, $\mathbf{w}_i$ is a vector of covariates, which may be a subset of $\mathbf{x}_i$, and $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2, \alpha_3)'$. This is a non-ignorable model as (3.8) depends on the off-protocol responses, $(y_{iS_i}, y_{iS_i+1}, \dots, y_{i,K-1})'$, in which $y_{iI_i}, y_{i,I_i+1}, \dots, y_{i,K-1}$ are not observed.

The proposed multivariate response model defined by (3.2)–(3.4) assumes a general covariance structure for the repeated measurements from the assigned standard treatment and the repeated measurements from the standard treatment over time. It captures the dependence between per- and off-protocol responses by random intercept v_i , which may be extended to other random effect model to allow for more general dependency. Furthermore, the joint distribution of three types of missing data mechanisms in the longitudinal process (R_i, S_i, L_i) and the outcome Y_i via (3.1) and (3.6)–(3.8) is modeled by a sequence of the conditional distributions via $[R_i] [Y_i|R_i] [S_i|Y_i, R_i] [L_i|R_i, Y_i, S_i]$. These models are quite new, as modeling such data has not been attempted in the statistical literature.

3.3. The Likelihood Functions

Let $\psi = (\gamma, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \tau, \Sigma_1, \Sigma_2)$ denote the collection of all model parameters. Write $\mathbf{v} = (v_1, \dots, v_n)'$. Denote the complete data by $D_c = \{\mathbf{y}_i, z_i, \mathbf{x}_i, v_i, \mathbf{w}_i, r_i, s_i, \ell_i, i = 1, \dots, n\}$ and the observed data by $D_o = \{(y_{ik}(1 - g_{ik}), k = 1, \dots, K), z_i, \mathbf{x}_i, \mathbf{w}_i, r_i, s_i, \ell_i, i = 1, \dots, n\}$. Then, the complete data likelihood function is given by

$$\begin{aligned} \mathcal{L}_c(\psi|D_c) \\ = \prod_{i=1}^n \left[\frac{\exp\{-v_i^2/(2\tau^2)\}}{\sqrt{2\pi}\tau} f_R(r_i|z_i, \mathbf{x}_i, \boldsymbol{\theta}) f(\mathbf{y}_i|r_i, v_i, \gamma, \boldsymbol{\beta}, \Sigma) \times f_{S|R}(s_i|r_i, z_i, \mathbf{x}_i, y_{i,r_i-1}, \boldsymbol{\lambda}) f_{L|R,S}(\ell_i|r_i, s_i, \mathbf{t}, \mathbf{w}_i, \mathbf{y}, \boldsymbol{\alpha}) \right], \end{aligned} \quad (3.9)$$

where the density functions are defined in (3.5)–(3.8).

Let $\mathbf{y}_{i,mis} = (y_{ik}g_{ik}, k = 1, \dots, K)$ denote the vector of missing per-protocol or off-protocol outcomes. The observed data likelihood function is written as follows:

$$\begin{aligned} \mathcal{L}_o(\psi|D_o) = \prod_{i=1}^n f_R(r_i|z_i, \mathbf{x}_i, \boldsymbol{\theta}) f_{S|R}(s_i|r_i, z_i, \mathbf{x}_i, y_{i,r_i-1}, \boldsymbol{\lambda}) \int \left[\frac{\exp(-v_i^2/(2\tau^2))}{\sqrt{2\pi}\tau} \right. \\ \left. \times \int f(\mathbf{y}_i|r_i, v_i, \gamma, \boldsymbol{\beta}, \Sigma) f_{L|R,S}(\ell_i|r_i, s_i, \mathbf{t}, \mathbf{w}_i, \mathbf{y}, \boldsymbol{\alpha}) d\mathbf{y}_{i,mis} \right] dv. \end{aligned} \quad (3.10)$$

We can leave $f_{S|R}(s_i|r_i, z_i, \mathbf{x}_i, y_{i,r_i-1}, \boldsymbol{\lambda})$ outside the integration in (3.10). This is because y_{i,r_i-1} cannot be missing, otherwise, by definition the time point at which the i th patient stops taking the assigned treatment should take the value of $r_i - 1$ but not r_i .

4. Posterior Inference

4.1. Prior and Posterior

We specify independent priors for $\gamma, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \tau, \Sigma_1$ and Σ_2 . Thus, the joint prior is given by

$$\pi(\psi) = \pi(\gamma)\pi(\boldsymbol{\beta})\pi(\boldsymbol{\theta})\pi(\boldsymbol{\lambda})\pi(\boldsymbol{\alpha})\pi(\tau)\pi(\Sigma_1)\pi(\Sigma_2). \quad (4.1)$$

Specifically, for each of γ , β , θ , λ and α , we take a noninformative multivariate normal prior $N(\mathbf{0}, 1000I)$, where I is a generic notation denoting an identity matrix with the dimension matching each of these parameter vectors. Also, we take an Inverse-Wishart distribution, $\text{Inv-Wishart}_{d_0}(\Sigma_0^{-1})$, for Σ_j for $j = 1, 2$ with density $\pi(\Sigma_j) \propto |\Sigma_j|^{-(d_0 + K + 1)/2}$

$\exp\{-\frac{1}{2}\text{tr}(\Sigma_0 \Sigma_j^{-1})\}$ for $j = 1, 2$, where the degrees of freedom d_0 and the $K \times K$ positive definite matrix Σ_0 are prespecified. We further assume an inverse gamma priors Inv-

$\text{Gamma}(a, b)$ for τ^2 with density $\pi(\tau^2) \propto \frac{1}{(\tau^2)^{a+1}} \exp(-b/\tau^2)$. In Section 5, we took $d_0 = K + 1$, $\Sigma_0 = 0.001I_K$, and $a = 2.001$ and $b = 0.01$, which are fairly noninformative.

Write $\mathbf{y}_{mis} = (\mathbf{y}'_{1,mis}, \dots, \mathbf{y}'_{n,mis})'$. Then, the joint posterior distribution of $(\psi, \mathbf{v}, \mathbf{y}_{mis})$ given the observed data D_0 takes the form:

$$\pi(\psi, \mathbf{v}, \mathbf{y}_{mis} | D_0) \propto * \mathcal{L}_c(\psi | D_c) \pi(\psi), \quad (4.2)$$

where $\mathcal{L}_c(\psi | D_c)$ is defined by (3.9). Using the results established in [23] and [24], we can show that under some mild conditions, the posterior $\pi(\psi, \mathbf{v}, \mathbf{y}_{mis} | D_0)$ is proper when $\pi(\theta) \propto 1$, $\pi(\lambda) \propto 1$, and $\pi(\alpha) \propto 1$. Thus, the proposed models (3.1), (3.7), and (3.8) for R , S , and L are identifiable in this sense.

4.2. Posterior Computation

Due to the complexity of the random-effects regression models for repeated treatment outcomes and logistic regression models for r_i , s_i , and ℓ_i , it does not appear possible to obtain an analytic evaluation of the posterior distributions. Also, due to a large percentage of missing values in both per- and off-protocol outcome measurements, there is a need to develop an efficient Markov chain Monte Carlo (MCMC) sampling algorithm to sample $(\psi, \mathbf{v}, \mathbf{y}_{mis})$ from the posterior distribution $\pi(\psi, \mathbf{v}, \mathbf{y}_{mis} | D_0)$. In this regard, we use the modified collapsed Gibbs technique of Liu [25] as discussed in Chen et al. [26] to collapse out as many missing $y_{i,mis}$ as possible when we sample γ , β , $\mathbf{v}$, and Σ . Let $\mathbf{y}_{i,obs} = (y_{ik}(1 - g_{ik}), = 1, \dots, K)$ denote the vector of observed treatment outcomes for

$i = 1, \dots, n$, $\mathbf{y}_{.,obs} = (\mathbf{y}'_{1,obs}, \dots, \mathbf{y}'_{n,obs})'$, and $\mathbf{y}_{.,mis} = (\mathbf{y}'_{1,mis}, \dots, \mathbf{y}'_{n,mis})'$. The Gibbs sampling algorithm requires to sample from the following conditional posterior distributions in turn: (i) $[\theta | D_0]$; (ii) $[\lambda | D_0]$; (iii) $[\alpha | \mathbf{y}_{.,mis}, D_0]$; (iv) $[\gamma, \beta, \mathbf{v}, \Sigma | \mathbf{y}_{.,mis}, \tau, D_0]$; and (v) $[\tau | \mathbf{v}]$.

We briefly discuss how to sample from each of the above conditional distributions as follows. For (i), the posterior distribution of θ conditional on D_0 is given by

$\pi(\theta | D_0) \propto \pi(\theta) \prod_{i=1}^n f_R(r_i | z_i, \mathbf{x}_i, \theta)$, which is log-concave in each component of θ . For (ii), the posterior distribution of λ conditional on D_0 is given by

$$\pi(\lambda | D_0) \propto \pi(\lambda) \prod_{i=1}^n f_{S|R}(s_i | r_i, z_i, \mathbf{x}_i, y_{i,r_{i-1}}, \lambda),$$

which is also log-concave in each component of λ . For (iii), the conditional posterior distribution of α given $y_{\cdot, mis}(\mathbf{0})$ and D_o is of the form:

$$\pi(\alpha | y_{\cdot, mis}, D_o) \propto \pi(\alpha) \prod_{i=1}^n f_{L|R,S}(\ell_i | r_i, s_i, \mathbf{t}, \mathbf{w}_i, \mathbf{y}_i, \alpha),$$

which is again log-concave. Thus, we can use the adaptive rejection algorithm of Gilks and Wild [27] to generate θ , λ , and α from the above three conditional distributions.

For (iv), we run a sub-Gibbs sampling algorithm to sample $(\gamma, \beta, \mathbf{v}, \Sigma, y_{mis})$ from the joint conditional distribution. The conditional posterior distributions required by this sub-Gibbs sampling algorithm are given as follows:

(iva) $[\gamma | \beta, \mathbf{v}, \Sigma, y_{mis}, D_o]$;

(ivb) $[\beta | \gamma, \mathbf{v}, \Sigma, y_{mis}, D_o]$;

(ivc) $[\mathbf{v} | \gamma, \beta, \Sigma, y_{mis}, D_o]$;

(ivd) $[\Sigma | \gamma, \beta, \mathbf{v}, y_{mis}, -K, D_o]$, where $y_{mis}, -K$ denotes all missing treatment outcomes with exclusion of those $(y_{ik}$'s such that all $(y_{i1}, \dots, y_{iK})$ are missing and none of these K missing treatment outcomes are included in (3.7) and (3.8); and

(ive) $[y_{mis} | \gamma, \beta, \mathbf{v}, \Sigma, D_o]$.

From these conditional distributions, we see that we have collapsed out most of the missing treatment outcomes y_{ik} 's in (iva), (ivb) and (ivc). This sub-Gibbs sampling algorithm is called the modified collapsed Gibbs sampler in Chen et al. [26]. In addition, the conditional posterior distributions of the components of γ , β , and $\mathbf{v}$ are either univariate or multivariate normals, and the conditional posterior distributions of Σ_1 and Σ_2 are inverse Wisharts. Thus, sampling γ , β , $\mathbf{v}$, and Σ is straightforward. The conditional posterior distributions of the missing treatment outcomes in y_{mis} are either normals or log-concave. Therefore, we sample those missing treatment outcomes either from the normals or using the adaptive rejection algorithm of Gilks and Wild [27]. Finally, we mention that for (v), τ follows an inverse gamma distribution and hence, sampling τ is straightforward.

5. Analysis of the Mild Cognitive Impairment Data

We present in this section the descriptive statistics and a detailed analysis of the data from the MCI trial discussed in Section 2. As shown in Table 1, of the 1011 enrolled patients, 505 (50%) were assigned to Exelon treatment; 36 and 10 had intermittent missingness on the per- and off-protocol outcome measurements, respectively; 472 (46.7%) complied with the original assignment and stayed till the end of the study. For this particular dataset, we found that after the patients discontinued the assigned treatment, he/she either started the standard treatment and came back to the clinic right away, or never came back to the clinic ($L = R$). Using the notations defined in Section 3.1, Table 1 also provides the summary information for all the observed combinations of the per- and off-protocol schedule. For example, among the 28 patients stopping their assigned treatment at visit 2 and starting the standard treatment at the same time ($R = S = 2$), one completely withdrew from the study at visit 5 ($L = 5$), two

completely withdrew at visit 6 ($L = 6$), and 25 stayed till the end of the study ($L = 7$). Furthermore, among those 28 patients, nobody had intermittent missing for per-protocol outcome (IM $Y_1 = 0$); four had intermittent missing for off-protocol outcome (IM $Y_2 = 4$); 12 were assigned to Exelon treatment originally.

To implement the method proposed in Section 3 to MCI trial, we considered four covariates for $\mathbf{x}_i$ in (3.2) and (3.1), where x_{i1} = the baseline composite z-score (CZSCR0), x_{i2} = age, x_{i3} = sex (1 =female and 0 =male), and x_{i4} = EDUGRP (1 = High education group and 0 = otherwise). Thus, we have

$$\beta'_j \mathbf{x}_i = \beta_{j1} \text{CZSCR0}_i + \beta_{j2} \text{age}_i + \beta_{j3} \text{sex}_i + \beta_{j4} \text{EDUGRP}_i$$

for $j = 1, 2$. Similarly, in (3.1) and (3.7), we took

$$\theta'_2 \mathbf{x}_i = \theta_{21} \text{CZSCR0}_i + \theta_{22} \text{age}_i + \theta_{23} \text{sex}_i + \theta_{24} \text{EDUGRP}_i \text{ and}$$

$\lambda'_2(\mathbf{x}_i, y_{i,r_i-1}) = \lambda_{21} \text{age}_i + \lambda_{22} \text{sex}_i + \lambda_{23} \text{EDUGRP}_i + \lambda_{24} y_{i,r_i-1}$, where $y_{i0} = \text{CZSCR0}_i$. As shown in Table 1, there were only 10 ℓ_i values such that $s_i < \ell_i < K + 1$, which essentially implies that there were only 9 “events” for the model (3.8). To ensure the model identifiability, we took a single covariate for $\mathbf{W}_i$, namely, $w_i = \text{EDUGRP}_i$ in (3.8). We also tried CZSCR0, age, or sex as w_i and we found that EDUGRP was most related to L_i . In all of the computations, we standardized age for numerical stability and centerized t_i at 4, which corresponds to 24 months, for improving convergence of the Gibbs sampling algorithm.

The posterior means (Estimates), posterior standard deviations (SDs), and 95% highest posterior density (HPD) intervals of the model parameters are reported in Tables 2 to 4. Tables 2 to 4 as well as Tables A1 and A2 in the supplemental material reveal some very interesting results. When examining Table 4, in the case of both the models for per-protocol and off-protocol outcome measurements, it is clear that the drug shows no evidence of a positive treatment effect at any time point. Nevertheless, by examining the R , S and L models, it is possible to assess how covariates affect the pattern of withdrawal from study treatment and whether a patient is more likely to return to the clinic following withdrawal from the assigned study treatment.

In the case of the R model, patients in the Exelon group have a higher tendency to withdraw from the experimental treatment at the first visit (the 95% HPD interval for the log-odds excludes 0). As this effect seems not to be present in subsequent visits, this seems to reflect the problem of initial toleration. A further interesting relationship can be seen in subjects with high levels of education, who appear to be more likely to withdraw from study treatment, either Exelon or placebo, but only after at least 1 year. As there appears to be no evidence of a treatment effect in either of the education groups, it can be speculated that subjects with higher levels of education were more likely to be able to evaluate the treatment benefit and therefore withdraw from study treatment, if noticeable benefits were not achieved within a reasonable time frame.

Due to sparse data, the S and L models provide less insight into possible covariate relationships. In the case of the L model, higher education level seems to increase the

propensity to withdraw completely from the study. This is not surprising given that education level was a predictive covariate in the R model. In Table 3, we see evidence of nonignorable missing data in that the 95% HPD interval for $y_{i,k-1}$ does not contain 0, indicating its significance. Again, for the L model, the 95% HPD interval for Visit does not contain 0, which indicates different dropout patterns over time. For the S model, we see from Table 3 that the 95% HPD interval for y_{i,r_i-1} does not include zero, which implies that whether a patient starts the standard treatment after stopping the assigned treatment highly depends primarily on treatment outcome at the previous visit. When linking these results to Figure 2 in the supplemental file, which compares covariate relationships for the marginal L model between the treatment groups, we again see that the largest difference between Exelon and placebo is observed at the first visit in highly educated females.

To achieve the goodness-of-fit, we fit the largest identifiable models to the MCI data using all available covariates. For example, for the R model, we compared the proposed multivariate logistic regression model to a more parsimonious proportional odds regression model. Specifically, the values of AIC were 3158.52 for the proposed model and 3234.47 for the proportional odds regression model. Similar results were obtained based on the deviance information criterion (DIC). We would like to mention that with the likelihood function and the posterior distribution developed in Sections 3 and 4, other Bayesian model comparison criteria such as the conditional predictive ordinate (CPO) and the Bayes factor can be used to evaluate the model fit. Due to the length of the paper, we have not included those comparisons in this paper. In addition, we compared the results under our proposed model to those obtained from the mixed effects models for the outcome measurements collected when the patients were on-protocol and when they were off-protocol, separately. Specifically, we consider the following models:

$$y_{ik} = \begin{cases} \gamma_{10k} + \gamma_{11k}z_i + \beta_1' \mathbf{x}_i + \varepsilon_{i1k}, & \text{if } k < r_i, \\ \gamma_{20k} + \gamma_{21k}z_i + \beta_2' \mathbf{x}_i + \varepsilon_{i2k}, & \text{if } k \geq r_i, \end{cases} \quad (5.1)$$

where the notations are the same as in Section 3. Table 4 provides the estimates, standard errors, and 95% confidence intervals (CIs) of the parameters under model (5.1). From Table 4, we see that the estimates of the model parameters from the proposed model were slightly different than those from the standard mixed effects models particularly for the longitudinal process corresponding to the standard treatment. These modest differences are expected since the proposed model assumes that the drop-out (i.e., the L model) is nonignorable while the standard mixed effects models assume MAR for the missing outcomes from the standard treatment.

In all the Bayesian computations, we used 50,000 Gibbs samples after a burn-in of 2000 for each model, to compute all the posterior estimates, including posterior means, posterior standard deviations, and 95% HPD intervals. Codes were written for FORTRAN 95 compiler, and we used IMSL subroutines with double precision accuracy. The convergence of the Gibbs sampler was checked using several diagnostic procedures as recommended by Cowles and Carlin [28]. The autocorrelations for all model parameters disappeared before lag 10. The HPD intervals were computed by the Monte Carlo method developed by Chen and Shao [29].

6. Discussion

We have proposed a novel class of models for accommodating two different types of longitudinal processes in studies where patients stop taking the assigned treatment and then start a standard treatment. We have developed a model that links these two longitudinal processes together through a multivariate linear mixed model. In order to account for the different types of dropout patterns, three types of missingness indicators (R , L , S) are incorporated into the modeling structure as described in detail in the paper. The model is sophisticated and complex and provides the user with an analysis tool in analyzing such longitudinal processes for which the only other alternative would be to do an analysis ignoring the standard treatment data. Due to a large percentage of missing values in both per- and off-protocol outcomes, sampling from the posterior distribution is extremely challenging. To overcome such difficulty, we have developed an efficient Gibbs sampling algorithm based on the modified collapsed Gibbs technique of [26] to sample from the posterior distribution under such a complex model.

As pointed out by one reviewer, s is a deterministic function of r and ℓ , namely, $s = r$ if $r \neq \ell$ and $s = K + 1$ if $r = \ell$. The deterministic relationship between s and (r, ℓ) is in fact built in the proposed models for S_i and L_i in (3.7) and (3.8). This deterministic relationship leads to another view of our proposed model, in which we may model $[R_i][Y_i|R_i][L_i|R_i, Y_i]$ and ignore the model for S_i . Under this alternative model, we need to model L_i for $r_i = L_i = K + 1$ when $r_i < K + 1$ instead of $s_i + 1 = L_i < K + 1$ when $r_i = s_i < K$. The alternative model is equivalent to the proposed model if we assume

$$P(L_i = \ell_i | R_i = r_i, \mathbf{y}_i) = \begin{cases} P(L_i = r_i | R_i = r_i, \mathbf{y}_i), & \text{if } \ell_i = r_i \\ [1 - P(L_i = r_i | R_i = r_i, \mathbf{y}_i)]P(L_i = \ell_i | L_i > r_i, R_i = r_i, \mathbf{y}_i) & \text{if } \ell_i > r_i \end{cases}$$

and then assume

$$P(L_i = \ell_i | L_i > r_i, R_i = r_i, \mathbf{y}_i) = \begin{cases} \frac{\exp(\alpha_0 + \alpha_1 t_{\ell_i} + \alpha_2 y_{i, \ell_i - 1} + \alpha_3' \mathbf{w}_i)}{1 + \sum_{k=s_i+1}^K \exp(\alpha_0 + \alpha_1 t_k + \alpha_2 y_{i, k-1} + \alpha_3' \mathbf{w}_i)}, & \text{if } \ell_i < K+1, \\ \frac{1}{1 + \sum_{k=s_i+1}^K \exp(\alpha_0 + \alpha_1 t_k + \alpha_2 y_{i, k-1} + \alpha_3' \mathbf{w}_i)}, & \text{if } \ell_i = K+1. \end{cases}$$

It is easy to see that

$$P(L_i = r_i | R_i = r_i, \mathbf{y}_i) = P(S_i = K+1 | R_i = r_i, \mathbf{y}_i) = 1 - \text{logit}^{-1}[\lambda_{0r_i} + \lambda_{1r_i} z_i + \lambda_2'(\mathbf{x}_i, y_{i, r_i-1})].$$

A possible different model is to directly assume a multivariate logistic regression model for $[L_i|R_i = r_i, \mathbf{y}_i]$ for $r_i = L_i = K + 1$. However, the standard saturated multivariate logistic regression model is not identifiable due to the sparsity of the subset of the observed data with $r_i < \ell_i < K + 1$. Thus, the proposed model is preferred and it can also allow for building in different non-ignorable missing patterns since both $P(L_i = r_i | R_i = r_i, \mathbf{y}_i)$ and $P(L_i = \ell_i | L_i > r_i, R_i = r_i, \mathbf{y}_i)$ may depend on missing values of $\mathbf{y}_i$. This is one reason why we chose to model

$[R_i][Y_i|R_i][S_i|R_i, Y_i][L_i|S_i, R_i, Y_i]$ rather than $[R_i][Y_i|R_i][L_i|R_i, Y_i]$. Another reason is that the proposed model is easy to understand for practitioners since S_i naturally follows after R_i and L_i follows after S_i .

Future work for the proposed model includes allowing for discrete responses or time-to-event responses along with time-varying covariates. An additional area for future research could be based around developing an overall treatment effect summary based on an estimand that combines information from the per-protocol and off-protocol models. Work on an overall treatment effect estimate would be particularly important when planning a new clinical trial, as opposed to post-hoc exploratory analysis presented here. In the context of planning, the National Research Council report emphasizes the need for the clear identification of a target estimand (e.g. intention-to-treat estimand). The model developed in this paper contains separate treatment effect parameters associated with each study visit for both per-protocol and off-protocol data. However, in a drug development setting regulators would have to primarily base decision making on an overall treatment effect, possibly based on a weighted average of the per- and off-protocol models. These future research topics are under investigation currently.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank the Editor, the Associate Editor, and the two anonymous reviewers for their very helpful comments and suggestions, which have led to an improved version of the paper. Dr. Chen's work was partially supported by NIH grants 1R21HL097334 and UL1 RR024975-01. Dr. M.-H. Chen and Dr. J. G. Ibrahim's research was partially supported by NIH grants #GM 70335 and #CA 74015.

Appendix

A. The Predictive Distributions of R , S , and L

To characterize the patterns of treatment switching and dropout, it is of great clinical importance to derive the predictive distributions of R , S , and L . Let z^* denote the treatment indicator, and $\mathbf{x}^*$ and $\mathbf{w}^*$ be vectors of covariates. The predictive distribution of R is straight forward and similar to (3.1). Conditioning on R , the predictive distribution of S is given as follows. For $s < K + 1$,

$$P(S=s|z^*, \mathbf{x}^*, \boldsymbol{\lambda}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \tau^2) = \int P(S=s|R=s, z^*, \mathbf{x}^*, y_{s-1}, \boldsymbol{\lambda}) P(R=s|z^*, \mathbf{x}^*, \boldsymbol{\theta}) f(\mathbf{y}|R=s, z^*, \mathbf{x}^*; \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \tau^2) d\mathbf{y} \quad (\text{A.1})$$

and for $s = K + 1$,

$$P(S=K+1|z^*, \mathbf{x}^*, \boldsymbol{\lambda}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \tau^2) = \sum_{k=1}^K [\int P(S=K+1|R=k, z^*, \mathbf{x}^*, y_{k-1}, \boldsymbol{\lambda}) P(R=k|z^*, \mathbf{x}^*, \boldsymbol{\theta}) \times f(\mathbf{y}|R=k, z^*, \mathbf{x}^*; \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \tau^2) d\mathbf{y}] + P(R=K+1|z^*, \mathbf{x}^*, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \tau^2), \quad (\text{A.2})$$

where $P(S = s|R = s, z^*, \mathbf{x}^*, y_{s-1}, \boldsymbol{\lambda})$ is defined in (3.7) and $P(R = K + 1|z^*, \mathbf{x}^*, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \tau^2)$ is denned in (3.1).

Similarly, the predictive distribution of L is given as follows. For $\ell < K + 1$,

$$\begin{aligned} P(L=\ell|z^*, \mathbf{x}^*, \mathbf{t}, \mathbf{w}^*, \boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \tau^2) \\ = \sum_{k=1}^{\ell-1} \int \{P(S=k|R=k, z^*, \mathbf{x}^*, y_{k-1}, \boldsymbol{\lambda})P(L=\ell|R=k, S=k, \mathbf{t}, \mathbf{w}^*, \mathbf{y}, \boldsymbol{\alpha}) \times P(R=k|z^*, \mathbf{x}^*, \boldsymbol{\theta})f(\mathbf{y}|R=k, z^*, \mathbf{x}^*; \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \tau^2)\} d\mathbf{y} \\ + \int P(R \\ = \ell|z^*, \mathbf{x}^*, \boldsymbol{\theta})P(S \\ = K+1|R \\ = \ell, z^*, \mathbf{x}^*, y_{\ell-1}, \boldsymbol{\lambda})f(\mathbf{y}|R \\ = \ell, z^*, \mathbf{x}^*; \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \tau^2) d\mathbf{y}, \end{aligned} \quad \begin{matrix} (A. \\ 3) \end{matrix}$$

and for $\ell = K + 1$,

$$\begin{aligned} P(L=K+1|z^*, \mathbf{x}^*, \mathbf{t}, \mathbf{w}^*, \boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \tau^2) = \sum_{k=1}^{K-1} \int \{P(R=k|z^*, \mathbf{x}^*, \boldsymbol{\theta})P(S=k|R=k, z^*, \mathbf{x}^*, y_{k-1}, \boldsymbol{\lambda}) \times P(L=K+1|R=k, S= \\ + P(S=K|z^*, \mathbf{x}^*, \boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \tau^2) + P(R=K+1|z^*, \mathbf{x}^*, \boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \tau^2), \end{aligned} \quad \begin{matrix} (A. \\ 4) \end{matrix}$$

where $P(L = \ell|R = k, S = k, \mathbf{t}, \mathbf{w}^*, \mathbf{y}, \boldsymbol{\alpha})$ is defined in (3.8), $P(R = K + 1|z^*, \mathbf{x}^*, \boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \tau^2)$ is denned in (3.1) $P(S = K|z^*, \mathbf{x}^*, \boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \tau^2)$ is denned in (A.2), $f(\mathbf{y}|R = k, z^*, \mathbf{x}^*; \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \tau^2)$ is given in (3.6).

References

1. Mallinckrodt C, Watkin J, Molenberghs G, Carroll R. Choice of the primary analysis in longitudinal clinical trials. *Pharmaceutical Statistics*. 2004; 3:161–169.
2. Lane P. Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. *Pharmaceutical Statistics*. 2008; 7:93–106. [PubMed: 17351897]
3. Mallinckrodt C, Lane P, Schnell C, Mancuso PYJ. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. *Drug Information Journal*. 2008; 42:303–319.
4. Siddiqui O, Hung H, O'Neill R. MMRM vs. LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets. *Journal of Biopharmaceutical Statistics*. 2009; 19:227–246. [PubMed: 19212876]
5. The Prevention and Treatment of Missing Data in Clinical Trials. Washington, DC: The National Academies Press; 2010. Panel on Handling Missing Data in Clinical Trials.
6. Wu MC, Bailey KR. Analysing changes in the presence of informative right censoring caused by death and withdrawal. *Statistics in Medicine*. 1988; 7:337–346. [PubMed: 3281208]
7. Wu MC, Bailey KR. Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics*. 1989; 45:939–955. [PubMed: 2486189]
8. Wu MC, Carroll RJ. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*. 1988; 44:175–188.
9. Creemers A, Molenberghs G, Verbeke G, Kenward MG. A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Biometrical Journal*. 2010; 52:111–125. [PubMed: 19937996]
10. Fitzmaurice, GM.; Laird, NM.; Ware, JH. *Applied Longitudinal Analysis*. New York: Wiley; 2004.
11. Molenberghs, G.; Verbeke, G. *Models for Discrete Longitudinal Data*. New York: Springer; 2005.

12. Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis (with Discussion). *Journal of the Royal Statistical Society. Series C (Applied Statistics)*. 1994; 43:49–93.
13. Little RJA. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*. 1995; 90:1112–1121.
14. Ibrahim JG, Chen MH, Lipsitz SR. Missing responses in generalized linear mixed models when the missing data mechanism is nonignorable. *Biometrika*. 2001; 88:551–564.
15. Little RJA, Wang Y. Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics*. 1996; 52:98–111. [PubMed: 8934587]
16. Hogan JW, Laird NM. Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine*. 1997; 16:239–257. [PubMed: 9004395]
17. Thijs H, Molenberghs G, Michiels B, Verbeke G, Curran D. Strategies to fit pattern-mixture models. *Biostatistics*. 2002; 3:245–265. [PubMed: 12933616]
18. Troxel AB, Lipsitz SR, Harrington DP. Marginal models for the analysis of longitudinal measurements with non-ignorable non-monotone missing data. *Biometrika*. 1998; 85:661–672.
19. Ibrahim JG, Molenberghs G. Missing data methods in longitudinal studies: a review (with discussion). *Test*. 2009; 18:1–75. [PubMed: 21218187]
20. Heitjan DF. Causal inference in a clinical trial: A comparative example. *Controlled Clinical Trials*. 1999; 20:309–18. [PubMed: 10440558]
21. Heitjan DF. Ignorability and bias in clinical trials. *Statistics in Medicine*. 1999; 18:2421–34. [PubMed: 10474150]
22. Grundman M, Petersen R, Ferris SH, Thomas RG, Aisen PS, Bennett DA, Foster NL, Jack CJ, Galasko DR, Doody R. Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Archives of Neurology*. 2004; 61:59–66. [PubMed: 14732621]
23. Chen MH, Shao QM. Propriety of posterior distribution for dichotomous quantal response models with general link functions. *Proceedings of the American Mathematical Society*. 2001; 129:293–302.
24. Chen MH, Ibrahim JG, Shao QM. Posterior propriety and computation for the cox regression model with applications to missing covariates. *Biometrika*. 2006; 93:791–807.
25. Liu JS. The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*. 1994; 89:958–966.
26. Chen, MH.; Shao, QM.; Ibrahim, JG. *Monte Carlo Methods in Bayesian Computation*. New York: Springer; 2000.
27. Gilks WR, Wild P. Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*. 1992; 41:337–348.
28. Cowles MK, Carlin BP. Markov chain monte carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*. 1996; 91:883–904.
29. Chen MH, Shao QM. Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*. 1999; 8:69–92.

Figure 1. Selected On-/Off-protocol Patterns with Exemplified Outcome Observations from Both Placebo (P) and Treatment (T) Groups (top) and boxplots of On- and Off-protocol Outcomes (bottom). For top panel, values were labeled as “NA”s if responses were missing intermittently and as “UO”s (unobserved) if patients completely withdrew from study.

Figure 2.

Plots of the estimated predictive probabilities along with 95% HPD intervals of R , where three columns correspond to three quartiles ($-4.738, 0.491, 5.024$) of the baseline measure; and four rows correspond to male and low education group, male and high education group, female and low education group, and female and high education group.

Table 1

Summary of the Mild Cognitive Impairment (MCI) Data

N	R	S	L Value(N)	IM Y ₁	IM Y ₂	Exelon Trt	Score Baseline [*]	Age [*]	Sex (F)	Edu (High)
43	1	1	2(1), 6(2), 7(40)	0	5	26	2.0 (-1.1,7.1)	70 (64,74.5)	27	8
105	1	7	1(105)	0	0	78	1.3 (-4.9,5.8)	72 (66,76)	69	28
28	2	2	5(1), 6(2), 7(25)	0	4	12	1.5 (-1.7,5.7)	70.5 (65,73.5)	15	6
75	2	7	2(75)	0	0	35	2.6 (-7.2,5)	71 (66.5,78)	42	14
15	3	3	6(1), 7(14)	1	1	10	3.3 (-5.6,5.6)	74 (62.5,81)	7	4
83	3	7	3(83)	2	0	36	-3.3 (-8.1,8)	73 (68,77.5)	49	26
5	4	4	6(1), 7(4)	0	0	2	1.8 (-0.6,4.8)	74 (64,74)	1	3
59	4	7	4(59)	1	0	29	-3.1 (-6.5,0.4)	72 (66.5,76)	26	19
12	5	5	6(2), 7(10)	0	0	8	4.2 (0.9,5.9)	68 (64,70.8)	7	3
48	5	7	5(48)	3	0	19	-3.6 (-8.7,1.4)	74 (68,76)	24	14
12	6	6	7(12)	0	0	6	2.3 (-1.6,4.2)	70.5 (66.2,73)	7	3
54	6	7	6(54)	1	0	28	-3.3 (-6.3,2.6)	74.5 (70,77)	29	18
472	7	7	7(472)	28	0	216	1.8 (-2.4,6.2)	70 (64,75)	226	110
1011				36	10	505	0.5 (-4.7,5.0)	71 (65,76)	529	256

IM Y₁/Y₂: number of intermittent missing for on/off treatment outcome;

^{*} median (interquartile)

Table 2

Posterior estimates of the parameters for the R-model

Visit	Variable	Estimate	SD	95% HPD Interval
1	Intercept	-2.240	0.223	(-2.681, -1.812)
	Treatment	1.042	0.205	(0.639, 1.443)
	CZSCR0	-0.004	0.017	(-0.037, 0.028)
	age	0.118	0.104	(-0.086, 0.323)
	sex	0.732	0.202	(0.324, 1.116)
	EDUGRP	0.224	0.240	(-0.253, 0.691)
2	Intercept	-1.637	0.211	(-2.058, -1.230)
	Treatment	-0.027	0.224	(-0.461, 0.415)
	CZSCR0	-0.072	0.018	(-0.108, -0.036)
	age	-0.054	0.119	(-0.288, 0.179)
	sex	0.176	0.229	(-0.270, 0.630)
	EDUGRP	0.067	0.289	(-0.502, 0.635)
3	Intercept	-2.081	0.234	(-2.536, -1.623)
	Treatment	0.034	0.228	(-0.422, 0.471)
	CZSCR0	-0.100	0.019	(-0.137, -0.064)
	age	0.164	0.124	(-0.071, 0.416)
	sex	0.317	0.239	(-0.162, 0.776)
	EDUGRP	0.856	0.272	(0.320, 1.379)
4	Intercept	-2.271	0.268	(-2.802, -1.755)
	Treatment	0.057	0.275	(-0.467, 0.610)
	CZSCR0	-0.126	0.023	(-0.170, -0.082)
	age	-0.122	0.147	(-0.410, 0.167)
	sex	-0.337	0.290	(-0.897, 0.240)
	EDUGRP	0.967	0.316	(0.354, 1.595)
5	Intercept	-2.326	0.276	(-2.859, -1.775)
	Treatment	-0.064	0.283	(-0.617, 0.495)
	CZSCR0	-0.106	0.023	(-0.151, -0.061)
	age	-0.054	0.151	(-0.354, 0.239)
	sex	0.049	0.292	(-0.519, 0.627)
	EDUGRP	0.686	0.336	(0.017, 1.334)
6	Intercept	-2.485	0.275	(-3.011, -1.940)
	Treatment	0.222	0.268	(-0.307, 0.743)
	CZSCR0	-0.082	0.022	(-0.125, -0.039)
	age	0.127	0.145	(-0.149, 0.417)
	sex	0.258	0.278	(-0.282, 0.808)
	EDUGRP	0.810	0.314	(0.184, 1.411)

Table 3

Posterior estimates of the parameters for the S - and L -models

Mode	Variable	Visit	Estimate	SD	95% HPD Interval
S	Intercept	1	-0.500	0.357	(-1.209, 0.195)
		2	-0.771	0.343	(-1.433, -0.084)
		3	-1.908	0.573	(-3.011, -0.871)
		4	-2.392	1.335	(-4.177, -0.736)
		5	-1.713	0.663	(-2.957, -0.524)
		6	-1.197	0.700	(-2.288, -0.134)
Treatment		1	-0.622	0.402	(-1.425, 0.153)
		2	-0.158	0.477	(-1.088, 0.780)
		3	0.811	0.674	(-0.434, 2.098)
		4	0.087	1.747	(-2.834, 2.728)
		5	1.132	0.805	(-0.382, 2.624)
		6	-0.075	0.886	(-1.556, 1.296)
age			0.004	0.124	(-0.233, 0.251)
			-0.071	0.239	(-0.531, 0.408)
			-0.352	0.280	(-0.908, 0.188)
			0.081	0.273	(0.046, 0.116)
L	Intercept		-5.541	0.967	(-7.316, -3.839)
		Visit	0.996	0.428	(0.208, 1.852)
		$y_{i,k} - 1$	-0.107	0.045	(-0.195, -0.020)
		EDUGRP	1.634	0.938	(-0.067, 3.241)

Table 4

Posterior estimates of the parameters for the on- and off-protocol outcome models and the restricted maximum likelihood estimates for model (5.1).

Model	Variable	Proposed Model in (3.2)			Mixed Effect Model in (5.1)		
		Est	SD	95% HPD	Est	SD	95% CI
On-Protocol Outcome Model	Intercept						
	1	0.028	0.192	(-0.356, 0.394)	0.004	0.190	(-0.369, 0.378)
	2	0.036	0.205	(-0.368, 0.434)	0.006	0.195	(-0.376, 0.388)
	3	0.352	0.239	(-0.109, 0.822)	0.313	0.224	(-0.127, 0.752)
	4	-0.211	0.254	(-0.705, 0.289)	-0.278	0.236	(-0.740, 0.185)
	5	-0.355	0.276	(-0.896, 0.189)	-0.474	0.251	(-0.967, 0.019)
	6	-0.012	0.314	(-0.631, 0.600)	-0.289	0.281	(-0.841, 0.263)
	Treatment						
	1	0.217	0.212	(-0.205, 0.626)	0.212	0.211	(-0.203, 0.627)
	2	0.087	0.236	(-0.379, 0.544)	0.090	0.219	(-0.340, 0.521)
	3	-0.167	0.293	(-0.745, 0.406)	-0.167	0.273	(-0.702, 0.368)
	4	-0.181	0.322	(-0.815, 0.444)	-0.172	0.294	(-0.749, 0.406)
	5	-0.085	0.358	(-0.781, 0.619)	-0.063	0.320	(-0.692, 0.565)
	6	-0.328	0.419	(-1.141, 0.495)	-0.278	0.371	(-1.007, 0.451)
Off-Protocol Outcome Model	CZSCR0	0.996	0.016	(0.965, 1.028)	0.992	0.015	(0.962, 1.023)
	age	-0.480	0.103	(-0.683, -0.277)	-0.487	0.102	(-0.686, -0.287)
	sex	-0.279	0.196	(-0.670, 0.095)	-0.258	0.193	(-0.638, 0.121)
	EDUGRP	0.119	0.232	(-0.323, 0.585)	0.188	0.228	(-0.258, 0.635)
	Intercept						
	1	0.257	0.882	(-1.500, 1.964)	0.418	0.799	(-1.166, 2.002)
	2	1.003	0.636	(-0.269, 2.234)	0.910	0.651	(-0.379, 2.199)
	3	1.260	0.696	(-0.102, 2.621)	1.469	0.697	(0.087, 2.850)
	4	0.607	0.662	(-0.660, 1.938)	0.893	0.669	(-0.432, 2.218)
	5	0.148	0.614	(-1.064, 1.335)	0.171	0.706	(-1.227, 1.570)
	6	0.280	0.658	(-0.983, 1.601)	0.314	0.751	(-1.174, 1.803)

Model	Variable	Proposed Model in (3.2)			Mixed Effect Model in (5.1)		
		Est	SD	95% HPD	Est	SD	95% CI
Treatment							
	1	0.809	1.056	(-1.211, 2.935)	0.435	0.911	(-1.371, 2.240)
	2	-0.073	0.741	(-1.515, 1.392)	-0.429	0.704	(-1.825, 0.967)
	3	-0.207	0.799	(-1.758, 1.378)	-0.702	0.755	(-2.198, 0.793)
	4	-0.638	0.739	(-2.088, 0.816)	-1.098	0.697	(-2.479, 0.283)
	5	0.120	0.653	(-1.182, 1.384)	-0.132	0.760	(-1.639, 1.375)
	6	-0.557	0.724	(-2.004, 0.839)	-0.865	0.833	(-2.516, 0.787)
both	CZSCR0	1.116	0.047	(1.025, 1.207)	1.072	0.047	(0.978, 1.166)
	age	-0.563	0.289	(-1.129, 0.012)	-0.627	0.292	(-1.206, -0.048)
	sex	-0.505	0.571	(-1.618, 0.622)	-0.390	0.591	(-1.561, 0.781)
	EDUGRP	-1.200	0.679	(-2.498, 0.156)	-0.470	0.698	(-1.853, 0.913)
	τ^2	4.254	0.459	(3.378, 5.177)	-	-	-