Journal of Multivariate Analysis 143 (2016) 94–106

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Bayesian regression analysis of data with random effects covariates from nonlinear longitudinal measurements

Rolando De la Cruz [a,*], Cristian Meza [b], Ana Arribas-Gil [c], Raymond J. Carroll [d]

[a] Institute of Statistics, Pontificia Universidad Católica de Valparaíso, Av. Errazuriz 2734, Valparaíso, Chile

[b] Centro de Investigación y Modelamiento de Fenómenos Aleatorios—CIMFAV, Faculty of Engineering, Universidad de Valparaíso, General Cruz 222, Valparaíso, Chile

[c] Departamento de Estadística, Universidad Carlos III de Madrid, Getafe, Spain

[d] Department of Statistics, Texas A&M University, College Station, TX 77843, USA

ARTICLE INFO

Article history:
Received 9 October 2014
Available online 14 September 2015

AMS subject classification:
62F15
62J02
62J12

Keywords:
Autocorrelated errors
Generalized linear models
Joint modelling
Longitudinal data
MCMC methods
Nonlinear mixed-effects model

ABSTRACT

Joint models for a wide class of response variables and longitudinal measurements consist on a mixed-effects model to fit longitudinal trajectories whose random effects enter as covariates in a generalized linear model for the primary response. They provide a useful way to assess association between these two kinds of data, which in clinical studies are often collected jointly on a series of individuals and may help understanding, for instance, the mechanisms of recovery of a certain disease or the efficacy of a given therapy. When a nonlinear mixed-effects model is used to fit the longitudinal trajectories, the existing estimation strategies based on likelihood approximations have been shown to exhibit some computational efficiency problems (De la Cruz et al., 2011). In this article we consider a Bayesian estimation procedure for the joint model with a nonlinear mixed-effects model for the longitudinal data and a generalized linear model for the primary response. The proposed prior structure allows for the implementation of an MCMC sampler. Moreover, we consider that the errors in the longitudinal model may be correlated. We apply our method to the analysis of hormone levels measured at the early stages of pregnancy that can be used to predict *normal* versus *abnormal* pregnancy outcomes. We also conduct a simulation study to assess the importance of modelling correlated errors and quantify the consequences of model misspecification.

1. Introduction

In many biomedical studies longitudinal biomarker profiles carry important information about the outcome of a therapy, a disease or a particular condition. In such cases, the association between the response or outcome and a series of longitudinal measurements is of primary interest. In Fig. 1 we illustrate one example that motivates the current paper. The longitudinal measurements of this data set represent beta human chorionic gonadotropin (β-HCG) levels measured over time on 173 pregnant woman during the first 80 days of gestation. Here, the response of interest for each woman is given by her pregnancy outcome: *normal*, if she had a normal delivery or *abnormal* if she had any complication resulting in a nonterminal delivery and loss of the foetus. In such a framework a relevant question is how the variation of hormone concentration

* Corresponding author.
E-mail addresses: rolando.delacruz@pucv.cl (R. De la Cruz), cristian.meza@uv.cl (C. Meza), ana.arribas@uc3m.es (A. Arribas-Gil), carroll@stat.tamu.edu (R.J. Carroll).

http://dx.doi.org/10.1016/j.jmva.2015.08.020
0047-259X/

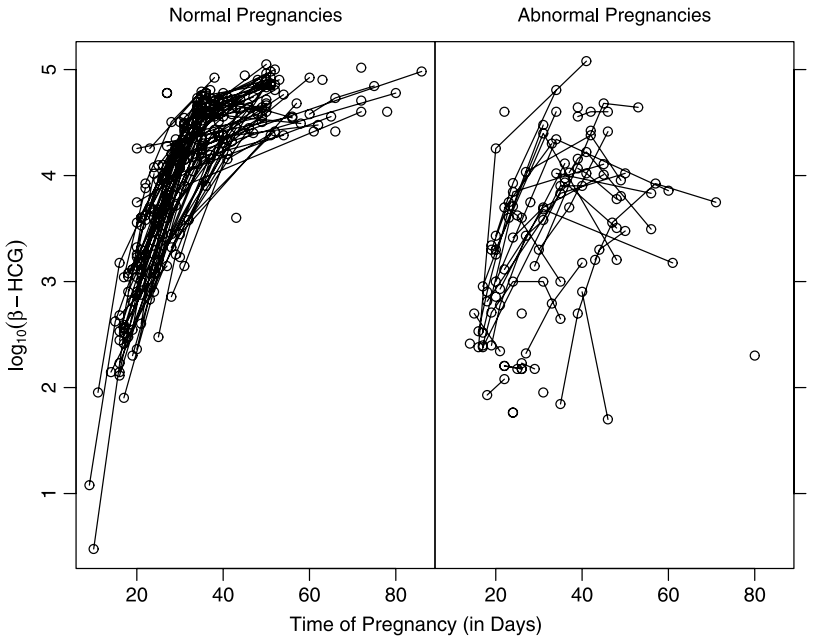

Fig. 1. Observed β-HCG time profiles in the log scale for women with normal and abnormal pregnancy outcomes.

during the early stages of pregnancy may affect its outcome. In this case we are interested in a binary outcome but in a general setting we may be dealing with any kind of response.

If we observed longitudinal measurements without noise on a dense grid of time points this problem could be addressed from a functional perspective by using a logistic functional regression model with functional predictor and scalar response [28,8] or, more generally, a generalized functional linear model [17,26]. However, this is an unrealistic setting in many biometrical applications in which the design for longitudinal data is irregular and sparse with very few observations available per individual and measurements are subject to experimental error. This is for instance the case in the β-HCG data set in which the number of observations per women varies from 1 to 6, with a median of 2.

Therefore, when dealing with noisy and highly sparse longitudinal trajectories a natural way of measuring their impact on the response of interest consists on extracting relevant latent information that could be used as covariates of a generalized linear model. Several authors have studied this problem focusing mainly on two types of response: binary outcomes and survival data. They propose joint models in which longitudinal data are fitted with a mixed-effects model whose random effects are covariates in a generalized linear model (GLM) for the response of interest. If we consider a linear mixed-effects (LME) model for the longitudinal trajectories, several solutions, including likelihood-based and Bayesian approaches, exist ([36,13,23,15,32]; see [27] for an overview).

However, when fitting nonlinear mixed-effects (NLME) models to the longitudinal data, an estimation approach that can efficiently deal with the complexity of the joint model is yet to be proposed. Note that the use of NLME models vs LME is necessary in some cases in which the evolution of the longitudinal trajectories over time is clearly nonlinear, as in the pregnancy data set that motivates this work. De la Cruz et al. [4] compared several estimation methods for the analysis of this data set with a NLME–GLM joint model, including the naive two-step approach, BLUP and likelihood approximation methods based on several numerical integration techniques. They verified that as in the LME–GLM joint model, the first two procedures yield biased estimates. The third method seemed to work better for some particular approximation techniques, namely Laplacian and adaptive Gaussian approximations. However, these methods can be computationally inefficient in practice. Wu et al. [38] also considered the problem of joint likelihood inference in the NLME–GLM model, although focusing on the case in which the primary outcome is the time to a given event, and not a binary response, and encountered similar implementation problems. Wu et al. [39] proposed a fast and accurate joint estimation procedure for that model relying on the Laplace approximation. However, considering the findings of [19] about the asymptotic bias of estimators based on Laplace approximation for GLM with discrete response, these authors acknowledged that the performance of their method might be less satisfactory when dealing with binary outcomes instead of survival data.

To overcome these drawbacks, in this article we consider a Bayesian estimation approach for the NLME–GLM joint model. The aim is to provide a framework for the implementation of an MCMC sampler in this kind of models. Indeed, we propose a prior structure that allows for calculating some of the conditional distributions explicitly, which facilitates the construction of a Metropolis-within-Gibbs algorithm. Although in its application to the pregnancy data set we focus in the prediction binary outcomes, the general estimation framework that we describe is flexible enough to be used with any kind of response of interest. Moreover, motivated by our real data set, we assume that we may have autocorrelated error terms in the NLME.

The rest of the paper is organized as follows. In Section 2 we present the detailed specifications of the proposed joint model. In Section 3 we describe the MCMC algorithm for Bayesian estimation. A model comparison strategy is discussed in Section 4.1 and in Section 5 we conduct a simulation study to assess the performance of our method and the importance of model misspecification in the presence of autocorrelated errors. In Section 6 we apply our method to the β-HCG data set, comparing the results to previous analyses on this data set. Finally, we offer a general discussion in Section 7.

2. Joint model

The structure of interest here can be described by two components. The first component contains repeated observed measurements that are assumed to follow a nonlinear mixed-effects model over possibly unequally spaced times. The second component contains the primary outcome, which is assumed to follow a generalized linear model where the random coefficients of the nonlinear mixed-effects models are used as covariates.

Denote by $y_{ij}, i = 1, \ldots, m,\ j = 1, \ldots, n_i$, the observation of a continuous response for individual i at time t_{ij}. Let $y_i = (y_{i1}, \ldots, y_{in_i})'$ be the observed vector of longitudinal measurement data at times $t_i = (t_{i1}, \ldots, t_{in_i})'$. Assume that y_i follows the nonlinear mixed-effects model

$$y_i = g(\alpha, X_i; t_i) + \epsilon_i, \quad i = 1, \ldots, m, \tag{1}$$

where α is a vector of p unknown fixed effects parameters, X_i is a vector of q unobservable random effects, g is a real-valued nonlinear function of the fixed and random effects, and $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})'$ is the within individual random error vector. We assume that the random effects X_i's are independent and identically normally distributed with mean vector μ_X and covariance matrix Σ_X. Typically, the error terms ϵ_i's are assumed to be normal with zero mean vector and covariance matrix $\Sigma_{\epsilon_i} = \sigma_\epsilon^2 I_{n_i}$, i.e. independent measurements errors, where I_a denotes the identity matrix of dimension a. However, in longitudinal data, measurements taken over time on individuals usually show a highly unbalanced structure (i.e. measurement times may be unequally spaced within an individual and may differ across individuals) and may be serially related. To take this into account we assume $\Sigma_{\epsilon_i} = \Sigma_{\epsilon_i}(\sigma_\epsilon^2, \rho)$, with σ_ϵ^2 being a scalar parameter and ρ a vector of parameters describing the correlation structure. Depending on the context, various assumptions about the matrix $\Sigma_{\epsilon_i}(\sigma_\epsilon^2, \rho)$ can be made (see [35, Chap. 7]). In the following we consider that $\Sigma_{\epsilon_i}(\sigma_\epsilon^2, \rho) = \sigma_\epsilon^2 \Sigma_i(\rho)$, where $\Sigma_i(\rho)$ is an $n_i \times n_i$ scaled matrix with (k_1, k_2)th element equal to $\rho^{|t_{ik_1} - t_{ik_2}|}$ though other choices are possible. This matrix has a continuous time first-order autoregressive, CAR(1), structure (see [5]), which can cope with nonequally spaced measurements. We also assume that the X_i's and ϵ_i's are mutually independent.

Now, assume that in addition to the n_i-dimensional vector of longitudinal measurements y_i, a primary response D_i, and a k-vector of observed covariates, W_i, are observed on the ith individual. We assume that the primary response and the random effects covariates are related via a GLM in canonical form; i.e., the conditional distribution of D_i given X_i (and W_i; conditioning on W_i is dropped throughout) is

$$f(D_i|X_i; \theta) = \exp\left\{\frac{D_i(\beta_0' W_i + \beta_1' X_i) - b(\beta_0' W_i + \beta_1' X_i)}{a(\phi)} + c(D_i, \phi)\right\}, \tag{2}$$

where $\theta = (\beta, \phi)'$, with $\beta = (\beta_0', \beta_1')$, are the parameters of primary interest; β_0 and β_1 are regression parameters, ϕ is a dispersion parameter and $a(\cdot)$, $b(\cdot)$, $c(\cdot, \cdot)$ are known functions. In our context, β_1 is of particular interest because it represents the relationship between the primary response and features of longitudinal profiles. As discussed in [36], we can further assume that y_i and D_i are conditionally independent given X_i, in which case

$$f(y_i, D_i, X_i) = f(y_i, D_i|X_i) f(X_i) = f(y_i|X_i) f(D_i|X_i) f(X_i).$$

The likelihood for the *joint* model (y_i, D_i) is given by

$$f(y, D) = \prod_{i=1}^{m} \int_{\mathbb{R}^q} f(y_i|X_i) f(D_i|X_i) f(X_i) d X_i, \tag{3}$$

where $y = (y_1, \ldots, y_m)$ and $D = (D_1, \ldots, D_m)$. Note that the *joint* model (y_i, D_i) is nonlinear in X_i, thus the integral in (3) does not have a closed-form expression. However, approximation methods can be used to help the estimation. De la Cruz et al. [4] discuss methods based on numerical integration techniques to obtain the MLE of the *joint model* in the special case for which the primary response is binary. In this paper we propose to estimate the model parameters using a Bayesian approach which is implemented using MCMC methods.

3. Estimation via MCMC methods

Bayesian fitting of the *joint model* described in Section 2 involves, as usual in the Bayesian framework, the updating from prior to posterior distributions for the parameters via appropriate likelihood functions. However, closed-form exact expressions for most of the relevant joint and marginal posterior distributions are not available. Instead, we rely here on

sampling-based approximations to the distributions of interest via Markov chain Monte Carlo (MCMC) methods: we use a Gibbs sampler or a Metropolis-within-Gibbs algorithm to explore the posterior.

We now consider the problem of choosing prior information for the parameters $\beta, \alpha, \mu_X, \Sigma_X, \sigma_\epsilon^2, \rho$, and ϕ of the *joint model*. The specification of the priors on the model parameters is often guided by mathematical and practical convenience. We assume prior independence for parameters and

$$\begin{aligned} &\alpha \sim N_p(a_1, A), \qquad \mu_X \sim N_q(c_1, C), \qquad \Sigma_X \sim IW(v, vV), \qquad \sigma_\epsilon^2 \sim IG(v_1, v_2), \\ &\rho \sim \pi(\rho), \qquad \beta \sim N_r(s, S), \quad \text{and} \quad \phi \sim \pi(\phi). \end{aligned} \tag{4}$$

Here $IG(h, k)$ denotes the inverse gamma distribution, with shape parameter h and scale parameter k, and mean $(h-1)^{-1}k^{-1}$. By $V \sim IW(d, D)$, we mean that the random matrix V follows an inverse Wishart distribution with scalar parameter d and matrix parameter D (by letting $V \sim IW(d, dD)$ we ensure that the mean of V^{-1} equals D^{-1}). Also, $N_p(\mu, \Sigma)$ represents the p-variate normal distribution with vector mean μ and covariance matrix Σ, and $\pi(\cdot)$ stands for a general prior distribution to be specified in each case, as we discuss below. For the variance components, Σ_X and σ_ϵ^2, we choose conjugate priors.

In (4) the hyperparameters $(a_1, A, c_1, C, v, V, v_1, v_2, s, S)$, and those involved in the prior for ρ and ϕ, are all assumed to be known and chosen so that the priors are proper. In practice the elicitation of hyperparameters may be difficult. Thus, the choice of the values of the hyperparameters can be based on strong prior knowledge or be chosen to reflect diffuse prior information.

Note that in (2), for binomial and Poisson primary responses, the dispersion parameter is $\phi = 1$. In that case no prior specification is required for ϕ in (4). For normal primary response, ϕ is σ^2, and we can follow common practice in choosing an inverse gamma prior, $IG(r_1, r_2)$, for σ^2, i.e. $\pi(\sigma^2) = IG(r_1, r_2)$. In (4) we assume a uniform prior for ρ.

We now present the posterior density associated with the joint model. We will note f_N, f_{IG}, f_U and f_{IW} the multivariate normal, inverse gamma, uniform and inverse Wishart densities, respectively. Furthermore, f_{GLM} denotes the primary response in the generalized linear model (2). The joint posterior density of $X, \beta, \alpha, \mu_X, \Sigma_X, \sigma_\epsilon^2, \rho$, and ϕ given the observed data $d_m = \{(y_i, D_i)\}_{i=1}^m$ is

$$\pi(X, \beta, \alpha, \mu_X, \Sigma_X, \sigma_\epsilon^2, \rho, \phi | d_m) = \frac{\pi^*(X, \beta, \alpha, \mu_X, \Sigma_X, \sigma_\epsilon^2, \rho, \phi; d_m)}{m^*(d_m)}, \tag{5}$$

where the unnormalized posterior density is

$$\begin{aligned} \pi^*(X, \beta, \alpha, \mu_X, \Sigma_X, \sigma_\epsilon^2, \rho, \phi; d_m) = &\left[\prod_{i=1}^m f_N(y_i; g(\alpha, X_i; t_i), \sigma_\epsilon^2 \Sigma_i(\rho)) f_{GLM}(D_i; X_i, \theta) f_N(X_i; \mu_X, \Sigma_X)\right] \\ &\times f_N(\alpha; a_1, A) f_{IG}(\sigma_\epsilon^2; v_1, v_2) f_U(\rho) f_N(\mu_X; c_1, C) f_{IW}(\Sigma_X; v, vV) \\ &\times f_N(\beta, s; S) \pi(\phi) \end{aligned}$$

and the normalizing constant (which is also the marginal density of the data) is

$$m^*(d_m) = \int \pi^*(X, \beta, \alpha, \mu_X, \Sigma_X, \sigma_\epsilon^2, \rho, \phi; d_m) dX\, d\beta\, d\alpha\, d\mu_X\, d\Sigma_X\, d\sigma_\epsilon^2\, d\rho\, d\phi.$$

The full conditionals to implement the MCMC procedure can be easily derived from (5). Indeed, we have

$$\pi(X | \text{rest}, d_m) = \prod_{i=1}^m \pi(X_i | \text{rest}, d_m), \tag{6}$$

$$\begin{aligned} \pi(\alpha | \text{rest}, d_m) &\propto \pi(\alpha) \prod_{i=1}^m f(y_i | X_i) \\ &\propto \exp\left\{-\frac{1}{2}\text{tr}\left(\frac{1}{\sigma_\epsilon^2}\Sigma^{-1}(\rho)(y_i - g(\alpha, X_i; t_i))'(y_i - g(\alpha, X_i; t_i)) + A^{-1}(\alpha - a)'(\alpha - a)\right)\right\}, \end{aligned} \tag{7}$$

$$\pi(\beta | \text{rest}, d_m) \propto \exp\left\{-\frac{1}{2}\text{tr}(S^{-1}(\beta - s)'(\beta - s)) + \sum_{i=1}^n \frac{D_i\theta_i - b(\theta_i)}{a(\phi)}\right\}, \tag{8}$$

$$\pi(\mu_X | \text{rest}, d_m) \propto \pi(\mu_X) \prod_{i=1}^m f(X_i), \tag{9}$$

$$\pi(\Sigma_X | \text{rest}, d_m) \propto \pi(\Sigma_X) \prod_{i=1}^m f(X_i), \tag{10}$$

$$\pi(\sigma_\epsilon^2|\text{rest}, d_m) \propto \pi(\sigma_\epsilon^2)\prod_{i=1}^{m} f(y_i|X_i), \tag{11}$$

$$\begin{aligned}\pi(\rho|\text{rest}, d_m) &\propto \pi(\rho)\prod_{i=1}^{m} f(y_i|X_i)\\ &\propto \exp\left\{-\frac{1}{2}\text{tr}\left(\frac{1}{\sigma_\epsilon^2}\Sigma^{-1}(\rho)(y_i - g(\alpha, X_i; t_i))'(y_i - g(\alpha, X_i; t_i))\right)\right\},\end{aligned} \tag{12}$$

$$\pi(\phi|\text{rest}, d_m) \propto \pi(\phi)\prod_{i=1}^{m} f(D_i|X_i), \tag{13}$$

where $\theta_i = \beta_0' W_i + \beta_1' X_i$ and *rest* denotes the remaining components of the model to which we are conditioning in each case. Some of these densities have a closed-form expression. Indeed, from (9)–(11) it is easy to check that $\mu_X|\text{rest}, d_m$ is multivariate normal with mean

$$(m\Sigma_X^{-1} + C^{-1})^{-1}\left(\Sigma_X^{-1}\sum_{i=1}^{m} X_i + c_1 C^{-1}\right)$$

and covariance matrix $(m\Sigma_X^{-1} + C^{-1})^{-1}$. Also, $\Sigma_X|\text{rest}, d_m$ follows an inverse Wishart distribution with scale parameter $v + \sum_{i=1}^{m} n_i$ and matrix parameter

$$vV + \sum_{i=1}^{m}(y_i - g(\alpha, X_i; t_i))'(y_i - g(\alpha, X_i; t_i)).$$

Finally, $\sigma_\epsilon^2|\text{rest}, d_m$ follows an inverse gamma distribution with shape parameter $N/2 + v_1$ and scale parameter

$$\left(\frac{1}{v_2} + \frac{\sum_{i=1}^{m} RSS_{y_i}}{2}\right)^{-1},$$

where $RSS_{y_i} = (y_i - g(\alpha, X_i; t_i))'\Sigma_i^{-1}(\rho)(y_i - g(\alpha, X_i; t_i))$. Due to the fact that $g(\cdot)$ is a nonlinear function of X_i, the full conditional density in (6), $\pi(X|\text{rest}, d_m)$, cannot be written explicitly. However, the full conditional density of X_i can be written, up to a proportionality constant, as

$$\begin{aligned}&\exp\left\{-\frac{1}{2}\text{tr}\left(\frac{1}{\sigma_\epsilon^2}\Sigma^{-1}(\rho)(y_i - g(\alpha, X_i; t_i))'(y_i - g(\alpha, X_i; t_i)) + \Sigma_X^{-1}(X_i - \mu_X)'(X_i - \mu_X)\right)\right.\\ &\quad\left. + \frac{D_i(\beta_0' W_i + \beta_1' X_i) - b(\beta_0' W_i + \beta_1' X_i)}{a(\phi)}\right\}.\end{aligned} \tag{14}$$

In this case, to simulate from this full conditional we use a Metropolis–Hastings algorithm within each Gibbs step. Because (14) is known up to a normalization constant, we can compute its mode $X_i^\star$ and Hessian $V_i^\star$ using numerical optimization techniques. This yields a natural choice of the proposal distribution, a multivariate normal distribution with mean vector $X_i^\star$ and variance–covariance matrix $cV_i^{\star -1}$, denoted by $f_N(X_i; X_i^\star, cV_i^{\star -1})$, where c is a known suitable tuning parameter chosen to get sure that the acceptance rate is satisfactory (typically between 0.2 and 0.5), see [10]. Then we can implement the Metropolis–Hastings algorithm as follows. Denote $X_i^{(r)}$ the current value of X_i at the rth iteration. A new candidate value X_i^{new} is drawn from the proposal distribution $f_N(X_i; X_i^\star, cV_i^{\star -1})$. The acceptance probability is computed as:

$$\min\left\{1, \frac{f_N(X_i^{\text{new}}; X_i^\star, cV_i^{\star -1})}{f_N(X_i^{(r)}; X_i^\star, cV_i^{\star -1})}\frac{f_N(y_i; g(\alpha, X_i^{\text{new}}; t_i), \sigma_\varepsilon^2\Sigma_i(\rho))f_{GLM}(D_i; X_i^{\text{new}}, \theta)f_N(X_i^{\text{new}}; \mu_X, \Sigma_X)}{f_N(y_i; g(\alpha, X_i^{(r)}; t_i), \sigma_\varepsilon^2\Sigma_i(\rho))f_{GLM}(D_i; X_i^{(r)}, \theta)f_N(X_i^{(r)}; \mu_X, \Sigma_X)}\right\}.$$

Note that there is no need to compute the normalization constant because it cancels out in the acceptance probability. For the remaining full conditionals, no such closed-form expression exists either and the same Metropolis–Hastings within Gibbs algorithm is used to obtain draws from them. Note that the full conditional of the dispersion parameter ϕ of the GLM is only required depending on the kind of the primary response. For instance, for the binomial and Poisson model we have $\phi = 1$.

The Markov chain associated with the MCMC algorithm is denoted by $\Phi\{(X^{(n)}, \alpha^{(n)}, \beta^{(n)}, \mu_X^{(n)}, \Sigma_X^{(n)}, \sigma_\epsilon^{2(n)}, \rho^{(n)}, \phi^{(n)})\}_{n=0}^{\infty}$ and has the posterior density (5) as its stationary density. To run the algorithm, given the current state, $(X^{(n)}, \alpha^{(n)}, \beta^{(n)}, \mu_X^{(n)}, \Sigma_X^{(n)}, \sigma_\epsilon^{2(n)}, \rho^{(n)}, \phi^{(n)})$, we draw each of the $X_i^{(n+1)}$'s independently and form $X^{(n+1)}$. Then, the following series of steps is conducted: given $X^{(n+1)}$, $\sigma_\epsilon^{2(n)}$ and $\rho^{(n)}$, we draw $\alpha^{(n+1)}$; given $X^{(n+1)}$ and $\phi^{(n)}$ we draw $\beta^{(n+1)}$; given $X^{(n+1)}$ and $\Sigma_X^{(n)}$ we draw $\mu_X^{(n+1)}$; given $X^{(n+1)}$, $\alpha^{(n+1)}$ and $\mu_X^{(n+1)}$ we draw $\Sigma_X^{(n+1)}$; given $X^{(n+1)}$, $\alpha^{(n+1)}$ and $\rho^{(n)}$ we draw $\sigma_\epsilon^{2(n+1)}$; given $X^{(n+1)}$ and $\alpha^{(n+1)}$ we draw $\rho^{(n+1)}$; and finally, given $X^{(n+1)}$ and $\beta^{(n+1)}$ we draw $\phi^{(n+1)}$.

4. Model comparison

4.1. Conditional predictive ordinate

The conditional predictive ordinate (CPO) statistics introduced by [9] is a popular and useful model assessment tool based on the marginal posterior predictive density of the response for individual i given the observed data from the rest of the individuals. Let $\theta = (\beta, \alpha, \mu_X, \Sigma_X, \sigma_\epsilon^2, \rho, \phi)$ be the parameters of the *joint model*, let d_m be the observed data for all individuals, and let $d_{-(i)}$ and $X_{-(i)}$ denote the observed data and random-effects vector, respectively, of the whole sample excluding individual i. Further, let us note $d_i = (y_i, D_i)$ where, for individual i, y_i is the observed vector of longitudinal measurements and D_i is the primary response of the GLM. Then, the *CPO* statistic for individual i for our *joint model* is defined as

$$\begin{aligned} CPO_i &= f(d_i|d_{-(i)}) \\ &= \left[\frac{f(d_{-(i)})}{f(d_m)}\right]^{-1} \\ &= \left[E_{\theta,X|d_m}\left(\frac{1}{f(y_i|X_i,\theta)f(D_i|X_i,\theta)f(X_i|\theta)}\right)\right]^{-1}. \end{aligned}$$

A Monte Carlo estimate of CPO_i can be obtained by using a single MCMC sample from the posterior distribution $\pi(X, \theta|d_m)$. Let $(\theta^{(1)}, X_i^{(1)}), \ldots, (\theta^{(R)}, X_i^{(R)})$ be a sample of size R, for corresponding parameters and individual-specific random effect, drawn from $\pi(\theta, X|d_m)$ after the burn-in phase. A natural Monte Carlo approximation of CPO_i is given by

$$\widehat{CPO_i} \approx \left[\frac{1}{R}\sum_{r=1}^{R}\frac{1}{f(y_i|X_i^{(r)},\theta)f(D_i|X_i^{(r)},\theta^{(r)})f(X_i^{(r)}|\theta^{(r)})}\right]^{-1}.$$

For each individual, larger values of CPO imply a better fit of the model. As a summary statistic of CPO over all individuals, we use the logarithm of the pseudomarginal likelihood (LPML; [16]), which is defined by

$$LPML = \frac{1}{m}\sum_{i=1}^{m}\log\left(\widehat{CPO_i}\right). \tag{15}$$

LPML measure focuses on both prediction of the primary response and longitudinal trajectory. Models with greater LPML values represent a prediction accuracy. The LPML is well defined under the posterior predictive density and it is computationally stable. LPML has been extensively used in Bayesian analysis for model selection in situations of simpler and more complicated models and has a long history in statistics literature (see [3,1,2,12]). Model comparison can also be performed using summary measures. Suppose that we have two models $\mathcal{M}_1$ and $\mathcal{M}_2$ under consideration. The pseudo-Bayes factor (PsBF), a surrogate for the Bayes factor [18,21] based on the CPO for comparing the two models, is defined as

$$PsBF(\mathcal{M}_1, \mathcal{M}_2) = \prod_{i=1}^{m}\frac{\widehat{CPO_i}(\mathcal{M}_1)}{\widehat{CPO_i}(\mathcal{M}_2)} = \exp\{LPML(\mathcal{M}_1) - LPML(\mathcal{M}_2)\}$$

where $\widehat{CPO_i}(\mathcal{M}_\ell)$ and $LPML(\mathcal{M}_\ell)$, with $\ell = 1, 2$, correspond, respectively, to the CPO and LPML estimated for model ℓ.

4.2. Testing for zero correlation

In model (1)–(2), it may be interesting to test for zero correlation of the errors. Let $\mathcal{M}_0$ and $\mathcal{M}_1$ denote the restrictions that $\rho = 0$ and $\rho > 0$, respectively. Following [25], we can apply the Savage–Dickey method to compute the Bayes factor associated to this restriction:

$$B_{01} = \frac{\pi(\rho = 0|d_m; \theta)}{\pi(\rho = 0)} \tag{16}$$

where $\pi(\rho = 0|d_m; \theta)$ represents the marginal posterior density function for ρ at 0, and $\pi(\rho = 0)$ represents the marginal prior density for ρ at 0 within the unrestricted model, i.e. with correlated errors.

In order to compute the Savage–Dickey density ratio, the posterior density must be estimated at the desired restriction using approximate samples from the posterior distribution of the unrestricted model. Following [22], we can use the output from our Gibbs sampler and the prior to estimate the required ratio

$$B_{01} = \frac{1/R\sum_{r=1}^{R}\pi(\rho = 0|X^{(r)}, \beta^{(r)}, \alpha^{(r)}, \mu_X^{(r)}, \Sigma_X^{(r)}, \sigma_\epsilon^{2(r)}, \phi^{(r)}, d_m)}{\pi(\rho = 0)}$$

where $r = 1, \ldots, R$ indicates the number of the (post burn-in) Gibbs sampler iteration and superscript (r) denotes the replications themselves. Nevertheless, when the full conditional posterior density does not have a closed form, an approximation strategy is required. In our context, following [37], we may use a normal approximation to the posterior distribution. The mean and variance of the posterior density of ρ are computed from the MCMC results. Another alternative method consists on generating a marginal posterior sample for ρ, fitting a density (e.g. kernel density estimator) to the sample, and evaluating it at the point of interest (in our case $\rho = 0$). This might be faster and more accurate than the normal approximation to the posterior distribution. Note also that since ρ is compactly supported one has to be careful about the use of normal approximations.

5. Simulation study

In order to assess the performance of our estimation method, we conducted a simulation study with highly sparse and unbalanced synthetic data. Indeed, we used the joint model (17)–(18) to simulate observations that replicate the structure of the real data set analysed in Section 6, keeping the same number of individuals in each group and for each individual, the same number of observations as well as the same observation time points. We simulated 500 data sets using the following parameter values:

$$\mu_X = 4$$
$$\alpha = (\alpha_q, q = 1, 2) = (15, 7)$$
$$\beta = (\beta_h, h = 1, 2) = (-22, 5)$$
$$\sigma_X^2 = 0.2$$
$$\sigma_\epsilon^2 = 0.2$$
$$\rho = 0.9.$$

To assess the importance of considering correlation in the error term of the NLME on synthetic data, the generated data sets were analysed using the estimation procedure presented in Section 3 assuming both correlated and independent error terms (i.e. $\Sigma_{\epsilon_i}(\sigma_\epsilon^2, \rho) = \sigma_\epsilon^2 I_{n_i}$). This strategy allows us to analyse the bias introduced by the misspecified model which does not consider the correlation structure of the data.

Implementing Gibbs sampling requires adopting specific values for the hyperparameters $(a_1, A, c_1, C, v, V, v_1, v_2, s, S)$. The values for the hyperparameters were taken as follows: $a_1 = s = (0, 0), A = S = 1\,000 I_2, c_1 = 0, C = 1\,000, v = 6, V = 0.00083, v_1 = 3$ and $v_2 = 0.01$.

We performed 2 000 000 iterations of the MCMC procedure. After the first 10 000 iterations, samples were collected, at a spacing of 50 iterations, to obtain approximately independent samples. We ended up with $R = 39\,800$ samples to calculate posterior quantities of interest. Summary statistics for the Bayesian estimates obtained for these 500 simulated data sets are given in Table 1. The true values of the parameters used in the simulation, the means and the medians with their respective standard errors, and individual coverage probability are provided. It can be seen that the mean and median values for the logistic submodel parameters present important biases. Specifically, when we use the misspecified model, we observe an important overestimation for β_1 and an underestimation for β_2. Instead, as expected, we get much better results when we consider correlated errors. On the contrary, the nonlinear model parameter estimates are very close to the simulated values for both models. We can observe the same behaviour in terms of coverage probabilities. Figs. 2 and 3 provide a graphical representation of these results displaying the distribution of estimates of the longitudinal and logistic submodel parameters. This simulation study shows that not taking into account correlation among errors in the longitudinal measurements of the joint model may introduce large bias in GLM parameter estimates.

6. Analysis of pregnant women data

The main objective of the analysis of the pregnant women data set presented in Section 1 is to investigate the effects of the β-HCG longitudinal process on pregnancy outcomes, and in particular the association between normal pregnancy and features of longitudinal β-HCG profiles. The data were collected from a total of 173 young pregnant women over a period of 2 years in a private fertilization obstetrics clinic in Santiago, Chile. The resulting data set consists of 124 patients whose pregnancies developed without any complications and 49 patients with abnormal pregnancies. Let $D_i = 1$ and 0 denote normal and abnormal pregnancy outcomes, respectively, for woman i, $i = 1, \ldots, m$, $(m = 173)$. For the longitudinal β-HCG concentrations, the 173 women altogether contribute a total of 375 observations, where the number of observations n_i per woman ranges from 1 to 6 (median 2). Approximately 30% of the 173 women have only one β-HCG measurement, 31% have two, 33% have three, and only 6% have four or more measurements.

As discussed in previous work [24,6,7], a reasonable representation of the log β-HCG profile (y_i) for the ith woman is

$$y_i = \frac{X_i}{1 + \exp\{-(t_i - \alpha_1)/\alpha_2\}} + \epsilon_i \tag{17}$$

Table 1
Results obtained on 500 simulated data sets for a joint model with independent and correlated errors.

	True value	Mean	$SD_{\bar{X}}$	Median	SD_{Median}	Coverage Prob.
Independent errors						
Longitudinal submodel						
μ_X	4.00	3.998	0.067	3.998	0.065	0.95
α_1	15	14.86	1.239	14.88	0.546	0.91
α_2	7	7.137	0.738	7.095	0.698	0.91
σ^2_ϵ	0.2	0.147	0.022	0.144	0.015	0.12
σ^2_X	0.2	0.285	0.044	0.282	0.044	0.44
ρ	0.9	–	–	–	–	–
Logistic submodel						
β_1	−22	−13.06	3.330	−12.823	3.067	0.31
β_2	5	2.861	0.797	2.805	0.725	0.30
Correlated errors						
Longitudinal submodel						
μ_X	4.00	4.004	0.067	4.003	0.066	0.938
α_1	15	14.90	0.519	14.92	0.515	0.942
α_2	7	7.163	0.639	7.130	0.633	0.930
σ^2_ϵ	0.2	0.186	0.033	0.177	0.032	0.850
σ^2_X	0.2	0.229	0.050	0.235	0.052	0.896
ρ	0.9	0.853	0.049	0.862	0.041	0.850
Logistic submodel						
β_1	−22	−19.28	3.578	−19.561	3.043	0.996
β_2	5	5.15	0.935	5.218	0.797	0.998

where time is measured in days and the measurement errors ϵ_i are Gaussian. Indeed, levels of β-HCG in the log scale clearly evolve nonlinearly with time and present a threshold after 50 days of pregnancy, which is well captured by the horizontal asymptote of this function (see Fig. 1). For this data set, it seems reasonable to consider the error distribution $\epsilon_i \sim N_{n_i}(0, \Sigma_{\epsilon_i}(\sigma^2_\epsilon, \rho))$ where $\Sigma_\epsilon(\sigma^2_\epsilon, \rho)$ is a correlation structure with unknown σ^2_ϵ and ρ parameters. In particular, we consider the CAR(1) correlation structure described in Section 2. The woman-specific random effect X_i is assumed to satisfy $X_i \sim N(\mu_X, \sigma^2_X)$ and it represents the asymptotic behaviour of the log β-HCG profile. To describe the relation between the pregnancy outcome and X_i, we consider the primary logistic regression model

$$\Pr(D_i = 1|X_i) = [1 + \exp\{-(\beta_1 + \beta_2 X_i)\}]^{-1}. \tag{18}$$

We used the Bayesian approach described in Section 3 to estimate the parameters of this joint model. To illustrate the gain obtained by considering correlated errors, we also fitted the same joint model with independent errors in (17). We also considered separate fitting, i.e. we estimated independently the NLME (17) and the GLM (18), assuming both independent and correlated errors.

Implementing Gibbs sampling requires adopting specific values for the hyperparameters $(a_1, A, c_1, C, v, V, v_1, v_2, s, S)$. The values for the hyperparameters were taken as follows: $a_1 = s = (0, 0), A = S = 1\,000 I_2, c_1 = 0, C = 1\,000, v = 6, V = 0.00083, v_1 = 3$ and $v_2 = 0.01$. We also performed the analysis with different hyperparameter values, obtaining very similar results. This suggests robustness to the hyperparameter choices. Always, the choice of the hyperparameter values was made to use diffuse proper priors. For the standard deviations σ_ϵ and σ_X, other prior formulations were also considered, including uniform and half-Cauchy distributions, with no substantive effect on the posterior distributions for this data set.

We performed 2 000 000 iterations of the MCMC procedure. After the first 10 000 iterations, samples were collected, at a spacing of 50 iterations, to obtain approximately independent samples. We ended up with $R = 39\,800$ samples to calculate posterior quantities of interest. In the implementation of the Metropolis–Hastings step, we obtained an acceptance rate around 0.30. The program used to fit the model was written in Fortran (and it is available under request), but let us point out that the model for the i.i.d. case can be fitted in OpenBUGS. To diagnose convergence, we suggest any of the convergence criteria discussed in the literature, for example, those included in the BOA package [33]. We prefer to use diagnostics which do not require multiple parallel chains, as proposed by [11]. This criterion consists on testing the equality of the means of the first (10%) and last parts (50%) of the chain. The test statistic is simply the difference between the two sample means divided by the estimated standard error (adjusted for autocorrelation), which, under the null hypothesis that both parts of the sample are drawn from the stationary distribution of the chain, has an asymptotically standard normal distribution. In this analysis, applying Geweke's convergence criterion separately to each model parameter, where the absolute value of the Z statistic was less than 1.6 in all cases, showed that convergence had been achieved.

Table 2 presents the results obtained by fitting the joint model (17)–(18) by the procedure described in this article and also the estimates provided by MCMC methods for the separate fitting. For both strategies, we considered independent and correlated errors for the NLME model. For each parameter and each model, the posterior mean, the standard error and the posterior median together with a 95% credible interval are given.

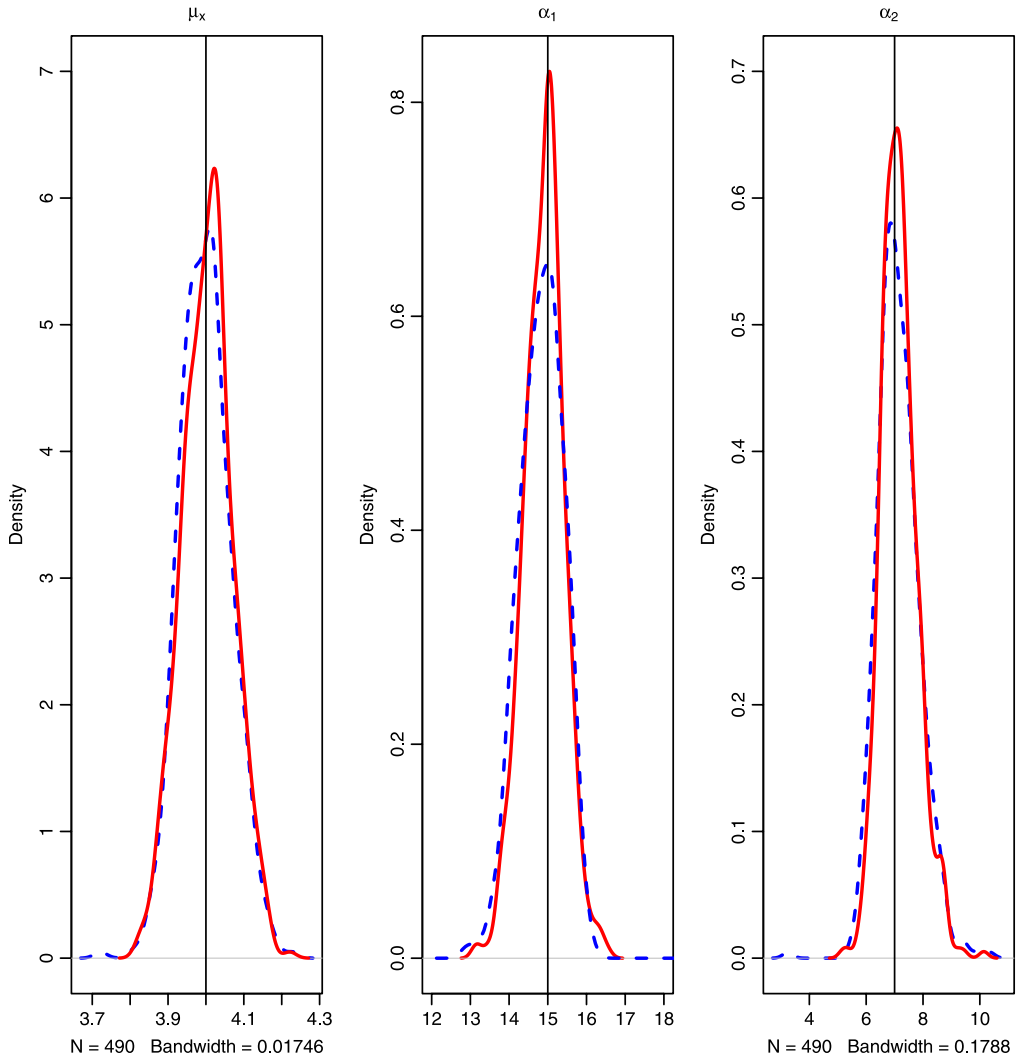

Fig. 2. Longitudinal submodel: Distribution of fixed effects parameter estimates over 500 simulated data sets using a joint model with independent (dashed line) and correlated (solid line) errors. Vertical lines represent true values.

From Table 2, we can see that there are no important differences between the parameter estimates obtained from joint and separate fitting under the assumption of independent errors. However, if we assume correlation in the error term, we obtain, as expected, a significant difference in the GLM parameter estimates β_1 and β_2 obtained from joint and separate fitting. If we now compare the estimates obtained when fitting a joint model with correlated and independent errors, we observe that the estimate of β_1 with correlated errors decreases in almost 50% in comparison with the estimate obtained under the independent error assumption whereas for β_2 we observe an increase of almost 50%. This is consistent with the estimation biases detected for these parameters in the simulation study of Section 5.

Now, from the estimated parameter values we get estimates of $P(D_i|X_i)$, which allows us to consider the underlying classification problem and compare the four model performances. To do so, we calculated the confusion matrix of classification which contains information about correspondence between actual and predicted classes. A probability cut-off value of 0.5 was considered as classification rule. The results are presented in Table 3 and are summarized in the first three columns of Table 4. The best classification is obtained with the joint model with correlated errors for which all individuals with normal pregnancy outcomes are correctly classified and only 13 individuals with abnormal pregnancy outcomes are incorrectly classified.

Table 4 shows the error rate, the sensitivity, and the specificity of the classification rule with a probability cut-off value of 0.5 for the four models. It also presents the area under the Receiver Operating Characteristic (ROC) curve (AUC) and its standard deviation. The ROC curve represents the sensitivity versus 1 minus the specificity for any cut-off value from 0 to 1. Then, a larger value of AUC means a better classifying performance. In the case of independent errors, we found an error rate estimation of approximately 13.3% and 17.3% for the joint and separate models respectively. As discussed before by [4], the

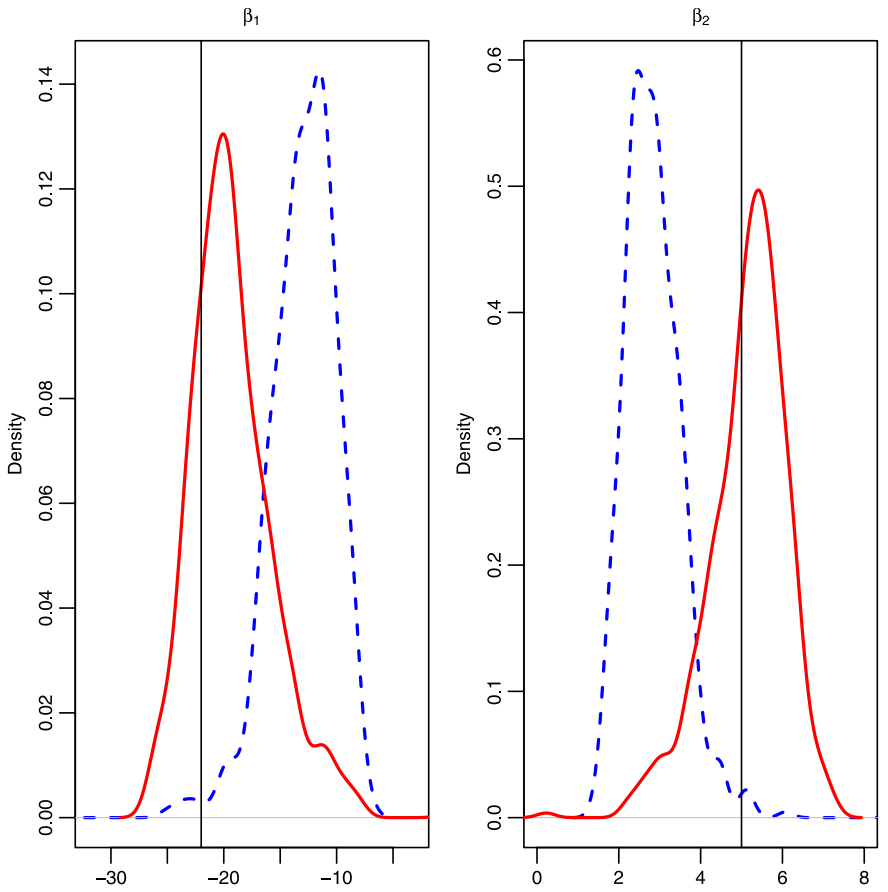

Fig. 3. Logistic submodel: Distribution of fixed effects parameter estimates over 500 simulated data sets using a joint model with independent (dashed line) and correlated (solid line) errors. Vertical lines represent true values.

Table 2
Parameter estimates for joint and separate modelling.

	Joint model					Separate model				
	Mean	SD	2.5%	Median	97.5%	Mean	SD	2.5%	Median	97.5%
Independent errors										
	Longitudinal submodel									
μ_X	4.495	0.063	4.375	4.494	4.620	4.513	0.065	4.388	4.512	4.643
α_1	14.850	0.400	14.040	14.870	15.590	15.000	0.392	14.190	15.020	15.740
α_2	7.467	0.520	6.510	7.446	8.551	7.482	0.527	6.515	7.461	8.581
σ^2_ϵ	0.132	0.014	0.108	0.131	0.161	0.131	0.014	0.107	0.130	0.161
σ^2_X	0.290	0.045	0.211	0.287	0.388	0.294	0.047	0.212	0.291	0.395
	Logistic submodel									
β_1	−15.280	3.957	−24.340	−14.850	−8.788	−14.460	2.868	−20.450	−14.320	−9.224
β_2	3.682	0.902	2.204	3.576	5.737	3.443	0.638	2.279	3.413	4.777
Correlated errors										
	Longitudinal submodel									
μ_X	4.495	0.063	4.373	4.494	4.621	4.521	0.064	4.399	4.519	4.649
α_1	15.180	0.409	14.340	15.190	15.940	15.330	0.433	14.460	15.340	16.160
α_2	7.211	0.487	6.311	7.193	8.228	7.278	0.504	6.361	7.256	8.331
σ^2_ϵ	0.187	0.025	0.143	0.185	0.240	0.250	0.053	0.162	0.245	0.359
σ^2_X	0.223	0.046	0.141	0.220	0.322	0.127	0.075	0.003	0.128	0.275
ρ	0.924	0.017	0.884	0.927	0.951	0.944	0.017	0.903	0.947	0.968
	Logistic submodel									
β_1	−22.860	5.474	−34.790	−22.400	−13.400	−39.040	6.965	−53.530	−38.730	−26.450
β_2	5.431	1.259	3.261	5.325	8.174	8.885	1.546	6.088	8.815	12.110

Table 3
Confusion matrix of classification for the joint and separate fitting with independent and correlated errors. Rows indicate the actual groups whereas columns represent the classification into groups obtained with each model.

Group	Joint model		Separate Model		Total
	Normal	Abnormal	Normal	Abnormal	
	Independent errors				
Normal	122	2	120	4	124
Abnormal	21	28	26	23	49
Total	143	30	146	27	173
	Correlated errors				
Normal	124	0	119	5	124
Abnormal	13	36	21	28	49
Total	137	36	140	33	173

Table 4
Error-rate, sensitivity, specificity and area under curve (AUC) for joint and separate models. In parenthesis, the standard deviation of AUC.

		Error-rate	Sensitivity	Specificity	AUC (s.d.)
Joint model:	Errors				
	Independent	13.3%	98.4%	71.8%	0.908 (0.032)
	Correlated	7.5%	100%	73.5%	0.988 (0.007)
Separate model:	Errors				
	Independent	17.3%	96.8%	46.9%	0.792 (0.046)
	Correlated	15.0%	96.0%	57.1%	0.815 (0.044)

joint model seems to improve classification. Now, considering a CAR correlation structure in the errors, we obtained an error rate estimation of approximately 7.5% and 15.0% for the joint and separate models, respectively. Therefore, it is clear that the inclusion of correlation structure allows to significantly improve the classification results in this data set. We observe the same kind of improvement for the sensitivity, the specificity and the AUC for the joint correlated model versus the other three models. It then appears evident that the joint strategy with correlation structure in the error term globally improves the sensitivity and the specificity for predicting a normal pregnancy outcome for this population of women.

To further compare the two joint models, this time in terms of predictive performance, we calculated for each one its *LPML*, as defined in Section 4.1 (see Eq. (15)). Models with greater *LPML* values will indicate a prediction accuracy. We found $LPML = -321.03$ for the joint model with correlated errors and $LPML = -350.26$ for the joint model assuming independent errors. This suggests that the joint model with a correlation structure in the errors provides a marginally better prediction performance to this specific data set. The $2 \times \log_{10}(\text{PsBF})$ for the independent errors model versus the correlated errors model was -25.39. Therefore, the improvement on fitting the data of the joint model with correlated errors over the joint model with independent errors is noticeable. Additionally, let us point out that in accordance with the credible interval for ρ in Table 2, the Savage–Dickey density ratio is close to zero, providing strong evidence against the hypothesis of independent errors. This ratio was evaluated following the two strategies described in Section 4.2 obtaining similar results.

We compare our results with those found using the Bayesian longitudinal discriminant analysis (BLDA) approach (see [6]) in which case the reported error rate was approximately 16% which is greater than under the joint model with correlated errors, 7.5%. The same happens with the sensitivity and the specificity: with the BLDA approach the sensitivity was found to be 95% and the specificity 57%.

7. Discussion

In this paper we have proposed inferential strategies for a generalized linear model for a primary outcome with covariates that are underlying individual-specific random effects in a nonlinear random effects model for longitudinal data, considering correlated errors in the NLME. We use an MCMC procedure to jointly estimate all parameters in the model. The proposed approach provides a general framework for estimation in joint NLME–GLM models that circumvents some problems related with likelihood approximations.

In the analysis of the pregnancy data set that motivates this work, we only use as the covariate for the logistic regression model the latent random effects of β-HCG profiles, but other covariates, such as age, number of previous normal and abnormal pregnancies and smoking status, could be useful for targeting specific individuals in future analysis. In our particular data set, however, a number of women had missing values for many of these covariates.

All the proposed estimators assume normality of random effects and within-individual errors. The latter is often reasonable, perhaps on a transformed scale. However, some authors (e.g., [34] among others), have shown that violation of this assumption can compromise inference in mixed-effects models, which raises similar concerns for the proposed joint

model. Further research on methods that go beyond traditional normality assumption on random effects would be useful. These topics are the subject of current research to be reported elsewhere.

Although in the examples studied in Sections 5 and 6 our MCMC sampler is fast and does not exhibit convergence problems, in some cases it may be difficult to choose the hyperparameter values in the Metropolis step. Thus, an interesting future work is to consider an adaptive Metropolis step (see for example [29]) for the joint model proposed in this paper in order to automatize the choice of the parameters of the proposal distribution. We are afraid that this method could be slower than our MCMC procedure.

Another important issue in the field of MCMC procedures is about the theoretical convergence of the MCMC sampler. In the context of Bayesian linear mixed-effects model, there have been several studies on the convergence properties of the Gibbs sampler (see [14,20,30,31]) that have resulted in easily-checked sufficient conditions for geometric ergodicity of the underlying Markov chain. However, our joint model turns out to be more complex since some full conditional distributions do not have a closed form, so providing similar results might be the object of future work.

Acknowledgments

The authors are grateful to two anonymous referees for their insightful comments and valuable suggestions, which led to substantial improvements in the presentation of this work. We also thank to Guillermo Marshall for facilitating us the β-HCG data set. Rolando de la Cruz thanks Fondo Nacional de Desarrollo Científico y Tecnológico — FONDECYT, CONICYT, Chile, for partially supporting this research under grant 1120739, and Programa de Investigación Asociativa — PIA, CONICYT, for partially supporting this research under grant ANILLOS ACT–87. Cristian Meza was supported by project FONDECYT 1141256 and grant ANILLOS ACT–1112, PIA, CONICYT, Chile. Ana Arribas-Gil was supported by projects MTM2010-17323 and ECO2011-25706, Spain. Carroll's research was supported by a grant from the National Cancer Institute (U01-CA057030).

References

[1] E.R. Brown, J.G. Ibrahim, Bayesian approaches to joint-cure rate and longitudinal models with application to cancer vaccine trials, Biometrics 59 (2003) 686–693.
[2] E.R. Brown, J.G. Ibrahim, V. DeGruttola, A flexible B-spline model for multiple longitudinal biomakers and survival, Biometrics 61 (2005) 64–73.
[3] M.H. Chen, Q.M. Shao, J.G. Ibrahim, Monte Carlo Methods in Bayesian Computation, Springer-Verlag, New York, 2000.
[4] R. De la Cruz, G. Marshall, F.A. Quintana, Logistic regression when covariates are random effects from a nonlinear mixed model, Biom. J. 53 (2011) 735–749.
[5] R. De la Cruz-Mesía, G. Marshall, Non-linear random effects models with continuous time autoregressive errors: a Bayesian approach, Stat. Med. 25 (2006) 1471–1484.
[6] R. De la Cruz-Mesía, F.A. Quintana, A model-based approach to Bayesian classification with applications to predicting pregnancy outcomes from longitudinal β-hCG profiles, Biostatistics 8 (2007) 228–238.
[7] R. De la Cruz-Mesía, F.A. Quintana, P. Müller, Semiparametric Bayesian classification with longitudinal markers, J. R. Stat. Soc. Ser. C 56 (2007) 119–137.
[8] M. Escabias, A.M. Aguilera, M.J. Valderrama, Principal component estimation of functional logistic regression: discussion of two different approaches, Nonparametr. Stat. 16 (3–4) (2004) 365–384.
[9] A.E. Gelfand, D.K. Dey, H. Chang, Model determination using predictive distributions with implementation via sampling-based methods (with discussion), in: J.M. Bernardo, J.O. Berger, A.P. Dawid, A.F.M. Smith (Eds.), Bayesian Statistics 4, Oxford University Press, Oxford, 1992, pp. 147–167.
[10] A. Gelman, G.O. Roberts, W.R. Gilks, Efficient metropolis jumping rules, in: J.M. Bernardo, J.O. Berger, A.P. Dawid, A.F.M. Smith (Eds.), Bayesian Statistics 5, Oxford University Press, Oxford, 1996, pp. 599–607.
[11] J. Geweke, Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, in: J.M. Bernardo, J.O. Berger, A.P. Dawid, A.F.M. Smith (Eds.), Bayesian Statistics 4, Oxford University Press, Oxford, 1992, pp. 169–194.
[12] P. Ghosh, W. Tu, Assessing sexual attitudes and behaviors of young women: a joint model with nonlinear time effects, time varying covariates, and dropouts, J. Amer. Statist. Assoc. 104 (2009) 474–485.
[13] X. Guo, B.P. Carlin, Separate and joint modeling of longitudinal and event time data using standard computer packages, Amer. Statist. 58 (2004) 1–9.
[14] J.P. Hobert, C.J. Geyer, Geometric ergodicity of Gibbs and block Gibbs samplers for a hierarchical random effects model, J. Multivariate Anal. 67 (1998) 414–430.
[15] J. Horrocks, M.J. van Den Heuvel, Prediction of pregnancy: A joint model for longitudinal and binary data, Bayesian Anal. 4 (2009) 523–538.
[16] J. Ibrahim, M. Chen, D. Sinha, Bayesian Survival Analysis, Springer-Verlag, New-York, 2001.
[17] G.M. James, Generalized linear models with functional predictors, J. R. Stat. Soc. Ser. B 64 (2002) 411–432.
[18] H. Jeffreys, Theory of Probability, third ed., Oxford University Press, New-York, 2011.
[19] H. Joe, Accuracy of Laplace approximation for discrete response mixed models, Comput. Statist. Data Anal. 52 (2008) 5066–5074.
[20] A.A. Johnson, G.L. Jones, Gibbs sampling for a Bayesian hierarchical general linear model, Electron. J. Stat. 4 (2010) 313–333.
[21] R. Kass, A. Raftery, Bayes factors, J. Amer. Statist. Assoc. 90 (1995) 773–795.
[22] G. Koop, R. Leon-Gonzalez, R. Strachan, Bayesian inference in a cointegrating panel data model, in: Bayesian Econometrics, in: Advances in Econometrics, vol. 23, Emerald Group Publishing Limited, WA, UK, 2008, pp. 433–469.
[23] E. Li, D. Zhang, M. Davidian, Conditional estimation for generalized linear models when covariates are subject-specific parameters in a mixed model for longitudinal measurements, Biometrics 60 (2004) 1–7.
[24] G. Marshall, A.E. Barón, Linear discriminant models for unbalanced longitudinal data, Stat. Med. 19 (2000) 1969–1981.
[25] R.D. Morey, J. Rouder, M. Pratte, P. Speckman, Using MCMC chain outputs to efficiently estimate Bayes factors, J. Math. Psychol. 55 (2011) 368–378.
[26] H.G. Müller, U. Stadtmüller, Generalized functional linear models, Ann. Statist. 33 (2005) 774–805.
[27] A. Neuhaus, T. Augustin, C. Heumann, D. Daumer, A review on joint models in biometrical research, J. Stat. Theory Pract. 3 (2009) 855–868.
[28] S.J. Ratcliffe, G.Z. Heller, L.R. Leader, Functional data analysis with application to periodically stimulated foetal heart rate data. II: Functional logistic regression, Stat. Med. 21 (2002) 1115–1127.
[29] G.O. Roberts, J.S. Rosenthal, Examples of adaptive mcmc, J. Comput. Graph. Statist. 18 (2009) 349–367.
[30] J.C. Román, J.P. Hobert, Convergence analysis of the Gibbs sampler for Bayesian general linear mixed models with improper priors, Ann. Statist. 40 (2012) 2823–2849.
[31] J.C. Román, J.P. Hobert, Geometric ergodicity of Gibbs samplers for Bayesian general linear mixed models with proper priors, Linear Algebra Appl. 473 (2015) 54–77.

[32] D. Ryu, E. Li, B. Mallick, Bayesian nonparametric regression analysis when covariates are subject-specific parameters in a random effect model for longitudinal measurement, Biometrics 67 (2011) 454–466.
[33] B.J. Smith, Bayesian Output Analysis Program (BOA) for MCMC. R package version 1.1.2-1, 2004. Available at http://www.public-health.uiowa.edu/boa.
[34] G. Verbeke, E. Lesaffre, A linear mixed-effects model with heterogeneity in the random-effects population, J. Amer. Statist. Assoc. 91 (1996) 217–221.
[35] E.F. Vonesh, V.M. Chinchilli, Linear and Nonlinear Models for the Analysis of Repeated Measurements, Marcel Dekker, New York, 1997.
[36] C.Y. Wang, N. Wang, S. Wang, Regression analysis when covariates are regression parameters of a random effects model for observed longitudinal measurements, Biometrics 56 (2000) 487–495.
[37] R. Wetzels, J.G.W. Raaijmakers, E.-J. Wagenmakers, How to quantify support for and against the null hypothesis: a flexible winbugs implementation of a default bayesian *t*-test, Psychon. Bull. Rev. 16 (2009) 752–760.
[38] L. Wu, X.J. Hu, H. Wu, Joint inference for nonlinear mixed-effects models and time-to-event at the presence of missing data, Biostatistics 9 (2008) 308–320.
[39] L. Wu, W. Liu, X.J. Hu, Joint inference on HIV viral dynamics and immune suppression in presence of measurement errors, Biometrics 66 (2010) 327–335.