Title
Multiple Partial Imputation for Longitudinal Data with Missing Values in Clinical Trials

Authors
Xiaowei Yang
Jinhui Li
Steven Shoptaw

Publication Date
2011-10-25

Title:

Multiple Partial Imputation for Longitudinal Data with Missing Values in Clinical Trials

Short Running Title

Multiple Partial Imputation for Longitudinal Data with Missing Values

Authors

Xiaowei Yang[1], Jinhui Li[2], and Steven Shoptaw[3]

1. University of California, Davis, School of Medicine – Division of Biostatistics
   Med Sci 1-C, Suite 200, Davis, CA 95616
2. Department of Statistics, UCLA
   8130 Math Sciences Bldg, PO Box 95155, Los Angeles, CA 90095-1554
3. UCLA, Integrated Substances Abuse Programs
   11075 Santa Monica Blvd., Suite 200, Los Angeles, CA 90025

Contact Author

Xiaowei Yang, Ph.D.,
Biostatistics Unit, Department of Public Health Sciences
Med. Sci. 1-C, University of California, Davis, CA 95616
Tel: +01-530-754-9472, Fax +01-530-792-1425
E-mail: XDYang@UCDavis.edu

SUMMARY: Biomedical research is plagued with problems of missing data, especially in clinical trials of medical and behavioral therapies adopting longitudinal design. After a comprehensive literature review on modeling incomplete longitudinal data based on the full-likelihood functions, this paper proposes a set of imputation-based strategies for implementing three advanced models for handling intermittent missing values and dropouts that are potentially nonignorable according to various criteria. In multiple partial imputation (MPI), intermittent missing values are first imputed several times. Then, each partially imputed data set is analyzed using selection, pattern-mixture, or shared-parameter models to deal with dropouts. If imputations are additionally created for dropouts, it is a 2-stage version of MPI. Depending on models used for making imputations, various strategies can offer a framework for analyzing the sensitivity of parameter estimation to the assumptions of the missingness mechanism. For illustration, both continuous and dichotomized binary data from a smoking cessation clinical trial are analyzed. Both likelihood-based and Markov Chain Monte Carlo (MCMC) based inferences are also described.

KEY WORDS: Multiple Partial Imputation; Selection Model; Pattern-mixture Model; Markov Transition Model, Noingnorable Dropout; Intermittent Missing Values.

1. INTRODUCTION

In many fields of biomedical research, longitudinal studies are necessary because disorders frequently last many years. Only by measuring each subject repeatedly throughout a period of time can the efficacy of a treatment be investigated. Missing data are common in longitudinal

studies, often reflecting the chaotic behavior of participants, e.g., drug addiction [1,2] and mental health problems [3,4]. The proportion of missing data is sometimes noticeably large, e.g., as large as 70% at termination in a randomized study of buprenorphine versus methadone [5]. Although investigators may devote substantial efforts to minimize the number of missing values, some amount of missing data is inevitable in the practice of randomized medical clinical trials.

There are a number of different methods for analyzing longitudinal data containing missing values, with or without imputations. According to Little and Rubin [6], more than ten options of imputation methods can be used to analyze incomplete longitudinal data, e.g., last observation carried forward, mean imputation, hot-deck imputation, etc. [4]. Likelihood-based or semi-parametric methods without imputation are of most common and include complete-case analysis, generalized linear mixed models [7], marginal models using generalized estimating equations [8], and Markov transition models [9]. These methods have been implemented into mainstream software packages (e.g., SAS, S-plus/R, and SPSS), which treat incomplete longitudinal data as sequences of unequal length [10].

In all above methods, missing values are ignored from consideration. Analyses based on "*ignorabiltiy*" could result in biased estimates of treatment effects unless ignorable missingness mechanisms have been identified. Unfortunately, there is often evidence that missing values in longitudinal data are "*nonignorable*" [11-13].

In dealing with nonigborable missing values, advanced methods have been developed in the past decade by modeling the joint distribution of the indicators of missingness pattern and the values of observed and potentially-observed repeated measures. Accordingly, the likelihood function based on this joint distribution is called the full-likelihood function [14]. Literature reviews [12, 15] indicate that there are at least three ways to model this joint distribution: (1)

outcome-dependence, where missingness indicators are conditioned on the values of repeated measures; (2) pattern-dependence, where the distribution of repeated measure values is a mixture of distributions for subjects within distinct sub-groups determined by the patterns of missingness; and (3) parameter-dependence, where repeated measure values and missingness indicators are conditional independent given a group of parameters shared by the two parties. Correspondingly, there are three modeling strategies: selection models, pattern-mixture models, and shared-parameter models.

In practice, the limitation with any of these models is that they are sensitive to the assumptions made on repeated-measures models and missingness mechanisms [16-18]. As shown by Molenberghs et al. [10], different analysis methods can have distinct impacts on the conclusions of the same study. In this article, the idea of using sensitivity analysis is adopted where, given a practical data set, various models with different missing-data assumptions are applied to the same data.

By differentiating intermittent missing values from dropouts, the strategy of multiple imputation [19] provides a useful tool in conducting sensitivity analysis. A first approach is to make imputations only for the intermittent missing values, and then analyze the partially-imputed data sets using selection, pattern-mixture, or shared-parameter models to handle nonignroable dropouts. This approach was called multiple partial imputation (MPI) by Yang and Shoptaw [20]. The second approach is to conduct imputations for intermittent missing values and dropouts in a sequential order, a method called 2-stage MPI in this article. The intermittent missing values and dropouts are treated separately because there is usually empirical evidence suggesting different mechanisms for the two types of missingness. For example, there might be auxiliary information suggesting that intermittent missing values are ignorable while dropouts

are nonignorable. By specifying various assumptions in the development of imputation models, the sensitivity of the estimation of the parameters of interest is investigated. This imputation-based sensitivity offers a flexible framework within which the models for imputation and the ones for analysis may be different.

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]. The demonstration study tested the effectiveness of a relapse prevention program (RP) and a contingency management program (CM), alone and in combination, for improving smoking cessation outcomes using nicotine trans-dermal pharmacotherapy in methadone maintained cigarette smokers. A total of 174 participants who received nicotine replacement therapy during the study were randomly assigned into one of the four behavioral treatment groups: 42 subjects were assigned to a Control group that received no behavioral therapy; 42 subjects to RP-only; 43 subjects to CM-only; and 47 subjects to a combined RP+CM condition. Thirty-six measures of carbon monoxide levels, assessed from expired breath, were taken over the 12-week study period, thrice weekly.

Ignoring missing values, 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 analysis results with p-values smaller than 0.01. Nonetheless, the 36 p-values cannot be easily combined in making inferences on overall differences and the comparisons based on available values are potentially biased due to certain mechanisms of missingness. For example, if smokers in the three treatment groups dropped out with higher

probabilities given higher level of observed previous carbon monoxide, while smokers in the control group dropped out completely at random, then mean levels of carbon monoxide in the treatment groups would be lower than those in the control group at visit times close to the termination of the study, even though there is no actual treatment effects.

<INSERT FIGURE 1 HERE>

In Figure 2, the patterns of missingness in this study 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, while intermittent missingness does not associate with any specific simple patterns. At the termination of the study, up to 36% of the participants had withdrawn. An overall percentage of 4.3% of intermittent missing values are seen. The patterns and rates of missingness in this study are typical in substance abuse research.

<INSERT FIGURE 2 HERE>

3. INCOMPLETE LONGITUDIAL DATA ANALYSIS

For a longitudinal data set with balanced design, J repeated measures are potentially observed on each of the N subjects at times $t_{i1},...,t_{iJ}$ **(**$i = 1,...,N$ **;** $j = 1,...,J$ **)**. For the following discussion, we use capital symbols to represent variables, e.g., $Y_1,\ldots,$ and Y_J indicating response variables, and $X_1,\ldots,X_K$ indicating covariates or explanatory variables. Symbols in lower case represent observed or missing values: y_{ij} denoting the value of Y_j and x_{ijk} denotes the value of X_k recorded at time t_{ij} ($i = 1,...,N$; $j = 1,...,J$; $k = 1,...,K$). Bold symbols represent vectors or

matrices, e.g., the vector $\mathbf{y}_i = (y_{i1},...,y_{iJ})^T$ indicating values of repeated measures and the matrix $\mathbf{X}_i = [x_{ijk}]_{J\times K}$ consisting of values of time-varying or -independent 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 written 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. Determined by the parameterization of the covariance matrix, various forms of mixed models can be derived [22].

When some values of repeated measures are missing, we partition $\mathbf{y}_i$ into two parts, $\mathbf{y}_i = (\mathbf{y}_i^{obs}, \mathbf{y}_i^{mis})$, with $\mathbf{y}_i^{obs}$ indicating the observed values, and $\mathbf{y}_i^{mis}$ indicating values that would be observed if they were not missing. A vector of missingness indicators is defined as $\mathbf{r}_i = (r_{i1},...,r_{iJ})^T$ with elements $r_{ij} = 0$ (or 1) indicating whether y_{ij} is observed (or missing). Theoretically, the joint distribution of the observed data (i.e., $\mathbf{y}_i^{obs}$) and missingness patterns (i.e., $\mathbf{r}_i$) should be modeled in statistical analysis based on the full likelihood function,

$$L(\boldsymbol{\theta}, \boldsymbol{\varphi} \mid \mathbf{y}_i^{obs}, \mathbf{r}_i) \propto \prod_{i=1}^{N} \int f(\mathbf{y}_i, \mathbf{r}_i \mid \boldsymbol{\theta}, \boldsymbol{\varphi}) d\mathbf{y}_i^{mis} \tag{1}$$

where $\boldsymbol{\theta}$ represents parameters of the model for repeated measures, and $\boldsymbol{\varphi}$ represents parameters of the missingness mechanism. According to the possible causal path, there exist three ways to factor the joint distribution of the complete data and missingness indicators: outcome-dependent factorization, pattern-dependent factorization, and parameter-dependent factorization. Accordingly, there are three models available for incomplete longitudinal data analysis.

Selection Model factors the joint distribution $f(\mathbf{y}_i, \mathbf{r}_i \mid \mathbf{X}_i, \boldsymbol{\theta}, \boldsymbol{\varphi})$ into a marginal distribution for $\mathbf{y}_i$ and a conditional distribution of $\mathbf{r}_i$ given $\mathbf{y}_i$, i.e.,

$$f(\mathbf{y}_i, \mathbf{r}_i \mid \mathbf{X}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}) = f(\mathbf{y}_i \mid \mathbf{X}_i, \boldsymbol{\theta}) f(\mathbf{r}_i \mid \mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\varphi}), \tag{2}$$

where $f(\mathbf{r}_i \mid \mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\varphi})$ can be interpreted as the "self-selection of the i^{th} subject into a specific missingness group."

Pattern-Mixture Model, which is a pattern-dependent model, assumes that distribution of repeated measures varies with the missingness patterns and that joint distribution is factored as

$$f(\mathbf{y}_i, \mathbf{r}_i \mid \mathbf{X}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}) = f(\mathbf{y}_i \mid \mathbf{r}_i, \mathbf{X}_i, \boldsymbol{\theta}) f(\mathbf{r}_i \mid \mathbf{X}_i, \boldsymbol{\varphi}). \quad (3)$$

Assuming that there are P patterns of missingness in a data set, the marginal model of $\mathbf{y}_i$ is a mixture model,

$$f(\mathbf{y}_i) = \sum_{p=1}^{P} f(\mathbf{y}_i \mid r_i = p, \mathbf{X}_i, \boldsymbol{\theta}^{(p)}) \pi_p \quad (4)$$

where $\boldsymbol{\theta}^{(p)}$ represents the parameters of $f(\mathbf{y}_i)$ in the p^{th} pattern, $\pi_p = \Pr(r_i = p \mid \mathbf{X}_i, \boldsymbol{\varphi})$, and r_i numerates the P patterns. In pattern-mixture models, $\boldsymbol{\theta}^{(1)}$,…, and $\boldsymbol{\theta}^{(P)}$ can be different in dimensionality or in value.

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.,

$$f(\mathbf{y}_i, \mathbf{r}_i \mid \mathbf{X}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}) = \int f(\mathbf{y}_i \mid \boldsymbol{\xi}_i, \mathbf{X}_i, \boldsymbol{\theta}) f(\mathbf{r}_i \mid \boldsymbol{\xi}_i, \mathbf{X}_i, \boldsymbol{\varphi}) f(\boldsymbol{\xi}_i) d\boldsymbol{\xi}_i \quad (5)$$

From the point view of causation, shared "parameters" $\boldsymbol{\xi}_i$ play the role of a confounder for the relationship between $\mathbf{y}_i$ and $\mathbf{r}_i$; they can be either observable variables (e.g., gender) or latent variables (e.g., random-effects). For the case of observed confounders, model (5) is in fact a mixture model and the analysis can be conducted by a stratification analysis.

3.1. Ignorable versus Nonignorable

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 of the patients can be modeled jointly. In these scenarios, the above selection, pattern-mixture, and shared-parameter models can be applied directly or after some modification [23]. In a majority biomedical studies, however, only the parameters of repeated measures themselves are of interest, while parameters related to missing values are usually viewed as nuisance. In this latter case, it is desirable that missing data be ignored.

Within the setting of outcome-dependent missingness, the concept of "ignorability" was defined and extensively addressed. According to Rubn [24], missing values are ignorable when (i) r_i is independent of $\mathbf{y}_i^{mis}$, given $\mathbf{y}_i^{obs}$ and $\mathbf{X}_i$; (ii) $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ are distinct. Under this ignorability, the likelihood function for $\boldsymbol{\theta}$ can be separated from the log-likelihood function for $\boldsymbol{\varphi}$, i.e., $l(\boldsymbol{\theta},\boldsymbol{\varphi} \mid \mathbf{y}_i^{obs},\mathbf{r}_i) = l(\boldsymbol{\theta} \mid \mathbf{y}_i^{obs}) + l(\boldsymbol{\varphi} \mid \mathbf{y}_i^{obs},\mathbf{r}_i)$. In Little and Rubin [6], outcome-dependent missingness was further divided into sub-categories: missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR).

For intermittent missing values, ignorability can be interpreted as whether the values can be interpolated from neighborhood observed values. For dropouts, the assumption corresponds to whether missing values after dropout can be extrapolated from the previous observed values. In certain applications, occasional omission or nonresponses are due to reasons that are purely random in nature, e.g., schedule conflicts or bad weather, thus, can be assumed to be ignorable. Nonetheless, subjects withdraw from a study usually because of study-related reasons, e.g., being unsatisfactory with the intervention or notorious side effects of a medical therapy, hence are nonignorable [13, 25-26].

The definition of ignorability should be extended to meet the needs of pattern-mixture and shared-parameter models. Adopting an informal way, this article redefines "ignorability" as a condition under which observed data can be used to estimate $\boldsymbol{\theta}$ without bias. For selection and pattern-mixture models, so long as r_i and $\mathbf{y}_i^{mis}$ are independent of each other, given $\mathbf{y}_i^{obs}$ and $\mathbf{X}_i$, missing data can be ignored. For shared-parameter models, ignorability corresponds only to the case where $\boldsymbol{\xi}_i$ are observable confounders, which are usually viewed as a subset of $\mathbf{X}_i$. Unless r_i and $\mathbf{y}_i$ share no random-effects, a shared-parameter model would generally associate with a nonignorability assumption.

3.2 Selection and Pattern-Mixture Models for Modeling Dropouts

Because of the monotone format of dropout patterns, it is easier to handle dropouts than intermittent missing values. Here, we are especially interested in the parametric selection and pattern-mixture models for nonignorable dropouts.

3.2.1 A Selection Model for Dropouts

Let us denote t_{d_i} as the dropout time for the i^{th} subject, where $2 \le d_i \le J+1$ ($d_i = J+1$ indicates a subject who has completed the study). Then, missingness indicators $\mathbf{r}_i$ 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 [11] assumes: (i) $\Pr(r_{ij}=1 \mid j > d_i) = 1$; (ii) for $j \le d_i$, $\Pr(r_{ij}=1)$ depends on y_{ij} and its history $\mathbf{H}_{ij} = (y_{i1},...,y_{id_i-1})^T$; and (iii) the

conditional distribution of y_{ij} given $\mathbf{H}_{ij}$ is $f_{ij}(y \mid \mathbf{H}_{ij}, \boldsymbol{\theta})$. The full likelihood function for the i^{th} subject is expressed as

$$L_i(\boldsymbol{\theta}, \boldsymbol{\varphi} \mid \mathbf{y}_i^{obs}, \mathbf{r}_i) \propto \prod_{j=1}^{d_i-1} f(y_{ij} \mid \mathbf{H}_{ij}, \boldsymbol{\theta}) \prod_{j=1}^{d_i-1} \left[1 - p_j(y_{ij}, \mathbf{H}_{ij})\right] \Pr(r_{id_i} = 1 \mid \mathbf{H}_{id_i}), \tag{6}$$

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

$$\text{logit}(\Pr(r_{ij} = 1 \mid y_{ij}, \mathbf{H}_{ij}, \boldsymbol{\varphi})) = \phi_0 + \mathbf{H}_{ij}\boldsymbol{\varphi}_1 + y_{ij}\phi_2, \tag{7}$$

where $\phi_2 \neq 0$ implies that dropout process is outcome-dependent nonignorable.

The full log-likelihood function of the whole data set for ($\boldsymbol{\theta}$, $\boldsymbol{\varphi}$) can be partitioned into

$$l(\boldsymbol{\theta}, \boldsymbol{\varphi}) = l_1(\boldsymbol{\theta}) + l_2(\boldsymbol{\varphi}) + l_3(\boldsymbol{\varphi}, \boldsymbol{\theta}), \tag{8}$$

where $l_1(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log\{f(\mathbf{y}_i^{obs})\}$ corresponds to the observed-data log-likelihood function for $\boldsymbol{\theta}$, and $l_2(\boldsymbol{\varphi}) = \sum_{i=1}^{N} \sum_{j=1}^{d_i-1} \log\{1 - p_j(\mathbf{H}_{ij}, y_{ij})\}$ and $l_3(\boldsymbol{\varphi}, \boldsymbol{\theta}) = \sum_{i \le N; d_i \le J} \log\{\Pr(r_{id_i} = 1 \mid \mathbf{H}_{id_i})\}$ together determine the log-likelihood function of dropout process, which contains partial information on $\boldsymbol{\theta}$. If dropouts are ignorable, then $l_3(\boldsymbol{\theta}, \boldsymbol{\varphi})$ depends only on $\boldsymbol{\varphi}$ and can therefore be absorbed into $l_2(\boldsymbol{\varphi})$. Thus estimation of $\boldsymbol{\theta}$ can be solely derived from $l_1(\boldsymbol{\theta})$.

For a normal longitudinal data set, $\mathbf{y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta},\ \boldsymbol{\Sigma}(\boldsymbol{\alpha}))$ with parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$, the conditional distribution $f_{ij}(y \mid \mathbf{H}_{ij}, \boldsymbol{\theta})$ is a scalar normal distribution, and the marginal distribution $\prod_{j=1}^{d_i-1} f(y_{ij} \mid \mathbf{H}_{ij}, \boldsymbol{\theta}) = f(y_{i1},..., y_{id_i-1}) = f(\mathbf{y}_i^{obs})$ is a multivariate-normal distribution.

Selection models originated from the Tobit model of Heckman [27]. Verbeke and Molenburghs [13] addressed the theoretical translation from Tobit model to Diggle and Kenward's selection model. Subsequently, Troxel, Harrington, and Lipsitz [28] extended it to the non-monotone setting. Selection models for categorical and other type of measures were also developed; see Fitzmaurice, Molenberghs, and Lipsitz [29], Molenberghs Kenward, and Lesaffre [30], Nordheim [31], and Kenward and Molenburghs [32].

3.2.2 Pattern-Mixture Models for Dropouts

The high sensitivity of selection modeling has led to a growing interest in patter-mixture modeling [33-34]. After initial introduction [35-36], they are receiving more attention lately for continuous repeated measures (e.g., [15], [37-40]) and for categorical measures (e.g., [41-43]).

As seen in (3), a pattern-mixture model for dropouts factorizes the joint distribution $f(\mathbf{y}_i, d_i \mid \boldsymbol{\theta}, \boldsymbol{\varphi})$ into the product of the marginal distribution $f(d_i \mid \mathbf{X}_i, \boldsymbol{\varphi})$ and the conditional distribution $f(\mathbf{y}_i \mid \mathbf{X}_i, \boldsymbol{\theta}^{(d_i)})$, where $d_i = 2,..., J+1$ indicate the dropout time. The difficulty in pattern-mixture modeling of premature dropouts regards parameter identification. For any subject with $d_i < J+1$, the sub-vector of $\boldsymbol{\theta}^{(d_i)}$ describing $\mathbf{y}_i^{mis}$ is generally unidentified, unless certain restrictions are applied. Thijs et al. [34] proposed a framework for identifying restrictions. By suppressing the subscript "i", the complete data density for pattern j is given by

$$f_j(\mathbf{y}) = f_j(\mathbf{y}^{obs}) f_j(\mathbf{y}^{mis} \mid \mathbf{y}^{obs}), \tag{9}$$

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

$$f_j(y_s \mid y_1,...,y_{s-1}) = \sum_{t=s}^{J} \omega_{st} f_t(y_s \mid y_1,...,y_{s-1}), \quad s = j+1,...,J. \tag{10}$$

Using this restriction method, the full density in (10) can be expressed as

$$f_j(\mathbf{y}) = f_j(\mathbf{y}^{obs}) \prod_{s=0}^{J-j-1} \left[\sum_{t=J-s}^{J} \omega_{T-s,t} f_t(y_{T-s} \mid y_1,...,y_{T-s-1}) \right]. \tag{11}$$

Depending on the specification of ω_{st}'s, various schemes of identification can be implemented. For example, if all the weights are set to positive values, this will correspond to the identification using "available case missing values" (ACMV; [39]). It is of special interest to us because it is the natural counterpart of MAR in the outcome-dependence missingness framework.

The restriction called "complete-cases missing variable" (CCMV; [35]) identifies $f_j(y_s \mid y_1,...,y_{s-1})$ by borrowing information only from the completers, i.e.,

$$f_j(y_s \mid y_1,...,y_{s-1}) = f_J(y_s \mid y_1,...,y_{s-1}), \quad s = j+1,...,J, \tag{12}$$

which is a special case of (11) with $\omega_{sT} = 1$ and $\omega_{ss} = \omega_{s,s+1} = ... = \omega_{s,T-1} = 0$.

Another special case of identification is via "neighboring case missing values" (NCMV), which borrows information from neighbors with observed values on y_s, i.e.,

$$f_j(y_s \mid y_1,...,y_{s-1}) = f_s(y_s \mid y_1,...,y_{s-1}), \quad s = j+1,...,J, \tag{13}$$

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

The above identification approach based on (12) is convenient for sensitivity analysis. By varying the specification of weights (i.e., ω_{st}), it is equivalent to specify different assumptions on the dropout mechanism. When the number of dropout patterns is large, it would be tedious to apply the above identification strategy directly using a likelihood-based method to estimate the parameters for each pattern and then combine them. As mentioned earlier, the method of imputation would be a convenient tool.

3.3. A Shared-Parameter Model for Nonignorable Missing Values

When the dynamic features of the transition pattern in longitudinal data are of interest, an appropriate longitudinal approach is a Markov transition model. For binary repeated measures with nonigorable missing values, Albert and Follmann [44] developed a Markov transition model with random-effects that were shared by the sub-model on measurement and the sub-model on missingness indicators.

In the REMTM for incomplete binary repeated measures, the sub-model for measurement process assumes a first-order Markov chain for each series of binary measures. The transitional probabilities $P_{kl} = f(y_{ij} = l \mid y_{i,j-1} = k)$ ($k = 0$ or 1; $l = 0$ or 1) can be modeled by a logistic regression with random intercepts,

$$\begin{aligned} \text{logit}(P_{01}) &= \text{logit}(y_{ij} = 1 \mid y_{i,j-1} = 0, \mathbf{x}_{ij}, \xi_i) = \mathbf{x}_{ij}\boldsymbol{\beta}_{01} + \xi_i \\ \text{logit}(P_{10}) &= \text{logit}(y_{ij} = 0 \mid y_{i,j-1} = 1, \mathbf{x}_{ij}, \xi_i) = \mathbf{x}_{ij}\boldsymbol{\beta}_{10} + \nu\xi_i \end{aligned}, \tag{14}$$

where $\xi_i \overset{iid}{\sim} N(0, \sigma_\xi^2)$ denotes the random intercept, and ν is the heterogeneity parameter indicating the correlation between P_{10} and P_{01}.

The distribution of misisngness indicators $\mathbf{r}_i = (r_{i1}, ..., r_{iJ})^T$ can be modeled by another Markov transition model. Redefining $r_{ij} = \begin{cases} 0 & \text{if } y_{ij} \text{ is observed} \\ 1 & \text{if } y_{ij} \text{ is missing intermittently} \\ 2 & \text{if } y_{ij} \text{ is missing due to dropout} \end{cases}$, a first-order Markov process associates with 3×3 transition probabilities (i.e., $P_{kl} = \Pr(r_{ij} = l \mid r_{i,j-1} = k)$: $k = 0, 1, 2$; $l = 0, 1, 2$). Determined by certain restrictions, the following transition probabilities would be always equal to zero: $P_{12} = P_{20} = P_{21} = 0$. For other combinations of $r_{i,j-1}$ and r_{ij}, the transition probabilities are calculated in the following way. First, if the previous count measure is observed (i.e., $r_{i,j-1}$=0), then the current one could be observed, intermittent missing, or dropout; the 3-category multinomial-logit model [45] can be used to calculate the transition probabilities:

$$P(r_{ij} = k \mid \xi_i, \mathbf{x}_{ij}, r_{i,j-1} = 0) = \begin{cases} \dfrac{1}{1 + \sum_{l=1}^{2} \exp(\mathbf{x}_{ij}\boldsymbol{\eta}_l + \xi_i \gamma_l)} & \text{if k} = 0, \\ \dfrac{\exp(\mathbf{x}_{ij}\boldsymbol{\eta}_k + \xi_i \gamma_k)}{1 + \sum_{l=1}^{2} \exp(\mathbf{x}_{ij}\boldsymbol{\eta}_l + \xi_i \gamma_l)} & \text{if k} = 1 \text{or } 2. \end{cases} \tag{15}$$

Second, if the previous measure is intermittently missing, then the current one may only be observed or intermittently missing again. Correspondingly, a logistic regression model is used to calculate P_{10} and P_{11}, i.e.,

$$P(r_{ij} = k \mid \xi_i, \mathbf{x}_{ij}, r_{i,j-1} = 1) = \begin{cases} \dfrac{1}{1 + \exp(\mathbf{x}_{ij}\boldsymbol{\eta}_1 + \xi_i \gamma_1)} & \text{if k} = 0, \\ \dfrac{\exp(\mathbf{x}_{ij}\boldsymbol{\eta}_1 + \xi_i \gamma_1)}{1 + \exp(\mathbf{x}_{ij}\boldsymbol{\eta}_1 + \xi_i \gamma_1)} & \text{if k} = 1. \end{cases} \tag{16}$$

Third, for the absorbing state, we would always have $P(r_{ij} = 2 \mid \xi_i, \mathbf{x}_{ij}, r_{i,j-1} = 2) = 1$. Denoting T_i as the time for the last observed measurement for subject i, special considerations should also be given to y_{iT_i}, for which we always have $P(r_{iT_i} = 0 \mid r_{i,T_i-1} = 1) = 1$. In the above logit and logistic regression models, regression coefficients $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$, respectively, indicate whether intermittent missingness and dropout depend on covariates, while γ_1 and γ_2 respectively indicated whether the two types of missing values are nonignorable.

By combining the above sub-models for measurement and missingness, the likelihood function for parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}_{01}, \boldsymbol{\beta}_{10}, \nu, \sigma_\xi^2)$ and $\boldsymbol{\varphi} = (\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \gamma_1, \gamma_1)^T$ is expressed as

$$L(\boldsymbol{\theta}, \boldsymbol{\varphi}) \propto \int \prod_{i=1}^{N} \left\{ \prod_{j=1}^{T_i} p(y_{ij} \mid \mathbf{x}_{ij}, y_{i,j-1}, \xi_i, \boldsymbol{\theta}) \right\} \left\{ \prod_{j=1}^{J} p(r_{ij} \mid \mathbf{x}_{ij}, r_{i,j-1}, \xi_i, \boldsymbol{\varphi}) \right\} p(\xi_i) d\xi_i \,. \qquad (17)$$

It should be remarked here that the above REMTM can be easily extended to deal with other types of repeated measures. Li et al. [46] applied the REMTMs to Poisson-distributed repeated measures with nonignorable missing values. The random-intercept in (15)-(18) can be replaced with other types of random effects, including random slopes and random cohort effects. The REMTM is only one specific example of shared-parameter models, other longitudinal models such as marginal model or random-effects models can be also used to implement shared-parameters modeling. The shared-parameter model was first developed by Wu and Caroll [47] where certain parameters are shared by the measurement model and a censoring process. Other examples of shared-parameter models are seen in Little [15], Wu and Bailey [48], Wu and Follmann [49], Albert [9], Albert et al. [50], Follaman and Wu [5], Pulkstenis, et al. [51], and Ten Have et al. [52].

4. MULTIPLE PARTIAL IMPUTATION

For incomplete longitudinal data sets, the method of multiple imputation [19] is especially useful. Accurately predicting missing values is possible because repeated measures are often highly correlated to each other. When imputing, all above three modeling options can be used. In longitudinal data sets, missingness patterns and mechanisms for intermittent missing values and dropouts are apt to be distinct, thus requiring different treatment. Our empirical experiences suggest that in certain clinical trials intermittent missing values are ignorable due to factors that are non-related to the theme of the study, while dropouts should not be simply ignored. In Yang and Shoptaw [20], a partial version of multiple imputation, MPI, was first proposed, within which only intermittent missing values are imputed. As seen in the application of pattern-mixture model, imputation methods can be further employed to implement various schemes of identification of restriction for managing dropouts. This leads to a further extension of multiple imputation, which we term 2-stage MPI here. Depending on the assumptions of the mechanism of intermittent missingness and dropout, there exist many specific forms of MPI and 2-stage MPI. MCMC algorithms for creating imputations are described in Section 5.

4.1 MPI and 2-Stage MPI

We further partitioned $\mathbf{y}_i^{mis}$ into ($\mathbf{y}_i^{IM}$, $\mathbf{y}_i^{DM}$) to denote intermittent missing values and dropouts. For MPI, we draw $m > 1$ independent values $\mathbf{y}_i^{IM(1)}$, $\mathbf{y}_i^{IM(2)}$, …, $\mathbf{y}_i^{IM(m)}$ using the posterior predictive distribution $p(\mathbf{y}_i^{im} \mid \mathbf{y}_i^{obs}, \mathbf{r}_i)$. Within the 2-stage MPI, for each of the partial imputation for intermittent missing values, n conditionally independent values $\mathbf{y}_i^{DM(j,1)}$, $\mathbf{y}_i^{DM(j,2)}$, …,

$\mathbf{y}_i^{DM(j,n)}$ are additionally drawn from the predictive distribution $p(\mathbf{y}_i^{DM(j,1)} \mid \mathbf{y}_i^{obs}, \mathbf{y}_i^{IM(j)})$, $j = 1,...,m$. As mentioned earlier, the 2-stage MPI provides a natural framework for fitting pattern-mixture models by identifying restrictions with information borrowed from completers, neighboring cases, or available cases. If we use selection models or REMTMs, imputations for dropouts can be similarly conducted by applying appropriate MCMC algorithms.

4.2 Consolidating Results from Post-Imputation Analyses

A main concern for multiple imputation is how to combine the multiple point estimators to make an overall inferential statement. A set of rules for combination was originally developed by Rubin and Schenker [53], which can be used directly for MPI. In Shen [54], the idea was extended for the case of 2-step multiple imputation, which can be viewed as a general approach for a 2-stage MPI. More specifically, $m * n$ complete data sets are obtained eventually in the 2-stage MPI, $\mathbf{y}_i^{(j,k)} = (\mathbf{y}_i^{obs}, \mathbf{y}_i^{IM(j)}, \mathbf{y}_i^{DM(j,k)})$: $j = 1,...,m$, $k = 1,...,n$. A noticeable problem with these complete data sets is that they are not independent from each other, because each block or nest ($\mathbf{y}_i^{DM(j,1)}$, $\mathbf{y}_i^{DM(j,2)}$, …, $\mathbf{y}_i^{DM(j,n)}$) contains identical values for $\mathbf{y}_i^{IM(j)}$. By denoting $\hat{Q}^{(j,k)}$ and $\hat{U}^{(j,k)}$ as the point and variance estimates for Q from the $(j,k)^{th}$ completed data set, the overall point estimate for Q is still the simply grand average, i.e.,

$$\overline{Q} = \frac{1}{mn} \sum_{j=1}^{m} \sum_{k=1}^{n} \hat{Q}^{(j,k)} . \tag{18}$$

The associated variance for $\overline{Q}$ involves three components, i.e.,

$$T = \overline{U} + (1 - \frac{1}{n})W + (1 + \frac{1}{m})B \tag{19}$$

where $\overline{U} = \frac{1}{mn}\sum_{j=1}^{m}\sum_{k=1}^{n} U^{(j,k)}$ estimates the complete-data variance, $B = \frac{1}{m-1}\sum_{j=1}^{m}(\overline{Q}^{(j,\cdot)} - \overline{Q})$ indicates the between-nest variance, $W = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n-1}\sum_{k=1}^{n}(\hat{Q}^{(j,k)} - \hat{Q}^{(j,\cdot)})^2$ represents the within-nest variance, and $\overline{Q}^{(j,\cdot)} = \frac{1}{n}\sum_{k=1}^{n}\hat{Q}^{(j,k)}$. Inferences about Q are based on the Student's t-distribution $\frac{(Q-\overline{Q})}{\sqrt{T}} \sim t_\nu$ with d.f. $\nu = \frac{1}{m(n-1)}\left[\frac{(1-1/n)W}{T}\right]^2 + \frac{1}{m-1}\left[\frac{(1+1/m)W}{T}\right]^2$. Other formulas such as rates of missing information and relative efficiency are seen in Shen [54]. Examples of 2-stage multiple imputation in cross-sectional studies are seen in Harel [55] and Rubin [56].

5. MCMC Algorithms for Model Fitting and Imputation

For the longitudinal models based on full-likelihood functions, parameter estimation based on asymptotic normal theory is difficult to apply mainly because of the complicated form of likelihood functions. Without an analytical solution for the score function and Hessian matrix, optimization is challenging. Diggle and Kenward [11] resort to the simplex algorithm [57], which does not depend on derivatives. Unfortunately it converges unacceptably slowly and provides no Fisher information matrix. We implemented the nonlinear optimization algorithms of Dennis and Schnabel [58] with numerical derivatives, but found that global maximum remained difficult to obtain in practical settings. When calculating the dropout probabilities in the selection model or integrating over the random-effects in the REMTMs, time-consuming numerical integration is demanded, such as the method of Gauss-Hermit. Bayesian inferences

based on MCMC provide a more affordable and appropriate alternative. By sampling parameters and missisng values, the method of Monte Carlo using Gibbs sampler or Metropolis-Hasting algorithm offers a natural option for integration and optimization, without relying on fully determined density functions or expressive derivatives.

5.1 A Hybrid Gibbs Sampler for Fitting REMTM

For the REMTM, we denote $\boldsymbol{\psi} = (\boldsymbol{\theta}, \varphi, \xi_i, \sigma_\xi^2)$ and $\mathbf{y}_i = (\mathbf{y}_i^{obs}, \mathbf{y}_i^{mis})$ without differentiating intermittent missing values from dropouts. By setting "$t = 0$" and initializing the parameters and missing values with $\boldsymbol{\psi} = \boldsymbol{\psi}^{(0)}$ and $\mathbf{y}_i^{mis} = \mathbf{y}_i^{mis(0)}$, we repeat the following Gibbs steps.

1. Imputation Step*: Draw imputations for missing values. For $i = 1,...,n$ and $j = 1,...,J$, if y_{it} is missing, an imputation would be drawn using*
$$y_{ij}^{mis(t+1)} \sim f(y_{ij} \mid y_{i,j-1}^{(t)}, y_{i,j+1}^{(t)}, \boldsymbol{\psi}^{(t)})$$
where $y_{i,j-1}^{(t)}$ and $y_{i,j+1}^{(t)}$ are observed or imputed values at the previous iteration.
2. Estimation Step*: draw parameters and random effects in the following order,*
$$\boldsymbol{\theta}^{(t+1)} \propto f(\boldsymbol{\theta}) \prod_{i=1}^{n} f(\boldsymbol{\theta} \mid \mathbf{y}_i^{obs}, \mathbf{y}_i^{mis(t+1)}, \varphi^{(t)}, \xi_i^{(t)}, \sigma_\xi^{2(t)})$$
$$\varphi^{(t+1)} \propto f(\varphi) \prod_{i=1}^{n} f(\varphi \mid \mathbf{y}_i^{obs}, \mathbf{y}_i^{mis(t+1)}, \boldsymbol{\theta}^{(t+1)}, \xi_i^{(t)}, \sigma_\xi^{2(t)})$$
$$\xi_i^{(t+1)} \propto f(\xi_i) \prod_{i=1}^{n} f(\xi_i \mid \mathbf{y}_i^{obs}, \mathbf{y}_i^{mis(t+1)}, \boldsymbol{\theta}^{(t+1)}, \varphi^{(t+1)}, \sigma_\xi^{2(t)})$$
$$\sigma_\xi^{2(t+1)} \propto f(\sigma_\xi^2) \prod_{i=1}^{n} f(\sigma_\xi^2 \mid \mathbf{y}_i^{obs}, \mathbf{y}_i^{mis(t+1)}, \boldsymbol{\theta}^{(t+1)}, \varphi^{(t+1)}, \xi_i^{(t+1)})$$
where $f(\boldsymbol{\theta})$, $f(\varphi)$, $f(\xi_i)$, and $f(\sigma_\xi^2)$ are prior distributions of the parameters, and $\mathbf{y}_i^{mis(t+1)}$ represents the current imputed values.
3. *Set $t = t + 1$ and go to step 1.*

Running this Gibbs sampler with large enough iterations, the procedure would converge under regularity conditions and we obtain a series $\boldsymbol{\psi}^{(0)}, \mathbf{Y}_{mis}^{(0)}, ..., \boldsymbol{\psi}^{(T)}, \mathbf{Y}_{mis}^{(T)}$. By discarding the first T_0

burning samples (e.g., $T_0 = 10\% * T$), ($\boldsymbol{\psi}^{(T_0)},...,\boldsymbol{\psi}^{(T)}$) can be used to estimate the posterior distribution of $\boldsymbol{\psi}$ and further inferences can be made accordingly. This algorithm provides a structure for fitting any forms of REMTM depending on the types of repeated measure. For Poisson-distributed count data, see implementations in Li et al., [46].

Depending on the option chosen for modeling measurements and missingness mechanisms, different version of Gibbs samplers are conceivable. For example, the selection model for continuous repeated measures with nonignorable dropouts can be implemented with $\boldsymbol{\psi} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\varphi})$, a Gaussian distribution (i.e., $f(\mathbf{y}_i^{mis} \mid \mathbf{y}_i^{obs}, \boldsymbol{\psi})$) for drawing missing values, and logistic regression modeling the dropouts [59].

5.2 Sampling Conditional Distributions and Prior Specification

The above algorithm is called a hybrid Gibbs sampler because various sampling schemes can be embedded to simulate parameters from the fully conditional densities. If a conditional distribution has a known form, the corresponding parameter vector is sampled directly. For example, missing values of continuous repeated measures in a selection model can be sampled from a Gaussian distribution. If the conditional distribution has a log-concave form after proper transformation (e.g., the density for regression coefficients and residual variance), the efficient scheme called *adaptive rejection sampling* [60] can be applied. Otherwise, less efficient but more robust methods can be applied, e.g., Metropolis-Hasting or the giddy Gibbs sampler [61].

When there is no historical data or auxiliary information at hand, a convenient choice is the flat or non-informative prior. According to our experience, flat uniform distributions for covariance parameters or diffused normal distributions for regression coefficients usually work

well. In certain situation, conjugate priors should be adopted if possible, e.g., a Normal-Wishart distribution for the joint distribution of mean and covariance matrix of a multivariate normal repeated measures.

5.3 Creating Imputations in the MPI and the 2-stage MPI:

The above hybrid Gibbs sampler can be used directly to make model-based imputations. Running the Gibbs sampler, we obtain $(\boldsymbol{\psi}^{(0)}, \mathbf{Y}_{mis}^{(0)}, ..., \boldsymbol{\psi}^{(T)}, \mathbf{Y}_{mis}^{(T)})$, from which a subset of $(\mathbf{Y}_{mis}^{(T_0)}, ..., \mathbf{Y}_{mis}^{(T)})$ can be selected as multiple imputations (T_0 represents the burning period). Again, depending on the model used for analysis, various imputation schemes can be realized by the specific version of the hybrid Gibbs sampler. For imputation based on the pattern-mixture model with restriction identification, the procedure of Thijs et al. [34] can be used to draw imputations. For details of implementation, refer to the data augmentation of Schafer [14].

If there is empirical evidence suggesting an ignorable mechanism for intermittent missingess, the data augmentation can be applied directly. More specifically, the PROC MI of SAS for close-to-monotone missingness patterns can be applied. Otherwise, the model-based hybrid Gibbs sampler can be used to make partial imputations in MPI. A simple way is to keep only the imputed $\mathbf{y}_i^{IM}$ after obtaining m imputations for $\mathbf{y}_i^{mis} = (\mathbf{y}_i^{IM}, \mathbf{y}_i^{DM})$.

The 2-stage MPI is mainly used for conducting sensitivity analysis. By sequentially applying a Gibbs sampler for imputing $\mathbf{y}_i^{IM}$ and a potentially different Gibbs sampler for imputing $\mathbf{y}_i^{DM}$, we can obtain multiple complete data sets that incorporate various assumptions on the mechanism of missingness and dropout.

6. APPLICATION

6.1. Selection Model for Continuous Carbon Monoxide Levels

As seen from Figure 1, the mean carbon monoxide levels declined quickly within the first week from the same beginning level and then remain leveling off at different levels through the rest of the study period. Using a piecewise linear mixed-effects model with ignorability, Shoptaw et al. [21] reported a significantly effective treatment of CM. Here, we reanalyzed the data using only the carbon monoxide levels after the first week. For carbon monoxide levels, the mixed model with AR(1) covariance was used. After a step-wise selection procedure, the following mean structure was chosen to make inferences on treatment effects for contingency management and relapse prevention,

$$y_{ij} = \beta_0 + \beta_1 CM_i + \beta_2 RP_i + \beta_3 RP_i * CM_i + \beta_4 BaseCO_i + \beta_5 Patches_i \qquad (20)$$

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 dropouts, the logistic regression model was used,

$$\log it(p_{id_i}) = \phi_0 + \phi_1 Y_{i,d_i-1} + \phi_2 Y_{i,d_i}\,, \qquad (21)$$

where d_i indicates the dropout time of the i^{th} participant.

PROC MI with monotone option was run to generate four partially-imputed data sets. Then, the above selection model was applied to each of the four data sets. The estimates of interesting parameters are shown in Table 1, which shows that the between-imputation variances are very

small for all the parameters. In other words, the fraction of missing information due to intermittent missingness is low. After consolidating the four sets of estimates, it is 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_2 is significantly larger than zero ($\hat{\phi}_2 = 1.28$; $T_{2024} = 3.86$ with p=0.0002), suggesting that the higher the underlying missing value is, the larger probability of dropping out. In other words, the dropouts are outcome-dependent nonignorable.

<INSERT TABLE 1 HERE>

6.2. Pattern-Mixture Models for Continuous Carbon Monoxide Levels

Mainly to illustrate the application of 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, contingency management seems to be less effective for the early terminators. The mixed model with AR(1) covariance structure and predictors, CM_i, $BaseCO_i$, and $Patches_i$, was selected for analyzing the carbon monoxide levels starting from the second week,

$$y_{ij} = \beta_0 + \beta_1 CM_i + \beta_2 BaseCO_i + \beta_3 Patches_i . \quad (22)$$

This model was applied separately for the completers and early terminators. Let $\hat{\beta}_1^c$ and $\hat{\beta}_1^w$ denote the point estimators of β_1 respectively for the completers and early terminators, and $\hat{\pi}_c = 64\%$ the estimated probability of being completion, then the overall pointer estimator

across dropout patterns 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 [62].

Since the fractions of missing information due to intermittent missing were low, only 3 instead of 4 partial imputations were created. The pattern-averaged point estimators and standard errors for the treatment effect of β_1 are listed in Table 2. After consolidation using Rubin's rule, the overall mean is $\bar{\hat{\beta}}_1 = -0.25$ with standard deviation $\sqrt{\text{Var}(\bar{\hat{\beta}})} = 0.13$. The test based on a t-test suggests a p-value of 0.06.

<INSET FIGURE 3 HERE>

<INSET TABLE 2 HERE>

When the number of target dropout patterns becomes large, the application of pattern-mixture models without imputation becomes less useful. For example, the mean profiles of carbon monoxide levels across four dropout patterns are plotted in Figure 4, from which we observe notable variances across and within the four patterns. As the number of patterns increases, the number of pattern-specific subjects becomes smaller, and it becomes tedious even infeasible to conduct pattern-specific analysis and then combine the results across patterns. Hence, the approach of 2-stage MPI with restriction identification provides a reasonable alternative solution.

Adopting the procedure described in Thijs et al. [34], three restriction schemes (CCMV, NCMV, and ACMV) were used to impute the dropouts. Within this 2-stage MPI, the numbers of partial imputations were set as $m = 2$ for the first stage and $n = 3$ for the second stage. So together, six complete data sets were generated. For each, the AR(1) mixed model with predictors *CM* , *BaseCO* , and *Patches* was applied. Using the consolidation procedure as

described in Section 4.2, the overall point estimates and fractions of missing information for the treatment effect of CM are shown in Table 3 along with the p-vales of a one-sided hypothesis test using the t-statistics. It is seen that the overall fraction of missing information due to intermittent and dropout is much higher then that due to intermittent missingness alone. Two out of three identification strategies strong support the favorable treatment efficacy of CM.

<INSET FIGURE 4 HERE>

<INSET TABLE 3 HERE>

6.3. REMTMs for Dichotomized Carbon Monoxide Data

We reanalyzed the same group of carbon monoxide levels using the REMTM after dichotomizing them to indicate use or non-use of cigarettes. This dichotomized version of carbon monoxide data was analyzed by Yang et al. [12] using a REMTM with maximum likelihood estimation. Here it was reanalyzed using the hybrid Gibbs sampler for REMTM with predictors: CM_i, RP_i, and $RP*CM_i$ as defined in Section 6.1. Table 4 depicts the estimated posterior means, standard deviations, and 95% credible intervals (C.I.) for all the parameters of interest.

The estimated parameters for σ^2, γ_1, and γ_2 jointly suggest that both intermittent missingness and dropout are parameter-dependent nonignorable. The introduced random intercept effects (i.e., ξ_i) capture the heterogeneity on missingness across the subjects. Among all the estimated parameters of $\boldsymbol{\eta}_1$, only the one corresponding to CM is significantly different from zero (i.e., $\xi_1^3 = -1.19$), indicating that smokers receiving CM were less likely to miss their clinic visits occasionally than subjects who did not receive CM. The negative value of estimated ν suggests that individuals with large transition probabilities from non-use to use had lower

transition probabilities from use to non-use. In other words, individuals had an affinity for staying at either the "use" or "abstinence" state, and the shift of use to non-use or from non-use to use were less common (see Figure 5). Of most interest, the fitted REMTM confirmed a strongly favorable treatment efficacy of CM by increasing the probability of instilling smoking abstinence P_{10} (i.e., $\hat{\beta}_{10}^{3} = 2.61$ with 95% C.I.=(1.69, 3.53)) and the probability for maintaining abstinence P_{00} (i.e., $\hat{\beta}_{01}^{3} = -1.19$ with 95% C.I.=(-1.86, -0.52)).

<INSET FIGURE 5 HERE>

<INSET TABLE 4 HERE>

7. DISCUSSION

This paper reviews three modeling strategies for incomplete longitudinal data using full-likelihood functions and demonstrates their application within the MPI framework using a carbon monoxide data set. Selection, pattern-mixture, and shared-parameter models are generalized versions of standard longitudinal models (marginal models using GEE, mixed-effects models, and transition models). For example, the mixed-effects model ignoring missing values is in fact a selection model with ignorable missingness mechanisms. Though continuous and binary repeated measures are used in this illustration, the three modeling strategies can suggest a wide range of modeling solutions for various formats of repeated measures.

The most notable limitation with practical data analysis with missing values is that the true model and mechanism for measurements and missingness are usually unverifiable. Thus, in many settings, selection, pattern-mixture, or shared-parameter models should be viewed as models with rich assumptions. Our guideline is to always investigate the sensitivity of the

inferences on fixed parameters to varying assumptions. Inspired by the idea of pattern-mixture with restriction identification, the MPI-based frameworks were proposed. In the statistical literature, model-based sensitivity analyses are sometimes seen; e.g., selection model with local influence [13] and pattern-mixture model with varying restrictions [34]. However, extending use of a specific model further beyond assumptions might not be supported by the practical data. This is especially true in phase I, II, and III clinical trials, where sample sizes are usually not large enough to support over-fitting of the model.

For the same set of data from the smoking trial, we applied various models to analyze the treatment efficacy of two behavioral therapies: contingency management and relapse prevention. Selection models and pattern-mixture models were jointly applied to the original continuous carbon monoxide levels. After dichotomizing the data, the Markov transition model with random intercepts was applied. Overall results depict a consistent image in supporting the favorable efficacy of the contingency management.

Our previous endeavors in methodology development for incomplete longitudinal data analysis mainly focused on software development and MCMC-based Bayesian computations. Simulation studies and practical applications for full-likelihood models have proved acceptable performance; see Yang and Shoptaw [20], Yang et al [12], Li et al. [46], and Yang and Li [59]. Currently, we are conducting simulation studies to evaluate and compare the selection, pattern-mixture, and shared-parameter models. We developed a software package to implement all the three modeling functions within the framework of one or two stage MPI; see www.Bayessoft.com/MPI. This package also provides tools for visual data exploration and formal assessment on missing-data assumptions; see an example from Yang and Shoptaw [20].

Currently, we are implementing more functions to this package so that most types of repeated measures with distribution from the exponential family can be modeled.

When describing the hybrid Gibbs sampler for model fitting and imputation, only the general structure and ideas were presented. For details related to the theoretical basis and technical implementation, refer to Yang and Li [59] and Li et al. [46]. Results of simulation studies for selection models and random-intercept Markov transition models are also presented in these two articles.

ACKNOWLEDGEMENTS

This work was partially 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.

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.

2. Collins LM, Seitz L. New methodology for substance abuse prevention research. *Advances in data analysis for prevention research.* National Institute on Drug Abuse Research Monograph, 1994.

3. Hedeker D, Gibbons RD. A random effects ordinal regression model for multilevel analysis. *Biometrics* 1994; **50**: 933-944.

4. Tang L, Song J, Belin TR, Unutzer, J. A comparison of imputation methods in a longitudinal randomized clinical trial. *Statistics in Medicine* Online: 12 May 2005, DOI: 10.1002/sim.2099.

5. Follmann D. Wu M. An approximate generalized linear model with random effects for informative missing data, *Biometrics* 1995; **51**:151-168.

6. Little RJA, Rubin DB. Statistical Analysis with Missing Data, 2nd edition, New York: John Wiley， 2002.

7. McCulloch CE， Searle SR. *Generalized, Linear, and Mixed Models*. John Wiley and Son, 2001.

8. Hardin JW. Hilbe JM. *Generalized Estimating Equations*. Chapman & Hall/CRC, 2002.

9. Albert PS. A transitional model for longitudinal binary data subject to nonignorable missing data, *Biometrics* 2000; **56**: 602-608.

10. Molenberghs G, Thijs H, Jansen I, Beunckens C, Kenward MG, Mallinckrodt C, Carroll RJ. Analyzing incomplete longitudinal clinical trial data. *Biostatistics* 2004; **5(3)**: 445-464.
11. Diggle PJ, Kenward MG. Informative drop-out in longitudinal data analysis. *Applied Statistics* 1994; **43**(1), 49-93.

12. Yang X, Nie K, Belin T, Liu J, Shoptaw S. Markov Transition Models for Binary Repeated Measures with Ignorable and Nonignorable Missing Values. *Resubmitted to Statistical Methods in Medical Research* 2005.

13. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data.* New York: Springer-Verlag, 2000.

14. Schafer JL. *Analysis of Incomplete Multivariate Data*, London: Chapman & Hall, 1997.

15. Little RJA. Modeling the Drop-Out Mechanism in Longitudinal Studies. *Journal of the American Statistical Association* 1995; **90**: 1112-1121.

16. Vach W, Blettber M. Logistic regression with incompletely observed categorical covariates--investigating the sensitivity against violation of the missing at random assumption. *Statistics in Medicine* 1995; **14(12)**:1315-29.

17. Verbeke G, Lesaffre E, Spiessens B. The practical use of different strategies to handle dropout in longitudinal studies. *Drug Information Journal* 2001; **35**: 419-439.

18. Molenberghs G, Kenward MG, Goetghebeur E. (2001). Sensitivity analysis for incomplete contingency tables: the Slovenian plebiscite case. *Applied Statistics* 2001; **50**: 15-29.

19. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons 1987.

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

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.

22. Jennrich RI, Schluchter, MD. Unbalanced repeated measures model with structural covariance matrices. *Biometrics* 1986; **42**: 805-820.

23. Hogan JW, Laird NM. Mixture models for the joint distribution of repeated measures and event times. *Statistics in medicine*, 1997; **16**: 239-258.

24. Rubin DB. Inference and missing data. *Biometrika* 1976; **63**, 581-582.

25. Murray GD, Findlay JG. Correcting for the bias caused by drop-outs in hypertension trials. *Statistics in Medicine* 1988; **7(9)**: 941-946.

26. Diggle P, Heagerty P, Liang, K-Y, Zeger S. Analysis of Longitudinal Data. Oxford University Press, 2nd Edition, 2002.

27. Heckman JJ. The common structure of statistical models of trucation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of economic and Social Measurement*, 1976; **5**: 475-492.

28. Troxel AB, Harrington DP, Lipsitz SR. Analysis of longitudinal data with non-ignorable non-monotone missing values. *Applied Statistics* 1998; **47**:425-438.

29. 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.

30. Molenberghs G, Kenward MG, Lesaffre E. The analysis of longitudinal ordinal data with non-random dropout. *Biometrika* 1997; **84**: 33-44.

31. 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.

32. Kenward MG, Molenberghs G. Parametric models for incomplete continuous and categorical longitudinal studies data. *Statistical Methods in Medical Research* 1999; **8**, 51-83.

33. Glynn RJ, Laird NM, Rubin DB. Selection modeling versus mixture modeling with non-ignorable nonresponse. In: Drawing Inferences from Self Selected Samples, H. Wainer (Ed.). New York: Springer-Verlag, 1986; pp. 115-142.

34. Thijs H, Molenberghs G, Michiels B, Verbeke G, and Curran D. Strategies to fit pattern-mixture models. *Biostatistics* 2002; **3-2**: 245-265.

35. Little RJA. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 1993; **88**: 125-134.

36. Little RJA. A class of pattern-mixture models for normal incomplete data. *Biometrika* 1994; **81**: 471-483.

37. Ekholm A, Skinner, C. The Muscatine children's obesity data reanalysed using pattern mixture models. *Applied Statistics*, 1998; **47**: 251-263.

38. Hogan JW, Laird NM. Intent-to-treat analysis for incomplete repeated measures. *Biometrics* 1996; **52**:1002-7.

39. Molenberghs G, Michiels B, Kenward MG, Diggle PJ. Missing data mechanisms and pattern-mixture models. *Statistica Neerlandica* 1998; **52**: 153-161.

40. Michiels B, Molenberghs G, Lispsitz SR. A pattern-mixture odds ratio model for incomplete categorical data. *Communication in Statistics: Theory and Methods* 1999; **28**, 2893-2870.

41. Molenberghs G, Michiels B, Lipsitz SR. Selection models and pattern-mixture models for incomplete categorical data with covariates. *Biometrics***,** 1999; **55**: 978-983.

42. Birmingham J, Fitzmaurice GM. A Pattern-Mixture Model for Longitudinal Binary Responses with Nonignorable Nonresponse. *Biometrics* 2002; **58**:4, 989-996

43. 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

44. Albert PS, Follman DA. A Random E_ects Transition Model for Longitudinal Binary Data with Informative Missingness. *Statistica Neerlandica* 2003; **57**: 100–111.

45. Agresti A. *Categorical Data Analysis*. John Wiley and Sons, 2002.

46. Li J, Yang X, Wu Y, Shoptaw S. A Random-e_ects Markov Transition Model for Poisson-Distributed Repeated Measures with Nonignorable Missing Values. *Submitted to Statistics in Medicine* 2005.

47. 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.

48. Wu MC, Bailey KR. Estimation and Comparison of changes in the pressure of informative right censoring: conditional linear model. *Biometrics* 1989; **45**, 939-955.

49. 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.

50. Albert PS, Follmann DA, Wang SA, Suh EB. A latent autoregressive model for longitudinal binary data subject to informative missingness, *Biometrics* 2002; **58**: 631 ‐ 642.

51. Pulksteinis EP, The 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.

52. Ten Have TR, Kunselman AR, Pulksteinis EP, Landis R. Mixed Effects Logistic Regression Models for Longitudinal Binary Repeated Reponse Data with Informative Drop-out. *Biometrics* 1998; **54**: 367-383.

53. 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.

54. Shen ZJ. *Nested multiple imputation*. Ph.D dissertation, Department of Statistics, Harvard University, Cambridge, MA, 2000.

55. Harel O. *Strategies for data analysis with two types of missing values*. Ph.D. dissertation, Department of Statistics, Pensylvania State University, University Park, PA, 2003.

56. Rubin DB. Nested multiple imputation of NMES via partially incompatible MCMC. *Statistica Neerlandica*, 2003; **57**: 3-18.

57. Nelder JA, Mead R. A Simplex Method for Function Minimization. *The Computer Journal* 1965; **7**: 308-313.

58. Dennis JE, Schnabel RB. Numerical methods for unconstrained optimization and nonlinear equations. Philadelphia: Society for Industrial and Applied Mathematics 1996.

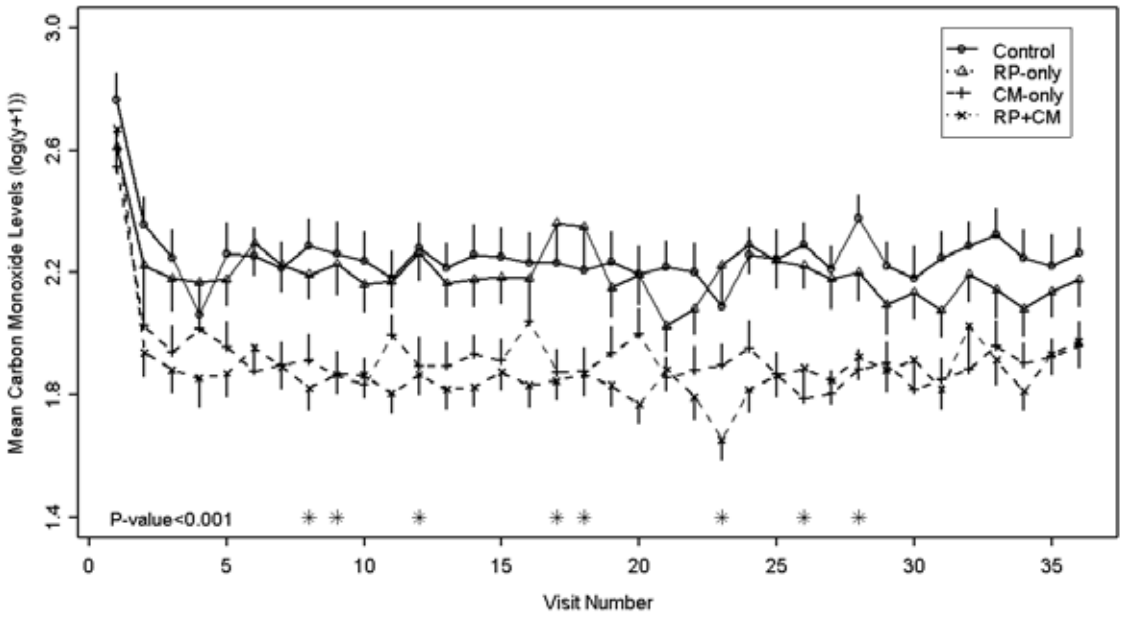
59. Yang X, Li J. A Hybrid Gibbs Sampler for Selection Models in Dealing with Outcome-dependent Nonignorable Dropouts. *Submitted to Statistics in Medicine* 2005.

60. Gilks WR, Wild P. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* 1992; **41**: 337-348.

61. Ritter C, Tanner MA. Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler. *Journal of the American Statistical Society* 1992; **87**: 861–868.

62. 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. Average and SD Curves for the Log-scaled Carbon Monoxide Levels



Notes: (1) The vertical bars indicate the estimated standard errors for the average carbon monoxide levels.

(2) The stares over the x-axis indicates the time points where the p-value of the point-wise ANOVA is smaller than 0.001.

Figure 2. Missingness Patterns for the Smoking Cessation Study

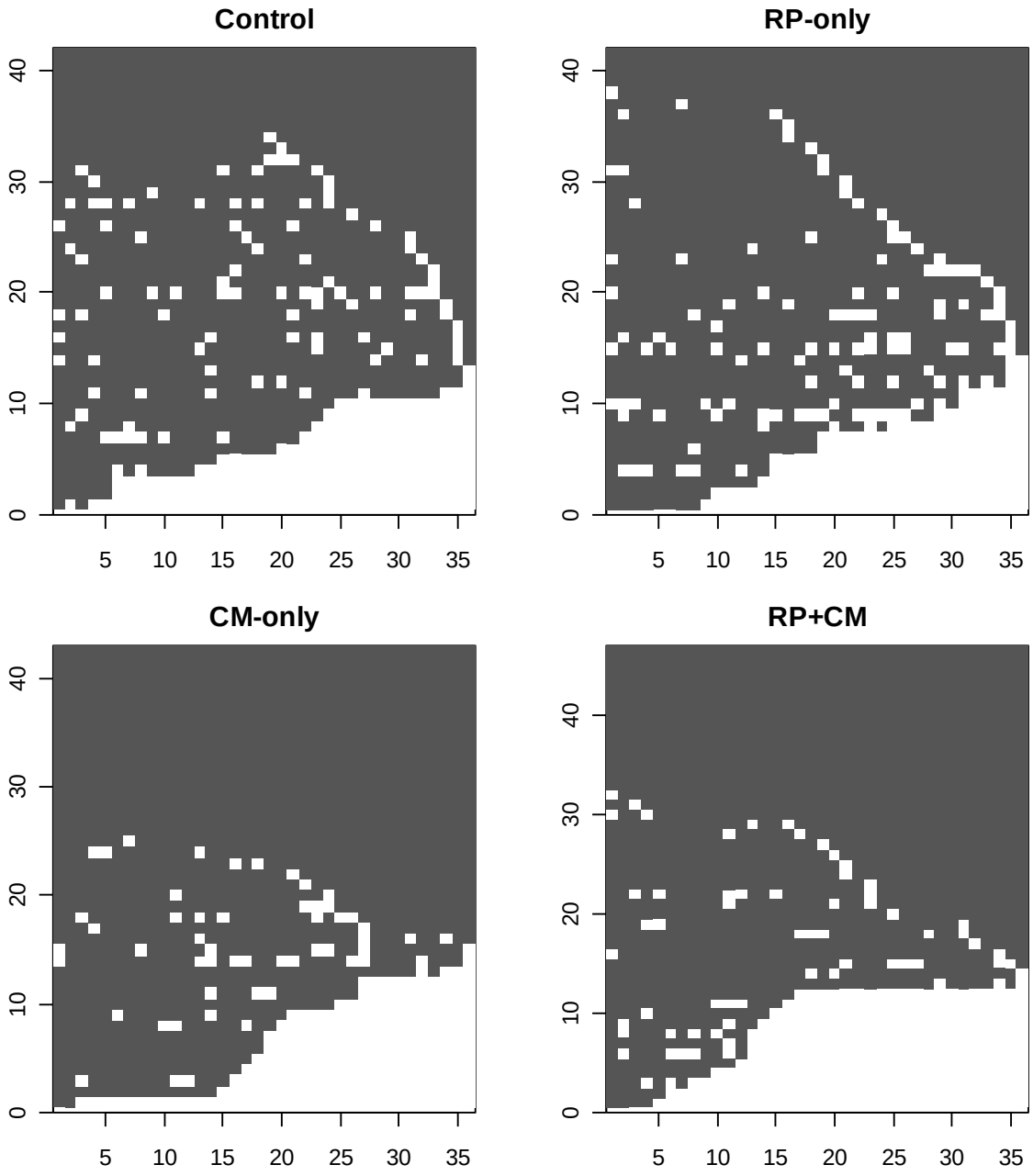


Figure 3. Mean Carbon Monoxide Levels for Completers and Early terminators.

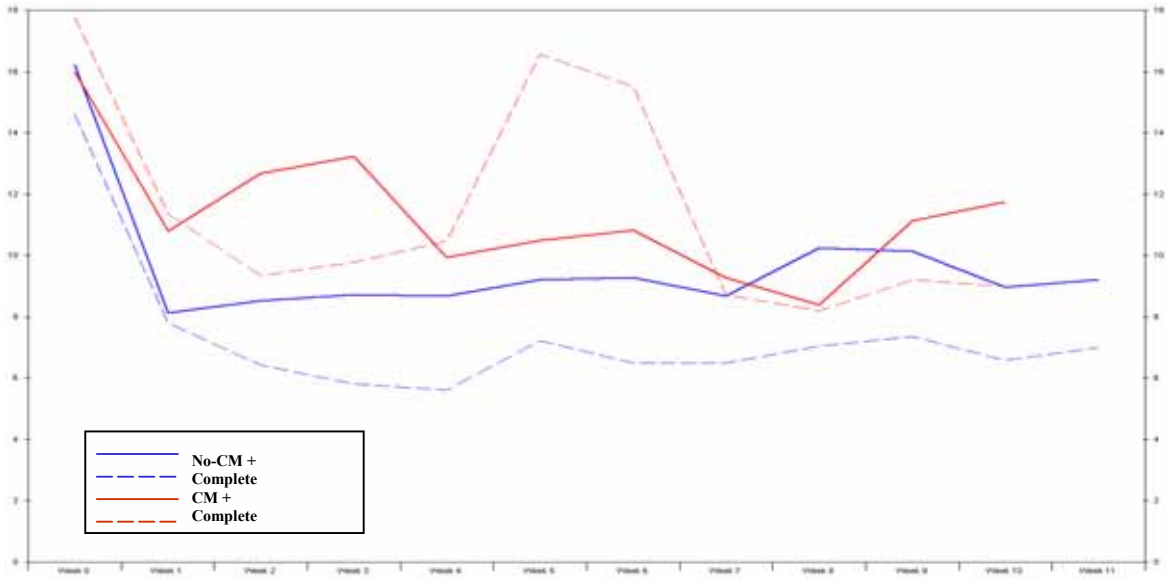


Figure 4. Pattern-Dependent Distribution Carbon Monoxide Levels

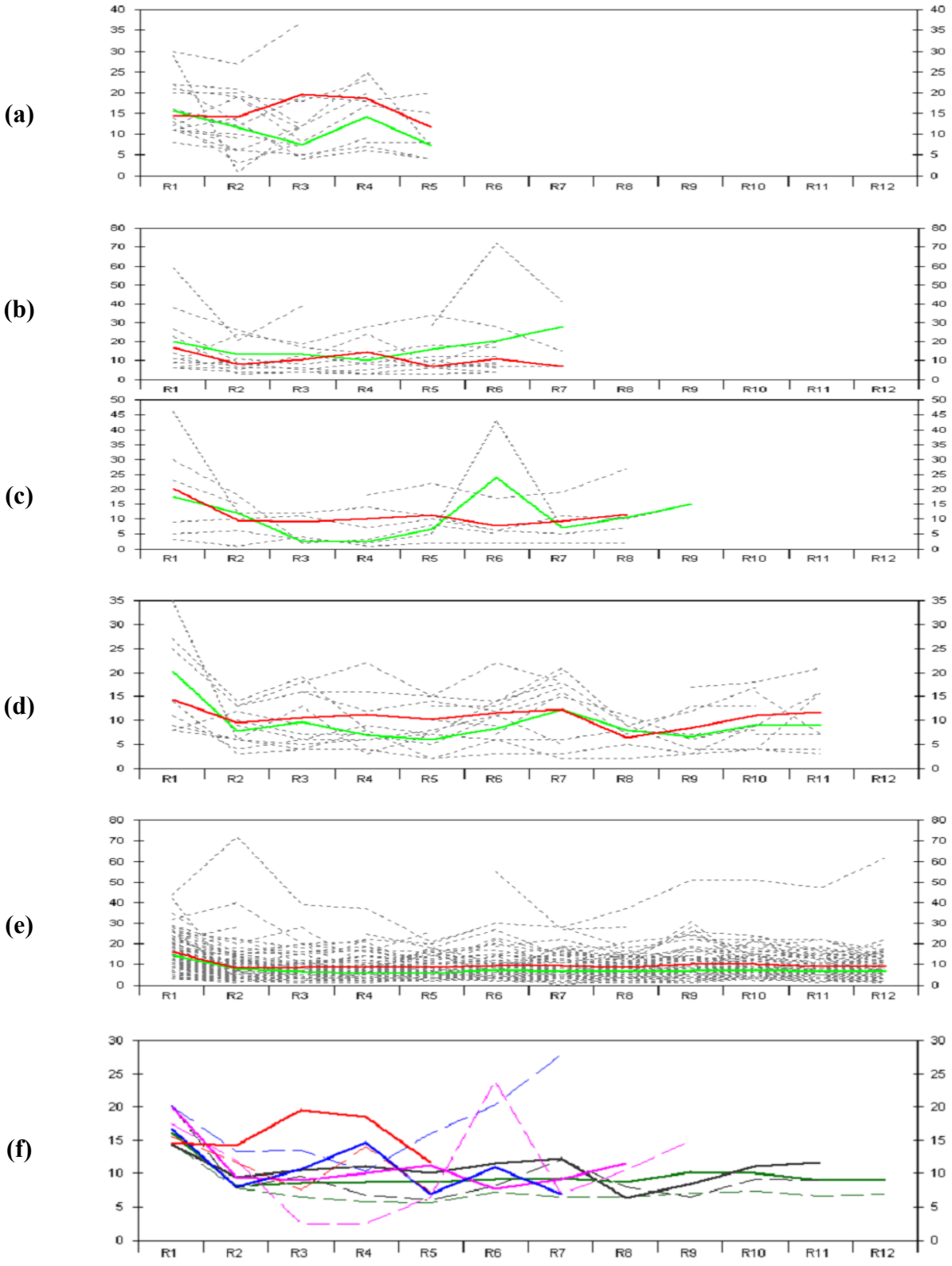


Figure 5. Transition Probabilities for the Dichotomized Carbon Monoxide Levels

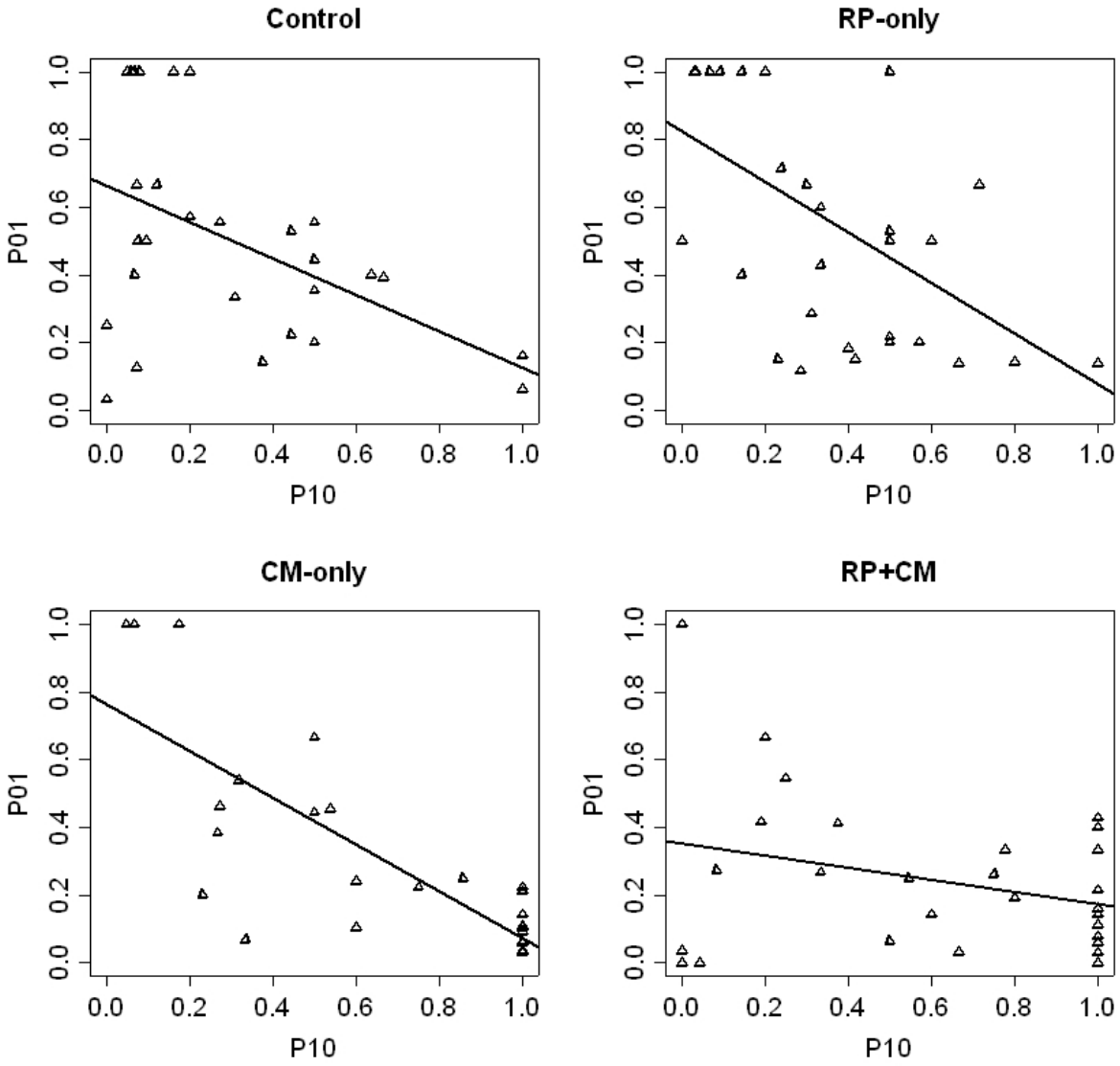


Figure 1. The average and SD curves for the log-scaled carbon monoxide levels. On this plot, the four mean curves of the log-scaled carbon monoxide levels and the corresponding point-wise standard errors are drawn for each of the four treatment conditions: Control, RP-only, CM-only, and RP+CM (RP=Relapse 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 point-wise 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).

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 diving 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. 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 before week 5 (a), 7 (b), 9 (c), 11 (d), and 12 (e). In plots (a) to (e), 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 depicts the profiles of all the subjects within each group. The plot (f) depicts all the mean profiles corresponding to the five dropout patterns.

Figure 5. Transition Probabilities for the Dichotomized Carbon Monoxide Levels. Within each the four treatment groups, the scatter plot displays the transition probabilities calculated from the observed repeated measures for each subject. Because $P_{00} + P_{01} = P_{10} + P_{11} = 1$, only P_{01} and P_{10} are displayed.

Partial Imputations	**1**	**2**	**3**	**4**	**Overall**
$\hat{\beta}_1$ (S.D.)	-0.29 (0.05)	-0.27 (0.05)	-0.28 (0.05)	-0.28 (0.05)	-0.28 (0.05)
$\hat{\beta}_2$ (S.D.)	0.01 (0.05)	0.02 (0.05)	0.02 (0.05)	0.02 (0.05)	0.02 (0.05)
$\hat{\beta}_1$ (S.D.)	-0.08 (0.06)	-0.10 (0.06)	-0.08 (0.07)	-0.08 (0.06)	-0.08 (0.07)
$\hat{\phi}_1$ (S.D.)	-0.02 (0.24)	-0.08 (0.24)	-0.00 (0.23)	-0.02 (0.23)	-0.03 (0.23)
$\hat{\phi}_2$ (S.D.)	1.27 (0.37)	1.37 (0.28)	1.24 (0.34)	1.25 (0.31)	1.28 (0.33)

Table 1. Estimates Treatment Effects and Parameters of the Dropout Model for the Four Partially Imputed Carbon Monoxide Data Sets.

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 2. Estimated Treatment Effect of Contingency Management ($\hat{\beta}_1$ (S.D.)) using the Pattern-Mixture Model with two Patterns (Complete vs. Dropout)

	Overall Estimate (S.D)	FMI	P-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 3. Estimated Treatment Effect of Contingency Management using the Pattern-Mixture Models within the Framework of 2-stage MPI

Parameter	**Estimates**	**Std. De.**	**95% C.I.**
Transition Probability (P_{01})			
--Intercept (β_{01}^{1})	-0.39	0.16	(-0.70, -0.07)
--RP (β_{01}^{2})	0.14	0.50	(-0.84, 1.12)
--CM (β_{01}^{3})	-1.19	0.34	(-1.86, -0.52)
--RP*CM (β_{01}^{4})	0.16	0.61	(-1.04, 1.36)
Transition Probability (P_{10})			
--Intercept (β_{10}^{1})	-1.69	0.37	(-2.42, -0.96)
--Baclofen (β_{10}^{2})	0.90	0.66	(0.39, 2.19)
--Baseline (β_{10}^{3})	2.61	0.47	(1.69, 3.53)
--RP*CM (β_{10}^{4})	-1.21	0.81	(-2.80, 0.38)
Variance of Random Intercept (σ^{2})	3.32	0.64	(2.07, 4.57)
Heterogeneity Parameter (ν)	-1.43	0.14	(-1.70, -1.15)
Nonignorable Missingness			
--Intermittent Missingness (γ_1)	0.93	0.11	(0.71, 1.14)
--Dropout (γ_2)	0.56	0.14	(0.29, 0.83)
Covariate-Dependent Missingness for Intermittent Missing (η_1)			
--Intercept (η_1^{1})	-2.28	0.25	(-2.77, -1.79)
--RP (η_1^{2})	-0.48	0.47	(-1.40, 0.44)
--CM (η_1^{3})	-1.19	0.34	(-1.85, -0.52)
--RP*CM (η_1^{4})	1.14	0.57	(0.02, 2.26)
Dropout (η_2)			
--Intercept (η_2^{1})	-4.76	0.37	(-5.49, -4.03)
--RP (η_2^{2})	0.14	0.52	(-0.88, 1.16)
--CM (η_2^{3})	-0.21	0.50	(-1.19, 0.77)
--RP*CM (η_2^{4})	-0.01	0.70	(-1.38, 1.36)

Table 4. Posterior Parameter Estimation with Standard Deviation and 95% Credible Intervals Using the REMTM to the Dinchotomized Carbon Monoxide Data