Biometrical Journal **0** (2011) 0, zzz–zzz / DOI: 10.1002/

An exploration of fixed and random effects selection for longitudinal binary outcomes in the presence of non-ignorable dropout

Ning Li[*,1], **Michael J. Daniels** [2], **Gang Li** [3] , and **Robert M. Elashoff** [1,3]

[1] Department of Biomathematics, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA 90095, U.S.A.

[2] Department of Statistics, University of Florida, Gainesville FL 32611, U.S.A.

[3] Department of Biostatistics, School of Public Health, University of California at Los Angeles, Los Angeles, CA 90095, U.S.A.

We explore a Bayesian approach to selection of variables that represent fixed and random effects in modeling of longitudinal binary outcomes with missing data caused by dropouts. We show via analytic results for a simple example that non-ignorable missing data lead to biased parameter estimates. This bias results in selection of wrong effects asymptotically, which we can confirm via simulations for more complex settings. By jointly modeling the longitudinal binary data with the dropout process that possibly leads to non-ignorable missing data, we are able to correct the bias in estimation and selection. Mixture priors with a point mass at zero are used to facilitate variable selection. We illustrate the proposed approach using a clinical trial for acute ischemic stroke.

Key words: Bayesian variable selection; Bias; Dropout; Missing data; Model selection.

1 Introduction

In models with a large number of candidate predictors, criterion-based strategies for variable selection such as Aikake's information criterion or the Bayesian information criterion are often used. However, these strategies can be computationally prohibitive because the number of possible sub-models to consider grows exponentially with the number of effects. Model selection for longitudinal data has also been studied using both penalized likelihood (Fu, 2003; Fan and Li, 2004) and Bayesian methods (George and McCulloch, 1993; Chen and Dunson, 2003; Cai and Dunson, 2006; Kinney and Dunson, 2007; Park and Casella, 2008; Chen *et al*., 2009).

However, missing data due to dropout or death during follow-up in longitudinal studies could compromise the validity of the aforementioned variable selection procedures. According to Little and Rubin (2002), missing data mechanisms are classified into three types: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). The first two mechanisms assume that the probability of missingness is not related to the missing values; they are called *ignorable* because valid statistical inference can be obtained by analyzing the available data without considering missing data mechanisms. In the MNAR mechanism, the probability of missingness is related to the missing values. As an example, a study that evaluates the impact of a new treatment on disease outcome may have MNAR values at the longitudinal endpoint because patients die due to worsening disease or drop out due to poor treatment efficacy. The probability of missingness could be related to the poor outcome that would have been observed in the patients' final measurements in the absence of death or dropout. Statistical analysis that ignores the MNAR mechanism can produce biased estimates (we will use MNAR and non-ignorable interchangeably here although we recognize that non-ignorable MAR is considered in some work focusing on mixture models).

*Corresponding author: nli@biomath.ucla.edu, Phone: 1-310-794-7322, Fax: 1-310-267-2611

The bias caused by analyzing data with MNAR values as if the missing data mechanism is ignorable can be corrected by modeling the association between the longitudinal measurements and the missing data mechanism. A growing literature focuses on joint analysis of normally-distributed longitudinal outcomes (Diggle and Kenward, 1994; Little, 1994; Wulfshon and Tsiatis, 1997; Henderson et. al., 2000; Hogan and Daniels, 2002; Elashoff et. al., 2007). Follman and Wu (1995), Ibrahim et. al. (2001), and Roy and Daniels (2008) extend joint models to generalized linear mixed models for exponential family distributions.

Yet, research on variable selection in the presence of MNAR values is limited. Mitra and Dunson (2010) developed a stochastic search variable selection approach for generalized linear models with missing predictors. This method cannot be used for longitudinal studies. Paddock (2007) proposed a pattern-mixture model to characterize the effects of informative censoring on the trajectory of longitudinal data. The functional form of the longitudinal trajectory was selected by using a stochastic search variable selection procedure. However, the approach does not focus on estimation and selection bias for fixed and random effects at the longitudinal endpoint.

In this paper, we address the selection of fixed and random effects for longitudinal binary outcomes with monotone missing data patterns due to dropout or death. We analytically examine the estimation bias and asymptotic behavior of posterior selection probabilities in a simplified setting. We show that a selection procedure that ignores missing data mechanisms asymptotically chooses the wrong effects when the missing data mechanism is non-ignorable.

We use a Bayesian joint analysis framework to correct the bias in estimation and selection. In the joint analysis, the dropout hazard is modeled by a logistic regression with current and prior responses of the longitudinal outcome included as covariates. Applying the ideas of Kinney and Dunson (2007), we select fixed and random effects in the logistic mixed model for the longitudinal binary outcome by imposing mixture priors with a point mass at zero. The parameters of the random effects are presented using a Cholesky decomposition of the covariance structure. We facilitate selection of the missing data mechanism by applying similar mixture priors to the parameters in the logistic model for dropout hazard. Conditional linearity of the parameters in the logistic models is achieved by a data augmentation strategy and by approximating the logistic density using the t-distribution (Albert and Chib, 1993; Kinney and Dunson, 2007). The mixture prior has the appealing feature that, by setting a positive probability at zero, effective exclusion of fixed and random effects is allowed. This prior is a conjugate for the proposed joint model under reparameterization and approximation and eliminates the need to compute model selection criteria such as DIC (Spiegelhalter et al, 2002). We use simulation studies to examine this modeling framework and the problem of non-ignorable missing data in complex analyses.

The motivation for this paper was a clinical trial of intravenous recombinant tissue plasminogen activator (rt-PA) in patients with acute ischemic stroke (National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group, 1995). The Barthel Index measures performance in daily living activities. Information on whether or not the patients had a favorable outcome on the Barthel Index was recorded at 7-10 days, 3 months, 6 months, and 12 months post stroke onset. Our goal is to select the baseline factors that predict post stroke performance from a pool of candidate variables, including health variables such as history of diabetes, hypertension, angina at baseline, demographic characteristics, and lifestyle variables. We are also interested in estimation and selection of the model components that describe the trajectory of the outcome. 30% of the Barthel Index data was missing due to dropout and death at 12 months. We apply our Bayesian framework to the rt-PA data.

The rest of the article is organized as follows. In Section 2, we present analytic results for the bias in variable selection caused by non-ignorable missing data. A specific framework for implementing variable selection using a selection model factorization is given in Section 3. Section 4 reports on simulation results for more complex settings than those examined analytically in Section 2. Section 5 illustrates the approach using data from the rt-PA clinical trial for acute ischemic stroke. Section 6 concludes with a discussion.

2 Analytical results to investigate the bias in posterior selection probabilities

In this section we show via a simple example that non-ignorable missing data leads to biased estimation. We further prove that the variable selection procedure ignoring the missing data mechanism selects the wrong effects asymptotically. We consider a scenario in which there are only two measurements, Y_{1i} and Y_{2i}, for the binary longitudinal outcome Y_i at visit 1 and visit 2, where i denotes the i^{th} individual and $i = 1, \ldots, n$. Let $Y = (Y_1, Y_2)$ and assume that all Y_{1i}'s are observed and Y_{2i}'s are possibly missing with a missingness mechanism that takes the form

$$logit\{pr(R_{2i} = 1|Y_i, \eta)\} = \eta_0 + \eta_1 Y_{1i} + \eta_2 Y_{2i}, \tag{1}$$

where $\eta = (\eta_0, \eta_1, \eta_2)$ are model parameters and R_{2i} is an indicator variable; its value is equal to 1 when Y_{2i} is observed and equal to 0 when Y_{2i} is missing. This mechanism indicates that the probability of observing Y_{2i} is a function of Y_{1i} and Y_{2i}. Let Y_{obs} and Y_{mis} represent observed and missing components of Y, respectively. When $\eta_2 \neq 0$, the missing data are MNAR.

The distribution of Y can be parameterized as follows:

$$\begin{aligned} p_{00} &= pr(Y_1 = 0, Y_2 = 0), \\ p_{10} &= pr(Y_1 = 1, Y_2 = 0), \\ p_{01} &= pr(Y_1 = 0, Y_2 = 1), \\ p_{11} &= 1 - p_{00} - p_{10} - p_{01} = pr(Y_1 = 1, Y_2 = 1). \end{aligned}$$

Denote $\theta = (p_{00}, p_{10}, p_{01})$. Let $\hat{\theta} = (\hat{p}_{00}, \hat{p}_{10}, \hat{p}_{01})$ be the maximum likelihood estimator of θ derived from the observed data likelihood, i.e., $L(\theta|Y_{obs})$. It can be shown that $\hat{\theta}$ is an inconsistent estimator of θ when $\eta_2 \neq 0$ (see Appendix A Theorem A.1).

Next we study the impact of MNAR missing data on variable selection. Suppose that $\hat{\theta}$ converges to $\theta^* = (p^*_{00}, p^*_{10}, p^*_{01})$ as n tends to infinity. We know that $\theta^* \neq \tilde{\theta}$ if $\eta_2 \neq 0$, where $\tilde{\theta} = (\tilde{p}_{00}, \tilde{p}_{10}, \tilde{p}_{01})$ is the true value of θ. We then reparameterize the distribution of (Y_1, Y_2) using the following regression model

$$logit\{pr(Y_t = 1|Y_{t-1}, \beta)\} = \beta_0 + \beta_1(t-1) + \beta_2(t-1)Y_{t-1}$$

for $t = 1, 2$ and $Y_0 = 0$. The parameters β_0, β_1 and β_2 are functions of p_{00}, p_{10} and p_{01}. Suppose that the maximum likelihood estimator of β derived from the observed data likelihood converges to $\beta^* = (\beta^*_0, \beta^*_1, \beta^*_2)$ as n tends to infinity. It is easy to show that $\beta^* \neq \tilde{\beta}$ if $\eta_2 \neq 0$, where $\tilde{\beta}$ is the true value of β (see Appendix A).

For simplicity, we assume that $\tilde{\beta}_0$ and $\tilde{\beta}_1$ are known (i.e., $\tilde{p}_{00}$, $\tilde{p}_{01}$ are known). The estimator of p_{10} converges almost surely to p^*_{10} and $p^*_{10} \neq \tilde{p}_{10}$ if $\eta_2 \neq 0$. Let $\Delta = \beta^*_2 - \tilde{\beta}_2$. It can be shown that $\Delta = logit\frac{\tilde{p}_{10}+\tilde{p}_{11}-p^*_{10}}{\tilde{p}_{10}+\tilde{p}_{11}} - logit\frac{\tilde{p}_{11}}{\tilde{p}_{10}+\tilde{p}_{11}} \neq 0$ if $\eta_2 \neq 0$. When the sample size is large, the likelihood for β_2 conditional on Y_{obs} can be approximated by a normal density up to a constant factor C, that is,

$$L(\beta_2|Y_{obs}) \approx C \times f_N(\beta_2; \beta^*_2, \sigma^2_{\beta^*_2}),$$

where $f_N(\beta_2; \beta^*_2, \sigma^2_{\beta^*_2})$ is a normal density function with mean β^*_2 and variance $\sigma^2_{\beta^*_2}$. We assume the prior for β_2 is a zero-inflated normal $ZI\text{-}N(\pi_0; 0, \sigma^2_0)$. Its density function can be expressed as $\pi_0 I(\beta_2 = 0) + (1 - \pi_0) I(\beta_2 \neq 0) f_N(\beta_2; 0, \sigma^2_0)$, in which π_0 is the prior probability that $\beta_2 = 0$ (please refer to Section 3.3 for more details on this prior). It can be shown that the posterior probability that $\beta_2 = 0$ is $\pi^* = \frac{\pi_0}{\pi_0 + (1-\pi_0)\frac{\sigma}{\sigma_0} exp(\frac{\sigma^2}{2\sigma^4_{\beta^*_2}}\beta^{*2}_2)}$, where $\sigma^2 = \frac{1}{\frac{1}{\sigma^2_0} + \frac{1}{\sigma^2_{\beta^*_2}}}$. The posterior probability π^* is a function of $\tilde{\beta}_2$, Δ and $\sigma^2_{\beta^*_2}$. As $n \to \infty$, it can be shown that

(a) When $\eta_2 = 0$ and $\tilde{\beta}_2 = 0$, $\pi^* \to 1$.

(b) When $\eta_2 = 0$ and $\tilde{\beta}_2 \neq 0$, $\pi^* \to 0$.
(c) When $\eta_2 \neq 0$ and $\tilde{\beta}_2 = 0$, $\pi^* \to 0$.
(d) When $\eta_2 \neq 0$ and $\tilde{\beta}_2 \neq 0$, $\beta_2^* \neq 0$ in general, and $\pi^* \to 0$.

Therefore, asymptotically the procedure selects the correct effects if the missing data are ignorable ($\eta_2 = 0$; scenarios (a) and (b)), but picks the wrong ones if $\eta_2 \neq 0$ and the missing data mechanism is ignored. For scenario (d), π^* tends to be larger than the true posterior exclusion probability in finite samples if $|\beta_2^*| < |\tilde{\beta}_2|$.

In what follows, we propose a Bayesian approach to variable selection while taking into account possible non-ignorable missing values in Y by joint modeling the missingness mechanism.

3 Variable selection framework and models

3.1 Kinney and Dunson model/approach for full data response model

In this section we extend the simple example discussed in Section 2 to longitudinal data with multiple observations over time. Let Y_{ij} denote the longitudinal binary outcome on subject i at time t_j, $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$. We assume that all the subjects follow the same schedule of visits. Some may drop out of the study early (or die) so that $n_i \leq m$, where m is the maximum number of visits. Let a $p \times 1$ vector $X_{ij}^{(1)}$ and a $q \times 1$ vector Z_{ij} collect covariates associated with Y_{ij}. The logistic mixed model can be expressed as

$$logit(p_{ij}) = logit\{pr(Y_{ij} = 1|X_{ij}^{(1)}, Z_{ij}, u_i, \beta, \Sigma)\} = X_{ij}^{(1)T}\beta + Z_{ij}^T u_i, \tag{2}$$

where β is a $p \times 1$ vector of unknown parameters to represent the fixed effects of $X_{ij}^{(1)}$ and $u_i = (u_{i1}, \ldots, u_{iq})$ are random effects to characterize inter-subject heterogeneity. We assume that $u_i \sim N_q(0, \Sigma)$.

In the variable selection context, the vectors $X_{ij}^{(1)}$ and Z_{ij} usually consist of all the candidate predictors. A stochastic search variable selection approach can be applied by choosing mixture priors that permit dropping predictors through setting their coefficients at zero. To facilitate random effects selection, Dunson and colleagues (Chen and Dunson, 2003; Cai and Dunson, 2006; Kinney and Dunson, 2007) proposed a modified Cholesky decomposition of Σ

$$\Sigma = \Lambda\Gamma\Gamma^T\Lambda,$$

where $\Lambda = diag(\lambda_1, \ldots, \lambda_q)$ with $\lambda_l \geq 0$ for $l = 1, \ldots, q$ and Γ is a lower triangular $q \times q$ matrix with the diagonal elements equal to 1 and $q(q-1)/2$ free off diagonal elements $\gamma = (\gamma_{21}, \gamma_{31}, \gamma_{32}, \ldots, \gamma_{q1}, \ldots, \gamma_{qq-1})$. It can be shown that λ_l is proportional to the standard deviation of the l^{th} random effect, so $\lambda_l = 0$ is equivalent to dropping the random effect from the model. Under this decomposition, model (2) can be rewritten in the following form:

$$logit\{pr(Y_{ij} = 1|X_{ij}^{(1)}, Z_{ij}, b_i, \beta, \lambda, \gamma)\} = X_{ij}^{(1)T}\beta + Z_{ij}^T\Lambda\Gamma b_i, \tag{3}$$

where $b_i \sim N_q(0, I)$ with I being the identity matrix. We now augment this model with a missing data mechanism to correct for bias and propose similar variable selection for this component of the model.

3.2 Model for missing data mechanism

The dropout hazard is modeled through logistic regression. We denote the response indicator $R_{ij} = I\{Y_{ij}$ is observed$\}$ for subject i at time t_j. That is, R_{ij} is equal to 1 if Y_{ij} is observed and equal to 0 if Y_{ij} is missing. Set $R_{i0} = 1$ for all i. Given that the binary outcome is observed at time t_{j-1}, the probability that it is observed at time t_j is

$$logit\{pr(R_{ij} = 1|R_{ij-1} = 1, Y_i, \eta)\} = h(\bar{Y}_{ij}; \eta),$$

where $h(\cdot)$ is assumed to be a known function, $Y_i = \{Y_{ij} : j = 1, \ldots, n_i\}$, $\bar{Y}_{ij} = (Y_{i1}, \ldots, Y_{ij})$, and η is an unknown (possibly vector-valued) parameter associated with $\bar{Y}_{ij}$. Here we assume that the probability of observing Y_{ij} given the subject has not dropped out at time t_{j-1} is only related to the responses up to time t_j, conditioning on all the observations of Y_i. Such non-future dependence assumptions for the missing data mechanism have been considered by Fitzmaurice *et al*. (1995), Baker (1995), Albert (2000), and Kenward *et al*. (2003). Various functional forms of $h(\bar{Y}_{ij}; \eta)$ have been proposed to characterize the relationship between the longitudinal outcome and the dropout process.

An advantage of using a selection model factorization is that there is a direct correspondence between the functional form of $h(\bar{Y}_{ij}; \eta)$ and missing data mechanisms as defined in Little and Rubin (2002): (1) when $h(\bar{Y}_{ij}; \eta) \neq h(\bar{Y}_{i,j-1}; \eta)$, which indicates that the dropout probability is related to the current (possibly missing) response Y_{ij}, the data are missing not at random (MNAR); (2) when $h(\bar{Y}_{ij}; \eta) = h(\bar{Y}_{i,j-1}; \eta)$, the data are missing at random (MAR) because the dropout probability is a function of observed values, not the missing components. Note that it would be a simple extension to allow for different types of dropout by having a separate missing data mechanism for each dropout type. This model can be further generalized to a setting in which a $1 \times \kappa$ vector of baseline covariates $X_i^{(2)}$ may affect the dropout probability:

$$logit\{pr(R_{ij} = 1 | R_{ij-1} = 1, Y_i, X_i^{(2)}, \eta, \alpha)\} \quad = \quad h(\bar{Y}_{ij}, X_i^{(2)}; \eta, \alpha), \tag{4}$$

where $\alpha = (\alpha_1, \ldots, \alpha_\kappa)$ are unknown parameters associated with $X_i^{(2)}$. An example of the functional form of h is given in equation (6) in Section 4.

3.3 Priors

Next we describe a variable selection procedure which involves zero-inflated mixture priors with a point mass at zero. We use this procedure to select fixed and random effects in (3) as well as the parameters η and α in model (4) for dropout hazard. To facilitate posterior variable selection via Gibbs sampling, conditional linearity of the parameters in the logistic models is achieved by decompositing Σ and approximating the logistic density using the t-distribution.

An advantage of the decomposition (3) is that we can obtain conditional linearity of λ given γ and b_i in the logit scale of p_{ij}, and similarly, the conditional linearity of γ given λ and b_i. We adopt mixture priors for β and λ as proposed by Cai and Dunson (2006) to allow a subset of fixed and random effects to be omitted from the model. To be specific, the prior for β is $\prod_{l=1}^{p} f_{ZIN}(\beta_l; \pi_{0l}^{\beta}, \mu_{0l}^{\beta}, \sigma_{0l}^{\beta 2})$, where $f_{ZIN}(\beta_l; \pi_{0l}^{\beta}, \mu_{0l}^{\beta}, \sigma_{0l}^{\beta 2})$ is a zero-inflated normal density defined as follows:

$$f_{ZIN}(\beta_l; \pi_{0l}^{\beta}, \mu_{0l}^{\beta}, \sigma_{0l}^{\beta 2}) = \pi_{0l}^{\beta} I(\beta_l = 0) + (1 - \pi_{0l}^{\beta}) I(\beta_l \neq 0) f_N(\beta_l; \mu_{0l}^{\beta}, \sigma_{0l}^{\beta 2}),$$

in which $\pi_{0l}^{\beta} \in [0, 1]$ and $f_N(\beta_l; \mu_{0l}^{\beta}, \sigma_{0l}^{\beta 2})$ is the normal density function with mean μ_{0l}^{β} and variance $\sigma_{0l}^{\beta 2}$. The prior probability that the l^{th} predictor is not selected into the model is thus π_{0l}^{β}. Because $\lambda_l \geq 0$, $l = 1, \ldots, q$, a zero-inflated positive normal density $f_{ZIN^+}(\lambda_l; \pi_{0l}^{\lambda}, \mu_{0l}^{\lambda}, s_{0l}^2)$ is chosen as the prior for λ_l. That is,

$$f_{ZIN^+}(\lambda_l; \pi_{0l}^{\lambda}, \mu_{0l}^{\lambda}, s_{0l}^2) = \pi_{0l}^{\lambda} I(\lambda_l = 0) + (1 - \pi_{0l}^{\lambda}) I(\lambda_l > 0) \frac{f_N(\lambda_l; \mu_{0l}^{\lambda}, s_{0l}^2)}{\Phi(0; -\mu_{0l}^{\lambda}, s_{0l}^2)},$$

where $\pi_{0l}^{\lambda} \in [0, 1]$ and Φ is the normal cumulative density function. The prior for λ is assumed to be $p(\lambda) = \prod_{l=1}^{q} f_{ZIN^+}(\lambda_l; \pi_{0l}^{\lambda}, \mu_{0l}^{\lambda}, s_{0l}^2)$. Similar to π_{0l}^{β}, π_{0l}^{λ} is the prior probability that the l^{th} random effect is excluded. The variance-covariance parameters have the joint prior $p(\lambda, \gamma) = p(\gamma|\lambda)p(\lambda)$, and we choose $p(\gamma|\lambda) = f_N(\gamma; \gamma_0, C_0) I(\gamma \in H_\lambda)$, where H_λ sets the value of γ_{tk} at zero if $\lambda_t = 0$ or $\lambda_k = 0$. We use a similar zero-inflated normal density $f_{ZIN}(\eta_l; \pi_{0l}^{\eta}, \mu_{0l}^{\eta}, \sigma_{0l}^{\eta 2})$ to facilitate selection of parameters pertaining to missing data mechanisms. If one is also interested in selecting baseline covariates in model (4), a zero-inflated normal prior $f_{ZIN}(\alpha_l; \pi_{0l}^{\alpha}, \mu_{0l}^{\alpha}, \sigma_{0l}^{\alpha 2})$, $l = 1, \ldots, \kappa$, can be specified for α. The impact of hyperparameters in these priors are explored via simulations (Section 4).

3.4 Posterior computation

The joint posterior distribution for $\Omega = (\beta, \lambda, \gamma, \eta, \alpha)$ takes a complex form primarily due to nonlinearity of the parameters in the logistic function. We approximate the logistic density by a t-density and use a data augmentation approach to obtain conditional linearity of Ω. These strategies help establish conditional conjugacy to simplify the Gibbs sampler. Similar approaches have been proposed by Albert and Chib (1993) and Kinney and Dunson (2007). The full conditional posterior distributions of the parameters and latent variables are given in Appendix B.

The posterior probability of selecting each of the fixed (β) and random effects (λ) as well as η and α can be estimated by calculating the proportion of non-zero posterior samples. Selection of the effects can proceed in two ways. At first, a threshold can be set for the posterior selection probability. If the probability is greater than that threshold, the corresponding effect (fixed or random) is selected. We recommend using 0.5 as a general rule of thumb. However, a more stringent selection criterion can be set if needed. Another approach is based on the probability of visiting a particular model (with a specific set of fixed and random effects) out of all possible candidate models and identifying the ones with highest probabilities. The probability can be estimated by calculating the proportion of that specific model in posterior samples.

4 Simulation studies

Via simulations, we examine variable selection in the presence of non-ignorable missing data using the modeling framework in Section 3 for more complex settings than those examined analytically in Section 2. We mimic a placebo controlled study where the research interest is to evaluate the between-group difference in the response variable. The longitudinal binary data Y_{ij} were generated from the following random intercept and slope model:

$$\begin{aligned} logit\{pr(Y_{ij}=1|x_{ij},u_i,\beta,\Sigma)\} &= \beta_0 x_{0i} + \beta_1 x_{1ij} + \beta_2 x_{2i} + \beta_3 x_{3ij} + \sum_{k=4}^{10} \beta_k x_{ki} \\ &\quad + u_{0i} x_{0i} + u_{1i} x_{1ij}, \end{aligned} \tag{5}$$

where $x_{0i} = 1$ is the intercept, $x_{1ij} = 0, 0.2, 0.4, \ldots$, up to 2.0 is the visit time, $x_{2i} \sim Bernoulli(0.5)$ is the indicator for the treatment group, $x_{3ij} = x_{1ij} \times x_{2i}$ is the interaction between group and time, and x_{4i} to x_{10i} are baseline covariates with a zero-mean multivariate normal distribution and AR(1) covariance structure (correlation $\rho = 0.5$). We set $\beta = (\beta_0, \ldots, \beta_{10})^T = (-1, 2, -1, -1.5, 0, 1, 0, 0, 0, 1, 0)^T$. The random effects $u_i = (u_{0i}, u_{1i}) \sim N(0, \Sigma)$ with $\Sigma = \Lambda\Gamma\Gamma^T\Lambda$, where $\Lambda = diag(0.5, 0.5)$ and $\gamma = (\gamma_{21}) = (0.1)$. Denote $x_{ij} = (x_{0i}, x_{1ij}, x_{2i}, x_{3ij}, x_{4i}, \ldots, x_{10i})$.

The dropout time was simulated using the logistic model

$$\begin{aligned} &logit\{pr(R_{ij}=1|R_{ij-1}=1, Y_{ij-1}, Y_{ij}, x_{ij}, \eta, \alpha)\} \\ &= \eta_0 + \eta_1 Y_{ij-1} + \eta_2 Y_{ij} + \eta_3 Y_{ij} x_{1ij} + \eta_4 Y_{ij} x_{1ij} x_{2i} + \alpha_1 x_{2i} + \sum_{k=2}^{8} \alpha_k x_{(k+2)i}, \end{aligned} \tag{6}$$

for $j \geq 2$; that is, all Y_{ij}'s are observed in the first two visits ($j = 0, 1$) for all i. The parameter $\alpha = (\alpha_1, \ldots, \alpha_8)^T = (2, 0, -0.5, 0, 0, -1, 0, 0)^T$. Missing data were generated in the following two scenarios and we simulated 100 data sets in each scenario:

(a) $(\eta_1, \ldots, \eta_4)$ = (0, 0, -1.5, 0.1), i.e., the missing data are non-ignorable. For subjects who respond (Y_{ij} = 1), the dropout rate increases with time, and the placebo group has a higher increase in the dropout rate than the treated group. On average there are about 23% and 55% missing data in the treated group and the placebo group, respectively;

(b) $(\eta_1, \ldots, \eta_4)$ = (-0.4, 0, 0, 0), that is, the missing data are ignorable. The treated and placebo groups have around 19% and 50% missing data, respectively.

Kinney and Dunson (2007) suggested that a relatively small prior variance of γ will facilitate stable estimation while containing the flexibility of allowing posterior adjustment by the data. Therefore, the hyperparameters in the prior for γ were set to be $\gamma_0 = 0$, $C_0 = 1$ (it is a scalar in the simulations), with the constraint of γ possibly being degenerate at zero due to λ. We chose a vague prior for η_0, a normal distribution with mean equal to 0 and variance equal to 1000. To explore the impact of choices of hyperparameters, we studied a range of priors in the form $f_{ZIN}(\pi_0^{\beta}, 0, \sigma_0^{\beta 2})$, $f_{ZIN^+}(\pi_0^{\lambda}, 0, s_0^2)$, $f_{ZIN}(\pi_0^{\eta}, 0, \sigma_0^{\eta 2})$, and $f_{ZIN}(\pi_0^{\alpha}, 0, \sigma_0^{\alpha 2})$,for β, λ, η (not including η_0), and α, respectively, with π_0^{β}, π_0^{λ}, π_0^{η}, and π_0^{α} = 0.5 and 0.8, $\sigma_0^{\beta 2}$, s_0^2, $\sigma_0^{\eta 2}$, and $\sigma_0^{\alpha 2}$ = 5, 10, and 100.

Each simulated data set was analyzed using either the joint model (5) and (6), or model (5) alone; the latter is referred to as the "ignorable analysis". We implemented the Gibbs sampling algorithm described in Appendix B. For each of the data sets, we first ran a burn-in of 5000 iterations, which was determined by the potential scale reduction factor (Gelman and Rubin, 1992) using three parallel chains with overdispersed starting values to ensure convergence of the Gibbs sampler. We generated 15,000 iterations after the burn-in and then thinned the chain by a factor of 6 to compute posterior means, standard deviations, and selection probabilities.

Tables 1A and 1B summarize the simulation results for scenario (a) with $\sigma_0^{\beta 2}$ = s_0^2 = $\sigma_{0l}^{\eta 2}$ = $\sigma_{0l}^{\alpha 2}$ = 10. We compare posterior means, standard deviations, and selection probabilities (SP) of the fixed and random effects from the joint analysis to those from the ignorable analysis. In reporting posterior means, to save space in decimal places for very small values in the intervals (-0.001, 0) and (0, 0.001), -0.001 and 0.001 are reported, respectively. In the presence of non-ignorable missing data, the ignorable analysis produces biased estimates for β_1 and β_3. Because patients who respond (Y = 1, indicating worse disease conditions) at the occasion j are more likely to drop out of the study at that occasion, the time trend is biased downward in the ignorable analysis. In addition, since the treated group has a lower dropout rate than the control group, the between group difference in the time trend is underestimated. We observe a larger bias in the estimates for β_1 and β_3 as we increase π_0 to 0.8, and the posterior selection probabilities for β_1 and β_3 decrease along with the increase of π_0. The missing data also lead to a biased estimate for the random effect λ_1 and its posterior selection probability drops to 0.39 when $\pi_0 = 0.8$. In contrast, the joint analysis has a much smaller bias in the estimates of these parameters, although there is some extent of shrinkage in β_1 and β_3 as π_0 gets large. Nevertheless, the bias in these parameters is mostly corrected as compared to the ignorable analysis. Also note that the most prominent parameter, η_3, in the dropout out submodel, was correctly identified, with the posterior selection probability greater than 0.95.

The analyses were repeated at a much larger sample size $n = 1000$ (see Supplementary Tables S1(A)-(B)). In the ignorable analysis, the bias in the estimates for β_1, β_3 and λ_1 still exists. Because this is systematic bias introduced by non-ignorable missing data, it cannot be corrected by increasing the sample size. Despite the estimation bias, the posterior selection probabilities of these parameters are uniformly higher than those in Tables 1A and 1B. This is consistent with our discussion about the behavior of π^* as $n \to \infty$ in Section 2. In addition, for both the joint and ignorable analyses, we observe reduced impact of the prior exclusion probability π_0 compared to the results for $n = 300$.

We also examined the bias in estimation and selection due to non-ignorable missing data when the actual time trend in both groups were zero, i.e., $\beta_1 = \beta_3 = 0$, (Supplementary Tables S2(A)-(B)). Again there is a large bias in estimated β_1 and β_3 in the ignorable analysis, which results in a relatively high probability of selecting them even though the true values are zero. This pattern is consistent with our discussion about the asymptotic behavior of π^*. The joint analysis leads to almost unbiased estimates for these parameters and the posterior selection probabilities reduce to below 0.05 when $\pi_0 = 0.8$.

When the missing data are ignorable (Supplementary Tables S3(A)-(B)) we observe comparable results between the two methods, which is expected since under this scenario the ignorable analysis is valid. The impact of π_0 is not as strong as in scenario (a), which implies that under MAR we may need smaller samples to overcome the influence of priors. It is also noted that, under both scenarios (a) and (b), the

missing data mechanism does not affect the point estimates or posterior selection probabilities of the fixed effects $\beta_2, \beta_4, \ldots, \beta_7$ and α's.

The results to examine the impact of hyperparameters are reported in Table 2. To reduce space we only show the results for three fixed effects in β (β_1, β_2, and β_3), the two random effects (λ_1 and λ_2), the parameters in η that are associated with Y_{ij}, and one effect in α. Under a given prior exclusion probability π_0 (0.5 or 0.8), we do not observe significant impact of the prior variance on the posterior inference, although the posterior inclusion probabilities tend to decrease with increasing prior variance. Since zero-inflated priors are spike and slab-type priors with a spike at zero, this is in line with results for such priors in Mitchell and Beauchamp (1988) and Dey *et al*. (2008). Consistent with Tables 1, S1 and S2, there is shrinkage in β_1, β_3 and λ for a higher π_0, associated with a small inflation in the posterior variability and a decrease in the posterior selection probability. In all the cases, the method is able to identify the right component in η that contributes most to the missing data mechanism.

We did further simulations to study the impact of model misspecification in the dropout process. In particular, data were simulated from a missingness mechanism similar to Equation (6), but with one additional term, -0.5$Y_{ij} \times Y_{ij-1}$. This model indicates that patients who respond (Y = 1) at two consecutive occasions have a higher probability to dropout. The simulated data were then analyzed using model (6), that is, the product term between Y_{ij} and Y_{ij-1} was ignored. Table 3 reports the variable selection results for the fixed and random effects at the longitudinal endpoint. Selection of baseline covariates in the dropout model is shown as well. Estimation and selection of β and α seem robust to this model misspecification, but the variance of the random intercept is underestimated, with the posterior selection probability dropping below 50% when π_0 = 0.8. It suggests that misspecification of the dropout model could lead to biased estimates at the longitudinal endpoint and thus misleading variable selection results.

5 Example

We illustrate the proposed variable selection procedure using data from a double-blinded, randomized clinical trial of intravenous recombinant tissue plasminogen activator (rt-PA) in patients with acute ischemic stroke (National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group, 1995). A total of 624 patients were enrolled and randomized to receive either intravenous recombinant t-PA or placebo. There were 312 patients in each treatment arm. Repeated measurements of Barthel Index, a scale in the rage of 0 to 100 for performance in daily living, were recorded at 7–10 days, 3 months, 6 months, and 12 months post stroke onset. Hacke *et al*. (1998) suggested an unfavorable outcome be defined as a score of less than 95 on Barthel Index. We use this dichotomized measure as the outcome of interest in this example. Out of the 624 patients, 25 dropped out before 12 months (14 in rt-PA group and 11 in the placebo) and 168 died (78 in rt-PA group and 90 in the placebo group, including those died after 12 months). The average number of measurements per patient was 3.2, and 30% data were missing at 12 months.

Table 4 summarizes the frequency of patients with unfavorable Barthel Index at each of the follow-up times. The placebo group had a higher rate of missing data during follow-up, and the observed data suggest that the rt-PA group tended to have fewer patients with unfavorable Barthel Index than the placebo group.

We analyzed the data using both the joint model as outlined in Section 3 and the ignorable analysis in which we ignored the missing data mechanism. The prior variance was set to 10 for all the fixed and random effects and $\pi_0 = 0.8$. The results were summarized using 60,000 iterations thinned by a factor of 10 after 40,000 burn-in iterations. The 40,000 burn-in iterations was determined by the potential scale reduction factor after we ran three parallel chains with over-dispersed starting values.

The results from the joint and ignorable analyses are given in Table 5. We adopted an unstructured time trend; three binary variables time3, time6, and time12 were generated to indicate the measurement taken at 3 months, 6 months or 12 months post stroke onset. We considered a random intercept and fit three models, in each of which, a second random effect was assumed for one of the three time indicator variables. All three models suggested that the time effects showed minimal between-subject variations. For illustration

purposes, the model that assumed a random effect of time3 is shown. The following baseline covariates (mostly are binary) were considered: treatment group, age, gender (1 = male), smoking status (1 = yes), drinking (1 = yes), abnormal baseline CT, and history of diabetes, hypertension, and angina at baseline. The dropout process model with baseline covariates (corresponding to α) was initially fit, but none of the baseline variables showed a significant relationship with dropouts. Therefore, we omitted these variables and only included the terms associated with η_0, η_1, and η_2.

The results from the joint analysis show that, conditional on the random effects, the probability of obtaining an unfavorable Barthel Index score decreased over time in the placebo group. The odds ratios were 0.05 (the 95% CI (0.02, 0.09)), 0.05 (0.02, 0.09), and 0.02 (0.007, 0.05), comparing the measurements at 3, 6, and 12 months post stroke onset with those at 7-10 days, respectively. The two groups showed similar trend over time (i.e., the interaction between group and time was close to zero). On average the odds ratio of an unfavorable Barthel index was 0.17 (the 95% CI (0.05, 0.53)) comparing the treated group versus the control. Among the baseline covariates, only history of hypertension showed higher than 50% chance of being selected. There was strong evidence for a large variation in the intercept, which has a 100% chance of being selected. However, the variation of time3 effect was relatively small, with a selection probability of 0.20. In the dropout sub-model, the posterior selection probability for η_1 was 1.00, but was only 0.57 for η_2. Assuming we have modelled the dropout process and full data response reasonably well, this result indicates that the missing data mechanism is likely MAR. This is also consistent with the observation that the results from the ignorable analysis were very similar to the joint analysis. The estimate of η_1 was negative, indicating that patients with unfavorable Barthel Index had a higher probability of missing the next visit either due to dropout or death.

6 Discussion

We have examined variable selection in the presence of non-ignorable missing data. Our approach fills in the gap in existing literature with the attempt to elucidate bias in variable selection associated with non-ignorable missing data. We propose to resolve this problem by jointly modeling the missing data mechanism. A Bayesian approach is implemented for fixed and random effects selection as well as selection of missing data mechanisms. To facilitate the posterior variable selection, a zero-inflated mixture prior proposed by Dunson and colleagues is employed. Conditional linearity is obtained for the parameters in the logistic models by approximation of the logistic density using the t-distribution. Although the proposed method is illustrated by a randomized clinical trial, we need to point out that in general variable selection is more frequently applied in early stages of drug development as an exploratory tool for hypothesis generation.

We emphasize the importance of joint analysis whenever the MAR assumption is questionable. In longitudinal studies, terminating events such as death and dropout could lead to non-ignorable missing data if these events are *informative*, i.e. the event incidence is related to the longitudinal outcome of interest. As explained in Section 3.2, the proposed model includes the MAR missingness as a special case, and thus can be used as a tool to investigate the possible underlying missingness mechanism. Since the MNAR assumption is untestable, it is always helpful to perform the joint model analysis and compare it with the ignorable analysis whenever the ignorable missingess is questionable. If the results from the two approaches are similar, then it is very likely the missing data are ignorable assuming the modelling assumptions are correct. On the other hand, if the two analyses produce very distinct results, then one should question the validity of the ignorable analysis. In this case the joint analysis is recommended.

The simulation results indicate that the estimation and selection at the longitudinal endpoint could be influenced by model misspecification for the dropout process. Since it is usually not clear what functional form of the missing data mechanism would be for a given study, it is recommended to fit a range of plausible models and evaluate sensitivity of the results to model specifications among those that provide similar fit to the observed data (Daniels and Hogan, 2008).

We use zero-inflated mixture priors for variable selection because they have the following advantages: (1) By setting a positive probability at zero, the prior allows effective exclusion of fixed and random effects from the model; (2) It is a conjugate prior for the proposed model because under reparameterization and approximation the model is conditionally linear in the parameters; (3) It is a proper prior to avoid possible impropriety of the posterior; (4) It avoids the need to compute model selection criteria such as DIC (Spiegelhalter *et al*., 2002); this is convenient for model selection as the observed data likelihood is typically difficult to compute (Daniels and Hogan, 2008).

The variable selection procedure in conjunction with joint analysis developed in this article can be extended to continuous longitudinal outcomes and the results here are *not specific* to the joint model given in Section 3. The whole procedure can be thought of as Bayesian model averaging and as such, accurately reflects uncertainty about variable selection. An alternative approach to our fixed prior selection probabilities would be to make these quantities random (Ley and Steel, 2009). Our method assumes monotonic missing data patterns and regular observation times for the longitudinal outcome. One future direction is to study variable selection for longitudinal data with irregular observation times and/or intermittent missing values. Although in practice binning time axis to convert irregular observation times to regular ones is not so rare, but it may not always be feasible. We expect similar results in these settings as well.

Joint models are generally classified as selection models and pattern-mixture models; our model is a parametric selection model. The hazard of dropout is modeled by logistic regression under a non-future dependent mechanism (Kenward *et al*., 2003). This approach has the advantages that it allows identification of all parameters in the model and that the interpretation of the coefficients in the missing data mechanism has a direct correspondence with the missing data mechanism classifications defined by Little and Rubin (2002). However, we need to point out that, in the presence of non-ignorable missing data, the inference on longitudinal measurements is dependent on the posited functional form of the missing data mechanism as well as the parametric form for the full data response. In addition, parametric selection models do not allow sensitivity analysis to investigate the impact of modeling assumptions for missing data. An alternative formulation to the selection model proposed here would be a mixture model; such a formulation would more easily accommodate sensitivity analysis. We would expect similar biases in selection probabilities and coefficients in this formulation. For a comprehensive discussion of sensitivity analysis and related issues, see Daniels and Hogan (2008).

Acknowledgements Michael Daniels work was supported in part by NIH grant CA85295, Gang Li's work was supported in part by NIH grants CA016042 and P01AT003960, and Robert Elashoff's work was supported in part by NIH grant H0089901-02.

Conflict of Interest

The authors have declared no conflict of interest.

A Appendix A

Theorem A.1 *$\hat{\theta}$ is an inconsistent estimator of θ when $\eta_2 \neq 0$*

Proof. The likelihood for θ based on the observed data, which ignores the missing data mechanism, is defined as

$$\begin{aligned} & L(\theta|Y_{obs}) = \int f(Y_{obs}, Y_{mis}|\theta) dY_{mis} \\ = \;& p_{00}^{n_{00}} p_{10}^{n_{10}} p_{01}^{n_{01}} (1 - p_{00} - p_{01} - p_{10})^{n_{11}} (p_{00} + p_{01})^{n_{0m}} (1 - p_{00} - p_{01})^{n_{1m}}, \end{aligned} \tag{7}$$

where n_{00} is the number of individuals with ($Y_{1i} = 0$, $Y_{2i} = 0$), and n_{10} is the number of individuals with ($Y_{1i} = 1$, $Y_{2i} = 0$); n_{01} and n_{11} are defined similarly. We use n_{0m} (n_{1m}) to denote the number of

individuals with Y_{2i} missing and $Y_{1i} = 0$ ($Y_{1i} = 1$). The maximum likelihood estimator of θ derived from (7) is

$$\begin{aligned}
\hat{p}_{00} &= \frac{n_{00} + n_{01} + n_{0m}}{n} \times \frac{n_{00}}{n_{00} + n_{01}}, \\
\hat{p}_{10} &= \frac{n_{10} + n_{11} + n_{1m}}{n} \times \frac{n_{10}}{n_{10} + n_{11}}, \\
\hat{p}_{01} &= \frac{n_{00} + n_{01} + n_{0m}}{n} \times \frac{n_{01}}{n_{00} + n_{01}}.
\end{aligned}$$

We show that these are biased estimates of θ. It is easy to see that $\frac{n_{00}+n_{01}+n_{0m}}{n} \to pr(Y_1 = 0) = \tilde{p}_{00}+\tilde{p}_{01}$ and $\frac{n_{10}+n_{11}+n_{1m}}{n} \to pr(Y_1 = 1) = \tilde{p}_{10} + \tilde{p}_{11}$, regardless of the missing data mechanism. We know that from (1),

$$\begin{aligned}
&\frac{n_{00}}{n_{00} + n_{01}} \to pr(Y_1 = 0, Y_2 = 0 | Y_1 = 0, R_2 = 1) \\
&= \frac{pr(Y_1 = 0, Y_2 = 0, R_2 = 1)}{pr(Y_1 = 0, R_2 = 1)} \\
&= \frac{pr(R_2 = 1|Y_1 = 0, Y_2 = 0)pr(Y_1 = 0, Y_2 = 0)}{pr(R_2 = 1|Y_1 = 0, Y_2 = 0)pr(Y_1 = 0, Y_2 = 0) + pr(R_2 = 1|Y_1 = 0, Y_2 = 1)pr(Y_1 = 0, Y_2 = 1)} \\
&= \frac{\frac{exp(\eta_0)}{1+exp(\eta_0)}\tilde{p}_{00}}{\frac{exp(\eta_0)}{1+exp(\eta_0)}\tilde{p}_{00} + \frac{exp(\eta_0+\eta_2)}{1+exp(\eta_0+\eta_2)}\tilde{p}_{01}} \\
&\neq \frac{\tilde{p}_{00}}{\tilde{p}_{01} + \tilde{p}_{00}},
\end{aligned}$$

if $\eta_2 \neq 0$. Therefore, $\hat{p}_{00}$ is not a consistent estimator of p_{00} if $\eta_2 \neq 0$, i.e., if the missing data are MNAR. Similarly, when $\eta_2 \neq 0$, it can be shown that $\hat{p}_{01}$ and $\hat{p}_{10}$ are not consistent estimators for p_{01} and p_{10}, respectively, because

$$\frac{n_{10}}{n_{10} + n_{11}} \to \frac{\frac{exp(\eta_0+\eta_1)}{1+exp(\eta_0+\eta_1)}\tilde{p}_{10}}{\frac{exp(\eta_0+\eta_1)}{1+exp(\eta_0+\eta_1)}\tilde{p}_{10} + \frac{exp(\eta_0+\eta_1+\eta_2)}{1+exp(\eta_0+\eta_1+\eta_2)}\tilde{p}_{11}} \neq \frac{\tilde{p}_{10}}{\tilde{p}_{11} + \tilde{p}_{10}},$$

and

$$\frac{n_{01}}{n_{00} + n_{01}} \to \frac{\frac{exp(\eta_0+\eta_2)}{1+exp(\eta_0+\eta_2)}\tilde{p}_{01}}{\frac{exp(\eta_0)}{1+exp(\eta_0)}\tilde{p}_{00} + \frac{exp(\eta_0+\eta_2)}{1+exp(\eta_0+\eta_2)}\tilde{p}_{01}} \neq \frac{\tilde{p}_{01}}{\tilde{p}_{00} + \tilde{p}_{01}}.$$

□

Suppose asymptotically $\hat{\theta}$ converges to $\theta^* = (p^*_{00}, p^*_{10}, p^*_{01})$. We have shown that $\theta^* \neq \tilde{\theta}$ if $\eta_2 \neq 0$. However, it still holds that $p^*_{00} + p^*_{01} = \tilde{p}_{00} + \tilde{p}_{01}$ and $p^*_{10} + p^*_{11} = \tilde{p}_{10} + \tilde{p}_{11}$ even if the missing data are MNAR.

For the regression model that reparameterizes the joint distribution of Y_1 and Y_2, because $logit\{pr(Y_1 = 1|\beta)\} = \beta_0 = logit(\tilde{p}_{10} + \tilde{p}_{11}) = logit(p^*_{10} + p^*_{11})$, we have $\beta^*_0 = \tilde{\beta}_0$. However, it can be shown that $\beta^*_1 \neq \tilde{\beta}_1$ and $\beta^*_2 \neq \tilde{\beta}_2$ if $\eta_2 \neq 0$ since

$$\begin{aligned}
logit\{pr(Y_2 = 1|Y_1 = 0, \beta)\} &= \tilde{\beta}_0 + \tilde{\beta}_1 = logit\frac{\tilde{p}_{01}}{\tilde{p}_{01} + \tilde{p}_{00}} \\
&\neq logit\frac{p^*_{01}}{p^*_{01} + p^*_{00}} = \beta^*_0 + \beta^*_1,
\end{aligned}$$

and

$$\begin{aligned} logit\{pr(Y_2=1|Y_1=1,\beta)\} &= \tilde{\beta}_0+\tilde{\beta}_1+\tilde{\beta}_2 = logit\frac{\tilde{p}_{11}}{\tilde{p}_{11}+\tilde{p}_{10}} \\ &\neq logit\frac{p^*_{11}}{p^*_{11}+p^*_{10}} = \beta^*_0+\beta^*_1+\beta^*_2. \end{aligned}$$

Appendix B

We outline the steps in the Gibbs sampler to draw samples from the posterior distribution of Ω, assuming the dropout process model takes the form of (6). By introducing a latent variable W_{ij}, the logistic mixed model (3) can be rewritten as follows:

$$W_{ij} = X_{ij}^{(1)T}\beta + Z_{ij}^T\Lambda\Gamma b_i + \epsilon_{ij},$$

where ϵ_{ij}'s are $i.i.d.$ logistic random variables with density function

$$p(\epsilon_{ij}) = \frac{exp(-\epsilon_{ij})}{\{1+exp(-\epsilon_{ij})\}^2}.$$

The outcome $Y_{ij}=1$ if $W_{ij}>0$ and $Y_{ij}=0$ otherwise. We approximate the distribution of ϵ_{ij} by a zero mean t random variable, $t_\nu(0,\tilde{\sigma}^2)$, where the degrees of freedom $\nu=7.3$ and the scale parameter $\tilde{\sigma}^2=\pi^2(\nu-2)/3\nu$. The distribution is equivalent to a mixture of normals such that conditional on the scale parameter d_{ij}, $\varepsilon_{ij}\sim N(0,\tilde{\sigma}^2/d_{ij})$, and $d_{ij}\sim Gamma(\nu/2,2/\nu)$ whose density is proportional to $d_{ij}^{\nu/2-1}exp(-\nu d_{ij}/2)$. Kinney and Dunson (2007) showed that approximation of the logistic distribution by $t_\nu(0,\tilde{\sigma}^2)$ was nearly exact using importance weighting. Similarly, for the dropout process model, we define

$$Q_{ij} = \eta_0+\eta_1Y_{ij-1}+\eta_2Y_{ij}+\eta_3Y_{ij}t_{ij}+\eta_4Y_{ij}\tau_it_{ij}+X_i^{(2)}\alpha+\varepsilon'_{ij},$$

where τ_i is the group indicator, $\varepsilon'_{ij}\sim N(0,\tilde{\sigma}^2/a_{ij})$, and $a_{ij}\sim Gamma(\nu/2,2/\nu)$. The response indicator $R_{ij}=1$ if $Q_{ij}>0$ and $R_{ij}=0$ if $Q_{ij}\leq 0$.

Let Y_{obs} denote the observed components in Y and Y_{mis} the missing values. Assume each subject has at least $r-1$ observations, $r\geq 2$. Conditional on the augmented data (W,d,Q,a) and the unobserved data (Y_{mis},b), the joint posterior density of Ω is

$$\begin{aligned} & L(\Omega;Y_{obs},Y_{mis},b,R,W,d,Q,a) \\ = & \prod_{i=1}^{n}[\prod_{j=1}^{min(n_i+1,m)}[\{I(W_{ij}>0)I(Y_{ij}=1)+I(W_{ij}\leq 0)I(Y_{ij}=0)\} \\ & \times\sqrt{\frac{d_{ij}}{2\pi\tilde{\sigma}^2}}\exp\{-\frac{d_{ij}}{2\tilde{\sigma}^2}(W_{ij}-X_{ij}^{(1)T}\beta-Z_{ij}^T\Lambda\Gamma b_i)^2\}d_{ij}^{\nu/2-1}\exp(-\nu d_{ij}/2)] \\ & \prod_{j=r}^{min(n_i+1,m)}[\{I(Q_{ij}>0)I(R_{ij}=1)+I(Q_{ij}\leq 0)I(R_{ij}=0)\} \\ & \times\sqrt{\frac{a_{ij}}{2\pi\tilde{\sigma}^2}}\exp\{-\frac{a_{ij}}{2\tilde{\sigma}^2}(Q_{ij}-\eta_0-\eta_1Y_{ij-1}-\eta_2Y_{ij}-\eta_3Y_{ij}t_{ij}-\eta_4Y_{ij}\tau_it_{ij}- \\ & X_i^{(2)}\alpha)^2\}a_{ij}^{\nu/2-1}\exp(-\nu a_{ij}/2)]\frac{1}{\sqrt{(2\pi)^q}}\exp(-\frac{1}{2}b_i^Tb_i)]p(\beta)p(\lambda,\gamma)p(\eta)p(\alpha). \end{aligned}$$

We assume that $X_{ij}^{(1)}$ and Z_{ij} are observable even if subject i drops out at occasion j, i.e., $R_{ij}=0$. Posterior computation relies on a Gibbs sampler which iteratively samples from the full conditional distributions

of the parameters β, λ, γ, η and α, the latent variables b, W, d, Q, and a, and the missing data Y_{mis}. Let $N(\mu, \sigma^2)$ stand for normal distribution, and $ZI\text{-}N(\pi, \mu, \sigma^2)$ and $ZI\text{-}N^+(\pi, \mu, \sigma^2)$ stand for zero-inflated normal distribution and zero-inflated positive normal distribution, respectively, as defined in Section 3.3. The full conditional distributions are given as follows.

- The full conditional posterior for β_l is $ZI\text{-}N(\pi_l^\beta, \mu_l^\beta, \sigma_l^{\beta 2})$, $l = 1, \ldots, p$, with

$$\begin{aligned}
\sigma_l^{\beta 2} &= \frac{1}{1/\sigma_{0l}^{\beta 2} + \sum_i \sum_j d_{ij} x_{ijl}^2/\tilde{\sigma}^2}, \\
\mu_l^\beta &= \sigma_l^{\beta 2}\{\frac{\mu_{0l}^\beta}{\sigma_{0l}^{\beta 2}} + \frac{1}{\tilde{\sigma}^2}\sum_i \sum_j d_{ij} x_{ijl}(W_{ij} - X_{ij(-l)}^{(1)T}\beta_{(-l)} - Z_{ij}^T \Lambda\Gamma b_i)\}, \\
\pi_l^\beta &= \frac{\pi_{0l}^\beta}{\pi_{0l}^\beta + (1-\pi_{0l}^\beta)\frac{\sigma_l^\beta}{\sigma_{0l}^\beta} exp(-\frac{\mu_{0l}^{\beta 2}}{2\sigma_{0l}^{\beta 2}} + \frac{\mu_l^{\beta 2}}{2\sigma_l^{\beta 2}})},
\end{aligned}$$

  where x_{ijl} is the l-th element of the vector $X_{ij}^{(1)}$, $X_{ij(-l)}^{(1)}$ is the vector $X_{ij}^{(1)}$ with the l-th element omitted, and $\beta_{(-l)}$ is defined similarly.

- Let $b_i^T = (b_{i1}, \ldots, b_{iq})$ and $Z_{ij}^T = (z_{ij1}, \ldots, z_{ijq})$. Define $t_{ijl} = z_{ijl}(b_{il} + \sum_{k=1}^{l-1} b_{ik}\gamma_{lk})$ for $l = 1, \ldots, q$. We have $W_{ij} = X_{ij}^{(1)T}\beta + \sum_{l=1}^q t_{ijl}\lambda_l + \epsilon_{ij}$. $\lambda_l \sim ZI\text{-}N^+(\pi_l^\lambda, \mu_l^\lambda, s_l^2)$, where

$$\begin{aligned}
s_l^2 &= \frac{1}{1/s_{0l}^2 + \sum_i \sum_j d_{ij} t_{ijl}^2/\tilde{\sigma}^2}, \\
\mu_l^\lambda &= s_l^2\{\frac{\mu_{0l}^\lambda}{s_{0l}^2} + \frac{1}{\tilde{\sigma}^2}\sum_i \sum_j d_{ij} t_{ijl}(W_{ij} - X_{ij}^{(1)T}\beta - t_{ij(-l)}^T \lambda_{(-l)})\}, \\
\pi_l^\lambda &= \frac{\pi_{0l}^\lambda}{\pi_{0l}^\lambda + (1-\pi_{0l}^\lambda)\frac{s_l(1-\Phi(-\mu_l^\lambda/s_l))}{s_{0l}(1-\Phi(-\mu_{0l}^\lambda/s_{0l}))} exp(-\frac{\mu_{0l}^{\lambda 2}}{2s_{0l}^2} + \frac{\mu_l^{\lambda 2}}{2s_l^2})},
\end{aligned}$$

  and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, $t_{ij(-l)}$ is the vector $t_{ij}^T = (t_{ij1}, \ldots, t_{ijq})$ with the l-th element omitted, and $\lambda_{(-l)}$ is λ with the l-th element omitted.

- Let $g_{ij}^T = (b_{il}\lambda_k Z_{ijk}; l = 1, \ldots, q-1, k = l+1, \ldots, q)$, so $W_{ij} - X_{ij}^{(1)T}\beta = g_{ij}^T\gamma + \epsilon_{ij}$. The full posterior conditional distribution for γ is given by $N(\hat{\gamma}, \hat{C})I(\gamma \in H_\lambda)$, where

$$\begin{aligned}
\hat{C} &= (\sum_i \sum_j d_{ij} g_{ij} g_{ij}^T/\tilde{\sigma}^2 + C_0^{-1})^{-1}, \\
\hat{\gamma} &= \hat{C}\{\sum_i \sum_j d_{ij} g_{ij}(W_{ij} - X_{ij}^{(1)T}\beta)/\tilde{\sigma}^2 + C_0^{-1}\gamma_0\}.
\end{aligned}$$

- $W_{ij} \sim N(X_{ij}^{(1)T}\beta + Z_{ij}^T\Lambda\Gamma b_i, \tilde{\sigma}^2/d_{ij})$ truncated at the left by zero if $Y_{ij} = 1$ and at the right by zero if $Y_{ij} = 0$.

- $d_{ij} \sim Gamma(\frac{\nu+1}{2}, \frac{2}{\nu + \frac{1}{\tilde{\sigma}^2}(W_{ij} - X_{ij}^{(1)T}\beta - Z_{ij}^T\Lambda\Gamma b_i)^2})$.

- The full conditional posterior for η_l is $ZI\text{-}N(\pi_l^\eta, \mu_l^\eta, \sigma_l^{\eta 2})$, $l = 1, \ldots, 4$, with

$$\begin{aligned}
\sigma_l^{\eta 2} &= \frac{1}{1/\sigma_{0l}^{\eta 2} + \sum_i \sum_j a_{ij} t_{ijl}^2/\tilde{\sigma}^2}, \\
\mu_l^\eta &= \sigma_l^{\eta 2}\{\frac{\mu_{0l}^\eta}{\sigma_{0l}^{\eta 2}} + \frac{1}{\tilde{\sigma}^2}\sum_i \sum_j a_{ij} t_{ijl}(Q_{ij} - \eta_0 - \sum_{u \neq l} \eta_u t_{iju} - X_i^{(2)}\alpha)\}, \\
\pi_l^\eta &= \frac{\pi_{0l}^\eta}{\pi_{0l}^\eta + (1-\pi_{0l}^\eta)\frac{\sigma_l^\eta}{\sigma_{0l}^\eta} exp(-\frac{\mu_{0l}^{\eta 2}}{2\sigma_{0l}^{\eta 2}} + \frac{\mu_l^{\eta 2}}{2\sigma_l^{\eta 2}})},
\end{aligned}$$

  for $j = r, \ldots, \min(n_i + 1, m)$, where $t_{ij1} = Y_{ij-1}$, $t_{ij2} = Y_{ij}$, $t_{ij3} = Y_{ij}t_{ij}$, and $t_{ij4} = Y_{ij}\tau_i t_{ij}$.

- The full conditional posterior for α_l is $ZI\text{-}N(\pi_l^\alpha, \mu_l^\alpha, \sigma_l^{\alpha 2})$, $l = 1, \ldots, \kappa$, with

$$\begin{aligned}
\sigma_l^{\alpha 2} &= \frac{1}{1/\sigma_{0l}^{\alpha 2} + \sum_i \sum_j a_{ij} x_{il}^{(2)2}/\tilde{\sigma}^2}, \\
\mu_l^\alpha &= \sigma_l^{\alpha 2}\{\frac{\mu_{0l}^\alpha}{\sigma_{0l}^{\alpha 2}} + \frac{1}{\tilde{\sigma}^2}\sum_i \sum_j a_{ij} x_{il}^{(2)}(Q_{ij} - \eta_0 - \eta_1 Y_{ij-1} - \eta_2 Y_{ij} - \eta_3 Y_{ij}t_{ij} - \\
&\quad \eta_4 Y_{ij}\tau_i t_{ij} - X_{i(-l)}^{(2)}\alpha_{(-l)})\}, \\
\pi_l^\alpha &= \frac{\pi_{0l}^\alpha}{\pi_{0l}^\alpha + (1-\pi_{0l}^\alpha)\frac{\sigma_l^\alpha}{\sigma_{0l}^\alpha} exp(-\frac{\mu_{0l}^{\alpha 2}}{2\sigma_{0l}^{\alpha 2}} + \frac{\mu_l^{\alpha 2}}{2\sigma_l^{\alpha 2}})},
\end{aligned}$$

  for $j = r, \ldots, \min(n_i + 1, m)$, where $x_{il}^{(2)}$ is the l-th element of the vector $X_i^{(2)}$, $X_{i(-l)}^{(2)}$ is the vector $X_i^{(2)}$ with the l-th element omitted, and $\alpha_{(-l)}$ is α without the l-th element.

- $Q_{ij} \sim N(\eta_0 + \eta_1 Y_{ij-1} + \eta_2 Y_{ij} + \eta_3 Y_{ij}t_{ij} + \eta_4 Y_{ij}\tau_i t_{ij} + X_i^{(2)}\alpha, \tilde{\sigma}^2/a_{ij})$ truncated at the left by zero if $R_{ij} = 1$ and at the right by zero if $R_{ij} = 0$.

- $a_{ij} \sim Gamma(\frac{\nu+1}{2}, \frac{2}{\nu + \frac{1}{\tilde{\sigma}^2}(Q_{ij} - \eta_0 - \eta_1 Y_{ij-1} - \eta_2 Y_{ij} - \eta_3 Y_{ij}t_{ij} - \eta_4 Y_{ij}\tau_i t_{ij} - X_i^{(2)}\alpha)^2})$.

- $b_i \sim N(\hat{b}_i, \hat{\Sigma}_b)$, where

$$\begin{aligned}
\hat{\Sigma}_b &= (\sum_j d_{ij} u_{ij} u_{ij}^T/\tilde{\sigma}^2 + I)^{-1}, \\
\hat{b}_i &= \hat{\Sigma}_b \sum_j d_{ij} u_{ij}(W_{ij} - X_{ij}^{(1)}\beta)/\tilde{\sigma}^2,
\end{aligned}$$

  and $u_{ij}^T = Z_{ij}^T \Lambda\Gamma$.

- If $n_i < m$, the missing observation $Y_{i(n_i+1)}$ for subject i is $Bernoulli(\frac{p_{1(n_i+1)}}{p_{0(n_i+1)} + p_{1(n_i+1)}})$, where

$$\begin{aligned}
p_{1(n_i+1)} &= \frac{exp(X_{i(n_i+1)}^{(1)}\beta + Z_{i(n_i+1)}^T \Lambda\Gamma b_i)}{1 + exp(X_{i(n_i+1)}^{(1)}\beta + Z_{i(n_i+1)}^T \Lambda\Gamma b_i)} \\
&\quad \times \frac{1}{1 + exp(\eta_0 + \eta_1 Y_{in_i} + \eta_2 + \eta_3 t_{i(n_i+1)} + \eta_4 \tau_i t_{i(n_i+1)} + X_i^{(2)}\alpha)}, \\
p_{0(n_i+1)} &= \frac{1}{\{1 + exp(X_{i(n_i+1)}^{(1)}\beta + Z_{i(n_i+1)}^T \Lambda\Gamma b_i)\}\{1 + exp(\eta_0 + \eta_1 Y_{in_i} + X_i^{(2)}\alpha)\}}.
\end{aligned}$$

The algorithm was implemented in C. The program is available upon request.

References

[1] Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669–679.

[2] Albert, P. S. (2000). A transitional model for longitudinal binary data subject to nonignorable missing data. *Biometrics* **56**, 602–608.

[3] Baker, S. G. (1995). Marginal regression for repeated binary data with outcome subject to non-ignorable nonresponse. *Biometrics* **51**, 1042–1052.

[4] Cai, B. and Dunson, D. B. (2006). Bayesian covariance selection in generalized linear mixed models. *Biometrics* **62**, 446–457.

[5] Chen, W., Ghosh, D., Raghunathan, T. E. and Sargent, D. J. (2009). Bayesian variable selection with joint modeling of categorical and survival outcomes: an application to individualizing chemotherapy treatment in advanced colorectal cancer. *Biometrics* **65**, 1030–1040.

[6] Chen, Z. and Dunson, D. B. (2003). Random effects selection in linear mixed models. *Biometrics* **59**, 762–769.

[7] Daniels, M. J. and Hogan, J. W. (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Chapman & Hall/CRC.

[8] Dey, T., Ishwaran, H. and Rao, S. J. (2008). An in-depth look at highest posterior model selection. *Econometric Theory* **24**, 377–403.

[9] Diggle, P. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis (with discussion). *Applied Statistics* **43**, 49–93.

[10] Elashoff, R. M., Li, G. and Li, N. (2008). A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics* **64**, 762–771.

[11] Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association* **99**, 710–723.

[12] Fitzmaurice, G. M., Molengerghs, G. and Lipsitz, S. R. (1995). Regression models for longitudinal binary responses with informative drop-outs. *Journal of the Royal Statistical Society, Series B* **57**, 691–704.

[13] Follmann, D. and Wu, M. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics* **51**, 151–168.

[14] Fu, W. J. (2003). Penalized estimating equations. *Biometrics* **59**, 126–132.

[15] Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457–511.

[16] George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **88**, 881–889.

[17] Hacke, W., Bluhmki, E., Steiner, T., Tatlisumak, T., Mahagne, M.-H., Sacchetti, M.-L. and Meier, D., for the ECASS Study Group (1998). Dichotomized efficacy end points and global end-point analysis applied to the ECASS intention-to-treat data set. *Stroke* **29**, 2073–2075.

[18] Henderson, R., Diggle, P. and Dobson, A. (2000). Joint modeling of longitudinal measurements and event time data. *Biostatistics* **4**, 465–480.

[19] Hogan, J. W. and Daniels, M. J. (2002). A hierarchical modelling approach to analysing longitudinal data with drop-out and non-compliance, with application to an equivalence trial in paediatric acquired immune defficiency syndrome. *Applied Statistics* **51**, 1–21.

[20] Ibrahim, J. G., Chen, M.-H. and Lipsitz, S. R. (2001). Missing responses in generalised linear mixed models when the missing data mechanism is nonignorable. *Biometrika* **88**, 551–564.

[21] Kenward, M. G., Molenberghs G. and Thijs, H. (2003). Pattern-mixture models with proper time dependence. *Biometrika* **90**, 53–71.

[22] Kinney, S. K. and Dunson, D. B. (2007). Fixed and random effects selection in linear and logistic models. *Biometrics* **63**, 690–698.

[23] Ley, E. and Steel, M. F. J. (2009). On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics* **24**, 651–674.

[24] Little, R. J. A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika* **81**, 471–483.

[25] Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. Wiley: Hoboken, 2nd ed.

[26] Mitchell, T. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association* **404**, 1023–1032.

[27] Mitra, R. and Dunson, D. (2010). Two-level stochastic search variable selection in GLMs with missing predictors. *The International Journal of Biostatistics* **6**, Iss. 1, Article 33.

[28] National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group (1995). Tissue plasminogen activator for acute ischemic stroke. *The New England Journal of Medicine* **333**, 1581–1587.

[29] Paddock, S. M. (2007). Bayesian variable selection for longitudinal substance abuse treatment data subject to informative censoring. *Applied Statistics* **56**, 293–311.

[30] Park, T. and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association* **103**, 681–686.

[31] Roy, J. and Daniels, M. J. (2008). A general class of pattern mixture models for nonignorable dropout with many possible dropout times. *Biometrics* **64**, 538–545.

[32] Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* **64**, 583–616.

[33] Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error.*Biometrics* **53**, 330–339.

Table 1A. Non-ignorable missing data in longitudinal binary measurements due to dropout (scenario (a), n = 300; $\sigma_0^{\beta 2} = s_0^2 = \sigma_{0l}^{\eta 2} = \sigma_{0l}^{\alpha 2} = 10$, $\pi_0 = 0.5$).

		Joint			Ignorable		
Parameter	True	Mean	SD	Posterior SP	Mean	SD	Posterior SP
$\beta_1^\dagger$	2	1.906	0.308	1.00	1.354	0.439	0.95
β_2	-1	-0.955	0.260	0.98	-0.964	0.295	0.95
$\beta_3^\dagger$	-1.5	-1.422	0.357	0.97	-1.196	0.516	0.90
β_4	0	0.001	0.017	0.06	0.001	0.026	0.07
β_5	1	0.972	0.097	1.00	0.948	0.088	1.00
β_6	0	0.003	0.023	0.06	-0.003	0.018	0.06
β_7	0	0.001	0.030	0.07	-0.002	0.015	0.05
β_8	0	-0.001	0.013	0.06	-0.022	0.052	0.12
β_9	1	0.978	0.089	1.00	0.978	0.081	1.00
β_{10}	0	0.002	0.021	0.06	0.002	0.016	0.06
λ_1	0.5	0.442	0.245	0.75	0.340	0.243	0.57
λ_2	0.5	0.451	0.257	0.77	0.492	0.284	0.73
η_0	1	0.817	0.201	-			
η_1	0	-0.001	0.058	0.11			
η_2	0	-0.206	0.312	0.28			
η_3	-1.5	-1.453	0.323	0.99			
η_4	0.1	0.083	0.233	0.22			
α_1	2	2.049	0.238	1.00			
α_2	0	0.003	0.034	0.07			
α_3	-0.5	-0.522	0.083	1.00			
α_4	0	-0.001	0.025	0.07			
α_5	0	0.002	0.013	0.06			
α_6	-1	-0.987	0.113	1.00			
α_7	0	-0.011	0.035	0.08			
α_8	0	-0.005	0.035	0.07			

† β_1: the time trend in the control group;
β_3: the between group difference in the time trend.

Table 1B. Non-ignorable missing data in longitudinal binary measurements due to dropout (scenario (a), n = 300; $\sigma_0^{\beta 2} = s_0^2 = \sigma_{0l}^{\eta 2} = \sigma_{0l}^{\alpha 2} = 10$, $\pi_0 = 0.8$).

		Joint			Ignorable		
Parameter	True	Mean	SD	Posterior SP	Mean	SD	Posterior SP
$\beta_1^{\dagger}$	2	1.892	0.335	1.00	1.186	0.558	0.86
β_2	-1	-0.955	0.314	0.94	-1.038	0.329	0.95
$\beta_3^{\dagger}$	-1.5	-1.418	0.398	0.94	-1.010	0.598	0.76
β_4	0	0.001	0.005	0.02	-0.001	0.004	0.02
β_5	1	0.969	0.093	1.00	0.946	0.101	1.00
β_6	0	0.001	0.009	0.02	-0.002	0.018	0.02
β_7	0	0.001	0.010	0.02	-0.002	0.010	0.02
β_8	0	0.001	0.005	0.02	-0.018	0.047	0.07
β_9	1	0.979	0.086	1.00	0.963	0.081	1.00
β_{10}	0	0.002	0.018	0.02	0.001	0.030	0.03
λ_1	0.5	0.408	0.283	0.60	0.245	0.239	0.39
λ_2	0.5	0.429	0.291	0.63	0.504	0.326	0.67
η_0	1	0.858	0.161	-			
η_1	0	-0.008	0.060	0.05			
η_2	0	0.067	0.158	0.10			
η_3	-1.5	-1.385	0.261	0.98			
η_4	0.1	0.040	0.112	0.07			
α_1	2	2.027	0.191	1.00			
α_2	0	-0.001	0.005	0.02			
α_3	-0.5	-0.487	0.105	0.98			
α_4	0	0.001	0.014	0.02			
α_5	0	0.001	0.011	0.02			
α_6	-1	-0.968	0.112	1.00			
α_7	0	-0.001	0.008	0.02			
α_8	0	-0.001	0.006	0.02			

† β_1: the time trend in the control group;
β_3: the between group difference in the time trend.

Table 2. Posterior estimate under different priors (scenario (a), n = 300; posterior standard deviation and selection probability in the parenthesis).

	$\pi_0 = 0.5$			$\pi_0 = 0.8$		
Prior Var	5	10	100	5	10	100
β_1	1.90 (0.33,1.00)	1.91 (0.31,1.00)	1.81 (0.38,0.99)	1.89 (0.35,0.99)	1.89 (0.34,1.00)	1.94 (0.37,0.95)
β_2	-0.99 (0.22,0.99)	-0.96 (0.26,0.98)	-1.02 (0.31,0.96)	-0.93 (0.30,0.93)	-0.96 (0.31,0.94)	-0.90 (0.37,0.88)
β_3	-1.42 (0.35,0.98)	-1.42 (0.36,0.97)	-1.34 (0.46,0.93)	-1.44 (0.38,0.97)	-1.42 (0.40,0.94)	-1.47 (0.49,0.94)
λ_1	0.43 (0.25,0.76)	0.44 (0.25,0.75)	0.43 (0.27,0.74)	0.41 (0.28,0.62)	0.41 (0.28,0.60)	0.40 (0.32,0.61)
λ_2	0.45 (0.25,0.77)	0.45 (0.26,0.77)	0.47 (0.30,0.75)	0.42 (0.30,0.65)	0.43 (0.29,0.63)	0.43 (0.32,0.62)
η_2	-0.16 (0.28,0.29)	-0.21 (0.31,0.28)	-0.09 (0.21,0.13)	0.08 (0.19,0.12)	0.07 (0.16,0.10)	0.03 (0.16,0.05)
η_3	-1.41 (0.31,0.99)	-1.45 (0.32,0.99)	-1.40 (0.32,0.97)	-1.31 (0.26,0.98)	-1.39 (0.26,0.98)	-1.32 (0.26,0.98)
η_4	0.08 (0.22,0.26)	0.08 (0.23,0.22)	-0.02 (0.25,0.12)	0.02 (0.16,0.09)	0.04 (0.11,0.07)	-0.01 (0.18,0.04)
α_1	2.00 (0.21,1.00)	2.05 (0.24,1.00)	2.06 (0.21,1.00)	2.01 (0.19,1.00)	2.03 (0.19,1.00)	2.05 (0.23,1.00)

Table 3. Variable selection under model misspecification (n = 300; $\sigma_0^{\beta 2} = s_0^2 = \sigma_{0l}^{\eta 2} = \sigma_{0l}^{\alpha 2} = 10$).

		Joint (π_0 = 0.5)			Joint (π_0 = 0.8)		
Parameter	True	Mean	SD	Posterior SP	Mean	SD	Posterior SP
β_1	2	1.798	0.365	1.00	1.862	0.348	1.00
β_2	-1	-0.988	0.267	0.98	-0.903	0.291	0.93
β_3	-1.5	-1.311	0.438	0.95	-1.399	0.413	0.96
β_4	0	0.001	0.018	0.06	-0.002	0.030	0.03
β_5	1	0.958	0.091	1.00	0.953	0.100	1.00
β_6	0	0.001	0.017	0.06	-0.002	0.011	0.02
β_7	0	-0.001	0.013	0.05	-0.001	0.005	0.02
β_8	0	-0.004	0.020	0.07	-0.001	0.011	0.02
β_9	1	0.961	0.091	1.00	0.956	0.091	1.00
β_{10}	0	0.001	0.012	0.05	-0.001	0.004	0.02
λ_1	0.5	0.309	0.237	0.55	0.274	0.238	0.43
λ_2	0.5	0.451	0.271	0.72	0.402	0.279	0.60
α_1	2	2.032	0.228	1.00	2.089	0.209	1.00
α_2	0	-0.003	0.015	0.06	-0.001	0.025	0.03
α_3	-0.5	-0.512	0.110	0.98	-0.539	0.112	0.98
α_4	0	-0.003	0.035	0.08	-0.001	0.003	0.01
α_5	0	0.003	0.026	0.06	-0.002	0.023	0.02
α_6	-1	-0.984	0.108	1.00	-0.993	0.110	1.00
α_7	0	-0.009	0.028	0.08	-0.007	.026	0.03
α_8	0	-0.002	0.029	0.07	-0.002	0.011	0.02

Table 4. The rt-PA stroke trial: frequency of patients with unfavorable Barthel Index at each of the follow-up times.

	7-10 days		3 months		6 months		12 months	
rt-PA group (N positive)	179	60.6%	96	37.7%	86	35.5%	67	30.2%
Total observed	296		255		242		222	
Missing	16		57		70		90	
placebo group (N positive)	218	76.5%	129	52.0%	116	50.2%	96	44.4%
Total observed	285		248		231		216	
Missing	27		64		81		96	

Table 5. Analyses of unfavorable Barthel Index in the rt-PA stroke trial ($\pi_0 = 0.8$).

	Joint Analysis			Ignorable Analysis		
	Mean	SD	Posterior SP	Mean	SD	Posterior SP
$intercept$	2.39	0.68	-	2.35	0.71	-
$time3$	-3.07	0.34	1.00	-2.90	0.31	1.00
$time6$	-3.07	0.33	1.00	-2.96	0.31	1.00
$time12$	-3.93	0.52	1.00	-3.66	0.34	1.00
$t-PA$	-1.78	0.58	0.97	-1.88	0.56	0.99
$time3 \times t-PA$	0.004	0.07	0.03	0.003	0.08	0.03
$time6 \times t-PA$	0.01	0.11	0.04	0.01	0.10	0.04
$time12 \times t-PA$	0.02	0.15	0.05	0.02	0.13	0.05
age	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$gender$	-0.46	0.66	0.39	-0.21	0.48	0.19
$smoking$	-0.05	0.22	0.08	-0.07	0.28	0.09
$drinking$	-0.13	0.37	0.15	-0.29	0.53	0.29
$diabetes$	0.56	0.82	0.38	0.63	0.86	0.41
$hypertension$	1.37	0.75	0.85	1.71	0.60	0.98
$angina$	-0.15	0.42	0.15	-0.24	0.56	0.21
$abnormal\ CT$	0.25	0.50	0.24	0.27	0.54	0.25
$\lambda_1(intercept)$	4.69	0.53	1.00	5.00	0.44	1.00
λ_2 ($time3$)	0.20	0.53	0.20	0.12	0.31	0.18
η_0	3.37	0.22	-			
η_1	-2.19	0.55	1.00			
η_2	0.94	0.93	0.57			

