Published in final edited form as:

Stat Biosci. 2012 November 1; 4(2): 213–234. doi:10.1007/s12561-011-9053-2.

A Semiparametric Approach to Joint Modeling of A Primary Binary Outcome and Longitudinal Data Measured at Discrete Informative Times

Song Yan,

Department of Statistics, North Carolina State University, Raleigh, NC 27695, U.S.A

Daowen Zhang,

Department of Statistics, North Carolina State University, Raleigh, NC 27695, U.S.A

Wenbin Lu,

Department of Statistics, North Carolina State University, Raleigh, NC 27695, U.S.A

James A. Grifo, and

NYU Fertility Center, NYU Langone Medical Center, New York, NY 10016, U.S.A

Mengling Liu

Division of Biostatistics, School of Medicine, New York University, New York, NY 10016, U.S.A

Song Yan: syan@ncsu.edu

Abstract

In a study conducted at the New York University Fertility Center, one of the scientific objectives is to investigate the relationship between the final pregnancy outcomes of participants receiving an in vitro fertilization (IVF) treatment and their β -human chorionic gonadotrophin (β -hCG) profiles. A common joint modeling approach to this objective is to use subject-specific normal random effects in a linear mixed model for longitudinal β -hCG data as predictors in a model (e.g., logistic model) for the final pregnancy outcome. Empirical data exploration indicates that the observation times for longitudinal β -hCG data may be informative and the distribution of random effects for longitudinal β -hCG data may not be normally distributed. We propose to introduce a third model in the joint model for the informative β -hCG observation times, and relax the normality distributional assumption of random effects using the semi-nonparametric (SNP) approach of Gallant and Nychka (1987) [8]. An EM algorithm is developed for parameter estimation. Extensive simulation designed to evaluate the proposed method indicates that ignoring either informative observation times or distributional assumption of the random effects would lead to invalid and/or inefficient inference. Applying our new approach to the data reveals some interesting findings the traditional approach failed to discover.

Keywords

EM algorithm; Informative observation times; IVF; Joint model; Longitudinal data; Maximum likelihood estimation; SNP density

1 Introduction

For patients undergoing in vitro fertilization and embryo transfer (IVF-ET), the early determination of pregnancy prognosis is of critical importance [9]. From a medical standpoint, there are increased risks of adverse outcomes in IVF pregnancies compared with natural conceptions, including ectopic pregnancies and spontaneous abortions [9]. The incidence of ectopic pregnancy after IVF is nearly 2 to 5 times higher than that in natural

pregnancies [17]. In naturally conceived cycles, among a variety of markers, human chorionic gonadotropin β subunit (β -hCG) has been found highly predictive of normal pregnancy [5]. In fact, because implantation can be timed more accurately in IVF pregnancies, β -hCG curves for IVF pregnancies should be even more accurate [5]. In the fertility literature, various cutoff levels of initial β -hCG measured some specific number of days after ET were proposed to predict viability (e.g., Lambers et al.(2006) [12]). However, a significant number of normal pregnancies often may have β -hCG levels below the established cut-offs [16]. Others suggested that the rising rate of early β -hCG after ET may also have a very strong positive correlation with the pregnancy outcome [4, 16]. Shamonki et al. (2009) [16] depicted logarithmic curves of initial β -hCG level and the rise of early β -hCG and the live delivery outcomes with IVF, and suggested that using the β -hCG curves could give both the clinician and the patient a more accurate assessment of the pregnancy since both the initial β -hCG value and the rising rate of β -hCG curve seem highly correlated with the final pregnancy outcome. Furthermore, Shamonki et al. (2009) [16] demonstrated there is a strong correlation among age, early β -hCG values and pregnancy outcome.

To understand the relationship between final pregnancy outcomes and early antenatal hormonal characteristics among IVF patients, a study conducted at the New York University Fertility Center collected data from patients who underwent IVF treatment between 2001 and 2003 (This retrospective study was approved by the Institutional Review Board at the NYU School of Medicine). For each patient, besides the baseline covariates and the primary binary pregnancy outcome (viable or nonviable) at the end of the study, β -hCG values are repeatedly measured at six potential follow-up time intervals between day 7 and day 23 after ET. Fig. 1 shows the observed log β -hCG profiles of 30 randomly selected patients with viable pregnancies and another 30 randomly selected patients with non-viable pregnancies. From Fig. 1, we can observe that the log β -hCG profiles of viable pregnancies show a clearly increasing trend, while those of nonviable pregnancies have a very diverse pattern, a mixture of increasing and decreasing trends, which suggests that patients' β -hCG profiles may be correlated with their final pregnancy outcomes. In this paper, we take a joint modeling approach to studying the relationship between the pregnancy outcome and repeated measures of early β -hCG values after ET. In the statistical literature, there have been abundant works on joint models for a primary outcome (discrete or continuous) and longitudinal covariates. A popular approach is the latent variable approach that uses subject-specific random effects characterizing longitudinal covariate profiles as covariates in the model for the primary outcome ([11,22,10,21,19,13,20], among many others).

One particular challenge in the analysis of NYU-IVF data is that the measurement mechanism of β -hCG profiles might also be informative for the primary pregnancy outcome. Table 1 shows the proportion of women having β -hCG measurements at each of the six intervals in the viable and non-viable pregnancy group separately. From Table 1 and Fig. 1, we observe that women with viable pregnancies tend to have less β -hCG measurements at intermittent time intervals 2 (days 10–11), 3 (days 12–13), 5 (days 17–19) and 6 (days 20–23) compared to those with non-viable pregnancies. To take into account the possible informative pattern of β -hCG measurements, we introduce extra subject-specific logistic submodels, one at each potential observation time (the middle point of a time interval) to model the probability that a β -hCG value will be measured at that time point for a study subject. The correlation among the β -hCG profile, binary pregnancy outcome and discrete observation times are then characterized via latent subject-specific random variables.

In addition, most joint modeling strategies usually assume a normal or other parametric distributions for unobserved subject-specific random effects shared by submodels. This assumption can be restrictive and easily violated in many applications. To informally explore the distributions of subject-specific random effects in the linear mixed model for the

β -hCG measurements from NYU-IVF data, we fit individual β -hCG profiles by simple linear regression over observation times (here we define the first observation time as time 0 and the unit is week). Fig. 2 presents the Q-Q plots of subject-specific intercept and slope estimates from individual least squares fits. Both plots suggest some discrepancy from the normal distribution, while the Q-Q plot of subject-specific slopes shows a larger deviation from the straight line. To take into account this departure from normality, we adopt the seminonparametric (SNP) approach of Gallant and Nychka (1987) [8] to model the distributions of subject-specific random effects. Similar approaches were also used by Zhang and Davidian (2001) [24] and Song, Davidian, and Tsiatis (2002) [18] in other contexts. We develop a maximum likelihood estimation method for the parameters in the joint model via an expectation maximization (EM) algorithm [7], and use an EM-aided numerical differentiation method to compute the variance-covariance matrix of the estimators of interest.

In this paper we propose a joint model to study the association between a primary binary outcome and longitudinal covariates that are measured at informative discrete occasions. In this joint model, the random effects can follow a flexible distribution as well. To the best of our knowledge, this has not been studied in the literature. The rest of this article is organized as follows. Section 2 introduces notation and describes the proposed joint model. Section 3 presents the inference procedure using an EM algorithm. Section 4 applies our method to NYU-IVF data, in which we also investigate the consequences of ignoring the violation of normality assumption and the informative observation times. Section 5 provides simulation studies to further justify our method, followed by some discussion in Section 6.

2 Joint Model

2.1 Data Notation

For NYU-IVF data, let Y_i denote the binary pregnancy outcome of subject i , $i = 1, \dots, n$, obtained at the end of the study, taking the value of 1 if a viable pregnancy is achieved and 0 otherwise. Let X_i denote the baseline covariates, including age and BMI (appropriately standardized for computational stability). Denote the potential measurement occasions of β -hCG profiles by $\mathbf{t} = (t_1, \dots, t_m)'$, where $t_1 = 0$ is the first observation time point, m is the maximum number of measurements and $m = 6$ for NYU-IVF data. Moreover, denote the complete log β -hCG profile by $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{im})'$. Define $\mathbf{R}_i = (R_{i1}, \dots, R_{im})'$, where R_{ij} indicates whether a measurement of the β -hCG value occurs at the j th time point: $R_{ij} = 1$ if a measurement is taken and $R_{ij} = 0$ otherwise. Therefore, Z_{ij} is observed only if $R_{ij} = 1$. The observed data consist of $\{Y_i, X_i, R_{ij}, Z_{ij}R_{ij}; i = 1, \dots, n; j = 1, \dots, m\}$.

2.2 Model Specifications

Fig. 1 indicates that there might be a quadratic trend over time in the longitudinal log β -hCG profile. To allow for the possible quadratic trend, we consider the following linear mixed model for the longitudinal log β -hCG profile Z_{ij} in the ideal situation of no missingness,

$$Z_{ij} = \gamma_0 + \gamma_1 t_j + \gamma_2 t_j^2 + \gamma_3' X_i + u_{i0} + u_{i1} t_j + \varepsilon_{ij}, \quad j = 1, \dots, m, \quad (1)$$

where $\mathbf{u}_i \equiv (u_{i0}, u_{i1})'$ are mean-zero random effects, measuring deviation in the intercept and the changing rate of subject-specific log β -hCG profile at $t_1 = 0$ from the population profile. It is assumed that the error term vector $\boldsymbol{\varepsilon}_i \equiv (\varepsilon_{i1}, \dots, \varepsilon_{im})'$ is independent of $\mathbf{u}_i$ and follows a multivariate normal distribution $\mathcal{N}(0, \Sigma)$, where the variance-covariance matrix Σ has the

form $\sum \sigma_\varepsilon^2 \Gamma(\rho)$ with σ_ε^2 being a positive scalar and $\Gamma(\rho)$ a $m \times m$ Markov structured correlation matrix indexed by ρ . Define $\gamma = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)'$.

Note that we could add $u_{i2}t_j^2$ in (1) to model subject-specific random effect u_{i2} of t_j^2 . A preliminary analysis of NYU-IFV data indicates that u_{i2} may not be necessary (the P-value from the correct likelihood ratio test assuming normal distribution for $\mathbf{u}_i$ is 0.29). Furthermore, adding u_{i2} will make the computation challenging for the SNP approach for $\mathbf{u}_i$. Therefore, we restrict two random effects in model (1).

For the binary pregnancy outcome Y_i , we assume a logistic model,

$$\text{logit}\{P(Y_i=1|X_i, \mathbf{u}_i)\} = \beta_0 + \beta_1' X_i + \alpha_1 u_{i0} + \alpha_2 u_{i1}, \quad (2)$$

where α_1 and α_2 are effects of u_{i0} and u_{i1} on the binary outcome Y_i respectively. Define $\beta = (\beta_0, \beta_1)'$. Finally, to account for the informative measurement pattern of β -hCG profiles as discussed before, we introduce extra logistic models for the measurement indicators R_{ij} 's

$$\text{logit}\{P(R_{ij}=1|X_i, \mathbf{u}_i)\} = \eta_{0j} + \eta_{1j}' X_i + \alpha_{3j} u_{i0} + \alpha_{4j} u_{i1}, \quad j=1, \dots, m, \quad (3)$$

where α_{3j} and α_{4j} are the time-specific effects of u_{i0} and u_{i1} on the measurement indicator.

Define $\eta_j = (\eta_{0j}, \eta_{1j})'$ and $\lambda_j = (\eta_j, \alpha_{3j}, \alpha_{4j})'$, $j=1, \dots, m$.

Note that the three submodels (1), (2) and (3) are correlated through shared subject-specific random effects $\mathbf{u}_i$ and the magnitude of the association is controlled by the parameters α_1 , α_2 , α_{3j} and α_{4j} . Besides $\mathbf{u}_i$ extra randomness may exist. For example, the Markov structured variance-covariance matrix of the error terms in submodel (1) describes the additional association in repeated β -hCG measurements that can not be explained by the random effects $\mathbf{u}_i$. Following the convention in joint modeling, we assume that $\mathbf{Z}_i$, $\mathbf{R}_i$ and Y_i are mutually independent given X_i and $\mathbf{u}_i$.

To flexibly model the distribution of the random effects $\mathbf{u}_i$, we adopt the SNP approach proposed by Gallant and Nychka (1987) [8]. This approach assumes that the random effects $\mathbf{u}_i$ belongs to a class of densities that are sufficiently smooth so that they do not exhibit unusual behavior such as kinks, jumps, or oscillation. However the densities are very flexible and are allowed to be skewed, multi-modal, and fat- or thin-tailed (relative to the normal distribution); furthermore, the class includes the normal distribution as a special case [24]. Specifically, SNP approach re-formulates the random effects $\mathbf{u}_i$ as

$$\mathbf{u}_i = \mu + D\mathbf{b}_i \quad (4)$$

where $\mu = (\mu_0, \mu_1)'$, D is a 2×2 lower triangular matrix with $\xi \equiv \text{vech}(D) = (d_{00}, d_{10}, d_{11})'$ and the “transformed” random effects $\mathbf{b}_i = (b_{i1}, b_{i2})'$ have a smooth density function

$$h_K(\mathbf{b}) = P_K^2(\mathbf{b})\phi(\mathbf{b}) = \left(\sum_{0 \leq i_1 + i_2 \leq K} a_{i_1 i_2} b_1^{i_1} b_2^{i_2} \right)^2 \phi(\mathbf{b}), \quad (5)$$

with $\mathbf{b} = (b_1, b_2)'$, where $\phi(\cdot)$ is the standard bivariate normal density, K is a non-negative integer, $P_K(\mathbf{b})$ is a bivariate polynomial function of order K with coefficients $a_{i_1 i_2}$, $i_1, i_2 = 0, 1, \dots, K$, and $0 \leq i_1 + i_2 \leq K$. To ensure $h_K(\mathbf{b})$ to be a proper bivariate density function, the coefficients $\{a_{ij}\}$ of $P_K(\mathbf{b})$ must be chosen to satisfy the equality $\int h_K(\mathbf{b}) d\mathbf{b} = 1$. Zhang and Davidian (2001)[24] showed that the above condition is equivalent to $E\{P_K^2(\mathbf{U})\} = \mathbf{a}' A \mathbf{a} = 1$, where $\mathbf{U} = (U_1, U_2)'$ is a standard bivariate normal random variable, $\mathbf{a}$ is a $d \times 1$ vector containing all the coefficients $a_{i_1 i_2}$'s, and A is the corresponding matrix with the elements of $E(U_1^{i_1+j_1} U_2^{i_2+j_2})$ with $0 \leq i_1 + i_2 \leq K$ and $0 \leq j_1 + j_2 \leq K$. In Appendix B we give more details on the computation of the matrix A . Note that $P_0(\cdot) \equiv 1$, which corresponds to the standard bivariate normal density of $\mathbf{b}$. As demonstrated in Zhang and Davidian (2001) [24], in many cases the SNP approach with a K as small as one or two can adequately approximate complicated shapes, including multimodality and skewness. Since A is a symmetric positive definite matrix, hence it has a decomposition $A = BB'$ with some square matrix B . Write $\mathbf{c} = B'\mathbf{a}$, then we require $\mathbf{a}' A \mathbf{a} = \mathbf{c}' \mathbf{c} = 1$. As in [24], we consider a polar coordinate transformation of $\mathbf{c} = (c_1, \dots, c_d)'$, i.e. $c_1 = \sin(\varphi_1)$, $c_2 = \cos(\varphi_1)\sin(\varphi_2)$, $\dots$, $c_{d-1} = \cos(\varphi_1)\cos(\varphi_2)\cos(\varphi_{d-2})\sin(\varphi_{d-1})$, and $c_d = \cos(\varphi_1)\cos(\varphi_2)\cos(\varphi_{d-2})\cos(\varphi_{d-1})$, where $-\pi/2 < \varphi_l < \pi/2$ for $l = 1, 2, \dots, d-1$. Then $\mathbf{c}' \mathbf{c} = 1$ and thus $\int h_K(\mathbf{b}) d\mathbf{b} = 1$ is automatically satisfied by parameterizing $h_K(\mathbf{b})$ in terms of $\varphi = (\varphi_1, \dots, \varphi_{d-1})'$.

For illustration, let us consider a univariate SNP density with $K = 2$. Using the results in [3], we have $a_0 = \cos(\varphi_1) - \sin(\varphi_1)\sin(\varphi_2)/\sqrt{2}$, $a_1 = \sin(\varphi_1)\cos(\varphi_2)$ and $a_2 = \sin(\varphi_1)\sin(\varphi_2)/\sqrt{2}$. Fig. 3 plots the resulting SNP densities for some selected values of (φ_1, φ_2) . It is seen clearly that the SNP density can be used to approximate many different densities with distinct features (such as multi-modal, skewed, etc).

3 Maximum Likelihood Estimation and Inference

For subject i , let $m_i = \sum_{j=1}^m R_{ij}$, and $\mathbf{t}_i^* = (t_{i1}, \dots, t_{im_i})'$ be the time points where β -hCG values are measured, and denote by $\mathbf{Z}_i^* = (Z_{i1}, \dots, Z_{im_i})'$ the observed log β -hCG values. The observed data for subject i is denoted as $\mathbf{O}_i$ and write $\mathbf{O} = (\mathbf{O}_i; i = 1, \dots, n)$. Let

$\Theta = \{\gamma, \sigma_\varepsilon^2, \rho, \beta, \alpha_1, \alpha_2, \lambda_1, \dots, \lambda_m, \mu, \xi, \varphi\}$ denote the unknown parameters in the proposed joint model. Then the joint likelihood contributed by the observed data is given by

$$L(\Theta) = \prod_{i=1}^n \left[\int f_L(\mathbf{Z}_i^* | \mathbf{u}_i) f_P(Y_i | \mathbf{u}_i) f_R(\mathbf{R}_i | \mathbf{u}_i) f_{\mathbf{u}}(\mathbf{u}_i) d\mathbf{u}_i \right], \quad (6)$$

where $f_L(\mathbf{Z}_i^* | \mathbf{u}_i)$, $f_P(Y_i | \mathbf{u}_i)$ and $f_R(\mathbf{R}_i | \mathbf{u}_i)$ are conditional density functions (given random effects $\mathbf{u}_i$) of the observed longitudinal log β -hCG values, primary binary outcome and measurement indicators for subject i , respectively, and $f_{\mathbf{u}}(\cdot)$ is the density function of the random effects $\mathbf{u}_i$ derived from the SNP representation (4). Here for simplicity, we omit the dependence of these density functions on the observed covariates. Specifically, we have SNP representation (4). Here for simplicity, we omit the dependence of these density functions on the observed covariates. Specifically, we have

$$\begin{cases} f_L(\mathbf{Z}_i^*|\mathbf{u}_i; \gamma, \Sigma_i) & = (2\pi)^{-m_i/2} |\Sigma_i|^{-1/2} \exp\{-(\mathbf{Z}_i^* - \mu_i^*)' \Sigma_i^{-1} (\mathbf{Z}_i^* - \mu_i^*)\}, \\ f_p(Y_i|\mathbf{u}_i; \beta, \alpha_1, \alpha_2) & = \frac{\exp\{Y_i(\beta_0 + \beta_1' X_i + \alpha_1 u_{i0} + \alpha_2 u_{i1})\}}{1 + \exp\{\beta_0 + \beta_1' X_i + \alpha_1 u_{i0} + \alpha_2 u_{i1}\}}, \\ f_R(\mathbf{R}_i|\mathbf{u}_i; \lambda_1, \dots, \lambda_m) & = \prod_{j=1}^m \frac{\exp\{R_{ij}(\eta_{0j} + \eta_{1j}' X_i + \alpha_{3j} u_{i0} + \alpha_{4j} u_{i1})\}}{1 + \exp\{\eta_{0j} + \eta_{1j}' X_i + \alpha_{3j} u_{i0} + \alpha_{4j} u_{i1}\}}, \\ f_{\mathbf{u}}(\mathbf{u}_i; \mu, \xi, \varphi) & = P_K^2 \{D^{-1}(\mathbf{u}_i - \mu)\} \phi\{D^{-1}(\mathbf{u}_i - \mu)\} |D|^{-1}, \end{cases} \quad (7)$$

where $\mu_i^* = \gamma_0 + \gamma_1 \mathbf{t}_i^* + \gamma_2 \mathbf{t}_i^{*2} + \gamma_3' X_i + u_{i0} + u_{i1} \mathbf{t}_i^* = E(\mathbf{Z}_i^* | \mathbf{u}_i, X_i)$, $\mathbf{t}_i^{*2}$ is the vector formed by squaring each element of $\mathbf{t}_i^*$, $\Sigma_i = \sigma_\varepsilon^2 \Gamma_i(\rho)$ is the associated Markov structured variance-covariance matrix for the observed $\mathbf{Z}_i^*$. Note that it is different from Σ for $\mathbf{e}_i$ in submodel (1).

The direct maximization of the likelihood function (6) is very challenging, mainly due to two reasons: (a) the integration in (6) does not have an analytical form; (b) the number of model parameters is large and may cause numerical instability. This motivates us to develop an EM algorithm to maximize the likelihood function (for a given K). A striking property of the EM algorithm is that the observed data likelihood will always increase during the parameter update process. In addition, the parameter update can be separately carried over for different but smaller subsets of parameters and some parameter updates may even have closed forms.

The details of E-step and M-step of the EM algorithm can be found in Appendix A and B. The proposed EM algorithm requires specifying initial values of the parameters. Here we use the naive regression calibration (RC) method to produce the initial values for Θ . To be more specific, we first fit the linear mixed submodel for longitudinal covariates without considering the informative observation times, to obtain the initial estimators

$(\hat{\gamma}^{(0)}, \hat{\sigma}_\varepsilon^{2(0)}, \hat{\rho}^{(0)})$ and the best linear unbiased predictors (BLUP) $\hat{u}_{i0}^{(0)}$ and $\hat{u}_{i1}^{(0)}$, $i = 1, \dots, n$. This can be easily done using standard software packages, e.g., PROC MIXED in SAS or lme in R. Then we fit the logistic models for the binary outcome and measurement

indicators with X_i and the BLUP estimates $\hat{u}_{i0}^{(0)}$ and $\hat{u}_{i1}^{(0)}$ as covariates to obtain the initial estimators of parameters in (2) and (3), which can also be implemented directly by standard software packages. To obtain initial values for the parameters (μ, ξ, φ) in the SNP distribution, similar to Song, Davidian, and Tsiatis (2002) [18], we calculate the pseudo-maximum-likelihood estimates of (μ, ξ, φ) by treating the BLUP estimators $\hat{\mathbf{u}}_i^{(0)} = (\hat{u}_{i0}^{(0)}, \hat{u}_{i1}^{(0)})$ as the true underlying random effects, and denote the resulting estimators by $(\hat{\mu}^{(0)}, \hat{\xi}^{(0)}, \hat{\varphi}^{(0)})$. For the convergence criterion of the proposed EM algorithm, we consider $\max |\hat{\Theta}^{(k+1)} - \hat{\Theta}^{(k)}| < \delta$ with $\delta = 0.0001$ in our numerical studies. We use the Gaussian quadrature method with 25 quadrature knots to approximate the two dimensional integrations used in the E-step.

In the SNP representation, the parameter K is a tuning parameter, controlling the flexibility of the random effects distribution, which is needed to be chosen based on the data. Here we select K based on various information criteria that all take the form of $-l(\Theta)/N + C(N)p_{net}/N$, where $l(\Theta)$ is the log likelihood, $N = \sum_{i=1}^n m_i$ and p_{net} is the number of free parameters in the joint model. For example, $C(N) = 1$ is for the Akaike Information Criterion (AIC), $C(N) = 0.5 \log N$ for the Schwarz Information Criterion (BIC), and $C(N) = \log \log(N)$ for the Hannan-Quinn criterion (HQ). As discussed by Zhang and Davidian (2001)[24], HQ criterion is often preferred in the selection of K . We will evaluate the empirical performance of these information criteria via simulations in Section 5.

For parametric models, Louis's formula [14] can be used to compute the variance of the estimates obtained by the EM algorithm. However, for our proposed SNP likelihood approach, the number of parameters in the joint model can be large and our interest mainly focuses on the regression parameters for the effects of the baseline predictors and random effects. The direct calculation of the variance for the EM estimates based on Louis's formula may be unstable. Here we calculate the variance for the estimates of the regression parameters by inverting the observed information matrix based on the corresponding profile log likelihood function. In general, the observed information matrix does not have a closed analytical form. We compute it using an EM-aided numerical differentiation method, which was studied by Meilijson (1989) [15] for parametric models and then extended to the proportional hazards model with missing covariates by Chen and Little (1999) [2]. Zeng and Cai (2005) [23] used a similar approach to variance estimation based on a profile likelihood function. The details of the EM-aided numerical differentiation method are given in Appendix C.

4 Application to NYU-IVF Data

The NYU-IVF data consist of 540 pregnancies obtained after IVF treatment at the New York University Fertility Center from 2001 to 2003. In this study, viable pregnancies are defined as pregnancies reaching the second trimester, including singleton, twins, higher order multiples and stillbirths; non-viable pregnancies include biochemical pregnancies, ectopic pregnancies and first trimester abortions [1]. The average age of participants was 35.27 years ($SD=4.40$ years) and the average body mass index (BMI) was 23.71 ($SD=5.05$). So for computational stability and ease of interpretation, age is centered at 35 and divided by 10, BMI is centered at 23 and divided by 10. The possible observation times of β -hCG are rounded to the median days of each six potential time intervals from day 7 to day 23 after ET. Interval 1 is days 7–9, interval 2 is days 10–11, interval 3 is days 12–13, interval 4 is days 14–16, interval 5 is days 17–19, and interval 6 is days 20–23. We define the first observation time as time 0 and transform the day unit to week unit. The total number of β -hCG observations is $N=1325$. We fit the NYU-IVF data by our proposed joint model using the SNP likelihood-based approach with $K=0, 1$ and 2 ($K=0$ means normality assumption for $\mathbf{u}_j$). For comparison, we also analyzed the NYU-IVF data by a reduced joint model in which the informative measurement mechanism of the β -hCG profiles is ignored. When choosing the K in the SNP representation, we use the AIC, BIC and HQ criteria. The log likelihood for full model for $K=0, 1, 2$ are $-1732.86, -1641.57$ and -1635.49 ; the AIC values are 1.3380, 1.2706 and 1.2683; the HQ values are 1.3674, 1.3014 and 1.3013; the BIC values are 1.4163, 1.3528 and 1.3564. The log likelihood, AIC, HQ and BIC values for reduced models are omitted here. Two of three criteria choose $K=2$ for the full joint models and $K=1$ for the reduced joint models, which again supports our previous discovery that the subject-specific random effects are not normally distributed. The parameter estimates with their estimated standard errors computed based on the proposed EM-aided numerical differentiation method for the full and reduced joint models are summarized in Tables 2 and 3, where Table 2 summarizes the results for the linear mixed model and the logistic model for the primary binary outcome while Table 3 for the logistic models for the measurement indicators.

Columns (A) and (B) in Table 2 are the results for our proposed joint model with $K=0$ and 2 , respectively. We make the following observations: (1) age is significantly negatively associated with the final pregnancy outcome, while BMI is not; (2) BMI is significantly negatively associated with the β -hCG profile, while age is not; (3) greater baseline value and stronger increasing trend of the β -hCG profile at baseline are associated with higher chance of viable pregnancy (e.g., $\hat{\alpha}_1 = 1.610$ and $\hat{\alpha}_2 = 1.527$ when $K=2$, with the p -values of both < 0.05); (4) the point estimates for $K=0$ and 2 are very comparable except for estimates of

α_2 , while the estimates for $K = 2$ show certain efficiency gain for some parameters. We will further investigate the efficiency of estimates for $K = 0$ and 2 by simulations in Section 5.

Columns (C) and (D) in Table 2 are the results for the reduced model ignoring the informative measurement times with $K = 0$ and 1, respectively. Comparing column (D) to column (B), we observe that there is some discrepancy in the estimated population curves of the log β -hCG profiles. This is further supported by the results from the simulation in Section 5, which demonstrates that ignoring the informative measurement pattern of longitudinal data may cause an estimation bias of the mean slope parameter in the linear mixed model. More importantly, we find that the estimate of α_1 (the effect of the random intercept on the final pregnancy outcome) becomes in-significant in the reduced model, due to a small point estimate and an increased standard error estimate.

From the results in Table 3 we observe that the estimates for the random intercepts (α_{3j} 's) are negatively significant at intermediate times t_2, t_3, t_5 and t_6 , while those for the random slopes (α_{4j} 's) are not significant. This suggests that doctors may require more intermediate measurements of β -hCG values for women with lower β -hCG values at t_1 . Furthermore, age and BMI do not have significant effects on most of the measurement indicators.

The estimated joint and marginal density functions of the random effects $\mathbf{u}_i$ for the full model with $K = 2$ are given in Fig. 4: (a) plots the estimated bivariate density function, which shows a bimodal distribution; (b) gives the contour plot corresponding to the bivariate density in (a) together with the posterior estimates of the $\mathbf{u}_i$'s, which also shows clearly two clusters; (c) and (d) are estimated marginal densities of u_{i0} and u_{i1} respectively, together with their histograms. It is seen that the estimated density of u_{i1} 's also shows a clear bimodal, implying the deviation from the normal distribution. The bimodal distribution of the subject-specific random effects indicates that the underlying study population may be a mixture of two populations, separated most by the latent variable u_{i1} . According to the results for the primary pregnancy outcome, we infer that one sub-population has a higher likelihood to have a viable pregnancy, while the other has a lower likelihood.

5 Simulation Studies

Simulation was conducted to evaluate the performance of the proposed method under practical settings. We used the same linear mixed model for longitudinal covariates (without a quadratic term of t_j for simplicity) and logistic model for the primary binary outcome. For the informative measurement indicators, we consider the following simpler logistic models:

$$\text{logit}\{P(R_{ij}=1|X_i, \mathbf{u}_i)\} = \eta_{0j} + \eta_{1j}'X_i + \alpha_3 u_{i0} + \alpha_4 u_{i1}, \quad j=1, \dots, m, \quad (8)$$

where the effects of covariates and subject-specific random variables on the measurement indicators are assumed to be the same over all time points. We consider two baseline covariates: a binary covariate X_{i1} sampled from the Bernoulli distribution with a success probability of 0.5 and a continuous covariate X_{i2} sampled from the standard normal distribution. Longitudinal covariates $\mathbf{Z}_i$ were generated at six possible observation times $(t_1, \dots, t_6) = (1, 2, 3, 4, 5, 6)$ in weeks. The subject-specific random effects $\mathbf{u}_i$'s were generated from a mixture of two bivariate normal distributions: $F(\mathbf{u}_i) = 0.7\Phi(\mathbf{u}_i; \mu_1, V) + 0.3\Phi(\mathbf{u}_i; \mu_2, V)$ with $\mu_1 = (1.56, 0)'$, $\mu_2 = (-3.64, 0)'$ and the elements of the covariance matrix V given by $v_{00} = 0.81$, $v_{01} = v_{10} = -0.0456$ and $v_{11} = 1.96$. The true values of the regression parameters are shown in Table 4. We run 100 simulation with sample size of $n = 300$.

We consider the SNP representation with $K = 0, 1$ and 2, and choose the optimal K values by the HQ criterion (the AIC and BIC criteria give similar results). Simulation results are

summarized in Table 4. Here we report the results for $K = 0$ corresponding to the bivariate normal distribution of the random effects, and the optimal estimates chosen by HQ information criterion. For comparison, we also report the estimation results based on the reduced model ignoring the informative measurement process. For our proposed method (part (a) in the table), the estimates of all the parameters are nearly unbiased in all cases, the SE's are close to the SD's, and all the CP's are close to the nominal level. This implies that the HQ information criterion performs well in selecting the tuning parameter K . Similar observations were also found in the literature [18]. However, the estimators assuming the incorrect normality assumption for the random effects ($K = 0$) may be less efficient than those estimated by SNP with the K value chosen by the HQ criterion (for example, the SD of $\hat{\gamma}_{2,1}$ is 0.156 for $K = 0$, but 0.079 for HQ; the SD for $\hat{\gamma}_{2,2}$ is 0.327 for $K = 0$, but 0.160 for HQ). In terms of the selection of the tuning parameter K by HQ, the values 0, 1 and 2 were chosen 2%, 84% and 14% out of 100 runs, indicating the ability of the HQ criterion to detect the departure of the random effects distribution from normality. These selection proportions by AIC and BIC are 1%, 78%, 21% and 2%, 88%, 10% respectively.

The results for the reduced model are shown in part (b) of Table 4. We can see that the estimates of the mean slope parameter (γ_1) in the linear mixed model for longitudinal covariates show bigger biases compared to those of the full model and the associated empirical coverage probability (CP) of the 95% Wald-type confidence interval is significantly below the nominal level (87%). We also see that the parameter estimates in the primary binary outcome model exhibit some degree of bias. In particular, there is about 10% bias in the regression parameter estimate $\hat{\beta}_{1,1}$, and 7% to 12% bias in the parameter estimate $\hat{\alpha}_1$. In addition, these parameter estimates tend to have greater variability than those from our proposed method. This suggests that ignoring the informative measurement process may cause invalid or less efficient inference for some parameters in the joint model.

In Fig. 5, we plot the true and estimated (based on the means of 100 simulations) joint and marginal densities of the subject-specific random effects $\mathbf{u}_i$: (a) and (b) are for the true and estimated bivariate densities, respectively; (c) and (d) are for the marginal densities of the random intercept and baseline slope, respectively. It can be seen that the SNP approach can estimate the true bivariate and marginal densities of random effects very well, demonstrating the flexibility of the SNP representation to capture the complex distribution of random effects that are different from the normal.

6 Discussion

In this article, we propose a joint model that can naturally study the association between a primary binary outcome and longitudinal covariates that are possibly measured at discrete informative observation times. The association is described by latent subject-specific random variables whose distribution is flexibly modeled by the SNP representation, allowing the departure from normal assumption. Our approach provides a great insight into the relationship between early β -hCG profiles and the final pregnancy outcomes after IVF treatment. The results demonstrate the power of using the latent characteristics of early stage β -hCG profiles for the prediction of viable pregnancies achieved by IVF treatment, and the importance to take into account the informative measurement process and non-normally distributed random effects for data analysis.

Other modeling strategies proposed in the literature may also be used to analyze NYU-IVF data. A particular one tries to use the β -hCG value measured at some specific time point as a covariate in the logistic model for the primary pregnancy outcome. This approach may be problematic for NYU-IVF data for the following reasons. First, the observation time points are different for different women. Therefore, it is not possible to select a common time point

where all women had β -hCG measured. Second, if we could find a common time point at which β -hCG measurements were available for all women, using the β -hCG value at this particular time point as a covariate may not have a scientific and/or clinical justification. Third, suppose we had β -hCG data at a clinically meaningful time point, there is usually some degree of measurement error in this hormone data. It is well-known that putting an error-prone variable as a covariate in a logistic model will lead to a biased result.

If we follow the convention in mixed model literature to interpret

$Z_i(t) = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \gamma_3' X_i + u_{i0} + u_{i1} t$ as the true log β -hCG of subject i (with covariates X_i) at time point t , our joint model then implies that the viable pregnancy probability has a logistic regression model as follows

$$\text{logit}\{P(Y_i=1|X_i, t, \mathbf{u}_i)\} = \beta_0(t) + \tilde{\beta}_1' X_i + \alpha_1 Z_i(t) + (\alpha_2 - \alpha_1 t) Z_i'(t).$$

where $\beta_0(t) = \beta_0 - \alpha_1 \gamma_0 - \alpha_2 \gamma_1 - 2\alpha_2 \gamma_2 t + \alpha_1 \gamma_2 t^2$, $\tilde{\beta}_1' = \beta_1' - \alpha_1 \gamma_3'$. That is, given the results from our joint model, we can simultaneously investigate the effects of the true log β -hCG and its changing rate at any time on the primary pregnancy outcome. Since both $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are positive, changing rate of the true log β -hCG at baseline is most predictive of the viable pregnancy outcome. Similarly, we can show from our joint model that the logit of a viable pregnancy probability is (linearly) related to the average of the true log β -hCG and its changing rate in any specific interval $[0, T]$.

Acknowledgments

The research of Zhang was partly supported by NIH grants R01-CA85848-08, R01-MH84022-01 and R37-AI031789-20. The research of Lu was partly supported by R01-CA140632. The authors gratefully acknowledge the comments and suggestions of the editor, associate editor, and two reviewers, which have greatly improved the paper.

References

1. Bjercke S, Tanbo T, Dale PO, Morkrid L, Abyholm T. Human chorionic gonadotrophin concentrations in early pregnancy after in-vitro fertilization. *Human Reproduction*. 1999; 14:1642–1646. [PubMed: 10357992]
2. Chen HY, Little RJA. Proportional hazards regression with missing covariates. *Journal of the American Statistical Association*. 1999; 94:896–908.
3. Chen J, Zhang D, Davidian M. A Monte Carlo EM algorithm for generalized linear mixed models with flexible random effects distribution. *Biostatistics*. 2001; 1:1–27.
4. Chung K, Sammel MD, Coutifaris C, Chalian R, Lin K, Castelbaum AJ, Freedman MF, Barnhart KT. Defining the rise of serum hCG in viable pregnancies achieved through use of IVF. *Human Reproduction*. 2006; 21:823–828. [PubMed: 16311298]
5. Confino E, Demir RH, Friberg J, Gleicher N. The predictive value of hCG β subunit levels in pregnancies achieved by in vitro fertilization and embryo transfer: an international collaborative study. *Fertility and Sterility*. 1986; 45:526–531. [PubMed: 2420652]
6. Davidian M, Gallant AR. The nonlinear mixed effects model with a smooth random effects density. *Biometrika*. 1993; 80:475–488.
7. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*. 1977; 39:1–38.
8. Gallant AR, Nychka DW. Semiparametric maximum likelihood estimation. *Econometrica*. 1987; 55:363–390.

9. Glatstein IZ, Hornstein MD, Kahana MJ, Jackson KV, Friedman AJ. The predictive value of discriminatory human chorionic gonadotropin levels in the diagnosis of implantation outcome in vitro fertilization cycles. *Fertility and Sterility*. 1995; 63:350–356. [PubMed: 7843442]
10. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics*. 2000; 1:465–480. [PubMed: 12933568]
11. Hogan J, Laird N. Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine*. 1997; 16:259–272. [PubMed: 9004396]
12. Lambers MJ, Weering HGIV, Grunewold MDV, Lambalk CB, Homburg R, Schats R, Hopes PGA. Optimizing hCG cut-off values: A single determination on day 14 or 15 is sufficient for a reliable prediction of pregnancy outcome. *European Journal of Obstetrics and Gynecology and Reproductive Biology*. 2006; 127:94–98. [PubMed: 16466846]
13. Li E, Zhang D, Davidian M. Conditional estimation for generalized linear models when covariates are subject-specific parameters in a mixed model for longitudinal measurements. *Biometrics*. 2004; 60:1–7. [PubMed: 15032767]
14. Louis TA. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*. 1982; 44:226–233.
15. Meilijson E. A fast improvement to the EM algorithm on its own terms. *Journal of the Royal Statistical Society, Series B*. 1989; 51:127–138.
16. Shamoni M, Frattarelli JL, Bergh PA, Scott RT. Logarithmic curves depicting initial level and rise of serum beta human chorionic gonadotropin and live delivery outcomes with in vitro fertilization: an analysis of 6021 pregnancies. *Fertility and Sterility*. 2009; 91:1760–1764. [PubMed: 18455162]
17. Strandell A, Thorburn J, Hamberger L. Risk factors for ectopic pregnancy in assisted reproduction. *Fertility and Sterility*. 1999; 71:282–286. [PubMed: 9988399]
18. Song X, Davidian M, Tsiatis AA. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics*. 2002; 58:742–753. [PubMed: 12495128]
19. Tsiatis AA, Davidian M. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*. 2001; 88:447–458.
20. Vonesh EF, Greene T, Schluchter MD. Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in Medicine*. 2006; 25:143–163. [PubMed: 16025541]
21. Wang CY, Wang N, Wang S. Regression analysis when covariates are regression parameters of a random-effects model for observed longitudinal measurements. *Biometrics*. 2000; 56:487–495. [PubMed: 10877308]
22. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics*. 1997; 53:330–339. [PubMed: 9147598]
23. Zeng D, Cai J. Simultaneous modeling of survival and longitudinal data with an application to repeated quality of life measures. *Lifetime Data Analysis*. 2005; 11:151–74. [PubMed: 15940822]
24. Zhang D, Davidian M. Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*. 2001; 57:795–802. [PubMed: 11550930]

Appendix A: The details of EM algorithm for Our Joint Model

The log complete-data likelihood is given by

$$l_c(\Theta) = \sum_{i=1}^n [\log\{f_L(\mathbf{Z}_i^*|\mathbf{u}_i)\} + \log\{f_p(Y_i|\mathbf{u}_i)\} + \log\{f_R(\mathbf{R}_i|\mathbf{u}_i)\} + \log\{f_u(\mathbf{u}_i)\}]. \quad (9)$$

In the k th E-step, we need to calculate the conditional expectation of $l_c(\Theta)$ with respect to the random effects $\mathbf{u}_i$ given the observed data $\mathbf{O}$ and the current parameter estimates $\hat{\Theta}^k$, commonly called the Q function and denoted as $Q(\Theta|\hat{\Theta}^k)$. From equation (9), the conditional expectation $Q(\Theta|\hat{\Theta}^k)$ can be expressed as

$Q(\Theta|\widehat{\Theta}^{(k)})=Q_1(\gamma, \sigma_\varepsilon^2, \rho; \mathbf{O}, \widehat{\Theta}^{(k)})+Q_2(\beta, \alpha_1, \alpha_2, \mathbf{O}, \widehat{\Theta}^{(k)})+Q_3(\lambda_1, \dots, \lambda_m; \mathbf{O}, \widehat{\Theta}^{(k)})+Q_4(\mu, \xi, \varphi; \mathbf{O}, \widehat{\Theta}^{(k)})$, where

$$\begin{aligned} Q_1(\gamma, \sigma_\varepsilon^2, \rho; \mathbf{O}, \widehat{\Theta}^{(k)}) &= \sum_{i=1}^n E[\log\{f_\varepsilon(\mathbf{Z}_i^*|\mathbf{u}_i)|\mathbf{O}_i, \widehat{\Theta}^{(k)}\}] \\ Q_2(\beta, \alpha_1, \alpha_2; \mathbf{O}, \widehat{\Theta}^{(k)}) &= \sum_{i=1}^n E[\log\{f_p(Y_i|\mathbf{u}_i)|\mathbf{O}_i, \widehat{\Theta}^{(k)}\}] \\ Q_3(\lambda_1, \dots, \lambda_m; \mathbf{O}, \widehat{\Theta}^{(k)}) &= \sum_{i=1}^n E[\log\{f_R(\mathbf{R}_i|\mathbf{u}_i)|\mathbf{O}_i, \widehat{\Theta}^{(k)}\}] \\ Q_4(\mu, \xi, \varphi; \mathbf{O}, \widehat{\Theta}^{(k)}) &= \sum_{i=1}^n E[\log\{f_u(\mathbf{u}_i)|\mathbf{O}_i, \widehat{\Theta}^{(k)}\}], \end{aligned} \quad (10)$$

and the expectation $E(\cdot|\mathbf{O}_i, \widehat{\Theta}^{(k)})$ is taken with respect to the conditional density function $f(\mathbf{u}_i|\mathbf{O}_i, \widehat{\Theta}^{(k)})$ of $\mathbf{u}$ given $\mathbf{O}$ evaluated at $\widehat{\Theta}^{(k)}$. By simple calculation, we have

$$\begin{aligned} f(\mathbf{u}_i|\mathbf{O}_i, \widehat{\Theta}^{(k)}) &= \frac{f(\mathbf{O}_i, \mathbf{u}_i|\widehat{\Theta}^{(k)})}{\int f(\mathbf{O}_i, \mathbf{u}_i|\widehat{\Theta}^{(k)})d\mathbf{u}_i} \\ &= \frac{f_p(Y_i|\mathbf{u}_i, \widehat{\Theta}^{(k)})f_R(\mathbf{R}_i|\mathbf{u}_i, \widehat{\Theta}^{(k)})p_K^2\{D^{-1}(\mathbf{u}_i-\mu)\}g(\mathbf{u}_i|\mathbf{Z}_i^*, \widehat{\Theta}^{(k)})}{\int f_p(Y_i|\mathbf{u}_i, \widehat{\Theta}^{(k)})f_R(\mathbf{R}_i|\mathbf{u}_i, \widehat{\Theta}^{(k)})p_K^2\{D^{-1}(\mathbf{u}_i-\mu)\}g(\mathbf{u}_i|\mathbf{Z}_i^*, \widehat{\Theta}^{(k)})d\mathbf{u}_i}, \end{aligned} \quad (11)$$

where $g(\mathbf{u}_i|\mathbf{Z}_i^*, \widehat{\Theta}^{(k)})$ is the density of a bivariate normal distribution with mean

$$\mu_i^{(k)} = \mathbf{W}_i \left\{ \mathbf{G}_i' \sum_i^{-1} (\mathbf{Z}_i^* - \gamma_2' X_i) + (DD')^{-1} \mu \right\}$$

and variance-covariance matrix

$$\mathbf{W}_i^{(k)} = \left\{ (DD')^{-1} + \mathbf{G}_i' \sum_i^{-1} \mathbf{G}_i \right\}^{-1},$$

with $\mathbf{G}_i = (\mathbf{1}, \mathbf{t}_i^*)$ being an $m_i \times 2$ matrix. This distribution can be viewed as the “working” conditional distribution of $\mathbf{u}_i$ given $\mathbf{Z}_i^*$ and $\widehat{\Theta}^{(k)}$. Then the integrations in (10) and (11) can be computed using numerical integration techniques. Here we use the Gaussian quadrature method.

Next, in the M-step, it is clear that the maximization of $Q(\Theta|\widehat{\Theta}^{(k)})$ can be conducted separately for each component in $Q(\Theta|\widehat{\Theta}^{(k)})$.

Specifically, $Q_1(\gamma, \sigma_\varepsilon^2, \rho; \mathbf{O}, \widehat{\Theta}^{(k)})$, the conditional expectation of log-likelihood of $(\gamma, \sigma_\varepsilon^2, \rho)$, takes the following form:

$$-\frac{1}{2} \sum_{i=1}^n \log \left| \sum_i \right| - \frac{1}{2} \sum_{i=1}^n (\mathbf{Z}_i^* - \gamma_0 - \gamma_1 \mathbf{t}_i^* - \gamma_2' X_i - u_{i0}^{(k)} - u_{i1}^{(k)} \mathbf{t}_i^*)' \sum_i^{-1} \quad (12)$$

$$\times (\mathbf{Z}_i^* - \gamma_0 - \gamma_1 \mathbf{t}_i^* - \gamma_2' X_i - u_{i0}^{(k)} - u_{i1}^{(k)} \mathbf{t}_i^*) \quad (13)$$

where $u_{i0}^{(k)} = E(u_{i0} | \mathbf{O}_i, \widehat{\Theta}^{(k)})$ and $u_{i1}^{(k)} = E(u_{i1} | \mathbf{O}_i, \widehat{\Theta}^{(k)})$, and $\sum_i = \sigma_\varepsilon^2 \Gamma_i(\rho)$, it can be shown by simple algebra that

$$\widehat{\gamma}^{(k+1)} = \left(\sum_{i=1}^n X_i' (\widehat{\sum}_i^{(k)})^{-1} X_i \right)^{-1} \sum_{i=1}^n X_i' (\widehat{\sum}_i^{(k)})^{-1} (\mathbf{Z}_i^* - u_{i0}^{(k)} - u_{i1}^{(k)} \mathbf{t}_i^*) \quad (14)$$

where $(\widehat{\sum}_i^{(k)})^{-1} = (\widehat{\sigma}_\varepsilon^2)^{(k)} \Gamma_i(\widehat{\rho}^{(k)})$, $u_{i0}^{(k)} = E(u_{i0} | \mathbf{O}_i, \widehat{\Theta}^{(k)})$ and $u_{i1}^{(k)} = E(u_{i1} | \mathbf{O}_i, \widehat{\Theta}^{(k)})$.

The updated estimators $(\widehat{\sigma}_\varepsilon^{2(k+1)}, \widehat{\rho}^{(k+1)})$ at $(k+1)$ th step can be obtained by solving the “score equation” of $Q_1(\gamma, \sigma_\varepsilon^2, \rho; \mathbf{O}, \widehat{\Theta}^{(k)})$ with respect to σ_ε^2 and ρ . For example, if subject i has $m_i = 4$ observations. Let $d_{jk} = |t_{ij} - t_{ik}|$ be the length of time between times t_{ij} and t_{ik} for all $j, k = 1, \dots, m_i$, then

$$\Gamma_i(\rho) = \begin{pmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} & \rho^{d_{14}} \\ \rho^{d_{12}} & 1 & \rho^{d_{23}} & \rho^{d_{24}} \\ \rho^{d_{13}} & \rho^{d_{23}} & 1 & \rho^{d_{34}} \\ \rho^{d_{14}} & \rho^{d_{24}} & \rho^{d_{34}} & 1 \end{pmatrix}.$$

The derivative of $Q_1(\gamma, \sigma_\varepsilon^2, \rho; \mathbf{O}, \widehat{\Theta}^{(k)})$ with respect to σ_ε^2 and ρ have the following expressions:

$$S_{(k)}(\sigma_\varepsilon^2) = -\frac{1}{2} \sum_{i=1}^n \text{tr} \left\{ \sum_i^{-1} \Gamma_i(\rho) \right\} + \frac{1}{2} \sum_{i=1}^n (\mathbf{Z}_i^* - \gamma_0 - \gamma_1 \mathbf{t}_i^* - \gamma_2 X_i - u_{i0}^{(k)} - u_{i1}^{(k)} \mathbf{t}_i^*)' \sum_i^{-1} \Gamma_i(\rho) \sum_i^{-1} \times (\mathbf{Z}_i^* - \gamma_0 - \gamma_1 \mathbf{t}_i^* - \gamma_2 X_i - u_{i0}^{(k)} - u_{i1}^{(k)} \mathbf{t}_i^*) \quad (15)$$

$$S_{(k)}(\rho) = -\frac{1}{2} \sum_{i=1}^n \text{tr} \left(\sum_i^{-1} \sigma_\varepsilon^2 \frac{\partial \Gamma_i(\rho)}{\partial \rho} \right) + \frac{1}{2} \sum_{i=1}^n (\mathbf{Z}_i^* - \gamma_0 - \gamma_1 \mathbf{t}_i^* - \gamma_2 X_i - u_{i0}^{(k)} - u_{i1}^{(k)} \mathbf{t}_i^*)' \times \sum_i^{-1} \sigma_\varepsilon^2 \frac{\partial \Gamma_i(\rho)}{\partial \rho} \sum_i^{-1} (\mathbf{Z}_i^* - \gamma_0 - \gamma_1 \mathbf{t}_i^* - \gamma_2 X_i - u_{i0}^{(k)} - u_{i1}^{(k)} \mathbf{t}_i^*) \quad (16)$$

Let $S_{(k)}(\sigma_\varepsilon^2, \rho) = (S_{(k)}(\sigma_\varepsilon^2), S_{(k)}(\rho))'$, $S_{(k)}(\rho)$. We then solve the following equation

$$S_{(k)}(\sigma_\varepsilon^2, \rho) = 0 \quad (17)$$

to obtain the update for $\psi = (\sigma_\varepsilon^2, \rho)$. However, the “score equation” (17) does not have a closed form solution. We then propose to update $\psi = (\sigma_\varepsilon^2, \rho)$ by one-step Newton-Raphson algorithm

$$\widehat{\psi}^{(k+1)} = \widehat{\psi}^{(k)} + I_{(k)}^{-1}(\widehat{\psi}^{(k)}) S_{(k)}(\widehat{\psi}^{(k)}), \quad (18)$$

where $S_{(k)}(\widehat{\psi}^{(k)}) = S_{(k)}(\widehat{\sigma}_\varepsilon^{2(k)}, \widehat{\rho}^{(k)})$ is $S_{(k)}(\sigma_\varepsilon^2, \rho)$ evaluated at $\widehat{\psi}^{(k)} = (\widehat{\sigma}_\varepsilon^{2(k)}, \widehat{\rho}^{(k)})$, and $I_{(k)}(\psi) = I_{(k)}(\sigma_\varepsilon^2, \rho)$ is the pseudo Fisher score matrix of ψ and defined by

$$I(\psi) = \begin{pmatrix} \frac{\partial S_{(k)}(\sigma_\varepsilon^2)}{\partial \sigma_\varepsilon^2} & \frac{\partial S_{(k)}(\sigma_\varepsilon^2)}{\partial \rho} \\ \frac{\partial S_{(k)}(\sigma_\varepsilon^2)}{\partial \rho} & \frac{\partial S_{(k)}(\rho)}{\partial \rho} \end{pmatrix} \quad (19)$$

where $\frac{\partial S_{(k)}(\sigma_\varepsilon^2)}{\partial \sigma_\varepsilon^2} = \frac{1}{2} \text{tr} \left(\sum_i^{-1} \frac{\partial \Sigma_i}{\partial \sigma_\varepsilon^2} \sum_i^{-1} \frac{\partial \Sigma_i}{\partial \sigma_\varepsilon^2} \right)$, $\frac{\partial S_{(k)}(\sigma_\varepsilon^2)}{\partial \rho} = \frac{1}{2} \text{tr} \left(\sum_i^{-1} \frac{\partial \Sigma_i}{\partial \sigma_\varepsilon^2} \sum_i^{-1} \frac{\partial \Sigma_i}{\partial \rho} \right)$ and $\frac{\partial S_{(k)}(\rho)}{\partial \rho} = \frac{1}{2} \text{tr} \left(\sum_i^{-1} \frac{\partial \Sigma_i}{\partial \rho} \sum_i^{-1} \frac{\partial \Sigma_i}{\partial \rho} \right)$. The updated estimators $\widehat{\beta}^{(k+1)}$, $\widehat{\alpha}_1^{(k+1)}$ and $\widehat{\alpha}_2^{(k+1)}$ can also be obtained by the Newton-Raphson (or one-step Newton-Raphson) method based on the “score function” and “information matrix” of $Q_2(\beta, \alpha_3, \alpha_4; \mathbf{O}, \widehat{\Theta}^{(k)})$. The updated estimators $(\widehat{\lambda}_1^{(k+1)}, \dots, \widehat{\lambda}_m^{(k+1)})$ can be obtained in a similar way based on the “score function” and “information matrix” of $Q_3(\lambda_1, \dots, \lambda_m; \mathbf{O}, \widehat{\Theta}^{(k)})$. The $(k+1)$ th update of μ , ξ and φ can be found by using any optimization software, e.g., nlm in R, to maximize $Q_4(\mu, \xi, \varphi; \mathbf{O}, \widehat{\Theta}^{(k)})$ under the constraint $E(\mathbf{u}_j) = 0$. The details can be found in the Appendix B. We iterate the E-step and M-step until a pre-specified convergence criterion is met. Let $\widehat{\Theta}$ denote the final estimator at convergence.

Appendix B: Updating Parameters μ , D and φ in the EM Algorithm

At the k th step in the EM algorithm, $Q_4(\mu, \xi, \varphi; \mathbf{O}, \widehat{\Theta}^{(k)})$ takes the following form:

$$\sum_{i=1}^n E \left[\log \left\{ P_K^2 \{ D^{-1}(\mathbf{u}_i - \mu) \} \right\} \right] - \frac{1}{2} \sum_{i=1}^n E \left[(\mathbf{u}_i - \mu)' (DD')^{-1} (\mathbf{u}_i - \mu) \right] - n \log |D| \quad (20)$$

To satisfy $E(\mathbf{u}_j) = 0$, let $E(\mathbf{u}_j) = \mu + D \cdot E(\mathbf{b}_j) = 0$, then

$$\widehat{\mu}(D, \varphi) = -D \cdot E(\mathbf{b}_i) \quad (21)$$

Plugging the above $\widehat{\mu}(D, \varphi)$ into (20) we have

$$Q_4(\widehat{\mu}(D, \varphi), \xi, \varphi; \mathbf{O}, \widehat{\Theta}^{(k)}) = \sum_{i=1}^n E \left[\log \left\{ P_K^2 \{ D^{-1}(\mathbf{u}_i - \widehat{\mu}(D, \varphi)) \} \right\} \right] - \frac{1}{2} \sum_{i=1}^n E \left[(\mathbf{u}_i - \widehat{\mu}(D, \varphi))' (DD')^{-1} (\mathbf{u}_i - \widehat{\mu}(D, \varphi)) \right] - n \log |D| \quad (22)$$

then the updated estimators $(\widehat{D}^{(k+1)}, \widehat{\varphi}^{(k+1)})$ at $(k+1)$ th step can be obtained by using standard optimization software to maximize $Q_4(\widehat{\mu}(D, \varphi), \xi, \varphi; \mathbf{O}, \widehat{\Theta}^{(k)})$. The updated estimator for μ is then $\widehat{\mu}^{(k+1)} = \widehat{\mu}(\widehat{D}^{(k+1)}, \widehat{\varphi}^{(k+1)})$.

$E(\mathbf{b}_j)$ varies for different K values. To ensure $\int g_K(\mathbf{b}) d\mathbf{b} = 1$, we need the quadratic constraint $a^T A a = 1$. Because A is positive definite, there exists B positive definite such that $A = BB^T$. Let $c = B^T a$, then $c^T c = 1$ and $a = (B^T)^{-1} c$.

For $K = 1$,

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (23)$$

and $P_1^2(\mathbf{b}) = (a_{00} + a_{10}b_1 + a_{01}b_2)^2$. Let $a = (a_{00}, a_{10}, a_{01})'$, define $c = (c_1, c_2, c_3)'$, where $c_1 = \sin\varphi_1$, $c_2 = \cos\varphi_1 \sin\varphi_2$ and $c_3 = \cos\varphi_1 \cos\varphi_2$. $a = c$ and

$$E(\mathbf{b}_i) = E \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 2a_{00}a_{10} \\ 2a_{00}a_{01} \end{pmatrix} \quad (24)$$

For $K = 2$,

$$A = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 3 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 3 \end{pmatrix}, \quad (25)$$

and $P_2^2(\mathbf{b}) = (a_{00} + a_{10}b_1 + a_{01}b_2 + a_{20}b_1^2 + a_{11}b_1b_2 + a_{02}b_2^2)^2$. Let $a = (a_{00}, a_{10}, a_{01}, a_{20}, a_{11}, a_{02})'$, define $c = (c_1, c_2, c_3, c_4, c_5, c_6)'$ where $c_1 = \sin\varphi_1$, $c_2 = \cos\varphi_1 \sin\varphi_2$, $c_3 = \cos\varphi_1 \cos\varphi_2 \sin\varphi_3$, $\dots$, $c_6 = \cos\varphi_1 \cos\varphi_2 \dots \cos\varphi_5$. Let $c = B'a$, then $a = (B')^{-1}c$ and

$$E(\mathbf{b}_i) = E \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 2a_{00}a_{10} + 2a_{10}a_{02} + 2a_{01}a_{11} + 6a_{10}a_{20} \\ 2a_{00}a_{01} + 2a_{10}a_{11} + 2a_{01}a_{20} + 6a_{01}a_{02} \end{pmatrix} \quad (26)$$

Appendix C: EM-Aided Algorithm for Variance Estimates

Define $\theta = (\gamma, \beta, \alpha_1, \alpha_2, \lambda_1, \dots, \lambda_m)$, the regression parameters of our main interest, and ζ the remaining nuisance parameters in Θ . Let $\hat{\theta}$ denote the MLE of θ obtained by the

proposed EM algorithm. Moreover, let $\tilde{\zeta}(\theta) = (\tilde{\sigma}_\varepsilon^2(\theta), \tilde{\rho}(\theta), \tilde{\mu}'(\theta), \tilde{\xi}'(\theta), \tilde{\varphi}'(\theta))'$ denote the maximizer of the observed likelihood function over ζ with θ fixed, which can be obtained by a similar EM algorithm as proposed. Then we calculate the information matrix for θ as follows:

- i. Perturb the j th component of $\hat{\theta}$ by a small amount h from both sides, i.e. $\hat{\theta}_j + h$ and $\hat{\theta}_j - h$, while keeping other components of $\hat{\theta}$ unchanged; denote the resulting parameter vectors by $\hat{\theta}^{j+}$ and $\hat{\theta}^{j-}$ respectively, and run the EM algorithm to obtain $\hat{\zeta}(\hat{\theta}^{j+})$ and $\hat{\zeta}(\hat{\theta}^{j-})$ accordingly.
- ii. Compute the j th row of the information matrix of θ by

$$\frac{1}{2h} \left[E \{ l_c^\theta(\hat{\theta}^{j+}, \hat{\zeta}(\hat{\theta}^{j+})) | \mathbf{O}, \hat{\theta}^{j+}, \hat{\zeta}(\hat{\theta}^{j+}) \} - E \{ l_c^\theta(\hat{\theta}^{j-}, \hat{\zeta}(\hat{\theta}^{j-})) | \mathbf{O}, \hat{\theta}^{j-}, \hat{\zeta}(\hat{\theta}^{j-}) \} \right]$$

where $l_c^\theta(\widehat{\theta}^{j+}, \zeta(\widehat{\theta}^{j+}))$ and $l_c^\theta(\widehat{\theta}^{j-}, \zeta(\widehat{\theta}^{j-}))$ denote the derivatives of the complete-data log likelihood $l_C(\theta, \zeta)$ with respect to θ evaluated at $(\widehat{\theta}^{j+}, \zeta(\widehat{\theta}^{j+}))$ and $(\widehat{\theta}^{j-}, \zeta(\widehat{\theta}^{j-}))$ respectively.

The derivation of $l_c^\theta(\theta, \zeta)$ is straightforward and hence omitted here. For the perturbation size h , [2] suggested to use $h = c/n$, where c is a positive constant. We found that $c = 0.15$ would produce satisfactory results in all of our numerical studies.

Fig. 1. Log transformed $\beta\text{-hCG}$ profiles for 30 randomly selected patients from the non-viable pregnancy group (left) and viable pregnancy group (right), respectively.

Fig. 2.

(a) Q-Q plot of subject-specific intercept estimates from individual least squares fits. (b) Q-Q plot of subject-specific slope estimates from individual least squares fits.

Fig. 3.

Univariate SNP densities with some selected φ 's for $K=2$: (a) $\varphi_1 = \pi/2$, $\varphi_2 = 3\pi/10$; (b) $\varphi_1 = \pi/2$, $\varphi_2 = \pi/10$; (c) $\varphi_1 = \pi/10$, $\varphi_2 = 3\pi/10$; (d) $\varphi_1 = \pi/10$, $\varphi_2 = \pi/2$.

Fig. 4.

(a) Estimated bivariate density of $\mathbf{u}_i$; (b) Contour plots of the density in (a) with posterior estimates of $\mathbf{u}_i$ superimposed (contours are 5, 50 and 95%); (c) and (d) Estimated marginal densities for the random intercept and slope, superimposed by the histograms for posterior estimates of $\mathbf{u}_i$ respectively.

Fig. 5. (a) and (b) True and estimated bivariate density of $\mathbf{u}_i$ by HQ; (c) and (d) True marginal density (solid line) and estimated marginal density chosen by HQ (dash-dotted line) for the random intercept an slope, respectively.

\$watermark-text

\$watermark-text

\$watermark-text

Table 1

Proportion of women having β -hCG measures at each time interval.

	1	2	3	4	5	6
Viable	91.7%	15.9%	7.3%	87.0%	15.7%	5.1%
Non-viable	91.7%	50.8%	25.0%	73.5%	37.9%	36.4%

Table 2

Estimation of parameters in the β -hCG longitudinal submodel and binary pregnancy outcome submodel in the proposed full model and reduced model for NYU-IVF data

Para	Full model $K=0$ (A)			Full model $K=2$ (B)			Reduced model $K=0$ (C)			Reduced model $K=1$ (D)		
	Est	SE	p-value	Est	SE	p-value	Est	SE	p-value	Est	SE	p-value
longitudinal submodel for β -hCG												
γ_0 (intercept)	4.605	0.071	<0.0001	4.606	0.066	<0.0001	4.613	0.057	<0.0001	4.729	0.048	<0.0001
γ_1 (slope)	2.952	0.170	<0.0001	2.881	0.144	<0.0001	2.593	0.153	<0.0001	2.610	0.126	<0.0001
γ_2 (quadratic)	-0.432	0.122	0.0003	-0.481	0.111	<0.0001	-0.357	0.101	0.0004	-0.349	0.093	0.0001
$\gamma_{3,1}$ (age)	-0.185	0.191	0.33	-0.190	0.155	0.22	-0.189	0.167	0.25	-0.183	0.141	0.19
$\gamma_{3,2}$ (BMI)	-0.385	0.156	0.01	-0.371	0.124	0.003	-0.396	0.132	0.002	-0.378	0.115	0.001
σ_e^2	0.395			0.385			0.392			0.367		
ρ	0.276	0.061	<0.0001	0.303	0.066	<0.0001	0.288	0.063	<0.0001	0.305	0.058	<0.0001
binary outcome submodel for viable pregnancy												
β_0	1.633	0.216	<0.0001	1.583	0.199	<0.0001	1.609	0.211	<0.0001	1.642	0.209	<0.0001
$\beta_{1,1}$ (age)	-1.180	0.403	0.003	-1.104	0.361	0.002	-1.283	0.398	0.001	-1.083	0.342	0.001
$\beta_{1,2}$ (BMI)	-0.352	0.260	0.17	-0.359	0.229	0.12	-0.347	0.253	0.17	-0.360	0.223	0.10
a_1	1.523	0.447	0.0006	1.610	0.324	<0.0001	1.415	0.747	0.058	1.278	0.692	0.064
a_2	2.059	0.720	0.004	1.527	0.415	0.0002	1.762	0.639	0.005	1.609	0.627	0.010

SE is the estimated standard error.

\$watermark-text

\$watermark-text

\$watermark-text

Table 3
Estimation of parameters in β -hCG observation times submodel of the proposed full model for NYU-IVF data

Para	Full model $K=0$ (A)			Full model $K=2$ (B)		
	Est	SE	p-value	Est	SE	p-value
observation times model						
η_0	-2.233	0.745	0.002	-2.412	0.415	<0.0001
$\eta_{2,1}(\text{age})$	0.110	0.429	0.79	0.175	0.436	0.68
$\eta_{2,2}(\text{BMI})$	0.258	0.350	0.46	0.367	0.327	0.26
α_{32}	-5.066	1.470	0.0005	-4.684	1.022	<0.0001
α_{42}	1.320	0.513	0.01	0.636	0.345	0.07
η_0	-2.668	0.259	<0.0001	-2.674	0.221	<0.0001
$\eta_{3,1}(\text{age})$	0.384	0.380	0.31	0.397	0.361	0.27
$\eta_{3,2}(\text{BMI})$	-0.259	0.318	0.41	-0.249	0.308	0.41
α_{33}	-2.649	1.059	0.012	-2.321	0.476	<0.0001
α_{43}	0.771	0.610	0.20	0.383	0.305	0.20
η_0	2.697	0.921	0.003	2.178	0.217	<0.0001
$\eta_{4,1}(\text{age})$	-0.401	0.409	0.32	-0.366	0.310	0.24
$\eta_{4,2}(\text{BMI})$	-0.532	0.338	0.11	-0.492	0.275	0.073
α_{34}	4.083	1.434	0.004	2.137	0.460	<0.0001
α_{44}	-1.413	1.763	0.42	-0.331	0.211	0.11
η_0	-2.053	0.553	0.0002	-1.894	0.190	<0.0001
$\eta_{5,1}(\text{age})$	0.837	0.399	0.035	0.698	0.318	0.028
$\eta_{5,2}(\text{BMI})$	0.562	0.286	0.049	0.531	0.207	0.01
α_{35}	-3.539	1.675	0.03	-2.068	0.513	<0.0001
α_{45}	1.180	1.402	0.40	0.664	0.441	0.13
η_0	-2.438	0.228	<0.0001	-2.425	0.213	<0.0001
$\eta_{6,1}(\text{age})$	0.148	0.345	0.66	0.190	0.317	0.55
$\eta_{6,2}(\text{BMI})$	0.237	0.255	0.35	0.286	0.243	0.23
α_{36}	-2.059	0.491	<0.0001	-1.994	0.527	0.0001

Para	Full model $K=0$ (A)			Full model $K=2$ (B)		
	Est	SE	p-value	Est	SE	p-value
$\alpha_{4,6}$	0.242	0.273	0.37	0.188	0.187	0.31

SE is the estimated standard error.

Table 4

Simulation results: $K = 0$ denotes the estimation by assuming normal random effects. HQ is the estimation when K is selected by the HQ criterion.

Para	true	Est	K=0				HQ			
			SD	SE	CP	Est	SD	SE	CP	
(a)Proposed full joint model:										
longitudinal submodel										
γ_0	1	0.987	0.237	0.220	93 %	0.986	0.175	0.177	96 %	
γ_1	0.5	0.498	0.080	0.087	97 %	0.502	0.080	0.085	96 %	
$\gamma_{2,1}(X_1)$	0.5	0.483	0.156	0.157	94 %	0.502	0.079	0.077	93 %	
$\gamma_{2,2}(X_2)$	-0.5	-0.490	0.327	0.312	94 %	-0.504	0.160	0.161	96 %	
σ_e^2	1	0.993				1.005				
ρ	0.3	0.296				0.304				
binary outcome submodel										
β_0	0.8	0.783	0.396	0.375	93 %	0.814	0.373	0.361	96 %	
$\beta_{1,1}(X_1)$	-1	-1.048	0.277	0.286	98 %	-1.050	0.237	0.245	97 %	
$\beta_{1,2}(X_2)$	-1.5	-1.568	0.522	0.531	95 %	-1.550	0.434	0.442	93 %	
α_1	-1	-1.071	0.173	0.171	96 %	-1.041	0.152	0.156	98 %	
α_2	1	1.061	0.222	0.209	97 %	1.048	0.202	0.204	92 %	
informative times submodel										
η_0	1.6	1.637	0.213	0.216	96 %	1.617	0.199	0.205	96 %	
η_2	1	1.037	0.218	0.206	93 %	1.015	0.212	0.201	93 %	
η_3	1.5	1.520	0.209	0.213	95 %	1.498	0.195	0.189	94 %	
η_4	1	1.006	0.222	0.205	95 %	0.984	0.199	0.193	94 %	
η_5	1	1.040	0.219	0.206	94 %	1.019	0.189	0.190	96 %	
η_6	1.4	1.402	0.220	0.211	93 %	1.385	0.192	0.183	94 %	
$\eta_{1,1}(X_1)$	1	0.991	0.104	0.110	95 %	1.002	0.082	0.087	96 %	
$\eta_{1,2}(X_2)$	-0.5	-0.502	0.228	0.204	94 %	-0.513	0.157	0.149	94 %	
α_3	0.5	0.526	0.043	0.041	94 %	0.510	0.040	0.040	96 %	
α_4	-0.5	-0.504	0.052	0.058	98 %	-0.502	0.051	0.059	98 %	

\$watermark-text

\$watermark-text

\$watermark-text

Para	true	K=0				HQ			
		Est	SD	SE	CP	Est	SD	SE	CP
% K									
(2, 84, 14)									
(b)Reduced model:									
longitudinal submodel									
γ_0	1	1.014	0.237	0.218	91%	0.982	0.178	0.169	90%
γ_1	0.5	0.401	0.079	0.085	84%	0.408	0.078	0.086	87%
$\gamma_{2,1}(X_1)$	0.5	0.479	0.157	0.156	94%	0.501	0.080	0.085	96%
$\gamma_{2,2}(X_2)$	-0.5	-0.488	0.326	0.310	94%	-0.504	0.164	0.175	97%
σ_e^2	1	0.992				1.022			
ρ	0.3	0.290				0.295			
binary outcome submodel									
β_0	0.8	0.724	0.395	0.383	93 %	0.819	0.381	0.330	88%
$\beta_{1,1}(X_1)$	-1	-1.096	0.312	0.304	99 %	-1.102	0.249	0.256	99%
$\beta_{1,2}(X_2)$	-1.5	-1.564	0.535	0.551	96 %	-1.527	0.448	0.457	94%
a_1	-1	-1.124	0.212	0.197	100 %	-1.073	0.157	0.169	98%
a_2	1	1.088	0.260	0.234	96 %	1.074	0.210	0.210	96%
% K									
(1, 77, 22)									

SD is the sample standard deviation; SE is the mean of estimated standard errors; CP is the empirical coverage probability of 95% Wald-type confidence intervals; %K represents the proportions of $K = 0, 1$ or 2 out of 100 runs selected by the HQ criterion.