Title

Selection Models with Augmented Gibbs Samplers for Continuous Repeated Measures with Nonignorable Dropout

Authors

Yang, Xiaowei
Li, Jinhui

Publication Date

2005

Title:

Selection Models with Augmented Gibbs Samplers for Continuous Repeated Measures with Nonignorable Dropout

Short Running Title

Augmented Gibbs Sampler for Nonignorable Dropout

Authors:

Xiaowei Yang[1,2] **and Li, Jinhui**[1,3]

1. BayesSoft, Inc.,

   1153 San Gallo Terrace, Davis, CA 95616.

2. Division of Biostatistics, School of Medicine

   University of California-Davis, MS1C, Davis, CA 95616 .

3. UCLA-Department of Statistics

   8130 Math Science Bldg, PO Box 95155, Los Angeles, CA 90025.

Contact Author:

Xiaowei Yang, Ph.D., BayesSoft, Inc.,

1153 San Gallo Terrace, Davis, CA 95616, U.S.A.

Tel 530-754-9472, Fax 530-792-1425

E-mail: XYang@BayesSoft.com

SUMMARY

Premature withdrawal is a notable problem in biomedical research with longitudinal design. It would complicate statistical analysis with biased or invalid inferences if missing values due to dropout are simply ignored. As one solution to potentially nonignorable droput, the selection model assumes a mechanism of outcome-dependent dropout and jointly models the process of repeated measures and the mechanism of dropout. Previous applications of selection models mainly resort to likelihood-based inferences using optimization methods such as simplex or EM-algorithm. This paper implements the modeling strategy using mixed-effects repeated-measures models with Bayesian inferences. Specifically, the selection model with random-effects and the one with autoregressive covariance structure are introduced. Markov Chain Monte Carlo (MCMC) algorithms based on augmented Gibbs samplers are developed in fitting the models. For demonstration, both simulated and practical data sets are analyzed.

Key words: Augmented Gibbs Sampler, Selection Model, Nonignorable Dropout, Mixed Models, Repeated Measures

1 INTRODUCTION

In biomedical research with longitudinal design, each subject is measured repeatedly throughout a period of time. In many longitudinal studies, dropouts (i.e., missing values due to earlier withdrawal) are common. In certain research areas, such as drug addiction [1] and cancer [2], the proportion of subjects who drop out prematurely is extremely high, e.g., as large as 70% in a randomized buprenorphine versus methadone study [3]. Reasons for dropout are

usually study-related, e.g., negative side-effects of the testing medicines, ineffectiveness of the intervention, and inappropriate conduction of the therapy [4]. Without careful handling of dropouts, either biased parameter estimates or invalid inferences would end up.

By ignoring dropouts, most available software packages treat incomplete repeated measures as sequences of unequal length, and make inferences based on the observed-data likelihood functions [5]. For example, mixed models [6] and marginal models using generalized estimating equations [7] have been implemented into SAS, R/Splus, and SPSS. Nonetheless, the assumption of ignorability is rarely the case within practical settings [8]. Depending on the nature of dropout mechanism, various models for nonignorable models have been proposed in the past decade (e.g., [9]-[14]).

One strategy is to model the joint distribution of the indicators of dropout and the complete values and then integrate out the missing values. In contrast to the observed-data likelihood function, likelihood based on this joint distribution is called full-likelihood functions [15]. According to [12], [16], and [17], at least three ways can be used to model the joint distribution. The first way conceives the assumption of outcome-dependent dropout, where distribution of dropout indicators is conditioned on the values of repeated measures. The second way assumes a pattern-dependent mechanism, where the distribution of repeated measures is a mixture of distributions for subjects within distinct sub-groups determined by the patterns of dropout. And the third way adopts the mechanism of parameter-dependent dropout, where repeated measure values and dropout indicators are conditionally independent given a group of parameters share by the two parties. Correspondingly, we have three modeling techniques: selection models [9], pattern-mixture models [18], and shared-parameter models [19].

Among the three modeling strategies, selection model is relatively more intuitive. In many practical settings, the assumption of outcome-dependent dropout mechanism is appealing, and the case of censoring over cutoff values is a typical example. Selection models originated from the Tobit model of Heckman [20] and then fully developed by Diggle and Kenward [9] for continuous repeated measures [8]. Subsequently, it was extended to the non-monotone setting to deal with intermittent missing values [21]. Selection models for other type of measures were also developed, e.g., binary repeated measures [22] and ordinal categorical measures [23].

To evaluate the full likelihood function and obtain likelihood-based inferences, numerical integration methods and optimization algorithms are necessary. Without expressive forms of score function and hessian matrix, Diggle and Kenward [9] resorted to the simplex algorithm [24], a method employing no derivatives. But the algorithm converges unacceptably slowly and does not provide precision estimators. For categorical repeated measures, Molenberghs, Kenward, and Lesaffre [23] implemented the EM algorithm for selection models, but it still suffers similar symptoms of the simplex algorithm. According to [25], the Newton-Raphson algorithms can be applied with numerical derivation and integration [26], but they are very sensitive to poor initial choice of parameters and often converges to local maxima or values on or outside the boundary of the parameter space. Instead of using likelihood-based inference, this paper advocates the application of MCMC-based Bayesian inference. Taking advantages of the MCMC method in specifying prior distributions, imputing missing values, and conducting Monte Carlo integration, many computational difficulties could be alleviated. For demonstration, both simulated and practical data sets with continuous repeated measures and nonignorbale dropouts are analyzed using the selection models with MCMC

fitting algorithms.

2 MODEL AND METHOD

2.1 Modeling Repeated Measures

For a longitudinal data set with balanced design, J repeated measures are potentially observed on each of the N subjects at times $t_{i1}, \ldots, t_{iJ}$ $(i = 1, \ldots, N;\ j = 1, \ldots, J)$. For the following discussion, capital symbols are used to represent variables: e.g., $Y_1, \ldots, Y_J$ indicate response variables and $X_1, \ldots, X_K$ indicate covariates or explanatory variables. Symbols in lower case are used to represent observed or missing values: y_{ij} denote the value of Y_j and x_{ijk} denoting the value of X_k recorded at time t_{ij} $(i = 1, \ldots, N;\ j = 1, \ldots, J;\ k = 1, \ldots, K)$. Further, we let bold symbols represent vectors or matrices. For example, $\mathbf{y}_i = (y_{i1}, \ldots, y_{iJ})^T$ is vector of values for the repeated measures and $\mathbf{x}_i = [x_{ijk}]_{J\times K}$is a matrix of values of time-varying or time-independent covariates on the i^{th} subject. Within this paper, we assume that repeated measures are distributed as multivariate normal, and a repeated-measures model with structured covariance matrix can be written as

$$\mathbf{y}_i = \mathbf{x}_i\beta + \varepsilon_i$$

where $\varepsilon_i \sim N(\mathbf{0}, \mathbf{\Sigma}_i(\alpha))$ independently, α represents the parameters of the covariance matrix, and β represents the vector of fixed-effects regression coefficients. Determined by the way of parameterization of the covariance matrix, various forms of mixed models can be derived

[27]. For example, an AR(1) model assumes that

$$cov(y_{ij}, y_{ik}) = \sigma^2 \rho^{|j-k|}$$

where $\alpha = (\sigma^2, \rho)^T$. A standard random-effects model is usually expressed as

$$\mathbf{y}_i = \mathbf{x}_i\beta + \mathbf{z}_i\gamma_i + \varepsilon_i$$

where $\varepsilon_i \sim \sigma^2\mathbf{I}$, and $\mathbf{z}_i$ represents covariates associated with random effects $\gamma_i \sim N(\mathbf{0}, \mathbf{D})$. In this model, $\alpha = (\gamma, \sigma^2, \mathbf{D})^T$ and the model can be re-parameterized as the above structured-covariance model with covariance $\mathbf{\Sigma}_i = \mathbf{z}_i\mathbf{D}\mathbf{z}_i^T + \sigma^2\mathbf{I}$.

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. When missing values are introduced by dropout, the pattern of missingness can be indicated by a scalar r_i ($r_i = 2, \ldots, J+1$), which represents the actual time of withdrawal for subject i and "$r_i = J+1$" indicates the case of completion of the study. Since a subject who drops out at baseline does not contribute to the likelihood function, the case of "$r_i = 1$" is excluded from our consideration.

2.2 Full-Likelihood Functions for Incomplete Longitudinal Data

Ideally, the joint distribution of the complete repeated measures (i.e., $\mathbf{y}_i^{obs}$ and $\mathbf{y}_i^{mis}$) and dropout patterns (i.e., r_i) should be modeled jointly. Correspondingly, we have the full

likelihood function,

$$L(\theta, \phi|\mathbf{y}_i, \mathbf{x}_i, r_i) \propto \prod_{i=1}^{N} f(\mathbf{y}_i, r_i|\mathbf{x}_i, \theta, \phi)$$

where $\theta = (\alpha, \beta)$ represents parameters of the model for repeated measures, and ϕ represents the parameters of the dropout mechanism. In practice, after $f(\mathbf{y}_i, r_i|\mathbf{x}_i, \theta, \phi)$ is modeled, the missing values may be integrated out and the actual full-likelihood function for inference manifests as

$$L(\theta, \phi|\mathbf{y}_i^{obs}, \mathbf{x}_i, r_i) \propto \prod_{i=1}^{N} \int f(\mathbf{y}_i, r_i|\mathbf{x}_i, \theta, \phi) d\mathbf{y}_i^{mis}.$$

According to the possible pathways between $\mathbf{y}_i$ and r_i, there exist three ways in factoring their joint distribution: outcome-dependent factorization, pattern-dependent factorization, and parameter-dependent factorization. Thus, three modeling approaches exist for incomplete longitudinal data analysis.

(i) Selection models factor the joint distribution into a marginal distribution for $\mathbf{y}_i$ and a conditional distribution of r_i given $\mathbf{y}_i$, i.e.,

$$f(\mathbf{y}_i, r_i|\mathbf{x}_i, \theta, \phi) = f(\mathbf{y}_i|\mathbf{x}_i, \theta) f(r_i|\mathbf{y}_i, \mathbf{x}_i, \phi)$$

where the conditional distribution can be interpreted as "self-selection of the i^{th} subject into a specific dropout group."

(ii) Pattern-mixture models assume that distribution of repeated measures varies with the dropout pattern and the joint distribution is factored as

$$f(\mathbf{y}_i, r_i|\mathbf{x}_i, \theta, \phi) = f(\mathbf{y}_i|r_i, \mathbf{x}_i, \theta) f(r_i|\mathbf{x}_i, \phi).$$

In other words, for a data set with P dropout patterns, the marginal distribution of $\mathbf{y}_i$ is a mixture, $f(\mathbf{y}_i|\mathbf{x}_i,\theta) = \sum_{p=1}^{P} f(\mathbf{y}_i|r_i = r^p, \mathbf{x}_i, \theta^{(p)})\pi_p$, where $\theta^{(p)}$ represents the parameters of $f(\mathbf{y}_i)$ in the p^{th} pattern and $\pi_i = Pr(r_i = r^p|\mathbf{x}_i, \phi)$.

(iii) Shared-parameter models assume that $\mathbf{y}_i$ and r_i are conditional independent of each other, given a group of parameters, ξ_i,

$$f(\mathbf{y}_i, r_i|\mathbf{x}_i, \theta, \phi) = \int f(\mathbf{y}_i|\xi_i, \mathbf{x}_i, \theta) f(r_i|\xi_i, \mathbf{x}_i, \phi) f(\xi_i) d\xi_i.$$

From the point view of causation, shared parameters ξ_i play the role of a confounder for the relationship between ξ_i and r_i, thus can be either observable variables (e.g., gender) or unobserved variables (e.g., random-effects or latent scores).

In the rest of this paper, we focus on the strategy of selection models. For applications of other two strategies, please refer to [16], [17] and [8].

2.3 Selection Models for Nonignorable Dropouts

Before describing the selection model, let us first introduce a symbol, d_i, to denote the "possible" dropout time for the i^{th} subject who actually drops out at time r_i, i.e., $d_i = 1, \ldots, r_i$. Suppressing the dependence on covariates, a selection model assumes: (i) if $r_i < J+1$, $Pr(d_i = j)$ depends on y_{ij} and its history $\mathbf{H}_{ij} = (y_{i1}, \ldots, y_{i,j-1})^T$ $(j = 1, \ldots, r_i)$; (ii) if $r_i = J+1$, $Pr(d_i = r_i) = 1$; and (iii) the conditional distribution of y_{ij} given $\mathbf{H}_{ij}$ is $f_{ij}(y|\mathbf{H}_{ij}, \theta)$. For this selection model with outcome-dependent dropout, the full likelihood

function for the i^{th} subject is

$$L_i(\theta, \phi|\mathbf{y}_i^{obs}, r_i) \propto \prod_{j=1}^{r_i-1} f(y_{ij}|\mathbf{H}_{ij}, \theta) \prod_{j=2}^{r_i-1} [1 - Pr(d_i = j|y_{ij}, \mathbf{H}_{ij}, \phi)] Pr(d_i = r_i|\mathbf{H}_{i,r_i})$$

where the dropout probability at r_i is

$$Pr(d_i = r_i|\mathbf{H}_{i,r_i}) = \begin{cases} \int Pr(d_i = r_i|\mathbf{H}_{i,r_i}, \phi) f_{i,r_i}(y|\mathbf{H}_{i,r_i}, \theta) dy & \text{if } r_i < J+1 \\ 1 & \text{if } r_i = J+1 \ . \end{cases}$$

Using the chain rue of conditional probability, we have $\prod_{j=1}^{r_i-1} f(y_{ij}|\mathbf{H}_{ij}, \theta) = f(\mathbf{y}_i^{obs}|\theta)$.

A natural choice for calculating $Pr(d_i = j|y_{ij}, \mathbf{H}_{ij}, \phi)$ is the logistic regression,

$$\text{logit} P(d_i = j|y_{ij}, \mathbf{H}_{ij}, \phi) = \phi_0 + \phi_1 y_{ij} + \sum_{k=2}^{j} y_{i,j+1-k} \phi_k.$$

from which "$\phi_1 \neq 0$" implies a nonignorable dropout mechanism. Here, we restrict that $Pr(d_i = 1) = 0$.

The full log-likelihood function of the whole data set for (θ, ϕ) can be partitioned into

$$l(\theta, \phi) = l_1(\theta) + l_2(\phi) + l_3(\theta, \phi)$$

where $l_1(\theta) = \sum_{i=1}^{N} \log f(\mathbf{y}_i^{obs})$ corresponds to the observed-data likelihood function for θ, $l_2(\phi) = \sum_{i=1}^{N} \sum_{j=2}^{r_i-1} [1 - Pr(d_i = j|y_{ij}, \mathbf{H}_{ij}, \phi)]$ and $l_3(\theta, \phi) = \sum_{i \leq N; r_i \leq J} log(Pr(d_i = r_i|\mathbf{H}_{i,r_i}))$ together correspond to the likelihood function for the dropout process. For nonignorable dropouts, $l_3(\theta, \phi)$ contains information on θ, thus cannot be ignored. If dropouts are ig-

norable, then $l_3(\theta, \phi)$ depends only on ϕ, thus can be absorbed into $l_2(\phi)$. In this case, estimation of θ can be solely derived from $l_1(\theta)$.

For a normal longitudinal data set, $\mathbf{y}_i \sim N(\mathbf{x}_i\beta, \mathbf{\Sigma}_i(\alpha))$, which has the feature of congeniality. In other words, the conditional distribution, $f_{ij}(y|\mathbf{H}_{ij}, \theta)$, is a scalar normal, and the marginal distribution, $\prod_{j=1}^{r_i-1} f(y_{ij}|\mathbf{H}_{ij}, \theta) = f(\mathbf{y}_i^{obs})$, is also a multivariate normal. Depending on the choice of mixed models for repeated measures, different selection models can be implemented. In this paper, we specifically consider the selection models with AR(I) covariance and the one with random effects.

3 BAYESIAN INFERENCE

For selection models, Bayesian inference based on MCMC provides an appealing alternative to the likelihood-based inferences. By sampling parameters and drawing missing values, the method of Monte Carlo using Gibbs sampler or Metrapolis-Hasting algorithms offers a natural option for intergration and optimization, without relying on fully-determined density functions or analytical derivatives. In the application of the Bayesian inference to the selection model, each element of the parameter vector $\psi = (\theta, \phi)^T$ is viewed as a variable instead of a constant, certain prior distributions $f(\psi)$ are specified, and the posterior distribution of the parameters is obtained using Bayes' theorem, i.e.,

$$P(\psi|\mathbf{Y}, \mathbf{R}) \propto \left[f(\mathbf{y}_i^{obs}|\theta) \prod_{j=2}^{r_i-1} [1 - Pr(d_i = j|y_{ij}, \mathbf{H}_{ij}, \phi)] Pr(d_i = r_i|\mathbf{H}_{i,r_i}) \right] \times f(\psi).$$

Using the method of MCMC, the posterior distribution is obtained through sampling. Inferences are then made as the summary of the statistical features of this posterior distribution, e.g., median, mean, and standard deviation.

3.1 An Augmented Gibbs Sampler

When evaluating the likelihood function of the selection model, the crucial part of computation is to calculate the actual dropout probabilities, $Pr(d_i = r_i|\mathbf{H}_{i,r_i}, \phi)$, which requires integrating out the missing value y_{i,r_i} over $(-\infty, +\infty)$. Only for simple cases, is it feasible to maximize the likelihood function analytically. Using an augmented Gibbs sampler, we can alleviate the computation difficulty by first impute the missing values and then draw parameters one by one conditionally on the observed and imputed data. More specifically, the augmented Gibbs sampler is an iterative procedure with each iteration consisting of two steps.

(I) *Imputation-Step*, where the missing values are updated by drawing from the conditional predictive distribution. That is, for $i = 1$ to N, draw y_{i,r_i}^{mis} from

$$y_{i,r_i}^{mis} \sim f_{i,r_i}(y|\mathbf{y}_i^{obs}, \mathbf{x}_i, r_i, \psi).$$

For multivariate-normally distributed measures, this predictive condition is a scalar normal distribution.

(II) *Estimation-Step*, where the parameters are drawn from the posterior distribution $\psi \sim P(\psi|\mathbf{Y}^*, \mathbf{X}, \mathbf{R})$ in the following order according to the decomposition of the joint distribution

into full-conditional distributions,

$$\theta \sim f(\theta|\psi_{\backslash\theta}, \mathbf{y}_1^*, \ldots, \mathbf{y}_N^*, \mathbf{X}, \mathbf{R})$$

$$\phi \sim f(\phi|\psi_{\backslash\phi}, \mathbf{y}_1^*, \ldots, \mathbf{y}_N^*, \mathbf{X}, \mathbf{R})$$

where $\mathbf{y}_i^* = (y_{i1}, \ldots, y_{i,r_i-1}, y_{i,r_i}^{mis})^T$, $\mathbf{Y}^* = (\mathbf{y}_1^*, \ldots, \mathbf{y}_N^*)^T$, and "\" means "excluding" (e.g., $\psi_{\backslash\alpha} = (\beta, \phi)^T$).

In the above algorithm, missing values are treated as another group of parameters in an approximate sense, but missing values are in fact unobserved values controlled by the parameters. They are simulated from the predictive function, instead of full-conditional distributions. In order to differentiate from standard Gibbs sampler, we call the above algorithm "augmented" Gibbs sampler. Similar ideas were adopted by Schafer [15] in his data augmentation algorithms for creating imputations of missing values seen in multivariate data sets.

Starting from an initial point and repeating the two steps with large enough iterations, the procedure would converge to its stationary distribution, i.e., the joint distribution of parameters and the missing values. Thus, after a long enough burning period, the simulated missing values and parameters can be used for parameter estimation. Note that, this algorithm can be used for conducting multiple imputation [28], where multiple imputed data sets are first created and then analyzed using standard longitudinal models for complete data. For the purpose of imputing all missing values (i.e., $\mathbf{y}_i^{mis} = (y_{i,r_i}, \ldots, y_{iJ})^T$), we only need to replace y_{i,r_i}^{mis} with $\mathbf{y}_i^{mis}$ in the imputation step. The estimation-step remains unchanged, because the dropout probability at r_i only depends on the current and previous values (i.e.,

y_{i,r_i} and $\mathbf{H}_{i,r_i}$).

3.2 Sampling Full-Conditional Distributions

The estimation step of the above algorithm is itself a hybrid Gibbs sampler, which involves other MCMC sampling schemes for simulating parameters using full-conditional distributions. If a conditional distribution has a known form, the corresponding parameter vector can be sampled directly. For example, missing values (y_{i,r_i}^{mis} or $\mathbf{y}_i^{mis}$) can be simulated directly using a scalar or multivariate normal distribution. Otherwise, the Metropolis sampling method and its variants can be applied [29]. Nonetheless, depending on the form of the density function, we recommend the following two approaches, which are usually more efficient.

If the conditional distribution has a log-concave from, the most efficient method called *adaptive rejection sampling* [30] is applied in the following two steps: (i) set the upper hull and lower hall functions, which are piecewise linear functions respectively consisting of tangent lines and cords of the logarithmic density function at selected points; (ii) sample a point from the cumulative density function determined by the up hall, and then update the upper and lower halls depending on whether the sampled value is accepted or rejected.

When the conditional density function does not have a log-concave form, the intuitive *griddy Gibbs sampler* [31] can be applied. This sampler is based on the empirical distribution method and consists if three steps: (i) determine the range of the conditional density function up to a constant; (ii) divide the range with or without respect to the probability change to form a grid; and (iii) sample from the grid points by a simple or sophisticated version of inverse sampling.

3.3 Prior Specification

When there is no actual prior information or historical data, non-informative priors can be adopted, which usually end up with results that are consistent to those based on the maximum likelihood estimate. More specifically, normal distribution with infinite variance are used for all the regression coefficients, random-effects, and the logrithm of residual variance. Flat uniform distributions are used for space restricted parameters, such as the correlation parameter in the AR(1) structure, $-1 < \rho < 1$.

3.4 The Selection Model with AR(1) Covariance

For the AR(1) covariance matrix, we have

$$\boldsymbol{\Sigma}_J^{-1} = \frac{1}{\sigma^2(1-\rho^2)} \begin{pmatrix} 1 & -\rho & 0 & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & -\rho & 1+\rho^2 & -\rho \\ 0 & \dots & 0 & 0 & -\rho & 1 \end{pmatrix}$$

and $\det(\boldsymbol{\Sigma}_J) = (\sigma^2)^J(1-\rho^2)^{J-1}$. The augmented Gibbs sampler consists of the following steps.

I. Imputation-Step: draw missing values from

$$\begin{aligned} y_{i,r_i} \sim f_{i,r_i}(y|\mathbf{y}_i^{obs}, \mathbf{x}_i^{obs}, \psi) &\propto \frac{1}{\sqrt{2\pi\nu_i}} \exp\{-\frac{(y - \sum_{k=1}^{K} x_{i,r_i,k}\beta_k - \mu_i)^2}{2\nu_i}\} \\ &\times \frac{\exp(\phi_0 + \phi_1 y + \sum_{k=2}^{r_i} y_{i,r_i+1-k}\phi_k)}{1 + \exp(\phi_0 + \phi_1 y + \sum_{k=2}^{r_i} y_{i,r_i+1-k}\phi_k)} \end{aligned}$$

where $\mu_i = C_{(r_i-1)}^T \mathbf{\Sigma}_{r_i-1}^{-1}(\mathbf{y}_i^{obs} - \mathbf{x}_i^{obs}\beta)$ and $\nu_i = \sigma^2(1 - C_{(r_i-1)}^T \mathbf{\Sigma}_{r_i-1}^{-1} C_{(r_i-1)})$, with $\mathbf{x}_i^{obs} = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{i,r_i-1})^T$ and $C_{(j)} = (\rho^{j-1}, \ldots, \rho)^T$. This can be derived from regressing y_{i,r_i} to $\mathbf{y}_i^{obs}$.

II. Estimation-Step: draw parameters one by one in the following order

(1) For $i = 1, \ldots, K$, draw fixed parameters:

$$\beta_k \sim f(\beta_k|\psi_{\setminus\beta_k}, \mathbf{Y}^*, \mathbf{X}^*) \propto \prod_{i=1}^{N} \exp\{-\frac{(\mathbf{y}_i^* - \mathbf{x}_i^*\beta)\Sigma_{r_i}^{-1}(\mathbf{y}_i^* - \mathbf{x}_i^*\beta)^T}{2}\}$$

(2) Draw variance parameter:

$$\rho \sim f(\rho|\psi_{\setminus\rho}, \mathbf{Y}^*, \mathbf{X}^*) \propto \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi \det(\Sigma_{r_i})}} \exp\{-\frac{(\mathbf{y}_i^* - \mathbf{x}_i^*\beta)^T\Sigma_{r_i}^{-1}(\mathbf{y}_i^* - \mathbf{x}_i^*\beta)}{2}\}$$

(3) Draw variance of residuals:

$$\sigma^2 \sim f(\sigma^2|\psi_{\setminus\sigma^2}, \mathbf{Y}^*, \mathbf{X}^*) \propto \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi \det(\Sigma_i)}} \exp\{-\frac{(\mathbf{y}_i^* - \mathbf{x}_i^*\beta)^T\Sigma_i^{-1}(\mathbf{y}_i^* - \mathbf{x}_i^*\beta)}{2}\}$$

(4) For $k = 1, \ldots, J$, draw parameters of the dropout mechanism:

$$\begin{aligned} \phi_k \sim f(\phi_k|\psi_{\setminus\phi_k}, \mathbf{Y}^*) &\propto \prod_{i=1}^{N}\prod_{j=2}^{r_i-1} \frac{1}{1+\exp(\phi_0+\phi_1 y_{ij}+\sum_{k=2}^{j} y_{i,j+1-k}\phi_k)} \\ &\times \frac{\exp(\phi_0+\phi_1 y_{i,r_i}+\sum_{k=2}^{r_i} y_{i,r_i+1-k}\phi_k)}{1+\exp(\phi_0+\phi_1 y_{i,r_i}+\sum_{k=2}^{r_i} y_{i,r_i+1-k}\phi_k)} \end{aligned}$$

In the above density functions, $\mathbf{x}_i^* = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{i,r_i})^T$ and $\mathbf{X}^* = (\mathbf{x}_1^*, \ldots, \mathbf{x}_N^*)^T$. When draw each parameter, all other parameters are viewed as known constants. As shown in Appendix 1, all above conditional distributions except the one for ρ have log-concave forms, directly or after some transformation, thus can be simulated using the method of adaptive rejection sampling. A convenience sampling method for ρ is the Metroplis-Hasting algorithm.

3.5 The Selection Model with Random-Effects

In the selection model with random-effects, the random-effects model for repeated measures can be rewritten as $y_{ij} = \sum_{k=1}^{K} x_{ijk}\beta_k + \sum_{k=1}^{q} z_{ijk}\gamma_{ik} + \epsilon_{ij}$, where $\epsilon_{ij} \sim N(0, \sigma^2)$. Here, we restrict that random-effects are independent of each other with $\gamma_{ik} \sim N(0, \sigma_{\gamma_k}^2)$ $(k = 1, \ldots, q)$. In the augmented Gibbs sampler, the missing values $(y_{1,r_1}, \ldots, y_{N,r_N})$ and parameters $(\psi = (\beta, \gamma, \sigma^2, \sigma_\gamma^2, \phi))$ are simulated in the following steps.

I. Imputation-Step: Draw missing values:

$$\begin{aligned} y_{i,r_i} \sim f_{i,r_i}(y|\mathbf{y}_i^{obs}, \mathbf{x}_{i,r_i}, \mathbf{z}_{i,r_i}, \psi) &\propto \exp\{-\frac{(y-\sum_{k=1}^{K} x_{i,r_i,k}\beta_k - \sum_{k=1}^{q} z_{i,r_i,k}\gamma_{ik})^2}{2\sigma^2}\} \\ &\times \frac{\exp(\phi_0+\phi_1 y+\sum_{k=2}^{r_i} y_{i,r_i+1-k}\phi_k)}{1+\exp(\phi_0+\phi_1 y+\sum_{k=2}^{r_i} y_{i,r_i+1-k}\phi_k)} \end{aligned}$$

II. Estimation-Step: draw parameters one by one:

(1) For fixed-effect parameters:

$$\beta_k \propto f(\beta_k|\psi_{\setminus\beta_k}, \mathbf{Y}^*, \mathbf{X}^*, \mathbf{Z}^*) \propto \prod_{i=1}^{N}\prod_{j=1}^{r_i} \exp\{-\frac{(y_{ij} - \sum_{k=1}^{K} x_{ijk}\beta_k - \sum_{k=1}^{q} z_{ijk}\gamma_{ik})^2}{2\sigma^2}\}.$$

(2) For $i = 1, \ldots, N; k = 1, \ldots, q$, simulate random effects:

$$f(\gamma_{ik}|\psi_{\setminus\gamma_{ik}}, \mathbf{Y}^*, \mathbf{X}^*, \mathbf{Z}^*) \propto \prod_{j=1}^{r_i} \exp\{-\frac{(y_{ij} - \sum_{k=1}^{K} x_{ijk}\beta_k - \sum_{k=1}^{q} z_{ijk}\gamma_{ik})^2}{2\sigma^2}\} \exp\{-\frac{{\gamma_{ik}}^2}{2\sigma^2_{\gamma_k}}\}$$

(3) For $k = 1, \ldots, q$, draw variance of random-effects:

$$f(\sigma^2_{\gamma_k}|\psi_{\setminus\sigma^2_{\gamma_k}}) \propto \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2_{\gamma_k}}} \exp\{-\frac{{\gamma_{ik}}^2}{2\sigma^2_{\gamma_k}}\}$$

(4) For the residual variance:

$$f(\sigma^2|\psi_{\setminus\sigma^2}, \mathbf{Y}^*, \mathbf{X}^*, \mathbf{Z}^*) \propto \prod_{i=1}^{N}\prod_{j=1}^{r_i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(y_{ij} - \sum_{k=1}^{K} x_{ijk}\beta_k - \sum_{k=1}^{q} z_{ijk}\gamma_{ik})^2}{2\sigma^2}\}$$

(5) For the parameters of dropout model:

$$\begin{aligned} f(\phi_k|\psi_{\setminus\phi_k}, \mathbf{Y}^*) \quad &\propto \quad \prod_{i=2}^{N}\prod_{j=1}^{r_i-1} \frac{1}{1 + \exp(\phi_0 + \phi_1 y_{ij} + \sum_{k=2}^{j+1} y_{i,j+1-k}\phi_k)} \\ &\times \quad \frac{\exp(\phi_0 + \phi_1 y_{ir_i} + \sum_{k=2}^{r_i+1} y_{i,r_i+1-k}\phi_k)}{1 + \exp(\phi_0 + \phi_1 y_{ir_i} + \sum_{k=2}^{r_i+1} y_{i,r_i+1-k}\phi_k)} \end{aligned}$$

In above densities, $\mathbf{z}_i^* = (\mathbf{z}_{i1}, \ldots, \mathbf{z}_{i,r_i})^T$ and $\mathbf{Z}^* = (\mathbf{z}_1^*, \ldots, \mathbf{z}_N^*)^T$. As shown in Appendix 2, all above conditional distributions have log-concave forms, directly or after some transformation, thus can be simulated using the method of adaptive rejection sampling.

4 APPLICATION

4.1 A Simulation Study

To compare with the performance of the maximum likelihood estimation of Diggle and Kenward [9], similar design was adopted for our simulation study. Each generated data set consists of two groups of equal number units: the "control" and "treatment" groups. Repeated measures are simulated using the mixed model with AR(1) covariance and linear trend: $E(y_{ij}) = \beta_0 + \beta_1 t_j + \beta_2 t_j g_i$ $(i = 1, \ldots, N; j = 1, \ldots, J)$, where group indicator g_i=1 (or 0) if the i^{th} unit belongs to the treatment (or control) group. Parameters were set as: $\beta_0 = 10$, $\beta_1 = 0$, $\beta_2 = -1$, and $\sigma^2 = 1$. The interest lies in the parameter β_2, which indicates the difference in slope between the two groups.

For each data set, sample size was chosen as either $N = 50$ or $N = 100$, correlation parameter ρ took values 0.5, 0.75, and 0.9, while number of repeated measures for completers was fixed at $J = 10$. For each combination of ρ and N, 100 sets of data were generated by using multivariate-normal pseudo-random number generator. Then, for each set, three mechanisms of dropout were applied according to the logistic model:

$$\text{logit}P(d_i = j|y_{ij}, \ldots, y_{i1}, \phi) = \phi_0 + \phi_1 y_{ij} + \phi_2 y_{i,j-1}.$$

Three cases were considered here: (i) DCAR (dropout completely at random), i.e., $\phi_1 = \phi_2 = 0$; (ii) DAR (dropout at random), i.e., $\phi_1 = 0$; and (iii) ND (nonignorable dropout), i.e., $\phi_2 = 0$. The parameters of the three dropout mechanisms were specified to achieve about 33%, 50%, and 66% of missing values in the data matrix.

The relative biases, expressed as $(\hat{\beta}_2 - \beta_2)/\beta_2 * 100\%$, in estimating the slope difference are shown in Table 1. The coverage rate of the estimated slope difference are shown in Table 2. Seen from Table 1, the relative bias in the estimate of slope difference is almost negligible for most data sets even for those with only 50 units. The estimate of the parameter of interest is very insensitive to the missing rates or correlation parameter ρ. The performance of augmented Gibbs sampler is quite satisfactory from the perspective of unbiasedness. From table 2, it is seen that the coverage rates are mostly around or above 95 percent for most data sets. The coverage rates are fairly insensitive across the the dropout mechanism and level of ρ. Using this criterion of coverage, the augmented Gibbs sampler, is proved to be very powerful for the AR(1) selection model. Compared with the results based on maximum likelihood estimation with simplex algorithm, our Bayesian inference with Gibbs sampler provides better solutions.

<INSERT TABLE 1 HERE>

<INSERT TABLE 2 HERE>

4.2 A Carbon Monoxide Data Set

A longitudinal data set containing carbon monoxide levels was analyzed by Shoptaw et al. [32]. The data set came from a smoking cessation clinical trial conducted on 175 methodone-maintained tobacco smokers to investigate the treatment efficacy of two behavioral therapies: contingency management (CM) and relapse prevention (RP). The smokers were randomized into four groups: Control ($N_1 = 43$), RP-only ($N2 = 42$), CM-only ($N_3 = 43$), and RP+CM ($N_4 = 47$). During the 12-week treatment, for each participant, up to 36 measures of carbon

monoxide levels were assessed from expired breath samples three time per week. Ignoring the 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 on the figure are the point wise standard deviations and ANOVA with p-values smaller than 0.001. Assuming that missing values are ignorable, a mixed model with AR(1) covariance and piecewise linear trend supported a favorable treatment efficacy of contingency management.

<INSERT FIGURE 1 HERE>

We reanalyzed the data by fitting an AR(1) selection model to the carbon monoxide levels after week one. After a step-wise selection procedure, the following mean structure was chosen to investigate the treatment effects of the two behavioral therapies,

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

where CM_i and RP_i respectively indicate whether the i^{th} smoker received contingency management or relapse prevention, $BaseCO_i$ indicates baseline carbon monoxide level, and $Patches_i$ represents the number of nicotine patches the smoker received during the study. For modeling the dropout mechanism, the logistic regression model was fitted with r_i depending on $y_{i,r_i}$and y_{i,r_i-1}.

Since there are moderate amount of intermittent missing values in the data set, the strategy of multiple partial imputation of Yang and Shoptaw [4] was adopted. This is a extended method of multiple imputation [28], where intermittent missing values are imputed first, then the partially-imputed data sets are analyzed using the selection model, and finally

the multiple estimated parameters are combined to make an overall inferential statement. Using PROC MI of SAS with the monotone option, four versions of intermittent missing values were imputed. By applying the AR(1) selection model to each of the four partially-imputed data set, the estimates of interesting parameters are shown in Table 3, from which we notice small between-imputation variances for all the parameters. In other words, the fraction of missing information due to intermittent missingness was low. After consolidating the four sets of estimates using a set of rules of Rubin (1987), it is seen that only the treatment effect of contingency management is significant ($\hat{\beta}_1 = -0.28$; $T_{2490} = -5.88$ with $p < 0.0001$). Relapse prevention turns to be ineffective and there is no interaction effect between the two behavioral therapies. The estimated regression coefficient $\hat{\phi}_1$ 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 is, the larger probability of dropping out. Therefore, the dropouts are outcome-dependent nonignorable.

<INSERT TABLE 3 HERE>

By diving all the smokers into two sub groups: completers (those who completed the study) and withdrawers (those who dropped out prematurely), we aimed to study the influence of dropout to the treatment efficacy. Using the same method to created three versions of intermittent missing values, then the AR(1) selection model was fitted to each of the three partially-imputed data sets. We fitted the model to the data collected on all the subjects and to the subset data collected on the withdrawers. For data from completers, only the counterpart of the AR(1) model for repeated measures was fitted since there are no dropouts in this sub population. The estimated parameters of interest are listed in Table 4. It is

seen that the results are very different across sub populations but do not vary much across versions of partial imputations. Contingency management turns out to be significantly effective for the completers. It may be still effective for withdrawers, but the significance level is less convincing. The dropout mechanism supported by all the subjects seems to be extremely outcome-dependent nonignorable (i.e., $\phi_1 \neq 0$, $\phi_2 = 0$), but the dropout mechanism estimated from the withdrawers seems to be intermediate between the case of "dropout at random" and the case of pure nonignorability. Although both $\hat{\phi}_1$ and $\hat{\phi}_2$ are not significant, their positivity suggests that the higher the current and the previous carbon monoxide levels, the larger probability that subject would drop out of the study. Because the number of subjects who dropped out is only about half of the number of completers and there are less observed data points in the withdrawers' data, the corresponding information is significantly lower for this smaller group of subjects.

<INSERT TABLE 4 HERE>

5 DISCUSSION

After a brief review on modeling strategies for incomplete longitudinal data analysis based on full-likelihood functions, this paper focused on the specific case of selection modeling to deal with potentially nonignorable dropouts. An augmented Gibbs sampler was proposed for making Bayesian inferences to two forms of selection model, one with AR(1) covariance structure and one with random-effects model in modeling repeated measures. Both simulation studies and practical applications confirmed the validity of the Bayesian inference using the augmented Gibbs sampler and and its superiority over the likelihood-based inference

using the method of maximum likelihood estimation.

The selection models with autoregressive and random-effects provide a simpler and convenient solution for nonignorable dropouts. These two mixed-effects models are popularly available in commercial software packages and practical experience have proved their validity and efficiency in modeling repeated measures in clinical trials, which usually come with small to moderate sample sizes. By implementing the AR(1) and the random-effects model with independent random effects, we aimed to provide two typical examples, so that other options of covariance-structured models for various types of repeated measures can be developed similarly within the framework of generalized linear mixed models [6]. When modeling the dropouts, the logistic regression model was used in the paper, but there are many other possibilities. For example, the length of the treatment before dropout can be modeled using survival analysis techniques. A two-level latent model can be applied, too, where the scalar continuous latent variable is regressed to the values of repeated measures, and then it is dichotomized to indicate whether the subject drops out.

When conducting Bayesian computation using the Gibbs sampler, either informative or noninformative prior distributions can be applied. For a practical data, if there are no informative priors, a general suggestion is to adopt the idea of empirical Bayes, where the space of parameters are restricted to the neighborhood of the values elicited from the estimation based on observed-data likelihood. For example, to determine the proper scope of regression coefficients of the repeated-measures model, we may run "PROC Mixed" in SAS to obtain maximum likelihood estimates assuming that dropouts are ignorable. From the fitted model, the possible scopes of values of fixed parameters can be elicited. Similar, we can model the missingness indicators to determine the neighborhood of the parameters of the logistic re-

gression model by assuming that dropouts are "dropout completely random" and "dropout at random" [4], [14]. This method can help us select initial points that are not far from the center of the true posterior distributions. When modeling the converges of the augmented Gibbs sampler, either the summary for simulated parameters or the summary for missing values can be used. Because the dimension of the parameter vector is usually much smaller than that of the missing values, it is usually more efficient to monitor the parameters.

Finally, we emphasize the problem of high sensitivity of the incomplete data modeling strategies. Because the true model and mechanism for measurement and missingness are usually unverifiable from practical contexts, it is recommended that multiple models should be fitted to the same set of data so that the influence of assumption on missingness or dropout could be investigated. When conducting sensitivity analysis, we recommend the joint application of selection, pattern-mixture, and shared-parameter models. For this purpose, the selection models proposed here along with the Bayesian inferences provide a useful device. Bayesain inferences for Pattern-mixture models and a shared-parameter model called random-effects Markov transition model are presented in Yang et al. [13] and Li, et al. [33]. A software package named MPI [25] is available at request. It has implemented selection, pattern-mixture, and shared-parameter models with MCMC-based Bayesian inferences.

ACKNOWLEDGEMENT

This work was partially supported by the National Institute of Drug Abuse through an SBIR contract N44 DA35513 and a research grants R03 DA016721. 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] Molenberghs G, Verbeke G. A review on linear mixed models for longitudinal data, possibly subject to dropout. *Statistical Modelling: An International Journal* 2001; **1**:235-269.

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

[4] 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.

[5] 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.

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

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

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

[9] Diggle P, Kenward MG. Informative dropout in longitudinal data analysis. *Applied Statistics* 1994; **43**:49C93 [With discussion].

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

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

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

[13] Yang X, Li J, Liu G, Zhang Q, Shoptaw S. Multiple Partial Imputation for Longitudinal Data with Missing Values in Clinical Trials. Submitted to *Statistics in Medicine* 2005.

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

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

[16] Yang X, Nie K, Belin T, Zhang Q, Shoptaw S. Markov Transition Models for Binary Repeated Measures with Missing Vaslues in Substance Abuse Studies. *Proceedings of the Joint Statistical Meetings* 2004, Toronto, Canada.

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

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

[19] Albert PS, Follman DA. A Random Effects Transition Model for Longitudinal Binary Data with Informative Missingness. *Statistica Neerlandica* 2003; **57**: 100-111.

[20] 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.

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

[22] 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.

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

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

[25] Yang X, Belin T, Zhang Q, Shoptaw S. Incomplete Longitudinal Data Analysis with Multiple Partial Imputation. *Technical Report of BayesSoft, Inc.* 2002.

[26] Dennis JE, Schnabel RB. *Numerical methods for unconstrained optimization and non-linear equations*. Philadelphia: Society for Industrial and Applied Mathematics, 1996.

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

[28] Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons, 1987.

[29] Robert RC, Casella G. *Monte Carlo Statistical Methods*. Springer Texts in Statistics, 2004.

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

[31] 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.

[32] 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.

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

Appendix: Proof of Log-Concave Form of the Condition Distributions

1. The Selection Model with AR(1) Covariance

(1) For the missing values, the predictive function is log-concave because

$$\partial^2 log f_{i,r_i}(y|\mathbf{y}_i^{obs}, \mathbf{x}_i^{obs}, \psi)/\partial y_{i,r_i}^2 = -\frac{1}{\nu_i} - \frac{\phi_1^2 \exp(\phi_0 + \phi_1 y + \sum_{k=2}^{r_i} y_{i,r_i+1-k}\phi_k)}{(1+\exp(\phi_0 + \phi_1 y + \sum_{k=2}^{r_i} y_{i,r_i+1-k}\phi_k))^2} < 0$$

(2) For the fixed-parameters, the conditional density function is log-concave because

$$\partial^2 log f(\beta_k|\psi_{\setminus\beta_k}, \mathbf{Y}^*, \mathbf{X}^*)/\partial\beta_k = \sum_{i=1}^{N} -(x_{i1k}, \ldots, x_{i,r_i,k})\Sigma_{r_i}^{-1}(x_{i1k}, \ldots, x_{i,r_i,k})^T < 0$$

(3) For the residual variance, the conditional density function after a logarithm transform, $s = log(\sigma^2)$, is log-concave

$$\partial^2 log f(s|\psi_{\setminus s}, \mathbf{Y}^*, \mathbf{X}^*)/\partial s^2 = -\sum_{i=1}^{N} \frac{(\mathbf{y}_i^* - \mathbf{x}_i^*\beta)^T\Sigma(\rho)_{r_i}^{-1}(\mathbf{y}_i^* - \mathbf{x}_i^*\beta)}{2e^s} < 0$$

(4) For the parameters of the dropout mechanism, the conditional density function is log-concave because

$$\begin{aligned}\partial^2 log f(\phi_k|\psi_{\setminus\phi_k}, \mathbf{Y}^*)/\partial\phi_k^2 &= \sum_{i=1}^{N}\sum_{j=2}^{r_i-1} -\frac{y_{i,j+1-k}^2 \exp(\phi_0 + \phi_1 y_{ij} + \sum_{k=2}^{j} y_{i,j+1-k}\phi_k)}{(1+\exp(\phi_0 + \phi_1 y_{ij} + \sum_{k=2}^{j} y_{i,j+1-k}\phi_k))^2} \\ &- \frac{y_{i,r_i+1-k}^2 \exp(\phi_0 + \phi_1 y_{i,r_i} + \sum_{k=2}^{r_i} y_{i,r_i+1-k}\phi_k)}{(1+\exp(\phi_0 + \phi_1 y_{i,r_i} + \sum_{k=2}^{r_i} y_{i,r_i+1-k}\phi_k))^2} < 0\end{aligned}$$

2. The Selection Model with Random-Effects

(1) For the missing values, the predictive function is log-concave because

$$\partial^2 log f_{i,r_i}(y|\mathbf{x}_{i,r_i}, \mathbf{z}_{i,r_i}, \psi)/\partial y_{ir_i}^2 = -\frac{1}{\sigma^2} - \frac{\phi_1^2 \exp(\phi_0 + \phi_1 y + \sum_{k=2}^{r_i} y_{i,r_i+1-k}\phi_k)}{(1+\exp(\phi_0 + \phi_1 y + \sum_{k=2}^{r_i} y_{i,r_i+1-k}\phi_k))^2} < 0$$

(2) For the fixed-parameters, the conditional density function is log-concave because

$$\partial^2 log f(\beta_k|\psi_{\setminus\beta_k}, \mathbf{Y}^*, \mathbf{X}^*, \mathbf{Z}^*)/\partial\beta_k^2 = \sum_{i=1}^{N}\sum_{j=1}^{r_i} -\frac{(x_{ijk})^2}{\sigma^2} < 0$$

(3) For the random effects, the conditional density function is log-concave because

$$\partial^2 log f(\gamma_{ik}|\psi_{\setminus\gamma_{ik}}, \mathbf{Y}^*, \mathbf{X}^*, \mathbf{Z}^*)/\partial\gamma_{ik}^2 = \sum_{j=1}^{r_i} -\frac{(z_{ijk})^2}{\sigma^2} - \frac{1}{\sigma_{\gamma_k}^2} < 0$$

(4) For the variance of random effects, the conditional density function after a logarithm transform, $s = \log(\sigma_{\gamma_k}^2)$, is log-concave,

$$\partial^2 log f(s|\psi_{\setminus s}^2, \mathbf{Y}^*)/\partial s^2 = -\sum_{i=1}^{N} \frac{{\gamma_{ik}}^2}{2e^s} < 0$$

(5) For the residual variance, the conditional density function after a logarithm transform, $s = log(\sigma^2)$, is log-concave

$$\partial^2 log f(s|\psi_{\setminus s}, \mathbf{Y}^*, \mathbf{X}^*, \mathbf{Z}^*)/\partial s^2 = -\sum_{i=1}^{N}\sum_{j=1}^{r_i} \frac{(y_{ij} - \sum_{k=1}^{K} x_{ijk}\beta_k - \sum_{k=1}^{q} z_{ijk}\gamma_{ik})^2}{2e^s} < 0$$

(6) For the parameters of the dropout mechanism, the conditional density function is log-concave because

$$\begin{aligned}
\partial^2 log f(\phi_k|\psi_{\setminus \phi_k}, \mathbf{Y}^*)/\partial \phi_k^2 &= \sum_{i=1}^{N}\sum_{j=2}^{r_i-1} -\frac{y_{i,t+1-k}^2 \exp(\phi_0 + \phi_1 y_{ij} + \sum_{k=2}^{j} y_{i,j+1-k}\phi_k)}{(1 + \exp(\phi_0 + \phi_1 y_{ij} + \sum_{k=2}^{j} y_{i,j+1-k}\phi_k))^2} \\
&- \frac{y_{i,r_i+1-k}^2 \exp(\phi_0 + \phi_1 y_{i,r_i} + \sum_{k=2}^{r_i} y_{i,r_i+1-k}\phi_k)}{(1 + \exp(\phi_0 + \phi_1 y_{i,r_i} + \sum_{k=2}^{r_i} y_{i,r_i+1-k}\phi_k))^2} < 0
\end{aligned}$$

Dropout Mechanism	DCAR			DAR			ND		
Average Missing Rate	33%	50%	66%	33%	50%	66%	33%	50%	66%
Sample size N=50									
$\rho = 0.50$	-0.05	-0.30	-0.49	0.06	0.22	0.22	0.01	-0.07	-0.35
$\rho = 0.75$	-0.33	-0.18	0.18	0.12	-0.06	0.28	-0.03	-0.06	-0.49
$\rho = 0.90$	0.11	-0.03	0.11	-0.24	0.07	0.22	0.09	-0.5	-0.64
Sample size N=100									
$\rho = 0.50$	-0.25	-0.03	0.03	0.08	0.11	-0.05	-0.57	0.12	0.01
$\rho = 0.75$	0.00	-0.12	0.03	-0.31	0.13	-0.22	0.04	0.05	-0.44
$\rho = 0.90$	0.08	0.22	0.05	-0.10	-0.18	-0.60	-0.39	-0.17	-0.10

Table 1: Averaged percentage bias (%) in the estimate of slope difference from the 100 simulated data sets.

Dropout Mechanism	DCAR			DAR			ND		
Average Missing Rate	33%	50%	66%	33%	50%	66%	33%	50%	66%
Sample Size N=50									
$\rho = 0.50$	96	95	96	98	93	95	96	96	95
$\rho = 0.75$	95	95	95	95	96	94	96	96	92
$\rho = 0.90$	95	96	94	94	98	95	96	95	94
Sample Size N=100									
$\rho = 0.50$	95	93	94	97	95	95	94	93	95
$\rho = 0.75$	95	91	92	90	98	93	95	94	94
$\rho = 0.90$	97	95	96	93	94	98	97	96	95

Table 2: Coverage rate (%) in the estimate of slope difference from the 100 simulated data sets.

Partial Imputation #	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)
$\hat{\beta}_3$(S.D.)	-0.08 (0.06)	-0.10 (0.07)	-0.08 (0.07)	-0.08 (0.06)	-0.08 (0.07)
ϕ_2(S.D.)	-0.02 (0.24)	-0.08 (0.20)	-0.00 (0.23)	-0.02 (0.23)	-0.03 (0.23)
ϕ_1(S.D.)	1.27 (0.37)	1.37 (0.28)	1.24 (0.34)	1.25 (0.31)	1.28 (0.33)

Table 3: Estimate of treatment effect and parameters for dropout probability on four imputed carbon monoxide data sets.

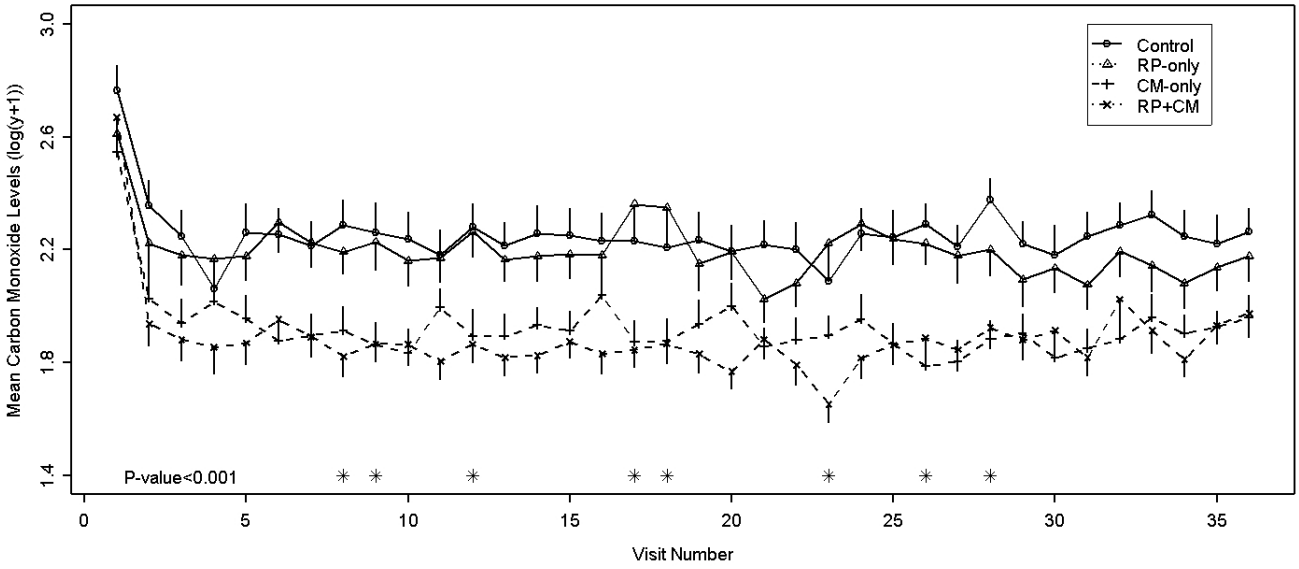
Partial Imputation #	subjects	$\hat{\beta}_1$(S.D.)	$\hat{\phi}_2$(S.D.)	$\hat{\phi}_1$(S.D.)
1	All units	-0.29(0.05)	-0.02(0.24)	1.27(0.37)
	Completers	-0.35(0.06)	–	–
	Dropouts	-0.11(0.09)	0.31(0.51)	0.27(0.30)
2	All units	-0.27(0.05)	-0.08(0.20)	1.37(0.28)
	Completers	-0.34(0.05)	–	–
	Dropouts	-0.07(0.10)	0.48(0.46)	0.14(0.27)
3	All units	-0.28(0.05)	-0.00(0.23)	1.24(0.34)
	Completers	-0.34(0.06)	–	–
	Dropouts	-0.10(0.09)	0.41(0.35)	0.22(0.25)

Table 4: Estimate of treatment effect and parameters of dropout mechanism on three imputed carbon monoxide data sets, repeated for three types of subjects: all units, completers-only, and withdrawers-only.

FIGURE CAPTIONS

Figure 1. The average and standard deviation (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+y) transform. X-axis represents number of clinic visit for study participants $(1, \ldots, 36)$.

Figure 1.


Mean Carbon Monoxide Levels (log(y+1))
3.0
2.6
2.2
1.8
1.4
0
5
10
15
20
25
30
35
Visit Number
Control
RP-only
CM-only
RP+CM
P-value<0.001