frontiers | Frontiers in Applied Mathematics and Statistics

TYPE Original Research
PUBLISHED 24 November 2022
DOI 10.3389/fams.2022.1052752

OPEN ACCESS

EDITED BY
Fabrizio Maturo,
Second University of Naples, Italy

REVIEWED BY
Mohd Tahir Ismail,
Universiti Sains Malaysia (USM),
Malaysia
Halim Zeghdoudi,
University of Annaba, Algeria

*CORRESPONDENCE
Ronald Wesonga
wesonga@wesonga.com
Muna Al-Shaaibi
m.shaaibi@gmail.com

SPECIALTY SECTION
This article was submitted to Statistics and Probability, a section of the journal Frontiers in Applied Mathematics and Statistics

RECEIVED 24 September 2022
ACCEPTED 01 November 2022
PUBLISHED 24 November 2022

CITATION
Al-Shaaibi M and Wesonga R (2022) Bias reduction in the logistic model parameters with the LogF(1,1) penalty under MAR assumption. *Front. Appl. Math. Stat.* 8:1052752. doi: 10.3389/fams.2022.1052752

Bias reduction in the logistic model parameters with the LogF(1,1) penalty under MAR assumption

Muna Al-Shaaibi[1,2]* and Ronald Wesonga[1,2]*

[1]Department of Statistics, Sultan Qaboos University, Muscat, Oman, [2]Data Science Analytics Lab, Department of Statistics, Sultan Qaboos University, Muscat, Oman

In this paper, we present a novel validated penalization method for bias reduction to estimate parameters for the logistic model when data are missing at random (MAR). Specific focus was given to address the data missingness problem among categorical model covariates. We penalize a logit log-likelihood with a novel prior distribution based on the family of the LogF(m,m) generalized distribution. The principle of expectation-maximization with weights was employed with the Louis' method to derive an information matrix, while a closed form for the exact bias was derived following the Cox and Snell's equation. A combination of simulation studies and real life data were used to validate the proposed method. Findings from the validation studies show that our model's standard errors are consistently lower than those derived from other bias reduction methods for the missing at random data mechanism. Consequently, we conclude that in most cases, our method's performance in parameter estimation is superior to the other classical methods for bias reduction when data are MAR.

KEYWORDS

penalization, parameter estimation, bias reduction, MAR, LogF(1,1) penalty

1. Introduction

Statistical model parameter estimation is a major process requiring improved methods because its a basis for informed decisions, which in most cases rely on data. Complete and incomplete data alike may suffer from biases that may arise from the methods of estimations. Careful scrutiny and efforts have been made to generate some remarkable reduction in bias of estimated model parameters [1, 2]. Penalizing the log-likelihood with a suitable prior distribution is known to reduce bias in model parameters. Bias reduction under conditions of missing data in the estimated parameters has received little research attention in the past, drawing a few studies, also with different focus in scope and coverage. Furthermore, recent studies that have attempted to derive bias reduction methods, under missing data conditions have tried only for the case of missingness in the outcome variable [3, 4].

The estimation process of the logistic regression requires that datasets are complete, besides observing a set of other model assumptions. However, in reality and for various reasons, it is difficult for some vital studies to generate complete datasets. Commonly, missing data are handled by using list-wise deletion method which drops any data item containing missing values. However, dropping cases reduces the sample size [5] and this plays negatively on the predictability and application of the fitted model. For the logistic model, it is well-known that with a small sample size the maximum likelihood estimates will be biased. In 1993, Firth proposed a method which resulted in the second order $[O(n^{-2})]$ bias reduction of the estimates [6]. Moreover, data separation, especially for the case of a binary outcome is yet another problem since the estimates are obtained using an iterative process such as the Newton-Raphson method. In 2006, Heinz found that Firth's method could also solve the separation problem to a certain level of efficiency [7–9]. Moreover in 2016, Carlisle noted that the separation problem could also be handled by using a well-tested prior distribution [10].

A recent study with a focus on bias reduction was by Little and Rubin [11]. Previously in 1990, Ibrahim [12] proposed an estimation method for incomplete covariates for a generalized linear model (GLM) by using the expectation-maximization (EM) algorithm. Followed by Ibrahim and Lipsitz [13], who applied the EM algorithm to estimate the logistic regression parameters when some of the responses were missing. Consequently, Das et al. [14] proposed a bias reduction for the logistic model under MAR assumption following the Firth approach by applying Ibrahim's EM method which achieved a second order bias reduction. In addition, Maity et al. adopted the Lipsitz and Ibrahim EM algorithm, and they came up with another bias reduction method for the logistic model under the incomplete response problem and the MNAR mechanisms, which attempted to reduce the bias as well as the separation problem [15, 16]. Overall, however, chronology shows sparsity of studies that focus specifically on achieving more accurate parameter estimates under the incomplete covariates' data since most of the attempts have focussed on the missing data in the outcome variable.

Our current study is guided by the success derived from other related complete data studies from which the choice of a penalty distribution is drawn. For the complete case, it was demonstrated that the LogF(m,m) family of distributions could provide a reliable penalty of the logit model [17]. Incidentally, no study has tested this hypothesis for the bias reduction under the missing at random (MAR) data problem.

2. Motivating example

In an exploratory analysis, we analyzed data from a survey meant to determine intensive care unit (ICU) admission risk factors in a study conducted at the Royal Hospital in the Sultanate of Oman [18]. To motivate the need of bias reduction method for estimated parameters, we considered only the significant covariates which are the independent predictors of the ICU admission for the COVID-19 patients. The selected covariates were; having conjuctivitis, being unconscious and having a neurological problem, whereas the response variable was whether the patient was admitted to the ICU. The proportion of missing data for each explanatory variable are; conjuctivitis (34%), unconscious (2%), and neurological (4%). We explored by fitting a logistic model for ICU admission. The results are summarized in Figure 1.

It was observed that, for example *being unconscious* could be interpreted as not significant under the Ibrahim's method yet it is significant under the Firth's method for bias reduction. Further, the predictor, *neurological*, under both the Ibrahim's and Firth's methods could not be considered significant with noticeably higher standard errors under the Firth than Ibrahim's method. Therefore, what method is best suited for drawing reliable statistical inference to such a problem? Why do the two methods, that are thought to reduce bias in the model parameters perform differently on the same data? Could penalization improve the performance in parameter estimation? To address these questions, we propose and validate a novel bias reduction method, as demonstrated in the preceeding sections.

Consequently, in this current study, we propose a novel method to reduce bias in the model parameters through penalizing the log-likelihood function, based on the EM method of weight for estimation of incomplete binary categorical covariates. Moreover, we also derive a closed form for the bias for estimated parameter by applying the Cox and Snell's equation, which is based on the Taylor's series [19]. The information matrix was developed using the Louis' method [20]. The main contribution of this study is the novel method, which is quite competitive in parameter estimation under the MAR data mechanism. The method is validated to respond to questions such as; can penalization result in bias reduction of the estimated parameters when the binary response is complete, but some categorical covariates are stochastically missing? What is the better penalty function to apply? How competitive is the novel approach compared to the available methods? Using the proposed method, and as an application, what are the determinants of ICU admission due to COVID-19?

The rest of the paper is organized as follows; in Section 3, we present theoretical methods as the building blocks for the study, Section 4 provides some results focussing on model validation through simulation as well as application using the real life data, and in Section 5, we present a discussion and draw conclusions for the study.

3. Methods

3.1. Definitions and notation

Three important missingness mechanisms have been defined [21] and these are mainly being attributed to the data rather than the distribution of missingness [22, 23]. Among them is the

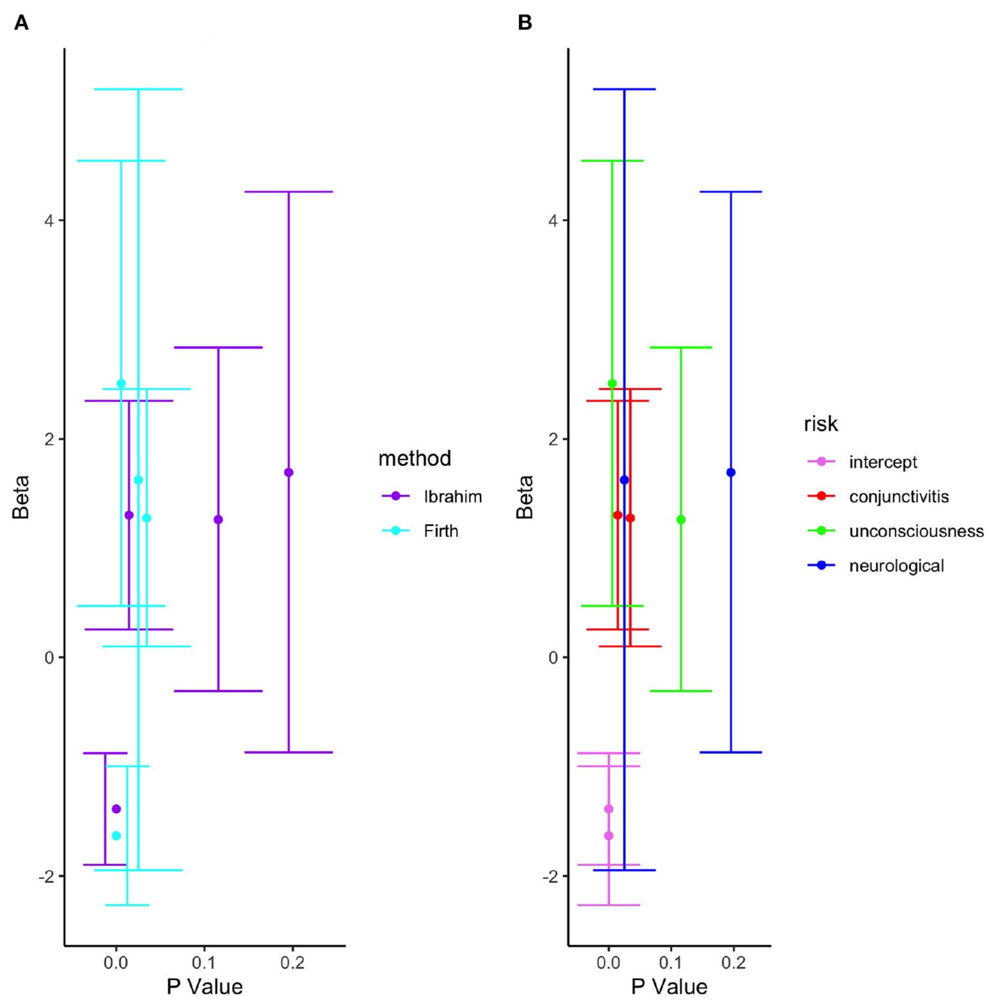

FIGURE 1
Bias reduction under the Ibrahim and firth methods. **(A)** Beta vs. P by method. **(B)** Beta vs. P by risk.

missing at random (MAR) mechanism, which is the main focus of this study, briefly defined here:

Definition 1. *Let* $y = (y_1, y_2, \cdots, y_n)$ *be a realization of the random sample from an infinite population with density* $f(y; \theta_0)$*. Let* δ_i *be an indicator function defined by*

$$\delta_i = \begin{cases} 1 & \text{if } y_i \text{ is observed} \\ 0 & \text{if } y_i \text{ otherwise} \end{cases}$$

Let's assume $P(\delta_i = 1|y_i) = \pi(y_i, \phi)$ *for some parameter* ϕ *and that* $\pi(.)$ *is a known function. Thus, if instead of observing* (δ_i, y_i) *directly, we do observe* $(\delta_i, y_{i,obs})$*, where the* $y_{obs} = y_{obs}(y, \delta)$ *and* $y_{i,obs}$ *is defined as*

$$y_{i,obs} = \begin{cases} y_i & \text{if } \delta_i = 1 \\ * & \text{if } \delta_i = 0 \end{cases}$$

The missing data mechanism is MAR if

$$p(\delta|y) = p(\delta|y_{obs})$$

for all y_1 *and* y_2 *satisfying* $y_{obs}(y_1, \delta) = y_{obs}(y_2, \delta)$.

This implies that under the MAR mechanism, the response mechanism $p(\delta|y)$ depends on y only through y_{obs}. If we let $y = (y_{obs}, y_{mis})$, then by the Bayes' theorem $p(y_{mis}|y_{obs,\delta}) = \frac{p(\delta|y_{mis} y_{obs})}{p(\delta|y_{obs})} = p(y_{mis}|y_{obs})$. Consequently, the two other mechanisms can be defined as missing completely at random (MCAR), which occur when $p(\delta|y) \not\propto y$, whereas not missing at random (NMAR) occurs when $p(\delta|y) \neq p(\delta|y_{obs})$. Our study focusses on the MAR data mechanism. However, though the MNAR may have been more plausible, it is known to rely mainly on unstable model assumptions. Moreover, every MCAR data are by default also MAR, and the MNAR data are not ignorable. Let θ be the parameter vector for data model, and Ψ is the parameter vector for missingness mechanisms, so the full-likelihood can be written as:

$$L(\theta, \Psi, y_{obs}, \delta) = \int f_Y(y_{obs}, y_{mis}|\theta) f_{(M|y)}(\delta|y_{(obs)}, y_{(mis)}, \Psi)\, dy_{(mis)}$$

Theorem 1. *The missingness mechanisms is ignorable for the direct likelihood inference*(m, y_{obs}) *if the following two condition hold:*

- *Parameter distinctness: two parameters* θ *and* Ψ *are distinct, in the sense that the joint parameter space* (θ, Ψ) *say,* $\Omega_{\theta,\Psi}$*, is the product of the parameter space* Ω_θ *of* θ *and the parameter space* Ω_Ψ *of* Ψ *that is,* $\Omega_{\theta,\Psi} = \Omega_\theta \times \Omega_\Psi$.
- *Factorization of the full likelihood with* (y_{obs}, δ) *factors as:*

$$L_{full}(\theta, \Psi|y_{obs,\delta}) = L_{ignore(\theta|y_{obs})} \times L_{rest}(\Psi|y_{obs}, \delta) \quad \forall \quad \theta, \Psi \in \Omega_{\theta,\Psi} \tag{1}$$

Corollary 1.1. *If the missing data are MAR at* (δ, y_{obs}) *with* θ *and* Ψ *distinct, then we can say that the missingness mechanisms is ignorable for the likelihood inference.*

Penalized regression methods have been used to assess the best fit for the complete-case [24].

Definition 2. *Assume a regression model,* $y = \boldsymbol{X}\beta$ *and a loss function of the parameter estimate* θ*, therefore* $Q(\theta)$*, the resulting loss function will be the summation of the squared error loss and a penalty term. The penalized regression is defined as*

$$Q(\theta) = L(\theta|\boldsymbol{X}) + \omega p(\theta) = (y - \boldsymbol{X}\beta)^{'}(y - \boldsymbol{X}\beta) + \omega p(\beta)$$

where, $Y \in \Re^n, \boldsymbol{X}_{nxp}, \beta \in \Re^p$*, and* ω *is a lagrange coefficient. The* $\omega p(\beta)$ *penalty term may also be referred to as a regularization term.*

3.2. The EM by the method of weights

Let $y_1, y_2, ..., y_n$ be independent Bernoulli random variables and $\mathbf{x}$, the matrix of regressors of order $n \times d$, where d is the number of regressors. For completely observed data, $\pi_i = (P(y_i = 1|\mathbf{x}, \beta)$ where $\beta = (\beta_0, \beta_1, \beta_2, ..., \beta_p)'$ are the unknown parameters [12]. The logistic model relates π_i to $\mathbf{x}$ as $log\dfrac{\pi_i}{(1-\pi_i)} = \mathbf{x}'\beta$, and the likelihood written as $L = \Pi_{i=1}^n P(y_i|\mathbf{x}, \beta)$.
Let $P(\mathbf{x}|\alpha)$ be the marginal density of $\mathbf{x}$ where α is the indexing nuisance parameter. Therefore, for $\theta = (\beta, \alpha)$ the complete log likelihood is:

$$\ell(\theta, \mathbf{x}, y) = \sum_i \ell(\theta, x_i, y_i) = \sum_i \ell_{y_i|x_i}(\beta) + \ell_{x_i}(\alpha) \tag{2}$$

Where, $\ell_{y_i|x_i}(\beta) = log(P(y_i|x_i, \beta))$ and $\ell_{x_i}(\alpha) = log(P(x_i|\alpha))$. Suppose further that $x_i = (x_{obs,i}, x_{mis,i})$, then the E-step at the $(t+1)th$ iteration of the log-likelihood using EM algorithm becomes:

$$\begin{aligned} Q(\theta|\theta^{(t)}) &= \sum_{i=1}^n \sum_{x_{mis},i(j)} w_{ij}^{(t)} \ell(\theta; x_i, y_i) \\ &= \sum_{i=1}^n \sum_{x_{mis},i(j)} w_{ij}^{(t)} log[P(y_i|x_i, \beta)] \\ &\quad + \sum_{i=1}^n \sum_{x_{mis},i(j)} w_{ij}^{(t)} log[P(x_i|\alpha)] \\ &= Q^{(1)}(\beta|\theta^{(t)}) + Q^{(2)}(\alpha|\theta^{(t)}) \end{aligned} \tag{3}$$

Where, $\theta^{(t)}$ is the estimate obtained at the tth iteration in the EM. Note that in the equation above, the inner sum is taken over all possible distinct missing pattern indexing j for each subject i. By using the Bayes' theorem, the weight $w_{(ij)}$ can be written as

$$\begin{aligned} w_{ij}^{(t)} &= P(x_{mis}, i(j)|x_{obs,i}, y_i, \theta^{(t)}) \\ &= \frac{P(y_i|x_{mis}, i(j), x_{obs,i}, \theta^{(t)}) P(x_{mis,i}(j), x_{obs,i}|\theta^{(t)})}{\sum_{xmis,i(j)} P(y_i|x_i, \theta^{(t)}) P(x_i|\theta^{(t)})} \end{aligned} \tag{4}$$

The E-step is a sum of two functions with different parameters, while the M-step involves two separate maximizations. Maximizing $Q^{(1)}$ gives the score function: $U(\beta) = \sum_{i=1}^n \sum_{xmis,i,(j)} w_{ij} x_i (y_i - \pi_i)$ and the information matrix $I(\beta) = x'\pi(1-\pi)wx$, where the maximization is done iteratively using the relation $\beta^{(t+1)} = \beta^{(t)} + I(\beta^{(t)})^{-1} U(\beta^{(t)})$.
Based on this approach, we proposed a novel method for bias reduction for the logistic model under MAR assumption as in the proceeding sections.

Theorem 2 (monotonicity property)**.** *Every EM algorithm increases* $\ell(\theta|y_{obs})$ *at each iteration that is:*

$$\ell(\theta^{(t+1)}|y_{obs}) \geq \ell(\theta^{(t)}|y_{obs})$$

This theorem implies that as the EM algorithm iterates, the $(t+1)th$ guess $(\theta)^{(t+1)}$ will never be less than tth guess θ^t. This satisfies the monotonicity property of the EM algorithm. That is $\ell(\theta|y_{obs})$ is non-decreasing on each iteration of the EM algorithm, and is strictly increasing on any iteration, such that the Q-function increases, that is $Q(\theta^{(t+1)}|\theta^{(t)}, y_{obs}) > Q(\theta^{(t)}|\theta^{(t)}, y_{obs})$

Proof: Let Y be the complete data which can be factored as:

$$f(Y|\theta) = f(y_{obs}, y_{mis}|\theta) = f(y_{obs}|\theta) f(y_{mis}|y_{obs}, \theta) \tag{5}$$

where $f(Y_{obs}|\theta)$ is the density of the Y_{obs} and $f(Y_{mis}|Y_{obs},\theta)$ is the density of the missing data given the observed data.
Then, the log-likelihood decomposition corresponding to (1) is:

$$\ell(\theta|Y) = \ell(\theta|Y_{obs}, Y_{mis}) = \ell(\theta|Y_{obs}) + lnf(Y_{mis}|Y_{obs},\theta) \quad (6)$$

We seek to estimate θ by maximizing the incomplete data likelihood $\ell(\theta|Y_{obs})$ with respect to θ for fixed Y_{obs}, (4) can be arranged as follows:

$$\ell(\theta|Y_{obs}) = \ell(\theta|Y) - lnf(Y_{mis}|Y_{obs},\theta) \quad (7)$$

Then, we take the expectation of both side in (5) over the distribution of the missing data Y_{mis}, given the observed data Y_{obs} and current estimate of θ say $\theta^{(t)}$ we end up with:

$$\ell(\theta|Y_{obs}) = Q(\theta|\theta^{(t)} - H(\theta|\theta^{t}) \quad (8)$$

where

$$Q(\theta|\theta^{t}) = \int [\ell(\theta|Y_{obs}, Y_{mis})]f(Y_{mis}|Y_{obs},\theta^{t})dY_{mis} \quad (9)$$

and

$$H(\theta|\theta^{t}) = \int [lnf(Y_{mis}|Y_{obs},\theta]f(Y_{mis}|Y_{obs},\theta^{t})dY_{mis}. \quad (10)$$

Note that by Jensen's inequality

$$H(\theta|\theta^{t}) \leq H(\theta^{t},\theta^{t}) \quad (11)$$

By considering a sequence of iterates $\theta^{(0)},\theta^{(1)},\cdots$, where $\theta^{(t+1)} = M(\theta^{(t)})$ for some $M(.)$. Then the difference between two successive iterations is:

$$\begin{aligned}\ell(\theta^{(t+1)}|Yobs) - \ell(\theta^{(t)}|Y_{obs}) &= [Q(\theta^{(t+1)}|\theta^{(t)}) - Q(\theta^{(t)}|\theta^{(t)})] \\ &\quad -[H(\theta^{(t+1)}|\theta^{(t)}) - H(\theta^{(t)}|\theta^{(t)})] \quad (12)\end{aligned}$$

Since the EM algorithm chooses $\theta^{(t+1)}$ to maximize $Q(\theta|\theta^{(t)})$ with respect to θ. Hence, the difference between the two Q function will always be positive and the difference between the H function will be always negative by (8). Therefore, we can conclude that: $\ell(\theta^{(t+1)}|y_{obs}) \geq \ell(\theta^{(t)}|y_{obs})$.

Corollary 2.1. *Suppose that for some θ^* in the parameter space of θ, $\ell(\theta^*|Y_{obs}) \geq \ell(\theta|Y_{obs})$ for all θ. Then for every generalized EM algorithm,*

$$\ell(M(\theta^*)|Y_{obs}) = \ell(\theta^*|Y_{obs})$$

$$Q(\delta(\theta^*)|\theta^*) = Q(\theta^*|\theta^*)$$

and

$$f(Y_{mis}|Y_{obs},\delta(\theta^*)) = f(Y_{mis}|Y_{obs},\theta^*)$$

almost everywhere.

3.3. LogF(1,1) penalty when data are missing at random

The $LogF(1,1)$ was derived from a generalized $LogF(a,b)$ distribution, which is a well known flexible family of distributions based on the log of an F-variate [25–27], where $a = b$ for an approximately normalized distribution. The generalized $LogF(a,b)$ distribution is defined as:

$$f_{a,b}(t) = \frac{1}{B(a,b)}\frac{e^{-bt}}{(1+e^{-t})^{a+b}} \quad -\infty < t < \infty \quad (13)$$

where, $B(a,b)$ is the beta function and both $a,b > 0$. The link with logistic order statistics makes things more apparent, in a way that a and b control skewness and affect tails. However, the $f_{a,b}(t)$ is symmetric if $a = b = 2m$, hence the *logF* distribution, maybe defined as below:

$$f_{2m,2m}(t) = \frac{1}{B(2m,2m)}\frac{e^{-2mt}}{(1+e^{-t})^{4m}} \quad -\infty < t < \infty \quad (14)$$

The model in Equation (14) is also referred to as Type III generalized logistic distribution. Thus, taking logs of the F random variable with equal degrees of freedom parameterized as $(2m, 2m)$ gives rise to an almost symmetric generalization of the logistic distribution. It should be noted in Equation (13) that when both a and b tend to infinity, then the logF distribution tends to the normal distribution [28]. Conversely, we can say that when m in Equation (14) tends to infinity, then the logF distribution tends to the normal distribution.

Greenland and Mansournia [17] proposed a method where the likelihood of the logistic model is penalized by a class of logF prior to reduce bias of order $O(n^{-1})$ for the complete-data. In our study, we introduced a similar class of LogF() prior, but for the MAR assumption and validated its performance in reducing bias in the estimation of the logistic model parameters. So we maximized

$$L^*(\beta) = L(\beta)\left(\frac{1}{\pi}\frac{e^{\left(\frac{\beta}{2}\right)}}{(1+e^{\beta})}\right), \text{ where } \frac{1}{\pi}\frac{e^{\left(\frac{\beta}{2}\right)}}{(1+e^{\beta})} \text{ is logF(1,1)}$$

prior with the penalized log-likelihood function of $\ell^*(\beta) = \ell(\beta) + log[logF(1,1)]$. Further, following the Firth approach, we penalized the likelihood function by the logF(1,1). Thus, applying the Expectation-Maximization (EM) principle, the E-step at the $(t+1)th$ iteration of the log-likelihood becomes:

$$\begin{aligned}Q(\theta|\theta^{(t)}) &= \sum_{i=1}^{n}\sum_{x_{mis},i(j)} w_{ij}^{(t)}\left(logP(y_i|x_i,\beta) + log\frac{1}{\pi} + \frac{\beta}{2} - log(1+e^{\beta})\right) \\ &\quad + \sum_{i=1}^{n}\sum_{x_{mis},i(j)} w_{ij}^{(t)} logP(x_i|\alpha) \quad (15)\end{aligned}$$

Given that y_i comes from a Bernoulli distribution, with variable π_i and penalty $logF(1,1)$, then Equation (15) becomes (16), which was then differentiated with respect to β to give the score function, $U(\beta)$ for the $Q^{(1)}(.)$ function from Equation (3).

$$Q^{(1)}(\beta|\theta^{(t)}) = \sum_{i=1}^{n} \sum_{x_{mis},i(j)} w_{ij}^{(t)} \left(logP(y_i|x_i,\beta) + log\frac{1}{\pi} + \frac{\beta}{2} - log(1+e^{\beta}) \right) \quad (16)$$

The derivation of the score function was done algebraically from $P(y|x,\beta) = (\pi_i)^{y_i}(1-\pi_i)^{1-y_i}$, where $\pi_i = \frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}$. For the M-step, the score function $U(\beta)$ is, thus:

$$U(\beta) = \sum_{i=1}^{n} \sum_{xmis,i,(j)} w_{ij} \left(x_i(y_i - \pi_i) + \frac{1-e^{\beta}}{2(1+e^{\beta})} \right)$$

while the maximization was done iteratively using a relation we derived using the Taylor expansion series for the log-likelihood function:

$$\beta^{(t+1)} = \beta^{(t)} + I(\beta^{(t)})^{-1} U(\beta^{(t)}) \quad (17)$$

The marginal distribution of x is multinomial, a generalization of the Bernoulli distribution. Thus, the value of a random variable can be one of the K mutually exclusive and exhaustive states with their probabilities. Consequently, using the Cox and Snell equation [19] we derived a closed form for the bias with the *sth* component $Bias(\hat{\beta}_s)$ of the estimates defined as:

$$Bias(\hat{\beta}_s) = \sum_{r,t,u} \frac{1}{2} I^{rs} I^{tu} \left(K_{rtu} + 2J_{t,ru}\right) \quad \text{where } r,t,u,s = 1,2,...,d \quad (18)$$

where:

$$K_{rst} = E\left(\sum_{j} \frac{\partial^3 logP_j(y_j,\beta)}{\partial\beta_r\beta_s\beta_t} \right)$$

and

$$J_{t,ru} = E\left(\sum_{j} \frac{\partial logP_j(y_j,\beta)}{\partial\beta_t} \frac{\partial^2 logP_j(y_j,\beta)}{\partial\beta_r\beta_u} \right)$$

and $I^{\cdot\cdot}$ is the inverse of the Fisher information matrix.

On developing the terms in respect to the $LogF(1,1)$ penalty and substituting in Equation (18), we obtain the bias expression as presented in Equation (19). This expression, basically represents an exact bias formula that is derived after a series of steps with the $LogF(1,1)$ penalized log-likelihood function.

$$\begin{aligned} Bias(\hat{\beta}_s) = \sum_{r,t,u} \Bigg\{ & \left(\sum_{i=1}^{n} \sum_{x_{mis},i(j)} w_{ij} \frac{x_{ir}x_{is}\pi_i}{(1+e^{x'\hat{\beta}_s})} + \frac{e^{\hat{\beta}_s}}{(1+e^{\hat{\beta}_s})^2} \right) \\ & + \sum_{i=1}^{n} \left(x_{ir}(y_i-\pi_i) + \frac{1-e^{\hat{\beta}_s}}{2(1+e^{\hat{\beta}_s})} \right) \left(x_{ir}(y_i-\pi_i) + \frac{1-e^{\hat{\beta}_s}}{2(1+e^{\hat{\beta}_s})} \right)' \\ & \left(\sum_{i=1}^{n} \sum_{x_{mis},i(j)} w_{ij} \frac{x_{it}x_{iu}\pi_i}{(1+e^{x_i'\hat{\beta}_s})} + \frac{e^{\hat{\beta}_s}}{(1+e^{\hat{\beta}_s})^2} \right) \\ & + \sum_{i=1}^{n} \left(x_{it}(y_i-\pi_i) + \frac{1-e^{\hat{\beta}_s}}{2(1+e^{\hat{\beta}_s})} \right) \left(x_{it}(y_i-\pi_i) + \frac{1-e^{\hat{\beta}_s}}{2(1+e^{\hat{\beta}_s})} \right)' \\ & \left(\sum_{i=1}^{n} \sum_{x_{mis},i(j)} w_{ij}x_{ir}x_{it}x_{iu} \frac{e^{x_{ir}'\hat{\beta}_s} + e^{2x_{ir}'\hat{\beta}_s}}{(1+e^{x_{ir}'\hat{\beta}_s})^4} \right) \\ & +2 \left(\sum_{i=1}^{n} \sum_{x_{mis},i(j)} w_{ij}x_{it}(y_i-\pi_i) \right) \left(\sum_{i=1}^{n} \sum_{x_{mis},i(j)} w_{ij} \frac{-x_{ir}x_{iu}e^{x'\hat{\beta}_s}}{(1+e^{x'\hat{\beta}_s})^2} \right) \Bigg\} \end{aligned} \quad (19)$$

From Equation (19), the bias notation involves the derivative of the log-likelihood function to the third derivative, so that we say the bias corrects the estimated parameters with logF(1,1) penalty up to $O(n^{-1})$. Further, our proposal for the LogF(1,1) is premised on some new derived theory. That is, when we vary the parameterization in Equation (13) such that $a \neq b \neq 2m$ we are led to a skewed distribution. Hence, when the $logF(a,b)$ is used as a penalty in the logistic model's log-likelihood function, it may not necessarily result in reduced bias of the parameters under the MAR assumption. Conversely, when penalizing the log-likelihood with $logF(2m,2m)$ the resultant parameters are said to be with reduced bias. Thus, the motivation to propose the $logF(2m,2m)$ penalization of the log-likelihood of the logit model, when data are missing at random. The bias expression is parameterized with $(r,t,u,s) \in 1:p$ parameters, the outcome y_i follows a Bernoulli density and π_i is the expected value. Furthermore, the $Bias(\hat{\beta}) = \hat{\beta} - \beta$ could have been derived from the first order ploynomial $L'(\beta) = 0$. However, for more accuracy, we preferred to use the second order polynomial function $L'(\beta) + (\hat{\beta} - \beta)L''(\beta) = 0$. Further simplification of this bias function, led us to an estimate of the exact bias expression presented in Equation (19). The idea is that with a better penalty term such as the one proposed in this study for the case of logistic regression, the estimated model parameters are generated with a reduced bias as estimated in Equation (19). We will show by numerical experimental and real life application in the subsequent Section 4. However, in Section 3.4 we briefly show how the information matrix of the penalized likelihood function was derived.

3.4. Estimating the information matrix using the Louis' method

The Louis method [20] was used to calculate the information matrix of the proposed method. This method gives the information gained from the observed and missing data, while maintaining the notation of the EM algorithm by using the following formula:

$$
\begin{aligned}
I(\theta|obs) =& E(I(\theta|com)|obs) - E(U(\theta|com)U'(\theta|com)|obs)+ \\
& E(U(\theta|com)|obs)E(U'(\theta|com)|obs) \\
=& -\ddot{Q}(\hat{\theta},\hat{\theta}) - E(U(\theta|com)U'(\theta|com)|obs)+ \\
& \sum_i \dot{Q}(\hat{\theta},\hat{\theta})\dot{Q}(\hat{\theta},\hat{\theta})'
\end{aligned} \tag{20}
$$

Where *comp* and *obs* represent complete and observed data, and $\theta = (\beta, \alpha)$. The dots on Q represent the order of the derivative. The function $U(\theta|com) = \sum_i \frac{\partial \ell(\theta, x_i, y_i)}{\partial \theta}$ is the score function based on the complete data and $\ddot{Q}(\hat{\theta},\hat{\theta}) = \sum_{i=1}^{n} \sum_{x_{mis,i(j)}} w_{ij}^{(t)} \partial^2 \frac{\ell(\theta, x_i, y_i)}{\partial\theta\partial\theta'}$, and $\dot{Q}(\hat{\theta},\hat{\theta}) = \sum_{i=1}^{n} \sum_{x_{mis,i(j)}} w_{ij}^{(t)} \partial \frac{\ell(\theta, x_i, y_i)}{\partial\theta}$.
Applying these formulae in Equation (20) with the proposed $LogF(1,1)$ penalty, the information matrix becomes:

$$
\begin{aligned}
I(\theta|obs) =& \sum_{i=1}^{n} \sum_{x_{mis,i(j)}} w_{ij}^{(t)} \left(\frac{x_i^2 \pi_i}{1+e^{x_i'\beta}} + \frac{e^{\beta}}{(1+e^{\beta})^2} \right) \\
+ \sum_{i=1}^{n} & \left(x_i(y_i - \pi_i) + \frac{1-e^{\beta}}{2(1+e^{\beta})} \right) \left(x_i(y_i - \pi_i) + \frac{1-e^{\beta}}{2(1+e^{\beta})} \right)'
\end{aligned} \tag{21}
$$

Since

$$E(U(\theta|com)U'(\theta|com)|obs) = x_i^3 \pi_i \left(\frac{E(y_i - \pi_i)}{1+e^{x_i\beta}} \right)$$

With

$$E(y_i) = \pi_i \longrightarrow E(U(\theta|com)U'(\theta|com)|obs) = 0$$

. Moreover, the observed information matrix can also be obtained by using the Hessian matrix, which is just the negative second derivative of the log-likelihood function.

4. Statistical validation results

4.1. Simulation study

In this section, we present results of the simulation studies to validate and compare the performance of our proposed method against the current classical methods developed by Firth and Ibrahim. The choice for the Firth Method is premised on his proposal to impose the Jeffreys prior penalty function in standard regression as a way of bias reduction of the MLE. The Ibrahim method was chosen because he showed that when data are MAR, the E-step of the EM algorithm can be expressed as a weighted complete log-likelihood when the unobserved covariates are assumed to come from a discrete distribution. We based the validation studies on examining the standard errors of the estimated parameters to avoid the complex computations involved in our derived bias expression [6, 12]. We employed simulation studies for several sample sizes and missingness proportions. Firstly, we simulated three categorical covariates, $x1$, $x2$, $x3$ from a discrete Bernoulli probability mass function $f(x) = p^x(1-p)^{(1-x)}; \quad \forall \ x = \{0,1\} \quad else \quad 0$ and obtained the response y from the Bernoulli distribution with the probability model $log(\pi_i/1-\pi_i) = -1 + (0.5)x_1 + (0.7)x_2 + (0.5)x_3$. Throughout the simulation study, the response variable is kept fully observed, while some missing values were imposed to the three categorical covariates. The missingness mechanism was MAR for all covariates, while the logistic model assumptions for categorical data were addressed before fitting all the models for the simulated data set. To ensure compliance with the model assumptions, the variance inflation factor was calculated and found to be less than five for all the covariates in all simulated cases. The other logit model assumption that requires large sample size was considered to exclude smaller sample ($n < 45$) as well as lower proportions of missingness because we required proportions of missingness that are large enough to create differences in the estimated model parameters. Therefore, during the simulation studies, we varied both the sample size and the missingness proportion (%) with $n = \{45, 60, 80, 120\}$ and $p = \{20, 25, 30, 35\}$, respectively.

The study followed the steps described in Algorithm 1 to validate the proposed method against the existing methods. The symbol n represents sample size, p percentage or proportion of missing data, y a vector of the outcome and $\mathbf{X}$ a matrix of categorical covariates.

Require: X, y, n, d **return** $\hat{\beta}, SE$

$data \leftarrow (X, y)$
do
$ampute \leftarrow data, p$
$complete.data \leftarrow expand.data$
$fit.model \leftarrow logF$
$calculate \leftarrow \hat{\beta}, SEs$
$n \leftarrow n$
$p \leftarrow p$
$repeat \quad until \quad convergence$

Algorithm 1. Validation algorithm for the LogF method.

From Table 1, the first column of the table represents the sample size (n), the second column represents, the proportion of the missingness, the parameter of interest, and the bias reduction method, while the last two columns show, the estimated parameter ($\hat{\beta}$), the standard error of the estimated parameters. The simulation results show that the *Ibrahim*'s method over estimates the model parameters in most of the scenarios, but as the sample size increases some estimates tend to get closer to the true values. For example, the true parameter for variable x_2 is $\beta_2 = 0.7$, while the results from the

Ibrahim's method were {0.8349, 1.0002, 0.4185, 0.6875}, implying better performance with increased sample size. This tendency is consistent throughout the various simulations conducted, except for very few cases, which may be attributed to the randomness used to generate the dataset and the proportion of the missing data.

Further, for the small sample size the *Firth* method performed better, and this may be mainly attributed to its ability to solve the separation problem in the smaller samples. For example, with $n = 45$, we found that the corresponding estimate for x_2 was (0.7023) which is very close to the true estimate ($\beta_2 = 0.7$), however with $n = 45$, the same estimate for x_2 was (0.5806). The estimates obtained from our proposed method, the *logF* were in many cases closer to the true estimates, especially as the sample size and proportion of missingness increased. For instance, estimates for $\beta_2 = 0.7$ as the sample size and missingness increase were $\hat{\beta_2} = \{0.8313, 1.002, 0.4185, 0.6875\}$. The ability to outperform other methods under situations when the sample size increases is a strong attribute toward applicability of our method, since this represents the real world situations.

Several methods have been developed to compute standard errors of parameter estimates by using the EM algorithm. Some methods are known to be specific on the incomplete data problem [29], whereas others are more general [20]. In Louis' method, the observed information matrix is computed on the last iteration of the EM procedure using only the complete-data gradient and second derivative matrix. The observed information matrix needs to be inverted in order to obtain estimates of the covariance matrix. However, since the Louis' method for estimating standard errors is quite technical [30] and given the fact that the purpose of our simulation is simply to validate the proposed method, we opted for a faster, and reliable alternative of examining the SEs of the estimates. The alternative approach uses the negative of the second derivative matrix of the incomplete-data log-likelihood as an estimate of the information matrix, which can be inverted in order to obtain an estimate of the covariance matrix. Overall, the smallest standard errors were observed from the proposed *logF* penalized method, thus demonstrating its superiority over both *Ibrahim* and *Firth* methods in reducing bias under the MAR assumption.

4.2. Application of bias reduction with LogF(1,1) prior on the COVID-19 data

In this section, we validated our method based on a real-life data problem for determining the potential risk factors of COVID-19 patient being admitted to the Intensive Care Unit (ICU). Thus, we compare between our proposed method *logF* against the proposed by *Ibrahim* and *Firth* when data are MAR. The survey data were collected between January 2020 and June 2020 from the Royal Hospital in the Sultanate of Oman [18]. The study aimed to assess the risk factors associated with admissions to the intensive care unit, ICU for the COVID-19 hospitalized patients. Details of this data were described in the section for the motivation of this study. As presented in Figure 1, the covariates include; having conjuctivitis, being unconscious and having a neurological problem. The response

TABLE 1 Validation of the LogF method using standard errors.

Sample	Missing (%)	Parameter	Method	Estimate ($\hat{\beta}$)	SE
45	20	β_0	Ibrahim	−1.1854	0.6968
			Firth	−1.0699	0.6978
			LogF	−1.1854	0.3000
		β_1	Ibrahim	0.5439	4.8729
			Firth	0.5255	0.7926
			LogF	0.5474	0.3076
		β_2	Ibrahim	0.8349	4.8873
			Firth	0.7023	0.8190
			LogF	0.8313	0.3075
		β_3	Ibrahim	0.6908	0.7120
			Firth	0.6245	0.7514
			LogF	0.6908	0.2962
60	25	β_0	Ibrahim	−1.4385	6.2972
			Firth	−1.2450	0.6752
			LogF	−1.4385	0.2720
		β_1	Ibrahim	0.4059	3.3549
			Firth	0.3225	0.7109
			LogF	0.4059	0.2516
		β_2	Ibrahim	1.0002	3.5720
			Firth	0.8887	0.6834
			LogF	1.0002	0.2590
		β_3	Ibrahim	0.9168	3.3415
			Firth	0.7548	0.6868
			LogF	0.9168	0.2501
80	30	β_0	Ibrahim	−0.8624	0.4669
			Firth	−1.1256	0.6064
			LogF	−0.8624	0.1994
		β_1	Ibrahim	0.3883	0.5229
			Firth	0.8095	0.6072
			LogF	0.3883	0.2050
		β_2	Ibrahim	0.4185	0.5138
			Firth	0.6287	0.6128
			LogF	0.4185	0.2051
		β_3	Ibrahim	0.5782	0.5182
			Firth	0.5288	0.6016
			LogF	0.5782	0.2047
120	35	β_0	Ibrahim	−1.0091	0.4647
			Firth	−1.0872	0.5351
			LogF	−1.0091	0.1832
		β_1	Ibrahim	0.4762	0.4306
			Firth	0.4110	0.5210
			LogF	0.4762	0.1638
		β_2	Ibrahim	0.6875	0.4332
			Firth	0.5806	0.5161
			LogF	0.6875	0.1646
		β_3	Ibrahim	0.6781	0.4281
			Firth	0.6032	0.5119
			LogF	0.6781	0.1627

TABLE 2 Performance of LogF against the Ibrahim and Firth methods.

Method	Variable	Estimate ($\hat{\beta}$)	SE	*P*-value
	Intercept	0.2442	0.0581	0.0001
CCA	Conjunctivitis	0.0442	0.0832	0.0596
	Unconscious	0.1453	0.1268	0.0254
	Neurological	−0.0188	0.1137	0.0869
	Intercept	−1.3876	0.2603	0.0001
Ibrahim	Conjunctivitis	1.3036	0.5327	0.0144
	Unconscious	1.2628	0.8026	0.1156
	Neurological	1.6943	1.3083	0.1953
	Intercept	−1.6314	0.3237	0.0001
	Conjunctivitis	1.2775	0.6016	0.0345
Firth	Unconscious	2.5074	1.0381	0.0057
	Neurological	1.6251	1.8230	0.0251
	Intercept	−1.0285	0.1304	0.0001
LogF	Conjunctivitis	0.1105	0.1781	0.0001
	Unconscious	0.6315	0.2210	0.0001
	Neurological	−0.2040	0.2102	0.0001

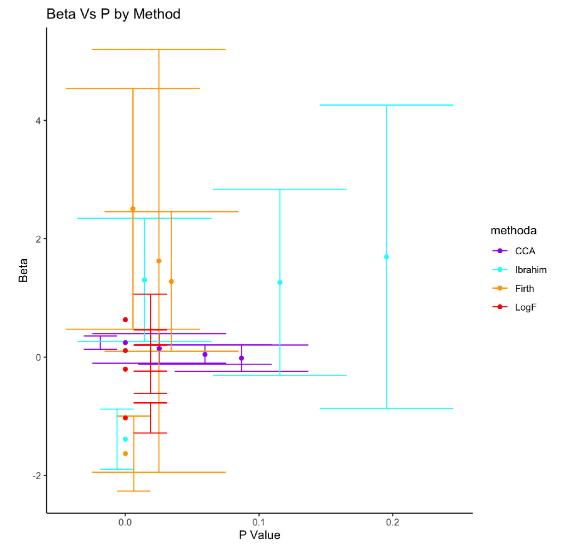

FIGURE 2
Validation of LogF against CCA, Ibrahim, and Firth methods.

variable was whether the patient was admitted to an intensive care unit or not. In the study, the proportions of missing data for each explanatory variable were as follows; conjuctivitis (34%), unconscious (2%), and neurological (4%). We explored by fitting a logistic model for the ICU admission and compared results with the complete case analysis (CCA). The main objective of the study was to obtain estimates for the model based on our proposed method, the *logF* penalty, and to compare the estimates with those obtained from *Ibrahim* and *Firth* methods, by examining their standard errors. The results are summarized in Table 2.

From Table 2, it can be observed that, the parameter estimates under all the methods; *CCA*, *Ibrahim*, *Firth*, and *logF* give different values of the coefficients. They show that lowest standard errors are achieved for the proposed method, implying that narrower confidence intervals for all the three covariates can be obtained using the proposed method, hence reduced bias. For the Ibrahim's method, every unit increase in conjuctivitis, being unconscious and having a neurological there is an increase in the log-odds of patient's ICU admission by 1.3036, 1.2628, and 1.6943, respectively. Whereas, using the *Firth*'s method, for every unit increase in conjuctivitis, being unconscious and having a neurological, the log-odds of ICU admission would increase by 1.2775, 2.5074, and 1.6251, respectively.

Figure 2 demonstrates the variation of estimated parameters ($\hat{\beta}$), their confidence intervals computed from the *SEs*, the probability values and the different methods of estimation. Clearly, the parameter estimates were smallest under the *logF* method of bias reduction as compared to the other methods. Conceptually, under the *logF* method of bias reduction, all variables were statistically significant, a distinct result from other comparable methods of bias reduction, besides the ones derived from the Firth method. Moreover, results of the complete case analysis (CCA) and Ibrahim's method each indicate only one different significant parameter, smaller SEs under the CCA than the Ibrahim's method. Interestingly, the SEs for the Intercept were smallest under the CCA, followed by the LogF method, possibly because CCA analyses only complete data, yet LogF first imputes before parameter estimation.

5. Discussion

In this study, we have demonstrated the importance of bias reduction of estimates for the logistic model under the missing at random mechanism. We conducted model validation studies using both simulation and real life COVID-19 patient data analysis and presented summary statistical results. Our novel method for bias reduction is robust as it relies on the penalization theory of the likelihood function to achieve the efficiency in the parameter estimates for the missing at random data mechanism [6, 31–33]. The fact that we penalized the log-likelihood function by the $LogF(1,1)$ prior and then, derived the closed form of the bias using the Cox and Snell's equation [19] shows how broadbased the novel method is. Besides, the information matrix was derived using the Louis' method [20, 34, 35]. The simulation studies conducted show results, which reveal that irrespective of the sample size, the proposed method based on the *LogF* penalty provides more accurate estimates, characterized by lower standard errors of the model estimates. Furthermore, the results from real life COVID-19 data also support the conclusion that the proposed *LogF* outperforms both the *Firth* and *Ibrahim* methods. These results from the application agree with the clinician's theory that admission to ICU for the COVID-19 patient is significantly explained by the three factors that is, conjuctivitis, unconsciousness and having a neurological condition [36, 37]. We also note that using the complete case analysis alone in modeling may lead to incorrect logit model, as it may be determined that a predictor is not significant, when actually the reverse is true or otherwise.

6. Conclusion

In this study, we have proposed a bias reduction method for the logit model when data are missing at random. Its theoretical aspects have been validated using simulation and real life data problems. The method is based on penalization of the log-likelihood function with Log-F(1,1) penalty chosen for its almost near to normal distribution for the missing at random mechanism. Its related information matrix and the exact form of the bias for that model were derived theoretically. The method is novel, it can efficiently be used to estimate logit model parameters when data are MAR. The results from statistical validation studies show better performance for the novel method when the categorical covariate data are missing at random than other methods. Therefore, we can conclude that the new method shows greater superiority in reducing the bias under MAR in contrast to the current bias reduction methods in estimated parameters by *Ibrahim* and *Firth*. The approach yields more efficient estimates in contrast to the existing classical methods, especially for larger sample sizes and missingness proportions. We recommend further research on bias reduction in parameters under other missing data mechanisms for other generalized linear models. The other research consideration is parameter estimation when the model covariates are not all categorical.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

MA-S conceptualized the study and contributed to the writing. MA-S and RW contributed substantially to the writing. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Santi F, Dickson MM, Giuliani D, Arbia G, Espa G. Reduced-bias estimation of spatial autoregressive models with incompletely geocoded data. *Comput Stat.* (2021) 36:2563–90. doi: 10.1007/s00180-021-01090-7

2. Lee SM, Pho KH, Li CS. Validation likelihood estimation method for a zero-inflated Bernoulli regression model with missing covariates. *J Stat Plann Infer.* (2021) 457:105–27. doi: 10.1016/j.jspi.2021.01.005

3. Jin J, Ma T, Dai J, Liu S. Penalized weighted composite quantile regression for partially linear varying coefficient models with missing covariates. *Comput Stat.* (2021) 36:541–75. doi: 10.1007/s00180-020-01012-z

4. Özkale MR, Lemeshow S, Sturdivant R. Logistic regression diagnostics in ridge regression. *Comput Stat.* (2018) 33:563–93. doi: 10.1007/s00180-017-0755-x

5. Chen YJ, DeMets DL, Gordon Lan K. Increasing the sample size when the unblinded interim result is promising. *Stat Med.* (2004) 23:1023–38. doi: 10.1002/sim.1688

6. Firth D. Bias reduction of maximum likelihood estimates. *Biometrika.* (1993) 80:27–38. doi: 10.1093/biomet/80.1.27

7. Heinze G. A comparative investigation of methods for logistic regression with separated or nearly separated data. *Stat Med.* (2006) 25:4216–26. doi: 10.1002/sim.2687

8. Gaucher S, Klopp O. Maximum likelihood estimation of sparse networks with missing observations. *J Stat Plann Infer.* (2021) 215:299–329. doi: 10.1016/j.jspi.2021.04.003

9. Kosmidis I, Firth D. A generic algorithm for reducing bias in parametric estimation. *Electron J Stat.* (2010) 4:1097–112. doi: 10.1214/10-EJS579

10. Rainey C. Dealing with separation in logistic regression models. *Polit Anal.* (2016) 24:339–55. doi: 10.1093/pan/mpw014

11. Little RJ, Rubin DB. *Statistical Analysis With Missing Data*. Vol. 793. Hoboken, NJ: John Wiley & Sons (2019). doi: 10.1002/9781119482260

12. Ibrahim JG. Incomplete data in generalized linear models. *J Am Stat Assoc.* (1990) 85:765–9. doi: 10.1080/01621459.1990.10474938

13. Ibrahim JG, Lipsitz SR. Parameter estimation from incomplete data in binomial regression when the missing data mechanism is nonignorable. *Biometrics.* (1996) 52:1071–8. doi: 10.2307/2533068

14. Das U, Maiti T, Pradhan V. Bias correction in logistic regression with missing categorical covariates. *J Stat Plann Infer.* (2010) 140:2478–85. doi: 10.1016/j.jspi.2010.02.018

15. Maity AK, Pradhan V, Das U. Bias reduction in logistic regression with missing responses when the missing data mechanism is nonignorable. *Am Stat.* (2018) 73:340–9. doi: 10.1080/00031305.2017.1407359

16. Karl AT, Zimmerman DL. A diagnostic for bias in linear mixed model estimators induced by dependence between the random effects and the corresponding model matrix. *J Stat Plann Infer.* (2021) 211:107–18. doi: 10.1016/j.jspi.2020.06.004

17. Greenland S, Mansournia MA. Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. *Stat Med.* (2015) 34:3133–43. doi: 10.1002/sim.6537

18. Al Awaidy ST, Khamis F, Mahomed O, Wesonga R, Al Shuabi M, Al Shabibi NS, et al. Epidemiological risk factors for acquiring severe COVID-19; prospective cohort study. *Oman Med J.* (2021) 36:e301. doi: 10.5001/omj.2021.127

19. Cox DR, Snell EJ. A general definition of residuals. *J R Stat Soc Ser B.* (1968) 30:248–65. doi: 10.1111/j.2517-6161.1968.tb00724.x

20. Louis TA. Finding the observed information matrix when using the EM algorithm. *J R Stat Soc Ser B*. (1982) 44:226–33. doi: 10.1111/j.2517-6161.1982.tb01203.x

21. Rubin DB. Inference and missing data. *Biometrika*. (1976) 63:581–92. doi: 10.1093/biomet/63.3.581

22. Kim JK, Shao J. *Statistical Methods for Handling Incomplete Data*. Boca Raton, FL: CRC Press (2013). doi: 10.1201/b13981

23. Li W, Yang S, Han P. Robust estimation for moment condition models with data missing not at random. *J Stat Plann Infer*. (2020) 207:246–54. doi: 10.1016/j.jspi.2020.01.001

24. Chen TH, Chatterjee N, Landi MT, Shi J. A penalized regression framework for building polygenic risk models based on summary statistics from genome-wide association studies and incorporating external information. *J Am Stat Assoc*. (2021) 116:133–43. doi: 10.1080/01621459.2020.1764849

25. Prentice RL. Discrimination among some parametric models. *Biometrika*. (1975) 62:607–14. doi: 10.1093/biomet/62.3.607

26. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons (2011).

27. Brown BW, Spears FM, Levy LB. The log F: a distribution for all seasons. *Comput Stat*. (2002) 17:47–58. doi: 10.1007/s001800200098

28. Aroian LA. A study of RA Fisher's z distribution and the related F distribution. *Ann Math Stat*. (1941) 12:429–48.

29. Baker SG. A simple method for computing the observed information matrix when using the EM algorithm with categorical data. *J Comput Graph Stat*. (1992) 1:63–76. doi: 10.1080/10618600.1992.10474576

30. Kang SM, Rafiq A, Kwun YC. A new second-order iteration method for solving nonlinear equations. In: Lee G, editor. *Abstract and Applied Analysis. Vol. 2013*. London: Hindawi (2013). doi: 10.1155/2013/487062

31. Bindele HF, Abebe A, Meyer NK. Robust confidence regions for the semi-parametric regression model with responses missing at random. *Statistics*. (2018) 52:885–900. doi: 10.1080/02331888.2018.1467419

32. Zou YY, Liang HY. Wavelet estimation of density for censored data with censoring indicator missing at random. *Statistics*. (2017) 51:1214–37. doi: 10.1080/02331888.2017.1336170

33. Wang Q. Probability density estimation with data missing at random when covariables are present. *J Stat Plann Infer*. (2008) 138:568–87. doi: 10.1016/j.jspi.2006.10.017

34. Le Cessie S, Van Houwelingen JC. Ridge estimators in logistic regression. *J R Stat Soc Ser C*. (1992) 41:191–201. doi: 10.2307/2347628

35. Cole SR, Chu H, Greenland S. Maximum likelihood, profile likelihood, and penalized likelihood: a primer. *Am J Epidemiol*. (2014) 179:252–60. doi: 10.1093/aje/kwt245

36. Gallo Marin B, Aghagoli G, Lavine K, Yang L, Siff EJ, Chiang SS, et al. Predictors of COVID-19 severity: a literature review. *Rev Med Virol*. (2021) 31:1–10. doi: 10.1002/rmv.2146

37. Elemam NM, Hammoudeh S, Salameh L, Mahboub B, Alsafar H, Talaat IM, et al. Identifying immunological and clinical predictors of COVID-19 severity and sequelae by mathematical modeling. *Front Immunol*. (2022) 13:865845. doi: 10.3389/fimmu.2022.865845