



# HHS Public Access

Author manuscript

*Paediatr Perinat Epidemiol.* Author manuscript; available in PMC 2018 September 01.

Published in final edited form as:

*Paediatr Perinat Epidemiol.* 2017 September ; 31(5): 468–478. doi:10.1111/ppe.12382.

## A two-step approach for analysis of nonignorable missing outcomes in longitudinal regression: an application to Upstate KIDS Study

Danping Liu<sup>a</sup>, Edwina H. Yeung<sup>b</sup>, Alexander C. McLain<sup>c</sup>, Yunlong Xie<sup>d</sup>, Germaine M. Buck Louis<sup>e</sup>, and Rajeshwari Sundaram<sup>a</sup>

<sup>a</sup>Biostatistics and Bioinformatics Branch, Division of Intramural Population Health Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA

<sup>b</sup>Epidemiology Branch, Division of Intramural Population Health Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA

<sup>c</sup>Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA

<sup>d</sup>Glotech, Inc., Division of Intramural Population Health Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA

<sup>e</sup>Office of the Director, Division of Intramural Population Health Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA

### Abstract

**Background**—Imperfect follow-up in longitudinal studies commonly leads to missing outcome data that can potentially bias the inference when the missingness is nonignorable; that is, the propensity of missingness depends on missing values in the data. In the Upstate KIDS Study, we seek to determine if the missingness of child development outcomes is nonignorable, and how a simple model assuming ignorable missingness would compare with more complicated models for a nonignorable mechanism.

**Methods**—To correct for nonignorable missingness, the shared random effects model (SREM) jointly models the outcome and the missing mechanism. However, the computational complexity and lack of software packages has limited its practical applications. This paper proposes a novel two-step approach to handle nonignorable missing outcomes in generalized linear mixed models. We first analyze the missing mechanism with a generalized linear mixed model and predict values

---

Correspondence to: Danping Liu, PhD, Biostatistics and Bioinformatics Branch, Division of Intramural Population Health Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, 6710B Rockledge Drive, Room 3214, MSC 7004, Bethesda, MD 20892-7004, USA. Telephone: (301) 443-7041 danping.liu@nih.gov.  
DR. DANPING LIU (Orcid ID : 0000-0001-5483-6303)  
DR. EDWINA YEUNG (Orcid ID : 0000-0002-3851-2613)

of the random effects; then, the outcome model is fitted adjusting for the predicted random effects to account for heterogeneity in the missingness propensity.

**Results**—Extensive simulation studies suggest that the proposed method is a reliable approximation to SREM, with a much faster computation. The nonignorability of missing data in the Upstate KIDS Study is estimated to be mild to moderate, and the analyses using the two-step approach or SREM are similar to the model assuming ignorable missingness.

**Conclusions**—The two-step approach is a computationally straightforward method that can be conducted as sensitivity analyses in longitudinal studies to examine violations to the ignorable missingness assumption and the implications relative to health outcomes.

### Keywords

Longitudinal data; maximum likelihood; nonignorable missing outcomes; shared random effect model; two-step estimation

---

Missing outcome data are a common problem in observational longitudinal studies due to imperfect follow-up. No matter how well a study is designed, participants may not respond or complete follow-up. For example, in studying children, parents may find it difficult to complete follow-up questionnaires or participate in clinical exams due to competing responsibilities. It is also possible that parents of children with poor developmental outcomes are likely to drop out of the study due to added responsibilities, or conversely, to be highly motivated to remain in the study due to their concerns. With the percentage of missing information being relatively high for many observational studies, researchers often ask questions whether the statistical inference remains valid ignoring missing data, and how sensitive the estimation is under different assumptions of missingness mechanism.

This paper focuses on missing outcomes, while covariates are assumed to be fully observed. A logical way to think of the underlying generation of missing data is through a “missing data mechanism” process, i.e., a model to explain the reasons for missingness. Little and Rubin<sup>1</sup> defined three missing mechanisms: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). The formal definitions and examples of these missing mechanisms are given in Appendix S1. Briefly, a missing data mechanism is MCAR if the propensity of missingness depends on the observed covariates, but not observed outcomes; MAR assumes that the propensity of missingness depends only on the observed components of the data (both the covariates and observed outcomes), not on the missing components; MNAR assumes that the propensity of missingness can depend on missing values in the data. MNAR is more likely to become a practical problem when the missing proportion is high, or when the reasons of missingness are either not adequately understood or unmeasured.

In practice, a simple and commonly used approach to deal with missing longitudinal outcomes is *available case analysis (ACA)*, i.e., including all the observed outcome values for each participant in the analysis. ACA with generalized estimating equations is valid under the MCAR mechanism. In contrast, ACA with a likelihood-based inference, such as generalized linear mixed models (GLMM), is valid under the MAR mechanism if all the variables associated with missingness are included in the model. Model estimation is

performed through maximizing the likelihood of the observed outcome, and the missing mechanism model does not need to be estimated. Therefore, MAR is also referred to as *ignorable missingness*, and MNAR as *nonignorable missingness*. The term “nonignorable” reflects that under MNAR the missing mechanism cannot be ignored and needs to be estimated together with the outcome model. Hereafter in this paper, we use nonignorable missingness to indicate MNAR, and ignorable missingness to indicate MAR.

Wu and Carroll<sup>2</sup> developed a shared random effects model (SREM) for analyzing nonignorable missing longitudinal outcomes. They proposed modeling the outcome and missing mechanism jointly by conditioning on the same set of random effects. The SREM is a plausible construct of the true data generating process, where the participants with higher propensity of missingness tend to also have larger (or smaller) expected outcomes. The idea of SREM was further developed to analyze nonignorable missing data in many scenarios<sup>3–5</sup>. Albert and Follmann<sup>6</sup> provided an extensive review of these methods. This joint modelling framework has also gained much attention in analyzing semicontinuous data<sup>7,8</sup>, in modelling correlated longitudinal and survival outcomes<sup>9</sup>, and in prediction models<sup>10</sup>.

Even though the SREM is well documented in statistical literature, it has two disadvantages that limit its applications in epidemiologic research. First, the likelihood of SREM involves intractable integration over the random effects. The numerical approximation of the integration by Gaussian quadrature is quite time-consuming even for two random effects. If additional random effects are included, the computation load can become prohibitively. Second, there are no standard software packages to implement SREM with intermittent missing patterns, so users have to write their own computer programs for model estimation. The “JM” package in R handles nonignorable dropout by jointly modelling the dropout time and longitudinal outcome<sup>11</sup>, but it cannot deal with intermittent missing patterns.

This paper proposes a novel two-step approach to handle nonignorable missing outcomes in generalized linear mixed models. We first estimate the missing mechanism by a generalized linear mixed model and calculate the predicted values of the random effects; then the outcome model is fitted adjusting for the predicted random effects, so as to account for unexplained heterogeneity in the missingness propensity. We evaluate this new analytic approach in comparison with existing methods in both simulated data sets and also in the recently completed Upstate KIDS Study, and offer recommendations for its practical use. We provide R code in Appendix S2 to facilitate practical use of this method.

## UPSTATE KIDS STUDY

The Upstate KIDS Study is a longitudinal birth cohort study ( $n = 4989$  infants) focusing on mode of conception (use of infertility treatment or not) and children’s growth and development through 3 years of age. Briefly, mothers delivering a live birth between 2008 and 2010 were sampled on mode of conception designed by infertility treatment being noted on the birth certificate and randomly selected 3 unassisted conceptions matched on geographic region; methods have been previously published<sup>12</sup>. Sampling weights were calculated to account for oversampling of mothers with infertility treatment and twin births. Participants were recruited at approximately 4 months postpartum, and were queried every

4–6 months through 3 years of age (totaling 7 times) to capture parental rating of children's development using the Ages and Stages Questionnaire (ASQ) that measures 5 domains: communication, gross motor, fine motor, problem solving and personal-social skills. Each domain ranges 0–60 points, and scores 2 standard deviations or more below the mean for normative data indicates failure.<sup>13,14</sup> The study outcome for this paper is developmental delay (yes/no) as measured by a fail score on the ASQ for any of the 5 domains. For illustration of the methodologic question under study, preterm birth or < 37 completed weeks gestation as noted on the birth certificate comprised the exposure. We hypothesized that preterm infants were more likely to fail the ASQ assessment across the seven time points under study in comparison to term infants. Other *a priori* defined covariates included: mother's age, race, education, marital status, smoking status before and during pregnancy, and health insurance status, which were collected from the maternal baseline questionnaire or birth certificates.

Previously, we examined the association between infertility treatment and developmental delay as measured by failing an ASQ domain under the ignorable missingness assumption<sup>15</sup>. After removing a small percentage of participants (4.4%) with missing covariates, our analysis sample contained  $n = 4767$  study participants. Although not unique to prospective cohorts<sup>16,17</sup>, an increasing percentage of missingness occurred over time (from 14.3% for the first to 65.3% for the seventh study follow-up). The mean number of follow-ups per participant is 3.5/7.0, with 48.2% of children having 1–3 follow-ups, 38.1% having 4–6 follow-ups, and 10.3% having all 7 follow-ups. Our analytic sample also included 3.5% of children with no ASQ observations who only contributed to the estimation of the missing mechanism and not outcome model. Given the percentage of missing information, we sought to determine if the propensity of missingness is correlated or not with the missing ASQ outcomes after conditioning on all the observed variables. We examined how a simple model assuming ignorable missingness would compare with more complicated models for the nonignorable mechanism by developing an approach relevant for observational research with missing data.

## METHODS

Let  $Y_{ij}$  denote the binary outcome variable for participant  $i$  at time point  $j$ , for  $i = 1, \dots, n$  and  $j = 1, \dots, M$ . Suppose the outcome is subject to intermittent missingness. Let  $R_{ij}$  be the missing indicator of  $Y_{ij}$ , where  $R_{ij} = 1$  if  $Y_{ij}$  is observed and 0 otherwise. Let  $\mathbf{X}_{ij}$  and  $\mathbf{W}_{ij}$  denote the design vector of observed covariates that are associated with  $Y_{ij}$  and  $R_{ij}$ , respectively. These covariates could include both time-constant and time-varying variables. Both SREM and our proposed two-step approach assume that only the longitudinal outcome  $Y_{ij}$  is subject to missingness, while covariates  $\mathbf{X}_{ij}$  and  $\mathbf{W}_{ij}$  are fully observed. This assumption seems reasonable in our application, since most of the covariates were collected at the baseline on the participant level. The only time-varying covariate is observation time itself. The parameters of interest are the odds ratios or the association between covariates  $\mathbf{X}_{ij}$  and the binary outcome  $Y_{ij}$ .

### Shared random effects model (SREM)

The shared random effects models (SREM) was first proposed by Wu and Carroll<sup>2</sup> to deal with nonignorable dropout in longitudinal data. We briefly describe the model here and more details are given in Appendix S3. The idea of SREM is that, two GLMMs are specified for  $Y_{ij}$  and  $R_{ij}$ , and their correlation is explained by sharing the random effects:

$$\text{logit } P(Y_{ij}=1 | \mathbf{X}_{ij}, b_i) = \mathbf{X}_{ij}^T \boldsymbol{\beta} + b_i, \quad (1)$$

$$\text{logit } P(R_{ij}=1 | \mathbf{W}_{ij}, c_i) = \mathbf{W}_{ij}^T \boldsymbol{\gamma} + c_i, \quad (2)$$

where  $\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$  is the link function. The random intercepts  $b_i$  and  $c_i$  follow a bivariate normal distribution:

$$\begin{pmatrix} b_i \\ c_i \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_b^2 & \rho\sigma_b\sigma_c \\ \rho\sigma_b\sigma_c & \sigma_c^2 \end{pmatrix} \right).$$

While others sometimes use the same random effects in both models, our setting is more realistic in describing the underlying missing mechanism<sup>18</sup>. The correlation parameter  $\rho$  dictates the strength of dependence between  $Y_{ij}$  and  $R_{ij}$ , and hence the degree of nonignorability, where  $\rho = 0$  implies an ignorable missing mechanism. The outcome model in equation (1) is the primary model of interest, where the coefficients  $\boldsymbol{\beta}$  are interpreted as the log odds ratios between various covariates and the outcome of interest. As shown in Appendix S1, the model parameters  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma_b, \sigma_c, \rho)$  are estimated by maximizing the observed data likelihood that integrates out the random effects. Because of the intractable integration, Gaussian-Hermite quadrature<sup>19</sup> is used to approximate the likelihood, which we implement in R.

### A two-step approach

The heavy computation load and lack of statistical software for evaluating the multidimensional integration in SREM is a serious obstacle for its practical use in epidemiological studies. We propose an alternative two-step approach to avoid direct maximization of the likelihood.

We note that the missing mechanism model (equation (2)) alone can be estimated from the observed data, because both  $R_{ij}$  and  $\mathbf{W}_{ij}$  are fully observed. The estimation for GLMM are readily available in many statistical software packages. In the first step, we estimate the model in equation (2) and compute the predicted random effects as  $\hat{c}_i = E(c_i | \mathbf{R}_i)$ , where  $\mathbf{R}_i = (R_{i1}, \dots, R_{iM})^T$  is the vector of missing indicators for participant  $i$ . In the second step we estimate the outcome model additionally adjusting for  $\hat{c}_i$  as if it were an observed covariate:

$$\text{logit } P(Y_{ij}=1 | \mathbf{X}_{ij}, b_i) = \mathbf{X}_{ij}^T \boldsymbol{\beta} + \alpha \hat{c}_i + b_i^* \quad (3)$$

In other words, we propose to estimate the models in equations (2) and (3) separately, instead of jointly estimating the models in equations (1) and (2). By additionally adjusting for the estimated propensity of missingness,  $\hat{c}_i$ , we hope to achieve valid inference on  $\boldsymbol{\beta}$  as if it were estimated from SREM.

We make a heuristic explanation why this can be effective. Note that  $b_i$  can be decomposed as two independent random variables:

$$b_i = E(b_i | c_i) + (b_i - E(b_i | c_i)) = \rho \frac{\sigma_b}{\sigma_c} c_i + b_i^* \quad (4)$$

The first variable  $E(b_i | c_i)$  is a linear function in  $c_i$  following the bivariate normal distribution; because the conditional expectation is an orthogonal projection, it can be shown that the second variable  $b_i^*$  is independent of  $c_i$ , which we can view as the new random effect term in equation (3). Substituting equation (4) into equation (1), we have

$$\text{logit } P(Y_{ij}=1 | \mathbf{X}_{ij}, b_i) = \mathbf{X}_{ij}^T \boldsymbol{\beta} + \alpha c_i + b_i^*,$$

where  $\alpha = \rho \frac{\sigma_b}{\sigma_c}$ , the coefficient of random effect  $c_i$  in the outcome model. This is very similar to equation (3) except that  $\hat{c}_i$  is replaced by  $c_i$ . If  $c_i$  were observed, we would have used  $c_i$  in equation (3). Without observing  $c_i$ ,  $\hat{c}_i$  is a reasonably good guess from the longitudinal  $R_{ij}$  when the sample size is large and the cluster size is not too small. Ideally, we may perform a regression calibration<sup>20</sup> to account for the estimation error in  $\hat{c}_i$ , similar to a model with measurement error in the covariates. But as we show in the simulation study, the model in equation (3) has good performance even without regression calibration. This is because the error in  $\hat{c}_i$  only impacts the estimation of the nuisance parameter  $\alpha$ , not  $\boldsymbol{\beta}$ , our parameters of interest.

Although in this paper we set up the model only for binary outcome with logit link, we would expect the two-step approach to work well under other regression models (such as Poisson random effect models to estimate the risk ratio). This is because a consistent estimation of the missing mechanism model alone would yield reasonable prediction of the random effects, and as shown in equations (3) – (4), the idea of decomposing the random effect and replacing it with the predicted value could apply.

### Design of simulation studies

We conducted three sets of simulation studies to investigate the performance of the proposed two-step approach in different scenarios. Additional simulations are presented in Appendix S4 and Tables S2–S5. All the simulations were repeated 1000 times.

The first set of simulation studies were performed to compare bias and confidence interval coverage of three different approaches to deal with missing data: (i) SREM, (ii) ordinary GLMM assuming ignorable missingness, and (iii) the proposed two-step approach. We considered two participant-level covariates,  $X_{i(1)} \sim \text{Binary}(0.5)$  and  $X_{i(2)} \sim N(0,1)$ , and fixed the sample size to be  $n = 1000$  and cluster size  $M = 7$ . Let  $t_{ij} = 1, 2, \dots, 7$  be the observation times. The outcome and missing indicator were generated from SREM:

$$\text{logit } P(Y_{ij}=1 | \mathbf{X}_{ij}, b_i) = \beta_0 + \beta_1 X_{i(1)} + \beta_2 X_{i(2)} + \beta_3 t_{ij} + b_i, \quad (5)$$

$$\text{logit } P(R_{ij}=1 | \mathbf{W}_{ij}, c_i) = \gamma_0 + \gamma_1 X_{i(1)} + \gamma_2 X_{i(2)} + \gamma_3 t_{ij} + c_i, \quad (6)$$

where  $(\beta_0, \beta_1, \beta_2, \beta_3) = (-1, 1, 0.8, -0.2)$ ,  $(\gamma_0, \gamma_1, \gamma_2, \gamma_3) = (3.5, -0.6, -1.2, -0.5)$ . We varied the variance component as  $(\sigma_b, \sigma_c) = (1.5, 1)$  and  $(2.5, 2)$  (moderate and large heterogeneity, respectively), and varied their correlation as  $\rho = -0.2$  and  $-0.6$  (weak and strong nonignorability, respectively). The negative correlation suggested that participants who were more likely to have an event (e.g., ASQ failure) in the outcome were less likely to be observed ( $R = 1$ ), so the missingness mechanism is nonignorable. The missing percentages of the simulation settings are shown in Table S1. The percentage of missingness varied over follow-up, ranging from 13–18% at first to 54–55% at last follow up, mimicking the data example from the Upstate KIDS Study.

In the second set of simulations, we compared the three approaches in terms of their power and type 1 error of testing the regression coefficients of interest,  $\beta_1$ ,  $\beta_2$  and  $\beta_3$ , respectively. Wald test was used for each of the regression coefficients. We fixed  $(\sigma_b, \sigma_c, \rho) = (1.5, 1, -0.6)$ ,  $(\gamma_0, \gamma_1, \gamma_2, \gamma_3) = (3.5, -0.6, -1.2, -0.5)$ , and varied  $\beta_1$  to be 0, 0.2, 0.4 and 0.6 (while fixing  $\beta_2 = 0.8$ ,  $\beta_3 = -0.2$ , same as the first set of simulations),  $\beta_2$  to be 0, 0.1, 0.2 and 0.3 (while fixing  $\beta_1 = 1$ ,  $\beta_3 = -0.2$ ), and  $\beta_3$  to be 0,  $-0.04$ ,  $-0.06$  and  $-0.08$  (while fixing  $\beta_1 = 1$ ,  $\beta_2 = 0.8$ ). For each of the parameters of interest,  $\beta_1$ ,  $\beta_2$  and  $\beta_3$ , when the true parameter value is 0, the type 1 error was computed as the proportion of simulations that falsely rejected the null hypothesis at 0.05 level; when the true parameter is non-zero, the power was computed as the proportion of simulations that correctly rejected the null hypothesis at 0.05 level.

The last set of simulations considered a mis-specified model and compared the performance of SREM and two-step approach. The outcome was generated in the same way as in equation (5) where  $(\beta_0, \beta_1, \beta_2, \beta_3) = (-1, 1, 0.8, -0.2)$  and  $\sigma_b = 1.5$ . The missing mechanism was generated from

$$\text{logit } P(R_{ij}=1 | Y_{ij}, \mathbf{W}_{ij}, c_i) = \gamma_0 + \gamma_1 X_{i(1)} + \gamma_2 X_{i(2)} + \gamma_3 t_{ij} + \delta Y_{ij} + c_i, \quad (7)$$

where  $(\gamma_0, \gamma_1, \gamma_2, \gamma_3) = (3.5, -0.6, -1.2, -0.5)$  and  $\sigma_c = 1$ . The missing mechanism was conditional on  $Y_{ij}$ , known as a “selection model”. We varied the correlation between  $b_j$  and  $c_j$  as  $\rho = -0.2$  and  $-0.6$ , and  $\delta = -0.2, -0.5, -1$ . A larger  $\rho$  or  $\delta$  in magnitude indicates stronger nonignorability in the missing mechanism. To apply SREM and two-step approach, we still assume the models in equations (5) and (6), so the missing mechanism model was mis-specified.

### Analysis of Upstate KIDS Study

In the Upstate KIDS Study, we estimated the odds that a child would fail the ASQ developmental assessment associated with the individual covariates, and compared the results from ordinary GLMM, SREM and two-step approach. The covariates that entered the outcome model included preterm birth, maternal age, race, education, marital status, insurance status, smoking, parity, and follow-up time. For the latter two approaches, the missing mechanism model additionally adjusted for plurality, since mothers with multiple births might be more likely to miss the ASQ assessment than mothers of singletons. All the models accounted for sampling weights by design.

The GLMM makes the ignorable missingness assumption, namely, missing ASQ assessment was only explained by the aforementioned covariates and previously failing an ASQ domain. The SREM and the two-step approach relaxed this assumption by introduced by the shared random effects.

## RESULTS

### Simulation results

We found in the first set of simulation studies that the proposed two-step approach performed well across all simulation settings, in terms of the bias and 95% confidence interval (CI) coverage rates (i.e., the percentage of simulations in which the 95% CI covers the true parameter value) (Table 1). The SREM was the correct model, and it showed good performance with low bias and nominal coverage rates. Under a weak nonignorability setting ( $\rho = -0.2$ ), the GLMM maintained reasonable CI coverage rates, despite a small amount of bias. With a strong nonignorable missing mechanism ( $\rho = -0.6$ ), the GLMM estimators for  $\beta$  were biased with lower than 95% CI coverage rates, especially  $\beta_2$  and  $\beta_3$ . With the two-step approach, the separate estimation of the missing mechanism model lead to valid inference on  $\gamma$  parameters. The bias of estimating  $\beta$  was much smaller than GLMM, and the coverage rates were mostly at the nominal level. Only under a large heterogeneity and strong nonignorability setting, the two-step approach showed slightly lower than nominal coverage. Regarding the computation time, on an Intel i7-4600U CPU, SREM takes 27 minutes to analyze one simulated data set in R, while the two-step approach takes 20 seconds.

In the second set of simulation studies, we found that the two-step approach had slightly inflated type 1 error (6.5–7.9%) and comparable power as the SREM (Table 2). The GLMM, on the other hand, had severely inflated type 1 error for testing  $\beta_2$  and  $\beta_3$ ; in the cases with



the severely inflated type 1 error, we do not report the power of the test since power is only meaningful for tests at the correct level.

When the missingness mechanism was mis-specified (Table 3), the GLMM was the most severely biased across all the settings. For SREM and two-step approach, the bias was small when  $\alpha = -0.2$ , and larger  $\alpha$  leads to more bias. While compared with SREM, the two-step approach produced slightly smaller bias for  $\beta_1$  and  $\beta_2$  (coefficients of subject-level covariates), similar bias in  $\beta_3$  (coefficient of time), and slightly more bias in  $\beta_0$  (intercept), suggesting a better robustness to this type of model mis-specification.

### Upstate KIDS Study

The participants' characteristics ( $n = 4767$ ) are described in the Table S6. Participating mothers were on average 30.5 years of age, mostly non-hispanic whites (80.9%), married (88.2%), with private health insurance (74.8%). The results of the outcome model using the SREM, GLMM, and two-step approach are shown in Table 4. The two-step approach and SREM yielded almost identical results, indicating that the proposed two-step approach closely approximates the SREM. We found that increased odds of ASQ failure was associated with preterm delivery (odds ratio [OR]: 2.90 [2.26, 3.71]), having a previous live birth (OR: 1.61 [1.35, 1.91]), and infants born to black mothers or mothers of "other" race (OR ranges between 1.61–1.72). Having private insurance (OR: 0.80 [0.65, 0.98]), and higher maternal education (OR ranges between 0.46–0.60) were associated with lower risk of ASQ failure. Maternal age, smoking, and marital status were not significantly associated with ASQ failure after adjustment for the other covariates. The SREM estimated the correlation coefficient  $\rho$  to be  $-0.14$  (95% CI:  $[-0.21, -0.06]$ ), suggesting a mild to moderate level of nonignorability in the data. The GLMM gives similar OR estimates to SREM for most of the covariates except time. SREM and two-step approach both suggested that ASQ failure had higher odds to occur at 8, 12, 24, and 30 months than at 4–6 months (OR ranges between 1.23 to 1.91), while the GLMM seemed to slightly underestimate the ORs.

Estimating the missing mechanism model separately or jointly with the outcome model yielded similar results (Table S7). Fewer missing data were observed among mothers who were older, white, married, non-smoking, nulliparous, and who had higher education, private insurance, and a singleton birth than their respective counterparts.

## DISCUSSION

We offer a new and simple approach for addressing a nonignorable missing outcome, which often arises in prospective cohort studies. The statistical inference using this approach was approximately unbiased and powerful, as evaluated in numerous simulation studies and an analysis of the Upstate KIDS Study. This new two-step approach is computationally much faster than SREM, and the implementation is quite straightforward. In extensive simulation studies, we found the two-step approach to be a reliable approximation of the SREM, in terms of the low bias, and similar power to SREM. While SREM is the true model, the ordinary GLMM assuming ignorable missingness often possesses sizable bias and inflated type 1 error. When SREM is misspecified, the bias in the two-step approach is a lot smaller

than GLMM; the two-step approach also had less bias than SREM in estimating the coefficients of subject-level covariates.

Using our newly developed method, the nonignorability of missing data in the Upstate KIDS Study was estimated to be mild to moderate. As a result, the ordinary GLMM assuming ignorable missingness is a reasonable choice for this particular analysis, and predictors of ASQ failure identified by GLMM were generally in line with those by SREM or two-step approach. As epidemiologists are often faced with increasing attrition and missing data over the course of follow up, we recommend that the two-step approximate inference should serve as an easy sensitivity analysis to a GLMM to examine possible violations of ignorable missingness assumption.

We only included random intercepts for the outcome model and the missing mechanism model, which is reasonable in the Upstate KIDS Study. However, the same approach can be applied to models with more random effect terms. The computational advantage of the two-step approach would be even more prominent, since SREM with high-dimensional integration over all the random effects would be extremely slow, if not infeasible.

## CONCLUSION

We show in simulation studies and with a practical example that the two-step approach has comparable performance as the SREM in various scenarios of cluster size, between-subject heterogeneity and strength of nonignorability, and is much faster and more straightforward to implement. We, therefore, recommend this approach for longitudinal studies with missing outcomes as a sensitivity analysis. Future work will examine instances where time-varying covariates are also missing intermittently, as one will also have to incorporate the distribution of the missing covariate into the joint estimation framework through a more complicated model.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The research was supported by the Intramural Research Program of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health. This work utilized the computational resources of the NIH HPC Biowulf cluster. (<http://hpc.nih.gov>) The authors thank Dr. Enrique F. Schisterman for helpful discussions. We thank the editor and two anonymous referees for their insightful suggestions that greatly improved the quality of the paper.

## References

1. Little, RJA., Rubin, DB. Statistical analysis with missing data. 2. Hoboken, N.J: Wiley; 2002. Wiley series in probability and statistics
2. Wu MC, Carroll RJ. Estimation and Comparison of Changes in the Presence of Informative Right Censoring by Modeling the Censoring Process. *Biometrics*. 1988; 44(1):175–188.
3. Pulkstenis P, Ten Have TR, Landis JR. Model for the analysis of binary longitudinal pain data subject to informative dropout through remedication. *Journal of the American Statistical Association*. 1998; 93(442):438–450.

4. Albert PS, Follmann DA. A random effects transition model for longitudinal binary data with informative missingness. *Statistica Neerlandica*. 2003; 57(1):100–111.
5. Gao S. A shared random effect parameter approach for longitudinal dementia data with non-ignorable missing data. *Stat Med*. 2004; 23(2):211–9. [PubMed: 14716723]
6. Albert PS, Follmann DA. Shared-parameter models. *Longitudinal Data Analysis*. 2009:433–452.
7. Tooze JA, Grunwald GK, Jones RH. Analysis of repeated measures data with clumping at zero. *Statistical Methods in Medical Research*. 2002; 11(4):341–355. [PubMed: 12197301]
8. Su L, Tom BDM, Farewell VT. Bias in 2-part mixed models for longitudinal semicontinuous data. *Biostatistics*. 2009; 10(2):374–389. [PubMed: 19136448]
9. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics*. 2000; 1(4):465–80. [PubMed: 12933568]
10. Albert PS. A linear mixed model for predicting a binary event from longitudinal data under random effects misspecification. *Statistics in Medicine*. 2012; 31(2):143–154. [PubMed: 22081439]
11. Rizopoulos D. JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data. *Journal of Statistical Software*. 2010; 35(9):1–33. [PubMed: 21603108]
12. Louis GMB, Hediger ML, Bell EM, Kus CA, Sundaram R, McLain AC, Yeung E, Hills EA, Thoma ME, Druschel CM. Methodology for Establishing a Population-Based Birth Cohort Focusing on Couple Fertility and Children’s Development, the Upstate KIDS Study. *Paediatric and Perinatal Epidemiology*. 2014; 28(3):191–202. [PubMed: 24665916]
13. Squires, J., Bricker, D. *Ages & Stages Questionnaires, Third Edition (ASQ-3)*. Baltimore, MD: Brookes Publishing; 2009.
14. Squires, J., Potter, L., Bricker, D. *The ASQ User’s Guide for the Ages & Stages Questionnaires: A Parent-Completed, Child-Monitoring System*. Baltimore, MD: Paul. H. Brookes Publishing Co; 1999.
15. Yeung EH, Sundaram R, Bell EM, Druschel C, Kus C, Ghassabian A, Bello S, Xie YL, Louis GMB. Examining Infertility Treatment and Early Childhood Development in the Upstate KIDS Study. *Jama Pediatrics*. 2016; 170(3):251–258. [PubMed: 26746435]
16. Nohr EA, Frydenberg M, Henriksen TB, Olsen J. Does low participation in cohort studies induce bias? *Epidemiology*. 2006; 17(4):413–8. [PubMed: 16755269]
17. Winding TN, Andersen JH, Labriola M, Nohr EA. Initial non-participation and loss to follow-up in a Danish youth cohort: implications for relative risk estimates. *J Epidemiol Community Health*. 2014; 68(2):137–44. [PubMed: 24072742]
18. Lin HZ, Liu DP, Zhou XH. A correlated random-effects model for normal longitudinal data with nonignorable missingness. *Statistics in Medicine*. 2010; 29(2):236–247. [PubMed: 19941316]
19. Pinheiro JC, Bates DM. Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model. *Journal of Computational and Graphical Statistics*. 1995; 4(1):12.
20. Carroll, RJ., Ruppert, D., Stefanski, LA., Crainiceanu, CM. *Measurement Error in Nonlinear Models : A Modern Perspective*. 2. 2012.

**Table 1**

Simulation results comparing the proposed two-step approach with shared random effects model, and generalized linear mixed model for different settings of parameter values, with 7 designed observations per subject.<sup>a</sup>

Parameter <sup>b</sup>	True	SREM	GLMM	Two-step
		Bias (coverage %) <sup>c</sup>	Bias (coverage %)	Bias (coverage %)
$\sigma_b = 1.5, \sigma_c = 1, \rho = -0.2$ (weak nonignorability)				
$\beta_0$	-1.0	-0.001 (94.6)	-0.007 (94.8)	-0.012 (94.9)
$\beta_1$	1.0	0.001 (94.0)	-0.010 (93.3)	0.005 (93.8)
$\beta_2$	0.8	0.001 (94.3)	-0.024 (92.6)	0.009 (94.4)
$\beta_3$	-0.2	0.000 (95.1)	-0.007 (93.4)	0.000 (95.2)
$\gamma_0$	3.5	0.004 (94.7)	-	0.004 (94.7)
$\gamma_1$	-0.6	-0.001 (95.3)	-	-0.001 (95.2)
$\gamma_2$	-1.2	0.000 (95.8)	-	0.000 (95.7)
$\gamma_3$	-0.5	0.000 (95.3)	-	0.000 (95.2)
$\sigma_b$	1.5	-0.005 (95.2)	-0.006 (94.8)	-
$\sigma_c$	1.0	-0.002 (96.4)	-	-0.003 (96.1)
$\rho$	-0.2	-0.003 (94.2)	-	-
$\sigma_b = 1.5, \sigma_c = 1, \rho = -0.6$ (strong nonignorability)				
$\beta_0$	-1.0	-0.003 (94.5)	-0.015 (94.6)	-0.031 (93.6)
$\beta_1$	1.0	0.000 (94.1)	-0.037 (93.2)	0.007 (93.3)
$\beta_2$	0.8	0.000 (94.3)	-0.079 (79.4)	0.014 (93.2)
$\beta_3$	-0.2	0.000 (94.6)	-0.023 (83.9)	0.000 (94.6)
$\gamma_0$	3.5	0.007 (95.1)	-	0.007 (95.2)
$\gamma_1$	-0.6	0.000 (95.2)	-	0.000 (95.2)
$\gamma_2$	-1.2	0.000 (94.6)	-	0.000 (94.7)
$\gamma_3$	-0.5	-0.001 (94.9)	-	-0.001 (95.2)
$\sigma_b$	1.5	-0.004 (95.2)	-0.027 (92.5)	-
$\sigma_c$	1.0	-0.001 (94.7)	-	-0.002 (94.7)
$\rho$	-0.6	-0.001 (95.9)	-	-
$\sigma_b = 2.5, \sigma_c = 2, \rho = -0.2$ (weak nonignorability)				
$\beta_0$	-1.0	0.000 (94.6)	-0.059 (94.7)	-0.038 (95.6)
$\beta_1$	1.0	-0.003 (94.4)	-0.028 (94.5)	0.012 (94.4)
$\beta_2$	0.8	0.000 (93.8)	-0.054 (91.0)	0.028 (93.5)
$\beta_3$	-0.2	0.000 (95.0)	-0.010 (92.8)	0.000 (95.0)

Parameter <sup>b</sup>	True	SREM	GLMM	Two-step
		Bias (coverage %) <sup>c</sup>	Bias (coverage %)	Bias (coverage %)
$\gamma_0$	3.5	0.013 (95.4)	-	0.013 (95.6)
$\gamma_1$	-0.6	-0.001 (95.3)	-	-0.001 (95.5)
$\gamma_2$	-1.2	-0.001 (94.3)	-	-0.001 (94.3)
$\gamma_3$	-0.5	-0.002 (95.1)	-	-0.002 (95.0)
$\sigma_b$	2.5	-0.006 (94.0)	-0.006 (93.9)	-
$\sigma_c$	2.0	0.002 (94.8)	-	0.002 (95.0)
$\rho$	-0.2	-0.001 (94.3)	-	-
$\sigma_b = 2.5, \sigma_c = 2, \rho = -0.6$ (strong nonignorability)				
$\beta_0$	-1.0	0.003 (94.7)	-0.142 (86.7)	-0.086 (89.7)
$\beta_1$	1.0	-0.004 (94.4)	-0.088 (91.2)	0.027 (91.7)
$\beta_2$	0.8	-0.002 (93.8)	-0.180 (57.5)	0.058 (89.4)
$\beta_3$	-0.2	-0.001 (95.1)	-0.033 (73.8)	-0.001 (95.2)
$\gamma_0$	3.5	0.007 (94.5)	-	0.007 (94.7)
$\gamma_1$	-0.6	0.004 (94.4)	-	0.004 (94.8)
$\gamma_2$	-1.2	-0.001 (93.7)	-	-0.001 (94.3)
$\gamma_3$	-0.5	-0.001 (94.8)	-	-0.001 (94.9)
$\sigma_b$	2.5	-0.011 (95.4)	-0.107 (84.5)	-
$\sigma_c$	2.0	0.000 (94.5)	-	0.001 (94.3)
$\rho$	-0.6	0.001 (93.8)	-	-

Abbreviations: GLMM, generalized linear mixed model; SREM, shared random effects model.

<sup>a</sup>Designed number of observations is 7.

<sup>b</sup>Parameters are defined in Equations (5)-(6).

<sup>c</sup>Bias is calculated as the difference between the estimated parameters over 1000 simulations and their true values. Coverage percentage is calculated as the percentage of simulations in which the 95% confidence interval covers the true parameter value.

**Table 2**

Simulation results comparing power and Type 1 errors using the proposed two-step approach with shared random effects model and generalized linear mixed model.

Parameter <sup>a</sup>		SREM	GLMM	Two-step
$\beta_1 = 0^b$	Type 1 error	0.065	0.063	0.078
	0.2 Power	0.322	0.232	0.371
	0.4 Power	0.843	0.764	0.876
	0.6 Power	0.992	0.984	0.994
$\beta_2 = 0^c$	Type 1 error	0.063	0.241	0.079
	0.1 Power	0.308		0.413
	0.2 Power	0.836		0.889
	0.3 Power	0.993		0.999
$\beta_3 = 0^d$	Type 1 error	0.064	0.201	0.065
	-0.04 Power	0.517		0.523
	-0.06 Power	0.825		0.828
	-0.08 Power	0.963		0.965

Abbreviations: GLMM, generalized linear mixed model; SREM, shared random effects model.

<sup>a</sup>Parameters are defined in Equation (5).

<sup>b</sup> $\beta_2$  and  $\beta_3$  are fixed as 0.8 and -0.2 while varying  $\beta_1$ .

<sup>c</sup> $\beta_1$  and  $\beta_3$  are fixed as 1 and -0.2 while varying  $\beta_2$ .

<sup>d</sup> $\beta_1$  and  $\beta_2$  are fixed as 1 and 0.8 while varying  $\beta_3$ .

**Table 3**

Simulation results comparing the proposed two-step approach with shared random effects model, and generalized linear mixed model when the missing mechanism model is mis-specified<sup>a</sup>.

Parameter <sup>b</sup>	True	SREM	GLMM	Two-step
		Bias (coverage %) <sup>c</sup>	Bias (coverage %)	Bias (coverage %)
$\delta = -0.2, \rho = -0.6$				
$\beta_0$	-1.0	0.001 (94.5)	-0.017 (94.2)	-0.027 (93.5)
$\beta_1$	1.0	-0.019 (93.4)	-0.060 (91.3)	-0.011 (92.4)
$\beta_2$	0.8	-0.033 (91.0)	-0.121 (60.3)	-0.018 (91.2)
$\beta_3$	-0.2	-0.014 (91.0)	-0.039 (62.9)	-0.014 (91.1)
$\delta = -0.5, \rho = -0.6$				
$\beta_0$	-1.0	0.000 (95.0)	-0.028 (94.2)	-0.028 (93.4)
$\beta_1$	1.0	-0.047 (92.3)	-0.096 (87.5)	-0.037 (90.8)
$\beta_2$	0.8	-0.087 (77.1)	-0.186 (26.3)	-0.070 (80.2)
$\beta_3$	-0.2	-0.037 (68.5)	-0.063 (24.0)	-0.037 (67.9)
$\delta = -1, \rho = -0.6$				
$\beta_0$	-1.0	-0.025 (94.4)	-0.076 (90.2)	-0.054 (90.9)
$\beta_1$	1.0	-0.094 (87.5)	-0.154 (77.2)	-0.082 (86.3)
$\beta_2$	0.8	-0.180 (32.1)	-0.298 (2.6)	-0.161 (38.9)
$\beta_3$	-0.2	-0.077 (13.4)	-0.106 (0.8)	-0.078 (13.1)
$\delta = -0.2, \rho = -0.2$				
$\beta_0$	-1.0	-0.001 (95.0)	-0.010 (95.0)	-0.014 (95.0)
$\beta_1$	1.0	-0.016 (93.3)	-0.029 (93.0)	-0.011 (93.2)
$\beta_2$	0.8	-0.030 (92.3)	-0.061 (85.2)	-0.022 (93.3)
$\beta_3$	-0.2	-0.013 (91.6)	-0.022 (84.3)	-0.013 (91.6)
$\delta = -0.5, \rho = -0.2$				
$\beta_0$	-1.0	-0.008 (94.5)	-0.023 (95.0)	-0.022 (94.9)
$\beta_1$	1.0	-0.044 (92.7)	-0.062 (91.2)	-0.038 (93.2)
$\beta_2$	0.8	-0.083 (78.1)	-0.121 (60.5)	-0.072 (81.6)
$\beta_3$	-0.2	-0.035 (69.7)	-0.045 (52.9)	-0.035 (69.5)
$\delta = -1, \rho = -0.2$				
$\beta_0$	-1.0	-0.042 (93.6)	-0.067 (91.4)	-0.057 (92.3)
$\beta_1$	1.0	-0.091 (87.0)	-0.118 (84.6)	-0.084 (87.9)
$\beta_2$	0.8	-0.175 (37.4)	-0.226 (14.7)	-0.161 (43.2)

Parameter <sup>b</sup>	True	SREM	GLMM	Two-step
		Bias (coverage %) <sup>c</sup>	Bias (coverage %)	Bias (coverage %)
$\beta_3$	-0.2	-0.073 (14.7)	-0.085 (6.4)	-0.073 (14.5)

Abbreviations: GLMM, generalized linear mixed model; SREM, shared random effects model.

<sup>a</sup>The mis-specified model is given in Equations (7).

<sup>b</sup>Parameters are defined in Equation (5).

<sup>c</sup>Bias is calculated as the difference between the estimated parameters over 1000 simulations and their true values. Coverage percentage is calculated as the percentage of simulations in which the 95% confidence interval covers the true parameter value.



**Table 4**

Estimated odds of failure in ages and stages questionnaire assessment associated with individual covariates from Upstate KIDS Study (2008–2010).

	GLMM	SREM <sup>a</sup>	Two-step <sup>b</sup>
	OR (95% CI)	OR (95% CI)	OR (95% CI)
Maternal age (per year)	1.01 (0.99, 1.03)	1.01 (0.99, 1.03)	1.01 (0.99, 1.03)
Time			
Baseline: 4–6 months	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
8 months	1.19 (0.98, 1.43)	1.23 (1.02, 1.49)	1.23 (1.02, 1.49)
12 months	1.41 (1.17, 1.72)	1.50 (1.24, 1.83)	1.50 (1.24, 1.83)
18 months	0.90 (0.72, 1.13)	0.97 (0.77, 1.22)	0.97 (0.77, 1.22)
24 months	1.74 (1.40, 2.15)	1.91 (1.53, 2.38)	1.91 (1.53, 2.38)
30 months	1.28 (1.02, 1.61)	1.40 (1.11, 1.77)	1.40 (1.11, 1.77)
36 months	1.13 (0.89, 1.43)	1.24 (0.97, 1.58)	1.24 (0.97, 1.58)
Maternal race			
White	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Black	1.71 (1.17, 2.49)	1.72 (1.18, 2.49)	1.72 (1.18, 2.49)
Asian	1.43 (0.90, 2.27)	1.43 (0.90, 2.25)	1.43 (0.90, 2.25)
Hispanic	1.12 (0.81, 1.54)	1.12 (0.82, 1.54)	1.12 (0.82, 1.54)
Other	1.59 (1.16, 2.17)	1.61 (1.18, 2.19)	1.61 (1.18, 2.19)
Maternal education			
Less than high school	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
High school	0.78 (0.55, 1.12)	0.79 (0.55, 1.12)	0.79 (0.55, 1.12)
Some college	0.60 (0.42, 0.84)	0.60 (0.43, 0.85)	0.60 (0.43, 0.85)
College graduate	0.48 (0.32, 0.71)	0.47 (0.32, 0.70)	0.47 (0.32, 0.70)
Advanced degree	0.46 (0.31, 0.69)	0.46 (0.31, 0.68)	0.46 (0.31, 0.68)
Preterm delivery			
Yes	2.93 (2.28, 3.77)	2.90 (2.26, 3.71)	2.90 (2.26, 3.71)
No	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Health insurance			
Private insurance	0.80 (0.65, 0.99)	0.80 (0.65, 0.98)	0.80 (0.65, 0.98)
No private insurance	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Marital status			
Married/Living as married	1.02 (0.79, 1.32)	1.01 (0.78, 1.30)	1.01 (0.78, 1.30)
Not married	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Previous live birth			
Yes	1.61 (1.35, 1.92)	1.61 (1.35, 1.91)	1.61 (1.35, 1.91)
No	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Smoking			
Yes	1.09 (0.87, 1.37)	1.09 (0.87, 1.37)	1.09 (0.87, 1.37)
No	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)

	GLMM	SREM <sup>a</sup>	Two-step <sup>b</sup>
	OR (95% CI)	OR (95% CI)	OR (95% CI)
$\sigma_b$ (random intercept)	1.52 (1.40, 1.65)	1.50 (1.39, 1.63)	1.49 (1.38, 1.62)

Abbreviations: GLMM, generalized linear mixed model; SREM, shared random effects model; OR, odds ratio; CI, confidence interval.

<sup>a</sup>SREM estimated the correlation between the random effects  $\rho$  to be  $-0.14$  (95% CI:  $[-0.21, -0.06]$ ).

<sup>b</sup>The two-step approach estimated the coefficient of the random effect  $\hat{e}_j$  to be  $-0.14$  (95% CI:  $[-0.24, -0.04]$ ).