

# Discriminant Analysis for Longitudinal Data with Multiple Continuous Responses and Possibly Missing Data

Guillermo Marshall<sup>1,2,3,\*</sup>, Rolando De la Cruz-Mesía<sup>2,\*</sup>,  
Fernando A. Quintana<sup>1,\*</sup>, and Anna E. Barón<sup>3,\*</sup>

<sup>1</sup>Departamento de Estadística, Facultad de Matemáticas,  
Pontificia Universidad Católica de Chile, Casilla 306, Correo 22, Santiago, CHILE.

<sup>2</sup>Departamento de Salud Pública, Facultad de Medicina,  
Pontificia Universidad Católica de Chile, Casilla 114D, Marcoleta 434, Santiago, CHILE.

<sup>3</sup>Department of Preventive Medicine and Biometrics,  
University of Colorado at Denver and Health Sciences Center,  
4200 E. 9th Ave., B-119, Denver, Colorado 80262, U.S.A.

\**email*: {gm,quintana}@mat.puc.cl; rolando@med.puc.cl; anna.baron@uchsc.edu

## SUMMARY

Multiple outcomes are often used to properly characterize an effect of interest. This paper discusses model-based statistical methods for the classification of units into one of two or more groups where, for each unit, repeated measurements over time are obtained on each outcome. We relate the observed outcomes using multivariate nonlinear mixed-effects models to describe evolutions in different groups. Due to its flexibility, the random-effects approach for the joint modeling of multiple outcomes can be used to estimate population parameters for a discriminant model that classifies units into distinct predefined groups or populations. Parameter estimation is done via the EM algorithm with a linear approximation step. We conduct a simulation study that sheds light on how the linear approximation affects the classification results. An example is presented using data from a study in 161 pregnant women in Santiago, Chile in order to predict normal versus abnormal pregnancy outcomes.

**KEY WORDS:** Discriminant analysis; Joint modeling; Missing data; Multivariate responses; Nonlinear mixed-effects models.

## 1. Introduction

Multivariate longitudinal data arise when a set of different responses on the same individual are measured repeatedly over time. Monitoring the extent or severity of disease over time using several clinical parameters is a common practice in the process of medical decision making. For instance patients meeting some threshold level of severity may be offered new or more aggressive treatments. Our motivating example concerns pregnant women. To assess risk factors for and to detect a number of complications during pregnancy, a variety of quantities or characteristics are measured at the prenatal examinations.

Some hormones produced during pregnancy are present in much larger amounts than in the non-pregnant state. In contrast, other hormones are unique in the sense of being present

only during pregnancy. The endocrine changes that a pregnant women undergoes are usually related to pregnancy maintenance and mainly for the benefit of the fetus, whose metabolic needs vary greatly during gestation. The maternal endocrine and metabolic environment must adapt to these varying fetal requirements.

First trimester miscarriages are common in both unassisted pregnancies and pregnancies that result from treatment with reproduction enhancing techniques. A reliable and inexpensive diagnostic test to differentiate between viable pregnancies and pregnancies with eventual early adverse outcome might reduce the psychological tension and anxiety present in many of these patients, and also reduce the cost by making the treatment more effective. On the other hand, a more careful follow up might reduce the risks associated with abnormal pregnancies for patients who are in a higher risk group according to such a test.

Ultrasound examination is effective for evaluation of ongoing pregnancies, but a gestational sac is not reliably visible until 33–37 days after the luteinizing hormone surge (Shapiro et al., 1992). As a result of this inability of ultrasound to identify very early pregnancy abnormalities, there is an ongoing effort to institute a method that can forecast pregnancy outcome. Various studies have investigated hormones like estradiol, beta-subunit human chorionic gonadotropin ( $\beta$ -HCG), among others, and their relationship to pregnancy outcome after in-vitro fertilization (see e.g. Yamashita et al., 1989).

For continuous longitudinal data, when only a single outcome is observed, extensions of classical discriminant analysis to longitudinal data have been considered. Verbeke and Lesaffre (1996) proposed a linear mixed-effects model with random-effects sampled from a mixture of normal distributions. Verbeke and Molenberghs (2000) indicated that the classification rule implied by Verbeke and Lesaffre’s model is equivalent to the discriminant function proposed by Tomasko, Helms and Snapinn (1999). Further developments along this direction have been discussed in Brant et al. (2003) and Wernecke et al. (2004). De la Cruz-Mesía and Quintana (2007) used Bayesian methods and Marshall and Barón (2000) developed a mixed-effects model for classification of hormone trajectories into pre-defined groups. With the goal of identifying Olympic athletes who use growth hormone injections, Brown, Kenward and Bassett (2000) developed a Bayesian method that defines trajectory classes based on a training dataset with known classification. Recently, De la Cruz-Mesía, Quintana and Müller (2007) developed a semiparametric Bayesian approach for classification of longitudinal markers. They defined a suitable extension of hierarchical models to allow such classification.

Some work has been done on longitudinal data with multiple outcomes using multivariate nonlinear mixed-effects models (M-NLMMs). Fitting separate models to each outcome is unsatisfactory because no correlation across responses is considered. By exploiting the correlation structure with a multivariate model, efficiency and power could be greatly increased (Marshall et al., 2006). The methodology proposed for parameter estimation in M-NLMMs is based on first-order (FO) approximations. Marshall et al. (2006) developed a Taylor series linear approximation to the marginal means and variance-covariance matrices of a M-NLMM, describing an EM-type algorithm for parameter estimation. Their EM algorithm extends the method described in Hirst et al. (1991) for parameter estimation in univariate

(single response variable) NLMM using the FO approximation obtained by expanding the conditional mean about the average random effect. The latter proposal was later generalized by Young et al. (1992) to include the Lindstrom and Bates (1990) FO approximation obtained by considering an expansion around the posterior mode of random-effects. Also, Hall and Clutter (2004) used the EM algorithm based on linear approximations to estimate multivariate multilevel nonlinear mixed-effects models.

The main objective of this article is to explore a classification technique for predicting class membership on the basis of a longitudinally measured multivariate response. The inference problem is formally described as a discriminant analysis based on a multivariate nonlinear mixed-effects model for longitudinal data. Additionally, missing data often occur in longitudinal studies because individuals miss some of their regular appointments or because some variables may not be measured at particular visits. Thus, we also consider the case where only a subset of the  $r$  responses may be observed at any occasion. Our approach provides the posterior probability of belonging to one of  $k$  classes based on having  $r$  different responses measured repeatedly over time. In the context of our motivating example, physicians would be able to make treatment or intervention decisions on the basis of these probabilities. Further, we consider a flexible and realistic data structure that allows dealing with joint modeling and classification for patients with very few or just one observation. This is especially relevant for the motivating example where 62 per cent of the patients had two or fewer measurements.

In our approach the classes or groups are predefined and the task is to understand the basis for the classification from a set of labeled units (training dataset). This information is then used to classify future units. In the case where classes or groups are unknown a priori and need to be estimated from the data, latent class models offer a fruitful approach. See additional developments along this direction in Lin et al. (2000), Muthén et al. (2002), and references therein.

The paper is organized as follows. In Section 2 we describe the statistical methodology for the classification of multivariate longitudinal data using multivariate nonlinear mixed-effects models, briefly discussing EM-type algorithms for parameter estimation. We illustrate the proposed multivariate longitudinal method in Section 3 using data from Santiago, Chile on  $\beta$ -HCG and estradiol measured in women with normal and abnormal pregnancy outcomes. In Section 4 we evaluate the performance of the classification procedure with two estimation methods using simulated data. Finally, we summarize and discuss implications in Section 5.

## 2. Discriminant Analysis with Multivariate Longitudinal Data

### 2.1 Model specification

Suppose that, for the  $i$ th of  $m$  units, we observe data at  $n_i$  time points. At the  $j$ th time point,  $t_{ij}$  ( $j = 1, \dots, n_i$ ), we have  $r$  continuous responses  $y_{ijk}$  ( $k = 1, \dots, r$ ). Now, let  $\mathbf{Y}_i = [\mathbf{y}_{i1}, \mathbf{y}_{i2}, \dots, \mathbf{y}_{in_i}]$  be the response matrix for unit  $i$  where  $\mathbf{y}_{ik}$  is an  $n_i \times 1$  vector response for variable  $k$ . Similarly, let  $\mathbf{E}_i = [\boldsymbol{\epsilon}_{i1}, \boldsymbol{\epsilon}_{i2}, \dots, \boldsymbol{\epsilon}_{in_i}]$  be the matrix of error terms associated with  $\mathbf{Y}_i$ . Let  $\mathbf{y}_i = \text{vec}(\mathbf{Y}_i)$  and  $\boldsymbol{\epsilon}_i = \text{vec}(\mathbf{E}_i)$  denote a stacked  $rn_i \times 1$  vector of all the response variables for unit  $i$  and error terms, respectively. Most stochastic models for

serial measurements can be classified either as full multivariate or multi-stage random-effects models. In the full multivariate model, it is assumed that each vector of multiple responses  $\mathbf{y}_i$ , inside the  $\ell$ th of  $g$  groups or populations, is multivariate normal with mean  $\boldsymbol{\mu}_{i\ell}$  ( $rn_i \times 1$ ) and an arbitrary  $rn_i \times rn_i$  dispersion matrix  $\mathbf{R}_\ell$ . The mean vector may depend upon the pattern of observations and also upon covariates.

When the design is balanced but observations are missing at random, traditional multivariate discriminant analysis based on the full multivariate model can be easily applied by using multivariate methods for missing observations (Dempster, Laird and Rubin, 1977; Schafer, 1997). However, when units are measured at arbitrary or unique times, or when the dimension of  $\mathbf{R}_\ell$  is large, this approach becomes unattractive, since a full multivariate model with unrestricted dispersion matrix requires a proliferation of variance parameters, many of which will be poorly estimated. In addition, the full multivariate model does not permit the definition and estimation of (random) unit-specific characteristics (Laird and Ware, 1982).

Two-stage random-effects models are based on explicit identification of unit-specific and population characteristics, and their form extends naturally to the unbalanced situation. The majority of work on methods for longitudinal data with multiple responses has focused on data that can be modeled by means of an expectation function that is assumed linear in its parameters (see, Shah, Laird and Schoenfeld, 1997; Schafer and Yucel, 2002; Fieuws and Verbeke, 2004; O'Brien and Fitzmaurice, 2005; and references therein). However, in many situations, we are concerned with data for which the assumption of normal errors is tenable but the proposed expectation function is nonlinear.

Let us consider in each group  $\ell$  ( $\ell = 1, \dots, g$ ) a M-NLMM, that is, the model for the  $rn_i \times 1$  multiple responses vector of the  $i$ th unit in group  $\ell$  can be formulated as

$$\mathbf{y}_i = \boldsymbol{\mu}_i(\boldsymbol{\phi}_{i\ell}, \mathbf{v}_i) + \boldsymbol{\epsilon}_{i\ell}, \quad i = 1, \dots, m, \quad (1)$$

where  $\boldsymbol{\mu}_i$  is a nonlinear vector-valued, differentiable function of a vector-valued mixed-effects parameter  $\boldsymbol{\phi}_{i\ell}$  and vector of covariates  $\mathbf{v}_i$ , and  $\boldsymbol{\epsilon}_{i\ell}$  is a vector containing the usual error components. The mixed-effects parameter  $\boldsymbol{\beta}_i$  can be incorporated into the model as  $\boldsymbol{\phi}_{i\ell} = \mathbf{X}_i \boldsymbol{\beta}_\ell + \mathbf{Z}_i \mathbf{b}_{i\ell}$ , where  $\mathbf{b}_{i\ell}$  is a  $q \times 1$  random-effects vector specific to the  $i$ th unit, and  $\mathbf{Z}_i$  is the associated design matrix. The fixed-effects design matrix and  $p$ -dimensional parameter vector specific to group  $\ell$  are  $\mathbf{X}_i$  and  $\boldsymbol{\beta}_\ell$ , respectively. We assume that  $\mathbf{b}_{i\ell} \sim \text{MVN}(\mathbf{0}, \mathbf{B}_\ell)$  where  $\mathbf{B}_\ell$  is a  $q \times q$  positive-definite covariance matrix,  $\boldsymbol{\epsilon}_{i\ell} \sim \text{MVN}(\mathbf{0}, \mathbf{R}_{i\ell})$ . Here,  $\mathbf{R}_{i\ell}$  is an  $rn_i \times rn_i$  covariance matrix of the error terms that depends on  $i$  through its dimension  $rn_i$  but with a corresponding set of unknown parameters that does not.  $\mathbf{B}_\ell$  is the covariance matrix of the random-effects in group  $\ell$  and allows for covariance between the random-effects within a given response variable as well as covariance among the random-effects of different response variables. The  $\mathbf{R}_{i\ell}$  covariance matrix in group  $\ell$  has a specified structure to reflect the multivariate nature of the data. We assume that the  $\boldsymbol{\epsilon}_{i\ell}$  at different time points are independent but that the multivariate responses at a particular time point are correlated with a  $r \times r$  covariance matrix  $\boldsymbol{\Sigma}_\ell$ , which is the same for all time points. Consequently,  $\text{Var}(\boldsymbol{\epsilon}_{i\ell}) = \mathbf{R}_{i\ell} = \boldsymbol{\Sigma}_\ell \otimes \mathbf{I}_{rn_i}$  where  $\otimes$  denotes the Kronecker products. In this model, the marginal distribution of  $\mathbf{y}_i$  can be difficult to find even in the case where the conditional distribution of  $\mathbf{y}_i$  given  $\mathbf{b}_{i\ell}$  is normal and the marginal distribution of  $\mathbf{b}_{i\ell}$  is normal.

When the random-effects are linear in the scale of the response vector the model can be formulated as

$$\mathbf{y}_i = \boldsymbol{\mu}_i(\boldsymbol{\beta}_\ell, \mathbf{v}_i) + \mathbf{Z}_i(\boldsymbol{\beta}_\ell)\mathbf{b}_{i\ell} + \boldsymbol{\epsilon}_{i\ell}, \quad (2)$$

where  $\mathbf{Z}_i(\boldsymbol{\beta}_\ell)$  is an  $rn_i \times q$  matrix of known functions of unknown parameters  $\boldsymbol{\beta}_\ell$ . In this case, the marginal distribution of the multiple response vector  $\mathbf{y}_i$  is given by

$$\mathbf{y}_{i|\text{group } \ell} \sim \text{MVN}(\boldsymbol{\mu}_i(\boldsymbol{\beta}_\ell, \mathbf{v}_i), \boldsymbol{\Psi}_{i\ell}) \quad (3)$$

with  $\boldsymbol{\Psi}_{i\ell} = \mathbf{Z}_i(\boldsymbol{\beta}_\ell)\mathbf{B}_\ell\mathbf{Z}_i'(\boldsymbol{\beta}_\ell) + \boldsymbol{\Sigma}_\ell \otimes \mathbf{I}_{n_i}$  and the corresponding density is denoted by  $f_{\ell i}(\mathbf{y}_i)$ .

Given prior probabilities  $\pi_\ell$ ,  $\ell = 1, \dots, g$ , for the  $g$  groups, and choosing a zero-one loss function (which minimizes the average error probability), the Bayes classification rule can be written:

$$\eta(\mathbf{y}_i) = \arg \max_{\ell=1, \dots, g} p_{\ell i}(\mathbf{y}_i) \quad (4)$$

where  $p_{\ell i}(\mathbf{y}_i)$  is the posterior probability of membership in group  $\ell$ , i.e.,

$$p_{\ell i}(\mathbf{y}_i) = \frac{\pi_\ell f_{\ell i}(\mathbf{y}_i)}{\sum_s \pi_s f_{s i}(\mathbf{y}_i)}. \quad (5)$$

Each unit is classified into the group for which the highest estimated posterior probability of membership is achieved. This is an optimal allocation rule based on the Neyman-Pearson lemma. Note that, although the M-NLMM (3) specifies the multiple response vector conditionally on a vector  $\mathbf{b}_{i\ell}$  of random-effects, classification is based on the marginal distribution obtained from integrating over the random-effects. In the case of model (1) the random-effects  $\mathbf{b}_{i\ell}$  are part of the nonlinear component. Even in the case where these random parameters are normally distributed, the resulting marginal distribution of the multiple responses vector  $\mathbf{y}_i$  is typically unknown and difficult to find analytically. In that case, a linear approximation to the model residuals yields a marginal distribution of the unit-specific observations vector that is approximately normal, i.e.,

$$\mathbf{y}_{i|\text{group } \ell} \overset{\sim}{\sim} \text{MVN}(\boldsymbol{\mu}_i(\boldsymbol{\beta}_\ell, \mathbf{b}_{i\ell} = \mathbf{0}), \boldsymbol{\Psi}_{i\ell}), \quad \text{with} \quad (6)$$

$$\boldsymbol{\Psi}_{i\ell} = \tilde{\mathbf{Z}}_i(\boldsymbol{\beta}_\ell, \mathbf{0})\mathbf{B}_\ell\tilde{\mathbf{Z}}_i'(\boldsymbol{\beta}_\ell, \mathbf{0}) + \boldsymbol{\Sigma}_\ell \otimes \mathbf{I}_{n_i} \quad (7)$$

where  $\tilde{\mathbf{Z}}_i(\boldsymbol{\beta}_\ell, \mathbf{0})$  is the Jacobian matrix  $\partial\boldsymbol{\mu}_i(\boldsymbol{\beta}_\ell, \mathbf{b}_{i\ell})/\partial\mathbf{b}_{i\ell}$  evaluated at  $\mathbf{b}_{i\ell} = \mathbf{0}$ .

In the remainder of this section the discussion will be based on a generic patient and thus we simplify the notation by dropping the subindex  $i$ . In classical discriminant analysis the Mahalanobis distance plays a central role in both the conceptual framework and the allocation rules. The squared Mahalanobis distance between the multiple response vector  $\mathbf{y}$  and the mean of the distribution of population  $\ell$ ,  $\boldsymbol{\mu}_\ell$ , with respect to  $\boldsymbol{\Psi}_\ell$  is

$$D_\ell(\mathbf{y}) = (\mathbf{y} - \boldsymbol{\mu}_\ell)' \boldsymbol{\Psi}_\ell^{-1} (\mathbf{y} - \boldsymbol{\mu}_\ell).$$

Having defined the Mahalanobis distance, the classification rule is to allocate  $\mathbf{y}$  to population  $s$  if  $\lambda_{\ell s}(\mathbf{y}) \leq 0$  for  $\ell = 1, \dots, g$  and  $\ell \neq s$ , where

$$\lambda_{\ell s}(\mathbf{y}) = D_s^*(\mathbf{y}) - D_\ell^*(\mathbf{y}) + 2 \log \frac{\pi_\ell}{\pi_s}, \quad (8)$$

with  $D_\ell^*(\mathbf{y}) = D_\ell(\mathbf{y}) + \log |\Psi_\ell|$ .

As elaborated in Marshall and Barón (2000), four discriminant models are possible according to the form of the variance of  $\mathbf{y}$  in the population  $\ell$ . Let  $\boldsymbol{\theta}$  denote the vector of all variance and covariance parameters found in  $\Psi$ , that is,  $\boldsymbol{\theta}$  consists of the different elements in  $\mathbf{B}$  and of all parameters in  $\Sigma$ . The variance of  $\mathbf{y}$  in the population  $\ell$ ,  $\Psi_\ell = \Psi(\boldsymbol{\beta}_\ell, \boldsymbol{\theta}_\ell) = \mathbf{Z}(\boldsymbol{\beta}_\ell)\mathbf{B}(\boldsymbol{\theta}_\ell)\mathbf{Z}'(\boldsymbol{\beta}_\ell) + \Sigma(\boldsymbol{\theta}_\ell) \otimes \mathbf{I}_n$ , is a function of the mean population-specific parameters  $\boldsymbol{\beta}_\ell$  and the variance components  $\boldsymbol{\theta}_\ell$ . For model (1),  $\mathbf{Z}(\boldsymbol{\beta}_\ell) = \tilde{\mathbf{Z}}(\boldsymbol{\beta}_\ell, \mathbf{0})$ . The *homoscedastic model* is obtained when  $\mathbf{Z}(\boldsymbol{\beta}_\ell) = \mathbf{Z}$  does not depend on the mean parameters  $\boldsymbol{\beta}_\ell$  and the variance components are homogeneous, that is,  $\boldsymbol{\theta}_\ell = \boldsymbol{\theta}$  for  $\ell = 1, 2, \dots, g$ . In this situation  $\Psi_\ell = \Psi$ . In particular, when  $g = 2$  the above classifier implies classification of a unit with multiple response vector  $\mathbf{y}$  in the first group, if and only if

$$\{\mathbf{y} - (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)/2\}'\Psi^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) > \log(\pi_2/\pi_1)$$

which is the linear discriminant function.

The *mean-heteroscedastic model* consists of a model in which the design matrix  $\mathbf{Z}(\boldsymbol{\beta}_\ell)$  depends on the mean parameters  $\boldsymbol{\beta}_\ell$  but the variance components  $\boldsymbol{\theta}_\ell = \boldsymbol{\theta}$  remain homogeneous among all populations  $\ell = 1, 2, \dots, g$ . In the *variance-heteroscedastic model* the design matrix  $\mathbf{Z}(\boldsymbol{\beta}_\ell)$  does not depend on the population parameters  $\boldsymbol{\beta}_\ell$ , but the variance components  $\boldsymbol{\theta}_\ell$  differ across populations  $\ell = 1, 2, \dots, g$ . This is the case where the between-unit variances  $\mathbf{B}_\ell$  are different among the groups, or the within-unit variances  $\Sigma_\ell$  vary among the  $g$  populations, or both covariance matrices are different. For the *fully-heteroscedastic model* the design matrix  $\mathbf{Z}(\boldsymbol{\beta}_\ell)$  depends on the population parameters  $\boldsymbol{\beta}_\ell$ , and the variance components  $\boldsymbol{\theta}_\ell$  differ across populations  $\ell = 1, 2, \dots, g$ . We will perform our own comparison of these four models later in Section 3.

Note that the above quantities, and hence the classification rule, will be different for each individual, insofar as they depend on unit-specific observation times.

## 2.2 Classification with Missing Data

The above discussion assumes that all  $r$  responses are observed at each occasion, although the number and timing of observations may differ from unit to unit; that is, the data are complete in responses but unbalanced in number of observations per experimental unit. In clinical trials missing data are common, in which case only a subset of the  $r$  responses may be observed at any occasion. We will now show how the classification rule (8) can be easily adapted to cover the case of responses that are missing completely at random (Little and Rubin, 1987).

Let  $\mathbf{y}_{\text{new}}$  denote the currently available partial responses vector for the new unit. Let  $\mathbf{S}$  be the matrix of zeros and ones which ‘selects’ the elements of  $\mathbf{y}_{\text{new}}$  which are actually observed; that is, the product  $\mathbf{S}\mathbf{y}_{\text{new}} = \mathbf{y}_{\text{new}}^o$  gives the observed components of  $\mathbf{y}_{\text{new}}$ . If all components of  $\mathbf{y}_{\text{new}}$  are observed, then  $\mathbf{S}$  is the  $rn^* \times rn^*$  identity matrix, where  $n^*$  is the number of measurements per response.

In the presence of missing data, the classification rule (8) implies allocating  $\mathbf{y}_{\text{new}}$  to

population  $s$  if  $\lambda_{\ell s}(\mathbf{y}_{\text{new}}^o) \leq 0$  for  $\ell = 1, \dots, g$  and  $\ell \neq s$ , where

$$\lambda_{\ell s}(\mathbf{y}_{\text{new}}^o) = D_s^*(\mathbf{y}_{\text{new}}^o) - D_\ell^*(\mathbf{y}_{\text{new}}^o) + 2 \log \frac{\pi_\ell}{\pi_s}, \quad (9)$$

and

$$D_\ell^*(\mathbf{y}_{\text{new}}^o) = (\mathbf{y}_{\text{new}}^o - \boldsymbol{\mu}_\ell^o)' \boldsymbol{\Psi}_\ell^{o^{-1}} (\mathbf{y}_{\text{new}}^o - \boldsymbol{\mu}_\ell^o) + \log |\boldsymbol{\Psi}_\ell^o|,$$

with  $\boldsymbol{\mu}_\ell^o = \mathbf{S}\boldsymbol{\mu}_\ell$ , and  $\boldsymbol{\Psi}_\ell^o = \mathbf{S}\boldsymbol{\Psi}_\ell\mathbf{S}'$ .

### 2.3 Parameter Estimation

The algorithms proposed for computing the maximum likelihood estimates (MLE) of  $(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{B})$  and empirical Bayes estimators (predictors) for the random-effects  $\mathbf{b}_i$  for M-NLMMs, rely on iteratively linearizing the conditional mean function and solving the resulting multivariate linear mixed-effects model (M-LMM). Marshall et al. (2006) extend the basic model to handle multivariate responses and used an EM-type algorithm to estimate the model parameters, also considering the case when missing data are present. They also show how to implement a variant of their algorithm using SAS Proc NLMIXED with the first-order approximation method of Beal and Sheiner (1988), which at each iteration evaluates the random-effects at  $\mathbf{b} = \mathbf{0}$  rather than at the current values  $\mathbf{b} = \hat{\mathbf{b}}$ . See details in Marshall et al. (2006).

Also, SAS Proc NLMIXED performs exact MLE using numerical integration. In principle, this approach could be extended to the multivariate case using the trick implemented in Marshall et al. (2006). In this case, the marginal density of the observation vector can be obtained using Monte Carlo integration. Details for implementing this procedure are given in Section 4.

Minor modifications to the algorithm developed in Marshall et al. (2006) are necessary to estimate the multivariate models described at the end of Section 2.1. Specifically, if some or all components of  $\boldsymbol{\beta}$  are hypothesized to be group-specific, the design matrix  $\mathbf{X}$  in the linearized version of M-NLMM is appropriately changed. When the variance components  $\boldsymbol{\Sigma}$  and  $\mathbf{B}$  differ across groups, we estimate them separately. All of these models can be straightforwardly estimated using SAS Proc NLMIXED. Details are provided in the next Section.

## 3. Data Analysis

We now apply the previously discussed methods to the motivating dataset. It is well known in obstetrics that  $\beta$ -HCG and estradiol are clinical variables that show dramatic changes in women during pregnancy (Yamashita et al., 1989). It has been also established that values of these variables are different in women who have normal pregnancies with terminal deliveries than in women who have spontaneous abortions or other types of adverse pregnancy outcomes. This association has made it possible to classify, with some degree of uncertainty, the outcome of pregnancy.

We consider here a follow-up study of 161 young women, representing different pregnancies over a period of 2 years in a private fertilization obstetrics clinic in Santiago, Chile. Estradiol and  $\beta$ -HCG concentrations on the 161 women were measured during the first

trimester of pregnancy. One of the main objectives of the study was to evaluate these concentrations at early stages of pregnancy, with the purpose of identifying women with a high risk of loss. Consequently, pregnancy outcomes were divided into two groups: normal and abnormal. The women were classified as normal pregnancies if they had a normal delivery, or as abnormal pregnancies if they had any complication resulting in a non-terminal delivery and loss of the fetus. The resulting dataset consists of 124 patients diagnosed with normal pregnancy and 37 patients with abnormal pregnancy outcome.

The responses that we analyze are the vectors of time-varying estradiol and  $\beta$ -HCG measurements for the 161 women. The 161 women altogether contribute a total of 348 observations per response, where the number the samples per woman ranged from 1 through 6, with median 2. There was some missing data but there was no reason to believe that the missingness was non-ignorable. Throughout we assume that the missing responses are missing completely at random. Missingness rates for the responses for normal and abnormal groups are 3% and 0% for  $\beta$ -HCG and 27% and 58% for estradiol, respectively. These data were originally presented in Marshall et al. (2006).

Figure 1 presents the patient-specific profiles of estradiol and  $\beta$ -HCG on the  $\log_{10}$  scale for both groups. The two populations appear clearly distinct when considering the ensemble of profiles. However, for any one of these profiles the classification into one or the other sub-population is far less certain, in particular when considering a series of partial responses. Our main inference goal in analyzing these data is to provide a classification rule for a new patient. The rule should allow sequential updating as data accrues for the new patient.

[Figure 1 about here.]

For the period of observation, roughly the first trimester of pregnancy, the mean gestational ages (days) in women with normal and abnormal pregnancies, were 34.5, and 32.6 days, respectively; thus, no significant difference in gestational age was found among the groups. We found that concentrations of estradiol and  $\beta$ -HCG were significantly lower in women with an abnormal pregnancy than in those with a normal pregnancy. In addition, we assumed prior probabilities of group membership to be proportional to the size of the groups in the training sample.

Marshall et al. (2006) proposed a nonlinear mixed and a linear mixed model to analyze the evolution of  $\beta$ -HCG and estradiol responses, respectively. The models were validated by fitting a series of more complex models and comparing them with respect to the corresponding values of minus twice the log-likelihoods. To test for differences between the two groups, they compared five alternatives and the best among these was the one having a different curve for each group and set of responses. We adopt the same model here. Letting  $y_{i1\ell}$  and  $y_{i2\ell}$  denote the  $\beta$ -HCG and estradiol responses, respectively, for patient  $i$  in group  $\ell$  taken at time  $t$ , the models are specified as

$$y_{i1\ell}(t) = \frac{\beta_{\ell 1} + b_{i1\ell}}{1 + \exp\{(\beta_{\ell 2} - t)/\beta_{\ell 3}\}} + e_{i1\ell}(t) \quad (10)$$

$$y_{i2\ell}(t) = \beta_{\ell 4} + \beta_{\ell 5}t + b_{i2\ell} + e_{i2\ell}(t) \quad (11)$$

in which  $t$  is time expressed in days and  $\ell = 1$  for the normal pregnancy and  $\ell = 2$  for the abnormal pregnancy group. The vector  $(\beta_{\ell 1}, \beta_{\ell 2}, \beta_{\ell 3}, \beta_{\ell 4}, \beta_{\ell 5})'$  of fixed effects describes the average evolution of the responses in group  $\ell = 1, 2$  and the vector  $(b_{i1\ell}, b_{i2\ell})$  of random-effects describes how the profile of the  $i$ th patient deviates from the average profiles in groups  $\ell = 1, 2$ . Both response trajectories are tied together through a joint distribution for the random-effects

$$\mathbf{b}_{i\ell} = [b_{i1\ell}, b_{i2\ell}]' \sim \text{MVN}(\mathbf{0}, \mathbf{B}_\ell),$$

where  $\mathbf{B}_\ell$ , the covariance matrix of the random-effects, has elements given by  $\left(B_{rs}^{(\ell)}\right)_{r=1,2;s=1,2}$ . The error components are correlated and not associated with the random-effects

$$\boldsymbol{\epsilon}_{i\ell} = [e_{i1\ell}, e_{i2\ell}]' \sim \text{MVN}(\mathbf{0}, \mathbf{R}_{i\ell})$$

with  $\mathbf{R}_{i\ell} = \boldsymbol{\Sigma}_\ell \otimes \mathbf{I}_{n_i}$ , and  $\boldsymbol{\Sigma}_\ell$  has elements  $\left(\Sigma_{rs}^{(\ell)}\right)_{r=1,2;s=1,2}$ . The latter implies that conditional on the random-effects, the response trajectories are dependent. Special cases can be obtained by making additional assumptions about the covariance matrices  $\mathbf{B}_\ell$  and  $\boldsymbol{\Sigma}_\ell$ . For example, if  $B_{12}^{(\ell)}$  and  $\Sigma_{12}^{(\ell)}$  are all equal to zero, the responses in each group are assumed to be completely independent at any point in time. In that case, parameters of the models can be obtained using likelihood based inference with, for example, the NLME software of Pinheiro and Bates (2000).

The marginal distribution of the response vector  $\mathbf{y}_i$  in the  $\ell$ th group is

$$\mathbf{y}_i \sim \text{MVN} \left( \boldsymbol{\mu}_{i\ell}(\mathbf{t}_i), \tilde{\mathbf{Z}}_{i\ell} \mathbf{B}_\ell \tilde{\mathbf{Z}}_{i\ell}' + \boldsymbol{\Sigma}_\ell \otimes \mathbf{I}_{n_i} \right) \quad (12)$$

where the mean vector  $\boldsymbol{\mu}_{i\ell}(\mathbf{t}_i)$  of dimension  $2n_i \times 1$  has elements

$$\mu_{i\ell}(t_{ij}) = \begin{pmatrix} \beta_{\ell 1} / (1 + \exp\{(\beta_{\ell 2} - t_{ij}) / \beta_{\ell 3}\}) \\ \beta_{\ell 4} + \beta_{\ell 5} t \end{pmatrix} \quad (13)$$

and represents the population curve at time  $t_{ij}$ , and  $\tilde{\mathbf{Z}}_{i\ell}$  is a  $2n_i \times 2$  working design matrix with rows made up of  $\tilde{\mathbf{z}}_{ij\ell}$  where

$$\tilde{\mathbf{z}}_{ij\ell} = \begin{pmatrix} v_{ij\ell} & 0 \\ 0 & 1 \end{pmatrix}$$

with  $v_{ij\ell} = 1 / (1 + \exp\{(\beta_{\ell 2} - t_{ij}) / \beta_{\ell 3}\})$  which depends on the values of the unknown population parameters  $\beta_{\ell 2}$  and  $\beta_{\ell 3}$ .

Expressions (12) and (13) constitute a fully-heteroscedastic model. If desired, a mean-heteroscedastic model can be obtained from expression (12) by introducing the restrictions  $\mathbf{B}_1 = \mathbf{B}_2$  and  $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ , and the variance-heteroscedastic model from expression (13) by introducing the restrictions  $\beta_{12} = \beta_{22}$  and  $\beta_{13} = \beta_{23}$ . The homoscedastic model is obtained by introducing both sets of restrictions.

We used the NLME software of Pinheiro and Bates (2000) to fit the models for each response separately to obtain initial estimates of model parameters. In this example, we

found that the better approximation corresponded to an expansion around  $\mathbf{b}_i = \mathbf{0}$  since this yields the correct marginal mean and variance-covariance. Implementing the EM algorithm with an expansion about the posterior mode may introduce some bias as the result of using the derivative matrix evaluated at the posterior mode rather than at  $\mathbf{0}$  (see Vonesh and Chinchilli, 1997). Thus, we used SAS Proc NLMIXED (method=firo) to estimate the final models.

[Table 1 about here.]

The results of fitting the four models are shown in Table 1. Likelihood ratio tests comparing the more restricted models to the fully-heteroscedastic model show that the differences in  $\beta$ -HCG and estradiol between the groups occur not only at the mean level but also at the within-group and patient-specific levels. From Table 1 we conclude that the fully-heteroscedastic model is the best for discriminating between normal and non-terminal deliveries.

Parameter estimates for the four models are shown in Table 2. Observed differences across the groups in the linear and logistic curve parameters reflect the changes in  $-2 \log L$  between, principally, the homoscedastic and mean-heteroscedastic models. Observed differences in the group-specific estimates of the variance components in the variance- and fully-heteroscedastic models suggest that there is significantly more between-patient variability in the abnormal group than in the normal group.

[Table 2 about here.]

Table 3(A) presents the classification results using the original sample. Seven of the 124 women having normal pregnancy were classified as abnormal whereas 16 of the 37 women having abnormal pregnancy were classified as normal. The observed misclassification rate is 14.3% (23/161). Other quantities of interest can be readily evaluated. Among these, “sensitivity” and “specificity” are popular ways to summarize the classification results. Letting  $A = \{ \text{the patient is classified as abnormal} \}$ , and  $I = \{ \text{the patient actually belongs to the abnormal group} \}$ , then sensitivity and specificity are defined respectively as  $\mathbb{P}[A|I]$  and  $\mathbb{P}[\bar{A}|\bar{I}]$ . From our results we found 56.8% sensitivity and 94.4% specificity.

It is well known that the error rate obtained by applying the classifier to the same data from which it has been formed tends to be biased downward as an estimate of the true error rate. For this reason we used the leave-one-out cross-validation to obtain more accurate estimates of the misclassification rate. Table 3(B) presents the corresponding results, with 24 women misclassified and an estimated misclassification rate of 14.9%.

[Table 3 about here.]

The Receiver Operating Characteristic (ROC) curves and the area under the ROC curve (AUC) for one multivariate and two univariate models are presented in Figure 2. Specifically, we present three curves showing the changes in sensitivity and specificity using only the estradiol variable, only the  $\beta$ -HCG variable, and using both the estradiol and  $\beta$ -HCG

variables. Using the bivariate responses improves the sensitivity and specificity for predicting an abnormal pregnancy outcome in this population of women, a result that is expected via the Neyman-Pearson lemma for comparison of the classification performance of a multivariate likelihood ratio based allocation rule vs. the best of the univariate likelihood ratio-based allocation rules (Pepe, 2003).

[Figure 2 about here.]

In the above analysis we have used all of the available information. However, it is interesting to assess the predictive power of our model as a function of the number of observations used or their timing. Thus we generated from the corresponding fitted distributions, one future patient for each group and evaluated the classifier defined through equations (4) and (5) for up to six possible observations. Time points were randomly chosen from the empirical distribution of observed times within each group. Figure 3 shows the evolution of posterior probabilities for classifying one normal and one abnormal pregnancy outcome patient in the future. For the normal patient, we observe a steady growth of the probabilities. For the abnormal patient, this probability decreases to values that leave no question about the classification. It is interesting to note that when using only the  $\beta$ -HCG profiles, the classification probabilities for abnormal patients require more observations than the normal ones to achieve correct classification (see De la Cruz-Mesía and Quintana, 2007 and De la Cruz-Mesía et al., 2007).

[Figure 3 about here.]

#### 4. Simulation Study

In this section we report the results of a simulation study motivated by a problem related to the pregnant women example. We investigate the impact of two estimation methods on the resulting parameter estimates and their influence on the performance of the classification rule. The effects of number of patients and number of measurements per patient are examined. In our example we have only one random effect in model (10), entering the model in a strictly linear fashion. The same is true for the random effect in model (11). Thus, the mean curves coincide with the population-averaged curves (see eq. 12). It is natural, however, to consider the case where the random-effects enter the models in a nonlinear fashion, so that the impact of the above approximation is far less clear on estimation and particularly on our classification rule. We describe next a specific scheme that will shed some light on these (and other) issues.

In the example we have 124 patients in the normal pregnancy group with the number of measurements per patient ranging from 1 to 4 (median 2) and in the abnormal group we have 37 patients with the number of measurements per patient ranging from 1 to 6 (median 2). We judge the number of observations in each group to be sufficiently large.

Our simulations use two groups, with 124 and 37 patients, respectively. We consider three cases for the number of measurements per patient within each group. The first corresponds to choosing the  $n_i$ 's to coincide with the observed values in the pregnancy dataset, the second has  $n_i = 5$ , and the last  $n_i = 10$  measurements per patient, respectively. A logistic

model similar to (10) was used to generate one response, but with random-effects entering in a nonlinear fashion. The other response was generated using model (11). To model the simulated data we use in all cases

$$y_{i1\ell}(t) = \frac{\beta_{\ell 1}}{1 + \exp\{(\beta_{\ell 2} + b_{i2\ell} - t)/\beta_{\ell 3}\}} + e_{i1\ell}(t)$$

$$y_{i2\ell}(t) = \beta_{\ell 4} + \beta_{\ell 5}t + b_{i4\ell} + e_{i2\ell}(t).$$

Assuming a set of values for the model parameters and the group membership probabilities, a training dataset is simulated. On this dataset, parameter estimation for all models is carried out using the first-order and Gaussian quadrature methods. Next, we simulate a validation dataset on which we run the following two classification procedures. The first one uses the MLEs obtained by Gaussian quadrature we implement the Bayes classifier defined through equations (4) and (5). This procedure involves evaluating the density of the observation vector by Monte Carlo integration with a large number of draws (we chose 10,000) from the random-effects distribution. We do this only once per subject. In the second procedure the Bayes classifier is implemented using density (6) with estimates obtained with the first-order approximation method. The entire process is replicated 1,000 times.

[Table 4 about here.]

[Table 5 about here.]

Tables 4 and 5 present the estimation results using the FO approximation and Gaussian quadrature. The column labeled “Estimate” denotes the average values for parameter estimates over the 1,000 replications; “SE” is the standard error of estimates. Our simulation results indicate that the two approximations produce similar estimates of the fixed effects  $\beta$  but yield some differences for the  $B$  parameters in the abnormal group with unbalanced number of measurements per patient. The FO approximation grossly overestimates  $B_{11}$ , the variance of  $b_{i2}$ , and underestimates  $B_{12}$ . When the number of patients is 124, the estimates from the FO and Gaussian approximations were almost unbiased and very similar. The FO approximation, particularly when the number of patients is 37, showed considerably more variation than in the Gaussian case.

We turn now to the assessment of the classification rule. For this, we calculated the misclassification error rate, sensitivity and specificity. Figure 4 shows the results obtained using both procedures (each box in the figures shows the median, quartiles, and extreme values within a category). All of these quantities are expressed as percentages. The results show that the number of measurements per patient does affect classification. This is not surprising because increasing the number of observations leads to a more accurate discrimination function and thus to a better classification procedure. For this reason, we only show the results when the number of measurements per patient is unbalanced and when  $n_i = 5$ . In view of Figure 4, we did not find convincing evidence that the estimation procedures have any significant effect on classification.

[Figure 4 about here.]

It is well known that as  $\mathbf{B} \rightarrow \mathbf{0}$ , the FO method will still yield a reasonably unbiased estimate of both the population parameter and the population mean response (see e.g. Vonesh and Chinchilli, 1997; pp. 352–357; Demidenko, 2004; pp. 455–462). This is what may be happening in the results of Tables 4 and 5, namely that the variance component  $\mathbf{B}$  may be small enough to make the two approximations appear to behave in a very similar fashion. From Table 4, the coefficient of variation (CV) for the nonlinear random-effect is of order 0.193 for the normal group ( $\beta_{22} = 11.6$ ,  $B_{11} = 1.03$ ). This is moderately small but not too small. Likewise, in Table 5, the CV associated with the nonlinear random-effect is 0.0875 and this is considerably smaller. In essence, when  $B_{11}$  is “small”, the random-effects will all be clustered near 0 and one can effectively approximate the likelihood by expanding around  $b_{i2} = 0$ . In order to show other situations, we carried out more simulations considering larger values for the CV associated with the random-effects. In doing so, however, we note that the previous simulation results suggest that, in general, the exact Gaussian quadrature method yields reasonable estimation results. Therefore, in what follows we restrict ourselves to studying the behavior of the FO approximation method.

Tables 6 and 7 show the results obtained when simulating samples with various CV values. We can clearly see that when the CV increases, the linearization approximation involved in the FO method gets progressively worse, especially when we consider 37 patients. In contrast with our previous simulation results, we note that increasing the CV has also a negative effect on the classification procedures (data not shown) basically because of the large variability of the nonlinear random-effects.

In summary, our results suggest that the likelihood approximation using the FO method performs well when the number of the intra-unit measurements is not small and the variability of the random-effects is not large. But when some of the units have either sparse data or a large variability of the associated random-effects, considerable errors are introduced when using the FO method. This may explain why the results of replicating the observed sample sizes from the pregnancy data (denoted as  $n_i = 6^*$  in Tables 5 and 7) differ so much from those obtained when fixing  $n_i = 5$ .

[Table 6 about here.]

[Table 7 about here.]

## 5. Comments and Conclusions

The main purpose of this paper was to propose an analytic approach to the development of discriminant functions when multiple responses are measured over time. Parameter estimates from nonlinear random-effects models for multiple responses are embedded into linear and quadratic discriminant models that are functions of time. We also consider the case when only a subset of the responses may be observed at any occasion.

Using simulation we compared the classification rule based on the squared Mahalanobis distance function, assuming the population mean vectors and variance-covariance matrices

for  $g$  groups of units, with that obtained under Taylor series linear approximations to the marginal means and variance-covariance matrices. We found no significant differences. Our simulation study also suggests that the ML estimates based on the first-order approximation are similar to “exact” ML estimates obtained using Gaussian quadrature when the inter-unit variability is small. From the simulation results we can see that the FO method performs poorly unless  $n_i$  is small and the time points are clustered in the same region of the design space. Otherwise, large variance components may lead to biased or inconsistent estimates of the parameters, and more generally, to a bad performance of the discrimination based on this approach.

It is also worth noting that the proposed procedure involves two Taylor approximations, one at the parameter estimation stage and the other at the classification stage. We explore the effect of each of these approximations on the results separately. When parameter estimation was carried out using exact MLE with the Gaussian quadrature, but the classification was done using the Taylor approximation, similar results were obtained.

The principal advantage of these discriminant models for unbalanced repeated measures is their ability to use all of the available information for classifying individuals over time, regardless of the number or timing of the observations. An area of concern and interest in longitudinal data analysis, but not pursued in this research, is the subject of informative censoring. This occurs when the response trajectory, or outcome, and the censoring mechanism for follow-up are not independent of each other. While the data from our motivating example did not appear to exhibit informative censoring, the general setting from which the study arose could have led to shorter follow-up times on average for the women with abnormal pregnancy outcomes. In that case, the random-effects would be correlated with follow-up time. Further research is called for to investigate the effects of informative censoring on parameter estimates and classification using our proposed models.

A second advantage of these longitudinal discriminant models is that the influence of both between-group variability and components of within-group variability on discrimination can be readily quantified. In the case of a single response, the discriminant model reduces to that proposed by Marshall and Barón (2000).

The multivariate model takes into account the correlation between responses at the same visit, and so it can discriminate much better than its univariate competitors. The results obtained from models incorporating additional biological and clinical information should help both physicians and patients to make more informed treatment decisions on the basis of objective staging data. Indeed, using the tools presented here, a reliable and inexpensive diagnostic test for early differentiation between normal and adverse outcome might reduce the psychological tension and anxiety present in many patients. For patients judged by this test to be at high risk of an unfavorable outcome, a more careful follow-up might lead to better patient management and interventions that reduce risk.

Finally, an extension of the discriminant procedure that would be of particular interest in many applications is the modeling of group membership probabilities as a function of covariates. In the context of our example, this could be accomplished using a logistic form for the population proportion of abnormal pregnancies,  $\pi = e^{\alpha \mathbf{x}} / (1 + e^{\alpha \mathbf{x}})$  where  $\mathbf{x}$  is a

vector of covariates for each woman and  $\alpha$  a vector of regression parameters. Identifying associations between women covariates, such as age, number of previous normal and abnormal pregnancies, smoking status and abnormal pregnancy tendencies, can be useful for targeting specific individuals in future analysis. Also the model for  $\mathbf{y}_i$  could depend on these covariates. In our example a number of women had missing covariate values.

#### ACKNOWLEDGEMENTS

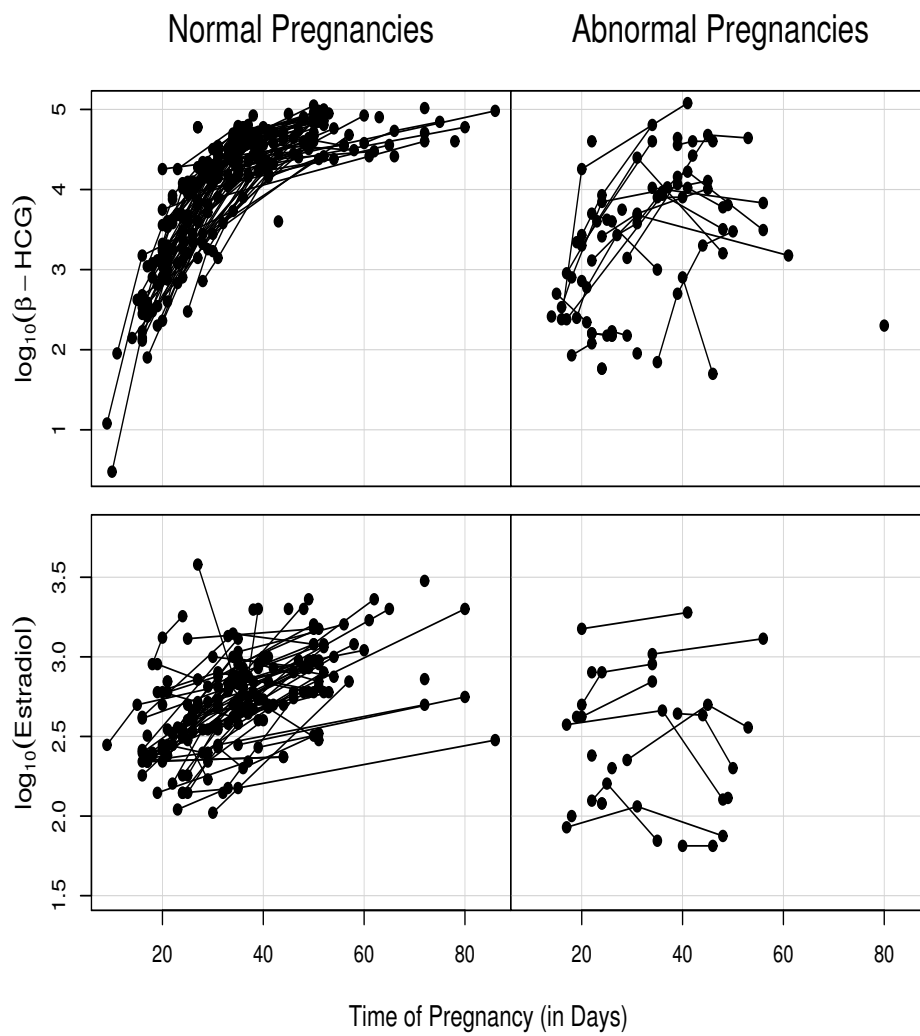
The first and second authors were partially supported by the National Fund for Scientific and Technological Development of Chile; grant number: Fondecyt 1010958 and Fondecyt 3060071, respectively. Also, the second author thanks the Comisión Nacional de Investigación Científica y Tecnológica - CONICYT, for partially supporting his Ph.D. studies at the Pontificia Universidad Católica de Chile. The third author was partially supported by grant Laboratorio de Análisis Estocástico PBCT-ACT13. We would also like to thank the editor, associate editor and the reviewers for their constructive comments, which helped to substantially improve this manuscript.

#### REFERENCES

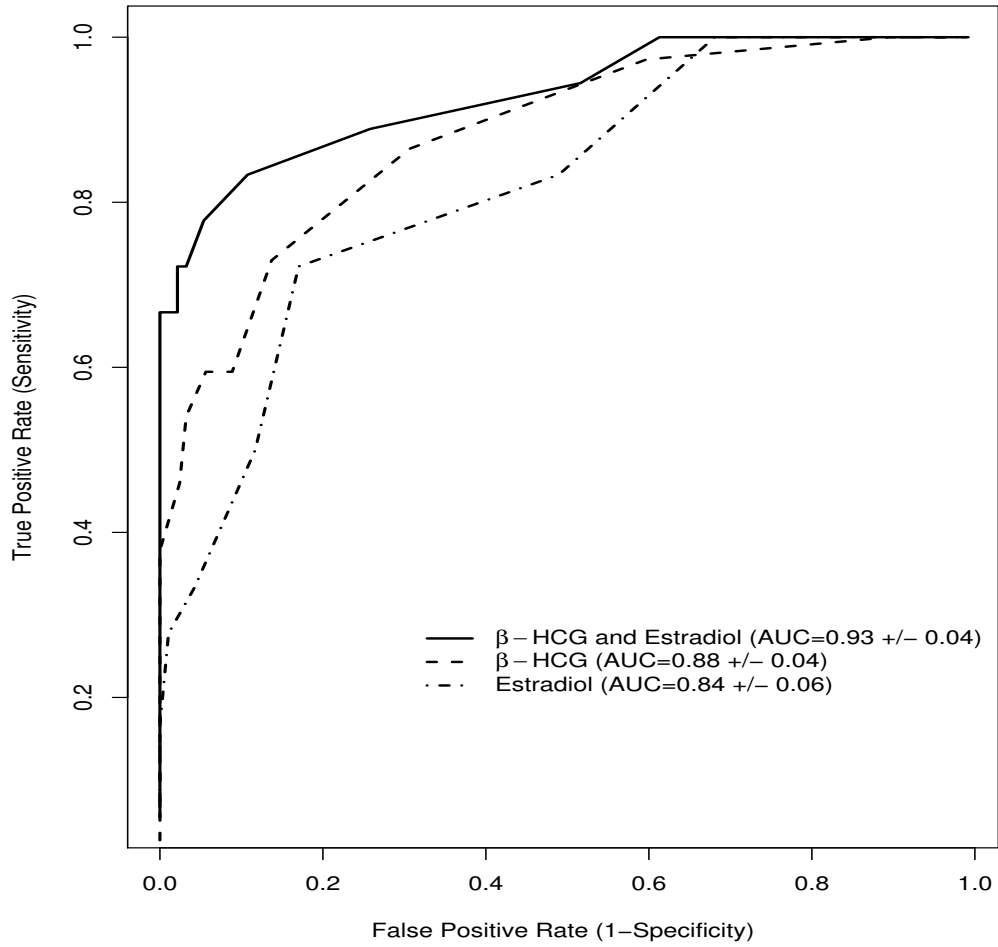
- Beal, S. L. and Sheiner, L. B. (1988). Heteroskedastic nonlinear regression. *Technometrics* **30**, 327–338.
- Brant, L. J., Sheng, S. L., Morrell, C. H., Verbeke, G. N., Lesaffre, E. and Carter, H. B. (2003). Screening for prostate cancer by using random-effects models. *Journal of the Royal Statistical Society, Series A* **166**, 51–62.
- Brown, P. J., Kenward, M. G. and Bassett, E. E. (2000). Bayesian discrimination with longitudinal data. *Biostatistics* **2**, 417–432.
- 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  $\beta$ -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 Statistical Society, Series C, (Applied Statistics)* **56**, 119–137.
- Demidenko, E. (2004). *Mixed Models: Theory and Applications*. Wiley, New York.
- Dempster, A. E., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood with incomplete data via the E-M algorithm. *Journal of the Royal Statistical Society, Series B* **39**, 1–38.
- Fieuws, S. and Verbeke, G. (2004). Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach. *Statistics in Medicine* **23**, 3093–3104.
- Hall, D. B. and Clutter, M. (2004). Multivariate multilevel nonlinear mixed effects models for timber yield predictions. *Biometrics* **60**, 16–24.
- Hirst, K., Zerbe, G. O., Boyle, D. W. and Wilkening, R. B. (1991). On nonlinear random effects models for repeated measurements. *Communications in Statistics B, Simulation and Computation* **20**, 463–478.

- Laird, N. M. and Ware, J. H. (1982). Random effects models for longitudinal data. *Biometric* **38**, 963–974.
- Lin, H. Q., McCulloch, C. E., Turnbull, B. W., Slate, E. H. and Clark, L. C. (2000). A latent class mixed model for analysing biomarker trajectories with irregularly scheduled observations. *Statistics in Medicine* **19**, 1303–1318.
- Lindstrom, M. J. and Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* **46**, 673–687.
- Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. Wiley, New-York.
- Marshall, G. and Barón, A. E. (2000). Linear discriminant models for unbalanced longitudinal data. *Statistics in Medicine* **19**, 1969–1981.
- Marshall, G., De la Cruz-Mesía, R., Barón, A. E., Rutledge, J. H. and Zerbe, G. O. (2006). Nonlinear random effects model for multivariate responses with missing data. *Statistics in Medicine* **25**, 2817–2830.
- Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S. T., Yang, C. C., Wang, C. P., Kellam, S. G., Carlin, J. B. and Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics* **3**, 459–475.
- O’Brien, L. M. and Fitzmaurice, G. M. (2005). Regression models for the analysis of longitudinal gaussian data from multiple sources. *Statistics in Medicine* **24**, 1725–1744.
- Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press, New York.
- Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer, New York.
- Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. Chapman and Hall, London.
- Schafer, J. L. and Yucel, R. M. (2002). Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of Computational and Graphical Statistics* **11**, 437–457.
- Shah, A., Laird, N. and Schoenfeld, D. (1997). A random-effects model for multiple characteristics with possibly missing data. *Journal of American Statistical Association* **92**, 775–779.
- Shapiro, B. S., Escobar, M., Makuch, R., Lavy, G. and DeCherney, A. H. (1992). A model-based prediction for transvaginal ultrasonographic identification of early intrauterine pregnancy. *American Journal of Obstetrics and Gynecology* **166**, 1495–1500.
- Tomasko, L., Helms, R. W. and Snapinn, S. M. (1999). A discriminant analysis extension to mixed models. *Statistics in Medicine* **18**, 1249–1260.
- Verbeke, G. and Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random effects population. *Journal of American Statistical Association* **91**, 217–221.
- Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New-York.
- Vonesh, E. F. and Chinchilli, V. M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. Marcel Dekker Inc, New York.
- Wernecke, K.-D., Kalb, G., Schink, B. and Wegner, B. (2004). A mixed model approach to

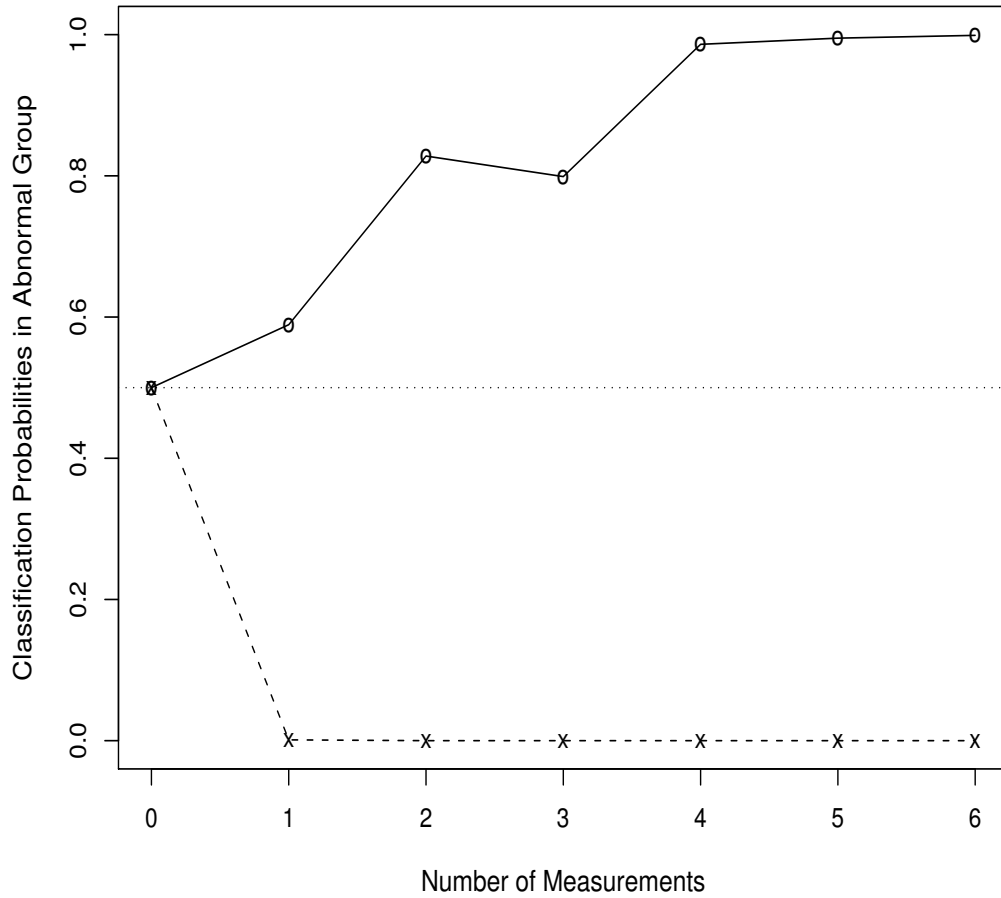
- discriminant analysis with longitudinal data. *Biometrical Journal* **46**, 246–254.
- Yamashita, T., Okamoto, S., Thomas, A., MacLachlan, V. and Healy, D. L. (1989). Predicting pregnancy outcome after in vitro fertilization and embryo transfer using estradiol, progesterone and human chorionic gonadotrophin  $\beta$ -subunit. *Fertil Steril* **51**, 304–309.
- Young, D. A., Zerbe, G. O. and Hay, W. W. (1992). Application of a non-linear random-effects model for comparing x-intercepts of linear regression lines. *Statistics in Medicine* **11**, 2039–2040.



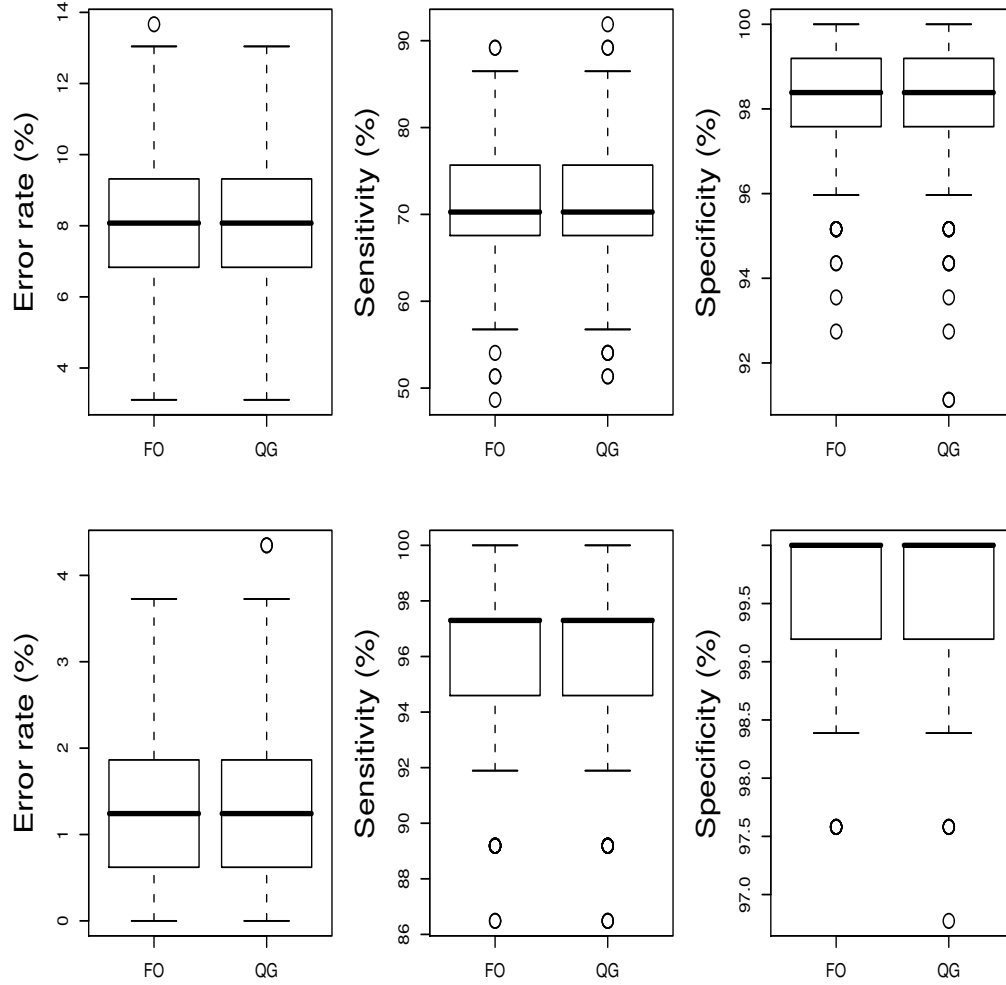
**Figure 1.** Time profiles for normal and abnormal patients.



**Figure 2.** ROC curves with areas and standard errors for three discriminant models.



**Figure 3.** Evolution of classification probabilities for one abnormal (solid line) and one normal (dashed line) future patient as a function of the number of observations.



**Figure 4.** Box-plot of classification results based on the parameters estimates obtained using the first-order (FO) and Gaussian quadrature (QG) methods. The first row represents the case when the number of measurements per patient is unbalanced and the second when  $n_i = 5$ . Box-plots are presented in terms of error rate, sensitivity and specificity.

**Table 1**  
*Summary of model fitting*

Model	d.f.	$-2 \log L$	AIC	$\chi^2$	$p$ -value
Homoscedastic	14	425.2	453.2	115.1	$< 0.01$
Mean-Heteroscedastic	16	403.0	435.0	92.9	$< 0.01$
Variance-Heteroscedastic	20	319.6	359.6	9.5	$< 0.01$
Fully-Heteroscedastic	22	310.1	354.1	–	–

**Table 2***Parameter estimates and standard errors (in parentheses) for the discriminant models.*

Parameter	Estimates of Model			
	Homoscedastic	Heteroscedastic		
		Mean	Variance	Fully
$\beta_{11}$	4.715 <sub>(0.064)</sub>	4.785 <sub>(0.068)</sub>	4.725 <sub>(0.046)</sub>	4.739 <sub>(0.047)</sub>
$\beta_{12}$	14.923 <sub>(0.415)</sub>	15.710 <sub>(0.405)</sub>	15.348 <sub>(0.342)</sub>	15.616 <sub>(0.340)</sub>
$\beta_{13}$	7.578 <sub>(0.538)</sub>	7.731 <sub>(0.549)</sub>	7.424 <sub>(0.431)</sub>	7.385 <sub>(0.432)</sub>
$\beta_{14}$	2.276 <sub>(0.047)</sub>	2.261 <sub>(0.046)</sub>	2.274 <sub>(0.045)</sub>	2.271 <sub>(0.045)</sub>
$\beta_{15}$	0.013 <sub>(0.001)</sub>	0.013 <sub>(0.001)</sub>	0.013 <sub>(0.001)</sub>	0.013 <sub>(0.001)</sub>
$\beta_{21}$	3.930 <sub>(0.094)</sub>	3.649 <sub>(0.107)</sub>	3.938 <sub>(0.163)</sub>	3.673 <sub>(0.182)</sub>
$\beta_{22}$	–	12.139 <sub>(1.344)</sub>	–	12.406 <sub>(1.792)</sub>
$\beta_{23}$	–	6.233 <sub>(1.253)</sub>	–	6.535 <sub>(1.917)</sub>
$\beta_{24}$	2.409 <sub>(0.117)</sub>	2.482 <sub>(0.112)</sub>	2.389 <sub>(0.126)</sub>	2.481 <sub>(0.124)</sub>
$\beta_{25}$	0.001 <sub>(0.003)</sub>	-0.001 <sub>(0.003)</sub>	0.002 <sub>(0.003)</sub>	-0.001 <sub>(0.003)</sub>
$B_{11}^{(1)}$	0.151 <sub>(0.032)</sub>	0.164 <sub>(0.031)</sub>	0.030 <sub>(0.013)</sub>	0.031 <sub>(0.013)</sub>
$B_{22}^{(1)}$	0.050 <sub>(0.010)</sub>	0.052 <sub>(0.010)</sub>	0.040 <sub>(0.008)</sub>	0.040 <sub>(0.008)</sub>
$B_{12}^{(1)}$	0.036 <sub>(0.016)</sub>	0.041 <sub>(0.017)</sub>	0.011 <sub>(0.009)</sub>	0.012 <sub>(0.009)</sub>
$B_{11}^{(2)}$	–	–	0.721 <sub>(0.237)</sub>	0.651 <sub>(0.199)</sub>
$B_{22}^{(2)}$	–	–	0.086 <sub>(0.035)</sub>	0.086 <sub>(0.035)</sub>
$B_{12}^{(2)}$	–	–	0.110 <sub>(0.080)</sub>	0.103 <sub>(0.075)</sub>
$\Sigma_{11}^{(1)}$	0.136 <sub>(0.015)</sub>	0.119 <sub>(0.013)</sub>	0.094 <sub>(0.011)</sub>	0.093 <sub>(0.011)</sub>
$\Sigma_{22}^{(1)}$	0.028 <sub>(0.004)</sub>	0.028 <sub>(0.004)</sub>	0.028 <sub>(0.004)</sub>	0.028 <sub>(0.004)</sub>
$\Sigma_{12}^{(1)}$	0.014 <sub>(0.007)</sub>	0.010 <sub>(0.006)</sub>	0.004 <sub>(0.006)</sub>	0.003 <sub>(0.006)</sub>
$\Sigma_{11}^{(2)}$	–	–	0.235 <sub>(0.053)</sub>	0.195 <sub>(0.044)</sub>
$\Sigma_{22}^{(2)}$	–	–	0.029 <sub>(0.010)</sub>	0.027 <sub>(0.009)</sub>
$\Sigma_{12}^{(2)}$	–	–	0.038 <sub>(0.018)</sub>	0.032 <sub>(0.015)</sub>

**Table 3**  
*Classification results within-sample (A) and using Cross-validation (B)*

<b>Groups</b>	<b>Classification</b>				
	<b>(A)</b>		<b>(B)</b>		
	Normal	Abnormal	Normal	Abnormal	
Normal	117	7	117	7	124
Abnormal	16	21	17	20	37
	133	28	134	27	161

**Table 4**

*Simulation results of the parameters estimates using first-order (FO) and Gaussian quadrature methods considering 124 patients and a different number of measurements per patient ( $n_i$ ). The notation  $4^*$  refers to values of  $n_i$  ranging from 1 to 4, exactly as in the pregnancy dataset.*

True value	$n_i$	First-order			Gaussian quadrature		
		Estimate	SE	Bias	Estimate	SE	Bias
$\beta_{11} = 4.8$	$n_i = 4^*$	4.804	0.034	0.004	4.801	0.035	0.001
	$n_i = 5$	4.806	0.020	0.006	4.801	0.021	0.001
	$n_i = 10$	4.805	0.014	0.005	4.800	0.014	0.000
$\beta_{12} = 15.1$	$n_i = 4^*$	15.130	0.489	0.030	15.038	0.456	-0.062
	$n_i = 5$	15.159	0.353	0.059	15.089	0.391	-0.011
	$n_i = 10$	15.144	0.308	0.044	15.065	0.447	-0.035
$\beta_{13} = 7.6$	$n_i = 4^*$	7.837	0.396	0.237	7.643	0.363	0.043
	$n_i = 5$	7.824	0.226	0.224	7.614	0.224	0.014
	$n_i = 10$	7.833	0.156	0.233	7.628	0.163	0.027
$\beta_{14} = 2.3$	$n_i = 4^*$	2.302	0.037	0.002	2.324	0.111	0.024
	$n_i = 5$	2.299	0.025	-0.001	2.313	0.065	0.013
	$n_i = 10$	2.299	0.023	-0.001	2.316	0.050	0.016
$\beta_{15} = 0.01$	$n_i = 4^*$	0.010	0.001	0.000	0.010	0.002	0.000
	$n_i = 5$	0.010	0.0005	0.000	0.010	0.001	0.000
	$n_i = 10$	0.010	0.0004	0.000	0.010	0.0007	0.000
$\Sigma_{11} = 0.05$	$n_i = 4^*$	0.050	0.005	0.000	0.050	0.005	0.000
	$n_i = 5$	0.050	0.003	0.000	0.050	0.002	0.000
	$n_i = 10$	0.050	0.002	0.000	0.051	0.002	0.001
$\Sigma_{22} = 0.03$	$n_i = 4^*$	0.030	0.003	0.000	0.030	0.004	0.000
	$n_i = 5$	0.030	0.002	0.000	0.030	0.002	0.000
	$n_i = 10$	0.030	0.001	0.000	0.030	0.001	0.000
$\Sigma_{12} = -0.0009$	$n_i = 4^*$	-0.0007	0.003	0.0002	-0.0004	0.002	0.0005
	$n_i = 5$	-0.0005	0.001	0.0004	-0.0003	0.001	0.0006
	$n_i = 10$	-0.0006	0.001	0.0003	-0.0003	0.001	0.0006
$B_{11} = 8.5$	$n_i = 4^*$	8.537	1.853	-0.037	8.292	0.667	-0.208
	$n_i = 5$	8.500	1.482	0.000	8.127	0.627	-0.373
	$n_i = 10$	8.398	0.888	-0.102	7.965	0.548	-0.535
$B_{22} = 0.04$	$n_i = 4^*$	0.041	0.008	0.001	0.040	0.007	0.000
	$n_i = 5$	0.040	0.006	0.000	0.040	0.006	0.000
	$n_i = 10$	0.040	0.005	0.000	0.038	0.006	-0.002
$B_{12} = -0.21$	$n_i = 4^*$	-0.213	0.103	-0.003	-0.204	0.089	0.006
	$n_i = 5$	-0.213	0.067	-0.003	-0.204	0.066	0.006
	$n_i = 10$	-0.208	0.057	0.002	-0.198	0.072	0.012

**Table 5**

*Simulation results of the parameter estimates using first-order (FO) and Gaussian quadrature methods considering 37 patients and a different number of measurements per patient ( $n_i$ ). The notation 6\* refers to values of  $n_i$  ranging from 1 to 6, exactly as in the pregnancy dataset.*

True value	$n_i$	First-order			Gaussian quadrature		
		Estimate	SE	Bias	Estimate	SE	Bias
$\beta_{21} = 3.9$	$n_i = 6^*$	3.965	0.308	0.065	3.953	0.126	0.053
	$n_i = 5$	3.928	0.133	0.028	3.950	0.090	0.050
	$n_i = 10$	3.918	0.085	0.018	3.945	0.053	0.047
$\beta_{22} = 11.6$	$n_i = 6^*$	11.945	2.800	0.345	11.875	0.881	0.275
	$n_i = 5$	11.543	1.470	-0.057	11.870	0.705	0.270
	$n_i = 10$	11.586	1.049	-0.014	11.952	0.300	0.352
$\beta_{23} = 8.5$	$n_i = 6^*$	9.177	4.240	0.678	8.958	0.784	0.458
	$n_i = 5$	8.854	1.442	0.354	8.947	0.512	0.447
	$n_i = 10$	8.746	1.027	0.246	8.987	0.249	0.487
$\beta_{24} = 2.3$	$n_i = 6^*$	2.291	0.100	-0.009	2.137	0.095	-0.163
	$n_i = 5$	2.298	0.076	-0.002	2.128	0.087	-0.172
	$n_i = 10$	2.298	0.064	-0.002	2.108	0.069	-0.192
$\beta_{25} = 0.002$	$n_i = 6^*$	0.002	0.003	0.000	0.005	0.002	0.003
	$n_i = 5$	0.002	0.001	0.000	0.005	0.001	0.003
	$n_i = 10$	0.002	0.001	0.000	0.005	0.001	0.003
$\Sigma_{11} = 0.61$	$n_i = 6^*$	0.575	0.094	-0.035	0.595	0.089	-0.015
	$n_i = 5$	0.600	0.064	-0.010	0.603	0.063	-0.007
	$n_i = 10$	0.608	0.045	-0.002	0.611	0.045	0.001
$\Sigma_{22} = 0.05$	$n_i = 6^*$	0.048	0.009	-0.002	0.047	0.009	-0.003
	$n_i = 5$	0.050	0.006	0.000	0.049	0.006	-0.001
	$n_i = 10$	0.050	0.004	0.000	0.050	0.004	0.000
$\Sigma_{12} = 0.12$	$n_i = 6^*$	0.113	0.025	-0.007	0.115	0.025	-0.005
	$n_i = 5$	0.118	0.017	-0.002	0.118	0.017	-0.002
	$n_i = 10$	0.119	0.011	-0.001	0.119	0.011	-0.001
$B_{11} = 1.03$	$n_i = 6^*$	2.948	9.137	1.918	1.002	0.231	0.028
	$n_i = 5$	1.095	1.158	0.065	1.007	0.293	-0.023
	$n_i = 10$	0.996	0.798	-0.034	0.991	0.127	-0.039
$B_{22} = 0.1$	$n_i = 6^*$	0.096	0.023	-0.004	0.103	0.032	0.003
	$n_i = 5$	0.095	0.022	-0.005	0.100	0.028	0.000
	$n_i = 10$	0.092	0.021	-0.008	0.098	0.025	-0.002
$B_{12} = -0.23$	$n_i = 6^*$	-0.062	0.234	0.168	-0.208	0.078	0.022
	$n_i = 5$	-0.221	0.370	0.009	-0.226	0.070	0.004
	$n_i = 10$	-0.220	0.262	0.010	-0.256	0.054	-0.026

**Table 6**

Simulation results of the parameters estimates using the FO method considering 124 patients, a different number of measurements per patient ( $n_i$ ), and different values of the coefficient of variation (CV) for the nonlinear random-effect,  $CV=\sqrt{B_{11}}/\beta_{12}$ . The last three columns correspond to the case when  $B_{11} = 25$  and the rest of the true parameters are the same. The notation 4\* refers to values of  $n_i$  ranging from 1 to 4, exactly as in the pregnancy dataset.

True value	$n_i$	CV=0.23			CV=0.33		
		Estimate	SE	Bias	Estimate	SE	Bias
$\beta_{11} = 4.8$	$n_i = 4^*$	4.804	0.032	0.004	4.814	0.032	0.014
	$n_i = 5$	4.807	0.020	0.007	4.815	0.021	0.015
	$n_i = 10$	4.807	0.014	0.007	4.815	0.014	0.015
$\beta_{12} = 15.1$	$n_i = 4^*$	15.262	0.530	0.162	15.292	0.675	0.192
	$n_i = 5$	15.238	0.389	0.138	15.357	0.547	0.257
	$n_i = 10$	15.215	0.354	0.115	15.436	0.551	0.336
$\beta_{13} = 7.6$	$n_i = 4^*$	7.844	0.405	0.244	8.246	0.416	0.646
	$n_i = 5$	7.907	0.246	0.307	8.278	0.271	0.678
	$n_i = 10$	7.926	0.157	0.326	8.281	0.188	0.681
$\beta_{14} = 2.3$	$n_i = 4^*$	2.292	0.288	-0.008	2.295	0.289	-0.005
	$n_i = 5$	2.303	0.287	0.003	2.299	0.289	-0.0007
	$n_i = 10$	2.300	0.290	0.000	2.290	0.289	-0.010
$\beta_{15} = 0.01$	$n_i = 4^*$	0.010	0.001	0.000	0.010	0.001	0.000
	$n_i = 5$	0.010	0.001	0.000	0.010	0.0005	0.000
	$n_i = 10$	0.010	0.0004	0.000	0.010	0.0004	0.000
$\Sigma_{11} = 0.05$	$n_i = 4^*$	0.051	0.006	0.001	0.052	0.006	0.002
	$n_i = 5$	0.050	0.003	0.000	0.052	0.003	0.002
	$n_i = 10$	0.050	0.002	0.000	0.052	0.002	0.002
$\Sigma_{22} = 0.03$	$n_i = 4^*$	0.030	0.003	0.000	0.030	0.004	0.000
	$n_i = 5$	0.030	0.002	0.000	0.030	0.002	0.000
	$n_i = 10$	0.030	0.001	0.000	0.030	0.001	0.000
$\Sigma_{12} = -0.0009$	$n_i = 4^*$	-0.0006	0.003	0.0003	-0.0005	0.002	0.0004
	$n_i = 5$	-0.0006	0.001	0.0003	-0.0005	0.001	0.0004
	$n_i = 10$	-0.0006	0.001	0.0003	-0.0004	0.0008	0.0005
$B_{11} = 12(25)^*$	$n_i = 4^*$	10.723	2.745	-1.277	24.764	5.144	-0.236
	$n_i = 5$	11.380	2.225	-0.620	25.071	4.122	0.071
	$n_i = 10$	11.660	1.869	-0.340	24.848	3.685	-0.152
$B_{22} = 10$	$n_i = 4^*$	9.955	1.326	-0.045	10.088	1.353	0.088
	$n_i = 5$	10.008	1.269	0.008	10.105	1.291	0.105
	$n_i = 10$	10.021	1.294	0.021	10.133	1.302	0.133
$B_{12} = -0.21$	$n_i = 4^*$	-0.205	1.404	0.004	-0.223	2.645	-0.013
	$n_i = 5$	-0.334	1.316	-0.124	-0.806	2.214	-0.596
	$n_i = 10$	-0.454	1.225	-0.244	-1.152	1.914	-0.942

**Table 7**

Simulation results of the parameters estimates using the FO method considering 37 patients, a different number of measurements per patient ( $n_i$ ), and different values of the coefficient of variation (CV) for the nonlinear random-effect,  $CV=\sqrt{B_{11}}/\beta_{12}$ . The last three columns correspond to the case when  $B_{11} = 25$  and the rest of the true parameters are the same. The notation 6\* refers to values of  $n_i$  ranging from 1 to 6, exactly as in the pregnancy dataset.

True value	$n_i$	CV=0.27			CV=0.43		
		Estimate	SE	Bias	Estimate	SE	Bias
$\beta_{11} = 3.9$	$n_i = 6^*$	3.952	0.376	0.052	3.976	0.348	0.076
	$n_i = 5$	3.928	0.171	0.028	3.935	0.207	0.035
	$n_i = 10$	3.910	0.095	0.010	3.920	0.110	0.020
$\beta_{12} = 11.6$	$n_i = 6^*$	11.975	3.248	0.375	12.175	3.533	0.575
	$n_i = 5$	11.739	1.960	0.139	11.929	2.391	0.329
	$n_i = 10$	11.741	1.326	0.141	11.859	1.772	0.259
$\beta_{13} = 8.5$	$n_i = 6^*$	9.335	4.536	0.835	9.728	5.242	1.228
	$n_i = 5$	8.941	2.261	0.441	9.233	2.731	0.733
	$n_i = 10$	8.744	1.340	0.244	9.133	1.841	0.633
$\beta_{14} = 2.3$	$n_i = 6^*$	2.298	0.582	-0.002	2.301	0.581	0.001
	$n_i = 5$	2.297	0.566	-0.003	2.303	0.566	0.003
	$n_i = 10$	2.341	0.635	0.041	2.314	0.619	0.014
$\beta_{12} = 0.002$	$n_i = 6^*$	0.002	0.003	0.000	0.002	0.003	0.000
	$n_i = 5$	0.002	0.002	0.000	0.002	0.002	0.000
	$n_i = 10$	0.002	0.001	0.000	0.002	0.001	0.000
$\Sigma_{11} = 0.61$	$n_i = 6^*$	0.615	0.123	0.005	0.652	0.113	0.042
	$n_i = 5$	0.622	0.070	0.011	0.648	0.083	0.038
	$n_i = 10$	0.623	0.050	0.012	0.630	0.057	0.020
$\Sigma_{22} = 0.05$	$n_i = 6^*$	0.050	0.012	0.000	0.050	0.011	0.000
	$n_i = 5$	0.050	0.006	0.000	0.050	0.006	0.000
	$n_i = 10$	0.050	0.004	0.000	0.050	0.004	0.000
$\Sigma_{12} = 0.12$	$n_i = 6^*$	0.120	0.036	0.000	0.125	0.031	0.005
	$n_i = 5$	0.121	0.017	0.001	0.123	0.020	0.003
	$n_i = 10$	0.121	0.011	0.001	0.119	0.013	-0.001
$B_{11} = 10(25)^*$	$n_i = 6^*$	1.823	5.722	-8.177	3.190	9.387	-21.810
	$n_i = 5$	2.759	5.847	-7.241	8.424	14.056	-16.576
	$n_i = 10$	3.374	4.972	-6.626	13.689	13.824	-11.311
$B_{22} = 12$	$n_i = 6^*$	11.647	2.790	-0.353	11.668	2.798	-0.332
	$n_i = 5$	11.647	2.683	-0.353	11.688	2.747	-0.312
	$n_i = 10$	11.593	2.717	-0.407	11.822	2.839	-0.178
$B_{12} = -0.23$	$n_i = 6^*$	-0.157	4.547	0.073	-0.020	4.940	0.210
	$n_i = 5$	-0.113	2.811	0.117	0.013	3.583	0.243
	$n_i = 10$	0.076	2.225	0.306	0.028	3.991	0.258