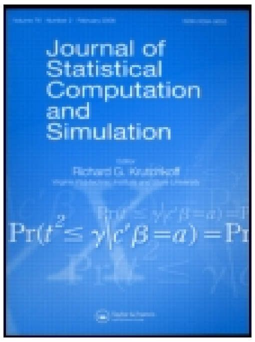

Taylor & Francis
Taylor & Francis Group

Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: https://www.tandfonline.com/loi/gscs20

Performances of Bayesian model selection criteria for generalized linear models with non-ignorably missing covariates

Zeynep Kalaylioglu

Journal of Statistical Computation and Simulation, 2014
Vol. 84, No. 8, 1670–1691, http://dx.doi.org/10.1080/00949655.2012.760089

Performances of Bayesian model selection criteria for generalized linear models with non-ignorably missing covariates

Zeynep Kalaylioglu*

Department of Statistics, Middle East Technical University, Ankara 06530, Türkiye

(*Received 9 February 2012; final version received 15 December 2012*)

This article deals with model comparison as an essential part of generalized linear modelling in the presence of covariates missing not at random (MNAR). We provide an evaluation of the performances of some of the popular model selection criteria, particularly of deviance information criterion (DIC) and weighted L (WL) measure, for comparison among a set of candidate MNAR models. In addition, we seek to provide deviance and quadratic loss-based model selection criteria with alternative penalty terms targeting directly the MNAR models. This work is motivated by the need in the literature to understand the performances of these important model selection criteria for comparison among a set of MNAR models. A Monte Carlo simulation experiment is designed to assess the finite sample performances of these model selection criteria in the context of interest under different scenarios for missingness amounts. Some naturally driven DIC and WL extensions are also discussed and evaluated.

Keywords: penalty; missing not at random; Bayesian inference; non-ignorable missingness model; identifiability

AMS Subject Classification: 62J12

1. Introduction

Model comparison is an essential stage in the generalized linear model (GLM) analysis in the presence of ignorably or non-ignorably missing data. As it is well known now in the missing data literature, missingness is defined by Little and Rubin [1] as being of three types: missing completely at random, missing at random (MAR), and missing not at random (MNAR). When the data are MAR, the mechanism that leads to the missingness can be ignored in the analysis. On the other hand, if the data are believed to be MNAR, then the underlying missingness mechanism should not be ignored and should be modelled. Since the data at hand inevitably lack information regarding the underlying mechanism that caused the missingness, the data analyst is bound to consider different ignorable and non-ignorable missingness modelling schemes and use an elaborate model selection period.

Model comparison criteria such as the deviance information criterion (DIC) [2] and weighted L (which we will call WL from hereon) measure [3] are extended to GLMs with missing covariates [4]. Basically, noting that deviance is based on the response data likelihood, the DIC = posterior deviance + (posterior deviance − a point estimate of deviance) and

*Email: kzeynep@metu.edu.tr

WL = (WL measure for subjects with missing covariate data) + (WL measure for subjects with full data), formal definitions and details of which are presented in Section 3. These criteria are shown to be performing satisfactorily for comparisons between ignorable missingness models and non-ignorable missingness models [4]. On the other hand, as strongly pointed out by Ibrahim et al. [5] and Ibrahim [6], it is dangerous to use any model selection criterion to choose among an aggregate set of MAR and MNAR models and they are to be used to select among a set of candidate MAR models or a set of candidate MNAR models. To the best of our knowledge, the performances of these criteria for comparisons within a class of MNAR models or that of MAR models are yet to be investigated.

The focus of this article in particular is a GLM with fully observed response and non-ignorably missing covariates. The GLM framework in the presence of non-ignorably missing covariate consists of three components: a GLM representing the association between the response and the covariates (the main model of interest), a model for covariates that are subject to missingness, and a model for probability of the covariate being missing conditional on the value of the covariate that would have been observed otherwise (MNAR model). The underlying physical mechanism that has led to missingness is reflected in the data analysis through the MNAR model. However, constructing a missingness model that can capture this mechanism is not an easy task as the data at hand lack sufficient information about the underlying missingness mechanism and henceforth the missingness model assumptions are not verifiable based on the data set at hand. For that reason, in practice a model comparison period is to be conducted in which several non-ignorable models are fit and compared through model comparison measures. Choosing one missingness model over another can change the main model parameter estimates and thus may result in a different statistical inference. Therefore, special care is required while considering missingness model selection period. In this article, we address the following questions. First, how are the performances of the DIC and WL used for comparisons among non-ignorable missing models? Second, can we improve the performances of the model selection criteria in this setting by somehow directly using statistics obtained from the missingness model or by introducing terms directly penalizing the missingness model?

Our first aim is to carry out an extensive simulation experiment that will portray the finite sample behaviours of the DIC and WL in the settings of a GLM with covariates that are MNAR. Both these measures that are employed for model comparison among non-ignorable missingness models are solely based on the response model likelihood although missingness models are the focus models in such comparisons. The missingness models are only indirectly involved in the computation of these criteria. Then, our second aim is to consider an adjustment to each criterion so that the missingness model is directly involved in the computations. Our third aim is to define alternative penalizing terms in DIC computation and assess the performance of the resulting DIC in the context of a GLM with MNAR covariates. The rest of the article is structured as follows: a GLM setting with covariates subject to MNAR is reproduced in Section 2. Section 2 also highlights modelling considerations in the context of interest. Section 3 describes the DIC and WL formulations that are readily used in missing data problems as well as our proposed adjustments to them. In Section 4, a simulation study is presented to evaluate and compare the performances of the criteria under different settings. Section 5 summarizes the findings, discusses the theoretical aspects, and indicates direction for future study.

2. Model

Consider a set of independent observations (y_i, x_i), $i = 1, \ldots, n$, where y_i is the response and $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$ is 1 by p set of covariates for the ith subject. GLMs have been widely used to model the association between y_i and x_i. The probability density function (pdf) of y_i coming from a GLM family was given by Nelder and Wedderburn [7] as

$$p(y_i|x_i, \beta) = \exp\left[\frac{y_i\theta_i - b(\theta_i)}{a_i(\tau)} + c(y_i, \tau)\right], \tag{1}$$

where $\theta_i = \theta(\eta_i)$ is the canonical parameter, τ is the dispersion parameter, and a, b, and c are known functions. For example, for a Bernoulli random variable Y_i, $\tau = 1$ and $a_i(\tau) = 1$. The systematic component of the model is the linear predictor given by $\eta_i = \boldsymbol{x}_i\boldsymbol{\beta}$ and $\boldsymbol{\beta}$ is the $(p + 1)$ by 1 vector of unknown regression coefficients including the intercept. When $\theta_i = x_i\boldsymbol{\beta}$, the link is said to be canonical and Model (1) is said to be the canonical-link model. In this case, we can find a link g so that $g(E(Y_i|x_i)) = \boldsymbol{x}_i\boldsymbol{\beta}$. For example, for a Bernoulli random variable Y_i, the link is the logit link. The parameter vector of focus is $\boldsymbol{\beta}$ as it is related to the association between the response variate and the covariates.

Let K be the number of covariates that are subject to MNAR and $\{k_1, k_2, \ldots, k_K\}$ denote the set of indices of these MNAR covariates. For instance, for an analysis in which X_1, X_2, and X_3 represent the covariates in the study and (X_1,X_3) are subject to MNAR and X_2 is observed for all the subjects, $K = 2$ and $\{k_1, k_2\} = \{1, 3\}$ is the aforementioned set of indices. For the moment, let $\boldsymbol{x}_{i,\text{miss}}$ and $\boldsymbol{x}_{i,\text{obs}}$, respectively, denote the vector of MNAR and fully observed covariates for the ith subject and let $\boldsymbol{x}_i = (\boldsymbol{x}_{i,\text{miss}}, \boldsymbol{x}_{i,\text{obs}})$. Also, let $\boldsymbol{r}_i = (r_{i,k_1}, r_{i,k_2}, \ldots, r_{i,k_K})$ and $r_{i,k_j} = I(x_{i,k_j}$ is observed) for $j = 1, \ldots, K$ denote the missingness indicators related to the covariates subject to MNAR for the ith subject. Also, note that $I(\cdot)$ is a binary indicator function. Then, adapting the selection model approach, the complete data likelihood of subject i can be broken down to three components as the joint pdf of the missingness indicators, the pdf of the response variable, and the joint pdf of the covariates in the following manner:

$$\begin{aligned} \boldsymbol{R}_i &\sim p(\boldsymbol{r}_i|\boldsymbol{x}_i, y_i, \boldsymbol{\phi}) \\ Y_i &\sim p(y_i|\boldsymbol{x}_i, \boldsymbol{\beta}) \\ \boldsymbol{X}_{i,\text{miss}} &\sim p(\boldsymbol{x}_{i,\text{miss}}|\boldsymbol{x}_{i,\text{obs}}, \boldsymbol{\alpha}) \end{aligned} \tag{2}$$

for $i = 1, \ldots, n$, where $\boldsymbol{\phi}$, $\boldsymbol{\beta}$, and $\boldsymbol{\alpha}$ are the parameter vectors characterizing the corresponding probability distributions. For instance, they can be the regression coefficients in each model. The top model in the hierarchy in selection model (2) is the missingness model, the middle one is the response model, and the model at the bottom is the covariate model. A model for the missing covariate of the ith subject is needed as the covariate takes up a random nature when it is missing. The joint distribution of each of the random vectors $\boldsymbol{R}_i$ and $\boldsymbol{X}_{i,\text{miss}}$ can be factorized further into a sequence of conditional densities [8]. For instance, $p(\boldsymbol{r}_i|\boldsymbol{x}_i, y_i, \boldsymbol{\phi}) = p(r_{i,k_K}|r_{i,k_1}, \ldots, r_{i,k_{K-1}}, \boldsymbol{x}_i, y_i, \boldsymbol{\phi})$ $p(r_{i,k_{K-1}}|r_{i,k_1}, \ldots, r_{i,k_{(K-2)}}, \boldsymbol{x}_i, y_i, \boldsymbol{\phi}) \cdots p(r_{i,k_2}|r_{i,k_1}, \boldsymbol{x}_i, y_i, \boldsymbol{\phi})p(r_{i,k_1}|\boldsymbol{x}_i, y_i, \boldsymbol{\phi})$ where $\boldsymbol{\phi}, j = 1, \ldots, K$ is the vector of associated regression coefficients. As r_{i,k_j} is a binary indicator variable, each conditional piece herein can be suitably modelled using a logistic or a probit regression approach. For an MNAR covariate, the linear predictor in its missingness model is necessarily a function of the covariate itself.

Unlike the response or the covariate models, the beliefs that are employed to construct the missingness models are unfortunately unverifiable by the data at hand, with the reason simply being that there is no sufficient information in the data set concerning the underlying mechanism that led to the missingness. As a result, the missingness model parameters may become unidentifiable if a larger missingness model is built. Therefore, in practice, analyses in the presence of non-ignorably missing data entail a model comparison period in which non-ignorable missingness models of different complexity are fit in Model (2). This way the sensitivity of $\boldsymbol{\beta}$ estimates (the main parameters of interest) to the missingness model structure is investigated and missingness models from which the identifiability issue arises are determined and eliminated for further inferential considerations. Since missingness model parameters can easily become unidentifiable, one needs to be meticulous in the model selection process. All these considerations imply that model selection criteria that are employed play an important role. In a broader sense, the study reported in [9] is an example for the necessity of taking account of this intertwining relationship of missing data and model selection in the analyses.

From an application standpoint, a model comparison period in general aims at one of the following: comparing selection models (by selection model, we mean Model (2) as a whole) with different non-ignorable missingness models and the same response and covariate models or comparing selection models with different response models and the same non-ignorable missingness model. The focus model in the first one is the missingness model, whereas in the latter one it is the response model. In this article, we focus on the first one, that is, comparisons of the selection models with different non-ignorable missingness models and the same response model.

3. Bayesian model selection criteria

We consider Bayesian model selection criteria, in particular, the DIC [2] and WL measure [3] that were originally developed for fully observed data. The DIC is based on the deviance function, whereas the WL measure is based on the quadratic loss function. Huang et al. [4] extended these measures to a GLM with missing covariates and investigated their performance for selecting the correct missingness model when the comparison was made between a non-ignorable missingness model and an ignorable one. In the subsequent parts, along with their DIC and WL measure, we reckon the variations of these criteria that were designed specifically for non-ignorable missingness model comparisons.

3.1. *Criteria based on deviance*

The underlying formulation of the DIC can be stated by the following simple combination:

$$\text{DIC} = \text{posterior deviance} + \text{penalty},$$

where the deviance and posterior deviance for a fully observed data set are, respectively, $D(\boldsymbol{\theta}) = -2\log L(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{X})$ and $\widehat{D(\boldsymbol{\theta})} = E(-2\log L(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{X})|\boldsymbol{Y},\boldsymbol{X})$, where $L(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{X})$ is the likelihood function. In the presence of missing data, the deviance can take different forms depending on the form of the likelihood function. In the case of data MAR, such alternative formulations of DIC are considered and evaluated in random-effect models and mixtures of distributions [10].

In the presence of missing covariates, Huang et al. [4] reexpressed the deviance conveniently as a function of the linear predictor $\boldsymbol{\eta} = (\eta_1,\ldots,\eta_n)^{\mathrm{T}}$, that is, $D(\boldsymbol{\eta}) = -2\log L(\boldsymbol{\eta}|\boldsymbol{Y},\boldsymbol{X})$ and $\widehat{D(\boldsymbol{\eta})} = E(-2\log L(\boldsymbol{\eta}|\boldsymbol{Y},\boldsymbol{X})|\boldsymbol{Y},\boldsymbol{X}_{\text{obs}},\boldsymbol{R})$, where $\boldsymbol{Y} = (Y_1, Y_2,\ldots,Y_n)^{\mathrm{T}}$, $\boldsymbol{R} = (\boldsymbol{R}_1^{\mathrm{T}},\boldsymbol{R}_2^{\mathrm{T}},\ldots,\boldsymbol{R}_n^{\mathrm{T}})^{\mathrm{T}}$, $\boldsymbol{X}_{\text{obs}}$ are the observed covariate data and $L(\boldsymbol{\eta}|\boldsymbol{Y},\boldsymbol{X}_{\text{obs}})$ is the observed likelihood function for the response model (middle model) in Equation (2). They penalized the model for model complexity using the penalty component defined by $\widehat{D(\boldsymbol{\eta})} - D(\hat{\boldsymbol{\eta}})$ in which $\hat{\boldsymbol{\eta}} = (\hat{\eta}_1,\ldots,\hat{\eta}_n)^{\mathrm{T}}$, where $\hat{\eta}_i$'s are the posterior means of $\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}$'s and $\hat{\eta}_i = E(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}|\boldsymbol{Y},\boldsymbol{X}_{\text{obs}},\boldsymbol{R})$ for $i = 1,\ldots,n$. Letting $\boldsymbol{D}_{\text{obs}} = (\boldsymbol{Y},\boldsymbol{X}_{\text{obs}},\boldsymbol{R})$ denote the observed data, the first type of DICs that will be investigated in this study for its ability of selecting the selection model with the true non-ignorable missingness model is

$$\text{DIC}_1 = E(-2\log L(\boldsymbol{\eta}|\boldsymbol{Y},\boldsymbol{X}_{\text{obs}})|\boldsymbol{D}_{\text{obs}}) + \text{penalty}_1,$$

where $\text{penalty}_1 = \widehat{D(\boldsymbol{\eta})} - D(\hat{\boldsymbol{\eta}})$ penalizing for model complexity.

In non-ignorably missing data situations, there is a vagueness in the target model that the penalty term is aiming to penalize. The aim of the penalty, whether it is to penalize the selection model in its entirety or just the missingness part, needs to be pinpointed. In this spirit, Mason et al. [11] developed for univariate data subject to MNAR two separate DIC strategies, one of which is for comparing selection models with different missingness models but with the same response model and the other is for comparing selection models with different response models but with the same missingness model. Since our interest currently lies in the comparison among selection models with the same response and covariate model but with different non-ignorable missingness models, the missingness model should, in particular, be penalized. In pursuit of this idea, a natural approach is to define a penalty term penalizing the missingness model for its possible shortcomings and thus the following DICs are proposed:

$$\begin{aligned} \mathrm{DIC}_2 &= E(-2\log L(\boldsymbol{\eta}|\boldsymbol{Y},\boldsymbol{X})|\boldsymbol{D}_{\mathrm{obs}}) + \mathrm{penalty}_2, \\ \mathrm{DIC}_3 &= E(-2\log L(\boldsymbol{\eta}|\boldsymbol{Y},\boldsymbol{X})|\boldsymbol{D}_{\mathrm{obs}}) + \mathrm{penalty}_3, \end{aligned} \tag{3}$$

where the terms $\mathrm{penalty}_2$ and $\mathrm{penalty}_3$ penalize the missingness models for the identifiability problem and are defined as

$$\begin{aligned} \mathrm{penalty}_2 &= \sum_{k=1}^{K} |\bar{I}_R(\hat{\boldsymbol{\phi}}_k)|^{-1}, \\ \mathrm{penalty}_3 &= \sum_{k=1}^{K} \|\bar{I}_R(\hat{\boldsymbol{\phi}}_k)\|\,\|\bar{I}_R^{-1}(\hat{\boldsymbol{\phi}}_k)\|, \end{aligned} \tag{4}$$

where the notations $|\cdot|$ and $\|\cdot\|$ are, respectively, the determinant and norm of a matrix. Here, K is the number of covariates subject to MNAR as stated earlier and $\hat{\boldsymbol{\phi}}_k$ is the vector of the estimated regression coefficients of the kth missingness model. For instance, in the case of two MNAR covariates, $K = 2$, $\hat{\boldsymbol{\phi}}_1$ and $\hat{\boldsymbol{\phi}}_2$ are the estimated coefficient vectors in the logistic (probit) regressions for $P(R_{i1} = 1|\cdot)$ and $P(R_{i2} = 1|R_{i1}, \cdot)$, respectively, where the dot in the conditional part represents other covariates or the response variate that might take part. These penalties are based on $\bar{I}_R(\hat{\boldsymbol{\phi}}_k)$, $k = 1, \ldots, K$, which is the estimated average observed Fisher information matrix for the missingness model corresponding to the kth MNAR covariate. More explicitly,

$$\bar{I}_R(\hat{\boldsymbol{\phi}}_k) = -\frac{\partial^2}{\partial\boldsymbol{\phi}_k\,\partial\boldsymbol{\phi}_k^{\mathrm{T}}}\frac{1}{n}\log L(\boldsymbol{\phi}_k|\boldsymbol{D}_{\mathrm{obs}},\boldsymbol{X}_{\mathrm{miss}}),$$

which is based on the complete data likelihood function for the missingness model evaluated at the posterior means $\hat{\boldsymbol{\phi}}_k = E(\boldsymbol{\phi}_k|\boldsymbol{D}_{\mathrm{obs}})$ and $\widehat{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\phi}_k} = E(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\phi}_k|\boldsymbol{D}_{\mathrm{obs}})$ for $i = 1, \ldots, n$. When $\boldsymbol{x}_i$ is completely observed, $\widehat{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\phi}_k} = \boldsymbol{x}_i^{\mathrm{T}}E(\boldsymbol{\phi}_k|\boldsymbol{D}_{\mathrm{obs}})$ for $i = 1, \ldots, n$.

In designing these penalties, we are motivated by the fact that the local identifiability of the parameters is equivalent to the non-singularity of the information matrix [12]. This condition for local identifiability is also employed as a first step towards checking the identifiability of non-ignorable models for incomplete binary responses [13]. The penalty named as $\mathrm{penalty}_3$ is the condition number for the information matrix, a large value (an ill-conditioned matrix) implying an almost-singular matrix. Both $\mathrm{penalty}_2$ and $\mathrm{penalty}_3$ penalize the missigness models for the proximity of the associated information matrix to singularity. The rationale behind these proposals stemmed from the reasoning that missingness models are the focus in sensitivity analyses conducted on various different selection models with different missingness models and the same response and covariate models and thus the missingness model should contribute to the comparison criteria directly. In spite of the sound conceptual argument, the performances of the proposed

DICs, namely DIC_2 and DIC_3, were mediocre in our simulation experiments and therefore will not be pursued in this article, but will be pursued for a possible improvement in a separate study.

The above notion is also adopted to develop a deviance-based criterion penalizing the response model for the identifiability issue in terms of the inverse-observed Fisher information matrix. It turns out that using the inverse-Fisher information matrix as a penalty term in deviance-based criteria was also considered in greater detail for non-missing data situations in [14] and other Bozdogan references therein. The proposed DICs are as follows:

$$\begin{aligned}\mathrm{DIC}_4 &= E(-2\log L(\boldsymbol{\eta}|\boldsymbol{Y},\boldsymbol{X})|D_{\mathrm{obs}}) + \mathrm{penalty}_4\\ \mathrm{DIC}_5 &= E(-2\log L(\boldsymbol{\eta}|\boldsymbol{Y},\boldsymbol{X})|D_{\mathrm{obs}}) + \mathrm{penalty}_5\end{aligned} \tag{5}$$

with

$$\begin{aligned}\mathrm{penalty}_4 &= |\bar{I}_Y(\hat{\boldsymbol{\beta}})|^{-1}\\ \mathrm{penalty}_5 &= \|\bar{I}_Y(\hat{\boldsymbol{\beta}})\|\|\bar{I}_Y^{-1}(\hat{\boldsymbol{\beta}})\|,\end{aligned} \tag{6}$$

where $\hat{\boldsymbol{\beta}}$ and $\bar{I}_Y(\hat{\boldsymbol{\beta}})$ are, respectively, the vector of the estimated regression coefficients and the estimated average observed Fisher information matrix of the response model. More explicitly,

$$\bar{I}_Y(\hat{\boldsymbol{\beta}}) = -\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}}\frac{1}{n}\log L(\boldsymbol{\beta}|\boldsymbol{Y},\boldsymbol{X}_{\mathrm{obs}},\boldsymbol{X}_{\mathrm{miss}}),$$

which is based on the complete data likelihood function for the response model evaluated at the posterior means $\hat{\boldsymbol{\beta}} = E(\boldsymbol{\beta}|\boldsymbol{D}_{\mathrm{obs}})$ and $\widehat{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}} = E(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}|\boldsymbol{D}_{\mathrm{obs}})$ for $i = 1,\ldots,n$. When $\boldsymbol{x}_i$ is completely observed, $\widehat{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\phi}_k} = \boldsymbol{x}_i^{\mathrm{T}}E(\boldsymbol{\phi}_k|\boldsymbol{D}_{\mathrm{obs}})$ for $i = 1,\ldots,n$.

3.2. *Criteria based on the quadratic loss function*

Weighted quadratic loss L measure is based on weighting the observations and assessing the models through the prediction abilities of the models [3]. Huang et al. [4] extended the WL measure to a GLM with missing covariates in the following manner:

$$\mathrm{WL}_1 = \mathrm{WL}_{1,\mathrm{miss}} + \mathrm{WL}_{1,\mathrm{obs}},$$

where $\mathrm{WL}_{1,\mathrm{miss}}$ and $\mathrm{WL}_{1,\mathrm{obs}}$ are the WL measures computed based on subjects with at least one missing covariate and the subjects with fully observed data, respectively, and defined as

$$\begin{aligned}\mathrm{WL}_{1,\mathrm{miss}} &= \nu\sum_{\{i:i\in A_{\mathrm{miss}}\}} w_{1,i}(\mu_i - y_i)^2 + \sum_{\{i:i\in A_{\mathrm{miss}}\}} w_{1,i}\,\mathrm{Var}(z_i|y_i)\\ \mathrm{WL}_{1,\mathrm{obs}} &= \nu\sum_{\{i:i\in A_{\mathrm{obs}}\}} w_{1,i}(E_{\boldsymbol{\beta}|D_{\mathrm{obs}}}(b'(\theta_i)) - y_i)^2\\ &\quad + \sum_{\{i:i\in A_{\mathrm{obs}}\}} w_{1,i}[E_{\boldsymbol{\beta}|D_{\mathrm{obs}}}(b''(\theta_i)) - \{E_{\boldsymbol{\beta}|D_{\mathrm{obs}}}(b'(\theta_i))\}^2 + E_{\boldsymbol{\beta}|D_{\mathrm{obs}}}(b'(\theta_i)^2)],\end{aligned} \tag{7}$$

where A_{miss} is the set of subjects with at least one missing covariate value, A_{obs} is the set of subjects whose covariate data are fully observed, $w_{1,i}$ is the weight function and is equal to $1/b''(\theta(\widehat{wx_i^{\mathrm{T}}\boldsymbol{\beta}}))$ for the members of A_{miss} and it is $1/b''(\theta(wx_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}))$ for the members of A_{obs}, $0 \leq w \leq 1$, $0 < \nu < 1$, $b'(\cdot)$ and $b''(\cdot)$ are, respectively, the mean and variance functions of the response GLM in Model (1),

z_i is a future response for the ith subject in a replicate experiment having the same sampling density of y_i, and $\mu_i = E(z_i|\boldsymbol{D}_{\text{obs}})$. In this formula, the prediction variance term plays the role of a penalty term. Each of these quantities is estimated using the posterior expectations approximated by the sample path average of the associated Markov chain once the samples from the posterior distribution are obtained. The details of the estimation procedure have been laid out previously in the literature and can be found in, for example, [4]. The performance of this measure in the comparison of the selection models with MAR models and those with MNAR models was investigated therein. The performance of the measure for comparison among a family of selection models with different non-ignorable missingness models is yet to be investigated. In the current article, we explore this issue.

The above formulation is solely based on the response model. Since the model of focus in model comparisons in this article is associated with the missingness part, one would naturally be curious to know if the WL measure can be adjusted so that the missingness model has a direct effect on the WL outcome and how well the adjusted WL measure would perform in comparing the selection models with different non-ignorable missingness models. As an adjustment, subjects that are substantially influenced by the badness of fit of the missingness model are proposed to be assigned more weight in WL computations. Towards this end, subjects are weighted according to the magnitude of their absolute residual from the regression model predicting the missingness probability (i.e. the missingness model). The motivation for this choice of weight function stems from the need for quantifying and evaluating the effect of problematic missingness model fit on the prediction qualities of the response model and comparing the selection models by taking this information into account. The proposed WL measure is as follows:

$$\mathrm{WL}_2 = \mathrm{WL}_{2,\text{miss}} + \mathrm{WL}_{2,\text{obs}},$$

where $\mathrm{WL}_{2,\text{miss}}$ and $\mathrm{WL}_{2,\text{obs}}$ are the WL measures computed based on subjects with at least one missing covariate and subjects with fully observed data, respectively, and defined as

$$\begin{aligned}
\mathrm{WL}_{2,\text{miss}} &= \nu \sum_{\{i:i\in A_{\text{miss}}\}} w_{2,i}(\mu_i - y_i)^2 + \sum_{\{i:i\in A_{\text{miss}}\}} w_{2,i}\,\mathrm{Var}(z_i|y_i) \\
\mathrm{WL}_{2,\text{obs}} &= \nu \sum_{\{i:i\in A_{\text{obs}}\}} w_{2,i}(E_{\boldsymbol{\beta}|D_{\text{obs}}}(b'(\theta_i)) - y_i)^2 \\
&\quad + \sum_{\{i:i\in A_{\text{obs}}\}} w_{2,i}[E_{\boldsymbol{\beta}|\boldsymbol{D}_{\text{obs}}}(b''(\theta_i)) - \{E_{\boldsymbol{\beta}|\boldsymbol{D}_{\text{obs}}}(b'(\theta_i))\}^2 + E_{\boldsymbol{\beta}|\boldsymbol{D}_{\text{obs}}}(b'(\theta_i)^2)].
\end{aligned} \tag{8}$$

Here, the weight function $w_{2,i}$ is based on Pearson's residual and defined as follows:

$$w_{2,i} = \sum_{k=1}^{K} \frac{|r_{i,k_j} - E(r_{i,k_j}|\boldsymbol{D}_{\text{obs}})|}{\sqrt{\mathrm{Var}(r_{i_j,k}|\boldsymbol{D}_{\text{obs}})}},$$

where r_{i,k_j} is the binary missingness indicator associated with the MNAR covariate k_j for subject i as described earlier and $E(r_{i,k}|\boldsymbol{D}_{\text{obs}})$ is estimated by $\Phi(wx_{ij}^{*\mathrm{T}}\hat{\boldsymbol{\phi}}_j)$ for subjects with fully observed data and by $\Phi(w\widehat{\boldsymbol{x}_{ij}^{*\mathrm{T}}\boldsymbol{\phi}_j})$ for subjects who have at least one missing covariate. Here, $\Phi(\cdot)$ is the standard normal cumulative distribution function, $\boldsymbol{x}_{ij}^{*\mathrm{T}}$ is the vector of explanatory variables involved in missingness model k_j (it may include the response variate and other missingness indicators in addition to the MNAR covariate x_{k_j} itself), and $\hat{\boldsymbol{\phi}}_j$ is the estimated coefficient in missingness model k_j and obtained as $E(\boldsymbol{\phi}_j|\boldsymbol{D}_{\text{obs}})$, that is, the posterior expectation of $\boldsymbol{\phi}_j$. Similarly, $\widehat{\boldsymbol{x}_{ij}^{*\mathrm{T}}\boldsymbol{\phi}_j}$ is obtained as $E(\boldsymbol{x}_{ij}^{*\mathrm{T}}\boldsymbol{\phi}_j|\boldsymbol{D}_{\text{obs}})$, that is, the sample path average of the Markov chain constructed for $\boldsymbol{x}_{ij}^{*\mathrm{T}}\boldsymbol{\phi}_j$. Also, $\mathrm{Var}(r_{i,k_j}|\boldsymbol{D}_{\mathrm{obs}})$ is estimated by $\Phi(w\boldsymbol{x}_{ij}^{*\mathrm{T}}\hat{\boldsymbol{\phi}}_j)(1-\Phi(w\boldsymbol{x}_{ij}^{*\mathrm{T}}\hat{\boldsymbol{\phi}}_j))$ and $\Phi(\widehat{w\boldsymbol{x}_{ij}^{*\mathrm{T}}\boldsymbol{\phi}_j})(1-\Phi(\widehat{w\boldsymbol{x}_{ij}^{*\mathrm{T}}\boldsymbol{\phi}_j}))$, respectively, for subjects with fully observed data and subjects with at least one missing covariate. Basically, $w_{2,i}$ is the amount of residuals accumulated over the K missingness models for subject i and thus can be regarded as a measure for the overall badness of fit of the non-ignorable missingness model structure for subject i.

The computation of the DIC is based on the posterior deviance and the computation of the WL measure is based on the bias and the sampling variance of the prediction. They possess attractive computational properties in both non-missing and missing data situations: they can be calculated using the output of the Markov chain Monte Carlo (MCMC) machinery that is used for the Bayesian analysis. Their computation can be automated with the addition of a couple of lines in a WinBUGS code that performs the Bayesian estimation using the MCMC technique and a code written in, for example, R or Matlab as a back-end to process the associated MCMC output into these measures.

4. Simulation study

In this section, we investigate and evaluate the performances of the DIC and the WL measure proposed by Huang et al. [4] and the intrinsically determined versions of these criteria as proposed in Section 3. Two metrics are used to evaluate the performances: (1) the average criterion value and (2) the percentage of times the criterion inferred to the true missingness model. We consider an always-observed binary response variable, y, and a logistic regression to model its dependence on the two covariates x_1 and x_2. For all our simulations, we generated $n = 250$ independent response variates from the model $\mathrm{logit}(P(Y_i = 1|x_{i1}, x_{i2})) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$, where X_{i1}'s are independently and identically generated from Normal$(\alpha_{11}, \alpha_{12})$ and X_{i2}'s are the dichotomous random variables generated from Bernoulli(p_i) with $\mathrm{logit}(p_i) = \alpha_{21} + \alpha_{22}x_{i1}$. The true values of the parameters are set at $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^{\mathrm{T}} = (1, 1, -0.5)^{\mathrm{T}}$, $\boldsymbol{\alpha}_1 = (\alpha_{11}, \alpha_{12})^{\mathrm{T}} = (0.5, 0.25)^{\mathrm{T}}$, and $\boldsymbol{\alpha}_2 = (\alpha_{21}, \alpha_{22})^{\mathrm{T}} = (1, -1)^{\mathrm{T}}$. Two scenarios are considered in terms of the number of covariates subject to MNAR. The first scenario assumes that only one covariate, namely x_1, is subject to MNAR and the covariate x_2 is fully observed. In the second scenario, both are assumed to be subject to being non-ignorably missing. For each scenario, three distinct sets of simulation experiments are carried out. In the first set of experiments, the true missingness mechanism is taken to be M_1. That is, missingness indicators for MNAR covariates are generated from Bernoulli distributions with the probability of missingness as given in M_1. Similarly, for the second and third sets of experiments, the true missingness mechanisms are taken to be M_2 and M_3, respectively, and the missingness indicators are simulated accordingly. Two different missingness proportions, namely 15% and 25%, are considered for each set of experiments. For the first scenario below, these are the proportions of subjects in the data set with missing x_1, whereas for the second one, they are the proportions of subjects who are missing at least one of the two covariate data. In the GLM analysis part of each experiment, three different missingness models are assumed and model comparison is conducted over a set of candidate selection models consisting of the same response and covariate models but different non-ignorable missingness models, namely M_1, M_2, or M_3. Probit regressions are used in the analyses to model these non-ignorable missingness models.

One hundred data sets are simulated for each scenario. For the Bayesian analysis of each simulated data set, we used improper uniform priors for the regression coefficients and the location parameters, namely $\boldsymbol{\beta}$, $\boldsymbol{\alpha}_2$, and α_{11}, as suggested in [4]. For these priors, the propriety of the joint posterior distribution and the conditions under which the propriety is satisfied are given in the same article. For the scale parameter α_{12}, the inverse of Gamma(0.01, 0.01) prior is used. A robust Multivariate Normal prior is used for $(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)$ where the hyperparameters, namely the mean vectors and the variance–covariance matrices, are obtained based on an empirical Bayes approach [4]. In the empirical Bayes procedure, we set the coefficient that accounts for the randomness of the procedure at 1000. As shown in [4] simulation wise, the DIC and the WL measure are robust to the choice of this quantity. In order to derive the posterior inferences, we employed the Gibbs sampling scheme described in [4], in which a vector of latent variables is introduced for each R_i (a technique introduced by Albert and Chib [15]). WinBUGS 1.4.3 is used to carry out the Gibbs sampling and hence obtain the posterior distributions. Other Bayesian technicalities are as follows: the period of the first 2000 iterations is used as burn-in. Further 200,000 iterations are run and every 10th of them is used for posterior inference. The convergence of the distributions of the full conditional draws to the target distributions is confirmed by various diagnostic tools. Autocorrelation plots ensure that the dependence of a draw from a full conditional density upon the previous draws vanishes satisfactorily. The Brooks–Gelman–Rubin statistic [16] ensures that the mean of each chain is a satisfactory approximation to the corresponding posterior expectation.

4.1. *Scenario I: one covariate subject to MNAR*

The non-ignorable missingness models considered are as follows:

$$\begin{aligned} M_1 &: P(R_{i1}=1)=\Phi(\phi_{10}+\phi_{11}x_{i1}) \\ M_2 &: P(R_{i1}=1)=\Phi(\phi_{10}+\phi_{11}x_{i1}+\phi_{12}x_{i2}+\phi_{13}y_i) \\ M_3 &: P(R_{i1}=1)=\Phi(\phi_{10}+\phi_{11}x_{i1}+\phi_{12}x_{i2}+\phi_{13}y_i+\phi_{14}x_{i1}y_i), \end{aligned} \tag{9}$$

where $\Phi(\cdot)$ is the standard Normal cumulative distribution function. Each model is nested in the succeeding model, that is, $M_1 \subset M_2 \subset M_3$. The dimensions of the parameter space of these models are 2, 4, and 5, respectively. Let $\boldsymbol{\phi}_1$ be the true coefficient vector for model $P(R_{i1}=1)$. For M_1, $\boldsymbol{\phi}_1=(\phi_{10},\phi_{11})^{\mathrm{T}}=(0.4,2.5)^{\mathrm{T}}$ corresponds to 15% and $\boldsymbol{\phi}_1=(\phi_{10},\phi_{11})^{\mathrm{T}}=(,)^{\mathrm{T}}$ corresponds to 25% missingness. For M_2, $\boldsymbol{\phi}_1=(\phi_{10},\phi_{11})^{\mathrm{T}}=(0.5,-1.5,1.1,1.45)^{\mathrm{T}}$ corresponds to 15% and $\boldsymbol{\phi}_1=(\phi_{10},\phi_{11})^{\mathrm{T}}=(,)^{\mathrm{T}}$ corresponds to 25%. For M_3, $\boldsymbol{\phi}_1=(\phi_{10},\phi_{11})^{\mathrm{T}}=(-0.6,-0.7,0.8,2.2,2.0)^{\mathrm{T}}$ correspond to 15% and $\boldsymbol{\phi}_1=(\phi_{10},\phi_{11})^{\mathrm{T}}=(-1.45,-0.7,0.8,2.2,2.0)^{\mathrm{T}}$ correspond to 25% missingness. This scenario is representative of the situations where only a single dimension of the covariate space is subject to MNAR.

4.2. *Scenario II: both covariates subject to MNAR*

The non-ignorable missingness models considered are as follows:

$$\begin{aligned} M_1 &: P(R_{i1}=1)=\Phi(\phi_{10}+\phi_{11}x_{i1}) \\ &\quad\; P(R_{i2}=1)=\Phi(\phi_{20}+\phi_{21}x_{i2}) \\ M_2 &: P(R_{i1}=1)=\Phi(\phi_{10}+\phi_{11}x_{i1}+\phi_{12}x_{i2}+\phi_{13}y_i) \\ &\quad\; P(R_{i2}=1)=\Phi(\phi_{20}+\phi_{21}x_{i1}+\phi_{22}x_{i2}+\phi_{23}y_i+\phi_{24}r_{i1}) \\ M_3 &: P(R_{i1}=1)=\Phi(\phi_{10}+\phi_{11}x_{i1}+\phi_{12}x_{i2}+\phi_{13}y_i+\phi_{14}x_{i1}y_i) \\ &\quad\; P(R_{i2}=1)=\Phi(\phi_{20}+\phi_{21}x_{i1}+\phi_{22}x_{i2}+\phi_{23}y_i+\phi_{24}r_{i1}+\phi_{25}x_{i2}y_i), \end{aligned} \tag{10}$$

with $\Phi(\cdot)$ representing the standard Normal cumulative distribution function. The dimensions of the parameter space of these models are 4, 9, and 11 respectively. Let $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ be the true coefficient vectors for models $P(R_{i1}=1)$ and $P(R_{i2}=1)$, respectively, used in missingness indicator generation. For M_1, $\boldsymbol{\phi}_1=(\phi_{10},\phi_{11})^{\mathrm{T}}=(0.4,2.5)^{\mathrm{T}}$ and $\boldsymbol{\phi}_2=(\phi_{20},\phi_{21})^{\mathrm{T}}=(2.3,2.5)^{\mathrm{T}}$

correspond to 15% and $\boldsymbol{\phi}_1 = (\phi_{10}, \phi_{11})^{\mathrm{T}} = (0.4, 2.5)^{\mathrm{T}}$ and $\boldsymbol{\phi}_2 = (\phi_{20}, \phi_{21})^{\mathrm{T}} = (0.5, 2.5)^{\mathrm{T}}$ correspond to 25% missingness. For M_2, $\boldsymbol{\phi}_1 = (\phi_{10}, \phi_{11}, \phi_{12}, \phi_{13})^{\mathrm{T}} = (0.5, -0.7, 0.8, 2.2)^{\mathrm{T}}$ and $\boldsymbol{\phi}_2 = (\phi_{20}, \phi_{21}, \phi_{22}, \phi_{23}, \phi_{24})^{\mathrm{T}} = (1.5, -1.2, 1.0, 0.3, -0.1)^{\mathrm{T}}$ correspond to 15% and $\boldsymbol{\phi}_1 = (\phi_{10}, \phi_{11}, \phi_{12}, \phi_{13})^{\mathrm{T}} = (0.5, -1.0, 1.2, 1.5)^{\mathrm{T}}$ and $\boldsymbol{\phi}_2 = (\phi_{20}, \phi_{21}, \phi_{22}, \phi_{23}, \phi_{24})^{\mathrm{T}} = (1.2, -1.0, 1.0, 0.3, -0.5)^{\mathrm{T}}$ correspond to 25% missingness. For M_3, $\boldsymbol{\phi}_1 = (\phi_{10}, \phi_{11}, \phi_{12}, \phi_{13}, \phi_{14})^{\mathrm{T}} = (-0.05, 1.9, 0.8, 2.2, 2.0)^{\mathrm{T}}$ and $\boldsymbol{\phi}_2 = (\phi_{20}, \phi_{21}, \phi_{22}, \phi_{23}, \phi_{24}, \phi_{25})^{\mathrm{T}} = (1.6, -1.2, 1.0, 0.3, -0.1, 0.2)^{\mathrm{T}}$ correspond to 15% and $\boldsymbol{\phi}_1 = (\phi_{10}, \phi_{11}, \phi_{12}, \phi_{13}, \phi_{14})^{\mathrm{T}} = (-0.45, -0.7, 0.8, 2.2, 2.0)^{\mathrm{T}}$ and $\boldsymbol{\phi}_2 = (\phi_{20}, \phi_{21}, \phi_{22}, \phi_{23}, \phi_{24}, \phi_{25})^{\mathrm{T}} = (1.3, -1.2, 1.0, 0.3, -0.1, 0.2)^{\mathrm{T}}$ correspond to 25% missingness. The case where the true missingness mechanism is M_2 and the missingness amount is 25% is the same scenario studied by Huang et al. [4]. In the tables that follow, DIC_1 and WL_1 results corresponding to (missingness = 25%, true missingness model = M_2) are thus comparable to their results.

4.3. *Results*

The box plots of the posterior means of $\boldsymbol{\beta}$ are constructed based on 100 simulated data sets. Only the ones corresponding to scenario II with 25% missingness are shown in Figures 1–3 as they provide a fairly good representation of the other situations considered herein. In each plot, a line across the true parameter value is drawn for convenience. According to the results, the inference on the intercept is insensitive to the assumed missingness modelling structure, whereas the inference on the coefficients associated with the covariates is affected by it. The three figures together imply that the posterior mean of β_2, that is, the coefficient associated with the binary covariate subject

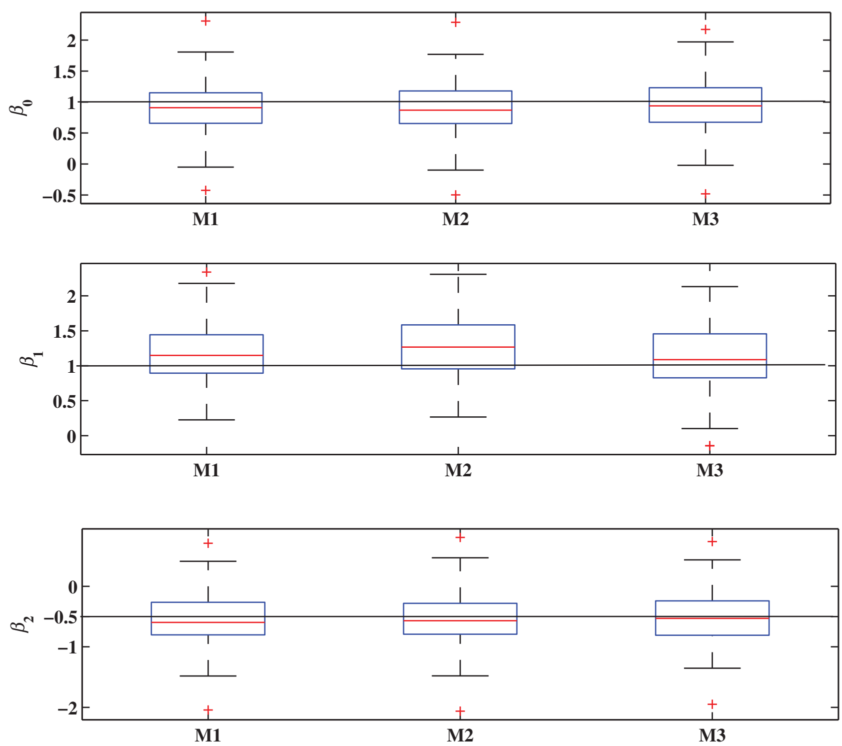

Figure 1. Box plots of the posterior means of β_0, β_1, and β_2 (25% missingness). The true missingness model is M_1. Two MNAR covariates.

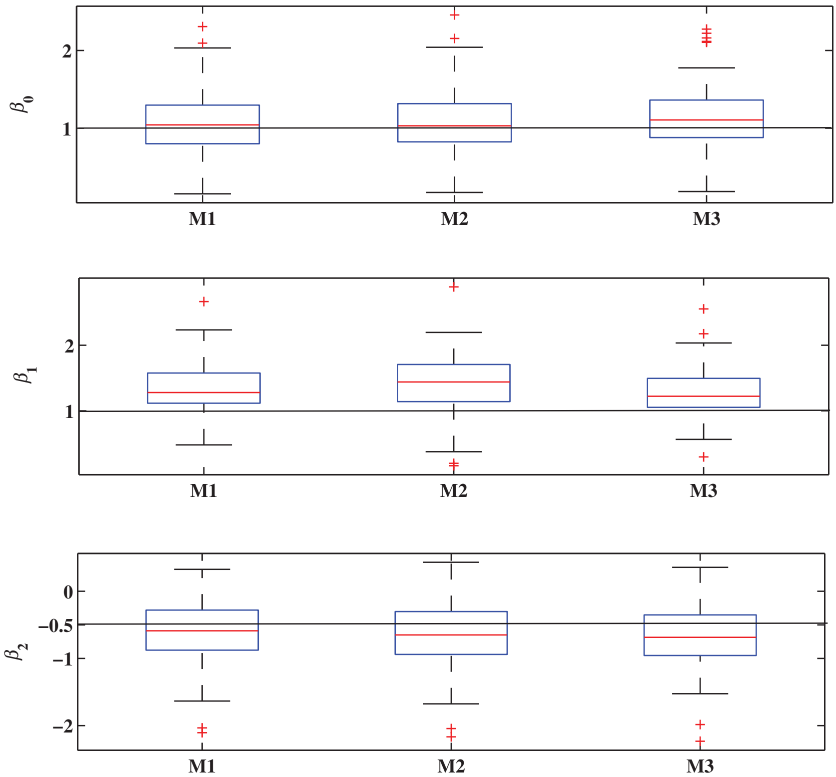

Figure 2. Box plots of the posterior means of β_0, β_1, and β_2 (25% missingness). The true missingness model is M_2. Two MNAR covariates.

to MNAR (X_2), is quite close to the true coefficient value when the assumed missingness model coincides with the true missingness model. On the other hand, the posterior mean of β_1, that is, the coefficient associated with the continuous covariate subject to MNAR (X_1), is closest to its true value when the assumed missingness model is the most saturated one in the set of candidate missingness models, namely M_3. In a simulation experiment, we found that this occurs when a highly non-informative prior such as the inverse of Gamma(0.01, 0.01) is used for the scale parameter α_{12} for which extreme values may be sampled for missing X_{i1}'s in the Gibbs sampling. This may also explain the low coverage probability of the 95% highest posterior density interval for β_1 that was observed by Huang et al. [4].

The sensitivity of posterior inference to the assumed missingness model is also observed for 15% missingness, but to a much lesser extent, implying that the posterior inference on $\boldsymbol{\beta}$ is rather robust to the underlying true missingness mechanism when the proportion of subjects with missing covariate information is moderate.

The results of the performances of the model selection criteria under scenarios I and II are presented in Tables 1–10 and 11–20, respectively. As the results are uniform over the scenarios and the missingness percentages, the following observations hold for all the scenarios considered. Tables 1 and 11 present the averages of deviance, penalty$_1$, penalty$_4$, penalty$_5$, DIC$_1$, DIC$_4$, and DIC$_5$ over 100 Monte Carlo replications. According to the results, deviance, penalty$_5$, DIC$_1$, and DIC$_5$ unequivocally yield the smallest values when M_1 is used in the analysis regardless of the true missingness model. This indicates that these criteria have a tendency for pointing to a smaller missingness model structure no matter what the true missingness structure is. On the other hand, overall, penalty$_1$, penalty$_4$, and DIC$_4$ tend to point to the largest missingness model structure.

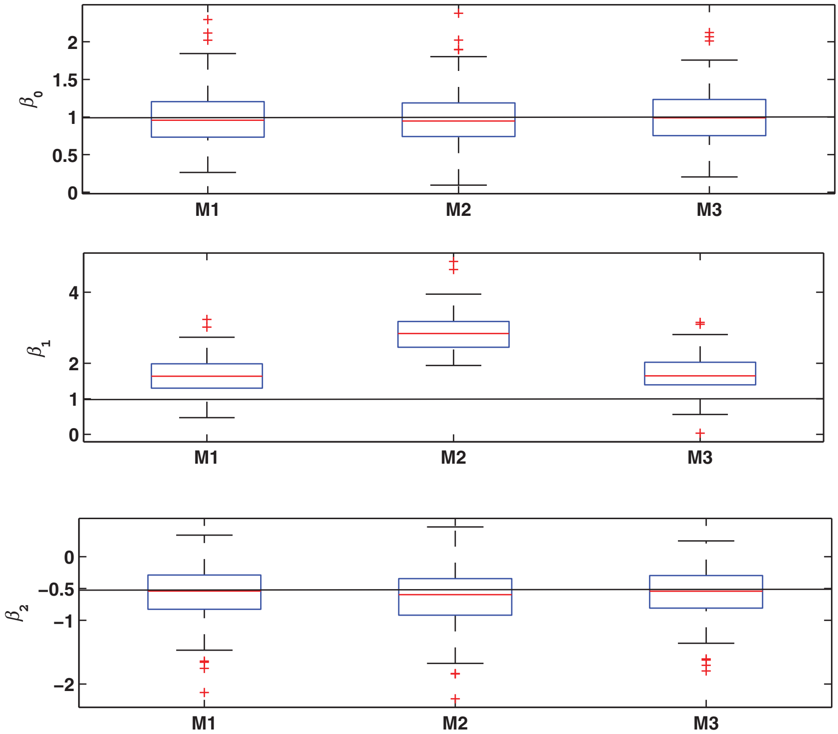

Figure 3. Box plots of the posterior means of β_0, β_1, and β_2 (25% missingness). The true missingness model is M_3. Two MNAR covariates.

Table 1. Monte Carlo averages of the deviance, penalty, and DIC estimates.

Missing	True model	Fitted model	Deviance	penalty_1	penalty_4	penalty_5	DIC_1	DIC_4	DIC_5
15%	M_1	M_1	265.50	5.25	7648.14	23.35	270.75	7913.64	288.85
		M_2	267.22	4.84	7125.20	23.51	272.06	7392.41	290.73
		M_3	267.97	5.51	7000.92	23.77	273.48	7268.89	291.73
	M_2	M_1	248.27	7.43	8483.06	14.84	255.71	8731.33	263.12
		M_2	251.58	6.96	8133.66	15.68	258.54	8385.24	267.25
		M_3	256.30	6.99	7382.59	15.89	263.29	7638.89	272.19
	M_3	M_1	188.82	15.50	19,634.75	13.55	204.31	19,823.57	202.36
		M_2	245.38	13.44	8966.56	17.90	258.82	9211.95	263.29
		M_3	243.86	15.67	9165.05	17.91	259.52	9408.91	261.77
25%	M_1	M_1	265.25	6.54	9163.54	28.23	271.79	9428.79	293.48
		M_2	267.82	5.67	8273.73	28.76	273.49	8541.55	296.58
		M_3	267.59	7.33	8258.50	29.33	274.92	8526.09	296.93
	M_2	M_1	226.46	32.77	16,659.95	20.35	259.23	16,886.40	246.81
		M_2	245.82	24.02	16,065.53	21.74	269.84	16,311.35	267.56
		M_3	234.59	35.59	21,313.20	21.80	270.18	21,547.79	256.39
	M_3	M_1	149.77	28.21	86,650.57	25.47	177.98	86,800.34	175.24
		M_2	224.45	27.52	17,765.98	20.46	251.97	17,990.43	244.92
		M_3	239.53	26.88	13,151.84	23.97	266.41	13,391.37	263.51

Note: One MNAR covariate.

Table 2. Monte Carlo averages of the WL estimates (15% missingness).

WL	w	Fitted model	$\nu = 0.2$	$\nu = 0.4$	$\nu = 0.5$	$\nu = 0.6$	$\nu = 0.8$
WL_1	0.4	M_1	221.99	258.34	276.52	294.69	331.05
		M_2	223.23	259.91	278.24	296.58	333.26
		M_3	224.27	260.94	279.27	297.61	334.27
	0.5	M_1	229.38	266.92	285.69	304.45	341.99
		M_2	230.58	268.41	287.33	306.25	344.08
		M_3	231.60	269.41	288.32	307.22	345.03
	0.6	M_1	238.86	277.92	297.44	316.97	356.03
		M_2	239.97	279.29	298.94	318.60	357.92
		M_3	240.96	280.23	299.86	319.50	358.77
WL_2	0.4	M_1	43.88	51.12	54.74	58.36	65.60
		M_2	42.93	50.27	53.94	57.60	64.94
		M_3	42.64	49.81	53.40	56.99	64.16
	0.5	M_1	42.37	49.38	52.88	56.38	63.39
		M_2	41.16	48.25	51.79	55.34	62.43
		M_3	40.74	47.63	51.08	54.52	61.41
	0.6	M_1	41.09	47.90	51.30	54.71	61.51
		M_2	39.63	46.50	49.94	53.37	60.25
		M_3	39.09	45.72	49.04	52.36	58.99

Notes: The true missingness model is M_1. One MNAR covariate.

Table 3. Monte Carlo averages of the WL estimates (25% missingness).

WL	w	Fitted model	$\nu = 0.2$	$\nu = 0.4$	$\nu = 0.5$	$\nu = 0.6$	$\nu = 0.8$
WL_1	0.4	M_1	222.50	258.58	276.62	294.65	330.73
		M_2	242.02	260.81	279.11	297.41	334.02
		M_3	224.96	261.27	279.42	297.58	333.89
	0.5	M_1	229.89	267.11	285.72	304.34	341.56
		M_2	231.53	269.26	288.13	306.99	344.73
		M_3	232.35	269.78	288.49	307.20	344.62
	0.6	M_1	239.36	278.04	297.39	316.73	355.42
		M_2	240.87	280.04	299.63	319.22	358.39
		M_3	241.79	280.63	300.05	319.47	358.31
WL_2	0.4	M_1	48.30	56.19	60.14	64.09	71.98
		M_2	47.57	55.59	59.59	63.60	71.61
		M_3	47.04	54.77	58.63	62.50	70.22
	0.5	M_1	47.55	55.33	59.22	63.11	70.88
		M_2	46.57	54.45	58.40	62.34	70.23
		M_3	45.88	53.42	57.20	60.97	68.52
	0.6	M_1	46.91	54.59	58.43	62.27	69.96
		M_2	45.62	53.46	57.34	61.23	69.00
		M_3	44.83	52.22	55.91	59.60	66.99

Notes: The true missingness model is M_1. One MNAR covariate.

Tables 2–7 and 12–17 present the simulation results of the WL measures for scenarios I and II, respectively. In all the tables, we can see that M_1 consistently yields the smallest WL_1, whereas M_3 yields the smallest WL_2 regardless of w and ν. This implies that WL_1 displays a consistent tendency towards selecting the smallest missingness model, whereas WL_2 consistently selects the largest missingness model. The results demonstrate that the behaviour of the WL measure is sensitive to the choice of the weight function.

Table 4. Monte Carlo averages of the WL estimates (15% missingness).

WL	w	Fitted model	$\nu = 0.2$	$\nu = 0.4$	$\nu = 0.5$	$\nu = 0.6$	$\nu = 0.8$
WL_1	0.4	M_1	208.63	242.22	259.02	275.81	309.40
		M_2	211.19	245.29	262.34	279.40	313.50
		M_3	215.24	249.97	267.34	284.71	319.45
	0.5	M_1	216.66	251.37	268.73	286.09	320.80
		M_2	219.13	254.32	271.92	289.25	324.71
		M_3	223.02	258.84	276.75	294.66	330.48
	0.6	M_1	227.15	263.31	281.39	299.48	335.64
		M_2	229.45	266.05	284.35	302.65	339.25
		M_3	233.08	270.29	288.89	307.50	344.71
WL_2	0.4	M_1	40.84	47.52	50.85	54.19	60.87
		M_2	37.43	43.83	47.02	50.22	56.62
		M_3	37.45	43.97	47.23	50.49	57.01
	0.5	M_1	39.39	45.84	49.06	52.29	58.73
		M_2	35.35	41.44	44.49	47.54	53.64
		M_3	35.17	41.38	44.48	47.59	53.80
	0.6	M_1	38.14	44.39	47.52	50.64	56.89
		M_2	33.58	39.42	42.34	45.25	51.09
		M_3	33.22	39.16	42.13	45.09	51.04

Notes: The true missingness model is M_2. One MNAR covariate.

Table 5. Monte Carlo averages of the WL estimates (25% missingness).

WL	w	Fitted model	$\nu = 0.2$	$\nu = 0.4$	$\nu = 0.5$	$\nu = 0.6$	$\nu = 0.8$
WL_1	0.4	M_1	204.47	231.33	244.75	258.18	285.03
		M_2	216.10	246.85	262.23	277.60	308.35
		M_3	213.95	241.99	256.02	270.04	298.09
	0.5	M_1	214.76	242.58	256.49	270.41	298.24
		M_2	225.17	256.77	272.57	288.37	319.97
		M_3	225.28	254.19	268.64	283.09	312.00
	0.6	M_1	228.43	257.54	272.09	286.64	315.75
		M_2	237.00	269.69	286.03	302.38	335.06
		M_3	240.49	270.52	285.53	300.54	330.57
WL_2	0.4	M_1	41.95	47.56	50.37	53.17	58.79
		M_2	41.90	47.87	50.85	53.83	59.80
		M_3	39.70	44.92	47.53	50.14	55.37
	0.5	M_1	41.00	46.49	49.24	51.98	57.47
		M_2	40.45	46.20	49.07	51.94	57.68
		M_3	38.08	43.06	45.55	48.05	53.03
	0.6	M_1	40.17	45.56	48.26	50.95	56.34
		M_2	39.22	44.77	47.55	50.33	55.87
		M_3	36.69	41.46	43.85	46.24	51.02

Notes: The true missingness model is M_2. One MNAR covariate.

We also examined what percentage of time that these criteria point to the true missingness model. This is calculated by ((number of times the criterion was smallest for the true missingness model)/100) and the results are given in Tables 8–10 and 18–20 for scenarios I and II, respectively. Our simulation results show that when comparison is made among a set of selection models with non-ignorable missingness models, the abilities of the criteria for selecting the true MNAR model are quite unsatisfactory.

Table 6. Monte Carlo averages of the WL estimates (15% missingness).

WL	w	Fitted model	$\nu = 0.2$	$\nu = 0.4$	$\nu = 0.5$	$\nu = 0.6$	$\nu = 0.8$
WL_1	0.4	M_1	162.20	186.52	198.67	210.83	235.15
		M_2	209.05	240.99	256.95	272.92	304.85
		M_3	209.17	240.64	256.37	272.10	303.57
	0.5	M_1	172.30	198.34	211.37	224.39	250.43
		M_2	217.21	250.01	266.42	282.82	315.63
		M_3	217.64	250.00	266.19	282.37	314.73
	0.6	M_1	186.40	214.89	229.15	243.39	271.89
		M_2	227.92	261.86	278.83	295.80	329.74
		M_3	228.83	262.36	279.12	295.88	329.40
WL_2	0.4	M_1	29.99	34.39	36.58	38.78	43.17
		M_2	28.69	34.38	37.22	40.06	45.75
		M_3	27.76	33.22	35.96	38.69	44.16
	0.5	M_1	28.67	32.85	34.95	37.04	41.23
		M_2	26.07	31.50	34.21	36.92	42.34
		M_3	25.09	30.27	32.86	35.45	40.63
	0.6	M_1	27.45	31.45	33.45	35.45	39.45
		M_2	23.96	29.16	31.76	34.36	39.56
		M_3	22.93	27.87	30.34	32.81	37.75

Notes: The true missingness model is M_3. One MNAR covariate.

Table 7. Monte Carlo averages of the WL estimates (25% missingness).

WL	w	Fitted model	$\nu = 0.2$	$\nu = 0.4$	$\nu = 0.5$	$\nu = 0.6$	$\nu = 0.8$
WL_1	0.4	M_1	137.75	155.62	164.56	173.50	191.37
		M_2	198.76	225.86	239.41	252.95	280.04
		M_3	212.15	241.80	256.62	271.44	301.09
	0.5	M_1	150.53	170.76	180.87	190.98	211.21
		M_2	208.17	235.91	249.79	263.66	291.41
		M_3	222.41	252.90	268.14	283.38	313.86
	0.6	M_1	169.22	193.11	205.06	217.00	240.89
		M_2	221.15	249.75	264.04	278.34	306.94
		M_3	236.84	268.41	284.20	299.98	331.56
WL_2	0.4	M_1	25.73	28.93	30.52	32.12	35.31
		M_2	31.93	36.73	39.14	41.54	46.34
		M_3	31.40	36.12	38.47	40.83	45.55
	0.5	M_1	24.89	28.00	29.55	31.11	34.21
		M_2	29.61	34.14	36.40	38.66	43.18
		M_3	28.53	32.86	35.03	37.20	41.53
	0.6	M_1	24.13	27.16	28.67	30.18	33.21
		M_2	27.65	31.93	34.07	36.21	40.49
		M_3	26.08	30.09	32.10	34.10	38.12

Notes: The true missingness model is M_3. One MNAR covariate.

Table 8. The proportion of time the correct model is selected.

Missing	True model	Deviance	penalty_1	penalty_4	penalty_5	DIC_1	DIC_4	DIC_5
15%	M_1	0.76	0.04	UA	0.66	0.87	UA	0.82
	M_2	0.11	0.55	0.09	0.13	0.11	0.09	0.12
	M_3	UA	0.15	0.46	0.03	UA	0.46	UA
25%	M_1	0.67	0.02	UA	0.70	0.85	UA	0.77
	M_2	UA	0.94	0.46	0.19	UA	0.45	UA
	M_3	0.01	0.47	0.70	0.02	UA	0.70	0.02

Notes: One MNAR covariate. UA, unable to select the model framework with the correct missingness model.

Table 9. The proportion of time the correct model is selected.

WL	w	True model	$\nu = 0.2$	$\nu = 0.4$	$\nu = 0.5$	$\nu = 0.6$	$\nu = 0.8$
WL_1	0.4	M_1	0.83	0.82	0.81	0.80	0.77
		M_2	0.11	0.12	0.12	0.12	0.12
		M_3	UA	UA	UA	UA	UA
	0.5	M_1	0.85	0.94	0.94	0.95	0.94
		M_2	0.12	0.12	0.12	0.12	0.11
		M_3	UA	UA	UA	UA	UA
	0.6	M_1	0.85	0.83	0.80	0.78	0.73
		M_2	0.13	0.13	0.12	0.12	0.09
		M_3	UA	UA	UA	UA	UA
WL_2	0.4	M_1	UA	0.02	0.02	0.02	0.04
		M_2	0.56	0.59	0.61	0.60	0.64
		M_3	0.51	0.45	0.43	0.39	0.34
	0.5	M_1	UA	UA	UA	0.01	0.02
		M_2	0.41	0.53	0.54	0.55	0.58
		M_3	0.68	0.57	0.55	0.51	0.45
	0.6	M_1	UA	UA	UA	UA	UA
		M_2	0.30	0.43	0.46	0.51	0.53
		M_3	0.73	0.71	0.66	0.60	0.53

Notes: Missingness is 15%. One MNAR covariate.

Table 10. The proportion of time the correct model is selected.

WL	w	True model	$\nu = 0.2$	$\nu = 0.4$	$\nu = 0.5$	$\nu = 0.6$	$\nu = 0.8$
WL_1	0.4	M_1	0.79	0.74	0.73	0.70	0.65
		M_2	UA	UA	UA	UA	UA
		M_3	UA	UA	UA	UA	UA
	0.5	M_1	0.82	0.90	0.91	0.93	0.93
		M_2	UA	UA	UA	UA	UA
		M_3	UA	0.01	0.01	0.01	0.02
	0.6	M_1	0.85	0.76	0.72	0.72	0.65
		M_2	UA	UA	UA	UA	UA
		M_3	UA	UA	UA	UA	UA
WL_2	0.4	M_1	0.02	0.03	0.03	0.04	0.04
		M_2	0.09	0.08	0.07	0.07	0.05
		M_3	0.12	0.08	0.08	0.07	0.07
	0.5	M_1	UA	0.02	0.02	0.03	0.03
		M_2	0.08	0.06	0.06	0.06	0.06
		M_3	0.22	0.18	0.17	0.14	0.13
	0.6	M_1	UA	UA	0.01	0.01	0.02
		M_2	0.07	0.07	0.07	0.07	0.07
		M_3	0.32	0.29	0.26	0.25	0.25

Notes: Missingness is 25%. One MNAR covariate.

5. Discussion

The focus of this article was the performances of the commonly used model selection criteria for comparisons among selection models with the same response and covariate models but with different non-ignorable missingness models. In addition to the traditional model selection criteria, novel extensions were also considered and their performances were evaluated. The sample size selected to study the finite sample performances of the criteria was $n = 250$ and a reasonable choice

Table 11. Monte Carlo averages of the deviance, penalty, and DIC estimates.

Missing	True model	Fitted model	Deviance	penalty_1	penalty_4	penalty_5	DIC_1	DIC_4	DIC_5
15%	M_1	M_1	269.05	5.18	7169.57	23.37	274.23	7438.62	292.42
		M_2	270.65	4.93	6943.96	23.89	275.58	7214.62	294.54
		M_3	270.43	5.79	6931.57	23.93	276.22	7202.00	294.36
	M_2	M_1	251.76	5.83	7534.10	17.95	257.58	7785.80	269.70
		M_2	262.24	5.30	6501.40	19.82	267.54	6763.60	282.06
		M_3	262.35	5.73	6468.50	19.86	268.07	6730.80	282.21
	M_3	M_1	250.65	5.76	7798.83	18.35	256.41	8049.48	268.99
		M_2	262.45	5.55	6688.22	20.52	267.99	6950.67	282.96
		M_3	262.27	5.78	6740.43	20.66	268.05	7002.69	282.93
25%	M_1	M_1	263.26	6.17	10,552.97	29.89	269.43	10,816.23	293.15
		M_2	265.05	5.88	10,328.18	30.67	270.93	10,593.23	295.72
		M_3	264.87	6.87	10,146.33	30.49	271.75	10,411.19	295.35
	M_2	M_1	258.91	6.29	8197.60	20.99	265.21	8456.50	279.90
		M_2	261.75	5.57	7896.50	21.87	267.33	8158.30	283.62
		M_3	262.28	6.12	7908.70	22.39	268.39	8171.00	284.68
	M_3	M_1	202.57	13.15	18,280.49	16.48	215.71	18,483.06	219.05
		M_2	247.13	11.12	10,088.97	21.07	258.25	10,336.10	268.20
		M_3	245.14	12.53	10,348.27	20.72	257.67	10,593.41	265.86

Note: Two MNAR covariates.

Table 12. Monte Carlo averages of the WL estimates (15% missingness).

WL	w	Fitted model	$\nu = 0.2$	$\nu = 0.4$	$\nu = 0.5$	$\nu = 0.6$	$\nu = 0.8$
WL_1	0.4	M_1	225.19	262.09	280.54	298.99	335.90
		M_2	226.43	263.61	282.21	300.80	337.98
		M_3	226.75	263.75	282.26	300.76	337.76
	0.5	M_1	232.29	270.32	289.33	308.34	346.37
		M_2	233.48	271.77	290.91	310.06	348.35
		M_3	233.83	271.93	290.98	310.03	348.13
	0.6	M_1	241.37	280.84	300.57	320.30	359.77
		M_2	242.47	282.17	302.02	321.87	361.57
		M_3	242.87	282.37	302.11	321.86	361.35
WL_2	0.4	M_1	61.86	72.05	77.15	82.25	92.44
		M_2	54.05	62.78	67.14	71.50	80.23
		M_3	52.65	61.01	65.20	69.38	77.75
	0.5	M_1	57.33	66.78	71.51	76.24	85.70
		M_2	50.61	58.90	63.05	67.20	75.50
		M_3	49.30	57.23	61.20	65.17	73.09
	0.6	M_1	53.51	62.35	66.77	71.19	80.03
		M_2	47.75	55.68	59.65	63.62	71.55
		M_3	46.48	54.04	57.82	61.60	69.16

Notes: The true missingness model is M_1. Two MNAR covariates.

of size according to similar simulation studies in the literature of interest. The main findings can be summarized as follows:

- Different criteria show a tendency towards different MNAR models.
- Deviance, penalty_5, DIC_1, and DIC_5 have a tendency to point to selection models with the smallest non-ignorably missingness model.

Table 13. Monte Carlo averages of the WL estimates (25% missingness).

WL	w	Fitted model	$\nu = 0.2$	$\nu = 0.4$	$\nu = 0.5$	$\nu = 0.6$	$\nu = 0.8$
WL_1	0.4	M_1	220.65	256.56	274.51	292.46	328.37
		M_2	221.99	258.20	276.31	294.41	330.62
		M_3	224.08	258.42	276.43	294.44	330.46
	0.5	M_1	228.16	265.23	283.77	302.31	339.38
		M_2	229.46	266.81	285.49	304.17	341.52
		M_3	229.93	267.08	285.66	304.24	341.39
	0.6	M_1	237.81	276.38	295.66	314.95	353.52
		M_2	239.02	277.83	297.24	316.65	355.47
		M_3	239.57	278.17	297.47	316.77	355.38
WL_2	0.4	M_1	83.59	97.20	104.00	110.81	124.42
		M_2	82.28	95.79	102.55	109.30	122.82
		M_3	80.89	93.81	100.26	106.72	119.63
	0.5	M_1	79.91	92.91	99.41	105.90	118.90
		M_2	78.39	91.28	97.72	104.16	170.04
		M_3	76.84	89.09	95.22	101.34	113.59
	0.6	M_1	76.74	89.20	95.44	101.67	114.14
		M_2	75.06	87.40	93.57	99.75	112.09
		M_3	73.34	85.01	90.85	96.69	108.36

Notes: The true missingness model is M_1. Two MNAR covariates.

Table 14. Monte Carlo averages of the WL estimates (15% missingness).

WL	w	Fitted model	$\nu = 0.2$	$\nu = 0.4$	$\nu = 0.5$	$\nu = 0.6$	$\nu = 0.8$
WL_1	0.4	M_1	210.75	245.08	262.25	279.42	313.75
		M_2	219.45	255.32	273.25	291.18	327.05
		M_3	219.79	255.60	273.50	291.41	327.22
	0.5	M_1	218.55	254.03	271.77	289.51	324.99
		M_2	226.89	263.82	282.28	300.74	337.67
		M_3	227.24	264.11	282.55	300.98	337.86
	0.6	M_1	228.75	265.71	284.19	302.67	339.63
		M_2	236.48	274.76	293.90	313.04	351.32
		M_3	236.85	275.08	294.19	313.30	351.52
WL_2	0.4	M_1	70.55	82.35	88.25	94.14	105.94
		M_2	61.74	73.27	79.04	84.81	96.35
		M_3	59.88	70.90	76.41	81.92	92.94
	0.5	M_1	65.46	76.49	81.99	87.51	98.53
		M_2	55.71	66.40	71.74	77.08	87.77
		M_3	53.89	64.06	69.15	74.23	84.40
	0.6	M_1	61.05	71.42	76.60	81.78	92.14
		M_2	50.70	60.67	65.65	70.63	80.60
		M_3	48.92	58.38	63.11	67.83	77.29

Notes: The true missingness model is M_2. Two MNAR covariates.

- Penalty_1, penalty_4, and DIC_4 have a tendency to point to selection models with the largest non-ignorably missingness model.
- Performances of the WL measures depend on the weight function; whether the WL measure tends to pick selection models with the smallest or the largest missingness model depends upon the weight function.

Our results showed that the existent as well as proposed model selection criteria were unable to fully satisfy the purpose of pinning down the correct missingness model, indicating a need

Table 15. Monte Carlo averages of the WL estimates (25% missingness).

WL	w	Fitted model	$\nu = 0.2$	$\nu = 0.4$	$\nu = 0.5$	$\nu = 0.6$	$\nu = 0.8$
WL_1	0.4	M_1	216.21	251.29	268.84	286.39	321.47
		M_2	218.12	253.71	271.50	289.30	324.88
		M_3	218.92	254.47	272.25	290.02	325.58
	0.5	M_1	223.92	260.11	278.21	296.30	332.49
		M_2	225.74	262.41	280.74	299.07	335.74
		M_3	226.55	263.18	281.50	299.82	336.45
	0.6	M_1	233.90	271.51	290.31	309.12	346.73
		M_2	235.57	273.62	292.64	311.67	349.71
		M_3	236.39	274.42	293.43	312.43	350.45
WL_2	0.4	M_1	80.61	93.99	100.68	107.38	120.76
		M_2	74.80	87.67	94.10	100.54	113.40
		M_3	73.92	86.36	92.58	98.80	111.24
	0.5	M_1	76.77	89.58	95.99	102.41	115.21
		M_2	69.97	82.12	88.20	94.28	106.44
		M_3	69.02	80.73	86.59	92.45	104.16
	0.6	M_1	73.44	85.77	91.94	98.10	110.43
		M_2	65.83	77.37	83.14	88.91	100.46
		M_3	64.81	75.90	81.44	86.98	98.07

Notes: The true missingness model is M_2. Two MNAR covariates.

Table 16. Monte Carlo averages of the WL estimates (15% missingness).

WL	w	Fitted model	$\nu = 0.2$	$\nu = 0.4$	$\nu = 0.5$	$\nu = 0.6$	$\nu = 0.8$
WL_1	0.4	M_1	209.65	243.84	260.93	278.02	312.20
		M_2	219.62	255.45	273.37	291.28	327.11
		M_3	219.62	255.39	273.27	291.15	326.92
	0.5	M_1	217.55	252.90	270.57	288.26	323.61
		M_2	227.09	263.99	282.43	300.88	337.77
		M_3	227.11	262.94	282.35	300.77	337.60
	0.6	M_1	227.86	264.74	283.18	301.61	338.49
		M_2	236.71	274.96	294.08	313.21	351.46
		M_3	236.75	274.93	294.03	313.12	351.30
WL_2	0.4	M_1	70.78	82.61	88.53	94.44	106.27
		M_2	52.27	63.18	68.63	74.08	84.99
		M_3	51.30	61.87	67.16	72.44	83.01
	0.5	M_1	65.83	76.91	82.45	87.99	99.07
		M_2	48.62	58.87	63.99	69.11	79.35
		M_3	47.64	57.55	62.50	67.45	77.36
	0.6	M_1	61.53	71.96	77.18	82.40	92.83
		M_2	45.52	55.20	60.03	64.87	74.55
		M_3	44.53	53.87	58.53	63.20	72.53

Notes: The true missingness model is M_3. Two MNAR covariates.

to improve the model selection criteria used in the GLM analysis with non-ignorably missing covariates. One such attempt towards improving the WL measures in this context and explaining the model selection tendency of the criterion as observed in the simulation experiment may involve examining whether a WL measure in this context is a special case of the general form introduced by Smith and Spiegelhalter [17]. They considered the general form given by $\Lambda(a) = \lambda - a(p_1 - p_0)$, where $\Lambda(a)$ is the log ratio of the specific model selection criterion computed for two competing models in which one model is nested in the other, λ is the likelihood ratio statistic, p_1 and p_0 are the

Table 17. Monte Carlo averages of the WL estimates (25% missingness).

WL	w	Fitted model	$\nu = 0.2$	$\nu = 0.4$	$\nu = 0.5$	$\nu = 0.6$	$\nu = 0.8$
WL_1	0.4	M_1	172.49	198.99	212.09	225.29	251.69
		M_2	209.50	242.01	258.27	274.53	307.04
		M_3	208.64	240.67	256.69	272.71	304.74
	0.5	M_1	182.15	210.12	224.10	238.09	266.06
		M_2	217.67	251.11	267.83	284.55	317.99
		M_3	216.97	249.92	266.40	282.88	315.84
	0.6	M_1	195.44	225.58	240.65	255.72	285.86
		M_2	228.37	263.00	280.32	297.64	332.28
		M_3	227.92	262.06	279.14	296.21	330.36
WL_2	0.4	M_1	61.25	70.50	75.13	79.75	89.00
		M_2	59.70	70.55	75.98	81.40	92.25
		M_3	58.04	68.40	73.59	78.77	89.13
	0.5	M_1	57.90	66.63	70.99	75.36	84.09
		M_2	55.03	65.26	70.38	75.50	85.73
		M_3	53.35	63.09	67.96	72.82	82.56
	0.6	M_1	54.90	63.17	67.30	71.43	79.69
		M_2	51.13	60.84	65.69	70.55	80.26
		M_3	49.44	58.65	63.25	67.85	77.05

Notes: The true missingness model is M_3. Two MNAR covariates.

Table 18. The proportion of time the correct model is selected.

Missing	True model	Deviance	penalty_1	penalty_4	penalty_5	DIC_1	DIC_4	DIC_5
15%	M_1	0.67	0.10	0.04	0.83	0.80	0.04	0.73
	M_2	UA	0.83	0.47	0.01	UA	0.47	UA
	M_3	0.02	0.08	0.39	0.03	0.01	0.39	UA
25%	M_1	0.61	0.10	0.11	0.66	0.79	0.11	0.69
	M_2	0.11	0.90	0.34	0.15	0.20	0.34	0.14
	M_3	UA	0.05	0.32	0.04	UA	0.32	UA

Notes: Two MNAR covariates. UA, unable to select the model framework with the correct missingness model.

number of covariates in models m_1 and m_0, respectively, such that $m_0 \subset m_1$, and the coefficient a is the penalty for overfitting quantified. They showed that a number of model comparison criteria such as the Akaike Information Criterion and Bayes Factor can be regarded as special cases of this general expression. Also, as they noted in their paper, the distribution of λ being asymptotically equivalent to $\chi^2_{(p_1-p_0)}$ results in $E[\Lambda(a)] \approx a(p_1 - p_0)$ and this expression helps in establishing the model selection tendency of the criterion based on whether the quantity $a \geq 1$ or not. Laud and Ibrahim [18] showed under non-informative priors that the WL measure in complete data situations is a special case of this general form: $2n\log(\text{WL}_{m_0}/\text{WL}_{m_1}) = \lambda - a(p_1 - p_0)$, where $a = (n/(p_1 - p_0))\log((n - p_0 - 2)/(n - p_1 - 2))$. Whether a WL measure in missing data situations too is a special case or not remains as an interesting question. If it can be shown that it is indeed a special case, then the magnitude of the coefficient a would shed theoretical light on the direction of the model tendency of the WL measure as observed in the simulation study.

In this article, we examined the performances of the commonly used model selection criteria and their possible extensions within the realm of GLMs with covariates subject to non-ignorable missingness. There are study designs such as longitudinal studies whereby especially response variables suffer from non-ignorable missingness. Investigating the performances of the present and proposed model selection criteria in such situations is a topic of interest for future research. Especially in epidemiological studies in which several explanatory variables are involved, the

Table 19. The proportion of time the correct model is selected.

WL	w	True model	$\nu = 0.2$	$\nu = 0.4$	$\nu = 0.5$	$\nu = 0.6$	$\nu = 0.8$
WL_1	0.4	M_1	0.79	0.73	0.72	0.72	0.68
		M_2	UA	UA	UA	UA	UA
		M_3	0.02	0.02	0.02	0.02	0.02
	0.5	M_1	0.82	0.92	0.92	0.92	0.92
		M_2	UA	UA	UA	UA	UA
		M_3	0.02	0.27	0.27	0.27	0.27
	0.6	M_1	0.84	0.76	0.73	0.69	0.67
		M_2	UA	UA	UA	UA	UA
		M_3	0.02	0.02	0.02	0.02	0.02
WL_2	0.4	M_1	UA	UA	UA	UA	UA
		M_2	0.07	0.05	0.05	0.05	0.05
		M_3	1.00	1.00	1.00	1.00	1.00
	0.5	M_1	UA	UA	UA	UA	UA
		M_2	0.05	0.05	0.05	0.04	0.02
		M_3	1.00	1.00	1.00	1.00	1.00
	0.6	M_1	UA	UA	UA	UA	UA
		M_2	0.02	UA	UA	UA	UA
		M_3	1.00	1.00	1.00	1.00	1.00

Notes: Missingness is 15%. Two MNAR covariates.

Table 20. The proportion of time the correct model is selected.

WL	w	True model	$\nu = 0.2$	$\nu = 0.4$	$\nu = 0.5$	$\nu = 0.6$	$\nu = 0.8$
WL_1	0.4	M_1	0.72	0.65	0.65	0.64	0.62
		M_2	0.27	0.23	0.20	0.17	0.14
		M_3	UA	UA	UA	UA	UA
	0.5	M_1	0.73	0.88	0.88	0.88	0.88
		M_2	0.28	0.23	0.21	0.19	0.14
		M_3	UA	UA	UA	UA	UA
	0.6	M_1	0.76	0.71	0.68	0.65	0.61
		M_2	0.27	0.22	0.20	0.18	0.15
		M_3	UA	UA	UA	UA	UA
WL_2	0.4	M_1	0.02	0.02	0.03	0.03	0.03
		M_2	0.20	0.20	0.20	0.20	0.19
		M_3	0.59	0.49	0.49	0.48	0.43
	0.5	M_1	0.01	0.01	0.01	0.02	0.02
		M_2	0.19	0.09	0.09	0.08	0.08
		M_3	0.70	0.64	0.64	0.55	0.52
	0.6	M_1	0.01	0.01	0.01	0.01	0.02
		M_2	0.11	0.08	0.08	0.08	0.08
		M_3	0.80	0.72	0.69	0.66	0.61

Notes: Missingness is 25%. Two MNAR covariates.

number of covariates subject to missingness is higher than that in the scenarios studied in the simulation experiment herein [19]. Because of the computational cost, Monte Carlo simulation studies usually are concerned with scenarios with a smaller number of covariates when the missing data are involved [4,20–23]. Nevertheless, the results of our simulation experiment point to a need for developing a model selection criterion that can be used to make a more accurate comparison among non-ignorable missingness models and constitute a reference for the practitioners in the fields where statistical modelling with missing covariates is involved.

References

[1] Little RJA, Rubin DB. Statistical analysis with missing data. New York: Wiley and Sons; 2002.
[2] Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde, A. Bayesian measures of model complexity and fit (with discussion). J. R. Stat. Soc. B. 2002;64:583–639.
[3] Chen MH, Dey DK, Ibrahim JG. Bayesian criterion based model assessment for categorical data. Biometrika. 2004;91:45–63.
[4] Huang L, Chen MH, Ibrahim JG. Bayesian analysis for generalized linear models with nonignorably missing covariates. Biometrics. 2005;61:767–780.
[5] Ibrahim JG, Zhu H, Tang N. Model selection criteria for missing-data problems using the EM algorithm. J. Amer. Statist. Assoc. 2008;103:1648–1658.
[6] Ibrahim JG. Comment: incomplete data in clinical studies: analysis, sensitivity, and sensitivity analysis. Drug. Inf. J. 2009;43:431–432.
[7] Nelder JA, Wedderburn WM. Generalized linear models. J. R. Stat. Soc. A. 1972;135:370–384.
[8] Ibrahim JG, Lipsitz SR, Chen MH. Missing covariates in generalized linear models when the missing data mechanism is nonignorable. J. R. Stat. Soc. B. 1999;61:173–190.
[9] Boone EL, Ye K, Smith EP. Assessing environmental stressors via Bayesian model averaging in the presence of missing data. Environmetrics. 2011;22:13–22.
[10] Celeux G, Forbes F, Robert CP, Titterington DM. Deviance information criteria for missing data models. Bayesian Anal. 2006;1:651–674.
[11] Mason A, Richardson S, Best N. Two-pronged strategy for using DIC to compare selection models with non-ignorable missing responses. Bayesian Anal. 2011;6:1–44.
[12] Rothenberg T. Identification in parametric models. Econometrica. 1971;39:577–591.
[13] Fitzmaurice GM, Laird NM, and Zahner GEP. Multivariate logistic models for incomplete binary responses. J. Amer. Statist. Assoc. 1996;91:99–108.
[14] Bozdogan H. Akaike's information criterion and recent developments in information complexity. J. Math. Psychol. 2000;44:62–91.
[15] Albert JH, Chib S. Bayesian analysis of binary and polychotomous response data. J. Amer. Statist. Assoc. 1993;88:669–679.
[16] Brooks SP, Gelman A. Alternative methods for monitoring convergence of iterative simulations. J. Comput. Graph. Statist. 1998;7:434–455.
[17] Smith AFM, Spiegelhalter DJ. Bayes factors and choice criteria for linear models. J. R. Stat. Soc. B. 1980;42:213–220.
[18] Laud PW, Ibrahim JG. Predictive model selection. J. R. Stat. Soc. B 1995;42:247–262.
[19] Marshall A, Altman DG, Royston P, Holder RL. Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. BMC Med. Res. Methodol. 2010;10. Available from: http://www.biomedcentral.com/1471-2288/10/7
[20] Lipsitz SR, Parzen M, Ewell M. Inference using conditional logistic regression with missing covariates. Biometrics. 1998;54:295–303.
[21] Sinha SK. Robust methods for generalized linear models with nonignorable missing covariates. Canad. J. Statist. 2008;36:277–299.
[22] Zhang N, Little RJ. A pseudo-Bayesian shrinkage approach to regression with missing covariates. Biometrics 2012;68:933–942.
[23] Seo B. A gradient-based algorithm for semiparametric models with missing covariates. J. Stat. Comput. Simul. 2011;81:381–390.