Joint modeling a primary endpoint and longitudinal data

Rolando De la Cruz[1,⋆], Guillermo Marshall[2,⋆], and Fernando A. Quintana[2,⋆]

[1]Departamento de Salud Pública, Escuela de Medicina

Pontificia Universidad Católica de Chile,

Marcoleta 434, Casilla 114D, Santiago, CHILE.

[2]Departamento de Estadística, Facultad de Matemáticas,

Pontificia Universidad Católica de Chile,

Casilla 306, Correo 22, Santiago, CHILE.

[⋆]Email: rolando@med.puc.cl gm;quintana@mat.puc.cl

Abstract

In many studies the association of longitudinal measurements of a continuous response and a primary endpoint are often of interest. A convenient framework for this type of problems is joint models, which are formulated to investigate the association between a primary endpoint and features of longitudinal measurements through a common set of latent random effects. The joint model, which is the focus of this article, is a logistic regression model with covariates defined as the individual–specific random effects in a non–linear random effects model for the longitudinal measurements. We discuss different estimation procedures, which include two-stage, best linear unbiased predictors, and various numerical integration techniques. The proposed methods are illustrated using a real dataset where the objective is to study the association between longitudinal hormone levels and the pregnancy outcome in a group of young women. The numerical performance of the estimating methods is also evaluated by means of simulation.

Key Words: Best linear unbiased predictor (BLUP); Gaussian quadrature methods; Laplace approximation; Logistic regression model; Non–linear mixed effects models; Two–stage estimator.

1 INTRODUCTION

In many scientific investigations, a primary endpoint and a number of longitudinal measurements of a continuous response are collected for each individual along with other covariates, the association between the primary endpoint and features of the longitudinal profiles being of interest. One such example is the pregnant women data of Section 5. Investigators wished to understand the association between the pregnancy outcome and features of hormonal patterns over the gestational period. During the pregnancy, longitudinal beta human chorionic gonadotropin (β–HCG) levels derived from the blood were obtained from 173 women. The β–HCG concentrations for the 173 women were measured during the first 80 days of gestational age. Consequently, pregnancy outcomes were divided into two groups: normal and abnormal. The women were classified as having a normal pregnancy if they had a normal delivery or as having an abnormal pregnancy if they had any complication resulting in a non-terminal delivery and loss of the fetus. Figure 1 exhibits time profiles in log scale for both groups. Unjoined points correspond to women for whom only one response was available. We can see that there is a great variation in β–HCG levels. Also, a non–linear relationship of the log β–HCG levels with gestational age (in days) is common for most women. Assessing whether log β–HCG levels at early stages rise and fall are associated with the pregnancy

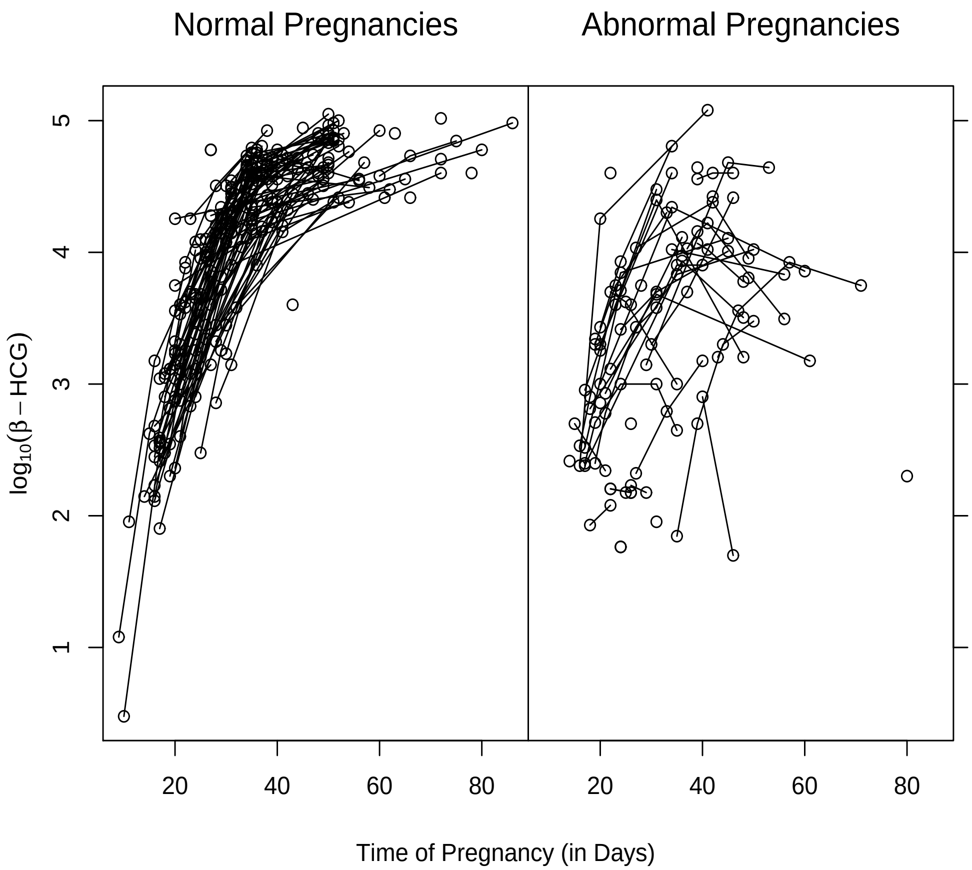

Figure 1: Observed time profiles of β–HCG for both normal and abnormal groups.

outcome would suggest the need for further evaluation and testing. However, these features are observed only through the log β–HCG measurements, which are subject to assay error among other variations.

A framework that accounts for this variation is the 'joint model', which assumes that the longitudinal data follow a non–linear random effects model whose random effects are covariates in a logistic regression model for the primary endpoint. For the pregnant women data, the woman–to–woman variation has been shown to be adequately accounted for by the introduction of random effects to model the asymptotic behavior of the log β–HCG using a three–parameter logistic curve model (see Marshall and Barón 2000), and for the primary endpoint normal/abnormal pregnancy, a logistic model incorporating these effects and other covariates appears to be a natural choice.

Much of the early literature on joint modeling focuses on models where the objective is to characterize the relationship between a longitudinal response process and a time–to–event. Here the primary endpoint is the time–to–event. Parametric and semiparametric survival models for event times together with linear or non–linear mixed effects model for longitudinal measurements are proposed for jointly modeling longitudinal data and event times, where a set of random effects is assumed to induce their interdependence (e.g. DeGruttola and Tu 1994; Wulfsohn and Tsiatis 1997; Henderson *and others* 2002; Guo and Carlin 2004; Vonesh et al 2006; Wu *and others* 2007). For a review of works in this direction see Tsiatis and Davidian (2004).

For the special case where a linear mixed effects model is used for the longitudinal response and the primary endpoint is a single binary response, count, or other continuous measure, a number of different approaches have been proposed to estimate the parameters of the joint model. Wang *and others* (2000) showed that replacing the latent random effects by individual–specific ordinary least squares estimators leads to biased estimation of primary model parameters. Wang *and others* (2000) also proposed regression calibration (Carroll *and others* 2006, Chapter 4), which in this context involves replacing the random effects by estimated best linear unbiased predictors (BLUP) from a fit of the linear mixed effects model to the longitudinal data. Because this method reduces but does not eliminate the bias, they also proposed complementary procedures to correct such problem. Following the strategy of Stefanski and Carroll (1987), Li *and others* (2004) proposed estimators in the case of generalized linear models for the primary endpoint. Their method requires no normality assumptions on the random effects and yields consistent inference regardless the true distribution. Li *and others* (2007) consider a semiparametric joint model for which the random effects distribution is left unspecified, but assumed to lie in a broad class whose members have smooth densities that render them likely candidates to represent variation in individual–specific random effects for continuous longitudinal responses. Extensions to a joint model framework that incorporates multiple longitudinal data are developed by Li *and others* (2007).

In this paper we develop a framework for the logistic regression model with covariates that are individual–specific random effects in a non–linear random effects model for the longitudinal measurements. We discuss different methods to estimate the model parameters, including two-stage, best linear unbiased predictor, and approximations based on various numerical integration techniques. The two-stage method fits the primary endpoint model by replacing the unobserved latent random effects with their estimators obtained by fitting individual non–linear regression models using each individual's data. The BLUP method replaces the unobserved latent random effects with the BLUP estimators obtained from fitting a non–linear mixed effects model to the longitudinal data. Within the numerical integration techniques we use adaptive Gaussian quadrature to approximate the marginal likelihood, which involves an integral that cannot be solved analytically, and employ the quasi-Newton algorithm to obtain the maximum likelihood estimation (MLE) of model parameters. The proposed methods are illustrated with an application to the pregnant women data and their performance is evaluated through a simulation study.

The rest of this paper is organized as follows: In Section 2 the 'joint model' is formulated. The estimation procedures are discussed in Section 3. In Section 4 a simulation study is carried out to evaluate the performance of the estimating methods. In Section 5 the methodology is exemplified using the pregnant women data set. Finally, we summarize and discuss implications in Section 6.

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 non–linear random effects model over possibly unequally spaced times. The second component is the primary logistic regression where the random coefficients of the random effects model are used as covariates.

Denote by Y_{ij} the observed longitudinal measurement data of a continuous response for individual i, $i = 1, \ldots, m$, recorded at times t_{ij} ($j = 1, \ldots, k_i$). Assume that Y_{ij} follow the non-linear random effects model

$$Y_{ij} = f(\boldsymbol{X}_i, t_{ij}) + U_{ij}, \tag{1}$$

where $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{iq})'$ represents the ith set of random effects. We further assume that $\boldsymbol{X}_i$ is a multivariate normal with mean $\boldsymbol{\mu_x}$ and variance-covariance matrix $\boldsymbol{\Sigma_x}$. Also let $\boldsymbol{U}_i \equiv (U_{i1}, \ldots, U_{ik_i})'$ the within–individual errors reflecting uncertainty in measuring $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{ik_i})'$, which is assumed to follow a multivariate normal distribution and satisfy $\mathbb{E}(\boldsymbol{U}_i|\boldsymbol{X}_i) = 0$ and $\text{Var}(\boldsymbol{U}_i|\boldsymbol{X}_i) = \sigma_u^2 \mathbf{I}_{k_i}$. The random error $\boldsymbol{U}_i$ is assumed to be independent of $\boldsymbol{X}_i$.

Now, let D_i be the binary response variable for the ith individual, $i = 1, \ldots, m$. We assume that the primary regression satisfies

$$\Pr(D_i = 1|\boldsymbol{X}_i) = H(\mathcal{X}_i'\boldsymbol{\beta}), \tag{2}$$

where $H(u) = \{1 + \exp(-u)\}^{-1}$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_q)'$ are the regression parameters and $\mathcal{X}_i = (1, \boldsymbol{X}_i')'$. Note that we can easily add a term like $\boldsymbol{\alpha}'\boldsymbol{Z}_i$ to $\mathcal{X}_i'\boldsymbol{\beta}$ in (2) to account for any additional covariates $\boldsymbol{Z}_i$ that may be observed.

Interest focuses on estimating $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \sigma_u^2, \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)'$. We consider maximum likelihood estimation of these parameters. As discussed in Wang *and others* (2000), we can further assume that $\boldsymbol{Y}_i$ and D_i are conditionally independent given $\boldsymbol{X}_i$, in which case

$$P(\boldsymbol{Y}_i, D_i, \boldsymbol{X}_i) = P(\boldsymbol{Y}_i, D_i|\boldsymbol{X}_i)P(\boldsymbol{X}_i) = P(\boldsymbol{Y}_i|\boldsymbol{X}_i)P(D_i|\boldsymbol{X}_i)P(\boldsymbol{X}_i).$$

The likelihood for the *joint* model $(\boldsymbol{Y}_i, D_i)$ is given by

$$P(\boldsymbol{Y}, \boldsymbol{D}) = \prod_{i=1}^{m} \int_{\boldsymbol{x}} P(\boldsymbol{Y}_i|\boldsymbol{X}_i)P(D_i|\boldsymbol{X}_i)P(\boldsymbol{X}_i)d\,\boldsymbol{x}_i. \tag{3}$$

Maximum likelihood estimation (MLE) requires maximizing the integrated log-likelihood

$$L(\boldsymbol{\Theta}|\boldsymbol{Y}, \boldsymbol{D}) = \sum_{i=1}^{m} \log \int_{\boldsymbol{x}} P(\boldsymbol{Y}_i|X_i)P(D_i|\boldsymbol{X}_i)P(\boldsymbol{X}_i)d\,\boldsymbol{x}_i. \tag{4}$$

Note that the *joint* model $(\boldsymbol{Y}_i, D_i)$ is nonlinear in $\boldsymbol{X}_i$, thus the integral in (4) does not have closed-form expression. However approximation methods can be used to help the estimation. These will be discussed next.

3 ESTIMATION PROCEDURES

Several procedures are available for estimating the parameters involved in the joint model. We discuss here the two-stage estimator, the approximate BLUP estimator and methods based on numerical integration techniques which include Gaussian quadrature, adaptive Gaussian quadrature and Laplace approximation. In Section 4 we describe the results of a simulation study designed to evaluate the performance of the estimation procedures under different configurations of m and k_i.

3.1 The Two-Stage Estimator

Should the latent variables $\boldsymbol{X}_i$ be observed, it would be natural to estimate $\boldsymbol{\beta}$ by fitting the logistic model (2). Because the latent variables $\boldsymbol{X}_i$ cannot be observed, a convenient method consists of imputing estimates of $\boldsymbol{X}_i$ in model (2).

One such method is the simple two-stage approach, also called the "naive" method. If the k_i are sufficiently large, we can estimate $\boldsymbol{X}_i$ under model (1) using standard nonlinear ordinary least squares, which can be easily implemented using existing software like SAS or R. Once we have individual estimates of $\boldsymbol{X}_i$, we can fit model (2) by replacing each $\boldsymbol{X}_i$ by its corresponding estimate. In case that the longitudinal measurements are assumed to follow a linear mixed model, Wang *and others* (2000) showed that the "naive" method may result in a significantly biased estimator.

3.2 The BLUP estimator

Because replacing individual estimates of $\boldsymbol{X}_i$ in model (2) results in biased estimates, it seems natural to consider an improvement in the estimation of $\boldsymbol{X}_i$ in stage one. A good candidate is the shrinkage or empirical Bayes estimator, also known as best linear unbiased predictors (BLUP) in the longitudinal literature. The BLUP estimator represents a compromise between estimates based only on an individual subject's data, and estimates based only on the population mean. Subjects with substantial data will have estimates closer to their individual least squares curves, whereas subjects with sparse data will have estimates closer to the population mean.

Since in general the latent variables $\boldsymbol{X}_i$ enter model (1) in a nonlinear fashion, the BLUP estimator can be approximately estimated using the conditional mean $\hat{\boldsymbol{X}}_i \approx \mathbb{E}(\boldsymbol{X}_i|\boldsymbol{Y}_i)$ (see Vonesh and Chinchilli 1997). Thus, we first fit the nonlinear mixed model (1) to obtain the BLUP estimates of $\boldsymbol{X}_i$, $\hat{\boldsymbol{X}}_i$, and then fit model (2) imputing the $\boldsymbol{X}_i$'s by these estimates. In the special case where the longitudinal measurements follow a linear mixed model, this method has been termed the regression calibration (RC) estimator by Wang *and others* (2000). The name comes from the fact that this method can be viewed as a particular case of measurement error problems. Wang *and others* (2000) show that their method reduces but does not eliminate bias.

3.3 Exact estimation using numerical integration techniques

Since the *joint model* will always be nonlinear in the latent variables $\boldsymbol{X}_i$, even when the conditional model for $\boldsymbol{Y}_i|\boldsymbol{X}_i$ is linear in $\boldsymbol{X}_i$, the integral (3) will generally have no closed form expression. A number of methods have been proposed for approximating the integrated log-likelihood, including various numerical integration techniques (Pinheiro and Bates 1995). One such technique is the Laplacian approximation, a method frequently used in Bayesian inference to estimate posterior densities and predictive distributions (see Tierney and Kadane 1986; Leonard *and others* 1989). The use of the Laplacian approximation for model (1) was described by Pinheiro and Bates (1995). Write

$$k_i \ell(\boldsymbol{X}_i) = \log\{P(\boldsymbol{Y}_i|\boldsymbol{X}_i)P(D_i|\boldsymbol{X}_i)\} - \frac{1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{x}}| - \frac{1}{2}\boldsymbol{X}_i'\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{X}_i,$$

where $P(\boldsymbol{Y}_i|\boldsymbol{X}_i)$ is the normal density function of $\boldsymbol{Y}_i|\boldsymbol{X}_i$ and $P(D_i|\boldsymbol{X}_i)$ is the Bernoulli density of $D_i|\boldsymbol{X}_i$. The Laplacian approximation entails taking a second–order expansion of the i^{th} individual's integrated likelihood, $L(\boldsymbol{Y}_i) = \int \exp[k_i\ell(\boldsymbol{X}_i)]\, d\boldsymbol{X}_i$, around the value of $\boldsymbol{X}_i$ that maximizes $k_i\ell(\boldsymbol{X}_i)$, holding all other parameters fixed.

Another numerical integration technique is the Gaussian quadrature. The Gaussian quadrature approximates the integral of a function, with respect to a given kernel, by a weighted sum over predefined abscissas for the random effects. Unlike other numerical integration techniques, the abscissas are spaced unevenly throughout the interval of integration. With a modest number of quadrature points, along with appropriate centering and scaling of the abscissas, the Gaussian quadrature approximation can be highly effective (see Abramowitz and Stegun 1964 for details).

Adaptive Gaussian quadrature for the integral over $\boldsymbol{X}_i$ centers the integral at the empirical Bayes estimate of $\boldsymbol{X}_i$, defined as the vector $\hat{\boldsymbol{X}}_i$ that minimizes

$$-\log\{P(\boldsymbol{Y}_i|X_i, \boldsymbol{\Theta})P(D_i|\boldsymbol{X}_i, \boldsymbol{\Theta})P(\boldsymbol{X}_i|\boldsymbol{\Theta})\},$$

with $\boldsymbol{\Theta}$ set equal to their current estimates. The final Hessian matrix from this optimization can be used to scale the quadrature abscissas. For $j = 1, \ldots, N_{GQ}$, suppose that (z_j, w_j) are, respectively, the standard Gauss–Hermite abscissas and weights for the Gaussian quadrature rule with N_{GQ} points (Golub and Welsch 1969). The adaptive Gaussian quadrature is then given by

$$
\begin{aligned}
\ell_i(\mathbf{\Theta}|\boldsymbol{Y}_i, D_i) &= \int_{\boldsymbol{x}} P(\boldsymbol{Y}_i|X_i, \mathbf{\Theta}) P(D_i|\boldsymbol{X}_i, \mathbf{\Theta}) P(\boldsymbol{X}_i|\mathbf{\Theta}) d\,\boldsymbol{X}_i \\
&\approx 2^{q/2}|\Gamma(\mathbf{\Theta})|^{-1/2} \sum_{j_1=1}^{N_{GQ}} \cdots \sum_{j_q=1}^{N_{GQ}} \Bigg\{ P(\boldsymbol{Y}_i|a_{j_1,\ldots,j_q}, \mathbf{\Theta}) P(D_i|a_{j_1,\ldots,j_q}, \mathbf{\Theta}) P(a_{j_1,\ldots,j_q}|\mathbf{\Theta}) \\
&\qquad \prod_{s=1}^{q} w_{j_s} \exp[z_{j_s}^2] \Bigg\},
\end{aligned}
$$

where q is the dimension of $\boldsymbol{X}_i$, $\Gamma(\mathbf{\Theta})$ is the Hessian matrix from the empirical Bayes minimization, $z_{j_1,\ldots,j_q}$ is a vector with elements $(z_{j_1}, \ldots, z_{j_r})$, and

$$
a_{j_1,\ldots,j_q} = \hat{\boldsymbol{X}}_i + 2^{1/2}\Gamma(\mathbf{\Theta})^{-1/2} z_{j_1,\ldots,j_q}.
$$

When $N_{GQ} = 1$, adaptive Gaussian quadrature approximation is simply the Laplacian approximation described previously, because in this case $z_1 = 0$ and $w_1 = 1$. This adaptive Gaussian quadrature approximation is similar to the approximation obtained from adaptive importance sampling; the basic difference is that the former uses fixed abscissas and weights, but the latter allows them to be determined by a pseudo–random mechanism. As with the importance sampling approximation, the adaptive Gaussian quadrature produces the exact log–likelihood when the model function is linear in random effects $\boldsymbol{X}_i$. In practice, $N_{GQ} \leq 7$ generally suffices and $N_{GQ} = 1$ often provides a reasonable approximation (Pinheiro and Bates 1995).

The objective function to minimize corresponds to the negative log–likelihood

$$
\mathcal{Q}_{AGQ}(\mathbf{\Theta}) = -\sum_{i=1}^{m} \log\{\ell_i(\mathbf{\Theta}|\boldsymbol{Y}_i, D_i)\}.
$$

We employ a Newton–type algorithm to minimize the negative log–likelihood.

4 SIMULATION STUDY

In this Section we discuss the results of a simulation study designed to compare the different estimation methods described in Section 3. The true parameter values are chosen to mimic what we find in the example presented below in Section 5. The longitudinal data were generated according to $y_{ij} = x_i[1 + \exp\{-(t_j - \mu_2)/\mu_3\}]^{-1} + U_{ij}$, where $i = 1, \ldots, m$, $j = 1, \ldots, k_i$, $U_{ij} \sim N(0, \sigma_u^2)$, and $x_i \sim N(\mu_1, \sigma_x^2)$. For the logistic primary model we assume $P(D_i = 1) = [1 + \exp\{-(\beta_0 + \beta_1 x_i)\}]^{-1}$. One thousand datasets were generated for each combination of the following number of individuals: $m = 200$ or 400, and the number of measurements within each individual $k_i = 5, 10,$ or 15. Time points were chosen from the empirical distribution of observed times of our example of Section 5.

Table 1 shows the simulation results for the primary logistic model parameters. Here, "Bias" corresponds to the average of the $(\hat{\beta} - \beta)$ differences, SD denotes the sample standard deviation of the estimators, mean(SE) denotes the average of the estimated standard errors of the estimators, and 95% CP denotes the 95% confidence interval coverage probabilities. We note here that it was not possible to implement the two–stage method for the $k_i = 5$ case, because convergence was not achieved when fitting the individual non–linear regression model.

When the estimation was based on (adaptive) Gaussian quadrature, our results suggest that the bias decreased when any of the number of individuals or longitudinal measurements within each individual increased. Indeed nearly unbiased parameter estimates were obtained when setting $k_i = 15$. In contrast biases observed for β_0 when applying the two–stage

Table 1: Simulation results. Estimators are: TS, two–stage; BLUP, best linear unbiased predictor; Laplacian, Laplacian approximation; GQ, Gaussian quadrature; AGQ, adaptive Gaussian quadrature. N_{GQ} indicates the number of quadrature points.

							GQ($N_{GQ}=q$)			AGQ($N_{GQ}=q$)		
m	k_i			TS	BLUP	Laplacian	$q=10$	$q=50$	$q=100$	$q=10$	$q=50$	$q=100$
200	5	β_0	Bias	–	-1.105	-0.227	-0.373	-0.187	-0.186	-0.186	-0.186	-0.186
			SD	–	2.361	2.297	2.471	2.502	2.501	2.501	2.501	2.501
			Mean(SE)	–	2.583	3.141	3.153	3.153	3.153	3.153	3.153	3.153
			95% CP	–	90.9	96.5	95.5	96.6	96.6	96.6	96.6	96.6
		β_1	Bias	–	0.267	0.051	0.086	0.041	0.041	0.041	0.041	0.041
			SD	–	0.548	0.583	0.576	0.584	0.583	0.583	0.583	0.583
			Mean(SE)	–	0.601	0.734	0.736	0.737	0.737	0.737	0.737	0.737
			95% CP	–	90.9	96.9	95.5	96.9	96.9	96.9	96.9	96.9
	10	β_0	Bias	-1.508	-0.563	-0.191	-0.396	-0.156	-0.198	-0.198	-0.198	-0.198
			SD	2.336	2.302	2.322	2.387	2.344	2.312	2.312	2.312	2.312
			Mean(SE)	2.503	2.641	2.872	2.842	2.879	2.872	2.872	2.872	2.872
			95% CP	87.6	95.6	97.1	96.3	97.1	97.2	97.2	97.2	97.2
		β_1	Bias	0.354	0.133	0.041	0.092	0.033	0.043	0.043	0.043	0.043
			SD	0.547	0.538	0.545	0.559	0.550	0.542	0.542	0.542	0.542
			Mean(SE)	0.584	0.616	0.671	0.664	0.673	0.671	0.671	0.671	0.671
			95% CP	87.8	95.2	97.0	95.6	97.0	97.0	97.1	97.1	97.1
	15	β_0	Bias	-0.757	-0.229	-0.029	-0.262	-0.053	-0.026	-0.024	-0.024	-0.024
			SD	2.353	2.357	2.346	2.376	2.319	2.348	2.347	2.347	2.347
			Mean(SE)	2.606	2.685	2.827	2.792	2.820	2.827	2.827	2.827	2.827
			95% CP	93.1	96.3	96.9	96.0	97.1	96.9	97.0	97.0	97.0
		β_1	Bias	0.178	0.056	0.006	0.063	0.012	0.006	0.005	0.005	0.005
			SD	0.549	0.550	0.548	0.555	0.541	0.548	0.548	0.548	0.548
			Mean(SE)	0.609	0.627	0.661	0.652	0.659	0.661	0.661	0.661	0.661
			95% CP	93.2	96.1	97.0	95.8	97.2	97.0	97.0	97.0	97.0
400	5	β_0	Bias	–	-1.034	0.105	0.037	0.160	0.160	0.160	0.160	0.160
			SD	–	1.834	2.012	2.044	2.021	2.021	2.021	2.021	2.021
			Mean(SE)	–	1.823	2.255	2.278	2.266	2.266	2.266	2.266	2.266
			95% CP	–	88.5	96.6	95.8	96.7	96.7	96.7	96.7	96.7
		β_1	Bias	–	0.253	-0.025	-0.006	-0.038	-0.038	-0.038	-0.038	-0.038
			SD	–	0.427	0.471	0.476	0.473	0.473	0.473	0.473	0.473
			Mean(SE)	–	0.425	0.527	0.532	0.530	0.530	0.530	0.530	0.530
			95% CP	–	87.9	96.7	95.9	96.9	96.9	96.9	96.9	96.9
	10	β_0	Bias	-1.496	-0.384	0.156	-0.119	0.161	0.167	0.167	0.167	0.167
			SD	1.714	1.747	1.833	1.833	1.833	1.835	1.835	1.835	1.835
			Mean(SE)	1.762	1.876	2.058	2.033	2.058	2.059	2.059	2.059	2.059
			95% CP	83.1	95.2	98.0	97.1	98.0	98.0	98.0	98.0	98.0
		β_1	Bias	0.353	0.093	-0.038	0.030	-0.039	-0.040	-0.040	-0.040	-0.040
			SD	0.400	0.408	0.429	0.428	0.430	0.430	0.430	0.430	0.430
			Mean(SE)	0.411	0.438	0.481	0.475	0.481	0.481	0.481	0.481	0.481
			95% CP	82.6	95.3	98.3	96.6	98.1	98.3	98.3	98.3	98.3
	15	β_0	Bias	-0.811	-0.256	0.078	-0.202	0.066	0.067	0.087	0.081	0.081
			SD	1.812	1.856	1.904	1.896	1.905	1.912	1.908	1.908	1.908
			Mean(SE)	1.829	1.886	1.999	1.970	1.998	2.001	2.000	2.000	2.000
			95% CP	92.3	95.7	96.8	96.0	96.9	96.8	96.9	96.9	96.9
		β_1	Bias	0.191	0.062	-0.019	0.049	-0.017	-0.022	-0.020	-0.020	-0.020
			SD	0.421	0.432	0.444	0.441	0.445	0.446	0.445	0.445	0.445
			Mean(SE)	0.427	0.440	0.467	0.460	0.467	0.468	0.468	0.468	0.468
			95% CP	96.2	95.5	96.8	96.1	96.8	96.5	97.0	97.0	97.0

estimator were quite large, regardless of the number of individuals. We also found fairly substantial biases for the BLUP estimator in the case of $k_i = 5$ observations per individual, thus leading to low coverage probabilities. When increasing k_i to 10 or 15, the BLUP estimation reduces but does not eliminate bias. These findings agree with the results reported in Wang *and others* (2000) in the context of longitudinal measurements following a liner mixed effects model. Estimation based on adaptive Gaussian quadrature and Laplacian approximations yields almost unbiased estimates and better coverage probabilities, specially when $k_i = 10,$ or 15. For the non–adaptive Gaussian quadrature approximation we need more than 10 quadrature points to obtain results that are comparable to those obtained using the adaptive Gaussian quadrature and Laplacian approximations.

In summary, the TS method has a serious bias problem, especially for β_0. The BLUP method does not perform well regardless of both the number of individuals and the number of observations per individual. Because the results obtained using the Laplacian approximation were very similar to those of the adaptive Gaussian approximation, we conclude that the first is to be preferred for its greater simplicity and computational efficiency, specially when $k_i > 10$.

5 ANALYSIS OF PREGNANT WOMEN DATA

The main interest of the analysis of pregnant women dataset discussed in Section 1 was to investigate the effects of the β–HCG longitudinal process on the pregnancy outcome, particularly the association between normal pregnancy and features of longitudinal β-HCG profiles. The data were collected for a total of 173 young women, representing different pregnancies over a period of 2 years in a private fertilization obstetrics clinic in Santiago, Chile. The resulting data set consists of 124 patients diagnosed with normal pregnancy and 49 patients with abnormal pregnancy. Let $D_i = 1$ and 0 denote the normal and abnormal pregnancy 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 k_i per woman ranged from 1 to 6 (median 2). Approximately 30% of the 173 women had one β–HCG measurement, 31% had two, 33% had three, and 6% had four or more measurements.

As discussed in earlier works (Marshall and Barón 2000; De la Cruz–Mesía and Quintana 2007; De la Cruz–Mesía *and others* 2007), a reasonable representation of the log β–HCG profile for the ith woman is

$$y_{ij} = \frac{x_i}{1 + \exp\{-(t_{ij} - \mu_2)/\mu_3\}} + U_{ij} \tag{5}$$

where time is measured in days, U_{ij} is a $\mathcal{N}(0, \sigma_u^2)$ measurement error, x_i is the woman–specific random effect, assumed to satisfy $x_i \sim \mathcal{N}(\mu_1, \sigma_x^2)$. The random effects x_i represent the asymptotic behavior of the log β-HCG. To describe the relation between pregnancy outcome and x_i, we consider the primary logistic regression model

$$\Pr(D_i = 1|x_i) = [1 + \exp\{-(\beta_1 + \beta_2 x_i)\}]^{-1}. \tag{6}$$

We use different quadrature points N_{GQ} ($N_{GQ} = 1, 10, 50, 200$) for the methods based on Gaussian quadrature. The case $N_{GQ} = 1$ corresponds to the Laplacian approximation. The standard errors of parameters are constructed from the vari ance–covariance matrix computed as the inverse negative Hessian matrix.

Table 2 shows results from fitting the joint model (5) and (6) by the methods described in Section 3. Table 2 shows the results for the Gaussian quadrature and adaptive Gaussian

Table 2: Parameter estimates for the pregnant women data via several methods, with standard errors within parentheses. Estimators are: BLUP, best linear unbiased predictor; Laplacian, Laplacian approximation; GQ, Gaussian quadrature; AGQ, adaptive Gaussian quadrature. N_{GQ} indicates the number of quadrature points.

Parameter	BLUP	Laplacian	GQ($N_{GQ}=10$)	GQ($N_{GQ}=50$)	AGQ($N_{GQ}=10$)	AGQ($N_{GQ}=50$)
	Longitudinal submodel					
	Est.(SE)	Est.(SE)	Est.(SE)	Est.(SE)	Est.(SE)	Est.(SE)
μ_1	4.506(0.064)	4.490(0.063)	4.476(0.054)	4.490(0.063)	4.490(0.063)	4.490(0.063)
μ_2	15.048(0.381)	14.882(0.388)	14.913(0.376)	14.882(0.388)	14.882(0.388)	14.882(0.388)
μ_3	7.375(0.506)	7.379(0.507)	7.329(0.485)	7.379(0.507)	7.379(0.507)	7.379(0.507)
σ_x^2	0.287(0.045)	0.277(0.044)	0.321(0.048)	0.278(0.044)	0.278(0.044)	0.278(0.044)
σ_u^2	0.128(0.013)	0.130(0.014)	0.122(0.013)	0.130(0.013)	0.130(0.013)	0.130(0.013)
	Logistic submodel					
	Est.(SE)	Est.(SE)	Est.(SE)	Est.(SE)	Est.(SE)	Est.(SE)
β_1	-13.852(2.818)	-17.398(5.049)	-15.603(4.082)	-17.213(4.772)	-17.213(4.772)	-17.214(4.772)
β_2	3.313(0.628)	4.169(1.155)	3.752(0.930)	4.121(1.089)	4.121(1.089)	4.121(1.089)

quadrature methods using $N_{GQ} = 50$ only, as those corresponding to $N_{GQ} = 100, 200$ are almost identical. Also, we remark that the two–stage estimator cannot be reliably implemented in this case because of the insufficient number of measurements per woman (recall k_i ranged from 1 to 6). As was to be expected from the simulation results, the BLUP and numerical integration techniques give different results. Also, the results obtained using the Laplacian approximation and adaptive Gaussian quadrature are very similar. We also note that the non–adaptive Gaussian quadrature method required more than 10 quadrature points to obtain similar results than the adaptive Gaussian quadrature. From the BLUP estimates, the odds ratio (e^{β_1}) associated with a 1–unit increase in the value of x_i is 27.5. From the adaptive Gaussian quadrature estimates the odds ratio is 61.6. The overall conclusion from the various estimation methods is that an abnormal pregnancy may be negatively associated with the asymptotic behavior of the log β–HCG profile. Also in agreement with the simulation results from Section 4, the adaptive Gaussian quadrature method with 10 points seems to provide cost-effective and reliable estimates for this particular application. Therefore, in what follows calculations are based exclusively on this method.

Having the $P(D_i|x_i)$ estimates available, we turn now to the underlaying classification problem. To do so we calculate the confusion matrix of classification. Table 3 gives the confusion matrix of classification for the adaptive Gaussian quadrature method. We found an error rate estimation of approximately 12.7%. We compare these results with those found using discriminant analysis (see Marshall and Barón 2000) in which case the reported error rate was approximately 18.5%. The joint model is thus seen to improve classification.

The Receiver Operating Characteristic (ROC) curves and the area under the ROC curve (AUC) for the joint modeling and discriminant analysis method of Marshall and Barón

Table 3: Confusion matrix of classification for the adaptive Gaussian quadrature method.

Clinical classification		Model classification		
	Groups	Normal	Abnormal	
	Normal	122	2	124
	Abnormal	20	29	49
		142	31	173

(2000) are presented in Figure 2. Specifically, we present two curves showing the changes in sensitivity and specificity using the joint modeling based on the adaptive Gaussian quadrature approximation with $N_{GQ} = 10$ and the results obtained using the longitudinal discriminant analysis method proposed by Marshall and Barón (2000). Again, using the joint modeling approach improves sensitivity and specificity for predicting a normal pregnancy outcome for this population of women.

6 DISCUSSION

In this paper we have proposed inferential strategies for logistic regression model for a primary outcome with covariates that are underlying individual–specific random effects in a non–linear random effects model for a longitudinal response. We discussed several estimation procedures for the model parameters.

Our results show that both the two–stage method and the BLUP method could yield biased estimates, although the former often performs better than the latter. Also, the BLUP

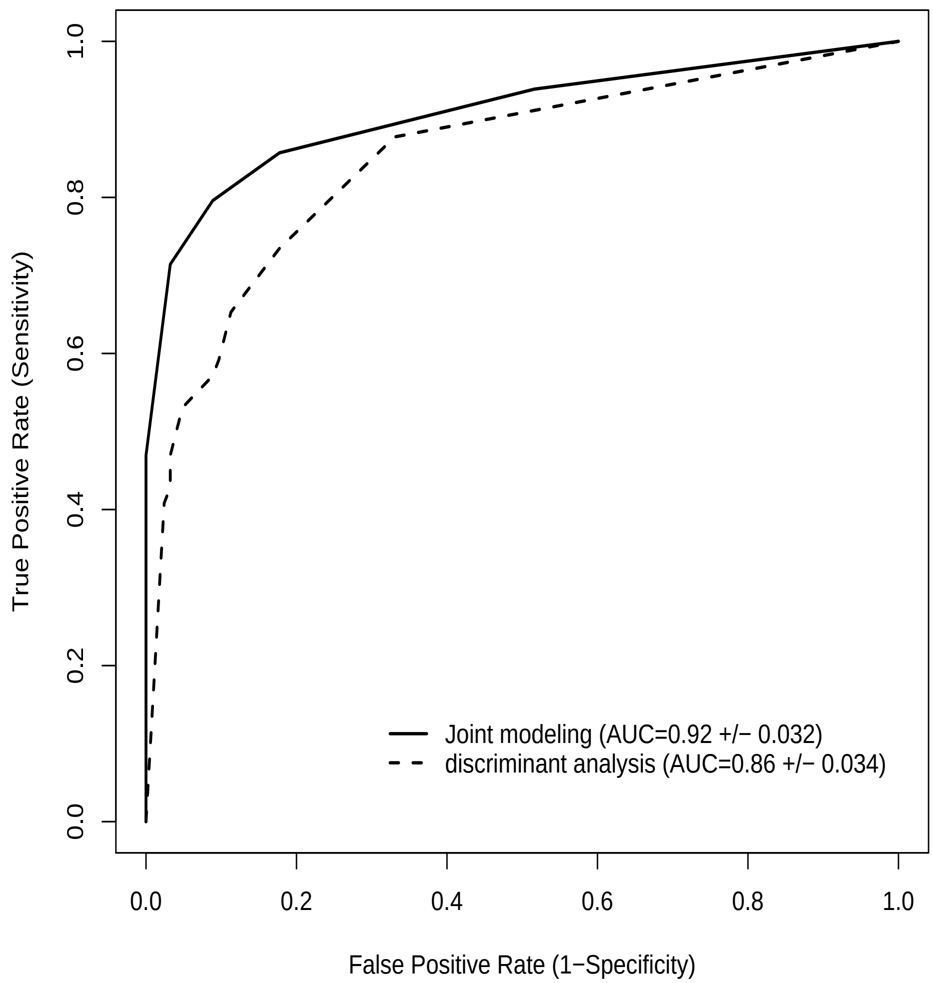

Figure 2: ROC curves with areas and standard errors for the joint modeling and the discriminant analysis approach.

method attenuates but does not eliminate bias. The Gaussian quadrature approximation only seems to give accurate results for a large number of quadrature points (> 10), which makes it very inefficient computationally. Of all approximations considered here, the Laplacian and adaptive Gaussian approximations probably give the best mix of efficiency and accuracy. The former can be regarded as a particular case of the latter, where just one quadrature point is used. For statistical analysis purpose we would recommend use the Laplacian approximation whether the number of observations per individual is at least 10, in other case, the adaptive Gaussian approximations with $N_{GQ} = 10$ seems preferable.

In our example, we only use as covariate the latent random effects in the logistic regression model, but other women's covariates, such as age, number of previous normal and abnormal pregnancies, smoking status and normal pregnancy tendencies, can be useful for targeting specific individuals in future analysis. In our example a number of women had missing covariate values.

All proposed estimators assume normality of random effects and within–individual errors. The latter is often reasonable, perhaps on a transformed scale. Some authors (e.g., Verbeke and Lesaffre 1997; among others) have shown that violation of this assumption can compromise inference, thus raising such 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.

Acknowledgements

The first author thanks the Comisión Nacional de Investigación Científica y Tecnológica – CONICYT, for partially supporting his Ph.D. studies at the Pontificia Universidad Católica de Chile, and Vicerrectoría Adjunta de Investigación y Doctorado – VRAID at the Pontificia Universidad Católica de Chile, for partially supporting this research under grant INICIO 06/2007. The second author was partially supported by grant Fondo Nacional de Desarrollo Científico y Tecnológico – FONDECYT 1060721. The third author was partially supported by grants Fondo Nacional de Desarrollo Científico y Tecnológico – FONDECYT 1060729 and Laboratorio de Análisis Estocástico PBCT-ACT13.

REFERENCES

ABRAMOWITZ, M. AND STEGUN, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York.

CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. AND CRAINICEANU, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective* (2nd ed), London: Chapman and Hall.

DE LA CRUZ-MESÍA, R. AND QUINTANA, F. A. (2007). A model.based approach to Bayesian classification with applications to predicting pregnancy outcomes from longitudinal β–hCG profiles. *Biostatistics* **8**, 228–238.

DE LA CRUZ-MESÍA, R., QUINTANA, F. A. AND MÜLLER, P. (2007). Semiparametric Bayesian classification with longitudinal markers. *Journal of the Royal Statitical Society, Series C (Applied Statistics)* **56**, 119–137.

DEGRUTTOLA, V. AND TU, X. M. (1994). Modeling progression of cd4.lymphocyte count and its relationship to survival time. *Biometrics* **50**, 1003–1014.

GOLUB, G. H. AND WELSCH, J. H. (1969). Calculation of gaussian quadrature rules. *Mathematical Computing* **23**, 221–230.

GUO, X. AND CARLIN, B. P. (2004). Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician* **58**, 1–9.

HENDERSON, R., DIGGLE, P. J. AND DOBSON, A. (2002). Joint modeling of longitudinal measurements and event time data. *Biostatistics* **1**, 465–480.

LEONARD, T., HSU, J. S. J. AND TSUI, K. W. (1989). Bayesian marginal inference. *Journal of the American Statistical Association* **84**, 1051–1058.

LI, E., WANG, N. AND WANG, N.-Y. (2007). Joint models for a primary endpoint and multiple longitudinal covariate processes. *Biometrics* **63**, 1068–1078.

LI, E., ZHANG, D. AND DAVIDIAN, M. (2004). Conditional estimation for generalized linear models when covariates are subject-specific parameters in a mixed model for longitudinal measurements. *Biometrics* **60**, 1-7.

LI, E., ZHANG, D. AND DAVIDIAN, M. (2007). Likelihood and pseudolikelihood methods for semiparametric joint models for a primary endpoint and longitudinal data. *Computational Statistics and Data Analysis* **51**, 5776-5790.

MARSHALL, G. AND BARÓN, A. E. (2000). Linear discriminant models for unbalanced longitudinal data. *Statistics in Medicine* **19**, 1969-1981.

PINHEIRO, J. C. AND BATES, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed effects models. *Journal of Computational and Graphical Statistics* **4**, 12-35.

STEFANSKI, L. A. AND CARROLL, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement models. *Biometrika* **74**, 703-716.

TIERNEY, L. AND KADANE, J. B. (1986). Accurate approximation for posterior moments and marginal densities. *Journal of the American Statistical Association* **81**, 82-86.

TSIATIS, A. A. AND DAVIDIAN, M. (2004). Joint modeling of longitudinal and timeto-event data: An overview. *Statistica Sinica* **14**, 809-834.

VERBEKE, G. AND LESAFFRE, E. (1997). The effect of misspecifying the randomeffects distribution in linear mixed effects models for longitudinal data. *Computational Statistics and Data Analysis* **23**, 541-556.

VONESH, E. F. AND CHINCHILLI, V. M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements.* Marcel Dekker.

VONESH, E. F., GREENE, T. AND SCHLUCHTER, M. D. (2006). Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in Medicine* **25**, 143-163.

WANG, C. Y., WANG, N. AND WANG, S. (2000). Regression analysis when covariates are regression parameters of a random effects model for observed longitudinal measurements. *Biometrics* **56**, 487-495.

WU, L., HU, X. J. AND WU, H. (2008). Joint inference for nonlinear mixed-effects models and time to event at the presence of missing data. *Biostatistics* **9**, 308-320.

WULFSOHN, M. S. AND TSIATIS, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330-339.