Statistics and Probability Letters 113 (2016) 16–22

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

A weighted simulation-based estimator for incomplete longitudinal data models

Daniel H. Li[a], Liqun Wang[b,*]

[a] Juno Therapeutics, Seattle, WA, USA
[b] University of Manitoba, Winnipeg, Canada

ARTICLE INFO

Article history:
Received 25 October 2015
Received in revised form 8 February 2016
Accepted 11 February 2016
Available online 23 February 2016

Keywords:
Generalized linear mixed model
Inverse probability weighting
Missing data
Simulation-based estimator

ABSTRACT

Recently, Li and Wang (2012a,b) and Wang (2007) have proposed a simulation-based estimator for generalized linear and nonlinear mixed models with complete longitudinal data. This estimator is constructed using the simulation-by-parts technique which leads to the unique feature that it is consistent even using finite number of simulated random points. This paper extends the methodology to deal with incomplete longitudinal data by applying the inverse probability weighting method for the monotone missing-at-random response data. The finite sample performance of this estimator is investigated through simulation studies and compared with the multiple imputation approach.

1. Introduction

In biomedical, environmental and social sciences research, longitudinal data analysis is widely used and constitutes the fundamental statistical research methodologies. Generalized linear mixed models (GLMM) have been widely used in the modeling of longitudinal data. Li and Wang (2012a) proposed a simulation-based estimator (SBE) for GLMM based on the first two marginal moments of the response variables, which does not rely on the normality distribution assumption for random effects. Li and Wang (2012b) extended the SBE to the GLMM where some covariates are measured with error. This approach was originally studied by Wang (2007) for nonlinear mixed effects models. The SBE is constructed using a novel simulation-by-parts technique to ensure its consistency by using finite number of simulated random points. This is the key difference from many other simulation-based estimators proposed in the literature, where they require the number of simulated random points go to infinity to achieve consistency. So far, the SBE is only studied under complete data settings although incomplete or missing data are common in longitudinal studies. For example, in clinical trials, missing data are almost inevitable because subjects may decide to withdraw from the study at anytime prior to completion or subjects are not compliant to protocol for scheduled assessments. Problems arise if the mechanism leading to the missing data depends on the response process. It is known that ignoring missing data and using naive methods may introduce bias, reduce the power of inference and lead to misleading conclusions (Little and Rubin, 2002).

The extension of the SBE to account for incomplete longitudinal data is non-trivial and needs to be addressed to allow this estimator used in more general settings. In this paper, we discuss the validity of SBE under different missing data mechanism and modify it for the data missing at random (MAR) with monotone missingness through the inverse probability weighting (IPW) method. The IPW is a general methodology for constructing parameter estimators in semi-parametric models with complete as well as missing data (Robins et al., 1995; Yi and Cook, 2002). Another popular approach to deal with missing

* Correspondence to: Department of Statistics, University of Manitoba, 186 Dysart Road, Winnipeg, Manitoba, Canada R3T 2N2.
E-mail address: liqun.wang@umanitoba.ca (L. Wang).

http://dx.doi.org/10.1016/j.spl.2016.02.004
0167-7152/

data is the multiple imputation (Rubin, 1987; Schafer, 1997). We also investigate the performance of the SBE using this strategy.

The structure of the paper is as follows. In Section 2, we introduce and review the missing data mechanism, pattern and estimation. Section 3 provides details on the proposed weighted simulation-based estimator, and addresses some practical computational issues. In particular, Section 3.2 handles the data missing completely at random (MCAR), while Section 3.3 focuses on the data MAR. Some simulation studies are conducted in Section 4 to examine the finite sample performance of the proposed estimator, and concluding remarks are given in Section 5.

2. Missing data framework and notation

2.1. Missing data mechanism

To obtain valid inferences, it is essential to consider the reason for missingness. Let Y_{ij} be the jth response for the ith subject, $R_i = (r_{i1}, r_{i2}, \dots, r_{in})'$ be the vector of missing data indicators for $Y_i = (y_{i1}, \dots, y_{in})'$, such that $r_{ij} = 1$ if response y_{ij} is observed, and 0 otherwise. We partition Y_i into Y_i^O and Y_i^M, where Y_i^O contains those y_{ij} for which $r_{ij} = 1$ and Y_i^M contains the remaining components. Assuming $X_i = (x_{i1}, \dots, x_{in})'$ to be a vector of covariates always observed, Little and Rubin (2002) classified missing data mechanism into three types: (1) MCAR, where the missingness is unrelated to the responses so that $P(R_i|Y_i, X_i) = P(R_i|X_i)$. (2) MAR, where the missingness depends only on the observed responses so that $P(R_i|Y_i, X_i) = P(R_i|Y_i^O, X_i)$. This is a weaker and more plausible assumption than MCAR. (3) MNAR, where the missingness depends on both observed and unobserved responses.

2.2. Missing data patterns

There are two broad classes of missing data patterns: intermittent missing and dropout. Intermittent missing pattern refers to the scenario that a subject completes the study but skips a few occasions in the middle of the study period. Dropout (attrition, lost of follow-up) is a particular example of monotone pattern of missingness, which means if one observation is missing, then all subsequent observations are unobserved. Intermittent missing is often easier to deal with because the subject is still participating in the study and the reason of missing values can be ascertained. Dropout is more serious because the subject is no longer available and it is not certain whether the dropout is related to the observed or unobserved outcome. MAR mechanisms are commonly assumed when the interest lies on the parameter estimation (Robins et al., 1995; Lindsey, 2000).

2.3. Estimation of missing data process

Let $\lambda_{ij} = P(r_{ij} = 1|r_{i,j-1} = 1, X_i, Y_i^O)$ be the conditional probability that subject i is observed at time j, given that the subject is present at time $j-1$; and $\pi_{ij} = P(r_{ij} = 1|X_i, Y_i^O)$ be the marginal probability that subject i is present at time j. Then $\pi_{ij} = \prod_{t=2}^{j} \lambda_{it}$. Generally it is assumed that all individuals are observed on the first occasion so that $r_{i1} = \lambda_{i1} = 1$. Further, let $\pi_{ijk} = P(r_{ij} = 1, r_{ik} = 1|X_i, Y_i^O)$ be the probability of observing both y_{ij} and y_{ik} given the response history and covariates. Usually λ_{ij} is estimated using a logistic regression model $\text{logit}\lambda_{ij} = A_{ij}'\alpha$, where A_{ij} is a vector consisting of information on X_i and response history, and α is the vector of parameters (Diggle and Kenward, 1994; Fitzmaurice et al., 1996; Molenberghs et al., 1997; Yi and Cook, 2002).

3. Weighted simulation-based estimator

3.1. GLMM formulation

Suppose subject i is measured repeatedly on n_i occasions. In a GLMM it is assumed that, given the covariates and random effects $b_i \in \mathbb{R}^q$, the responses y_{ij} are conditionally independent and have distribution from the exponential family

$$f(y_{ij}|b_i, X_i, Z_i) = \exp\left\{\frac{\omega_{ij}y_{ij} - a(\omega_{ij})}{\phi} + c(y_{ij}, \phi)\right\}, \quad i = 1, \dots, N,\ j = 1, \dots, n_i, \tag{3.1}$$

where ϕ is a dispersion parameter, ω_{ij} is the canonical parameter and $a(\cdot)$ and $c(\cdot)$ are known functions. The conditional mean and variance

$$\mu_{ij}^c = E(y_{ij}|b_i, X_i, Z_i) = a^{(1)}(\omega_{ij}) \tag{3.2}$$

$$v_{ij}^c = \text{Var}(y_{ij}|b_i, X_i, Z_i) = \phi a^{(2)}(\omega_{ij}) \tag{3.3}$$

satisfy $g^{-1}\{\mu_{ij}^c\} = x_{ij}'\beta + z_{ij}'b_i$ and $\upsilon_{ij}^c = \phi\nu(\mu_{ij}^c)$, where $a^{(d)}$ denotes the dth derivative against ω_{ij}, $g^{-1}(\cdot)$ and $\nu(\cdot)$ are known link and variance functions, respectively. The random effects are assumed to have mean zero and distribution $f_b(u;\theta)$ with unknown parameters $\theta \in \mathbb{R}^r$.

Our approach is motivated by the fact that all the model parameters of interest can be identified and consistently estimated using the first two marginal moments

$$E(y_{ij}|X_i, Z_i) = \int g(x_{ij}'\beta + z_{ij}'u)f_b(u;\theta)du, \tag{3.4}$$

$$E(y_{ij}y_{ik}|X_i, Z_i) = \int g(x_{ij}'\beta + z_{ij}'u)g(x_{ik}'\beta + z_{ik}'u)f_b(u;\theta)du + \delta_{jk}\phi\int \nu(g(x_{ij}'\beta + z_{ij}'u))f_b(u;\theta)du \; j \le k, \tag{3.5}$$

where $\delta_{jk} = 1$ if $j = k$ and 0 otherwise. For example, in a linear model $y_{ij} = x_{ij}'\beta + z_{ij}'b_i + \epsilon_{ij}$ with $b_i \sim N(0, D(\theta))$, the first two marginal moments are $E(y_{ij}|X_i, Z_i) = x_{ij}'\beta$ and $E(y_{ij}y_{ik}|X_i, Z_i) = (x_{ij}'\beta)(x_{ik}'\beta) + z_{ij}'D(\theta)z_{ik} + \delta_{jk}\phi$. Another example is a logistic model $\text{logit}P(y_{ij} = 1|b_i) = x_{ij}'\beta + z_{ij}'b_i$, where the moments are given in (3.4)–(3.5) with logit link function $g(w) = (1 + e^{-w})^{-1}$.

In general the integrals on the right-hand sides of (3.4)–(3.5) are intractable but can be approximated using the Monte Carlo simulation techniques such as importance sampling.

3.2. Simulation-based estimator for the data MCAR

Li and Wang (2012a,b) and Wang (2007) used a simulation-by-parts technique to construct two sets of simulated moments which are unbiased estimates of the true moments. First, a known density $h(u)$ is chosen such that its support covers that of the integrands in (3.4)–(3.5). Second, a set of random points u_{is}, $s = 1, 2, \ldots, 2S$ are generated from $h(u)$ which are used to construct the simulated moments as

$$\mu_{ij,1}(\psi) = \frac{1}{S}\sum_{s=1}^{S}\frac{g(x_{ij}'\beta + z_{ij}'u_{is})f_b(u_{is};\theta)}{h(u_{is})}, \tag{3.6}$$

$$\eta_{ijk,1}(\psi) = \frac{1}{S}\sum_{s=1}^{S}\frac{g(x_{ij}'\beta + z_{ij}'u_{is})g(x_{ik}'\beta + z_{ik}'u_{is})f_b(u_{is};\theta)}{h(u_{is})} + \frac{\delta_{jk}\phi}{S}\sum_{s=1}^{S}\frac{\nu(g(x_{ij}'\beta + z_{ij}'u_{is}))f_b(u_{is};\theta)}{h(u_{is})} \tag{3.7}$$

and $\mu_{ij,2}(\psi)$ and $\eta_{ijk,2}(\psi)$ are constructed similarly using the second half of the points u_{is}, $s = S+1, S+2, \ldots, 2S$. Finally the SBE for $\psi = (\beta', \theta', \phi)'$ is obtained by minimizing

$$Q_{N,S}(\psi) = \sum_{i=1}^{N}\rho_{i,1}'(\psi)W_i\rho_{i,2}(\psi) \tag{3.8}$$

within a compact parameter space Ψ, where $\rho_{i,t}(\psi) = (y_{ij} - \mu_{ij,t}(\psi),\; 1 \le j \le n_i, y_{ij}y_{ik} - \eta_{ijk,t}(\psi),\; 1 \le j \le k \le n_i)'$, $t = 1, 2$, and $W_i = W(X_i, Z_i)$ is a nonnegative definite weight matrix. As is shown in Li and Wang (2012a,b) that in the case of complete data the SBE is consistent and asymptotically normal as $N \to \infty$ for any finite S. Their simulation studies and real data applications have also shown that the SBE works well in the finite sample situations with moderately large S.

Now for the case of data MCAR, we define the SBE $\hat\psi_{N,S}$ as the solution of the score equation

$$\sum_{i=1}^{N}\frac{\partial\rho_{i,1}'(\psi)}{\partial\psi}W_i\Delta_i\rho_{i,2}(\psi) = 0, \tag{3.9}$$

where $\Delta_i = \text{diag}(r_{ij},\; 1 \le i \le n_i, r_{ij}r_{ik},\; 1 \le j \le k \le n_i)$. It is easy to see that (3.9) is an unbiased estimating equation because under MCAR Δ_i does not depend on Y_i and therefore

$$E\left[\frac{\partial\rho_{i,1}'(\psi_0)}{\partial\psi}W_i\Delta_i\rho_{i,2}(\psi_0)\right] = E\left[\frac{\partial\rho_{i,1}'(\psi_0)}{\partial\psi}W_i\Delta_i E(\rho_{i,2}(\psi_0)|X_i, Z_i)\right] = 0,$$

where $\psi_0 = (\beta_0', \theta_0', \phi_0)'$ denotes the true parameter value. Moreover, following Li and Wang (2012a) it is straightforward to show that for a finite S, $\sqrt{N}(\hat\psi_{N,S} - \psi_0) \xrightarrow{L} N(0, B^{-1}CB^{-1})$, where

$$B = E\left[\frac{\partial\rho_{i,1}'(\psi_0)}{\partial\psi}W_i\Delta_i\frac{\partial\rho_{i,2}(\psi_0)}{\partial\psi'}\right] \tag{3.10}$$

and

$$C = E\left[\frac{\partial\rho_{i,1}'(\psi_0)}{\partial\psi}W_i\Delta_i\rho_{i,2}'(\psi_0)\rho_{i,2}(\psi_0)\Delta_iW_i\frac{\partial\rho_{i,1}(\psi_0)}{\partial\psi'}\right]. \tag{3.11}$$

An approximately optimal choice of W_i as derived in Li and Wang (2012a) is given by

$$A(\hat{\psi}_{N1}) = \frac{1}{N}\sum_{i=1}^{N} \rho_{i,1}(\hat{\psi}_{N1})\Delta_i \rho_{i,2}'(\hat{\psi}_{N1}), \tag{3.12}$$

where $\hat{\psi}_{N1}$ is an initial consistent estimator of ψ.

3.3. Weighted simulation-based estimator for the data MAR

Now we modify the SBE to handle the data MAR by using the IPW method. The idea is to weight each subject's contribution in the estimation by the inverse probability that the subject drops out at the time of dropping out (Robins et al., 1995). The weights are obtained based on models for the missing data process as specified in Section 2.3.

Specifically, let $\tilde{\Delta}_i = \mathrm{diag}(r_{ij}/\pi_{ij}, 1 \le j \le n_i, r_{ij}r_{ik}/\pi_{ijk}, 1 \le j \le k \le n_i)$ be the weight matrix accommodating missingness. Then we define the weighted SBE (WSBE) $\tilde{\psi}_{N,S}$ as the solution of

$$\sum_{i=1}^{N} \frac{\partial \rho_{i,1}'(\psi)}{\partial \psi} W_i \tilde{\Delta}_i \rho_{i,2}(\psi) = 0. \tag{3.13}$$

This is an unbiased estimating equation because under MAR $E[\tilde{\Delta}_i|X_i, Z_i, Y_i]$ is an identity matrix and therefore by the law of iterated expectation we have

$$\begin{aligned} E\left[\frac{\partial \rho_{i,1}'(\psi_0)}{\partial \psi} W_i \tilde{\Delta}_i \rho_{i,2}(\psi_0)\right] &= E\left[\frac{\partial \rho_{i,1}'(\psi_0)}{\partial \psi} W_i E[\tilde{\Delta}_i|X_i, Z_i, Y_i]\rho_{i,2}(\psi_0)\right] \\ &= E\left[\frac{\partial \rho_{i,1}'(\psi_0)}{\partial \psi} W_i \rho_{i,2}(\psi_0)\right] = 0. \end{aligned}$$

It follows that $\tilde{\psi}_{N,S}$ is consistent and asymptotically normally distributed with asymptotic covariance matrix $B^{-1}CB^{-1}$, where B and C are given by (3.10) and (3.11) respectively with Δ_i replaced by $\tilde{\Delta}_i$. Similarly, the approximately optimal weight $\tilde{A}_i$ is calculated as in (3.12) with Δ_i replaced by $\tilde{\Delta}_i$.

For the computation of A_i or $\tilde{A}_i$, the moment estimator described in Li and Wang (2012a) needs to be modified because the length of $\tilde{\rho}_i(\psi)$ is different across subjects. The second-order marginal moments can be calculated using (3.7), and the third- and fourth-order moments can be calculated using the same simulation method by constructing the conditional moments first. For all j, k, l, t, $\mathrm{cov}(y_{ij}, y_{ik}y_{il}|b_i, X_i, Z_i) = E(y_{ij}y_{ik}y_{il}|b_i, X_i, Z_i) - E(y_{ij}|b_i, X_i, Z_i)E(y_{ik}y_{il}|b_i, X_i, Z_i)$, and $\mathrm{cov}(y_{ij}y_{ik}, y_{il}y_{it}|b_i, X_i, Z_i) = E(y_{ij}y_{ik}y_{il}y_{it}|b_i, X_i, Z_i) - E(y_{ij}y_{ik}|b_i, X_i, Z_i)E(y_{il}y_{it}|b_i, X_i, Z_i)$. Alternatively, one can adopt the idea of working variance matrix (Prentice and Zhao, 1991; Vonesh et al., 2002) to construct $\tilde{A}_i$. For example, assuming y_i is multivariate normal, then $\mathrm{cov}(y_{ij}, y_{ik}y_{il}|b_i, X_i, Z_i) = \mu_{il}^c\sigma_{ijk} + \mu_{ik}^c\sigma_{ijl}$, and $\mathrm{cov}(y_{ij}y_{ik}, y_{il}y_{it}|b_i, X_i, Z_i) = \sigma_{ijl}\sigma_{ikt} + \sigma_{ijt}\sigma_{ikl} + \mu_{ik}^c\mu_{il}^c\sigma_{ijt} + \mu_{ij}^c\mu_{il}^c\sigma_{ikt} + \mu_{ik}^c\mu_{it}^c\sigma_{ijl} + \mu_{ij}^c\mu_{it}\sigma_{ikl}^c$, where $\sigma_{ijk} = E[(y_{ij} - u_{ij})(y_{ik} - u_{ik})|b_i, X_i, Z_i]$. Thus, both third and fourth moments can be obtained through the first two moments. We can also assume independence among the elements of y_i, in which case the third and fourth moments of y_i are respectively given by

$$\mathrm{cov}(y_{ij}, y_{ik}y_{il}|b_i, X_i, Z_i) = \begin{cases} E[(y_{ij} - u_{ij}^c)^3] + 2\mu_{ij}^c\sigma_{ijj} - 2(\mu_{ij}^c)^3 & \text{if } j = k = l, \\ \sigma_{ijj}\mu_{ik}^c & \text{if } j = l \ne k, \\ \sigma_{ijj}\mu_{il}^c & \text{if } j = k \ne l, \\ 0 & \text{otherwise} \end{cases}$$

and

$$\mathrm{cov}(y_{ij}y_{ik}, y_{il}y_{it}|b_i, X_i, Z_i) = \begin{cases} E[y_{ij}^4] - (\mu_{ij}^c)^2 - \sigma_{ijj} & \text{if } j = k = l = t, \\ E[(y_{ij} - u_{ij}^c)^3]\mu_{it} + 2\mu_{ij}^c\mu_{it}^c\sigma_{ijj} & \text{if } j = k = l \ne t, \\ E[(y_{ij} - u_{ij})^3]\mu_{il}^c + 2\mu_{ij}^c\mu_{il}^c\sigma_{ijj} & \text{if } j = k = t \ne l, \\ 0 & \text{otherwise.} \end{cases}$$

If we further assume that the distribution of y_i is symmetric, then we have $E[(y_{ij} - u_{ij}^c)^3] = 0$.

4. Monte Carlo simulation studies

In this section we conduct simulation studies to assess the finite sample performance of the proposed WSBE under the MCAR and MAR scenarios with various amount of missing data. We consider two models for two types of response variables. In particular, we consider a linear mixed model for the continuous response $y_{ij} = \beta_0 + \beta_1 x_{ij} + b_i + \epsilon_{ij}$ with $\epsilon_{ij} \sim N(0, 1)$, and a mixed Poisson model for the count data $\log E(y_{ij}|b_i) = \beta_0 + \beta_1 x_{ij} + b_i$. In both models we set $\beta' = (1, 1)$ and $b_i \sim N(0, \theta)$ with $\theta = 0.25$. The covariate x_{ij} is generated from normal distribution $N(1, 1)$ in the linear model and $N(0.5, 1)$ in the Poisson model. The missing indicator r_{ij} is generated from the logistic model $\mathrm{logit}\lambda_{ij} = \alpha_0 + \alpha_1 y_{i,j-1}$, with $\alpha' = (2, 0), (2, 0.5), (2, 1)$

Table 1
Simulation results for the linear regression model.

Missingness	N		SBE		WSBE		MI-SBE	
			Bias	RMSE	Bias	RMSE	Bias	RMSE
(2, 0)	50	β_0	−0.085	0.208	−0.088	0.217	0.133	0.203
		β_1	0.035	0.123	0.036	0.126	−0.136	0.170
		θ	0.203	0.372	0.209	0.391	−0.020	0.143
		ϕ	−0.101	0.267	−0.104	0.271	0.159	0.271
	100	β_0	−0.047	0.151	−0.050	0.154	0.137	0.171
		β_1	0.020	0.085	0.021	0.086	−0.144	0.160
		θ	0.120	0.277	0.127	0.290	−0.006	0.113
		ϕ	−0.056	0.195	−0.059	0.198	0.228	0.281
	200	β_0	−0.023	0.107	−0.022	0.107	0.133	0.153
		β_1	0.009	0.064	0.008	0.063	−0.142	0.151
		θ	0.066	0.200	0.067	0.204	0.007	0.082
		ϕ	−0.035	0.142	−0.036	0.143	0.247	0.270
	500	β_0	−0.009	0.070	−0.010	0.072	0.132	0.140
		β_1	0.002	0.038	0.003	0.039	−0.140	0.144
		θ	0.021	0.131	0.023	0.138	0.017	0.061
		ϕ	−0.005	0.086	−0.004	0.084	0.257	0.268
(2, 0.5)	50	β_0	−0.068	0.186	−0.113	0.205	0.056	0.146
		β_1	0.031	0.114	0.038	0.113	−0.062	0.115
		θ	0.243	0.387	0.151	0.326	−0.023	0.132
		ϕ	−0.158	0.266	−0.022	0.226	0.052	0.212
	100	β_0	−0.040	0.139	−0.083	0.156	0.066	0.118
		β_1	0.021	0.078	0.028	0.079	−0.074	0.101
		θ	0.182	0.307	0.097	0.255	−0.015	0.105
		ϕ	−0.120	0.207	0.014	0.178	0.111	0.186
	200	β_0	−0.004	0.099	−0.046	0.107	0.073	0.101
		β_1	0.016	0.080	0.012	0.060	−0.073	0.085
		θ	0.118	0.218	0.024	0.172	−0.002	0.076
		ϕ	−0.099	0.158	0.037	0.134	0.121	0.160
	500	β_0	0.011	0.063	−0.031	0.070	0.070	0.082
		β_1	−0.011	0.085	0.006	0.036	−0.071	0.078
		θ	0.070	0.139	−0.028	0.123	0.002	0.048
		ϕ	−0.076	0.107	0.065	0.102	0.127	0.142
(2, 1)	50	β_0	−0.059	0.182	−0.134	0.237	0.053	0.149
		β_1	0.029	0.110	0.049	0.120	−0.051	0.104
		θ	0.266	0.402	0.160	0.333	−0.027	0.136
		ϕ	−0.173	0.263	−0.029	0.229	0.023	0.200
	100	β_0	−0.029	0.133	−0.108	0.196	0.051	0.103
		β_1	0.015	0.075	0.037	0.092	−0.054	0.086
		θ	0.181	0.309	0.090	0.282	−0.028	0.104
		ϕ	−0.120	0.203	0.027	0.194	0.070	0.170
	200	β_0	0.022	0.138	−0.084	0.146	0.054	0.088
		β_1	0.010	0.079	0.025	0.069	−0.052	0.071
		θ	0.136	0.230	0.042	0.194	−0.017	0.076
		ϕ	−0.109	0.162	0.046	0.147	0.085	0.136
	500	β_0	0.019	0.064	−0.059	0.091	0.053	0.069
		β_1	−0.004	0.076	0.006	0.041	−0.050	0.058
		θ	0.083	0.147	−0.030	0.132	−0.008	0.050
		ϕ	−0.087	0.114	0.084	0.098	0.092	0.115

for the linear model and $\alpha' = (3, 0), (0.5, 0.1), (0.5, 0.5)$ for the Poisson model respectively. Note that $\alpha_1 = 0$ represents the scenario of data MCAR, while $\alpha_1 \neq 0$ represents data MAR. For a given α_0, the smaller α_1 results in higher percentage of missing data. The combined choice of α_0 and α_1 leads to about 10%–40% drop-out, spread over time points 2–4. Therefore, these parameter setups not only lead to different missing data mechanism but also different percentage of missing data.

For comparisons, we also calculate the naive SBE that ignores the missing data and the SBE based on the multiple imputed data. The multiple imputation is done using the `R` package `MICE` with predictive mean matching method and iteration time as the seed for random number generation (Horton and Lipsitz, 2001). Further, we set the number of multiple imputations to be 5 which is generally sufficient to yield efficient results (Rubin, 1987). The sample sizes are $N = 50, 100, 200, 500$ and the number of observations per subject is $n_i = 4$. In each simulation we generate $M = 500$ datasets and report the average biases $(1/M)\sum_{i=1}^{M}\hat{\psi}_i - \psi_0$ and the root mean square errors (RMSE) $(\sum_{i=1}^{M}(\hat{\psi}_i - \psi_0)^2/M)^{-1/2}$.

Tables 1 and 2 contain the numerical results for two models respectively. These results show that in the case of MCAR ($\alpha_1 = 0$), the SBE and WSBE perform similarly, which is consistent with theory. However, in linear model the MI-SBE has

Table 2
Simulation results for the Poisson regression model.

Missingness	N		SBE		WSBE		MI-SBE	
			Bias	RMSE	Bias	RMSE	Bias	RMSE
(3, 0)	50	β_0	−0.155	0.526	−0.156	0.507	0.089	0.166
		β_1	−0.042	0.392	−0.046	0.379	−0.129	0.277
		θ	0.148	0.494	0.153	0.478	−0.169	0.203
	100	β_0	−0.143	0.389	−0.153	0.403	0.041	0.090
		β_1	−0.015	0.252	−0.016	0.247	−0.081	0.150
		θ	0.137	0.339	0.148	0.350	−0.041	0.080
	200	β_0	−0.090	0.283	−0.087	0.276	0.034	0.069
		β_1	−0.026	0.194	−0.025	0.195	−0.076	0.123
		θ	0.089	0.251	0.084	0.241	−0.027	0.060
	500	β_0	−0.087	0.254	−0.083	0.264	0.045	0.049
		β_1	−0.029	0.218	−0.030	0.216	−0.032	0.081
		θ	0.080	0.237	0.077	0.232	−0.037	0.056
(0.5, 0.1)	50	β_0	−0.275	0.555	−0.208	0.726	0.185	0.239
		β_1	−0.054	0.395	0.076	0.464	−0.299	0.382
		θ	0.407	0.620	0.079	0.536	−0.210	0.243
	100	β_0	−0.299	0.504	−0.145	0.410	0.026	0.086
		β_1	−0.034	0.255	0.023	0.280	−0.149	0.202
		θ	0.362	0.532	0.068	0.328	0.036	0.081
	200	β_0	−0.275	0.465	−0.093	0.333	0.016	0.061
		β_1	−0.044	0.256	0.008	0.228	−0.141	0.177
		θ	0.337	0.486	0.017	0.258	0.049	0.071
	500	β_0	−0.284	0.434	−0.043	0.247	0.018	0.032
		β_1	−0.081	0.225	0.003	0.127	−0.027	0.122
		θ	0.405	0.517	−0.110	0.209	0.026	0.067
(0.5, 0.5)	50	β_0	−0.182	0.460	−0.189	0.645	0.164	0.222
		β_1	−0.095	0.343	0.034	0.388	−0.214	0.316
		θ	0.310	0.519	0.050	0.527	−0.189	0.216
	100	β_0	−0.262	0.488	−0.150	0.518	0.014	0.087
		β_1	−0.034	0.246	0.042	0.277	−0.075	0.164
		θ	0.337	0.494	0.043	0.404	−0.025	0.065
	200	β_0	−0.205	0.392	−0.078	0.273	0.008	0.061
		β_1	−0.044	0.292	0.018	0.198	−0.071	0.130
		θ	0.285	0.421	−0.015	0.220	−0.003	0.047
	500	β_0	−0.160	0.344	−0.060	0.207	0.005	0.060
		β_1	−0.055	0.227	0.009	0.106	−0.009	0.100
		θ	0.261	0.369	−0.008	0.124	−0.003	0.042

larger bias and RMSE than the other two estimators except for the variance parameter θ due to the relatively high percentage of missing data. When we repeat the simulation with lower percentage of missing data, the MI-SBE performs actually slightly better than the SBE and WSBE. Moreover, the RMSE of all estimators reduce when the sample size increases. In the case of MAR ($\alpha_1 \neq 0$), the WSBE has smaller bias and RMSE than the SBE for variance parameters. For the regression parameters, the WSBE performs similarly as the SBE for small sample sizes, but improves fast and is significantly better for large sample sizes. In general, MI-SBE clearly outperforms the WSBE and SBE, especially in the Poisson model. This is not surprising since it is documented in the literature that MI is generally more efficient than the IPW method (Robins et al., 1995). More general discussions about the IPW and MI methods can be found in e.g., Carpenter et al. (2006) and Seaman et al. (2012). To improve efficiency, one may consider applying augmented inverse probability weight method (Robins et al., 1995). Furthermore, we notice that the numerical computation of the MI-SBE is more stable. We have repeated our simulations with various values of (α_0, α_1) and observed similar patterns as discussed above.

5. Concluding remarks

Incomplete longitudinal data are common in practical applications. For a valid analysis, a study of the missing mechanism is necessary. Although comprehensive theoretical work and application of the SBE for GLMM were discussed in Li and Wang (2012a,b), there is still a strong need to examine this approach when missing data are present. In this paper, we show that the SBE based on observed data is only valid for data MCAR, and hence we adopt the inverse probability weighting method to construct the WSBE for data MAR. We also investigate the performance of SBE for incomplete longitudinal data by the means of multiple imputation. Our simulation studies demonstrate that the proposed WSBE is feasible to compute, performs well under finite sample sizes, and is comparable to the multiple imputation approach in many cases. Furthermore, this paper suggests a few ways to compute the optimal weight matrix under the incomplete longitudinal data setting. Since the weight matrix contains the third and fourth moments, the computation can be cumbersome even using simulated moments. In our experience, diagonal weight works quite well and can reduce the computational burden substantially.

The proposed WSBE is formulated under the GLMM framework, however, it can be easily extended to nonlinear mixed effects models. In principle, the WSBE and MI-SBE can also be used for data with intermittent missing pattern or longitudinal data with unequally spaced repeated measures. Missing data and measurement error often arise simultaneously in a real world problem, so it would be valuable to develop the proposed methodology to cope with these situations. Another future research is to further extend the SBE to deal with the non-ignorable missing data problems.

Acknowledgments

The authors are grateful to the editor and two anonymous referees for their valuable comments and suggestions that helped to improve this paper. The research was supported by grant RGPIN/227197-2009 from the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

Carpenter, J.R., Kenward, M.G., Vansteelandt, S., 2006. A comparison of multiple imputation and doubly robust estimation for analyses with missing data. J. Roy. Statist. Soc. Ser. A 169, 571–584.
Diggle, P., Kenward, M.G., 1994. Informative drop-out in longitudinal data analysis (with discussion). Appl. Stat. 43, 49–93.
Fitzmaurice, G.M., Laird, N.M., Zahner, G.E.P., 1996. Multivariate logistic models for incomplete binary responses. J. Amer. Statist. Assoc. 91, 99–108.
Horton, N.J., Lipsitz, S.R., 2001. Multiple imputation in practice: comparison of software packages for regression models with missing variables. Amer. Statist. 55, 244–254.
Li, H., Wang, L., 2012a. A consistent simulation-based estimator in generalized linear mixed models. J. Stat. Comput. Simul. 82, 1085–1103.
Li, H., Wang, L., 2012b. Consistent estimation in generalized linear mixed models with measurement error. J. Biomet. Biostat. S7:007. http://dx.doi.org/10.4172/2155-6180.S7-007.
Lindsey, J.K., 2000. Dropouts in longitudinal studies: definitions and models. J. Biopharm. Statist. 10, 503–525.
Little, R.J.A., Rubin, D.B., 2002. Statistical Analysis with Missing Data, second ed. John Wiley & Sons, Hoboken, NJ.
Molenberghs, G., Kenward, M.G., Lesaffre, E., 1997. The analysis of longitudinal ordinal data with nonrandom drop-out. Biometrika 84, 33–44.
Prentice, R.L., Zhao, L.P., 1991. Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. Biometrics 47, 825–839.
Robins, J.M., Rotnitzky, A., Zhao, L.P., 1995. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. J. Amer. Statist. Assoc. 90, 106–121.
Rubin, D.B., 1987. Multiple Imputation for Nonresponse in Surveys. Wiley, New York.
Schafer, J.L., 1997. Analysis of Incomplete Multivariate Data. Chapman & Hall, New York.
Seaman, S.R., White, I.R., Copas, A.J., Li, L., 2012. Combining multiple imputation and inverse-probability weighting. Biometrics 68, 129–137.
Vonesh, E.F., Wang, H., Nie, L., Majumdar, D., 2002. Conditional second-order generalized estimating equations for generalized linear and nonlinear mixed effects models. J. Amer. Statist. Assoc. 97, 271–283.
Wang, L., 2007. A unified approach to estimation of nonlinear mixed effects and Berkson measurement error models. Canad. J. Statist. 35, 233–248.
Yi, G.Y., Cook, R.J., 2002. Marginal methods for incomplete longitudinal data arising in clusters. J. Amer. Statist. Assoc. 97, 1071–1080.