Published in final edited form as:

Stat Methods Med Res. 2019 January ; 28(1): 134–150. doi:10.1177/0962280217717760.

Regularized approach for data missing not at random

Chi-hong Tseng¹, Yi-Hau Chen²

¹University of California at Los Angeles

²Academia Sinica

Abstract

It is common in longitudinal studies that missing data occur due to subjects' no response, missed visits, dropout, death or other reasons during the course of study. To perform valid analysis in this setting, data missing not at random (MNAR) have to be considered. However, models for data MNAR often suffer from the identifiability issue and hence result in difficulty in estimation and computational convergence. To ameliorate this issue, we propose the LASSO and Ridge regularized selection models that regularize the missing data mechanism model to handle data MNAR, with the regularization parameter selected via a cross validation procedure. The proposed models can be also employed for sensitivity analysis to examine the effects on inference of different assumptions about the missing data mechanism. We illustrate the performance of the proposed models via simulation studies and the analysis of data from a randomized clinical trial.

Keywords

Missing at Random; LASSO Regression; Ridge Regression; Pseudo Likelihood; Selection Model

Introduction

Missing data problems arise frequently in clinical and observational studies. For example, in a longitudinal study where subjects are followed over time, the outcomes of interest and covariates may be missing due to subjects' no response, missed visits, dropout, death and other reasons during the course of study. A vast statistical literature exists on missing data problems. The fundamental problem of missing data is that the law of observed data is not sufficient to identify the distribution of outcomes of interest. The complete data can be expressed as a mixture of conditional distributions of observed data and unobserved data, and in general the later cannot be identified from the observed data. One way to facilitate the identification of the complete data distribution is to place assumptions on the missing mechanism. Three types of missing data mechanisms have been discussed:¹ missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). If the missingness is independent of both the observed and unobserved data, the missing data mechanism is considered to be MCAR. The MAR mechanism is defined when missingness is independent of unobserved data given observed data. With data MCAR or

Corresponding author: Yi-Hau Chen, Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan, R.O.C.
yhchen@stat.sinica.edu.tw.

MAR, the distribution of missing data can be ignored with likelihood based inference, and the missing data mechanism is ignorable.¹ Otherwise, with data MNAR, the distribution of missing data must play a role to make valid inferences, and hence the missing data mechanism is non-ignorable.

For instance, in our example of Scleroderma Lung Study, about 15% of subjects dropped out of the study before 12 months, and 30% of dropouts were due to death and treatment failures. Intermittent missed visits and missing outcome measures also occurred during the course of the study. It is likely that the missing data are due to the ineffectiveness of treatment and hence are related to the outcome of interest.

In general, to handle data MNAR requires the modelling of both the missing data mechanism and the outcomes of interest.² Three likelihood-based approaches are commonly used for MNAR problems: selection models, pattern mixture models and shared parameter models. Selection models provide a natural way to express the outcome process and the missing data mechanism.³ The models usually consist of an overall outcome model that specifies the distribution of outcomes, and a missing mechanism model that characterizes the dependence between missingness and outcomes of interest. For example, logistic regression model can be employed as the missing mechanism model.^{4, 5} The second approach is based on the pattern mixture models⁶ which consider the full data as a mixture of data from different missing data patterns. This is a flexible modeling approach that allows the outcome models to be different for subjects with different missing data patterns. Finally, the shared parameter models use latent variables, such as random effects, to capture the correlation between the outcome and missingness. For example, a joint modelling approach has been used to analyze the lung function outcomes in a Scleroderma study in the presence of non-ignorable dropouts.^{7, 8}

Although data MNAR may arise in many real applications, the model specifications in MNAR analyses are generally unverifiable with the observed data, and parameters in MNAR models mentioned above may be unidentifiable.^{9, 10, 11, 12} For example, in selection models, it is often impossible to distinguish the violations of assumptions of the distribution of outcomes and the functional form of the missing mechanism model.² In contrast, models that assume ignorable missing data do not require the knowledge of the unobserved data distribution and therefore are generally more identifiable and accessible for model checking.

To overcome the identifiability issues of selection models with data MNAR, we propose to use LASSO and Ridge regression techniques to regularize the missing data mechanism model. LASSO and Ridge regressions are common methods of regularization for ill-posed problems.^{13, 14} In statistical literature, the idea of regularization or shrinkage has been successfully applied to multi-collinearity,¹³ bias reduction,¹⁵ smoothing spline,¹⁶ model selection,¹⁴ high dimensional data analysis problems,¹⁷ and so forth to regularize the model parameters, and hence to ameliorate the identifiability issue and enhance stability in computation and inference. In addition, regularized regression models have Bayesian interpretations. For example, the LASSO estimates are equivalent to the posterior mode estimates in Bayesian analysis with Laplace priors, and the Ridge estimates are equivalent to the posterior mode estimates with Gaussian priors.^{18, 14} There is a rich statistical literature

that employs the Bayesian priors to provide stable estimates effectively in the ill-posed irregular problems.

In the missing data literature, regularized regression has been proposed to provide the estimation of smoothed and flexible covariate distribution.¹⁹ Our approach is different: the proposed regularized selection models impose regularization on the parameters in the missing data mechanism model that represent the strength of correlation between missingness and the outcome, and it aims to provide the computational stability and satisfactory inference under weakly identifiable models. Our approach is similar to the concept of partial prior for sensitivity analysis;²⁰ intuitively shrinkage effect moves the model specification in between ignorable and non-ignorable missing data mechanism. As a consequence, the proposed model may facilitate sensitivity analysis to investigate the impact of missing data mechanism assumptions on the conclusion of analysis.²

We organize the paper as follows. In Section 2, we consider the pseudo likelihood inference and formulate the regularized selection models. Section 3 gives the details of computation and inference procedures for the proposed model. In Section 4, we apply the proposed method to data from the Scleroderma Lung Study. In Section 5, simulation studies are carried out to demonstrate the performance of the proposed model. We conclude the paper, in Section 6, with a discussion.

The Regularized Selection Models

Consider a longitudinal study of n subjects with n_i study visits for the i th subject ($i = 1, \dots, n$). Let Y_{ij} denote the outcome of interest for subject i at the j th visit, and let $M_{ij} = 0, 1$, or 2 indicate respectively that Y_{ij} is observed, intermittently missing, or missing due to dropout. In particular,

$$M_{ij} = \begin{cases} 0 & \text{if } Y_{ij} \text{ is observed;} \\ 2 & \text{if } Y_{ij'} \text{ is missing for all } j', j \leq j' \leq n_i; \text{ (dropout)} \\ 1 & \text{otherwise (intermittent missingness).} \end{cases}$$

Namely, a missing outcome is referred to as “intermittent missingness” if there exist some outcome Y that is observed after the missing outcome. On the other hand, if there exist no outcome Y that is observed after a missing outcome, that missing outcome is defined to be a dropout. Let $X_{ij}(p \times 1)$ be the vector of covariates for subject i at the j th visit. The data available are (Y_{ij}, M_{ij}, X_{ij}) when $M_{ij} = 0$, and are (M_{ij}, X_{ij}) when $M_{ij} = 1$ or 2 for $i = 1, \dots, n, j = 1, \dots, n_i$. That is, only the outcome is subject to missingness while the missingness status and the covariates are always observed.

Under a selection model framework, the likelihood L_i of data for the i th subject ($i = 1, \dots, n$) is factored as the product of an outcome model and a missing mechanism model:

$$\begin{aligned} L_i &= f(Y_{i1}, \dots, Y_{in_i}, M_{i1}, \dots, M_{in_i} | X_i) \\ &= f(Y_{i1}, \dots, Y_{in_i} | X_i) f(M_{i1}, \dots, M_{in_i} | Y_{i1}, \dots, Y_{in_i}, X_i) \\ &\equiv L_{1i} L_{2i} \end{aligned}$$

with $X_i = (X'_{i1}, \dots, X'_{in_i})'$. Similar to,⁴ we consider the pseudo-likelihood type inference such that

$$L_{1i} = f(Y_{i1}, \dots, Y_{in_i} | X_i) = \prod_{j=1}^{n_i} f(Y_{ij} | X_{ij}), \quad (1)$$

Here a generalized linear model²¹ can be considered for $f(Y_{ij} | X_{ij})$ ($i = 1, \dots, n, j = 1, \dots, n_i$) with mean $E(Y_{ij} | X_{ij}) = g(\beta' X_{ij})$ and variance $\text{var}(Y_{ij} | X_{ij}) = \phi g'(\beta' X_{ij})$, where $g(\cdot)$ is some link function relating the covariate vector X_{ij} to the outcome Y_{ij} and $g'(t) = dg(t)/dt$.

We assume a first-order Markov model²² for the missingness model to accommodate missingness due to both missed visits and dropouts such that

$$L_{2i} = f(M_{i1}, \dots, M_{in_i} | Y_{i1}, \dots, Y_{in_i}, X_i) = \prod_{j=1}^{n_i} f(M_{ij} | Y_{ij}, X_{ij}, M_{i,j-1}), \quad (2)$$

namely the missingness status M_{ij} at time j depends on the missingness at past time points only through the missingness $M_{i,j-1}$ at the immediately previous time point given the current outcome Y_{ij} , which is possibly unobserved, and the current covariates X_{ij} .

The Markov-type missingness model can be specified as a multinomial logistic regression model

$$\Pr(M_{ij} = p | M_{i,j-1} = q, Y_{ij}, X_{ij}) = \frac{\phi_{ij}(p, q)}{\sum_{p^*=0}^2 \phi_{ij}(p^*, q)}, \quad (3)$$

with $\phi_{ij}(p, q) = \exp(\alpha_{p0} + \alpha_{p1}Y_{ij} + \alpha'_{p2}X_{ij} + \alpha_{p3}q)$ for $p, q = 0$ (data being observed), 1 (intermittent missingness), 2 (dropout), where for identifiability, $\alpha_{00} = \alpha_{01} = \alpha_{03} \equiv 0$ and α_{02} is a zero vector. Also, α_{23} is set to 0 since by definition there is no transition directly from intermittent missingness to dropout, and $\Pr(M_{ij} = 2 | M_{i,j-1} = 2, Y_{ij}, X_{ij}) \equiv 1$ by recalling that dropout is an absorbing state. Note that here for notational simplicity we assume the covariates involved in the outcome and the missingness models are the same, but in practical implementation they may well be different subsets of the covariate variables.

let $\theta = (\alpha', \beta')'$ and $\alpha' = (\alpha_{p0}, \alpha_{p1}, \alpha_{p2}, \alpha_{p3}; p = 1, 2)$. With above model specifications, the total log pseudo-likelihood is

$$\ell(\theta) = \log \prod_{i=1}^n L_i = \sum_{i=1}^n \sum_{j=1}^{n_i} \log L_{ij}(\theta) \quad (4)$$

where

$$L_{ij}(\theta) = f(Y_{ij} | X_{ij}; \beta) f(M_{ij} | M_{i,j-1}, Y_{ij}, X_{ij}; \alpha)$$

if Y_{ij} is observed, and

$$L_{ij}(\theta) = \int_{y_{ij}} f(y_{ij} | X_i; \beta) f(M_{ij} | M_{i,j-1}, y_{ij}, X_{ij}; \alpha) dy_{ij}$$

if Y_{ij} is missing. The parameter estimates can be obtained by solving the pseudo-score equation

$$\partial \ell(\theta) / \partial \theta = 0. \quad (5)$$

Nevertheless, the selection models often suffer from identifiability problems,^{9, 11, 12} which can result in unstable and unreliable estimates when solving the pseudo-score equation above. The parameters α_{p1} ($p = 1, 2$) represent the degree of missingness not at random: the more α_{p1} deviates from 0, the stronger dependence is between outcome and missingness, and when $\alpha_{p1}=0$ for $p=1, 2$ the model reduces to an MAR model. They have been called sensitivity parameters²³ or bias parameters.²⁴ Although the sensitivity parameters can not be identified from observed data, all parameters become identifiable when the sensitivity parameters are given. As a result, it has been a common practice to analyze data with various values of sensitivity parameters.²³ Theoretical results also imply that the parameters in some simplified selection models are identifiable if prior knowledge and restriction on sensitivity parameters are available.¹¹ Therefore, we consider using a regularized selection model which is based on the models (1) and (2) but with the LASSO (L^1 -norm) or Ridge (L^2 -norm) regularization on the magnitudes of the parameters α_{p1} ($p = 1, 2$). Specifically, the regularized log pseudo-likelihoods corresponding to the LASSO and Ridge regularized selection models are given respectively by:

$$\ell_1(\theta) = \ell(\theta) - N\lambda \|\alpha_1\|_1$$

and

$$\ell_2(\theta) = \ell(\theta) - N\lambda \|\alpha_1\|_2,$$

where $N = \sum_i n_i$ and $\|\alpha_1\|_r \equiv \sum_{p=1,2} |\alpha_{p1}|^r$. The constant λ in $\ell_1(\theta)$ and $\ell_2(\theta)$ is the regularization parameter which determines the degree of regularization of the parameters α_{p1} ($p = 1, 2$); a larger value of λ leads to a stronger degree of regularization on α_{p1} ($p = 1, 2$).

For a given value of λ , the proposed estimator $\hat{\theta}$ for the regularized selection model parameter θ is obtained by solving

$$\partial \ell_r(\theta) / \partial \theta = 0, \quad r = 1 \text{ or } 2, \quad (6)$$

which is expected to enjoy more stable computational performance than the estimator without regularization by solving (5). Our numerical studies shown later will provide empirical evidence supporting this.

In the context of the regularized selection models we proposed, the role of the regularization parameter λ can be twofold. Firstly, because regularized regression models have Bayesian interpretation¹⁴ and λ reflects one's belief on missing data mechanism. Therefore sensitivity analysis can be performed to obtain estimates of the parameter β over a range of λ . It allows us to examine the impact of missing data assumptions on the inference of outcome models, and addresses the issue of uncertainty in missing data mechanism when analyzing real data.² Secondly, λ can serve as a tuning parameter to facilitate the estimation of θ . To this aim, we propose using 5-fold cross validation to choose the value of λ that yields the minimum cross-validation mean squared error (CVMSE). Here the CVMSE for a fixed value of λ is defined as

$$\frac{1}{5} \sum_{K=1}^5 \frac{\sum_{i \in D_K} \sum_{j=1}^{n_i} I(M_{ij}=0) \{Y_{ij} - \hat{E}^{-K}(Y_{ij} | M_{ij}=0, M_{i,j-1}, X_{ij}; \lambda)\}^2}{\sum_{i \in D_K} \sum_{j=1}^{n_i} I(M_{ij}=0)},$$

where $K=1, \dots, 5$ denotes the folds of the sample, and D_K is the subject index set for the K th fold (i.e., subjects in the K th fold of the sample). The term $\hat{E}^{-K}(Y_{ij} | M_{ij}=0, M_{i,j-1}, X_{ij}; \lambda)$ is the mean of Y_{ij} given $M_{ij}=0$ and data on $M_{i,j-1}$ and X_{ij} based on the outcome model $f(Y_{ij} | X_{ij}; \hat{\beta}^{-K}, \lambda)$ and the missingness mechanism model $\Pr(M_{ij}=0 | M_{i,j-1}, Y_{ij}, X_{ij}; \hat{\alpha}^{-K}, \lambda)$ given in (3), with $\hat{\beta}^{-K}$ and $\hat{\alpha}^{-K}$ the estimates of β and α using only observed data outside the K -th fold of the sample for a given λ value. Explicitly,

$$\hat{E}^{-K}(Y_{ij} | M_{ij}=0, M_{i,j-1}, X_{ij}, \lambda) = \frac{\int y f(y | X_{ij}; \hat{\beta}^{-K}, \lambda) \Pr(M_{ij}=0 | M_{i,j-1}, y, X_{ij}; \hat{\alpha}^{-K}, \lambda) dy}{\int f(y | X_{ij}; \hat{\beta}^{-K}, \lambda) \Pr(M_{ij}=0 | M_{i,j-1}, y, X_{ij}; \hat{\alpha}^{-K}, \lambda) dy}.$$

In both the L^1 and L^2 regularized pseudo-likelihoods, our numerical studies suggest that the cross validation procedure given above produce satisfactory inference results on the regression parameter β with λ value being of order $O(1/\sqrt{n})$.

Computation and inference

For a given value of the regularization parameter λ , the Ridge (L^2) regularized log pseudo-likelihood $\ell_2(\theta)$ is smooth in θ and hence can be readily solved via a Newton-Raphson algorithm as in the usual ridge regression. For the LASSO (L^1) regularized log pseudo-likelihood $\ell_1(\theta)$, which is however non-smooth in θ , we follow the technique in²⁵(section 3.3) to approximate the L^1 regularized function $\|\alpha_1\|_1$ locally by a quadratic function, and then apply a Newton-Raphson algorithm to solve the resulting regularized pseudo-score equation. Specifically, let α_{p1}^* be a current estimate of α_{p1} , $p=1, 2$. We approximate $|\alpha_{p1}|$ by the quadratic function $\alpha_{p1}^2 / (2|\alpha_{p1}^*|)$ around α_{p1}^* , $p=1, 2$. Then, in each iteration of the Newton Raphson procedure, when the absolute difference between α_{p1}^* and 0 is smaller than a threshold value such as 10^{-8} , we set the estimate of α_{p1} to 0. This algorithm is very stable and fast in the considered setting.

The sandwich estimator for the variance-covariance matrix of $\hat{\theta}$ can provide statistical inference under the regularized selection models.²⁵ For the LASSO regularization, let Δ be a diagonal matrix of size equal to the length of θ with the diagonal elements corresponding to α_{11} and α_{21} being $1/|\alpha_{11}|$ and $1/|\alpha_{21}|$, respectively, and all the other diagonal elements being zero. For the Ridge regularization, Δ is similarly defined with both the diagonal elements corresponding to α_{11} and α_{21} being 2. Let

$$U_{ij}(\theta) = \frac{\partial}{\partial \theta} \log L_{ij}(\theta) - \lambda \Delta \theta$$

and

$$H(\theta) = \sum_{i,j} \frac{\partial^2}{\partial \theta \partial \theta'} \log L_{ij}(\theta) - N \lambda \Delta.$$

Then an variance estimate for $\hat{\theta}$ can be obtained by

$$\{H(\hat{\theta})\}^{-1} \left\{ \sum_{i,j} U_{ij}(\hat{\theta}) U_{ij}'(\hat{\theta}) \right\} \{H(\hat{\theta})\}^{-1}.$$

Example

In this section, we demonstrate the use of proposed method in the analysis of data from the Scleroderma lung study (SLS).²⁶ The Scleroderma lung study is a multi-center placebo-control double bind randomized study to evaluate the effects of oral cyclophosphamide (CYC) on lung function and other health-related symptoms in patients with evidence of active alveolitis and scleroderma-related interstitial lung disease. In this study, eligible participants received either daily oral cyclophosphamide or matching placebo for 12 months, followed by another year of follow-up without study medication.

A large portion of Scleroderma patients suffer from cough symptom.²⁷ Table 1 gives the percents of subjects with moderate or severe cough for CYC and Placebo groups. At baseline, about 30% patients had moderate or severe cough, and the percentages were reduced to 11% and 20% at 24 months for the intervention group and control group, respectively. However, about 50% subjects had intermittently missing data or dropped out by 24 months.

We applied the regularized selection model to examine the treatment effect on cough symptom in the SLS study. Since the outcome is binary (moderate/severe vs. mild/none cough), a logistic regression is used for the outcome model with the covariates of treatment (intervention vs. control), time and treatment-time interactions. For the missing mechanism model, a multinomial logistic regression model (3) is used to model the transition among 3 states of ‘observing outcome’, ‘intermittent missingness’, and ‘dropout’, with cough, treatment assignment, and missingness at the previous visit as covariates. Five fold cross validation method was used to choose the regularization parameters for LASSO and Ridge

regularized selection models such that the expected and observed data are closest. Table 2 provides the parameter estimates and inferences. Both the LASSO and Ridge regularized selection models show similar results that the intervention group has faster decline in percent of subjects with moderate or severe cough over time.

As a sensitivity analysis, we also perform analyses with various regularization parameters to investigate the influence of missing data assumption on the estimates. Table 3 gives the results of outcome model for various values of λ of order of $O(1/\sqrt{n})$. Without regularization ($\lambda = 0$), numerical convergence was not reached within the pre specified maximum number of iterations of 50. For the LASSO and Ridge selection models with various values of regularization parameter λ , the results are very similar.

When interpreting the analysis results of the SLS data, we should note that twelve patients died during the two years study follow-up. In this analysis, we assume that patient dropout merely censored the measures of cough and cough could have been measured after drop-out time. Although this assumption is consistent with the proposed analysis plan for other longitudinal endpoints of the study,²⁶ it seems unlikely when the drop out cause is death. To properly handle death, one possible approach is to make inferences about the subpopulation of individuals who would survive, or who have non-zero probability of surviving, to a certain time t .^{28, 29} Because this example aims to illustrate the use of our proposed method, the issue of death is not addressed in the analysis and caution is needed when interpreting the results.

Numerical Studies

We perform simulation studies to assess the performance of the proposed regularized selection models for the analysis of missing data. In this section we present the binary outcome logistic regression simulation. Normal outcome linear regression simulation is included in the supplemental materials.

Here we consider the binary outcome logistic regression problem, similar to the cough data in the SLS study. In particular, the covariate vector $X_{it} = (X_{ij,1}, X_{ij,2})'$ for subject i at time j ($1 \leq j \leq n_i, 1 \leq i \leq n$) is composed of a time-fixed covariate $X_{ij,1}$ which follows Bernoulli(0.5), and a time-varying covariate $X_{ij,2} = j - 1$. The number of visits n_i for each subject is a constant of 3 or 5. For the outcomes, the joint distribution of $Y_i = (Y_{i1}, \dots, Y_{in_i})$ is simulated from the Bahadur's representation:³⁰

$$f(y_i | \mu_i, \rho_i) = \left\{ \prod_{j=1}^{n_i} \mu_{ij}^{y_{ij}} (1 - \mu_{ij})^{1 - y_{ij}} \right\} \left(1 + \sum_{j < k} \rho_{ijk} e_{ij} e_{ik} \right), \quad (7)$$

where $\mu_i = E(Y_i | X_i) = (\mu_{i1}, \dots, \mu_{in_i})$ with

$$\log \left(\frac{\mu_{ij}}{1 - \mu_{ij}} \right) = \beta_0 + \beta_1 X_{ij,1} + \beta_2 X_{ij,2}.$$

$e_{ij} = \frac{(Y_{ij} - \mu_{ij})}{\sqrt{\mu_{ij}(1 - \mu_{ij})}}$, and $\rho_{ijk} = E(e_{ij}e_{ik})$ for $1 \leq j < k \leq n_i$. The parameter values in the true model are $\beta_0 = -0.25$, $\beta_1 = 0.25$, $\beta_2 = 0.25$, and $\rho_{ijk} = 0.25$, $1 \leq i \leq n$, $1 \leq j < k \leq n_i$. These parameter values make (7) a bona fide density when $n_i = 3$ or 5.

The missingness mechanism is determined by the Markov transition model given as (3). The missing status M_{ij} 's are simulated from model (3) with $\alpha_{p1} = 0$, which corresponds to ignorable missingness, $\alpha_{p1} = 0.5$, which corresponds to moderate non-ignorable missingness, or $\alpha_{p1} = 1$, which corresponds to strong non-ignorable missingness, for $p = 1, 2$. The value of α_{p2} is fixed at $(0.1, 0)'$ for $p = 1, 2$, and α_{23} is fixed at 1. The value of α_{p0} for $p = 1, 2$ is specified to yield the proportion of missing observations around 30%. Across the simulations, the sample size $n = 100$, $n_i = 3$ or 5, and the simulation replication is 1500. The maximum number of iterations is 50 in each simulation. The 95% Wald type confidence intervals for β 's are constructed with $\hat{\beta} \pm 1.96 \cdot \text{Std.Err}(\hat{\beta})$. Bias, mean square error (MSE) and 95% coverage probability (CP) are calculated to evaluate the performance of proposed methods.

Table 4 shows the simulation results for the parameters in the outcome model, with the regularization parameter λ determined by 5-fold cross validation. The parameters in the outcome model are often parameters of interest, and α 's are considered as nuisance parameters. For the data generated by ignorable missing data mechanism, the proposed model works well for both long ($n_i = 5$) and short ($n_i = 3$) follow-ups such that the estimates have minimal bias and their 95% coverage probability attaining the nominal level. The bias is minimal for β_0 and β_1 , they are generally within 10% of standard deviation. With increasing correlation between outcome and missingness among non-ignorable missingness, the simulations suggest bias and the mean square error of the estimates become larger, particularly for the coefficient of time-trend variable (β_2). This larger bias may be also due to the difficulty in estimating time-trend with non-ignorable outcome missingness. For example, with $n_i = 3$ and Ridge regularized selection models, the bias of β_2 increases from 0.010 for ignorable missing data to 0.072 and 0.094 for moderate and strong MNAR data. Although MSE also increases from 0.021 to 0.030 and 0.029, it appears stable between moderate and strong MNAR data. The bias is reduced significantly when more follow-up visits are available ($n_i = 5$). The coverage probability for the 95% Wald-type confidence interval is generally satisfactory and is over 85% in both the cases with $n_i = 3$ and 5 even with strong MNAR mechanism.

In the second simulation, we investigate the performance of the LASSO and Ridge regularized selection models with different regularization parameters of $\lambda = (0, 0.01, 0.05, 0.1, 1) / \sqrt{n}$ (Figure 1). Without regularization ($\lambda = 0$) or small regularization ($\lambda = 0.01 / \sqrt{n}$), the simulations show difficulty in identifying regression parameters and low percentage of convergence in computation. On the other hand, when $\lambda = 1 / \sqrt{n}$, the convergence rates are close to 100% in all cases. Figures 2, 3, and 4 give the bias, mean squared error (MSE) and 95% coverage probability of regression coefficients in the outcome model with 3 follow-ups ($n_i = 3$), among the simulation runs that reach numerical convergence. With no or small regularization ($\lambda = 0, 0.1 / \sqrt{n}$), larger bias and small coverage probability are generally observed, in particular for the coefficient of time-trend variable

(β_2). On the other hand, using $\lambda = 1 / \sqrt{n}$ generally provides more desirable parameter inferences with smaller bias, smaller MSE and coverage probability $> 90\%$. The results are similar when longer follow-up are available with $n_i = 5$, presented in the supplemental materials. Additional simulations with Normal outcomes are also provided in the supplemental materials.

Discussion

Selection models provide a natural way of specifying the overall outcome process and the relationship between missingness and outcome. However, with data MNAR, selection models often suffer from identifiability issues and difficulty in numerical convergence. In this paper, we use the LASSO and Ridge regression techniques to regularize the parameters that characterizes the MNAR mechanism. We have demonstrated by numerical simulations that the proposed regularized selection model is computationally more stable than the unregularized one and provides satisfactory inferences for the regression parameters. We note that our method does not solve the fundamental problem that missing data model assumptions are generally not verifiable and many models can have equally good fit to a set of observed data.³¹ Instead, we aim to provide one practical solution to the identifiability issues when fitting selection models. Our regularized approach provides the computational stability and satisfactory inference under weakly identifiable models. The theoretical property of proposed method, however, needs further investigation.

We have illustrated comparable and satisfactory performance of Ridge and LASSO regularization on weakly identifiable MNAR models. Alternative regularization methods, such as elastic net,³² have subtle but important difference from LASSO and Ridge regularization, and are readily applicable to our proposed approach. In addition, there is a rich statistical literature that employs the Bayesian priors to provide stable estimates effectively in the ill-posed irregular problems, and regularization approaches can usually be cast in the Bayesian framework. Although the regularization parameter(λ) is generally chosen using cross-validation, one can potentially express prior belief on the strength of MNAR by specifying the regularization parameter λ according to experts knowledge about the odds of dropout or missed visits for a proportional change in the outcome. For example, LASSO regressions are equivalent to the Bayesian analysis with Laplace priors, and one can use several quantiles to uniquely identify the prior distribution and regularization parameter.^{33, 20} Further research evaluating the use other regularization methods and Bayesian priors in MNAR models is worthwhile.

Our simulation results illustrate the excellent performance with the regularization parameter $O(1 / \sqrt{n})$, and suggest that the cross-validation method provides a viable way to choose the regularization parameter for the proposed regularized selection models. Missing data mechanisms are generally not testable and MNAR models rely on assumptions that cannot be verified empirically. It is crucial to execute and interpret missing data analysis with extra care. In practice, we recommend using cross validation to determine the regularization parameter, as well as using different values of the regularization parameter as sensitivity analyses to investigate the impact of missing data assumptions and the robustness of results under different missing data assumptions.^{23, 2} In addition, the region constraint approach³⁴

can provide further insight on both ignorance, which represents the uncertainty about selection bias or missing data mechanism, and imprecision, which represents sampling random error. Similarly the relaxation penalties and priors approach²⁰ can be applied to conduct sensitivity analysis and compare these two sources of uncertainty. By varying the regularization parameter λ in the regularized selection models, one can possibly perform sensitivity analysis over a region of parameter values that are consistent with the observed data model in the spirit of the region constraint approach.³⁴ This sensitivity approach will be a topic of our future work.

We used 5-fold cross validation in our numerical studies. It is recommended use five- or ten fold cross validation as a good compromise between bias and variance.³⁵ We have tried both 5 and 10 fold cross validations in the initial simulation runs, and the results are very similar. Other choices, such as leave-one-out cross validation, could also be viable options. Other models to handle MNAR data, such as the pattern mixture models, also suffer from the identifiability problem.⁶ The regularization technique similar to the proposed one may be useful in making the pattern mixture models more stable and providing reliable estimates. We are planning further work in this area.

Acknowledgment

This research is supported by National Science Council of Republic of China (NSC 104-2118-M-001-006-MY3), NIH/National Heart, Lung, and Blood Institute Grant Numbers U01HL060587 and R01HL089758, and NIH/National Center for Advancing Translational Science (NCATS) UCLA CTSI Grant Number UL1TR000124.

References

1. Little RJA and Rubin DB. Statistical Analysis with Missing Data. 2nd ed. New York: Wiley, 2002.
2. Daniels MJ and Hogan JW. Missing data in longitudinal studies: Strategies for Bayesian modeling and sensitivity analysis. CRC Press, 2008.
3. Wu MC and Carroll RJ. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* 1988; 44(1): 175–188.
4. Troxel AB, Lipsitz S and Harrington D. Marginal models for the analysis of longitudinal measurements with non-ignorable non-monotone missing data. *Biometrika* 1998; 85(3): 661–672.
5. Parzen M, Lipsitz S, Fitzmaurice G et al. Pseudo-likelihood methods for longitudinal binary data with non-ignorable missing responses and covariates. *Statistics in Medicine* 2006; 25: 2784–2796. [PubMed: 16345018]
6. Little R Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 1993; 88(421): 125–134.
7. Elashoff RM, Li G and Li N. An approach to joint analysis of longitudinal measurements and competing risks failure time data. *Statistics in Medicine* 2007; 26: 2813–2835. [PubMed: 17124698]
8. Elashoff RM, Li G and Li N. A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics* 2008; 64: 762–771. [PubMed: 18162112]
9. Rotnitzky A and Robins J. Analysis of semiparametric regression models with non-ignorable non-response. *Statistics in medicine* 1997; 16(1–3): 81. [PubMed: 9004385]
10. Wang S, Shao J and Kim JK. An instrumental variable approach for identification and estimation with nonignorable nonresponse. *Statistica Sinica* 2014; : 1097–1116.
11. Miao W, Ding P and Geng Z. Identifiability of normal and normal mixture models with non-ignorable missing data. *Journal of the American Statistical Association* 2015;.
12. Zhao J and Shao J. Semiparametric pseudo-likelihoods in generalized linear models with non-ignorable missing data. *Journal of the American Statistical Association* 2015; 110(512): 1577–1590.

13. Herl AE and Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 1970; 12(1): 55–67.
14. Tibshirani R Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B (Methodological)* 1996; : 267–288.
15. Firth D Bias reduction of maximum likelihood estimates. *Biometrika* 1993; 80(1): 27–38.
16. Wahba G Spline models for observational data, volume 59 Siam, 1990.
17. Wu B Differential gene expression detection using penalized linear regression models: the improved sam statistics. *Bioinformatics* 2005; 21(8): 1565–1571. [PubMed: 15598833]
18. Titterton DM. Common structure of smoothing techniques in statistics. *International Statistical Review / Revue Internationale de Statistique* 1985; 53(2): 141–170.
19. Chen Q and Ibrahim JG. Semiparametric models for missing covariate and response data in regression models. *Biometrics* 2006; 62(1): 177–184. [PubMed: 16542244]
20. Greenland S Relaxation penalties and priors for plausible modeling of nonidentified bias sources. *Statistical Science* 2009; 24(2): 195–210.
21. McCullagh P and Nelder JA. *Generalized Linear Models*, volume 37 CRC press, 1989.
22. Albert PS and Follmann DA. A random effects transition model for longitudinal binary data with informative missingness. *Statistica Neerlandica* 2003; 57(1): 100–111.
23. Molenberghs G, Kenward MG and Goetghebeur E. Sensitivity analysis for incomplete contingency tables: The slovenian plebiscite case. *Journal of the Royal Statistical Society Series C (Applied Statistics)* 2001; 50(1): 15–29.
24. Greenland S Multiple-bias modelling for analysis of observational data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2005; 168(2): 267–306.
25. Fan J and Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 2001; 96(456): 1348–1360.
26. Tashkin DP, Elashoff R, Clements PJ et al. Cyclophosphamide versus Placebo in Scleroderma Lung Disease. *New England Journal of Medicine* 2006; 354(25): 2655–2666. [PubMed: 16790698]
27. Theodore AC, Tseng CH, Li N et al. Correlation of cough with disease activity and treatment with cyclophosphamide in scleroderma interstitial lung disease: findings from the scleroderma lung study. *CHEST Journal* 2012; 142(3): 614–621.
28. Frangakis CE and Rubin DB. Principal stratification in causal inference. *Biometrics* 2002; 58(1): 21–29. [PubMed: 11890317]
29. Kurland BF, Johnson LL, Egleston BL et al. Longitudinal data with follow-up truncated by death: match the analysis method to research aims. *Statistical Science* 2009; 24(2): 211–222. [PubMed: 20119502]
30. Bahadur RR. A representation of the joint distribution of responses to n dichotomous items. *Studies in Item Analysis and Prediction* 1961; 6: 158–168.
31. Molenberghs G, Beunckens C, Sotito C et al. Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society Series B (Statistical Methodology)* 2008; 70(2): 371–388.
32. Zou H and Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2005; 67(2): 301–320.
33. Scharfstein DO, Daniels MJ and Robins JM. Incorporating prior beliefs about selection bias into the analysis of randomized trials with missing outcomes. *Biostatistics* 2003; 4(4): 495–512. [PubMed: 14557107]
34. Vansteelandt S, Goetghebeur E, Kenward MG et al. Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica* 2006; 16(3): 953–979.
35. Hastie T, Tibshirani R and Friedman J. *The elements of statistical learning*. Springer Series in Statistics, Springer, New York, 2009.

Figure 1.

Convergence percentage for simulations with ignorable, moderate non-ignorable, and strong non-ignorable data with $n_i=3$ and 5, $\lambda = \lambda_0 / \sqrt{n} (\lambda_0 = 0.01, 0.05, 0.1, \text{ and } 1)$

Figure 2.

Bias, mean square error (MSE) and 95% coverage probability (95% CP) for simulations with ignorable data with $n_I = 3$, $\lambda = \lambda_0 / \sqrt{n}$ ($\lambda_0 = 0.01, 0.05, 0.1$, and 1). Solid and dashed lines are for estimates from LASSO and ridge regularized selection models, respectively.

Figure 3.

Bias, mean square error (MSE) and 95% coverage probability (95% CP) for simulations with moderate non-ignorable data with $n_i = 3$, $\lambda = \lambda_0 / \sqrt{n}$ ($\lambda_0 = 0.01, 0.05, 0.1$, and 1). Solid and dashed lines are for estimates from LASSO and ridge regularized selection models, respectively.

Figure 4.

Bias, mean square error (MSE) and 95% coverage probability (95% CP) for simulations with strong non-ignorable data with $n_i = 3$, $\lambda = \lambda_0 / \sqrt{n}$ ($\lambda_0 = 0.01, 0.05, 0.1$, and 1). Solid and dashed lines are for estimates from LASSO and ridge regularized selection models, respectively.

Figure 5.

Bias, mean square error (MSE) and 95% coverage probability (95% CP) for simulations with ignorable data with $n_I = 5$, $\lambda = \lambda_0 / \sqrt{n}$ ($\lambda_0 = 0.01, 0.05, 0.1$, and 1). Solid and dashed lines are for estimates from LASSO and ridge regularized selection models, respectively.

Figure 6.

Bias, mean square error (MSE) and 95% coverage probability (95% CP) for simulations with moderate non-ignorable data with $n_i = 5$, $\lambda = \lambda_0 / \sqrt{n}$ ($\lambda_0 = 0.01, 0.05, 0.1$, and 1). Solid and dashed lines are for estimates from LASSO and ridge regularized selection models, respectively.

Figure 7.

Bias, mean square error (MSE) and 95% coverage probability (95% CP) for simulations with strong non-ignorable data with $n_i = 5$, $\lambda = \lambda_0 / \sqrt{n}$ ($\lambda_0 = 0.01, 0.05, 0.1$, and 1). Solid and dashed lines are for estimates from LASSO and ridge regularized selection models, respectively.

Table 1.

Summary of the number of observed data ($M=0$), percentage of the moderate or severe cough symptom (percent cough), percentages of intermittent ($M=1$) and dropout missingness ($M=2$), for the intervention and control groups in the SLS study

Month	Control				Intervention			
	$M=0$	percent cough	$M=1$	$M=2$	$M=0$	percent cough	$M=1$	$M=2$
0	79	27%	0%	0%	77	29%	0%	0%
3	72	15%	1%	8%	71	24%	4%	9%
6	72	24%	1%	8%	69	20%	5%	10%
9	64	19%	4%	15%	67	19%	6%	12%
12	61	36%	3%	20%	68	18%	1%	16%
15	50	20%	4%	33%	56	25%	1%	29%
18	44	30%	1%	43%	48	21%	3%	36%
21	39	38%	5%	46%	43	19%	1%	44%
24	35	20%	0%	56%	38	11%	0%	52%

Table 2.

Cough analysis for the SLS study with LASSO and Ridge regularized selection models.

A. LASSO Selection Model				
Outcome Model	Variable	Estimate	Std.err	p value
	Intercept	-1.290	0.222	< 0.001
	Treatment	0.323	0.322	0.316
	Time	0.018	0.013	0.175
	Time*Treatment	-0.051	0.021	0.016
Missing Mechanism Model				
Dropout	Variable	Estimate	Std.err	p value
	Intercept	-2.252	0.144	< 0.001
	Cough	0	0	-
Intermittent Missing	Treatment	-0.095	0.208	0.649
	Intercept	-3.539	0.302	< 0.001
	Cough	0	0	-
Previous Missing Status	Treatment	0.104	0.392	0.790
	Previous Missing Status	2.017	0.532	< 0.001
B. Ridge Selection Model				
Outcome Model	Variable	Estimate	Std.err	p value
	Intercept	-1.290	0.222	< 0.001
	Treatment	0.324	0.323	0.316
	Time	0.018	0.013	0.177
	Time*Treatment	-0.051	0.021	0.016
Missing Mechanism Model				
Dropout	Variable	Estimate	Std.err	p value
	Intercept	-2.250	0.143	< 0.001
	Cough	-0.010	0.025	0.690
Intermittent Missing	Treatment	-0.095	0.208	0.648
	Intercept	-3.549	0.324	< 0.001
	Cough	0.037	0.103	0.718
Previous Missing Status	Treatment	0.106	0.396	0.789
	Previous Missing Status	2.020	0.560	< 0.001

Table 3.

Sensitivity analysis of cough analysis in the SLS study with regularized selection models. The parameter estimates of the outcome model are presented with various values of the regularization parameter $\lambda = \lambda_0 / \sqrt{n}$ and $\lambda_0 = 0, 0.5, 1, 5$.

Selection Model	λ_0	variable	Estimate	Std.err	p value
no penalty	0	not convergent			
LASSO	0.5	Intercept	-1.290	0.222	< 0.001
		Treatment	0.323	0.322	0.316
		Time	0.018	0.013	0.175
		Time*Treatment	-0.051	0.021	0.016
LASSO	1	Intercept	-1.290	0.222	< 0.001
		Treatment	0.323	0.322	0.316
		Time	0.018	0.013	0.175
		Time*Treatment	-0.051	0.021	0.016
LASSO	5	Intercept	-1.290	0.222	< 0.001
		Treatment	0.323	0.322	0.316
		Time	0.018	0.013	0.175
		Time*Treatment	-0.051	0.021	0.016
Ridge	0.5	Intercept	-1.289	0.222	< 0.001
		Treatment	0.325	0.323	0.315
		Time	0.018	0.013	0.180
		Time*Treatment	-0.051	0.021	0.016
Ridge	1	Intercept	-1.290	0.222	< 0.001
		Treatment	0.324	0.323	0.316
		Time	0.018	0.013	0.177
		Time*Treatment	-0.051	0.021	0.016
Ridge	5	Intercept	-1.290	0.222	< 0.001
		Treatment	0.323	0.322	0.317
		Time	0.018	0.013	0.175
		Time*Treatment	-0.051	0.021	0.016

Table 4.

Simulation results for binary outcome with number of repeated measure $n_i = 3$ and 5. Bias, standard deviation of the estimate (Std), estimated standard error (E. ste), mean square error (MSE), and 95% coverage probability (95% CP) of confidence interval are presented.

A. $n_i = 3$							
Data	Model	Parameter	Bias	Std	E. ste	MSE	95% CP
Ignorable	LASSO	β_0	-0.017	0.243	0.247	0.059	95.2 %
		β_1	0.011	0.315	0.308	0.099	93.9 %
		β_2	0.005	0.145	0.142	0.021	93.8 %
	Ridge	β_0	-0.011	0.245	0.285	0.060	95.4 %
		β_1	0.008	0.304	0.358	0.093	95.1 %
		β_2	0.010	0.145	0.172	0.021	93.8 %
Moderate nonignorable	LASSO	β_0	-0.010	0.253	0.249	0.064	95.1 %
		β_1	0.001	0.323	0.313	0.104	94.1 %
		β_2	-0.086	0.150	0.147	0.030	90.5 %
	Ridge	β_0	-0.007	0.251	0.253	0.063	94.8 %
		β_1	0.004	0.319	0.323	0.102	95.1 %
		β_2	-0.072	0.156	0.161	0.030	92.1 %
Strong - nonignorable	LASSO	β_0	-0.025	0.247	0.245	0.062	95.5 %
		β_1	-0.006	0.308	0.305	0.095	94.3 %
		β_2	-0.110	0.136	0.138	0.031	87.8 %
	Ridge	β_0	-0.021	0.251	0.251	0.064	94.8 %
		β_1	-0.009	0.310	0.316	0.096	94.6 %
		β_2	-0.094	0.141	0.168	0.029	90.7 %
B. $n_i = 5$							
Data	Model	Parameter	Bias	Sth	E.ste	MSE	95% CP
Ignorable	LASSO	β_0	-0.007	0.226	0.228	0.051	95.8 %
		β_1	0.020	0.292	0.288	0.086	94.9 %
		β_2	-0.001	0.070	0.071	0.005	95.9 %
	Ridge	β_0	-0.004	0.224	0.231	0.050	95.2 %
		β_1	0.007	0.291	0.292	0.085	95.3 %
		β_2	0.005	0.072	0.076	0.005	96.1 %
Moderate nonignorable	LASSO	β_0	-0.028	0.238	0.231	0.057	93.8 %
		β_1	-0.019	0.294	0.293	0.087	95.3 %
		β_2	-0.041	0.076	0.074	0.007	90.2 %
	Ridge	β_0	-0.019	0.233	0.234	0.055	94.1 %
		β_1	-0.026	0.287	0.293	0.083	95.5 %
		β_2	-0.031	0.078	0.078	0.007	92.6 %
Strong nonignorable	LASSO	β_0	-0.053	0.231	0.225	0.056	93.6 %

A. $n_i = 3$							
Data	Model	Parameter	Bias	Std	E. ste	MSE	95% CP
		β_1	-0.014	0.285	0.282	0.082	94.9 %
		β_2	-0.058	0.068	0.067	0.008	85.4 %
	Ridge	β_0	-0.049	0.230	0.251	0.055	94.4 %
		β_1	-0.010	0.282	0.299	0.080	94.7 %
		β_2	-0.046	0.071	0.098	0.007	89.4 %

Table 5.

Simulation results for Normal outcome with number of repeated measures $n_i = 3$. Bias, mean square error (MSE) and 95% coverage probability (95% CP) for simulations with $\lambda = \lambda_0 / \sqrt{n} (\lambda_0 = 0, 0.01, 0.05, 0.1, \text{ and } 1)$

LASSO Models		Ignorable			Moderate nonignorable			Strong nonignorable		
λ_0	Parameter	Bias	MSE	95% CP	Bias	MSE	95% CP	Bias	MSE	95% CP
0	β_0	-0.003	0.017	94.5%	-0.007	0.016	95.5%	0.001	0.016	96.5%
	β_1	-0.014	0.032	93.9%	-0.010	0.030	96.4%	-0.023	0.030	95.0%
	β_2	0.012	0.009	86.4%	-0.022	0.008	88.8%	-0.043	0.011	84.8%
0.01	β_0	-0.002	0.017	94.3%	-0.006	0.016	95.9%	0.002	0.016	96.5%
	β_1	-0.014	0.032	94.1%	-0.012	0.029	95.7%	-0.024	0.030	94.8%
	β_2	0.011	0.009	86.5%	-0.024	0.008	88.5%	-0.045	0.011	84.9%
0.05	β_0	-0.002	0.017	93.7%	-0.006	0.016	95.1%	0.004	0.016	96.7%
	β_1	-0.013	0.032	93.9%	-0.011	0.029	96.2%	-0.030	0.030	94.2%
	β_2	0.009	0.008	87.6%	-0.031	0.008	88.2%	-0.058	0.012	80.5%
0.1	β_0	-0.002	0.017	93.8%	-0.003	0.016	95.0%	0.008	0.016	96.1%
	β_1	-0.011	0.032	93.4%	-0.017	0.029	95.9%	-0.039	0.030	94.0%
	β_2	0.009	0.008	88.6%	-0.041	0.008	88.0%	-0.073	0.013	75.5%
1	β_0	0.001	0.017	93.2%	0.009	0.016	95.2%	0.037	0.017	93.4%
	β_1	-0.011	0.032	92.8%	-0.037	0.029	94.8%	-0.087	0.033	91.3%
	β_2	0.001	0.005	94.0%	-0.094	0.014	71.9%	-0.157	0.030	34.8%
Ridge Models		Ignorable			Moderate nonignorable			Strong nonignorable		
λ_0	Parameter	Bias	MSE	95% CP	Bias	MSE	95% CP	Bias	MSE	95% CP
0	β_0	-0.003	0.017	94.5%	-0.007	0.016	95.1%	0.001	0.016	95.8%
	β_1	-0.014	0.032	93.9%	-0.010	0.030	96.4%	-0.023	0.030	95.0%
	β_2	0.012	0.009	86.4%	-0.022	0.008	88.8%	-0.043	0.011	84.8%
0.01	β_0	-0.003	0.017	94.7%	-0.006	0.016	95.6%	0.002	0.016	95.7%
	β_1	-0.013	0.032	93.9%	-0.011	0.029	96.6%	-0.026	0.030	94.4%
	β_2	0.011	0.009	86.8%	-0.024	0.008	88.2%	-0.049	0.011	84.6%
0.05	β_0	-0.002	0.017	94.2%	-0.003	0.016	95.3%	0.006	0.016	96.3%
	β_1	-0.012	0.032	94.2%	-0.014	0.029	96.1%	-0.035	0.029	94.0%
	β_2	0.010	0.008	88.5%	-0.034	0.008	89.1%	-0.067	0.012	81.5%
0.1	β_0	-0.002	0.017	93.2%	-0.001	0.016	95.3%	0.010	0.016	96.0%
	β_1	-0.012	0.032	94.0%	-0.018	0.029	96.1%	-0.043	0.029	93.6%
	β_2	0.009	0.008	89.2%	-0.042	0.008	89.1%	-0.081	0.014	79.0%
1	β_0	0.000	0.017	93.4%	0.005	0.016	95.2%	0.030	0.017	94.0%
	β_1	-0.011	0.032	93.0%	-0.030	0.029	94.8%	-0.076	0.031	92.8%
	β_2	0.004	0.005	94.6%	-0.078	0.011	79.7%	-0.137	0.024	45.9%

Table 6.

Simulation results for Normal outcome with number of repeated measures $n_i = 5$. Bias, mean square error (MSE) and 95% coverage probability (95% CP) for simulations with $\lambda = \lambda_0 / \sqrt{n}$ ($\lambda_0 = 0, 0.01, 0.05, 0.1$, and 1)

LASSO Models		Ignorable			Moderate nonignorable			Strong nonignorable		
λ_0	Parameter	Bias	MSE	95% CP	Bias	MSE	95% CP	Bias	MSE	95% CP
0	β_0	-0.011	0.017	93.4%	-0.015	0.016	94.6%	-0.006	0.016	93.3%
	β_1	0.012	0.026	94.4%	0.016	0.026	95.4%	-0.007	0.025	96.1%
	β_2	0.001	0.001	92.2%	-0.013	0.002	93.4%	-0.026	0.003	89.6%
0.01	β_0	-0.011	0.017	93.4%	-0.015	0.016	94.0%	-0.006	0.016	93.3%
	β_1	0.012	0.026	94.4%	0.016	0.026	95.4%	-0.008	0.025	96.1%
	β_2	0.001	0.001	92.2%	-0.014	0.002	93.2%	-0.026	0.003	89.6%
0.05	β_0	-0.011	0.017	93.4%	-0.016	0.016	93.8%	-0.006	0.016	93.2%
	β_1	0.012	0.026	94.4%	0.015	0.026	95.4%	-0.010	0.025	96.3%
	β_2	0.001	0.001	92.6%	-0.015	0.002	92.8%	-0.028	0.003	89.1%
0.1	β_0	-0.011	0.017	93.4%	-0.016	0.016	93.6%	-0.006	0.016	93.9%
	β_1	0.012	0.026	94.4%	0.014	0.026	95.2%	-0.012	0.025	97.2%
	β_2	0.001	0.001	92.4%	-0.016	0.002	92.8%	-0.031	0.003	88.7%
1	β_0	-0.009	0.017	93.6%	-0.015	0.016	93.2%	0.007	0.015	92.6%
	β_1	0.012	0.026	94.8%	-0.003	0.025	94.8%	-0.049	0.025	95.4%
	β_2	0.000	0.001	93.0%	-0.044	0.004	77.4%	-0.080	0.009	54.3%
Ridge Models		Ignorable			Moderate nonignorable			Strong nonignorable		
λ_0	Parameter	Bias	MSE	95% CP	Bias	MSE	95% CP	Bias	MSE	95% CP
0	β_0	-0.011	0.017	93.4%	-0.015	0.016	94.4%	-0.006	0.016	93.1%
	β_1	0.012	0.026	94.4%	0.016	0.026	95.4%	-0.007	0.025	96.1%
	β_2	0.001	0.001	92.2%	-0.013	0.002	93.4%	-0.026	0.003	89.6%
0.01	β_0	-0.011	0.017	93.0%	-0.015	0.016	94.2%	-0.006	0.016	93.1%
	β_1	0.012	0.026	94.4%	0.016	0.026	95.4%	-0.008	0.025	96.1%
	β_2	0.001	0.001	92.2%	-0.014	0.002	93.4%	-0.027	0.003	89.6%
0.05	β_0	-0.011	0.017	93.4%	-0.016	0.016	93.8%	-0.006	0.016	93.7%
	β_1	0.012	0.026	94.4%	0.016	0.026	95.2%	-0.012	0.025	96.9%
	β_2	0.001	0.001	92.4%	-0.015	0.002	93.0%	-0.031	0.003	89.1%
0.1	β_0	-0.011	0.017	93.2%	-0.016	0.016	93.6%	-0.006	0.016	93.8%
	β_1	0.012	0.026	94.4%	0.014	0.026	95.2%	-0.015	0.025	97.4%
	β_2	0.001	0.001	92.4%	-0.016	0.002	92.2%	-0.035	0.003	89.6%
1	β_0	-0.010	0.017	93.4%	-0.016	0.016	93.4%	0.003	0.015	92.5%
	β_1	0.012	0.026	94.4%	0.004	0.025	95.0%	-0.046	0.025	96.1%
	β_2	0.001	0.001	92.4%	-0.033	0.003	87.8%	-0.077	0.008	54.5%