Published in final edited form as:

Stat Med. 2008 July 10; 27(15): 2826–2849. doi:10.1002/sim.3111.

Imputation-based strategies for clinical trial longitudinal data with nonignorable missing values

Xiaowei Yang^{1,2,*†}, Jinhui Li^{2,3}, and Steven Shoptaw⁴

¹ Division of Biostatistics, School of Medicine, University of California, Med Sci 1-C, Suite 200, Davis, CA 95616, U.S.A

² BayesSoft Inc., 2221 Caravaggio Drive, Davis, CA 95618, U.S.A

³ Department of Statistics, UCLA, 8130 Math Sciences Bldg, P.O. Box 95155, Los Angeles, CA 90095-1554, U.S.A

⁴ Department of Family Medicine, UCLA, 10880 Wilshire Blvd., Suite 1800, Los Angeles, CA 90095-7087, U.S.A

SUMMARY

Biomedical research is plagued with problems of missing data, especially in clinical trials of medical and behavioral therapies adopting longitudinal design. After a literature review on modeling incomplete longitudinal data based on full-likelihood functions, this paper proposes a set of imputation-based strategies for implementing selection, pattern-mixture, and shared-parameter models for handling intermittent missing values and dropouts that are potentially nonignorable according to various criteria. Within the framework of multiple partial imputation, intermittent missing values are first imputed several times; then, each partially imputed data set is analyzed to deal with dropouts with or without further imputation. Depending on the choice of imputation model or measurement model, there exist various strategies that can be jointly applied to the same set of data to study the effect of treatment or intervention from multi-faceted perspectives. For illustration, the strategies were applied to a data set with continuous repeated measures from a smoking cessation clinical trial.

Keywords

multiple partial imputation; selection model; pattern-mixture model; Markov transition model; nonignorable dropout; intermittent missing values

1. INTRODUCTION

1.1. Background

In biomedical research with clinical trials, the effectiveness of a treatment or intervention method is often investigated by adopting a longitudinal design where each subject is repeatedly measured on the variables of interest throughout a period of time. In such studies, there are often missing values reflecting the problematic nature of the phenomenon under study, such as substance abuse [1] and mental health disorder [2]. The proportion of missingness is sometimes notably large, e.g. 70 per cent at termination in a randomized trial

*Correspondence to: Xiaowei Yang, Division of Biostatistics, Department of Public Health Sciences, Med. Sci. 1-C, University of California, Davis, CA 95616, U.S.A.

†XDYang@UCDavis.edu

of buprenorphine *versus* methadone in treating addiction to cocaine use [3]. Although investigators may devote substantial effort in minimizing the number of missing values, some amount of missingness is often inevitable in practice.

A convenient way for incomplete longitudinal data analysis is to ignore those missing values and fit a model based on all available data. Currently, three groups of longitudinal models are popularly used in this way: marginal models with generalized estimating equations (GEE) for studying the group-averaged characteristics [4], mixed models with random effects used to describe heterogeneity among individuals [5], and transition models, which model the sequence of responses from a person dynamically by conditioning on previous observations and baseline features [6]. Although different assumptions are required by each modeling option, they essentially require that missing values be at least '*ignorable*,' i.e. the indicator variable for whether a measure is missing is independent of the value of that measure given other observed measures and covariates [7,8]. Here, the independence between the missingness indicators and missing values also requires that the parameters in modeling the repeated measures and those modeling the mechanism of missingness are distinct. Without specifying likelihood functions, the method of GEE offers a flexible modeling strategy. Although the standard version of GEE requires a stronger assumption, various adjusting methods have been proposed to handle ignorable missing values [9].

Unfortunately, there are some studies where empirical evidence suggests that *ignorability* is implausible [10–12]. In this case, when standard modeling options are used, invalid statistics or biased estimators may be obtained. Hence, advanced modeling strategy assuming *nonignorable missingness* becomes desirable. In previous years, several lines of modeling development for nonignorable missingness have been initiated based on modeling the joint distribution of the indicators of missingness and the values of repeated measures (including observed and missing values). The likelihood function derived from the joint distribution is called the full-likelihood function [13]. As summarized by Li and coworkers [14], at least three factorizations of the joint distribution could be considered: (1) outcome-dependent factorization, where missingness indicators are assumed to be conditioned on the values of repeated measures; (2) pattern-dependent factorization, where the distribution of repeated measure values is a mixture of distributions for subjects within sub-groups determined by the patterns of missingness; and (3) parameter-dependent factorization, where repeated measure values and missingness indicators are conditionally independent of each other given a group of shared parameters. Correspondingly, three models could be conceived according to the way of factorization and are termed, respectively, as selection models, pattern-mixture models, and shared-parameter models.

Compared with intermittent missingness (occasional omission), dropout (premature withdrawal) usually leads to a larger proportion of missingness and the mechanism for dropout is often associated with both missing and observed values (as well as baseline covariates such as treatment assignment). Hence, dropout is more problematic. To deal with nonignorable dropouts (i.e. missing values after withdrawal), Diggle and Kenward [10] proposed the method of selection model while comparing diets of cows for increasing their milk protein content. Using the data set from a multi-center trial with a parallel group design to study the efficacy of Vorozole for treating breast cancer, Molenburghs and colleagues [15] developed strategies for fitting pattern-mixture models for dropouts. For repeatedly measured count data subject to dropout, Albert and Follmann [16] introduced the prototype of a shared-parameter model. For a detailed review on modeling dropout mechanisms, refer to Little [17]. The presence of intermittent missing values in addition to dropouts further complicates the modeling procedures. Unlike dropouts, the patterns of intermittent missingness can be of any nonmonotonic form. Therefore, intermittent missingness is technically difficult to handle, although conceptually easy. According to Troxel and

coauthors [18], the selection model was extended to include the case of intermittent missing values. A breakthrough contribution was given by Albert and Follmann [19], who proposed the shared-random-effects Markov transition model (REMTM) to deal with nonignorable missingness in longitudinal binary data. This model was further generalized by Li *et al.* [14] to accommodate Poisson-distributed count measures.

Since the subjects in question still remain in the study, we may be able to assume that intermittent missingness is ignorable [6] when the proportion of missingness is moderate. Adopting this point of view, Yang and Shoptaw [20] developed the idea of *multiple partial imputation* (MPI) to assess dropout mechanisms when there are intermittent missing values. Within the framework of MPI, intermittent missing values are first assumed to be ignorable and imputed using an outcome-dependent modeling technique; then the partially imputed data sets are analyzed to investigate the dropout mechanism; and finally, the multiple versions of assessment are combined to make one final set of inferential statements.

The approach of MPI provides a much more general framework, not only for data exploration but also for analysis purposes. This article proposes strategies for implementing incomplete longitudinal models to analyze continuous repeated measures with intermittent missing values and dropouts. The article is organized as follows. We begin by introducing a motivating practical data set from a smoking cessation clinical trial where a moderate amount of missing values is seen. In Section 2, we discuss modeling strategies based on full-likelihood functions for incomplete longitudinal data with special attention to handle nonignorable dropouts. In Section 3, we propose imputation-based strategies and Markov chain Monte Carlo (MCMC) algorithms for implementing various incomplete longitudinal models. The smoking cessation data set is then analyzed using the above strategies, and finally we give some practical guidelines for using the proposed imputation-based strategies.

1.2. A motivating study

The development of this work was closely related to the analysis of a clinical trial of smoking cessation in methadone-maintained tobacco smokers [21]. This study tested the effectiveness of a relapse prevention (RP) program and a contingency management (CM) program, alone and in combination, for improving smoking cessation outcomes using nicotine transdermal pharmacotherapy. A total of 174 participants were randomly assigned to one of the four behavioral treatment groups: a control group that received no behavioral therapy (42 subjects); RP-only (42 subjects); CM-only (43 subjects); and a combined RP +CM condition (47 subjects). Thirty-six measures of carbon monoxide levels in expired breath were scheduled to be taken on each participant over the 12-week study period, three times per week.

Figure 1 depicts the mean values of observed carbon monoxide levels for the four treatment groups, after a $\log(1 + y)$ transformation. Also depicted are standard deviations and point-wise ANOVA results with p -values smaller than 0.01 after ignoring missing values. A problem with this exploratory analysis is that the 36 p -values cannot be easily combined in making inferences regarding overall differences. Additionally, the comparison between treatment conditions when missing values are ignored may lead to biased conclusions when missingness is not *completely at random*. For example, if smokers in the three treatment groups dropped out with higher probabilities given a higher level of previously observed carbon monoxide while smokers in the control group dropped out completely at random, then mean levels of carbon monoxide in the treatment groups would turn out to be lower than those in the control group at visit times close to the termination of the study, even though there are no treatment effects at all.

In Figure 2, the patterns of missingness are plotted for each treatment group, after a sorting process on the dropout times. From the graphs, it is seen that missingness due to dropout corresponds to monotonic forms. At the termination of the study, up to 36 per cent of the participants had withdrawn. An overall percentage of 4.3 per cent of intermittent missing values is seen. The patterns and rates of missing values in this study are typical in substance abuse research. Assuming that intermittent missing values and dropouts were ignorable, random-effects models were applied to the whole incomplete data set and significantly favorable effect of CM was reported in Shoptaw *et al.* [21]. In Section 4, we will reanalyze this set of carbon monoxide levels to illustrate various modeling strategies introduced in the following two sections.

2. MODELING INCOMPLETE LONGITUDINAL DATA

For a longitudinal data set with balanced design, J repeated measures are potentially observed on each of the N independent subjects at times $t_{i1}, \dots, t_{iJ}$ ($i = 1, \dots, N; j = 1, \dots, J$). For the following discussion, we use capital letters to represent variables (e.g. $Y_1, \dots, Y_J$ indicate response variables, and $X_1, \dots, X_K$ indicate covariates or explanatory variables) and lower letters for values, which may be observed or missing (e.g. y_{ij} denotes the value of Y_j and x_{ijk} denotes the value of X_k recorded at time t_{ij} ; $k = 1, \dots, K$). Bold symbols are used to represent vectors or matrices: $\mathbf{y}_i = (y_{i1}, \dots, y_{iJ})^T$ indicates values of repeated measures; $\mathbf{X}_i = [x_{ijk}]_{J \times K}$ consists of possibly time-varying covariates for the i th subject. Assuming that repeated measures are distributed as multivariate normal, a repeated-measures model with structured covariance matrix can be expressed as $\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$, where $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$ and $\boldsymbol{\beta}$ is a vector of fixed-effects parameters. Various ways of parameterization of the covariance matrix $\boldsymbol{\Sigma}_i$ are conceivable, allowing various forms for the specification of a wide range of repeated-measures models [22].

2.1. Models with full-likelihood function

When values of some measures are missing, we partition $\mathbf{y}_i$ into two parts, $\mathbf{y}_i = (\mathbf{y}_i^{\text{obs}}, \mathbf{y}_i^{\text{mis}})^T$, where $\mathbf{y}_i^{\text{obs}}$ indicates the observed values and $\mathbf{y}_i^{\text{mis}}$ indicates values that would be observed if they were not missing. For convenience, we also introduce a vector of missingness indicators, $\mathbf{r}_i = (r_{i1}, \dots, r_{iJ})^T$, where $r_{ij} = 0$ (or 1) indicates whether y_{ij} is observed (or missing), and $\mathbf{R} = [\mathbf{r}_{ij}]_{N \times J}$ represents the missingness patterns for the whole data matrix $\mathbf{Y} = [\mathbf{y}_{ij}]_{N \times J}$. Theoretically, the joint distribution of the observed data and missingness patterns should be modeled in statistical analysis based on the full-likelihood function, i.e.

$$L(\boldsymbol{\theta}, \boldsymbol{\phi} | \mathbf{y}_i^{\text{obs}}, \mathbf{X}_i, \mathbf{r}_i) \propto \prod_{i=1}^N \int f(\mathbf{y}_i, \mathbf{r}_i | \mathbf{X}_i, \boldsymbol{\theta}, \boldsymbol{\phi}) d\mathbf{y}_i^{\text{mis}}$$

where vectors $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, respectively, represent the parameters of the measurement model and those of the missingness mechanism. Determined by possible causal pathways, there exist at least three approaches to decompose the joint distribution of the complete data and missingness indicators: outcome-dependent factorization, pattern-dependent factorization, and parameter-dependent factorization. Accordingly, we have the following models for incomplete longitudinal data.

Selection model factors the joint distribution $f(\mathbf{y}_i, \mathbf{r}_i | \mathbf{X}_i, \boldsymbol{\theta}, \boldsymbol{\phi})$ into a marginal distribution of $\mathbf{y}_i$ and a conditional distribution of $\mathbf{r}_i$ given $\mathbf{y}_i$ (i.e. outcome dependent),

$$f(\mathbf{y}_i, \mathbf{r}_i | \mathbf{X}_i, \boldsymbol{\theta}, \boldsymbol{\phi}) = f(\mathbf{y}_i | \mathbf{X}_i, \boldsymbol{\theta}) f(\mathbf{r}_i | \mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\phi})$$

where $f(\mathbf{r}_i | \mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\phi})$ can be interpreted as ‘self-selection of the i th subject into a specific missingness-pattern group.’

Pattern-mixture model is a pattern-dependent model, assuming that distribution of repeated measures varies with the missingness patterns and the joint distribution is factored as

$$f(\mathbf{y}_i, \mathbf{r}_i | \mathbf{X}_i, \boldsymbol{\theta}, \boldsymbol{\phi}) = f(\mathbf{y}_i | \mathbf{r}_i, \mathbf{X}_i, \boldsymbol{\theta}) f(\mathbf{r}_i | \mathbf{X}_i, \boldsymbol{\phi})$$

Assuming that there are P patterns of missingness in a data set, the marginal distribution of

$\mathbf{y}_i$ would be a mixture of pattern-specific distributions, $f(\mathbf{y}_i) = \sum_{p=1}^P f(\mathbf{y}_i | r_i=p, \mathbf{X}_i, \boldsymbol{\theta}^{(p)}) \pi_p$, where $\boldsymbol{\theta}^{(p)}$ represents the parameters of $f(\mathbf{y}_i)$ in the p th pattern, $\pi_p = \Pr(r_i = p | \mathbf{X}_i, \boldsymbol{\phi})$ and indicator r_i here indexes the missingness patterns.

Shared-parameter model assumes that $\mathbf{y}_i$ and $\mathbf{r}_i$ are conditionally independent of each other, given a group of parameters $\boldsymbol{\xi}_i$, i.e.

$$f(\mathbf{y}_i, \mathbf{r}_i | \mathbf{X}_i, \boldsymbol{\theta}, \boldsymbol{\phi}) = \int f(\mathbf{y}_i | \boldsymbol{\xi}_i, \mathbf{X}_i, \boldsymbol{\theta}) f(\mathbf{r}_i | \boldsymbol{\xi}_i, \mathbf{X}_i, \boldsymbol{\phi}) f(\boldsymbol{\xi}_i) d\boldsymbol{\xi}_i$$

From the viewpoint of causation, shared ‘parameters’ play the role of confounders for the relationship between $\mathbf{y}_i$ and $\mathbf{r}_i$ and, hence, can be either observable attributes (e.g. gender) or latent factors (e.g. random effects).

2.2. Ignorability

In certain biomedical studies, both missingness patterns and values of repeated measures are of interest. For example, in a heart-disease study, the repeatedly measured blood pressures and the survival lengths (a form of dropout patterns) of the patients are apt to be modeled jointly [23]. Within this scenario, the above selection, pattern-mixture, and shared-parameter models can be applied directly or after some modification. In the majority of biomedical research, however, only the parameters for the distribution of the repeated measures (i.e. $\boldsymbol{\theta}$) are of interest, while those related to missingness patterns are viewed as nuisance parameters. In this latter scenario, it would be desirable that we could ignore the missing values when making inferences regarding $\boldsymbol{\theta}$.

Within the setting of selection models, the concept of ‘ignorability’ was defined by Rubin [8] and extensively addressed thereafter. Missing values are said to be *ignorable* when two conditions hold: (i) $\mathbf{r}_i$ is independent of $\mathbf{y}_i^{\text{mis}}$, given $\mathbf{y}_i^{\text{obs}}$ and $\mathbf{X}_i$, and (ii) $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are distinct. Under *ignorability*, the log-likelihood function for $\boldsymbol{\theta}$ can be separated from the log-likelihood function for $\boldsymbol{\phi}$, i.e. $l(\boldsymbol{\theta}, \boldsymbol{\phi} | \mathbf{y}_i^{\text{obs}}, \mathbf{r}_i) = l(\boldsymbol{\theta} | \mathbf{y}_i^{\text{obs}}) + l(\boldsymbol{\phi} | \mathbf{y}_i^{\text{obs}}, \mathbf{r}_i)$. Little and Rubin [7] further classified this type of ignorability (based on outcome-dependent factorization) into two sub-categories: MCAR (i.e. $\Pr(\mathbf{r}_i | \mathbf{y}_i, \boldsymbol{\phi}) = \Pr(\mathbf{r}_i | \boldsymbol{\phi})$) and MAR (i.e. $\Pr(\mathbf{r}_i | \mathbf{y}_i, \boldsymbol{\phi}) = \Pr(\mathbf{r}_i | \mathbf{y}_i^{\text{obs}}, \boldsymbol{\phi})$). For intermittent missing values, ignorability or nonignorability could be interpreted by whether the missing values can be unbiasedly interpolated from neighborhood observed values. For dropouts, the diagnostics of ignorability corresponds to test whether missing values after dropout can be unbiasedly extrapolated from the previous observed values. In certain

applications, occasional omissions or nonresponses happen due to reasons unrelated to the outcome (e.g. schedule conflicts or bad weather) and the missing values could be unbiasedly imputed from observed values or follow-up inquiries. Therefore, ignorability can be assumed for them. Nonetheless, subjects withdraw from a study usually because of study-related reasons (e.g. being unsatisfied with the intervention or its side effects) with a mechanism similar to outcome-dependent censoring and, hence, are usually nonignorable [6,12,24].

Within the context of pattern-mixture or shared-parameter models, we define ignorability as a condition under which observed data can be used to estimate θ without bias. For pattern-mixture models, so long as $\mathbf{y}_i^{\text{mis}}$ does not depend on $\mathbf{r}_i$ (given $\mathbf{y}_i^{\text{obs}}$ and $\mathbf{X}_i$), missing data are thought to be ignorable. For shared-parameter models, ignorability corresponds only to the case where ξ_i are observable confounders, which are usually viewed as a subset of $\mathbf{X}_i$. Otherwise, when latent variables such as random effects are shared, a shared-parameter model would generally associate with the assumption of nonignorability. It is also noted here that ‘informative’ is sometimes used to describe a specific form of nonignorability, e.g. ‘informative dropout’ (within the context of selection model, [10]) and ‘informative process for missingness’ (within the context of shared-parameter models [19]).

2.3. Modeling nonignorable dropouts

As seen in Figure 2, dropout patterns display monotonic forms after sorting on the time of withdrawal. This feature makes it easier to characterize the dropout mechanism. Also, considering the fact that dropouts are more problematic in practice, we focus on modeling nonignorable dropouts in this paper.

2.3.1. Selection model—Let us denote t_{d_i} as the dropout time for the i th subject, where $2 \leq d_i \leq J+1$ ($d_i = J+1$ indicates a subject who has completed the study). Then, missingness indicator $\mathbf{r}_i$ is a vector of $d_i - 1$ consecutive zeros followed by $J+1 - d_i$ consecutive ones. Suppressing the dependence on covariates, the selection model of Diggle and Kenward [12] assumes: (i) $\Pr(r_{ij} = 1 | j > d_i) = 1$; (ii) for $j \leq d_i$, r_{ij} depends on y_{ij} and its history $\mathbf{H}_{ij} = (y_{i1}, \dots, y_{i,j-1})^T$; and (iii) the conditional distribution of y_{ij} given $\mathbf{H}_{ij}$ is $f_{ij}(y | \mathbf{H}_{ij}, \theta)$. The full-likelihood function for the i th subject can be expressed as

$$L_i(\theta, \phi | \mathbf{y}_i^{\text{obs}}, \mathbf{r}_i) \propto \prod_{j=1}^{d_i-1} f(y_{ij} | \mathbf{H}_{ij}, \theta) \prod_{j=1}^{d_i-1} [1 - p_j(y_{ij}, \mathbf{H}_{ij})] \Pr(r_{id_i} = 1 | \mathbf{H}_{id_i})$$

where $p_j(y_{ij}, \mathbf{H}_{ij}) = \Pr(r_{ij} = 1 | y_{ij}, \mathbf{H}_{ij}, \phi)$ represents the probability of dropout at time t_{ij} . Dropout probability $\Pr(r_{id_i} = 1 | \mathbf{H}_{id_i}) = \int \Pr(r_{id_i} = 1 | y, \mathbf{H}_{id_i}, \phi) f_{id_i}(y | \mathbf{H}_{id_i}, \theta) dy$ if $d_i < J+1$ (y represents the possible value of y_{id_i}) and $\Pr(r_{id_i} = 1 | \mathbf{H}_{id_i}) = 1$ if $d_i = J+1$. A natural choice for calculating $\Pr(r_{ij} = 1 | y_{ij}, \mathbf{H}_{ij}, \phi)$ is a logistic regression model:

$$\text{logit}(\Pr(r_{ij} = 1 | y_{ij}, \mathbf{H}_{ij}, \phi)) = \varphi_0 + \varphi_1 y_{ij} + \sum_{k=2}^j \varphi_k y_{i,j+1-k} \quad (i=1, \dots, N, j=1, \dots, J)$$

where φ_1 with a nonzero value implies an outcome-dependent nonignorable dropout mechanism.

The full log-likelihood function of the whole data set with sample size N for θ and ϕ can be partitioned into $l(\theta, \phi) = l_1(\theta) + l_2(\phi) + l_3(\phi, \theta)$, where $l_1(\theta) = \sum_{i=1}^N \log\{f(\mathbf{y}_i^{\text{obs}})\}$ corresponds to the observed-data log-likelihood function for θ , and $l_2(\phi) = \sum_{i=1}^N \sum_{j=1}^{d_i-1} \log\{1 - p_j(\mathbf{H}_{ij}, y_{ij})\}$ and $l_3(\phi, \theta) = \sum_{i \leq N; d_i \leq J} \log\{\Pr(r_{id_i} = 1 | \mathbf{H}_{id_i})\}$ together determine the log-likelihood function of the dropout process. If dropouts are ignorable, then $l_3(\theta, \phi)$ depends only on ϕ and can be absorbed into $l_2(\phi)$; thus estimation of θ can be solely derived from $l_1(\theta)$.

As shown by Verbeke and Molenburghs [12], the idea of the selection model could be originated from the Tobit model of Heckman [25]. Later, Troxel *et al.* [18] further extended it to handle nonmonotone missing values. Selection models for categorical and other types of repeated measures were also developed, see [26–29].

2.3.2. Pattern-mixture model—The high sensitivity of selection modeling to misspecification of measurement process and dropout mechanism has led to a growing interest in pattern-mixture modeling [30,31]. After initial introduction [32,33], they received more attention lately, e.g. for continuous repeated measures [17,34–37] and for categorical measures [15,38,39].

For dropouts, a pattern-mixture model factorizes the joint distribution $f(\mathbf{y}_i, d_i | \mathbf{X}_i, \theta, \phi)$ into the product of the marginal distribution $f(d_i | \mathbf{X}_i, \phi)$ and the conditional distribution $f(\mathbf{y}_i | \mathbf{X}_i, \theta^{(d_i)})$, where $d_i = 2, \dots, J+1$ indicates the dropout time. A big challenge in pattern-mixture modeling regards parameter identification. For any subject with early withdrawal ($d_i < J+1$), the sub-vector of $\theta^{(d_i)}$ characterizing the distribution of missing values ($\mathbf{y}_i^{\text{mis}}$) is generally unidentified, unless certain restrictions are applied. Thijs *et al.* [31] proposed a framework for identifying restrictions. By suppressing the subscript i and replacing d_i with j , we can express the full probability density function for the pattern with dropout time at t_j , i.e.

$$f_j(\mathbf{y}) = f_j(\mathbf{y}^{\text{obs}}) f_j(\mathbf{y}^{\text{mis}} | \mathbf{y}^{\text{obs}})$$

where $\mathbf{y}^{\text{obs}} = (y_1, \dots, y_j)^T$ and $f_j(\mathbf{y}^{\text{mis}} | \mathbf{y}^{\text{obs}})$ is the conditional distribution, which cannot be identified within the j th pattern. By borrowing information from the observed data in other patterns where the corresponding measurements $y_s \in \mathbf{y}^{\text{mis}}$ ($s = j+1, \dots, J$) are observed, it is possible to restrict $f_j(\mathbf{y}^{\text{mis}} | \mathbf{y}^{\text{obs}})$. After introducing some proper weights (i.e. $\sum_{t=s}^J \omega_{st} = 1$), we can identify $f_j(y_s | y_1, \dots, y_{s-1})$ by

$$f_j(y_s | y_1, \dots, y_{s-1}) = \sum_{t=s}^J \omega_{st} f_t(y_s | y_1, \dots, y_{s-1}), \quad s = j+1, \dots, J$$

Using this restriction method, the full density function $f_j(\mathbf{y})$ can be expressed as

$$f_j(\mathbf{y}) = f_j(\mathbf{y}^{\text{obs}}) \prod_{s=0}^{J-j-1} \left[\sum_{t=J-s}^J \omega_{J-s,t} f_t(y_{J-s} | y_1, \dots, y_{J-s-1}) \right]$$

Depending on the specification of the weights, several schemes of identification can be implemented. Setting all the weights to positive values corresponds to the identification scheme called *available case missing values* (ACMV [36]), which is the natural counterpart of the mechanism of MAR in the context of selection models. The restriction scheme called *complete-cases missing variable* (CCMV [32]) identifies $f_j(y_s|y_1, \dots, y_{s-1})$ by borrowing information only from the subjects who have completed the study, i.e.

$$f_j(y_s|y_1, \dots, y_{s-1}) = f_j(y_s|y_1, \dots, y_{s-1}), \quad s = j+1, \dots, J$$

In this case, weights are set as $\omega_{sJ} = 1$ and $\omega_{ss} = \omega_{s,s+1} = \dots = \omega_{s,J-1} = 0$. Another special case of identification scheme—*neighboring case missing values* (NCMV)—borrows information from neighbor patterns with observed values on y_s , i.e.

$$f_j(y_s|y_1, \dots, y_{s-1}) = f_s(y_s|y_1, \dots, y_{s-1}), \quad s = j+1, \dots, J$$

which corresponds to $\omega_{ss} = 1$ and $\omega_{s,s+1} = \omega_{s,s+2} = \dots = \omega_{s,J} = 0$.

2.3.3. Shared-parameter model—When the dynamic transition features in longitudinal data are of interest, an appropriate analytical approach is *via* transition models, which use previous observations to predict current ones. Here, we propose a shared-parameter model called REMTM to deal with continuous repeated measures subject to nonignorable dropout. Similar models have been proposed to analyze binary or count measures with nonmonotone missing values [14,19]. Within REMTM, for each subject the repeated measures ($\mathbf{y}_i$) are conditionally independent of the missingness indicators ($\mathbf{r}_i$), given the shared-parameters—random effects (ξ_i). Therefore, we can separately characterize the measurement process $p(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\theta}, \xi_i)$ and the dropout process $p(\mathbf{r}_i | \mathbf{x}_i, \boldsymbol{\phi}, \xi_i)$.

To model the measurement process, an order-1 Markov chain can be assumed for $\mathbf{y}_i = (y_{i1}, \dots, y_{iJ})^T$, where y_{ij} is independent of $(y_{i1}, \dots, y_{i,j-2})^T$ if the previous observation $y_{i,j-1}$ is given. To capture the baseline heterogeneity across subjects, we use a random-intercept effect (ξ_i). Therefore, the part of REMTM characterizing the measuring process is itself a regression type of linear transition model

$$y_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta} + (y_{i,j-1} - \mathbf{x}_{i,j-1}\boldsymbol{\beta})\alpha + \xi_i + \varepsilon_{ij}$$

where $\xi_i \stackrel{\text{iid}}{\sim} N(0, \sigma_\xi^2)$ denotes the random intercept for subject i , $\varepsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2)$ represents the residual errors as seen in standard linear regression models. To one's most interest, $\boldsymbol{\beta}$ contains the fixed parameters regarding treatment efficacy in clinical trials. The parameter α sets up the link between the previous and the current unexplained measurement effects, i.e. between $y_{i,j-1} - \mathbf{x}_{i,j-1}\boldsymbol{\beta}$ and $y_{ij} - \mathbf{x}_{ij}\boldsymbol{\beta}$.

To model the dropout process, a logistic transition model with random intercepts is used. For missingness indicators $\mathbf{r}_i = (r_{i1}, \dots, r_{iJ})^T$ with $r_{ij} = 0$ (or 1) if y_{ij} is observed (or missing due to dropout), a first-order Markov chain is assumed with a 2×2 matrix of transition probabilities: $P_{lk} = \Pr(r_{ij} = k | r_{i,j-1} = l)$ ($l = 0$ or 1 ; $k = 0$ or 1). By the definition of dropout, we always have $P_{10} = 0$ and $P_{11} = 1$. The transition probabilities P_{00} and P_{01} are calculated as

$$P(r_{ij}=k|\xi_i, \mathbf{x}_{ij}, r_{i,j-1}=0) = \begin{cases} \frac{1}{1+\exp(\mathbf{x}_{ij}\boldsymbol{\eta}+\xi_i\gamma)} & \text{if } k=0 \\ \frac{\exp(\mathbf{x}_{ij}\boldsymbol{\eta}+\xi_i\gamma)}{1+\exp(\mathbf{x}_{ij}\boldsymbol{\eta}+\xi_i\gamma)} & \text{if } k=1 \end{cases}$$

where parameters $\boldsymbol{\eta}$ calibrate the influence of covariates on the possibility of dropout. The parameter γ indicates whether the dropout process shares the random intercepts with the measurement process. Hence, it tells us whether dropout is informative.

By combining the above sub-models for measurement and dropout, we can write the full-likelihood function for parameters $\boldsymbol{\theta}=(\boldsymbol{\beta}^T, \sigma_\xi^2, \sigma_\varepsilon^2)^T$ and $\boldsymbol{\phi}=(\boldsymbol{\eta}, \gamma)^T$, i.e.

$$L(\boldsymbol{\theta}, \boldsymbol{\phi}) \propto \prod_{i=1}^N \int \left\{ \prod_{j=1}^{d_i-1} p(y_{ij}|\mathbf{x}_{ij}, y_{i,j-1}, \xi_i, \boldsymbol{\theta}) \right\} \left\{ \prod_{j=1}^{d_i} p(r_{ij}|\mathbf{x}_{ij}, r_{i,j-1}, \xi_i, \boldsymbol{\phi}) \right\} p(\xi_i) d\xi_i$$

where $p(\xi_i)$ is the density function for normally distributed random intercept ξ_i . This model could be naturally generalized to deal with more complicated cases, e.g. incomplete data with intermittent missingness and dropout, nested random effects, or other types of shared parameters [3,16,40–45].

3. IMPLEMENTATION VIA MPI

In practical settings, it is common to have longitudinal data sets with both intermittent missing values and dropouts. To deal with such data sets, this section proposes a series of imputation strategies to implement the full-likelihood-based models introduced earlier.

3.1. MPI and two-stage MPI

It has been popular to do statistical analysis with the method of imputation, which fills the empty cells in a data matrix by ‘plausible’ values predicted from empirical evidence or assumption-driven models. Imputation enforces an incomplete data set into a complete one. Thus, standard complete-data modeling techniques can be applied afterward. For repeated measures, the method of imputation is especially useful since repeated measures are often highly correlated. By imputing a data set once and treating it as the actual complete data set, the uncertainty in parameter estimation is apparently underestimated. To overcome this limitation, Rubin [46] proposed the method called *multiple imputation* within which multiple sets of imputed values are generated for the same set of missing values.

3.1.1. Inference via MPI—As argued earlier, intermittent missingness and dropout differ not only in pattern but also in mechanism and hence should be treated in different ways. Yang and Shoptaw [20] introduced the idea of MPI by conducting imputations only for intermittent missing values so that dropouts can be isolated and treated differently. MPI potentially offers a generic solution that can be applied throughout the whole spectrum of longitudinal data analysis. By partitioning $\mathbf{y}_i^{\text{mis}}$ into $(\mathbf{y}_i^{\text{IM}}, \mathbf{y}_i^{\text{DM}})$ to denote intermittent missing values and dropouts, the first step of MPI is to draw $m > 1$ sets of imputations for the intermittent missing values: $\mathbf{y}_i^{\text{IM}(1)}, \mathbf{y}_i^{\text{IM}(2)}, \dots, \mathbf{y}_i^{\text{IM}(m)}$ ($i=1, \dots, N$). Then, in the second step, each partially imputed data set $(\mathbf{Y}^{\text{obs}}, \mathbf{Y}^{\text{IM}(j)})$ ($j=1, \dots, m$) along with $\mathbf{X}$ (i.e. data on covariates) is treated with selection, pattern-mixture, shared-parameter, or any other modeling strategy to deal with potentially nonignorable dropouts. Finally, in the third step, multiple versions of analysis are consolidated to derive an overall inference.

For this final step, a set of rules for consolidation was originally developed by Rubin [46] and later improved by Rubin and Schenker [47]. For the j th ($j = 1, \dots, m$) imputed data set, we denote $\hat{Q}^{(j)}$ as the point estimate of Q (a parameter or quantity of interest) and $\hat{U}^{(j)}$ as the corresponding variance estimate. Then the MPI estimate (overall point estimate) of Q is $\bar{Q} = (1/m) \sum_{j=1}^m \hat{Q}^{(j)}$. The associated variance of $\bar{Q}$ is $T = \bar{U} + ((m+1)/m)B$, where $\bar{U} = (1/m) \sum_{j=1}^m \hat{U}^{(j)}$ and $B = (1/(m-1)) \sum_{j=1}^m (Q^{(j)} - \bar{Q})^2$, respectively, represent the *within-imputation variability* and the *between-imputation variability*. For hypothesis testing, we can use the statistic $(\bar{Q} - Q)T^{-1/2}$, which has approximately t -distribution with degrees of freedom $v = (m-1)[1 + \bar{U}/(1 + m^{-1})B]^2$. Making proper MPI inferences requires that the multiple imputations be created ‘independently.’ In Section 3.2, we will see how to create independent imputed values *via* MCMC algorithms.

3.1.2. Inference via two-stage MPI—Within the second step of the above MPI inferential procedure, additional multiple imputations can be made for the dropouts when fitting a selection, pattern-mixture, or shared-parameter model. That is, for each partially imputed data set, $(\mathbf{Y}^{\text{obs}}, \mathbf{Y}^{\text{IM}(j)})$, we draw $n > 1$ sets of imputations for the dropouts: $\mathbf{y}_i^{\text{DM}(j,1)}, \mathbf{y}_i^{\text{DM}(j,2)}, \dots, \mathbf{y}_i^{\text{DM}(j,n)}$ ($i=1, \dots, N$). For emphasis, this version of MPI with sequential imputations is called a *two-stage MPI* in this article. When creating imputations for dropouts, the predictive density functions $p(\mathbf{y}_i^{\text{DM}(j,k)} | \mathbf{y}_i^{\text{obs}}, \mathbf{y}_i^{\text{IM}(j)})$ have different forms depending on the modeling strategy used ($j = 1, \dots, m; k = 1, \dots, n$). In Section 4.2, we will see that this method is especially useful for implementing pattern-mixture models. After the two-stage imputation, we would obtain $m * n$ complete data sets with $\mathbf{y}_i^{(j,k)} = (\mathbf{y}_i^{\text{obs}}, \mathbf{y}_i^{\text{IM}(j)}, \mathbf{y}_i^{\text{DM}(j,k)})$ ($i=1, \dots, N; j=1, \dots, m; k=1, \dots, n$). Each can be analyzed by traditional longitudinal models, e.g. marginal models with GEE or linear mixed-effects models, since concerns regarding missingness and dropout mechanism have been dissolved during the model-based imputation processes. Similar to MPI, the last step of the two-stage MPI is to derive the overall inference by combining the multiple analytical results. Nonetheless, Rubin’s rules presented earlier cannot be simply used by just presuming that the number of imputations is now $m * n$ instead of m . Among all the $m * n$ complete data sets, imputed values are far from being independent of each other, because each block $(\mathbf{y}_i^{(j,1)}, \dots, \mathbf{y}_i^{(j,n)})$ contains identical imputed values $\mathbf{y}_i^{\text{IM}(j)}$. Adopting the idea of ANOVA with nested blocks, a modified set of Rubin’s rules was developed by Shen [48], which can be used for making two-stage MPI inference. Within other contexts (e.g. cross-sectional survey data with nonresponse), sequential imputation strategies are also seen; see Harel [49] and Rubin [50].

3.2. MPI via MCMC

One does not need to subscribe the Bayesian paradigm in developing a ‘proper’ imputation method so long as it satisfies a set of technical conditions [46] to guarantee frequency-valid inferences. An example of non-Bayesian imputation is seen in Section 3.4. Nonetheless, this set of conditions is useful in evaluating the properties of a given method but provides little guidance in practice to devising such a method. For this reason, a Bayesian process is often preferred. By specifying a parametric model based on the full-likelihood function and applying prior distributions to the unknown model parameters, we can simulate multiple independent draws from the conditional distribution of the missing data given the observed data using Bayes’ theorem. For the selection and REMTM models, such a conditional distribution would usually be too complicated to be directly simulated. As a solution, a collection of MCMC algorithms can be used. Within MCMC, parameters are drawn from a complicated distribution by forming a Markov chain that has this distribution as the

stationary distribution. One of the most popular MCMC methods is *Gibbs sampling*, which simulates the conditional distribution of each component of a multivariate random variable given the other components in a cyclic manner. A series of Gibbs samplers have been developed and implemented into software packages, such as R/S-plus and SAS, to deal with various types of incomplete multivariate data [13].

3.2.1. Gibbs sampling algorithm for MPI—Conceptually, one of the Gibbs sampling algorithms dealing with multivariate normal data [13] can be modified to conduct partial imputations within MPI. By iterating the following two steps, we have the Gibbs sampling algorithm for creating imputations for intermittent missing values:

- i. *I-Step*: Draw values of intermittent missing data from their conditional predictive distribution, i.e. for $i = 1, \dots, N$

$$\mathbf{y}_i^{\text{IM}(t+1)} \sim f(\mathbf{y}_i^{\text{IM}} | \mathbf{y}_i^{\text{obs}}, \mathbf{X}_i, \mathbf{r}_i, \boldsymbol{\psi})$$

- ii. *P-step*: Conditioning on the drawn values for intermittent missing data ($\mathbf{Y}^{\text{IM}(t+1)}$), draw parameters $\boldsymbol{\psi}$ from its posterior distribution based on partially imputed data,

$$\boldsymbol{\psi}^{(t+1)} \propto f(\boldsymbol{\psi} | \mathbf{Y}^{\text{obs}}, \mathbf{Y}^{\text{IM}(t+1)}, \mathbf{X}, \mathbf{R})$$

In this algorithm, $\boldsymbol{\psi} = (\boldsymbol{\theta}, \boldsymbol{\phi})$ represents a vector of all parameters in a model chosen to characterize the complete data and missingness mechanism. If nonignorability is assumed for intermittent missingness, we can use a selection or REMTM model. Although detailed modeling formulations for nonignorable intermittent missingness is not given in this article, they are conceptually natural extensions of the models given in Section 2.3. For ignorable missingness, we can choose a linear mixed-effects or a transition model. By applying Bayes' theorem, we have $f(\boldsymbol{\psi} | \mathbf{Y}^{\text{obs}}, \mathbf{Y}^{\text{IM}(t+1)}, \mathbf{X}, \mathbf{R}) \propto f(\boldsymbol{\psi}) f(\mathbf{Y}^{\text{obs}}, \mathbf{Y}^{\text{IM}(t+1)} | \boldsymbol{\psi}, \mathbf{X}, \mathbf{R})$, where $f(\boldsymbol{\psi})$ represents the prior distribution for $\boldsymbol{\psi}$. Depending on the choice of modeling strategy, the conditional predictive distribution $f(\mathbf{y}_i^{\text{IM}} | \mathbf{y}_i^{\text{obs}}, \mathbf{X}_i, \mathbf{r}_i, \boldsymbol{\psi})$ has different forms.

Starting from an initial value, $\boldsymbol{\psi}^{(0)}$ (i.e. $t = 0$), which can be any reasonable value obtained using an affordable way, the above I-step and P-step can be repeated with large enough number of iterations to yield a stochastic sequence $\{(\boldsymbol{\psi}^{(t)}, \mathbf{Y}^{\text{IM}(t)}): t = 1, \dots, T\}$. Provided that certain regularity conditions [51] hold, the empirical joint distribution of the parameter and missing values within this sequence will approach the stationary distribution $f(\boldsymbol{\psi}, \mathbf{Y}^{\text{IM}} | \mathbf{Y}^{\text{obs}})$ as $T \rightarrow \infty$. In practice, we would monitor the convergence properties of the process [52]. If diagnostics suggest that convergence is achieved after T_0 iterations, we would then retain simulated missing values every $(T - T_0)/m$ iteration starting from $t = T_0 + 1$ and treat them as multiple partial imputed values, which could be approximately viewed as being independent for large T . In this way, $m > 1$ sets of partial imputations are obtained, and each could be further analyzed to deal with dropouts.

3.2.2. Gibbs sampling algorithm for two-stage MPI—A two-stage MPI procedure conducts a second round of imputation regarding the dropouts within each partially imputed data set, $(\mathbf{Y}^{\text{obs}}, \mathbf{Y}^{\text{IM}(j)})$ ($j = 1, \dots, m$). With the same flow structure as that of the sampler in Section 3.2.1, a Gibbs sampler for two-stage MPI also consists of two steps at each iteration:

- i. *I-step*: Missing values due to dropout are drawn from their conditional predictive function, i.e.

$$\mathbf{y}_i^{\text{DM}(t+1)} \sim f(\mathbf{y}_i^{\text{DM}} | \mathbf{y}_i^{\text{obs}}, \mathbf{y}_i^{\text{IM}(j)}, \mathbf{X}_i, \mathbf{r}_i, \boldsymbol{\psi}), \quad i=1, \dots, N$$

- ii. *P-step*: Conditioning on the drawn values for dropouts ($\mathbf{Y}^{\text{DM}(t+1)}$), draw parameters of $\boldsymbol{\psi}$ from their complete-data posterior distribution,

$$\boldsymbol{\psi}^{(t+1)} \propto f(\boldsymbol{\psi} | \mathbf{Y}^{\text{obs}}, \mathbf{Y}^{\text{IM}(j)}, \mathbf{Y}^{\text{DM}(t+1)}, \mathbf{X}, \mathbf{R})$$

The model used in this algorithm aims at characterizing data only subject to dropout, since the intermittent missing values have been imputed in the first stage. Therefore, the parameters $\boldsymbol{\psi}$ should be different from that in the Gibbs sampler in Section 3.2.1, although the same symbol is used for presentation. A whole procedure of the two-stage MPI requires running both the above sampler and the one in Section 3.2.1. The imputation of $\mathbf{Y}^{\text{DM}}$ should be nested within each partially imputed data set ($\mathbf{Y}^{\text{obs}}, \mathbf{Y}^{\text{IM}(j)}$) ($j = 1, \dots, m$). By running this algorithm long enough, we can obtain $n > 1$ versions of imputation on $\mathbf{Y}^{\text{DM}}$ from the sampled stochastic sequence $\{(\boldsymbol{\psi}^{(t)}, \mathbf{Y}^{\text{IM}(t)}): t = 1, \dots, T\}$ and end up with $m \times n$ sets of complete data: $\{(\mathbf{Y}^{\text{obs}}, \mathbf{Y}^{\text{IM}(j)}, \mathbf{Y}^{\text{DM}(j,k)}): j = 1, \dots, m, k = 1, \dots, n\}$. Each can be analyzed using standard longitudinal models.

3.3. Implementing the selection and REMTM models within MPI

For the selection or the REMTM model, the full-likelihood function can be theoretically formulated and evaluated to make likelihood-based inferences. Unfortunately, the cost on computation is very high. When calculating the dropout probabilities in the selection model or integrating over the random effects in REMTM, time-consuming numerical solutions are demanded. Bayesian inferences based on MCMC, again, provide an affordable alternative way for fitting the models. By summarizing the sampled values on parameters, posterior distribution is naturally obtained to make exact inferences that do not count on large-sample approximation theories.

3.3.1. A Gibbs sampler for fitting a selection model with structured covariance matrix

—Recently, we developed a hybrid Gibbs sampling algorithm for selection models with structured covariance matrices [22]. The algorithm for such a model with AR(1) covariance structure is presented here. For complete continuous repeated measures with multivariate normal distribution ($\mathbf{y}_i \sim N(\mathbf{X}_i \boldsymbol{\beta}, \boldsymbol{\Sigma}(\boldsymbol{\alpha}))$) with an AR(1) structured covariance,

we have $\text{cov}(y_{ij}, y_{ik}) = \sigma^2 \rho^{|j-k|}$. It is also clear that $f(\mathbf{y}_i^{\text{obs}} | \boldsymbol{\theta}) = \prod_{j=1}^{d_i-1} f(y_{ij} | \mathbf{H}_{ij}, \boldsymbol{\theta})$ has a marginal multivariate normal distribution with the same AR(1) covariance structure (just with lower dimension). After specifying a prior distribution $f(\boldsymbol{\psi})$ for $\boldsymbol{\psi} = (\boldsymbol{\theta}^T, \boldsymbol{\phi}^T)^T$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$ and $\boldsymbol{\phi} = (\phi_0, \phi_1, \dots, \phi_J)^T$, the posterior distribution for the model parameters is

$$P(\boldsymbol{\psi} | \mathbf{Y}, \mathbf{R}) \propto \left[\prod_{i=1}^N \left\{ f(\mathbf{y}_i^{\text{obs}} | \boldsymbol{\theta}) \times \left(\prod_{j=1}^{d_i-1} [1 - p_j(y_{ij}, \mathbf{H}_{ij})] \right) \times \Pr(r_{i,d_i} = 1 | \mathbf{H}_{i,d_i}) \right\} \right] \times f(\boldsymbol{\psi})$$

This function looks fairly complicated and the integration within $\Pr(r_{i,d_i} = 1 | \mathbf{H}_{i,d_i}) = \int \Pr(r_{i,d_i} = 1 | y, \mathbf{H}_{i,d_i}, \boldsymbol{\phi}) f_{i,d_i}(y | \mathbf{H}_{i,d_i}, \boldsymbol{\theta}) dy$ requires numerical solutions. Using a Gibbs sampler, we can alleviate the computation by first imputing the missing values at withdrawal (i.e. $\{y_{i,d_i}: i = 1, \dots, N\}$) and then drawing parameters one by one conditionally on the observed and imputed data. More specifically, we have the following steps at each iteration:

- i. *I-step*: Draw missing values at withdrawal $\{y_{i,d_i}; i = 1, \dots, N\}$:

$$y_{i,d_i} \sim f_{i,d_i}(y_i^{\text{obs}} | \mathbf{x}_{i,d_i}, d_i, \boldsymbol{\psi}) \propto \frac{1}{\sqrt{2\pi v_i}} \exp \left\{ -\frac{(y_i - \mathbf{x}_{i,d_i} \boldsymbol{\beta} - \mu_i)^2}{2v_i} \right\} \times \frac{\exp(\varphi_0 + \varphi_1 y_i + \sum_{k=2}^{d_i} \varphi_k y_{i,d_i+1-k})}{1 + \exp(\varphi_0 + \varphi_1 y_i + \sum_{k=2}^{d_i} \varphi_k y_{i,d_i+1-k})}$$

where by regressing y_{i,d_i} to $\mathbf{y}_i^{\text{obs}}$, we have

$$\mu_i = \mathbf{C}_{(d_i-1)}^T \sum_{d_i-1}^{-1} (\mathbf{y}_i^{\text{obs}} - \mathbf{X}_i \boldsymbol{\beta}), v_i = \sigma^2 (1 - \mathbf{C}_{(d_i-1)}^T \sum_{d_i-1}^{-1} \mathbf{C}_{(d_i-1)}), \text{ and } \mathbf{C}_{(j)} = (\rho^{j-1}, \dots, \rho)^T. \text{ For AR(1) covariance structure, the inverse of the covariance matrix for } \mathbf{y}_i^{\text{obs}} \text{ (i.e. } \sum_{d_i-1}^{-1} \text{) has analytical expression.}$$

- ii. *P-step*: Draw parameters one by one in the following order:

1. For $i = 1, \dots, K$, draw regression coefficients:

$$\boldsymbol{\beta}_k \sim f(\boldsymbol{\beta}_k | \boldsymbol{\psi}_{\setminus \boldsymbol{\beta}_k}, \mathbf{Y}^*, \mathbf{X}, \mathbf{R}) \propto \prod_{i=1}^N \exp \left\{ -\frac{(\mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta})^T \sum_{d_i}^{-1} (\mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta})}{2} \right\}$$

where $\mathbf{y}_i^* = (y_{i1}, \dots, y_{i,d_i})^T$, $\mathbf{Y}^* = (\mathbf{y}_1^*, \dots, \mathbf{y}_N^*)^T$, and $\boldsymbol{\psi}_{\setminus \boldsymbol{\beta}_k}$ indicates all parameters except $\boldsymbol{\beta}_k$ (' $\setminus$ ' means 'except', similarly defined in the following).

2. Draw covariance parameter:

$$\rho \sim f(\rho | \boldsymbol{\psi}_{\setminus \rho}, \mathbf{Y}^*, \mathbf{X}, \mathbf{R}) \propto \prod_{i=1}^N \frac{1}{\sqrt{2\pi |\sum_{d_i}|}} \exp \left\{ -\frac{(\mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta})^T \sum_{d_i}^{-1} (\mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta})}{2} \right\}$$

3. Draw variance of residuals:

$$\sigma^2 \sim f(\sigma^2 | \boldsymbol{\psi}_{\setminus \sigma^2}, \mathbf{Y}^*, \mathbf{X}, \mathbf{R}) \propto \prod_{i=1}^N \frac{1}{\sqrt{2\pi |\sum_{d_i}|}} \exp \left\{ -\frac{(\mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta})^T \sum_{d_i}^{-1} (\mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta})}{2} \right\}$$

4. For $k = 0, \dots, J$, draw parameters of the dropout mechanism:

$$\varphi_k \sim f(\varphi_k | \boldsymbol{\psi}_{\setminus \varphi_k}, \mathbf{Y}^*, \mathbf{X}, \mathbf{R}) \propto \prod_{i=1}^N \left\{ \left[\prod_{j=2}^{d_i-1} \frac{1}{1 + \exp(\varphi_0 + \varphi_1 y_{i,j} + \sum_{k=2}^j \varphi_k y_{i,j+1-k})} \right] \times \frac{\exp(\varphi_0 + \varphi_1 y_{i,d_i} + \sum_{k=2}^{d_i} \varphi_k y_{i,d_i+1-k})}{1 + \exp(\varphi_0 + \varphi_1 y_{i,d_i} + \sum_{k=2}^{d_i} \varphi_k y_{i,d_i+1-k})} \right\}$$

In presenting the above algorithm, we have assumed noninformative prior distributions for all the parameters: normal distributions with infinite variance for $\boldsymbol{\beta}_k$, φ_k , and $\log(\sigma^2)$; and a uniform distribution for $-1 < \rho < 1$. The above algorithm is a hybrid Gibbs sampler within which other MCMC sampling methods such as Metropolis–Hastings can be applied to simulate the conditional distribution at each step. As shown by Yang and Li [53], all the conditional distributions of each sub-step (except the one for ρ) have log-concave forms and thus can be simulated using an efficient method called *adaptive rejection sampling* [54]. A convenience sampling method for ρ is the Metropolis–Hastings algorithm.

3.3.2. A Gibbs sampler for fitting REMTM—REMTM characterizes both the continuous measuring process and the dropout mechanism as Markov processes. Since the dropout indicators and the repeated measures are conditionally independent of each other given the shared-random effects, in the imputation step, it is sufficient to sample only the random effects (which can be viewed as a special group of ‘missing data’). Still, we use $\Psi = (\theta^T, \phi^T)^T$ to denote all parameters $\theta = (\alpha, \beta^T, \sigma_\xi^2, \sigma_\varepsilon^2)^T$ and $\phi = (\eta, \gamma)^T$. A Gibbs sampler for REMTM consists of the following steps:

- i. *I-step*: For $i = 1, \dots, N$, draw random intercepts:

$$\begin{aligned} \xi_i \sim f(\xi_i | \mathbf{y}_i^{\text{obs}}, \mathbf{x}_i, \Psi) &\propto \prod_{j=2}^{d_i-1} \exp \left\{ -\frac{(y_{ij} - (\mathbf{x}_{ij}\beta + (y_{i,j-1} - \mathbf{x}_{i,j-1}\beta)\alpha + \xi_i))^2}{2\sigma_\varepsilon^2} \right\} \\ &\times \prod_{j=2}^{d_i-1} \frac{1}{1 + \exp(\mathbf{x}_{ij}\eta + \xi_i\gamma)} \\ &\times \frac{\exp(\mathbf{x}_{i,d_i}\eta + \xi_i\gamma)}{1 + \exp(\mathbf{x}_{i,d_i}\eta + \xi_i\gamma)} \\ &\times \exp \left\{ -\frac{\xi_i^2}{2\sigma_\xi^2} \right\} \end{aligned}$$

- ii. *P-step*: Draw parameters one by one in the following order:

1. For $k = 1, \dots, K$, draw fixed-effects regression coefficients in the measurement model:

$$\beta_k \sim f(\beta_k | \Psi_{\setminus \beta_k}, \xi, \mathbf{Y}^*, \mathbf{X}, \mathbf{R}) \propto \prod_{i=1}^N \prod_{j=2}^{d_i-1} \exp \left\{ -\frac{(y_{ij} - (\mathbf{x}_{ij}\beta + (y_{i,j-1} - \mathbf{x}_{i,j-1}\beta)\alpha + \xi_i))^2}{2\sigma_\varepsilon^2} \right\}$$

2. Draw the parameter indicating transition from history in the measurement model:

$$\alpha \sim f(\alpha | \Psi_{\setminus \alpha}, \xi, \mathbf{Y}^{\text{obs}}, \mathbf{X}, \mathbf{R}) \propto \prod_{i=1}^N \prod_{j=2}^{d_i-1} \exp \left\{ -\frac{(y_{ij} - (\mathbf{x}_{ij}\beta + (y_{i,j-1} - \mathbf{x}_{i,j-1}\beta)\alpha + \xi_i))^2}{2\sigma_\varepsilon^2} \right\}$$

3. Draw variance of residuals in the measurement model:

$$\sigma_\varepsilon^2 \sim f(\sigma_\varepsilon^2 | \Psi_{\setminus \sigma_\varepsilon^2}, \xi, \mathbf{Y}^{\text{obs}}, \mathbf{X}, \mathbf{R}) \propto \prod_{i=1}^N \prod_{j=2}^{d_i-1} \exp \left\{ -\frac{(y_{ij} - (\mathbf{x}_{ij}\beta + (y_{i,j-1} - \mathbf{x}_{i,j-1}\beta)\alpha + \xi_i))^2}{2\sigma_\varepsilon^2} \right\}$$

4. Draw variance of random intercepts:

$$\sigma_\xi^2 \sim f(\sigma_\xi^2 | \Psi_{\setminus \sigma_\xi^2}, \mathbf{Y}^{\text{obs}}, \mathbf{X}, \mathbf{R}) \propto \prod_{i=1}^N \exp \left\{ -\frac{\xi_i^2}{2\sigma_\xi^2} \right\}$$

5. For $k = 1, \dots, K$, draw the regression coefficients in modeling the dropout mechanism:

$$\eta_k \sim f(\eta_k | \psi_{\eta_k}, \xi, Y^{\text{obs}}, \mathbf{X}, \mathbf{R}) \propto \prod_{i=1}^N \left[\prod_{j=2}^{d_i-1} \frac{1}{1 + \exp(\mathbf{x}_{ij} \boldsymbol{\eta} + \xi_i \gamma)} \times \frac{\exp(\mathbf{x}_{i,d_i} \boldsymbol{\eta} + \xi_i \gamma)}{1 + \exp(\mathbf{x}_{i,d_i} \boldsymbol{\eta} + \xi_i \gamma)} \right]$$

6. Draw the parameter indicating nonignorability:

$$\gamma \sim f(\gamma | \psi_{\gamma}, \xi, Y^{\text{obs}}, \mathbf{X}, \mathbf{R}) \propto \prod_{i=1}^N \left[\prod_{j=2}^{d_i-1} \frac{1}{1 + \exp(\mathbf{x}_{ij} \boldsymbol{\eta} + \xi_i \gamma)} \times \frac{\exp(\mathbf{x}_{i,d_i} \boldsymbol{\eta} + \xi_i \gamma)}{1 + \exp(\mathbf{x}_{i,d_i} \boldsymbol{\eta} + \xi_i \gamma)} \right]$$

Again, noninformative priors are used in the above algorithm [14]. It can be shown that all the above conditional distributions have log-concave forms, directly or after transformation. Thus, all conditional distributions can be simulated using the method of adaptive rejection sampling.

3.3.3. Implementing the selection and REMTM models within MPI and two-stage MPI

Within MPI, for each partially imputed data set, one of the above Gibbs samplers can be used to fit a selection or REMTM model to deal with dropouts. Simulated values of parameters can be summarized to obtain parameter estimates for each data set, which are then combined to make an MPI inference. An illustration of the strategy is seen in Sections 4.1 and 4.3.

For the two-stage MPI strategy, the above Gibbs samplers for the selection and REMTM models can be modified to create multiple imputations for dropouts. For the selection model Gibbs sampler, we only need to modify the I-step by drawing values on all the missing values after withdrawal $\{(y_{i,d_i}, \dots, y_{iJ}): i = 1, \dots, N\}$ instead of only those at withdrawal (i.e. $\{y_{i,d_i}: i = 1, \dots, N\}$). This is straightforward because of the congeniality feature of the multivariate normal distribution, i.e. conditional distribution of $(y_{i,d_i}, \dots, y_{iJ})^T$ given $\mathbf{y}_i^{\text{obs}}$ is also normal. For the REMTM Gibbs sampler, we need to add a sub-step within the I-Step to draw missing values given the current drawn random effects and the previous drawn parameters, that is, $\mathbf{y}_i^{\text{DM}} \sim f(\mathbf{y}_i^{\text{DM}} | \psi, \xi, \mathbf{y}_i^{\text{obs}}, \mathbf{y}_i^{\text{IM}(j)}, \mathbf{X}_i, d_i)$. Similarly, this conditional distribution is multivariate normal with parameters derived by regressing $\mathbf{y}_i^{\text{DM}}$ on $(\mathbf{y}_i^{\text{obs}}, \mathbf{y}_i^{\text{IM}(j)})$ (see I-step in Section 3.3.1), where $\mathbf{y}_i^{\text{IM}(j)}$ refers to the j th imputation for the intermittent missing values ($j = 1, \dots, n$). Running this modified Gibbs sampler for the selection or REMTM model, we can generate $m \times n$ complete data sets to be analyzed using a standard longitudinal model.

3.4. Implementing pattern-mixture models within MPI

As seen in Section 2.3.2, there are at least three schemes in identifying restrictions for the parameters in pattern-mixture models: CCMV, NCMV, and ACMV. MPI or two-stage MPI provides a convenient framework for implementing these identification schemes. Assuming that the intermittent missing values have been imputed multiple times in a previous round, now we use the pattern-mixture model to create imputations for the dropouts without employing the Bayesian paradigm.

First, we fit a model to the pattern-specific identifiable densities: $f_j(y_1, \dots, y_J)$ ($j = 2, \dots, J$ indicating the observed dropout times in a data set) and obtain maximum likelihood estimates $\hat{\theta}^{(j)}$. Second, we select an identification scheme to determine the conditional distributions of the unobserved measurements, given the observed ones: $f_{d_i}(y_{is} | y_{i1}, \dots, y_{i,d_i-1})$ ($i = 1, \dots, N; s = d_i, \dots, J$). As seen from Section 2.3.2, each of such conditional

distributions is a mixture of known normal densities for continuous repeated measures. An easy way to simulate values from the mixture distribution is to randomly select a component of the mixture (according to the weights ω_{st} 's defined in Section 2.3.2) and then draw from it. For details of implementation and examples, see Thijs *et al.* [31] and Section 4.2.

4. APPLICATION

In this section, we use the carbon monoxide data to illustrate the above imputation-based strategies. To deal with intermittent missing values, ignorability was assumed when generating MPIs. Then, for each partially imputed data set, we dealt with the possibly nonignorable dropouts using the selection, pattern-mixture, and REMTM models.

4.1. Application of the selection model with AR(1) covariance structure

Using a piecewise linear mixed-effects model with ignorability assumption on missingness, Shoptaw *et al.* [21] reported a significant treatment effect of CM. Here, we reanalyzed a subset of the data starting from the second week using the strategy of MPI. After taking the logarithmic transformation, repeated carbon monoxide levels for each participant were viewed as multivariate normally distributed (i.e. $\mathbf{y}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$). Specifying a normal prior distribution for the mean vector (i.e. $\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim N(\boldsymbol{\mu}_0, \tau^{-1}\boldsymbol{\Sigma})$) and an inverted Wishart distribution for the covariance matrix (i.e. $\boldsymbol{\Sigma} \sim W^{-1}(r, \boldsymbol{\Lambda})$), we began creating MPIs using the SAS procedure called *PROC MI* with the option of monotone missingness [55]. This procedure adopts the Gibbs sampling algorithm similar to the one described in Section 3.2.1. Since no prior information was available, Jeffery's invariance principle was used to derive the noninformative form for the normal-inverse-Wishart prior distribution. For details of implementation, please refer to Chapter 6 of Schafer [13]. The expectation maximization (EM) algorithm was first run to obtain the maximum likelihood estimates (i.e. $\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}$), which were set as the starting point to initiate the MPI Gibbs sampler. Various diagnostic tools suggested that the procedure converged within 200 iterations. By simulated one chain of parameters and missing values with a total of $T = 2200$ iterations and setting the first $T_0 = 200$ iterations as the burn-in period, four sets of imputations on intermittent missing values were obtained with an interval of 500 iterations.

For each of the partially imputed data set, we applied the selection model to analyze the carbon monoxide levels after transformation. As seen from Figure 1, the mean carbon monoxide levels decline quickly within the first week approximately from the same starting levels and then remain leveling off at different levels throughout the rest of the study period. For each partially imputed data set, we used PROC MIXED in SAS to fit linear mixed models with various predictors and covariance structures. By model comparison with AIC, the following mean structure with AR(1) covariance was supported by all the four data sets. Thus, the AR(1) selection model was used with the following mean structure for characterizing the carbon monoxide levels:

$$y_{ij} = \beta_0 + \beta_1 CM_i + \beta_2 RP_i + \beta_3 RP_i * CM_i + \beta_4 BaseCO_i + \beta_5 Patches_i$$

where CM_i and RP_i , respectively, indicate whether the i th smoker received CM or RP, $BaseCO_i$ indicates baseline carbon monoxide level, and $Patches_i$ represents the number of nicotine patches the smoker received during the study. To model the dropout process, the following logistic regression model was selected after sequential model comparisons:

$$\text{logit}(p_{d_i}(y_{id_i}, \mathbf{H}_{ij})) = \varphi_0 + \varphi_1 Y_{i,d_i} + \varphi_2 Y_{i,d_i-1}$$

where d_i indicates the dropout time of the i th smoker ($i = 1, \dots, 174$).

By running the Gibbs sampler with noninformative priors for the AR(1) selection model, we obtained the estimates of all the parameters for each partially imputed data set. Only interesting parameters are listed in Table I from which we see that the between-imputation variance is very small for each parameter. In other words, the fraction of missing information due to intermittent missingness is low. After consolidating the four sets of estimates using Rubin's rules, it is clearly seen that the treatment effect of CM is significant ($\hat{\beta}_1 = -0.28$; $T_{2490} = -5.88$ with $p < 0.0001$). RP turns out to be ineffective and there is no significant interaction effect between CM and RP. The regression coefficient ϕ_1 is significantly larger than zero ($\hat{\phi}_1 = 1.28$; $T_{2024} = 3.86$ with $p = 0.0002$), suggesting that the higher the underlying missing value, the larger the probability of dropping out. In other words, the dropouts are possibly outcome-dependent nonignorable.

4.2. Application of pattern-mixture models and two-stage MPI

For the purpose of demonstrating pattern-mixture models, only the efficacy of CM is investigated in the following analyses. We first clustered participants into two groups: completers ($n_1 = 112$) and early terminators ($n_2 = 62$). Then within each group, the efficacy of CM was investigated. As seen from Figure 3, CM seems to be less effective for the early terminators. A linear mixed model with AR(1) covariance structure was selected (with predictors CM_i , $BaseCO_i$, and $Patches_i$) for analyzing the carbon monoxide levels starting from the second week,

$$y_{ij} = \beta_0 + \beta_1 CM_i + \beta_2 BaseCO_i + \beta_3 Patches_i$$

This model was applied separately to the completers and the early terminators. Let $\hat{\beta}_1^c$ and $\hat{\beta}_1^w$ denote the point estimators of β_1 , respectively, for the completers and the early terminators, and $\hat{\pi}_c = 64$ per cent denote the estimated probability of being complete, then the overall pointer estimator is the weighted average, $\hat{\beta}_1 = \hat{\pi}_c \hat{\beta}_1^c + (1 - \hat{\pi}_c) \hat{\beta}_1^w$, with variance derived using the delta method [56].

Since the fraction of missing information due to intermittent missing was low, only three sets of imputations for intermittent missing values were created this time. When conducting partial imputation, the procedure described in Section 4.1 was applied. The pattern-averaged point estimators and standard errors for β_1 are listed in Table II. After consolidating, the

overall point estimate is $\hat{\beta}_1 = -0.25$ with standard deviation $\sqrt{\text{var}(\hat{\beta})} = 0.13$. The test based on the t -statistic gives a p -value of 0.06.

In the above preliminary analysis, a simple pattern-mixture modeling strategy with only two target dropout patterns (complete or incomplete) was used within the framework of MPI. In the following, we describe the application of restriction identification strategies within the framework of the two-stage MPI for pattern-mixture models with larger number of dropout patterns.

When the number of target dropout patterns becomes large, the application of pattern-mixture models without imputation becomes less attractive. For example, the mean profiles of carbon monoxide levels across five dropout patterns are plotted in Plate 1, from which we observe notable within-pattern and across-pattern variances regarding the trajectory of the carbon monoxide levels. As the number of patterns increases, the number of subjects within

each pattern becomes smaller, and it becomes tedious (even infeasible) to conduct pattern-specific analysis and then combine the results across patterns as we did above.

Adopting the procedure described in Section 3.4, three restriction schemes (CCMV, NCMV, and ACMV) were used to make multiple imputations for the dropouts. Within this two-stage MPI, the numbers of partial imputations were set as $m = 2$ for the imputation of intermittent missing values and $n = 3$ for the imputation of dropouts. Again, the Gibbs sampling process with multivariate normal assumption was used for imputing intermittent missing values (see Section 4.1). Hence, we ended up with totally six complete data sets. Then, each complete data set was analyzed using a linear mixed model with AR(1) covariance structure with predictors CM, BaseCO, and Patches. Using the consolidation procedure as described in Section 3.2.2, the final point estimates and fractions of missing information for the treatment effect of CM are shown in Table III. The p -values of a one-sided hypothesis test using the t -statistics are also listed. From these results, we can see that the fraction of missing information due to dropout is much higher than that due to intermittent missingness. Two out of three identification strategies strongly support the favorable treatment efficacy of CM.

4.3. Application of the REMTM

We reanalyzed the same subset of the four groups of carbon monoxide levels using the REMTM model as we did using the selection model. The carbon monoxide data after dichotomization were analyzed by Yang *et al.* [11] using an REMTM for binary repeated measures subject to intermittent missingness and dropout. Here, the continuous data were analyzed using the hybrid Gibbs sampler for REMTM with predictors: CM_i , RP_i , and $RP \cdot CM_i$. Still with the partially imputed data sets created in Section 4.1, Table IV depicts the combined estimated posterior means, standard deviations, and 95 per cent credible intervals (CI) for all parameters of the model.

The estimated parameters for σ_ξ^2 and γ jointly suggest that dropout is random-effects dependent, hence nonignorable. The introduced random-intercept effects (i.e. ξ_i 's) capture the heterogeneity on dropout and carbon monoxide levels across the subjects. Among all the estimated parameters of η , no one is significantly different from zero. The significantly positive value of estimated α suggests that the current carbon monoxide level of an individual positively depends on the previous one. This is reasonable and consistent with the result supported by the selection model. It is of most interest that the fitted REMTM confirmed a strongly favorable treatment efficacy of CM in reducing the levels of carbon monoxide (i.e. $\beta_2 = -0.25$ with 95 per cent CI = $(-0.46, -0.05)$). This result is also consistent with that of the selection model, where $\beta_2 = -0.28$.

5. DISCUSSION

This paper introduces alternative imputation-based strategies for implementing longitudinal models with full-likelihood functions in dealing with intermittent missing values and dropouts that are potentially nonignorable. Using the carbon monoxide data set from a smoking cessation clinical trial, we have demonstrated the application of MPI and two-stage MPI to implement selection, pattern-mixture, and shared-random-effects models. We emphasize that the framework of MPI or two-stage MPI provides a very flexible solution for incomplete longitudinal data analysis. When drawing imputation on the intermittent missing values, various modeling options can be employed, depending on the assumption on the missingness mechanism (e.g. MCAR, MAR, or nonignorable). Although the formulation of selection, pattern-mixture, and shared-parameter models is only presented for repeated measures subject to dropout, it is conceivable that similar ideas be developed for intermittent missingness. When handling dropouts, we can use another group of advanced modeling

options. The models used for the two types of missingness can be totally different. For example, a selection model can be used for imputing intermittent missing values while a pattern-mixture model is used for imputing dropouts. It is also possible that imputation of intermittent missing values and analysis on dropouts are conducted by different persons at different places and times.

Another advantage of imputation-based strategies regards sensitivity analysis. As discussed earlier, a notable limitation with incomplete data analysis is that the true model and mechanism for measurements and missing values (including dropouts) are usually unverifiable in practical settings. When making MPI inferences, as mentioned above, various combinations of modeling schemes can be implemented with various assumptions. Therefore, MPI provides a useful tool in studying the sensitivity of model-based analytical conclusions. For one data set, if different MPI modeling strategies end up with inconsistent conclusions regarding the efficacy of the same treatment or intervention in a study, then further investigation should be conducted. We should try to avoid reporting with confidence a strong conclusion based only on one modeling option, while other options suggest controversial results. For the same set of data from the smoking trial, we applied various models to analyze the treatment efficacy of two behavioral therapies: CM and RP. Overall results depict a consistent image in supporting the favorable efficacy of CM.

It should also be noted that selection, pattern-mixture, and shared-parameter models are generalized versions of standard longitudinal models (i.e. marginal models using GEE, linear mixed-effects models, and transition models). For example, the linear mixed-effects model ignoring missing values can be viewed as a selection model with the MAR assumption. Though only continuous repeated measures are targeted in this article, the modeling strategies based on the full-likelihood function can be extended for other formats of repeated measures.

When describing the various Gibbs sampling algorithms for model fitting or multiple imputation, we tried to present the technical implementation as detailed as we can, but a full description for all possible modeling techniques is not possible. Most technical details are seen in the user manual and the technical report of the MPI software package. When specifying the prior distribution for the parameters, we adopted the choice of noninformative priors because there were no historical data or empirical evidence that could guide us in eliciting proper prior options. As raised by an anonymous reviewer, it is possible that flat priors are problematic in practice especially for those parameters related to missingness or dropout mechanism. This is notable when there are very few missing values, providing very little information for estimating the missingness-related parameters. For our carbon monoxide data set, fortunately, this was not the issue. When monitoring the convergence of the Gibbs samplers, we basically tried two main approaches. The first one was to generate two or more Markov chains starting from different initial values and wait until they interweaved with each other and the between-chain variance was much smaller than the within-chain variance. The second approach was to work with one chain, retaining only every k th sample after the burn-in period, with k set as a large enough number (e.g. 10) such that the retained samples are approximately independent. Usually, we stopped simulation procedure when the quantiles of all or some selected parameters were stable.

Acknowledgments

This work was supported by the National Institute of Drug Abuse through an SBIR contract N44 DA35513 and three research grants: R03 DA016721, R01 DA09992, and P50 DA18185. We especially thank Hamutahl Cohen for her editorial assistance and the reviewers for their constructive comments.

References

1. Nich C, Carroll KM. 'Intention-to-treat' meets 'missing data': implications of alternate strategies for analyzing clinical trials data. *Drug and Alcohol Dependence* 2002;68:121–130. [PubMed: 12234641]
2. Hedeker D, Gibbons RD. A random effects ordinal regression model for multilevel analysis. *Biometrics* 1994;50:933–944. [PubMed: 7787006]
3. Follmann D, Wu M. An approximate generalized linear model with random effects for informative missing data. *Biometrics* 1995;51:151–168. [PubMed: 7766771]
4. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73:13–22.
5. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982;38:963–974. [PubMed: 7168798]
6. Diggle, P.; Heagerty, P.; Liang, K-Y.; Zeger, S. *Analysis of Longitudinal Data*. 2. Oxford University Press; Oxford: 2002.
7. Little, RJA.; Rubin, DB. *Statistical Analysis with Missing Data*. 2. Wiley; New York: 2002.
8. Rubin DB. Inference and missing data. *Biometrika* 1976;63:581–582.
9. Robins JM, Rotnitzky AG, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* 1995;90:106–120.
10. Diggle PJ, Kenward MG. Informative drop-out in longitudinal data analysis. *Applied Statistics* 1994;43(1):49–93.
11. Yang X, Nie K, Belin T, Liu J, Shoptaw S. Markov transition models for binary repeated measures with ignorable and nonignorable missing values. *Statistical Methods in Medical Research* 2007;16(4):347–364. [PubMed: 17715161]
12. Verbeke, G.; Molenberghs, G. *Linear Mixed Models for Longitudinal Data*. Springer; New York: 2000.
13. Schafer, JL. *Analysis of Incomplete Multivariate Data*. Chapman & Hall; London: 1997.
14. Li J, Yang X, Wu Y, Shoptaw S. A random-effects Markov transition model for Poisson-distributed repeated measures with nonignorable missing values. *Statistics in Medicine* 2007;26(12):2519–2532. [PubMed: 17106918]
15. Molenberghs G, Michiels B, Lipsitz SR. Selection models and pattern-mixture models for incomplete categorical data with covariates. *Biometrics* 1999;55:978–983. [PubMed: 11315038]
16. Albert PS, Follmann DA. Modeling longitudinal count data subject to informative dropout. *Biometrics* 2000;56:667–677. [PubMed: 10985201]
17. Little RJA. Modeling the drop-out mechanism in longitudinal studies. *Journal of the American Statistical Association* 1995;90:1112–1121.
18. Troxel AB, Harrington DP, Lipsitz SR. Analysis of longitudinal data with non-ignorable non-monotone missing values. *Applied Statistics* 1998;47:425–438.
19. Albert PS, Follmann DA. A random effects transition model for longitudinal binary data with informative missingness. *Statistica Neerlandica* 2003;57:100–111.
20. Yang X, Shoptaw S. Assessing missing data assumptions in longitudinal studies: an example using a smoking cessation trial. *Drug and Alcohol Dependence* 2005;77:213–225. [PubMed: 15734221]
21. Shoptaw S, Rotheram-Fuller E, Yang X, Frosch D, Nahom D, Jarvik ME, Rawson RA, Ling W. Smoking cessation in methadone maintenance. *Addiction* 2002;97:1317–1328. [PubMed: 12359036]
22. Jennrich RI, Schluchter MD. Unbalanced repeated measures model with structural covariance matrices. *Biometrics* 1986;42:805–820. [PubMed: 3814725]
23. Hogan JW, Laird NM. Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine* 1997;16:239–258. [PubMed: 9004395]
24. Murray GD, Findlay JG. Correcting for the bias caused by drop-outs in hypertension trials. *Statistics in Medicine* 1988;7(9):941–946. [PubMed: 3175393]

25. Heckman JJ. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* 1976;5:475–492.
26. Fitzmaurice GM, Molenberghs G, Lipsitz SR. Regression models for longitudinal binary responses with informative dropouts. *Journal of the Royal Statistical Society: Series B* 1995;57:691–704.
27. Molenberghs G, Kenward MG, Lesaffre E. The analysis of longitudinal ordinal data with non-random dropout. *Biometrika* 1997;84:33–44.
28. Nordheim EV. Inference from nonrandomly missing categorical data: an example from a genetic study on Turner's syndrome. *Journal of the American Statistical Association* 1984;79:772–780.
29. Kenward MG, Molenberghs G. Parametric models for incomplete continuous and categorical longitudinal studies data. *Statistical Methods in Medical Research* 1999;8:51–83. [PubMed: 10347860]
30. Glynn, RJ.; Laird, NM.; Rubin, DB. Selection modeling versus mixture modeling with non-ignorable nonresponse. In: Wainer, H., editor. *Drawing Inferences from Self Selected Samples*. Springer; New York: 1986. p. 115-142.
31. Thijs H, Molenberghs G, Michiels B, Verbeke G, Curran D. Strategies to fit pattern-mixture models. *Biostatistics* 2002;3-2:245–265. [PubMed: 12933616]
32. Little RJA. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 1993;88:125–134.
33. Little RJA. A class of pattern-mixture models for normal incomplete data. *Biometrika* 1994;81:471–483.
34. Ekholm A, Skinner C. The Muscatine children's obesity data reanalysed using pattern mixture models. *Applied Statistics* 1998;47:251–263.
35. Hogan JW, Laird NM. Intent-to-treat analysis for incomplete repeated measures. *Biometrics* 1996;52:1002–1007. [PubMed: 8805765]
36. Molenberghs G, Michiels B, Kenward MG, Diggle PJ. Missing data mechanisms and pattern-mixture models. *Statistica Neerlandica* 1998;52:153–161.
37. Michiels B, Molenberghs G, Lipsitz SR. A pattern-mixture odds ratio model for incomplete categorical data. *Communication in Statistics: Theory and Methods* 1999;28(12):2863–2870.
38. Birmingham J, Fitzmaurice GM. A pattern-mixture model for longitudinal binary responses with nonignorable nonresponse. *Biometrics* 2002;58(4):989–996. [PubMed: 12495154]
39. Birmingham J, Rotnitzky A, Fitzmaurice GM. Patternmixture and selection models for analysing longitudinal data with monotone missing patterns. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2003;65(1):275–297.
40. 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.
41. Wu MC, Bailey KR. Estimation and comparison of changes in the pressure of informative right censoring: conditional linear model. *Biometrics* 1989;45:939–955. [PubMed: 2486189]
42. Wu MC, Follmann DA. Use of summary measures to adjust for informative missingness in repeated measures data with random effects. *Biometrics* 1999;55:75–84. [PubMed: 11318181]
43. Albert PS. A transitional model for longitudinal binary data subject to nonignorable missing data. *Biometrics* 2000;56:602–608. [PubMed: 10877323]
44. Pulksteinis EP, Ten Have TR, Landis R. Model for the analysis of binary longitudinal pain data subject to informative dropout through remedication. *Journal of the American Statistical Association* 1998;93:438–450.
45. Ten Have TR, Kunselman AR, Pulksteinis EP, Landis R. Mixed effects logistic regression models for longitudinal binary repeated response data with informative drop-out. *Biometrics* 1998;54:367–383. [PubMed: 9544529]
46. Rubin, DB. *Multiple Imputation for Nonresponse in Surveys*. Wiley; New York: 1987.
47. Rubin DB, Schenker N. Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association* 1986;81:366–374.
48. Shen, ZJ. PhD Dissertation. Department of Statistics, Harvard University; Cambridge, MA: 2000. Nested multiple imputation.

49. Harel, O. PhD Dissertation. Department of Statistics, Pennsylvania State University; University Park, PA: 2003. Strategies for data analysis with two types of missing values.
50. Rubin DB. Nested multiple imputation of NMES via partially incompatible MCMC. *Statistica Neerlandica* 2003;57:3–18.
51. Gilks, WR.; Richardson, S.; Spiegelhalter, DJ. *Markov Chain Monte Carlo in Practice*. Chapman & Hall; London: 1996.
52. Cowles MK, Carlin BP. Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association* 1996;91:883–904.
53. Yang X, Li J. A hybrid Gibbs sampler for selection models in dealing with outcome-dependent nonignorable dropouts. *UCLA Statistics Electronic Publications*. 2005 Preprint 451.
54. Gilks WR, Wild P. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* 1992;41:337–348.
55. SAS Institute Inc. *SAS/STAT* Software Changes and Enhancements, Release 8.2*. SAS Institute Inc; 2001.
56. Hedeker D, Gibbons RD. Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods* 1997;2:64–78.

Figure 1.

The average and SD curves for the log-scaled carbon monoxide levels. On this plot, the four mean curves of the log-scaled carbon monoxide levels and the corresponding pointwise standard errors are drawn for each of the four treatment conditions: Control, RP-only, CM-only, and RP+CM (RP = relapse prevention, CM = contingency management). Vertical bars indicate the estimated standard errors of average carbon monoxide levels. The stars (*) over the x -axis mark the time points (i.e. visit numbers), where the carbon monoxide levels are significantly different indicated by a pointwise ANOVA (p -value <0.001). Y -axis indicates values of carbon monoxide levels after $\log(1+x)$ transform. X -axis represents number of clinic visit for study participants (1, ..., 36; three times per week).

Figure 2.

Missingness patterns for the carbon monoxide levels across treatment conditions. For each treatment condition, an image depicts the missingness indicators of carbon monoxide levels for each smoker at each research visit. Dark colored area indicates that the corresponding carbon monoxide levels were observed while white colored area indicates that the corresponding data were missing intermittently or missing after dropout. The four treatment conditions are control, RP-only, CM-only, and RP+CM (RP = relapse prevention, CM = contingency management).

Figure 3.

Mean carbon monoxide levels for completers and early terminators. By dividing the 174 smokers into two groups: completers ($n_1 = 112$) and early terminators ($n_1 = 62$), the mean curves of carbon monoxide levels for subjects receiving CM (contingency management) and for subjects receiving no CM are depicted within each of the two groups (completers and early terminators).

Figure 4.

Plate 1. Pattern-dependent distribution of carbon monoxide levels. Using the software package named 'MPI 2.0', profiles and mean curves of carbon monoxide levels are drawn within each of the five groups determined by the dropout times: dropout at or before week 5, 7, 9, 11, and 12. In plots, green curves correspond to the mean carbon monoxide levels of subjects who received CM (contingency management), red curves indicate the mean curves of the subjects who did not receive CM, and gray-colored dash-lines depict the profiles of all the subjects within each group. The bottom-right plot depicts all the mean profiles corresponding to the five dropout patterns.

Table 1

Estimates treatment effects and parameters of the dropout model for the four partially imputed carbon monoxide data sets.

Partial imputations	1	2	3	4	Overall
β_1 (SD)	-0.29 (0.05)	-0.27 (0.05)	-0.28 (0.05)	-0.28 (0.05)	-0.28 (0.05)
β_2 (SD)	0.01 (0.05)	0.02 (0.05)	0.02 (0.05)	0.02 (0.05)	0.02 (0.05)
β_3 (SD)	-0.08 (0.06)	-0.10 (0.06)	-0.08 (0.07)	-0.08 (0.06)	-0.08 (0.07)
$\hat{\varphi}_1$ (SD)	1.27 (0.37)	1.37 (0.28)	1.24 (0.34)	1.25 (0.31)	1.28 (0.33)
$\hat{\varphi}_2$ (SD)	-0.02 (0.24)	-0.08 (0.24)	-0.00 (0.23)	-0.02 (0.23)	-0.03 (0.23)

Table II

Estimated treatment effect of contingency management (β_1 (SD)) using the pattern-mixture model with two patterns (complete *versus* dropout).

Imputations	Completers	Early terminators	Average
1	−0.35 (0.06)	−0.11 (0.09)	0.26 (0.13)
2	−0.34 (0.05)	−0.07 (0.10)	0.24 (0.13)
3	−0.34 (0.06)	−0.10 (0.09)	0.25 (0.13)

Table III

Estimated treatment effect of contingency management using the pattern-mixture models within the framework of two-stage MPI.

	Overall estimate (SD)	FMI (per cent)	<i>p</i> -Value
CCMV	−0.46 (0.22)	11	0.02
ACMV	−0.42 (0.19)	9	0.01
NCMV	−0.43 (0.28)	16	0.06

Table IV

Posterior parameter estimation with standard deviation and 95 per cent credible intervals using the REMTM to the continuous carbon monoxide data.

Parameter	Estimates	Std. dev.	95 per cent CI
Transition probability			
Intercept (β_0)	2.30	0.07	(2.15, 2.44)
RP (β_1)	-0.03	0.10	(-0.21, 0.16)
CM (β_2)	-0.25	0.10	(-0.46, -0.05)
RP*CM (β_3)	-0.05	0.14	(-0.33, 0.22)
Dependence parameter (α)	0.21	0.01	(0.18, 0.24)
Covariate-dependent missingness for			
Dropout (η)			
Intercept (η_0)	-4.61	0.29	(-5.18, -4.04)
RP (η_1)	0.07	0.41	(-0.73, 0.87)
CM (η_2)	0.03	0.42	(-0.80, 0.85)
RP*CM (η_3)	-0.03	0.59	(-1.19, 1.13)
Nonignorable missingness			
Dropout (γ)			
	1.30	0.34	(0.64, 1.96)
Variance of random effect (σ_{ξ}^2)	0.15	0.02	(0.12, 0.19)
Variance of random error (σ_{ε}^2)	0.12	0.00	(0.11, 0.12)