T E C H N I C A L
R E P O R T

0450

PAIRWISE FITTING OF MIXED MODELS FOR THE JOINT MODELLING OF MULTIVARIATE LONGITUDINAL PROFILES

FIEUWS, S. and G. VERBEKE

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Pairwise Fitting of Mixed Models for the Joint Modelling of Multivariate Longitudinal Profiles

Steffen Fieuws and Geert Verbeke

Biostatistical Centre, Katholieke Universiteit Leuven, Leuven, Belgium

Abstract

A mixed model is a flexible tool for joint modelling purposes, especially when the gathered data are unbalanced. However, computational problems due to the dimension of the joint covariance matrix of the random effects arise as soon as the number of outcomes and/or the number of used random effects per outcome increases.

We propose a pairwise approach in which all possible bivariate models are fitted, and where inference follows form pseudo-likelihood arguments. The approach is applicable for linear, generalised linear and nonlinear mixed models, or for combinations of these. The methodology will be illustrated in the analysis of 22-dimensional, highly

[1]Correspondence to: Steffen Fieuws, Biostatistical Centre, Katholieke Universiteit Leuven, Kapucijnenvoer 35, B-3000 Leuven, Belgium. E-mail : steffen.fieuws@med.kuleuven.ac.be

unbalanced, longitudinal profiles of hearing thresholds.

Key words: Joint Modelling, Multivariate Longitudinal Profiles, Mixed Models, Pseudo-likelihood

1 Introduction

Many statistical models have been proposed in the literature for the analysis of longitudinal data. Most of these models restrict attention to the analysis of one single outcome variable, measured repeatedly over time. We will denote this standard situation as the analysis of univariate longitudinal data. Multivariate longitudinal data arise when, instead of a single outcome, a set of different outcomes on the same unit is measured repeatedly over time. As an example of multivariate longitudinal data, consider the situation where different indices of physical and mental health are measured repeatedly over time.

In many situations joint modelling of the multivariate longitudinal profiles is needed or has additional advantages over the separate analyses of the different outcomes. First, the association structure can be of importance. A possible question might be how the association between outcomes evolves over time or how outcome-specific evolutions are related to each other (Fieuws and Verbeke, 2004). In a second situation, the aim is to improve the results of a discriminant analysis by using more than one longitudinally measured outcome. In another situation, interest lies on estimation of the average trends. As an example, consider testing the difference in evolution between many outcomes or testing for a treatment effect common to a set of outcomes. All these situations require a joint model for all outcomes.

A flexible approach is to model the different outcomes jointly by using random effects models. In applied sciences, random-effects models have become the preferred tool to analyse various types of longitudinal data. With these models, the average evolution of a specific outcome is described using some function of time, and subject-specific deviations from this average evolution are introduced by using so-called random effects. The introduction by Laird and Ware (1982) of the linear mixed model, devised for continuous data, has been followed rapidly by extensions to deal with non-linear data (non-linear mixed effects model, Davidian and Giltinan, 1995) and with non-continuous measurements (generalized linear mixed model, Breslow and Clayton, 1993). In a joint-modelling approach using mixed models, random-effects are assumed for each outcome process, and by imposing a joint multivariate distribution on the random effects, the different processes are associated. This approach has many advantages and is applicable in a wide variety of situations. First, the data can be highly unbalanced. For example, it is not necessary that all outcomes are measured at the same timepoints. Moreover, the approach is applicable in situations where linear, nonlinear or generalised linear mixed models are used to describe the evolution of the individual outcome processes. Also, models can be constructed joining different types of mixed models. For example, a generalised linear mixed model for a discrete outcome combined with a nonlinear mixed model for a continuous outcome. The approach has been used in a non-longitudinal setting to validate surrogate endpoints in meta-analyses (Buyse et al. 2000, Burzykowski et al. 2001) or to model multivariate clustered data (Thum, 1997). Gueorguieva (2001a) used the approach for the joint modelling of a continuous and a binary outcome measure in a developmental toxicity study on mice. Also in a longitudinal setting, Chakraborty et al. (2003) obtained estimates of the correlation between blood and semen HIV-1 RNA by using a joint random-effects model. Other examples with longitudinal studies can be found in MacCallum et al. (1997), Thiébaut et al. (2002a, 2002b) and Shah et al. (1997). All these examples refer to situations where the number of different outcomes is low. This implies that the dimension of the joint distribution of the random effects does not lead to computational problems during the estimation process. In this paper we will focus on the situation where there is no limit on the number of different outcomes, such that the distribution for the random effects becomes high-dimensional.

Section 2 introduces data on pure-tone hearing thresholds as an example of high-dimensional multivariate longitudinal profiles. Section 3 introduces the joint modelling approach using mixed models and describes the dimensionality problem. Section 4 details the pairwise fitting approach as a solution to this problem. Section 5 illustrates the pairwise modelling approach using data on pure-tone hearing thresholds. Section 6 presents the results of a simulation study. Section 7 contains a discussion.

2 Example: Pure-Tone Hearing thresholds

In a hearing test, hearing threshold sound pressure levels (dB) are determined at different frequencies to evaluate the hearing performance of a subject. A hearing threshold is the lowest signal intensity a subject can detect at a specific frequency. In this study, hearing thresholds measured at eleven different frequencies (125Hz, 250Hz, 500Hz, 750Hz, 1000Hz, 1500Hz, 2000Hz, 3000Hz, 4000Hz, 6000Hz and 8000Hz), obtained on 603 male participants from the Baltimore Longitudinal Study of Aging (BLSA, Shock et al. 1984), are considered. Hearing thresholds are measured at the left as well as at the right ear, implying that the data are 22-variate longitudinal. All subjects have at least one measurement for each of the 22 different outcomes. Hearing thresholds are measured repeatedly over time, with the number of visits per subject varying from 1 to 15 (a median follow-up time of 6.9 years). Visits are unequally spaced. The age at first visit of the participants ranges from 17.2 to 87 years (with a median age at first visit of 50.2 years). Analyses of the hearing data collected in the BLSA study can be found in Brant and Fozard (1990), Morrell and Brant (1991) and Pearson et al. (1995). Ear- and frequency specific profiles (for all 22 outcomes) for one randomly chosen subject are shown in Figure 1. Individual profiles of 30 randomly selected subjects for two specific frequencies are shown in Figure 2. It is well known that the hearing performance deteriorates as one gets older. Therefore, a significant interaction between the evolution of the hearing threshold and age at entry is expected. But is this deterioration the same for all frequencies? Otherwise said, should frequency-specific interactions between the evolution and age at entry be allowed or can an interaction term common to all outcomes be assumed? Another question refers to the association structure of the frequency-specific evolutions within a subject. To what extent is the evolution of one frequency related to the evolution of another frequency? Both questions will be answered using a joint model for all

22 outcomes.

[Figure 1 about here.]

[Figure 2 about here.]

3 Joint Modelling : Random-Effects Approach

3.1 Univariate Mixed Models

Let m be the number of different outcomes to be modelled jointly. It will be assumed that each of the m longitudinally measured outcomes can be modelled using a mixed model. More specifically, for one outcome, let y_{ij} denote the jth measurement available for the ith subject, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$, and let $\boldsymbol{y_i}$ denote the vector of all measurements for the ith subject, i.e., $\boldsymbol{y_i}' = (y_{i1}, \ldots, y_{in_i})$. Our general model assumes that $\boldsymbol{y_i}$ (possibly appropriately transformed) satisfies

$$\boldsymbol{y_i}|\boldsymbol{b_i} \sim F_i(\boldsymbol{\psi}, \boldsymbol{b_i}), \tag{1}$$

i.e., conditional on $\boldsymbol{b_i}$, $\boldsymbol{y_i}$ follows a pre-specified distribution F_i, possibly depending on covariates, and parameterized through a vector $\boldsymbol{\psi}$ of unknown parameters, common to all subjects. Further, $\boldsymbol{b_i}$ is a q-dimensional vector of subject-specific parameters, called random effects, assumed to follow a normal distribution with mean zero and covariance matrix D. The $\boldsymbol{b_i}$ reflect the between-subject heterogeneity in the population with respect to the distribution of $\boldsymbol{y_i}$. In the presence of random effects, conditional independence is often assumed, under which the components y_{ij} in $\boldsymbol{y_i}$ are independent, conditional on $\boldsymbol{b_i}$. The distribution function F_i in (1) then becomes a product over the n_i independent elements in $\boldsymbol{y_i}$.

In general, unless a fully Bayesian approach is followed, inference is based on the marginal model for $\boldsymbol{y_i}$ which is obtained from integrating out the random effects, over their normal distribution. Let $f_i(\boldsymbol{y_i}|\boldsymbol{b_i})$ denote the density function corresponding to the distribution F_i, and let $g(\boldsymbol{b_i})$ denote the density of a normal distribution. We have that the marginal density function of $\boldsymbol{y_i}$ equals

$$f_i(\boldsymbol{y_i}) = \int f_i(\boldsymbol{y_i}|\boldsymbol{b_i})g(\boldsymbol{b_i})d\boldsymbol{b_i}, \qquad (2)$$

which depends on the unknown parameters in $\boldsymbol{\psi}$ and in D. Assuming independence of the subjects, estimates can be obtained from maximizing the likelihood function built from (2), and inferences immediately follow from classical maximum likelihood theory.

Special cases are linear and nonlinear mixed models for continuous data, and generalised linear mixed models for discrete data. Linear mixed models assume that the n_i-dimensional vector $\boldsymbol{y_i}$ satisfies

$$\boldsymbol{y_i}|\boldsymbol{b_i} \sim N(X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i}, \Sigma_i), \qquad (3)$$

where X_i and Z_i are $(n_i \times k)$ and $(n_i \times q)$ dimensional matrices of known covariates, $\boldsymbol{\beta}$ is a k-dimensional vector of regression parameters, called the fixed effects, and Σ_i is a $(n_i \times n_i)$ covariance matrix which depends on i only through its dimension n_i, i.e. the set of unknown parameters in Σ_i will not depend upon i. Marginally, $\boldsymbol{y_i}$ follows a normal distribution with mean $X_i\boldsymbol{\beta}$ and covariance matrix $V_i = Z_iDZ_i' + \Sigma_i$. Non-linear mixed models can be considered as an extension of model (3). Non-linear mixed models allow for non-linear relationships between the responses in $\boldsymbol{y_i}$ and the covariates in X_i and/or Z_i such that

$$\boldsymbol{y_i}|\boldsymbol{b_i} \sim N(h(X_i, Z_i, \boldsymbol{\beta}, \boldsymbol{b_i}), \Sigma_i) \tag{4}$$

for some known 'link' function h. The definition of X_i, Z_i, $\boldsymbol{\beta}$ and $\boldsymbol{b_i}$ remains unchanged. The generalized linear mixed model assumes that conditionally on random effects $\boldsymbol{b_i}$ the elements y_{ij} of $\boldsymbol{y_i}$ are independent, with density function of the form

$$f_i(y_{ij}|\boldsymbol{b_i}) = \exp\left[(y_{ij}\eta_{ij} - a(\eta_{ij}))/\phi + c(y_{ij}, \phi)\right],$$

with mean $\mathsf{E}(y_{ij}|\boldsymbol{b_i}) = a'(\eta_{ij}) = \mu_{ij}(\boldsymbol{b_i})$ and variance $\mathsf{Var}(y_{ij}|\boldsymbol{b_i}) = \phi a''(\eta_{ij})$, and where, apart from a link function h, a linear regression model with parameters $\boldsymbol{\beta}$ and $\boldsymbol{b_i}$ is used for the mean, i.e., $h(\boldsymbol{\mu_i}(\boldsymbol{b_i})) = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i}$. In contrast to the linear mixed model, for the generalized linear mixed model and the nonlinear mixed model, the marginal distibution of $\boldsymbol{y_i}$ cannot be derived analytically such that numerical approximations are required.

3.2 Joint Mixed Model

The joint model assumes a mixed model for each outcome, and models are combined through specification of a joint multivariate distribution for all random effects appearing in the univariate mixed models. Obviously, the joint model can be considered as a new mixed model of the form (1), but with a random-effects vector $\boldsymbol{b_i}$ of a higher dimension. Let $\boldsymbol{\Theta}^*$ the vector containing all parameters (fixed effects parameters as well as covariance parameters), then $l_i(\boldsymbol{Y}_{1i}, \boldsymbol{Y}_{2i}, \ldots, Y_{mi}|\boldsymbol{\Theta}^*)$ refers to the loglikelihood contribution of subject i to the full joint mixed model. Strictly speaking, standard software using likelihood based inference for fitting linear, non-linear or generalised linear mixed models can be used to obtain parameter estimates for this joint mixed model. Examples using the SAS-procedure PROC MIXED for joining linear mixed models can be found in Thiébaut et al. (2002b). However, computational problems will arise as the dimension of the random-effects vector $\boldsymbol{b_i}$ in the joint model increases. Obviously, a joint mixed model will suffer sooner from computational problems if numerical approximations are needed to calculate the marginal density in (2). But even if all outcome-specific models are linear mixed models and the marginal density in (2) can be calculated analytically, computational problems are also encountered rapidly. This will be shown in the application presented in Section 5.

4 Pairwise Modelling Approach

The general idea behind the pairwise modelling approach is straightforward. Instead of maximising the likelihood of the full joint model presented in the previous section, all pairwise bivariate models will be fitted separately in a first step. In a second step, the parameters obtained by fitting the pairwise models will be combined to obtain one single estimate for each parameter in the full joint model.

4.1 Fitting Pairwise Models

Instead of maximising the loglikelihood of the joint mixed model, loglikelihoods of the following form will be maximised separately

$$\sum_{i=1}^{N} l_{rsi}(\boldsymbol{Y}_{ri}, \boldsymbol{Y}_{si}|\boldsymbol{\Theta}_{r,s}), \tag{5}$$

$r = 1, \ldots, m-1, s = r+1, \ldots, m$. The total number of subjects is indicated by N. $\boldsymbol{\Theta}_{r,s}$ represents the vector of all parameters in the bivariate joint mixed model corresponding to the specific pair (r, s). To simplify the notation in the remainder, expression (5) can be rewritten as $\sum_{i=1}^{N} l_{pi}(\boldsymbol{\Theta}_p)$ with $p = 1, \ldots, P$ and $P = m(m-1)/2$ representing the total number of possible pairs. Finally, let $\boldsymbol{\Theta}$ then be the stacked vector combining all pair-specific parameter vectors $\boldsymbol{\Theta}_p$. Estimates for the elements in $\boldsymbol{\Theta}$ are obtained by maximising each of the P likelihoods separately.

It is important to note that the parameter vectors $\boldsymbol{\Theta}$ and $\boldsymbol{\Theta}^*$ are not equivalent. Indeed, some parameters in $\boldsymbol{\Theta}^*$ will have a single counterpart in $\boldsymbol{\Theta}$, e.g. the covariance between random effects from two different outcomes. Other elements in $\boldsymbol{\Theta}^*$ will have multiple counterparts in $\boldsymbol{\Theta}$, e.g. the covariance between random effects from the same outcome. In the latter case a single estimate is obtained by averaging all corresponding pair-specific estimates in $\widehat{\boldsymbol{\Theta}}$. Standard errors of the so-obtained estimates clearly cannot be obtained from averaging standard errors or variances. Indeed, the variability amongst the pair-specific estimates needs to be taken into account. Furthermore, two pair-specific estimates corresponding to two pairwise models with a common outcome are based on overlapping information and hence correlated. This correlation should also be accounted for in the sampling variability of the combined estimates in $\widehat{\boldsymbol{\Theta}}^*$. In the sequel of this section we will use pseudo-likelihood ideas to obtain standard errors for the estimates in $\widehat{\boldsymbol{\Theta}}^*$.

4.2 Inference for $\boldsymbol{\Theta}$

Borrowing ideas from the pseudo-likelihood framework, first a covariance matrix for the elements in $\widehat{\boldsymbol{\Theta}}$ will be constructed. The idea behind pseudo-likelihood estimation (Besag, 1975), is to replace the joint likelihood by a suitable product of marginal or conditional densities, where this product is easier to evaluate than the original likelihood. Examples of pseudo-likelihood estimation can be found in Arnold and Strauss (1991), Geys et al. (1997) and Renard et al. (2004). Although in the pairwise approach a set of likelihoods is maximised separately, the approach fits within the pseudo-likelihood framework. Indeed, fitting all possible pairwise models is equivalent to maximising a pseudo-likelihood function of the following form

$$\begin{aligned} pl(\boldsymbol{\Theta}) &= l(\boldsymbol{Y}_1, \boldsymbol{Y}_2|\boldsymbol{\Theta}_{1,2}) + l(\boldsymbol{Y}_1, \boldsymbol{Y}_3|\boldsymbol{\Theta}_{1,3}) + \ldots + l(\boldsymbol{Y}_{m-1}, \boldsymbol{Y}_m|\boldsymbol{\Theta}_{m-1,m}) \\ &= \textstyle\sum_{p=1}^{P} l_p(\boldsymbol{\Theta}_p). \end{aligned} \tag{6}$$

Note however that, in the classical examples of pseudo-likelihood estimation, the same parameter is present in the different parts of the pseudo-likelihood function whereas in (6) the set of parameters in $\boldsymbol{\Theta}_p$ is considered pair-specific (subscript p at this stage, only at a later stage the estimates will be combined). This separate parametrisation is needed to be able to maximise the different parts in expression (6) separately.

Since the pairwise approach fits within the pseudo-likelihood framework, an asymptotic multivariate normal distribution for $\widehat{\mathbf{\Theta}}$ can be derived. Asymptotic normality of the pseudo-likelihood estimator in the single parameter case and in the vector valued parameter case is shown in Arnold and Strauss (1991), and in Geys (1999), respectively. The asymptotic multivariate normal distribution for $\widehat{\mathbf{\Theta}}$ is given by

$$\sqrt{N}(\widehat{\mathbf{\Theta}} - \mathbf{\Theta}) \sim MVN(\mathbf{0}, J^{-1}KJ^{-1}) \tag{7}$$

where $J^{-1}KJ^{-1}$ is a "sandwich-type" robust variance estimator. J is a block-diagonal matrix containing second derivatives where each of the P blocks has the form

$$\frac{1}{N}\sum_{i=1}^{N} E\left(\frac{\partial^2 l_{pi}}{\partial \boldsymbol{\theta}_p \partial \boldsymbol{\theta}'_p}\right). \tag{8}$$

and K is a symmetric matrix containing $P \times P$ blocks, each block of the form

$$\frac{1}{N}\sum_{i=1}^{N} E\left(\frac{\partial l_{pi}}{\partial \boldsymbol{\theta}_p}\frac{\partial l_{qi}}{\partial \boldsymbol{\theta}'_q}\right), \tag{9}$$

$p, q = 1, \ldots, P$ Estimates for (9) and (8) are obtained by dropping the expectation and replacing the unknown parameters by their estimates.

4.3 Combining Information: Inference for $\mathbf{\Theta}^*$

In a final step, estimates for the parameters in $\mathbf{\Theta}^*$ can be calculated, as suggested before, by taking averages over all pairs. This is obtained by

$$\widehat{\mathbf{\Theta}}^* = A\widehat{\mathbf{\Theta}} \sim MVN(\mathbf{\Theta}^*, A\Sigma(\widehat{\mathbf{\Theta}})A') \tag{10}$$

with A a matrix containing the appropriate coefficients to calculate the averages and $\Sigma(\widehat{\mathbf{\Theta}})$ equals the covariance matrix for $\widehat{\mathbf{\Theta}}$ obtained by expression (7).

5 Application

For the data on the pure-tone hearing thresholds, introduced in Section 2, Verbeke and Molenberghs (2000) proposed the following linear mixed model to analyse the evolution of the hearing threshold for a single frequency. Let $Y_i(t)$ denote the hearing threshold at some frequency for a subject i taken at time t, the model is specified as

$$\begin{aligned} Y_i(t) &= (\beta_1 + \beta_2 Age_i + \beta_3 Age_i^2 + a_i) \\ &\quad + (\beta_4 + \beta_5 Age_i + b_i)t \\ &\quad + \beta_6 Visit1(t) + \varepsilon_i(t) \end{aligned} \tag{11}$$

in which t is time expressed in years from entry in the study and Age_i the age of subject i at the time of entry in the study. Since their is evidence for the presence of a learning effect from the first to the subsequent visits, a time-varying covariate $Visit1$ has been added. This covariate is defined to be one at the first measurement and zero for all other visits. Finally, the a_i are random intercepts, the b_i are the random slopes for time, and the ε_i represent the usual error components, independent of the random effects. The vector $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6)'$ of fixed effects describes the average evolution of the hearing threshold and the vector $(a_i, b_i)'$ of random effects describes how the profile of the ith subject deviates from the average profile. The classical normality assumptions apply for all random terms in this model.

Considering all outcomes, let $Y_{1i}(t)$, $Y_{2i}(t)$,..., $Y_{22i}(t)$ denote the hearing thresholds of the 11 frequencies at the left as well as at the right ear for a subject i taken at time t. Each of the 22 Frequency×Side combinations is described using the linear mixed-effects model (11). More specifically,

$$\begin{cases} Y_{1i}(t) = \mu_1(t) + a_{1i} + b_{1i}t + \varepsilon_{1i}(t) \\ \\ Y_{2i}(t) = \mu_2(t) + a_{2i} + b_{2i}t + \varepsilon_{2i}(t) \\ \\ \vdots \\ \\ Y_{22i}(t) = \mu_{22}(t) + a_{22i} + b_{22i}t + \varepsilon_{22i}(t) \end{cases}$$

where $\mu_1(t)$, $\mu_2(t)$, ..., $\mu_{22}(t)$ refer to the average evolutions. All 22 outcome trajectories are tied together through a joint distribution for the random effects

$$\left(a_{1i}, a_{2i}, \ldots, a_{22i}, b_{1i}, b_{2i}, \ldots, b_{22i} \right)' \sim N\left(\mathbf{0}, \boldsymbol{D}\right)$$

where $\boldsymbol{D}$ is the 44×44 covariance matrix of the random effects. The error components,

$$\left(\varepsilon_{1i}(t), \varepsilon_{2i}(t), \ldots, \varepsilon_{22i}(t) \right)' \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}\right) \quad \text{for all } t,$$

are assumed independent across timepoints and assumed independent of the random effects. Σ is the 22×22 covariance matrix of the error components at any specific point in time. If all off-diagonal elements in Σ equal zero then the error components are uncorrelated which implies that, conditionally on the random effects, the 22 outcome trajectories are independent. The covariance matrix of the random effects is 44 dimensional and the covariance matrix of the error components is 22 dimensional, leading to $990 + 253 = 1243$ covariance parameters to be estimated. Not suprisingly, this is computationally too demanding. Note that, in this particular example, without relying on the pairwise approach, we were not able to fit directly joint models with more than 4 outcomes. Besides the unstructured covariance matrices for the random effects and for the error components, all fixed effects parameters are assumed to be outcome-specific. As such, this joint mixed model can serve as a starting model in a search for more parsimonious models, which would assume common fixed effects, common or uncorrelated random effects, or common or uncorrelated error components. A

SAS-macro has been used to fit the 231 bivariate joint mixed models and to combine their results into estimates for $\mathbf{\Theta}^*$. The pairwise fitting results in 21 estimates for each fixed effect, each variance of the random effects, each covariance between random effects of the same outcome and each variance of the error components. For each covariance parameter between the error components and each covariance parameter between random effects of different outcomes, there is a single corresponding estimate in $\widehat{\mathbf{\Theta}}$.

As an example of the results for the fixed effects, Figure 3 shows the outcome-specific estimates and their 95% confidence interval for the interaction between the linear evolution and age at entry. We clearly observe an increasing trend implying that age accelerates hearing loss, but that this is more severe for higher frequencies. Wald-type tests indicate that these estimates are significantly different between the outcomes, at the left side ($\chi^2_{10} = 90.4$, $p < 0.0001$) as well as at the right side ($\chi^2_{10} = 110.9$, $p < 0.0001$)

[Figure 3 about here.]

To explore some results for the covariance parameters, Figure 4 shows the results of a principal component analysis on a subset of the 44×44 correlation matrix of the random effects, i.e. the 11×11 correlation matrix of the random slopes at the left side. The result indicates that the closer the frequencies, the stronger their evolutions are correlated. Very similar results were obtained for the right hand side (results not shown). Figure 5 shows the results of a principal component analysis on the 22×22 correlation matrix of the error components. Not surprisingly, errors of the same frequency are highly correlated between the left and the right side.

[Figure 4 about here.]

[Figure 5 about here.]

6 Simulation Study

A simulation study was conducted to evaluate the performance of the estimators from the pairwise approach and to compare it with the estimators obtained from the full multivariate mixed model. From the example on the hearing thresholds, three outcomes were selected (500 Hz, 1000 Hz and 2000 Hz, all taken at the right ear). A joint model using the random-effects approach, as outlined in the previous section, was fitted. Unstructured 6×6 and 3×3 covariance matrices were used for the random effects and the error components respectively. All fixed and random effects were considered to be outcome-specific. This model will be denoted as the full trivariate mixed model in the remainder. 1000 data sets were simulated from this full trivariate model with parameters obtained from the hearing data. Each time, the full trivariate model was fitted, and the pairwise approach was applied as well. Two versions of the model were used: The correct model and an incorrect model. The latter model does not contain random slopes and the error components are assumed to be uncorrelated. In the full trivariate approach, model-based as well as robust standard errors are calculated. The latter are the asymptotically consistent 'sandwich' estimators described in Liang and Zeger (1986) and in Diggle, Heagerty, Liang and Zeger (1994).

Figures 6 and 7 show the agreement between the estimates from the full trivariate and from the pairwise approach, for a subset of fixed effects and variance components. This shows that the estimates obtained with the pairwise approach are almost identical to the estimates of the full trivariate approach.

[Figure 6 about here.]

[Figure 7 about here.]

A comparison of the different approaches to assess the sampling variability of the estimates is summarized in Tables 1 and 2, for the three fixed slopes in the trivariate model. We first consider results under the correct model formulation (Table 1). Comparison of the first two columns shows comparable efficiency of the pairwise and the trivariate estimation procedures. Comparison of columns 2 and 5 indicates that the pseudo-likelihood approach indeed yields standard errors which reflect the true sampling variability. As was to be expected, the model-based and the robust standard errors from the trivariate model are very similar because the correct model was fitted. Under the misspecified model (Table 2), we still get equally efficient estimates under both approaches, the pseudo-likelihood standard errors are still valid, but the model-based ones obtained from the trivariate model no longer are.

[Table 1 about here.]

[Table 2 about here.]

7 Discussion

In this paper we have presented a joint-modelling approach designed to model high-dimensional multivariate longitudinal data. The approach is based on fitting bivariate mixed models for all pairs of outcomes. As long as each bivariate mixed model can be fitted, estimates can be obtained for the full multivariate mixed model. Obviously, this approach is advantageous whenever fitting the full multivariate mixed model is not possible or too time-consuming.

The approach can also be considered as an example of partition(ed) maximum likelihood, an example of which can be found in Poon and Lee (1987), in a completely different context. Their inference was based on simply averaging all pair-specific standard errors, which clearly does not reflect the true sampling variability.

The approach has been illustrated using data on hearing thresholds. The model used in this example is characterised by outcome-specific fixed and random effects. However, a joint model can be simplified by assuming random and/or fixed effects shared by all outcomes or by a set of the outcomes, such that more parsimonious models can be used or specific research questions can be answered.

Although the methodology has been presented in the context of longitudinal studies, the pairwise approach is also applicable if the data are gathered in a clustered setting. The example on the hearing data restricts attention to the case of normal data. However, the same strategy to circumvent computational problems can be applied in the context of generalized linear mixed models or nonlinear mixed models, or in combinations of those.

Acknowledgments

Both authors acknowledge support from the Interuniversity Attraction Poles Program P5/24 - Belgian State - Federal Office for Scientific, Technical and Cultural Affairs. They also wish to thank Larry Brant for providing the hearing data from the Baltimore Study of Aging.

References

Arnold, B.C. and Strauss, D. (1991), "Pseudolikelihood estimation: Some examples", *Sankyhya, Series B*, **53**, 233–243.

Besag, J. (1975), "Statistical analysis of non-lattice data", *The Statistician*, **24**(3), 179–195.

Brant, L.J. and Fozard, J.L. (1990), "Age changes in pure-tone hearing thresholds in a longitudinal study of normal human aging", *Journal of the Acoustical Society of America*, **88**, 813–820.

Breslow, N.E. and Clayton, D.G. (1993), "Approximate inference in generalized linear mixed models", *Journal of the American Statistical Association*, **88**, 9–25.

Burzykowski, T., Molenberghs, G., Buyse, M., Geys, H., and Renard, D. (2001), "Validation of surrogate endpoints in multiple randomized clinical trials with failure time end points", *Applied Statistics*, **50**, 405–422.

Buyse, M., Molenberghs, G., Burzykowski, T., Renard, D., and Geys, H. (2000), "The validation of surrogate endpoints in meta-analyses of randomized experiments", *Biostatistics*,

1, 49–67.

Chakraborty, H., Helms, R.W., Sen, P.K., and Cohen, M.S. (2003), "Estimating correlation by using a general linear mixed model: Evaluation of the relationship between the concentration of HIV-1 RNA in blood and semen", *Statistics in Medicine*, **22**, 1457–1464.

Davidian, M. and Giltinan, D.M. (1995), *Nonlinear models for repeated measurement data*, Chapman & Hall.

Diggle, P.J., Heagerty, P., Liang, K.Y., and Zeger, S.L. (2002), *Analysis of longitudinal data*, Clarendon Press, Oxford.

Fieuws, S. and Verbeke, G. (2004), "Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach", *Statistics in Medicine*, in press.

Geys, H., Molenberghs, G., and Ryan, L. (1997), "Pseudo-likelihood inference for clustered binary data", *Communications in Statistics: Theory and methods*, **26**(11), 2743–2767.

Gueorguieva, R. (2001), "A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family", *Statistical Modelling*, **1**(3), 177–193.

Laird, N.M. and Ware, J.H. (1982), "Random-effects models for longitudinal data", *Biometrics*, **38**, 963–974.

Liang, K.Y. and Zeger, S.L. (1986), "Longitudinal data analysis using generalized linear models", *Biometrika*, **73**, 13–22.

MacCallum, R., Kim, C., Malarkey, W., and Kiecolt-Glaser, J. (1997), "Studying multivariate change using multilevel models and latent curve models", *Multivariate Behavioral Research*, **32**, 215–253.

Morrell, C.H. and Brant, L.J. (1991), "Modelling hearing thresholds in the elderly", *Statistics in Medicine*, **10**, 1453–1464.

Pearson, J.D., Morrell, C.H., Gordon-Salant, S., Brant, L.J., Metter, E.J., Klein, L.L., and Fozard, J.L. (1995), "Gender differences in a longitudinal study of age-associated hearing loss", *Journal of the Acoustical Society of America*, **97**, 1196–1205.

Poon, W.Y. and Lee, S.Y. (1987), "Maximum likelihood estimation of multivariate polyserial and polychoric correlation coefficients", *Biometrika*, **52**, 409–430.

Renard, D., G.Molenberghs, and Geys, H. (2004), "A pairwise likelihood approach to estimation in multilevel probit models", *Computational statistics and data analysis*, **44**, 649–667.

Shah, A., Laird, N., and Schoenfeld, D. (1997), "A random-effects model for multiple characteristics with possibly missing data", *Journal of the American Statistical Association*, **92**, 775–779.

Shock, N.W., Greullich, R.C., Andres, R., Arenberg, D., Costa, P.T., Lakatta, E.G., and Tobin, J.D. (1984), "Normal human aging : The Baltimore Longitudinal Study of Aging", *National Institutes of Health publication 84-2450*.

Thiébaut, R., Jacqmin-Gadda, H., Chêne, G., and Commenges, D. (2002), "Bivariate longitudinal study of CD4+ cell count and HIV RNA taking into account informative drop-out and left censoring of HIV RNA values", in M. Stasinopoulos and G. Touloumi, editors, *Proceedings of the 17th International Workshop on Statistical Modelling*, 521–524, Odense, Denmark.

Thiébaut, R., Jacqmin-Gadda, H., Chêne, G., Leport, C., and Commenges, D. (2002), "Bivariate linear mixed models using SAS PROC MIXED", *Computer Methods and Programs in Biomedicine*, **69**, 249–256.

Thum, Y.M. (1997), "Hierarchical linear models for multivariate outcomes", *Journal of Educational and Behavioral Statistics*, **22**(1), 77–108.

Verbeke, G. and Molenberghs, G. (2000), *Linear mixed models for longitudinal data*, Springer Series in Statistics, Springer-Verlag, New-York.

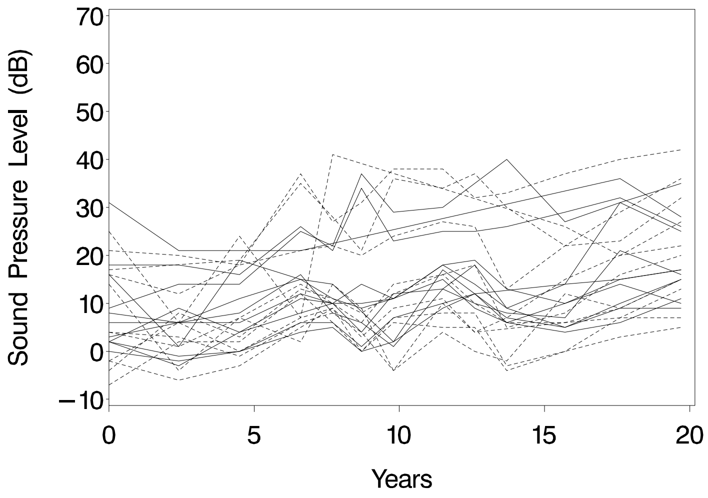

Figure 1: *Ear- and frequency specific profiles of hearing thresholds of 1 randomly selected subject (dotted and solid lines represent the frequency specific profiles at the left ear and right ear respectively)*

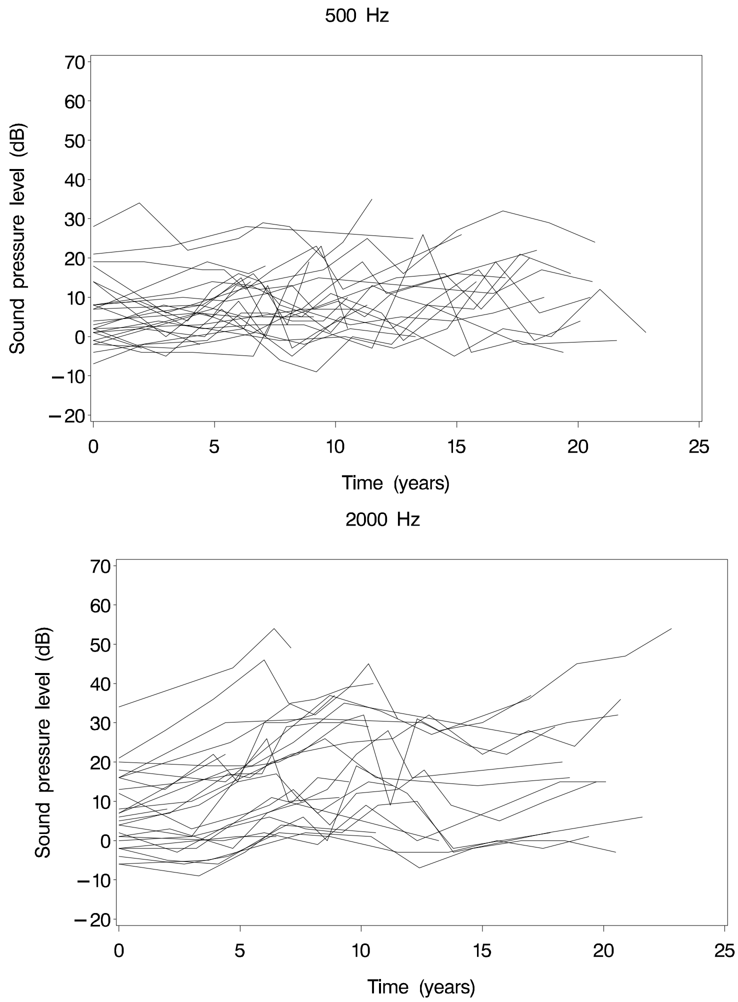

Figure 2: *Individual profiles of hearing thresholds at the right ear for 500 Hz and 2000 Hz (of 30 randomly selected subjects).*

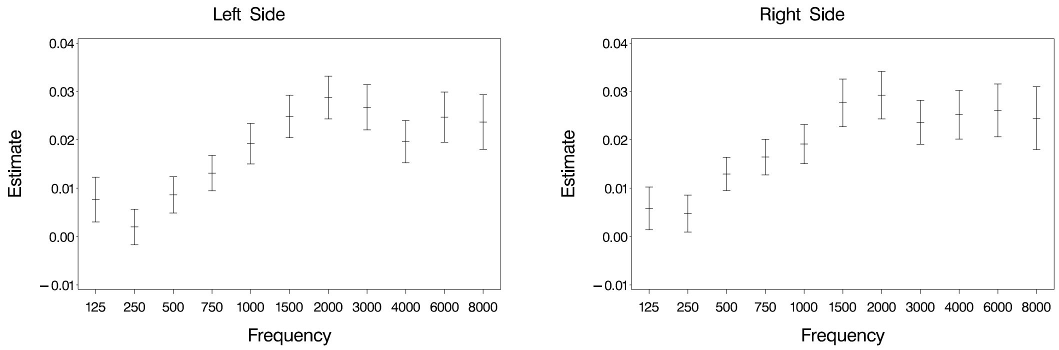

Figure 3: *Outcome-specific estimates and their 95% confidence interval for the interaction between the linear evolution and age at entry.*

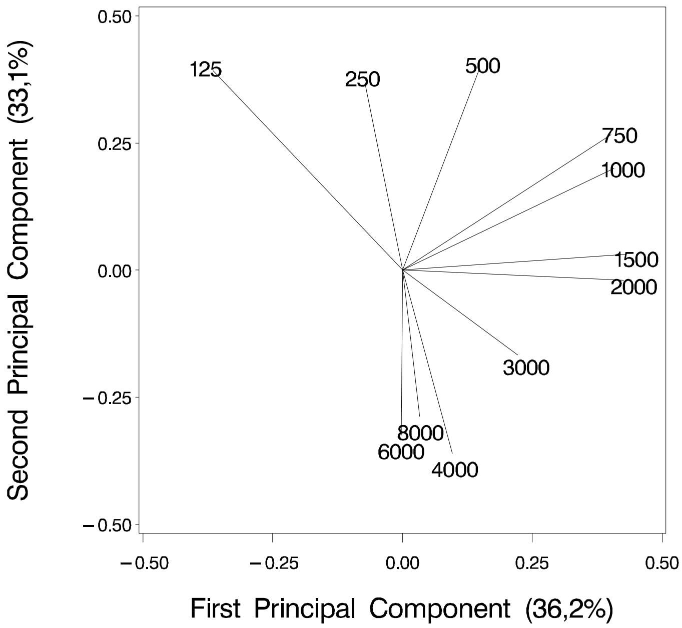

Figure 4: *Results (component loadings for first and second principal component) of a PCA on the* 11×11 *correlation matrix of the random slopes at the left side.*

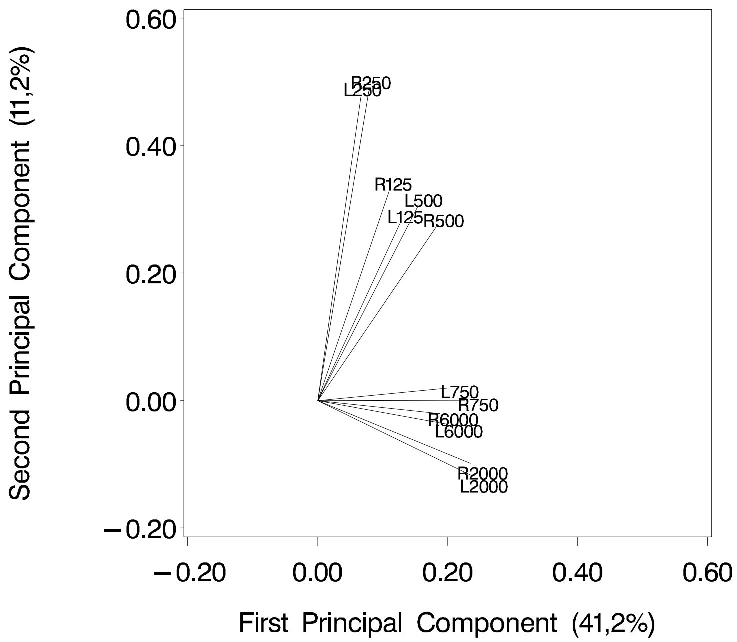

Figure 5: *Results of a PCA on the correlation matrix of the error components. L=left, R=right (note: not all frequencies are presented in this figure).*

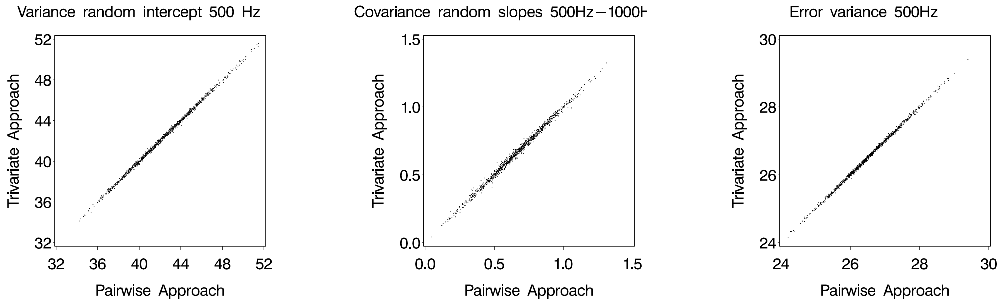

Figure 6: *Results Simulation Study. Agreement between the estimates obtained with the full trivariate and with the pairwise approach. Plots are given for three covariance parameters, (1) the variance of the random effects of 500 Hz, (2) the covariance between the random slopes of 500 Hz and 1000 Hz and (3) the variance of the error components of 500 Hz.*

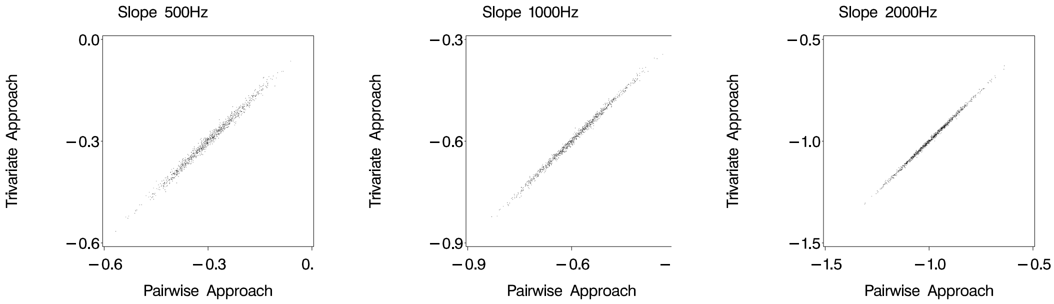

Figure 7: *Results Simulation Study. Agreement between the estimates obtained with the full trivariate and with the pairwise approach. Plots are given for three fixed effects, the slopes of the frequencies 500 Hz, 1000 Hz and 2000 Hz*

Table 1: *Results Simulation Study: Correct Model. Standard deviation (SD) of the estimates obtained from the trivariate and the pairwise approach, as well as the means of the standard errors (model-based and robust) obtained with the pairwise and with the trivariate approach.*

	SD(of estimate)		**Mean(of SE)**		
Effect	Trivariate	Pairwise	Trivariate model-based	Trivariate robust	Pairwise
slope 500 Hz	0.0795	0.0801	0.0783	0.0780	0.0785
slope 1000 Hz	0.0847	0.0848	0.0843	0.0840	0.0842
slope 2000 Hz	0.1082	0.1081	0.1101	0.1097	0.1099

Table 2: *Results Simulation Study: Incorrect Model. Standard deviation (SD) of the estimates obtained from the trivariate and the pairwise approach, as well as the means of the standard errors (model-based and robust) obtained with the pairwise and with the trivariate approach.*

	SD(of estimate)		**Mean(of SE)**		
Effect	Trivariate	Pairwise	Trivariate model-based	Trivariate robust	Pairwise
slope 500 Hz	0.0836	0.0842	0.0644	0.0813	0.0817
slope 1000 Hz	0.0912	0.0910	0.0628	0.0898	0.0895
slope 2000 Hz	0.1192	0.1188	0.0791	0.1202	0.1198