

## Mixed model approaches for joint modeling of different types of responses

Peer-reviewed author version

IVANOVA, Anna; MOLENBERGHS, Geert & VERBEKE, Geert (2015) Mixed model approaches for joint modeling of different types of responses. In: Journal of Biopharmaceutical Statistics, 26 (4), p. 601-618.

DOI: 10.1080/10543406.2015.1052487

Handle: <http://hdl.handle.net/1942/20856>

# Mixed Models Approaches for Joint Modeling of Different Types of Responses

Anna Ivanova<sup>1,2</sup>, Geert Molenberghs<sup>3,2</sup>, Geert Verbeke<sup>2,3</sup>

<sup>1</sup>Leuven Statistics Research Centre, KU Leuven, Belgium

<sup>2</sup>I-BioStat, KU Leuven, Belgium

<sup>3</sup>I-BioStat, Universiteit Hasselt, Belgium

Email: anna.ivanova@lstat.kuleuven.be

## Abstract

In many biomedical studies one jointly collects longitudinal continuous, binary, and survival outcomes, possibly with some observations missing. Random-effects models, sometimes called shared-parameter models or frailty models, received a lot of attention. In such models, the corresponding variance components can be employed to capture the association between the various sequences. In some cases, random effects are considered common to various sequences, perhaps up to a scaling factor; in others there are different but correlated random effects. Even though a variety of data types has been considered in the literature, less attention has been devoted to ordinal data. For univariate longitudinal or hierarchical data, the proportional odds mixed model (POMM) is an instance of the generalized linear mixed model (GLMM; Breslow and Clayton 1993). Ordinal data are conveniently replaced by a parsimonious dummies, which in the longitudinal setting leads to a repeated set of dummies. When ordinal longitudinal data are part of a joint model, the complexity increases further. This is the setting considered in this paper. We formulate a random-effects based model that, in addition, allows for overdispersion. Using two case studies, it is shown that the combination of random effects to capture association with further correction for overdispersion can improve the model's fit considerably and that the resulting models allow to answer research questions that could not be addressed otherwise. Parameters can be estimated in a fairly straightforward way, using the SAS procedure NLMIXED.

**Some Keywords:** Joint modeling; Linear mixed model; Generalized linear mixed model; Proportional odds mixed model; Maximum likelihood.

## 1 Introduction

When conducting longitudinal studies, investigators often find themselves faced with a combination of continuous, binary, ordinal, and even other data types. Many statistical models have been developed for analyzing longitudinal data and at a given time the random-effects approach became very popular. The introduction of linear mixed models for continuous data by Laird and Ware (1982) was extended to non-continuous data (generalized linear mixed models) by Breslow and Clayton (1993), Wolfinger and O'Connell (1993), and Engel and Keen (1994). The proportional odds mixed model can be regarded as a special case of the latter as it also considers ordinal responses.

In a longitudinal context, joint modeling is sometimes preferred over a separate analysis of the different responses given that it allows every outcome to have its own random effects and the association between different outcomes can be captured in terms of the correlation between random effects. The big advantage is that researchers can obtain answers to various research questions, all in one format. The average trend for different outcomes can be estimated and the difference in evolution between several outcomes tested.

A further advantage is that the effect of a covariate on the outcomes can be assessed simultaneously, rather than outcome by outcome or by assuming that the outcomes are independent of each other, usually contrary to fact. For example, in the case of drug treatment, the duration of the disease prior to the start of the trial can simultaneously diminish the effect of therapy and increase the severity of side effects. It is possible that each of these effects separately fail to show up significantly, while jointly they may pass the bar. In addition, the association between various responses can be measured, and its evolution over time monitored. Chakraborty *et al* (2003) considered joint modeling in a continuous-continuous setting and obtained estimates of the correlation between HIV-1 RNA concentration in blood and semen.

As already mentioned, very often longitudinal studies generate data that are of different types. In the literature, a wide array of techniques can be found to address specific instances. It is the main reason for the variety in modeling applications reported in the literature. For example, Thiébaut *et al* (2002) performed the joint modeling of two continuous outcomes and applied it to the analysis of some HIV infection markers. Gueorguieva (2001) considered the joint modeling of continuous-binary measures in a toxicity study of pregnant/non-pregnant mice. Also, Iddi and Molenberghs (2012) considered joint modeling of continuous visual activity and a binary vision-loss outcome in an age-related macular degeneration study; these authors also considered two binary longitudinal outcomes: the number of positive HCV and HIV cases in serological data. Molenberghs and Verbeke (2005) discuss a number of techniques that jointly model continuous and binary outcomes. Even the high-dimensionality issue, when including (substantially) more than two responses in the joint model, received a certain degree of attention in the research community. For example, Morrell *et al* (2012) considered three continuous responses for screening prostate data, and further used the outcome for classification purposes. Fieuws and Verbeke (2006) suggested a pairwise bivariate modeling approach for continuous data to resolve the computational complexity issue when jointly modeling random effects in the high-dimensional case. An extension of this approach was the application of the pairwise technique on binary data (Fieuws *et al* 2006) and on the combination of continuous-binary data types (Fieuws *et al* 2008). The latter approach was used for predicting renal graft failures based on a set of biochemical and physiological markers. The prediction improvement obtained when considering joint modeling can also be regarded as very important. When one wishes to improve the quality of the survival prediction of a patient, then one should have the option to include several of the available responses or even all of them in the analysis. An example is the work of Horrocks and van den Heuvel (2009). They used a joint continuous-binary mixed model to predict pregnancy in a group of women undergoing infertility treatment given measurements of adhesiveness of certain blood lymphocytes.

There are also many approaches to joint modeling of longitudinal and time-to-event data. For an excellent relatively early review, see Tsiatis and Davidian (2004). These techniques are intended to be flexible and cope with different types of longitudinal outcomes. We mention in particular the implementations developed by Xu and Zeger (2001), Yu *et al* (2004), Efendi *et al* (2013), and Njeru Njagi *et al* (2013), among others. Rizopoulos (2012) offers a recent review and considered applications using R for joint modeling longitudinal and time-to-event data. The important point of such a

joint setting is the correction for non-random dropouts. Due to the ability of joint modeling to provide valid inferences for cases where traditional techniques fail, they became popular in recent years. Li *et al* (2010) considered joint models for an ordinal outcome, with the partial proportional odds assumptions, and time-to-event outcomes, in a competing risks setting. These authors developed an EM algorithm for parameter estimation. Some packages with specific interest, such as OpenMx, can handle versions of such joint models. Our approach, in contrast, allows the easy use of standard software tools, such as the NLMIXED procedure in SAS. In particular, this procedure will allow us to formulate parts of the models in a multiplicative fashion; this is useful for scale parameters that allow sharing the same effect between different outcomes that are recorded on different measurement scales.

Combinations of continuous and binary, as well as continuous and time-to-event data types in joint modeling have been sufficiently covered in the literature. But what is lacking is the case of one or more ordinal responses in joint modeling. The focus of this article is on bivariate analysis with the modeling done by joining a linear mixed model and a proportional odds mixed model.

The remainder of the paper is organized as follows. Section 2 introduces two datasets that are further analyzed in Section 6. The existing methodology for single longitudinal responses is reviewed together with the joint modeling approach in Section 3. Joint models for various outcome types are discussed in Section 4. The estimation method is described in Section 5.

## 2 Case Studies

### 2.1 Diabetes Study

In Belgium, the diabetes project was conducted from January 2005 until December 2006, with the aim to study the effect of implementing a structured model for chronic diabetes care based on the patients' clinical outcomes. General practitioners (GP's) were offered assistance and could redirect patients to the diabetes care team, consisting of a nurse educator, a dietician, an ophthalmologist, and an internal medicine doctor. A total of 120 GP's and 2495 patients took part in the study.

During the project, several outcomes useful to evaluate how well diabetes is controlled were measured, at the moment the program was initiated (time  $T_0$ ) and one year later ( $T_1$ ). The most important outcomes were HbA1c (glycosylated hemoglobin), LDL-cholesterol (low-density lipoprotein cholesterol) and SBP (systolic blood pressure). Furthermore, experts specified cut-off values defining a so-called *clinical targets* for all outcomes:  $HbA1c < 7\%$ ,  $LDL\text{-cholesterol} < 100\text{ mg/dl}$  and  $SBP \leq 130\text{ mmHg}$ . As a result, for a particular time point, every patient could reach minimum 0 – maximum 3 clinical targets. If at least one measurement per patient was missing, the value for the number of clinical targets ( $CT$ ) was set to missing as well. The data are discussed in detail in Borgermans *et al* (2009). A summary is given in Table 1. Arguably, the definition of CT is debatable; one could consider alternative components, and even treat them jointly rather than in the form of a sum score, we will follow the original definition, for compatibility with the protocol and the study's basic analysis.

Several continuous outcomes were collected as well. In this paper, we will study the BMI (body mass index) and build a joint model for BMI and CT, in order to be able to study the association between both outcomes, and how it evolves over time, as well as to explore what sociological and demographic factors are related to both outcomes simultaneously. Note that BMI is not part of the outcomes producing clinical targets, which makes studying BMI and CT a meaningful endeavor.

**Table 1:** *Diabetes Data: Number of observations with the corresponding clinical targets reached at every time-point.*

# Clin. Targets	# Observations	
	$T_0$	$T_1$
0	307	128
1	828	598
2	726	834
3	281	501

**Table 2:** *Diabetes Data: Descriptive statistics for BMI and log of BMI at every time-point.*

	$T_0$				$T_1$			
	$Q_{0.25}$	Mean	Median	$Q_{0.75}$	$Q_{0.25}$	Mean	Median	$Q_{0.75}$
BMI	25.94	29.60	29.06	32.46	25.73	29.28	28.87	32.10
log(BMI)	3.26	3.37	3.37	3.48	3.25	3.36	3.36	3.47

Given that the distribution of BMI is skewed, a logarithmic transformation will be used in all our analyses. Some descriptive statistics, at both time points, are given in Table 2.

## 2.2 Fluvoxamine Trial

The data are from a multicentre study involving 315 patients that were treated by fluvoxamine for psychiatric symptoms described as possibly resulting from a dysregulation of serotonin in the brain. The data are discussed in Molenberghs and Lesaffre (1994), Kenward, Lesaffre, and Molenberghs (1994), Molenberghs, Kenward, and Lesaffre (1997), Michiels and Molenberghs (1997), and Molenberghs and Verbeke (2005). After recruitment to the study, patients were assessed at four visits. The therapeutic effect and the extent of side effects were scored at each visit on an ordinal scale. The side effect response is coded as 1: none; 2: not interfering; 3: interfering significantly with functionality; 4: side effects surpasses the therapeutic effect. Similarly, the effect of therapy is recorded on a four point ordinal scale: 1: no improvement or worsening; 2: minimal improvement; 3: moderate improvement and 4: important improvement. Thus, a side effect occurs if new symptoms occur while there is therapeutic effect if old symptoms disappear. A total of 299 patients have at least one measurement, including 242 completers. A summary is given in Table 3. There is also baseline covariate information on each subject, including gender, age, presence of psychiatric antecedents, initial severity of the disease, duration of the actual mental illness. Obviously, it would be expected that in the context of this study the improvement of the patient's condition measured by the therapeutic effect is correlated with the side effect of the therapy, and this association can only be studied if both outcomes are modeled jointly.

**Table 3:** *Fluvoxamine Trial. Number of observations with side and therapeutic effects categories for each of the four follow-up time points.*

Ther. Effect	# Observations			
	Week 2	Week 4	Week 8	Week 12
0	19	64	110	135
1	95	114	93	62
2	102	62	30	19
3	83	29	10	10
Side Effect	Week 2	Week 4	Week 8	Week 12
0	128	144	156	148
1	128	103	79	71
2	28	17	6	7
3	15	5	2	0

### 3 Models for a Single Longitudinal Response

#### 3.1 General Mixed Model

Assume that each of the longitudinal outcomes can be modeled using a mixed model. For one outcome, let  $Y_{ij}$  denote the  $j$ th measurement for subject  $i = 1, \dots, N$ ,  $j = 1, \dots, n_i$ . The  $n_i$  measurements are grouped into a vector  $\mathbf{Y}_i$ . The main model assumption is that the  $\mathbf{Y}_i$ , conditionally on a  $q$ -dimensional vector  $\mathbf{b}_i$  of normally distributed, zero mean random effects, follow a pre-specified distribution  $F_i$ :  $\mathbf{Y}_i|\mathbf{b}_i \sim F_i(\boldsymbol{\psi}, \mathbf{b}_i)$ , which is allowed to depend on known covariates, and parameterized in terms of a vector  $\boldsymbol{\psi}$  of unknown parameters, common to all subjects. Often, it is assumed that the components of  $\mathbf{Y}_i$  are independent, conditionally on  $\mathbf{b}_i$ , but extensions are possible. Inference is based on the marginal distribution of  $\mathbf{Y}_i$  with density:

$$f(\mathbf{y}_i) = \int f(\mathbf{y}_i|\mathbf{b}_i)f(\mathbf{b}_i)d\mathbf{b}_i. \quad (1)$$

in which  $f(\mathbf{y}_i|\mathbf{b}_i)$  and  $f(\mathbf{b}_i)$  are the conditional density of  $\mathbf{Y}_i$  given  $\mathbf{b}_i$  and the density of  $\mathbf{b}_i$ , respectively. Estimation and inference of the parameters in (1) is based on the maximum likelihood (ML) approach, assuming independence across subjects (Laird and Ware 1982, Verbeke and Molenberghs 2000, Molenberghs and Verbeke 2005). Important special cases of mixed models are the linear mixed model (LMM) for continuous outcomes, and the generalized linear mixed model (GLMM) for discrete outcomes, which will be briefly discussed in the next sections.

#### 3.2 The Linear Mixed Model

The general formulation of a LMM is that  $\mathbf{Y}_i$  satisfies

$$\mathbf{Y}_i|\mathbf{b}_i \sim N(X_i\boldsymbol{\xi} + Z_i\mathbf{b}_i, \Sigma_i),$$

where  $X_i$  and  $Z_i$  are  $(n_i \times k)$  and  $(n_i \times q)$  dimensional matrices of known covariates,  $\boldsymbol{\xi}$  a  $k$ -dimensional vector of regression parameters, and  $\Sigma_i$  an  $(n_i \times n_i)$  covariance matrix that depends on

$i$  only through its dimension  $n_i$ . In case of conditional independence, we have  $\Sigma_i = \sigma^2 I_{n_i}$ . Finally, the marginal distribution of  $\mathbf{Y}_i$  equals

$$\mathbf{Y}_i \sim N(X_i \boldsymbol{\xi}, Z_i D Z_i' + \Sigma_i),$$

in which  $D$  is the covariance matrix of the random effects  $\mathbf{b}_i$ .

### 3.3 The Generalized Linear Mixed Model

In case of a GLMM (Engel and Keen 1994, Breslow and Clayton 1993, Wolfinger and O'Connell 1993), it is assumed that the outcomes  $Y_{ij}$ , conditionally on the random effects  $\mathbf{b}_i$ , with densities that belong to the exponential family, i.e., of the form:

$$f_i(y_{ij}|\mathbf{b}_i, \boldsymbol{\xi}, \phi) = \exp \left\{ \phi^{-1} [y_{ij} \lambda_{ij} - \psi(\lambda_{ij})] + c(y_{ij}, \phi) \right\},$$

with

$$\eta[\psi'(\lambda_{ij})] = \eta(\mu_{ij}) = \eta[E(Y_{ij}|\mathbf{b}_i, \boldsymbol{\xi})] = \mathbf{x}_{ij}' \boldsymbol{\xi} + \mathbf{z}_{ij}' \mathbf{b}_i,$$

in which  $\mathbf{x}_{ij}$  and  $\mathbf{z}_{ij}$   $k$ -dimensional and  $q$ -dimensional vectors of known covariate values,  $\boldsymbol{\xi}$  a  $k$ -dimensional vector of unknown fixed regression coefficients, and  $\phi$  a scale parameter. In contrast to the above result for linear mixed models, the integral in (1) cannot be calculated analytically for a number of generalized linear mixed models. Even if there are analytical expressions, they tend to be cumbersome (Molenberghs, Verbeke, and Demétrio 2007, Molenberghs *et al* 2010). Hence, numerical approximations are needed (see Molenberghs and Verbeke 2005).

### 3.4 The Proportional Odds Mixed Model

A special case of GLMM, of particular interest in this work, is the proportional odds mixed model (POMM) for ordinal outcomes. Let  $Y_{ij}$  be ordinal, with values  $r = 1, \dots, R$ . We first define  $R$  indicator variables as

$$W_{r,ij} = \begin{cases} 1 & \text{if } Y_{ij} = r, \\ 0 & \text{otherwise.} \end{cases}$$

Evidently, these are redundant, but any subset of  $R - 1$  components is not. Group the dummies into vectors  $\mathbf{W}_{ij}$  for a specific subject  $i$  and occasion  $j$ , and further into  $\mathbf{W}_i$  for all dummies across all occasions for subject  $i$ . We assume a multinomial distribution  $\mathbf{W}_{ij} \sim \text{multinomial}(\boldsymbol{\pi}_{ij})$ , with  $\boldsymbol{\pi}_{ij} = (\pi_{1,ij}, \dots, \pi_{r,ij}, \dots, \pi_{R,ij})$ . The multinomial distribution at a given occasion is determined by the modeling choice made for the ordinal outcome. The probabilities can be written as:

$$\pi_{r,ij} = \begin{cases} \kappa_{1,ij} & \text{if } r = 1, \\ \kappa_{r,ij} - \kappa_{r-1,ij} & \text{if } 1 < r < R, \\ 1 - \kappa_{R-1,ij} & \text{if } r = R. \end{cases} \quad (2)$$

where, assuming proportional odds:

$$\kappa_{r,ij} = \frac{\exp(\xi_{0r} + \mathbf{x}_{ik}' \boldsymbol{\xi} + \mathbf{z}_{ij}' \mathbf{b}_i)}{1 + \exp(\xi_{0r} + \mathbf{x}_{ik}' \boldsymbol{\xi} + \mathbf{z}_{ij}' \mathbf{b}_i)}.$$

Here,  $\xi_{01} \leq \dots \leq \xi_{0(R-1)}$  are intercepts,  $\boldsymbol{\xi}$  are fixed regression coefficients,  $\mathbf{b}_i$  is a vector of normally distributed random effects, and  $\mathbf{x}_{ij}$  ( $\mathbf{z}_{ij}$ ) is the design vector for the fixed (random) effects at

occasion  $j$ . Note that these parameter constraints are identical to the ones routinely used with proportional odds logistic regression, i.e., they already commonly occur with univariate ordinal outcomes (Agresti, 1990).

## 4 Joint Models for Multiple Outcomes

Consider two longitudinal outcomes, for example, a continuous one and an ordinal one. Furthermore, let  $Y_{1ij}$  and  $Y_{2ij'}$  denote the  $j$ th and the  $j'$ th measurement on the  $i$ th subject for the continuous and the ordinal outcome, respectively ( $i = 1, \dots, N$ ,  $j = 1, \dots, n_{1i}$ , and  $j' = 1, \dots, n_{2i}$ ). A joint model is build by describing the joint density  $f(\mathbf{Y}_{1i}, \mathbf{Y}_{2i})$  of the continuous vector  $\mathbf{Y}_{1i}$  and ordinal vector  $\mathbf{Y}_{2i}$ . A very flexible way forward to this end is by considering a mixed model for both outcomes, and to allow the random effects in these to be correlated. This allows for joint modeling of outcomes of a different nature and does not alter the interpretation of the parameters in the sub-models for individual outcomes. This approach has been applied already by Fieuws and Verbeke (2006) and Iddi and Molenberghs (2012), but it has not been used in the context of ordinal outcomes.

The mixed models assumed for each outcome are as described in Section 3, i.e., a LMM for a continuous outcome and a POMM for an ordinal outcome. In case two ordinal outcomes are to be analyzed jointly, as in the fluvoxamine trial, two POMM's would be used. Note also that the covariates in both models may but do not need to be the same. Let  $\mathbf{b}_{1i}$  and  $\mathbf{b}_{2i}$  be the vectors of random effects for the first and second outcome, respectively. A joint model is now obtained by assuming a distribution for both sets of random effects  $\mathbf{b}_{1i}$  and  $\mathbf{b}_{2i}$  jointly. More specifically, it will be assumed that  $\mathbf{b}_i = (\mathbf{b}_{1i}, \mathbf{b}_{2i})'$  is normally distributed with mean zero and covariance matrix  $D$ , which contains components to model the association within each outcome sequence separately, as well as components to model the association between the sequences. It will also be assumed that, conditionally on  $\mathbf{b}_i$ , the outcome vectors  $\mathbf{Y}_{1i}$  and  $\mathbf{Y}_{2i}$  are independent, i.e., we assume that the association between the outcome vectors is completely captured by the association between the random effects.

Often, a general unstructured matrix  $D$  will be assumed, but specific restrictions can be imposed as well. For example, assuming perfect correlations between elements in  $\mathbf{b}_{1i}$  and elements in  $\mathbf{b}_{2i}$  would lead to a joint model in which some random effects are shared between the two outcomes (Molenberghs and Verbeke 2005). Evidently, this is equivalent to sharing components between  $\mathbf{b}_{1i}$  and  $\mathbf{b}_{2i}$ . When both outcomes are of a different nature, some rescaling may be necessary, leading to a model with perfectly correlated random effects but with variances that differ by a scaling factor.

## 5 Estimation and Inference

Assuming subjects to be independent, then it immediately follows from the independence of  $\mathbf{Y}_{1i}$  and  $\mathbf{Y}_{2i}$ , conditional on  $\mathbf{b}_i$ , that the likelihood function corresponding to the joint model is given by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^N \int f_{1i}(\mathbf{y}_{1i}|\mathbf{b}_{1i})f_{2i}(\mathbf{y}_{2i}|\mathbf{b}_{2i})f(\mathbf{b}_i)d\mathbf{b}_i, \quad (3)$$

in which  $\boldsymbol{\theta}$  is the vector of all parameters in the conditional distributions and the multivariate normal distribution for  $\mathbf{b}_i$ . Except for special cases (e.g., with linear models), the integral in (3) cannot be calculated analytically and numerical approaches are needed. In this paper, we will use numerical



integration, more specifically adaptive Gaussian quadrature, which has been implemented in the SAS procedure NLMIXED (Pinheiro and Bates 1995, 2000, Molenberghs and Verbeke 2005). The order  $Q$  of the integration is determined manually by fitting the model for increasing values of  $Q$  until numerical stability is obtained in the approximated likelihood value and parameter estimates. An example of the implementation can be found in the Appendix. Once the model has been fitted, inferences for all elements in  $\theta$  become available using standard likelihood theory (e.g., likelihood ratio tests, Wald tests, score tests).

## 6 Analysis of Case Studies

### 6.1 Diabetes Study

We will analyze the diabetes data, introduced in Section 2.1. In the analysis, the longitudinal continuous outcome  $\log(BMI)$  and the ordinal outcome *clinical targets* (CT) on a scale from 0 to 3, will be used. For simplicity, we will use the same covariate structure across all models implemented. For  $\log(BMI)$ , we assume the predictor to be of the form:

$$E(Y_{1ij}) = \xi_{0,1} + \xi_{1,1}t_{ij} + \xi_{1,2}X_{1,i} + \xi_{1,3}X_{2,i} + \xi_{1,4}X_{3,i},$$

and for CT:

$$\text{logit}[P(Y_{2ij} \leq r)] = \xi_{2,0r} + \xi_{2,1}t_{ij} + \xi_{2,2}X_{1,i} + \xi_{2,3}X_{2,i} + \xi_{2,4}X_{3,i},$$

where  $t_{ij}$  is the time point at which outcome  $j$  is measured, i.e.,  $t_{ij} = 0$  or  $1$ . In the corresponding conditional models, the following random effects are included:  $b_{1i}$  and  $b_{2i}$  are the random intercept and random slope for  $\log(BMI)$ , and  $b_{3i}$  is the random intercept for CT. To capture the correlation between the responses, various assumptions about the distribution of the random effects can be made. For example, if we want to implement Model 4 (as in Table 4), with: (a) uncorrelated random intercept and random slope for  $\log(BMI)$ , (b) random intercept for CT, and (c) further with the random intercepts taken to be correlated, then the following structure emerges:

$$\begin{pmatrix} b_{1i} \\ b_{2i} \\ b_{3i} \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} d_{11} & 0 & d_{13} \\ 0 & d_{22} & 0 \\ d_{13} & 0 & d_{33} \end{pmatrix} \right]. \quad (4)$$

Upon simplifying this structure, by backward model selection, and removing the random slope for  $\log(BMI)$  from the model, we obtain Model 2, with two correlated random intercepts. If we further remove the correlation between the random intercepts, then separate modeling of the responses follows, leading to Model 1. Also, a joint model was considered, fitted by making use of a random intercept shared between both outcomes, with of course an inflation factor ( $\lambda$ ) to account for the different scales at which they are measures; this is Model 3. In this case, the random effects vector in (4) reduces to a single, normally distributed random effect only. Evidently, rather than simplifying the model, one can also consider more elaborate models, for example, with correlated random intercept and slope for  $\log(BMI)$ , etc. From a data-analytic view-point, if the  $D$  matrix is overly simple, then incorrect inferences might result. In many cases, there will come a point where fitting such more and more complex models becomes computationally challenging.

Table 4 shows the modeling results. Because Models 1, 2, and 4 are nested, we can compare them using likelihood ratio tests. We start with the most complex Model 4 and compare it with the

simpler Model 2, and then Model 2 with the simplest Model 1. The deviance for the comparison of Model 4 with Model 2 equals 4.2. Care needs to be taken when selecting the null distribution. As explained in Verbeke and Molenberghs (2000, 2003) and Molenberghs and Verbeke (2007), two different points of view can be adopted: a hierarchical and a marginal view. In a hierarchical view, scientific interest is directed, at least in part, to the random effects. This means that a hierarchical (two-stage) view is maintained, with subjects and their individual random effects at the first stage, and repeated measurements given the subject at the second. Marginally, one is merely interested in the population-level effects (fixed effects) and variance components. The hierarchical model is then used merely as a vehicle to conveniently and ideally parsimoniously generate the marginal distribution for the outcomes.

Thus, in the hierarchical view, the variance components are taken to describe random effects and have the meaning of a variance. According to this, the null value lies on the boundary of the parameter space. Such a non-traditional situation for testing was discussed Stram and Lee (1994, 1995), Self and Liang (1987), Verbeke and Molenberghs (2003) and Molenberghs and Verbeke (2007). The null distributions of the likelihood ratio, score, and Wald tests then take the form of a mixture of  $\chi^2$  distributions. The precise form of the mixture depends on the geometry of the null space. For a single variance parameter, as in our case, it becomes a 50:50 mixture of a  $\chi_0^2$  (the degenerate distribution with mass in 0) and a  $\chi_1^2$  distribution. The likelihood ratio test yields  $p = 0.5P(\chi_0^2 \geq 4.2) + 0.5P(\chi_1^2 \geq 4.2) = 0.02$ . However, in the marginal view, only the marginal distribution needs to be valid (i.e., the one containing the fixed-effects to describe the mean and all variance components combined to describe the variance-covariance matrix  $V_i = Z_i D Z_i' + \Sigma_i$ ; not the distribution conditional on the random effects, for which  $D$  and  $\Sigma_i$  are needed to be positive definite separately, rather than merely as contributions to  $V_i$ ). This obviously is a weaker condition because then only the corresponding marginal variance-covariance matrix needs to be positive-definite and the boundary issue on the variance components describing the random effects vanishes. Variance parameters then reduce to a mere variance component. Hence, in the marginal case, the likelihood ratio produces  $p = P(\chi_1^2 \geq 4.2) = 0.04$ . Clearly, the  $p$ -value of the marginal test is in this case twice that of its hierarchical version. The comparison of Models 2 and 1 is different from the previous case, because here the models differ in the covariance, which is a non-boundary problem. The likelihood ratio test statistic of 28.27 is then simply compared with a  $\chi_1^2$  as the reference distribution, and produces  $p < 0.001$ . Hence, a highly significant association between  $\log(BMI)$  and *clinical targets* is obtained, a conclusion that could not have been reached without modeling both outcomes jointly.

Model 3 is not nested in any other model but the difference in log-likelihood with Models 1, 2, and 4 is very large: 205.55, 233.82, and 238.02, respectively. We can conclude with high confidence that the performance of Model 3 is much poorer in comparison to the other models considered.

We further consider the effect of the covariate *practice with more than two GPs*. To assess the importance of this covariate, a likelihood ratio test is applied. For example, for Model 1 we obtain  $38298.67 - 38294.27 = 4.4$  for the  $\chi_2^2$  distributed test statistic with corresponding  $p = P(\chi_2^2 \geq 4.40) = 0.11$ . The result of this test for Model 2 is  $p = P(\chi_2^2 \geq 4.64) = 0.10$ , for Model 3,  $p = P(\chi_2^2 \geq 5.89) = 0.05$ , and for Model 4,  $p = P(\chi_2^2 \geq 4.58) = 0.10$ . Clearly, we did not observe significance of the test for the new covariate in any of the aforementioned models. Of course, our results suggest that a change in significance could occur in different but similar situations, underscoring the need for joint modeling. However, by going from separate modeling (Model 1) to the models with associated random intercepts (Model 2 and 4) the  $p$ -value for testing the importance of the new covariate decreased by 0.01. This suggests that careful joint modeling is necessary so as to reach proper conclusions. The  $p$ -value of the test for Models 2 and 3 decreases from 0.10 to 0.05 and is not

too far away from significance of the covariate. Knowing that the assumption of perfect association in Model 3 overestimates the relationship between the responses, we conclude that the significance of the new covariate cannot be considered adequate. This is why we advocate the consideration of different but correlated random effects, that place less stringent restrictions on the association structure.

In addition, to assess the effect on model fitting when including the correlation between random intercept and random slope into the continuous part of the joint model, an extended model (Model 5, see Table 6 in Appendix A) was implemented. When comparing the point and precision estimates of the covariance structure related to the  $\log(\text{BMI})$  part, we observe the presence of collinearity: the standard errors of the covariance components related to  $\log(\text{BMI})$  are very high in Model 5. The effect disappeared after removing the correlation between the random intercept and random slope and by adhering to Model 4. There is also no difference in log-likelihood between the two models. Obviously, when including the correlation between random intercept and random slope in the continuous part, the  $D$  matrix is over-specified.

Further, to investigate the sample size effect on the precision of the covariance component estimates, a series of targeted simulations were conducted. The selected scenario was similar to the diabetes study and was performed for 100 and 500 subjects measured longitudinally with two outcomes: continuous and ordinal. For each setting, 500 random samples were generated from the joint population of continuous and ordinal variables with the following covariance structure: correlated random intercept and random slope for the continuous part, and correlated random intercepts for both variables. Both correlations were set to the same magnitude: 0.4. Then, two joint models, one with uncorrelated and another with correlated random intercept and random slope for the continuous variable were fitted. They can be considered similar to the aforementioned Models 4 and 5, respectively. For both sample sizes, the mean squared error (MSE) was calculated for each parameter. Finally, all results were summarized in Table 7 (see Appendix A). For the case of 100 subjects, for Models 4 and 5, the MSE of the covariance components turned out to be relatively large. However, when considering the case of 500 subjects, the MSE of the covariance components becomes very small. Hence, we can conclude that, when collecting sufficiently large data sets, the inferences of the covariance components completely fall under the asymptotic theory of estimation by maximization of the likelihood. Under this condition, the MSE for the covariance component becomes stable and at the same time there is no observable bias issue after simplifying the structure of the  $D$  matrix by removing the correlation between random intercept and random slope for the continuous variable.

## 6.2 Fluvoxamine Trial

Recall from Section 2.2 that the study encompasses four time points, at each of which side effects and therapeutic effect are assessed on a 4-point ordinal scale. Let  $Y_{1ij}$  be the score for therapeutic effect for patient  $i$  at time point  $j$ . Then, proportional odds logistic regression with random intercept can be expressed as follows:

$$\text{logit}[P(Y_{1ij} \leq r)] = \xi_{1,0r} + b_{1i} + \xi_{1,11}t_{1ij} + \xi_{1,12}t_{2ij} + \xi_{1,13}t_{3ij} + \xi_{1,2}X_{1i} + \dots + \xi_{1,5}X_{4i},$$

where  $t_{1ij}$ ,  $t_{2ij}$ , and  $t_{3ij}$  are dummies corresponding to weeks 4, 8, and 12, respectively. In a similar way, the model for proportional odds model with random intercept  $b_{2i}$  for side effect can be defined.

Clinically, one cannot separate therapeutic effect from side effects, because the first one refers to the disappearance of existing symptoms, whereas the second one corresponds to emerging symptoms. This calls for Model 1 in Table 5, with two correlated random intercepts. This model is based on the following assumptions for the random effects:

$$\begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix} \right] \quad (5)$$

Similar to the diabetes data, a shared random intercept version of a joint hierarchical model can be formulated and fitted with scale factor  $\lambda$  in the therapeutic-effect part of modeling. These models are non-nested. Again, the difference in log-likelihood is very large: 277.39. Based on this we can conclude that the best performance is for the model with correlated random effects.

As mentioned before, joint models have the advantage of allowing tests for covariate effects on both outcomes simultaneously. This is of particular importance in the fluvoxamine study, given the logically intertwined therapeutic and side-effects outcomes. For example, focus on the effect of the importance of the duration of the disease. Under Model 1 from Table 5, a likelihood ratio test for the effect of duration on both outcomes simultaneously produces a test statistic equal to 52.17, which is highly significant (2 degrees-of-freedom,  $p < 0.001$ ). To illustrate the importance of accounting for the association between the outcomes, we also applied the joint test assuming independence of the therapeutic and side effects. The test-statistic then equals 52.29, evidently also with 2 degrees-of-freedom, and  $p < 0.001$ . In this case, the inference of the joint duration effect is quite similar due to the fact that the log-likelihood values for the models, assuming dependent and independent outcomes, are very close to each other: 3462.90 and 3465.12, indicative of the lack of association between the responses.

To explore whether the strength of this association could be explained, and hence reduced, by covariates, we compared Model 1 with separate models for the two responses, with only *time* as covariate. The models differ in covariance, which is a non-boundary problem. The likelihood ratio test statistic is 2.5 and is referred to a  $\chi^2_1$  with  $p = 0.11$ . The results of this test again confirm the presence of weak association between the responses.

## 7 Concluding Remarks

In this paper, we have focused on two approaches for the modeling of bivariate longitudinal sequences: (1) assuming mixed models for both outcomes with separate random effects that are correlated; (2) formulating a model with shared random effects, perhaps up to a scale parameter. Emphasis is placed on an ordinal and a continuous sequence on the one hand, and two ordinal sequences on the other. The ordinal aspect is a natural extension of similar models for binary outcomes, such as formulated by Gueorgieva (2001), Molenberghs and Verbeke (2005), Fieuws and Verbeke (2006), Fieuws *et al* (2006), Fieuws *et al* (2008), and Iddi and Molenberghs (2012). A key difference is that ordinal data typically require the introduction of a set of non-redundant dummies, in the sense  $R - 1$  dummies are used for an  $R$ -category ordinal variable (Agresti, 1990).

The continuous-ordinal model was applied to a diabetes study, the bivariate ordinal model to the fluvoxamine study. Qualitatively similar results were obtained: the joint mixed models with correlated random effects yielded a better fit compared to the shared-parameter model. The estimates for the shared parameter in the two case studies differed in magnitude: for the diabetes study, we obtained an estimate with a high value. This is due to the difference in scale between continuous log(BMI)

and ordinal clinical targets. In contrast, in the fluvoxamine study, the same ordinal scale was used for side and therapeutic effects.

The models formulated here can easily be formulated and developed in various software packages. Example code of the implementation in the NLMIXED procedure in SAS are offered in the Appendix. There, we briefly discuss the most important details for re-formatting the data and how to correctly use the syntax. For the diabetes study, 10 quadrature points was sufficient; switching to 20 and 50 quadrature points did not show the estimates and standard errors, to four decimals. For the fluvoxamine study, numerical stability was reached from 20 quadrature points onwards.

Other models for categorical data, whether ordinal or not, such as the partial proportional odds model or the generalized logistic model, can be fitted in a similar fashion, thanks to the flexibility of the procedure NLMIXED. This will then require to reformulate the cell probabilities  $\pi_{r,ij}$  in (2).

The models formulated can be extended to more than two outcomes. There are two possibilities. First, one could go in the direction suggested by Morrell *et al* (2012) by trying to combine more than two responses in one model. Such models are computationally challenging: there is no analytical solution to integrate out the random effects so that approximate methods are required. Fieuws and Verbeke (2006) reported that they were not able to fit joint models for more than four outcomes. They also proposed a solution for this problem as the second possibility for high-dimensional modeling: the computational complexity of the full joint model can be reduced by fitting all possible bivariate mixed models. Estimates for the full joint model are obtained by averaging the estimates from all pairwise models. Evidently, it is also possible to combine an ordinal sequence with other data types, such as repeated count data or repeated time to events.

## Acknowledgments

Financial support from the IAP research network #P7/06 of the Belgian Government (Belgian Science Policy) and the Flemish Supercomputer Project are gratefully acknowledged. We are also grateful to Mr. Kris Bogaerts of I-Biostat for his expert advice.

## References

- Agresti, A. (1990). *Categorical Data Analysis*. New York: John Wiley & Sons.
- Borgermans L., Goderis G., Van Den Broeke C., Verbeke G., Carbonez A., Ivanova A., Mathieu C., Aertgeerts B., Heyrman J., and Grol R. (2009). Interdisciplinary diabetes care teams operating on the interface between primary and specialty care are associated with improved outcomes of care: Findings from the Leuven Diabetes Project, *BMC Health Services Research*, **9**, 179.
- Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9–25.
- Chakraborty, H., Helms, R., Sen, P., and Cohen, M. (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.

- Efendi, A., Molenberghs, G., Njeru Njagi, E., and Dendale, P. (2013). A joint model for longitudinal continuous and time-to-event outcomes with direct marginal interpretation. *Biometrical Journal*, **00**, 000–000.
- Engel, B. and Keen, A. (1994). A simple approach for the analysis of generalized linear mixed models. *Statistica Neerlandica*, **48**, 1–22.
- Fieuws, S. and Verbeke, G. (2006). Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles, *Biometrics*, **62**, 424–431.
- Fieuws, S., Verbeke, G., Boen, F., and Delecluse, C. (2006). High-dimensional multivariate mixed models for binary questionnaire data. *Applied Statistics*, **55**, 1–12.
- Fieuws, S., Verbeke, G., Maes B., and Vanrenterghem Y. (2008). Predicting renal graft failure using multivariate longitudinal profiles. *Biostatistics*, **9**, 419–431.
- Gueorguieva, R. (2001). A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. *Statistical Modeling*, **1**, 177–193.
- Horrocks, J. and van den Heuvel, M.J. (2009). Prediction of pregnancy: A joint model for lonitudinal and binary data. *Bayesian Analysis*, **4**, 523–538.
- Iddi, S. and Molenberghs, G. (2012). A joint marginalized multilevel model for continuous and binary longitudinal outcomes. *Journal of Applied Statistics*, **00**, 000–000.
- Kenward, M.G., Lesaffre, E., and Molenberghs, G. (1994). An application of maximum likelihood and generalized estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random. *Biometrics*, **50**, 945–953.
- Laird, N. and Ware, J. (1982). Random-effects models for longitudinal data. *Biometrics*, **38**, 963–974.
- Li, N., Elashoff, R.M., Li, G., and Saver, J. (2010). Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rt-PA stroke trial. *Statistics in Medicine*, **29**, 546–557.
- Michiels, B. and Molenberghs, G. (1997). Protective estimation of longitudinal categorical data with nonrandom dropout. *Communications in Statistics, Theory and Methods*, **26**, 65–94.
- Molenberghs, G., Kenward, M.G., and Lesaffre, E. (1997). The analysis of longitudinal ordinal data with non-random dropout. *Biometrika*, **84**, 33–44.
- Molenberghs, G., and Lesaffre, E. (1994). Marginal modelling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association*, **89**, 633–644.
- Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.
- Molenberghs, G. and Verbeke, G. (2007). Likelihood ratio, score, and Wald tests in a constrained parameter space. *The American Statistician*, **61**, 1–6.

- Molenberghs, G., Verbeke, G., and Demétrio, C. (2007). An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis*, **13**, 513–531.
- Molenberghs, G., Verbeke, G., Demétrio, C.G.B., and Vieira, A. (2010). A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science*, **25**, 325–347.
- Morrell, C., Brant, L., Sheng, S., and Metter E. (2012). Screening for prostate cancer using multivariate mixed-effects models. *Journal of Applied Statistics*, **39**, 1151–1175 .
- Njeru Njagi, E., Molenberghs, G., Verbeke, G., Kenward, M.G., Dendale, P., and Willekens, K. (2013). A flexible joint-modelling framework for longitudinal and time-to-event data with overdispersion. *Statistical Methods in Medical Research*, **00**, 000–000.
- Pinheiro, J.C., and Bates, D.M. (1995). Approximations to the log-likelihood function in the non-linear mixed-effects model. *Journal of Computation and Graphical Statistics* **4**, 12–35.
- Pinheiro, J.C., and Bates, D.M. (2000) *Mixed Effects Models in S and S-Plus*. New York: Springer.
- Rizopoulos, D. (2012). *Joint Models for Logitudinal and Time-to-Event Data with Applications in R*. Boca Raton: Chapman & Hall/CRC.
- Self, S.G. and Liang, K.Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, **82**, 605–610.
- Stram, D.O. and Lee, J.W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, **50**, 1171–1177.
- Stram, D.A. and Lee, J.W. (1995). Correction to: Variance components testing in the longitudinal mixed effects model. *Biometrics*, **51**, 1196.
- Thiébaud, 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.
- Tsiatis, A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica*, **14**, 809–834.
- Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.
- Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics*, **59**, 254–262.
- Wolfinger, R. and O'Connell, M. (1993). Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, **48**, 233–243.
- Xu, J. and Zeger, S.L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics*, **50**, 375–387.
- Yu, M., Law, N., Taylor, J., and Sandler, H. (2004). Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*, **14**, 835–862.

**Table 4:** *Diabetes Study. Parameter estimates (standard errors) from the regression coefficients in four Models. Model 1: two independent models with random intercepts; Model 2: model with two correlated random intercepts; Model 3: model with one random intercept and scale parameter for log(BMI); Model 4: model with uncorrelated random intercept and random slope for log(BMI), random intercept for clinical targets, with random intercepts taken as correlated. Estimation was done by maximum likelihood using numerical integration over the normal random effects, if present.*

Outcome	Effect	Par.	10 <sup>2</sup> ·Est.(10 <sup>2</sup> ·s.e.)	10 <sup>2</sup> ·Est.(10 <sup>2</sup> ·s.e.)
			Model 1	Model 2
log(BMI)	int.	$\xi_{1,0}$	338.15 (0.6010)	338.13 (0.6010)
	time	$\xi_{1,1}$	-1.4195 (0.1495)	-1.4185 (0.1495)
	second. edu.	$\xi_{1,2}$	-0.6714 (0.8389)	-0.6373 (0.8388)
	higher edu.	$\xi_{1,3}$	-1.2479 (1.1029)	-1.2426 (1.1028)
	gender	$\xi_{1,4}$	-0.7445 (0.7773)	-0.7508 (0.7773)
	res.sd.	$\sigma$	4.6605 (0.0750)	4.6602 (0.0750)
	RI sd.	$\sqrt{d_{11}}$	17.060 (0.2686)	17.063 (0.2686)
	RS sd.	$\sqrt{d_{22}}$	—	—
	Scale	$\lambda$	—	—
CT	int. 0	$\xi_{2,00}$	-2.2758 (0.1009)	-2.2770 (0.1010)
	int. 1	$\xi_{2,01}$	0.3635 (0.0824)	0.3650 (0.0824)
	int. 2	$\xi_{2,02}$	2.7786 (0.1055)	2.7819 (0.1055)
	time	$\xi_{2,1}$	-1.0382 (0.0683)	-1.0438 (0.0683)
	second. edu.	$\xi_{2,2}$	-0.0307 (0.1023)	-0.0324 (0.1023)
	higher edu.	$\xi_{2,3}$	-0.3064 (0.1352)	-0.3068 (0.1352)
	gender	$\xi_{2,4}$	-0.2810 (0.0950)	-0.2777 (0.0950)
	RI sd.	$\sqrt{d_{33}}$	1.5025 (0.0752)	1.5054 (0.0752)
log(BMI)& CT	Cov. RI's	$d_{1,3}$	—	4.2897 (0.8214)
-2 log-likelihood			38,298.67	38,270.40
			Model 3	Model 4
log(BMI)	int.	$\xi_{1,0}$	338.10 (0.6010)	338.15 (0.5991)
	time	$\xi_{1,1}$	-1.4185 (0.1496)	-1.4268 (0.1496)
	second. edu.	$\xi_{1,2}$	-0.5739 (0.8389)	-0.6774 (0.8387)
	higher edu.	$\xi_{1,3}$	-1.1318 (1.1028)	-1.2621 (1.1023)
	gender	$\xi_{1,4}$	-0.7923 (0.7773)	-0.7563 (0.7766)
	res.sd.	$\sigma$	4.6609 (0.0751)	4.0471 (0.3320)
	RI sd.	$\sqrt{d_{11}}$	—	17.079 (0.2690)
	RS sd.	$\sqrt{d_{22}}$	—	3.2638 (0.8072)
	Scale	$\lambda$	89.286 (14.688)	—
CT	int. 0	$\xi_{2,00}$	-1.7041 (0.0679)	-2.2771 (0.1009)
	int. 1	$\xi_{2,01}$	0.2905 (0.0573)	0.3647 (0.0824)
	int. 2	$\xi_{2,02}$	2.0570 (0.0667)	2.7817 (0.1055)
	time	$\xi_{2,1}$	-0.7857 (0.0598)	-1.0436 (0.0683)
	second. edu.	$\xi_{2,2}$	-0.0215 (0.0670)	-0.0314 (0.1023)
	higher edu.	$\xi_{2,3}$	-0.2373 (0.0883)	-0.3073 (0.1352)
	gender	$\xi_{2,4}$	-0.2118 (0.0621)	-0.2786 (0.0950)
	RI sd.	$\sqrt{d_{33}}$	0.1911 (0.0316)	1.5054 (0.0752)
log(BMI) & CT	Cov. RI's	$d_{13}$	—	4.3010 (0.8210)
-2 log-likelihood			38,504.22	38,266.20



**Table 5:** *Fluvoxamine Trial. Parameter estimates (standard errors) from the regression coefficients in two models. Model 1: model with 2 correlated random intercepts; Model 2: model with 1 random intercept and scale parameter. Estimation was done using maximum likelihood with numerical integration over the normal random effect, if present.*

Outcome	Effect	Par.	Est. (s.e.)	Est. (s.e.)
			Model 1	Model 2
Ther. Eff.	int. 0	$\xi_{1,00}$	-2.1912 (1.1783)	-1.2344 (0.4792)
	int. 1	$\xi_{1,01}$	0.9927 (1.1752)	0.6505 (0.4774)
	int. 2	$\xi_{1,02}$	3.3944 (1.1844)	2.0858 (0.4829)
	time (week = 4)	$\xi_{1,11}$	2.0725 (0.1998)	1.1952 (0.1607)
	time (week = 8)	$\xi_{1,12}$	3.6062 (0.2451)	2.1658 (0.1773)
	time (week = 12)	$\xi_{1,13}$	4.4439 (0.2774)	2.7192 (0.1902)
	antecedents	$\xi_{1,2}$	-0.3309 (0.3379)	-0.1601 (0.1302)
	age/30	$\xi_{1,3}$	-0.1755 (0.3650)	-0.0182 (0.1419)
	duration/100	$\xi_{1,4}$	-0.8565 (0.7732)	-0.5015 (0.3199)
	initial severity	$\xi_{1,5}$	-0.2614 (0.2071)	-0.1872 (0.0830)
	Scale	$\lambda$	—	0.0634 (0.0284)
	RI sd.	$\sqrt{d_{11}}$	2.3524 (0.1801)	—
Side Eff.	int. 0	$\xi_{2,00}$	-1.7769 (1.6748)	-1.7481 (1.6627)
	int. 1	$\xi_{2,01}$	3.2179 (1.6825)	3.2264 (1.6697)
	int. 2	$\xi_{2,02}$	5.5565 (1.7079)	5.5660 (1.6953)
	time (week = 4)	$\xi_{2,11}$	0.7256 (0.2278)	0.7224 (0.2272)
	time (week = 8)	$\xi_{2,12}$	1.5917 (0.2584)	1.5841 (0.2577)
	time (week = 12)	$\xi_{2,13}$	1.6562 (0.2671)	1.6444 (0.2665)
	antecedents	$\xi_{2,2}$	-0.0970 (0.4872)	-0.1158 (0.4845)
	age/30	$\xi_{2,3}$	-1.5512 (0.5362)	-1.5451 (0.5325)
	duration/100	$\xi_{2,4}$	-3.9801 (1.0796)	-3.9478 (1.0726)
	initial severity	$\xi_{2,5}$	0.7529 (0.2958)	0.7463 (0.2936)
	RI sd.	$\sqrt{d_{22}}$	3.3756 (0.2986)	3.3545 (0.2979)
	Cov. RI's	$d_{12}$	0.9245 (0.6322)	—
	-2 log-likelihood		3462.90	3740.29

# Mixed Models Approaches for Joint Modeling of Different Types of Responses

Anna Ivanova<sup>1,2</sup>, Geert Molenberghs<sup>3,2</sup>, Geert Verbeke<sup>2,3</sup>

<sup>1</sup>Leuven Statistics Research Centre, KU Leuven, Belgium

<sup>2</sup>I-BioStat, KU Leuven, Belgium

<sup>3</sup>I-BioStat, Universiteit Hasselt, Belgium

Email: anna.ivanova@lstat.kuleuven.be

## Appendix

### A Additional Tables

**Table 6:** *Diabetes Study. Parameter estimates (standard errors) from the regression coefficients. Model 4: model with uncorrelated random intercept and random slope for log(BMI) and random intercept for the clinical targets, with the random intercepts taken as correlated. Model 5: model with correlated random intercept and random slope for log(BMI) and random intercept for the clinical targets, with the random intercepts taken as correlated. Estimation was done by maximum likelihood using numerical integration over the normal random effects, if present.*

Outcome	Effect	Par.	10 <sup>2</sup> ·Est.(10 <sup>2</sup> ·s.e.)	10 <sup>2</sup> ·Est.(10 <sup>2</sup> ·s.e.)
			Model 4	Model 5
log(BMI)	int.	$\xi_{1,0}$	338.15 (0.5991)	338.15 (0.5991)
	time	$\xi_{1,1}$	-1.4268 (0.1496)	-1.4270 (0.1496)
	second. edu.	$\xi_{1,2}$	-0.6774 (0.8387)	-0.6829 (0.8387)
	higher edu.	$\xi_{1,3}$	-1.2621 (1.1023)	-1.2651 (1.1023)
	gender	$\xi_{1,4}$	-0.7563 (0.7766)	-0.7569 (0.7767)
	res.sd.	$\sigma$	4.0471 (0.3320)	4.1495 (9.2660)
	RI sd.	$\sqrt{d_{11}}$	17.079 (0.2690)	17.0538 (2.2710)
	RS sd.	$\sqrt{d_{22}}$	3.2638 (0.8072)	2.2957 (25.6681)
	Cov. RI & RS	$d_{1,2}$	—	0.9055 (76.9391)
CT	int. 0	$\xi_{2,00}$	-2.2771 (0.1009)	-2.2770 (0.1009)
	int. 1	$\xi_{2,01}$	0.3647 (0.0824)	0.3648 (0.0824)
	int. 2	$\xi_{2,02}$	2.7817 (0.1055)	2.7817 (0.1055)
	time	$\xi_{2,1}$	-1.0436 (0.0683)	-1.0435 (0.0683)
	second. edu.	$\xi_{2,2}$	-0.0314 (0.1023)	-0.0315 (0.1023)
	higher edu.	$\xi_{2,3}$	-0.3073 (0.1352)	-0.3075 (0.1352)
	gender	$\xi_{2,4}$	-0.2786 (0.0950)	-0.2787 (0.0950)
	RI sd.	$\sqrt{d_{33}}$	1.5054 (0.0752)	1.5054 (0.0752)
log(BMI)& CT	Cov. RI's	$d_{1,3}$	4.3010 (0.8210)	4.3023 (0.8211)
	-2 log-likelihood		38,266.20	38,266.20

**Table 7:** *Simulation Study: MSE for parameter estimates of the regression coefficients. Model 4: model with uncorrelated random intercept and random slope for continuous variable and random intercept for ordinal variable, with the random intercepts taken as correlated. Model 5: model with correlated random intercept and random slope for continuous variable and random intercept for ordinal variable, with the random intercepts taken as correlated. Estimation was done by maximum likelihood using numerical integration over the normal random effects, if present.*

Outcome	Effect	Par.	N Subjects = 100		N Subjects = 500	
			Model 4	Model 5	Model 4	Model 5
Continuous	int.	$\xi_{1,0} = 34.00$	0.3533	0.3533	0.0819	0.0819
	time	$\xi_{1,1} = -1.50$	0.1118	0.1118	0.0207	0.0207
	res.sd.	$\sigma = 3.00$	0.0169	0.0165	0.0036	0.0027
	RI sd.	$\sqrt{d_{11}} = 5.48$	0.2476	0.2254	0.0713	0.0420
	RS sd.	$\sqrt{d_{22}} = 3.16$	0.0631	0.0548	0.0240	0.0124
	Cov. RI & RS	$d_{1,2} = 6.93$	—	3.7129	—	0.7676
Ordinal	int. 0	$\xi_{2,00} = -2, 50$	0.1017	0.1017	0.0176	0.0176
	int. 1	$\xi_{2,01} = 0.50$	0.0503	0.0503	0.0103	0.0103
	int. 2	$\xi_{2,02} = 3.00$	0.0818	0.0818	0.0156	0.0156
	time	$\xi_{2,1} = -1.05$	0.0075	0.0075	0.0017	0.0017
	RI sd.	$\sqrt{d_{33}} = 1.41$	0.0296	0.0487	0.0063	0.0063
Continuous & Ordinal	Cov. RI's	$d_{1,3} = 3.10$	1.1472	1.0035	0.2612	0.2347

## B SAS Implementation

As an example, we present the routines for the case of a joint model for continuous-ordinal responses with different random effects, some of which are correlated. This procedure can easily be reformulated for different covariate structures per response. Also, the user can apply a similar syntax on the data with different numbers of time points per subject.

In line with the description on how to restructure the data and implement the models, given in Molenberghs and Verbeke (2005, Ch. 24), the NLMIXED program makes use of the so-called general-likelihood feature, i.e., a user-defined likelihood that can be invoked with the 'general()' option in the model statement. The data should a priori be restructured in the following way: in the first and the second positions of the new variable 'outcome,' we enter the values of the continuous variable of the original data, in the third and fourth positions the values of the ordinal variable. Then, the data will take the following form:

```
Obs patient time repeat outcome educ_secund educ_high gender
1 10-1      0    1    326.28 0          0          1
2 10-1      1    2    338.83 0          0          1
3 10-1      0    3     1        0          0          1
4 10-1      1    4     3        0          0          1
5 10-10     0    1    352.78 1          0          .
6 10-10     1    2    356.82 1          0          .
```

```

7  10-10    0    3    1    1    0    .
8  10-10    1    4    1    1    0    .
...

```

We use the following instance of the NL MIXED procedure in SAS for the joint continuous-ordinal model with normal random effects:

```

proc nlmixed data=diabetes Qpoints=10 maxiter=50;
parms beta1_0=338.15 beta1_1=-1.43 beta1_2=-0.68 beta1_3=-1.25 beta1_4=-0.75
sigma=4.05 tau11=17.08 tau22=3.26 beta2_00=-2.28 beta2_01=0.36 beta2_02=2.78
beta2_1=-1.04 beta2_2=-0.03 beta2_3=-0.31 beta2_4=-0.28 tau33=1.50 tau13=1;
title "Model 4: Transformed BMI (RI + RS (uncorr)) + Clinical Targets (RI):
      RI's are correlated";
if repeat in (1,2) then do;
mean = beta1_0 + beta1_1*time + beta1_2*educ_secund + beta1_3*educ_high +
      beta1_4*gender + a + b*time;
dens = -0.5*log(3.14) - log(sigma) - 0.5*(outcome-mean)**2/(sigma**2);
ll=dens;
end;
if repeat in (3,4) then do;
eta = beta2_1*time + beta2_2*educ_secund + beta2_3*educ_high +
      beta2_4*gender + c;
if outcome = 0 then do;
lik = exp(beta2_00+eta)/(1+exp(beta2_00+eta));
end;
if outcome = 1 then do;
lik = exp(beta2_01+eta)/(1+exp(beta2_01+eta)) -
      exp(beta2_00+eta)/(1+exp(beta2_00+eta));
end;
if outcome = 2 then do;
lik = exp(beta2_02+eta)/(1+exp(beta2_02+eta)) -
      exp(beta2_01+eta)/(1+exp(beta2_01+eta));
end;
if outcome = 3 then do;
lik = 1 - exp(beta2_02+eta)/(1+exp(beta2_02+eta));
end;
ll = log(lik);
end;
model outcome ~ general(ll);
random a b c ~ normal([0,0,0], [tau11*tau11,0,tau22*tau22,tau13,0,tau33*tau33])
      subject = patient;
run;

```

For a given data analysis, it is best to conduct a numerical sensitivity analysis to define the optimal number of quadrature points ('qpoints=' option). This can easily be done by progressively increasing their number until the parameter estimates and all related quantities (including standard errors, log-likelihood at maximum, etc.) stabilize.

Another example using the NLMIXED SAS program is for a shared random-intercepts model with a scale parameter  $\lambda$ :

```
proc nlmixed data=diabetes Qpoints=10 maxiter=100;
parms beta1_0=338.15 beta1_1=-1.43 beta1_2=-0.68 beta1_3=-1.25 beta1_4=-0.75
sigma=4.05 beta2_00=-2.28 beta2_01=0.36 beta2_02=2.78 beta2_1=-1.04
beta2_2=-0.03 beta2_3=-0.31 beta2_4=-0.28 tau=1.50 lambda=1;
title "Model 3: Shared RI and Scale Parameter Lambda";
if repeat in (1,2) then do;
mean = beta1_0 + beta1_1*time + beta1_2*educ_secund + beta1_3*educ_high +
      beta1_4*gender + lambda*b;
dens = -0.5*log(3.14) - log(sigma) - 0.5*(outcome-mean)**2/(sigma**2);
ll=dens;
end;
if repeat in (3,4) then do;
eta = beta2_1*time + beta2_2*educ_secund + beta2_3*educ_high +
      beta2_4*gender + b;
if outcome = 0 then do;
lik = exp(beta2_00+eta)/(1+exp(beta2_00+eta));
end;
if outcome = 1 then do;
lik = exp(beta2_01+eta)/(1+exp(beta2_01+eta)) -
      exp(beta2_00+eta)/(1+exp(beta2_00+eta));
end;
if outcome = 2 then do;
lik = exp(beta2_02+eta)/(1+exp(beta2_02+eta)) -
      exp(beta2_01+eta)/(1+exp(beta2_01+eta));
end;
if outcome = 3 then do;
lik = 1 - exp(beta2_02+eta)/(1+exp(beta2_02+eta));
end;
ll = log(lik);
end;
model outcome ~ general(ll);
random b ~ normal(0, tau*tau) subject = patient;
run;
```

For the ordinal-ordinal case, the data can be prepared in a similar way: 4 outcomes for therapeutic effect (THEFF) and 4 outcomes for side effect (SIDE). Then, the data will take the following form:

Obs	patient	week4	week8	week12	repeat	outcome	anteced	severit0	age1	duration1
1	1	0	0	0	1	3	0	4	1.47	0.06
2	1	1	0	0	2	.	0	4	1.47	0.06
3	1	0	1	0	3	.	0	4	1.47	0.06
4	1	0	0	1	4	.	0	4	1.47	0.06
5	1	0	0	0	5	1	0	4	1.47	0.06
6	1	1	0	0	6	.	0	4	1.47	0.06

7	1	0	1	0	7	.	0	4	1.47	0.06
8	1	0	0	1	8	.	0	4	1.47	0.06
9	2	0	0	0	1	2	1	.	0.93	0.01
10	2	1	0	0	2	1	1	.	0.93	0.01
11	2	0	1	0	3	0	1	.	0.93	0.01
12	2	0	0	1	4	0	1	.	0.93	0.01
13	2	0	0	0	5	0	1	.	0.93	0.01
14	2	1	0	0	6	0	1	.	0.93	0.01
15	2	0	1	0	7	0	1	.	0.93	0.01
16	2	0	0	1	8	0	1	.	0.93	0.01
...										

For the joint ordinal-ordinal model, the following SAS syntax can be used:

```
proc nlmixed data=fluvo QPOINTS=20 maxiter=100;
parms
beta1_00=-2.18 beta1_01=1.00 beta1_02=3.39 beta1_11=2.07 beta1_12=3.61
beta1_13=4.45 beta1_2=-0.32 beta1_3=-0.17 beta1_4=-0.82 beta1_5=-0.27
tau1=2.33 beta2_00=-1.70 beta2_01=3.17 beta2_02=5.45 beta2_11=0.73
beta2_12=1.59 beta2_13=1.66 beta2_2=-0.08 beta2_3=-1.52 beta2_4=-3.86
beta2_5=0.72 tau2=3.21 cov12=1;
title "Model 1: THEFF (RI) + SIDE (RI): RI's are correlated";
if repeat in (1,2,3,4) then do;
eta1 = beta1_11*week4 + beta1_12*week8 + beta1_13*week12+ beta1_2*anteced +
      beta1_3*age1 + beta1_4*duration1 + beta1_5*severit0 + a;
if outcome = 0 then do;
lik = exp(beta1_00+eta1)/(1+exp(beta1_00+eta1));
end;
if outcome = 1 then do;
lik = exp(beta1_01+eta1)/(1+exp(beta1_01+eta1)) -
      exp(beta1_00+eta1)/(1+exp(beta1_00+eta1));
end;
if outcome = 2 then do;
lik = exp(beta1_02+eta1)/(1+exp(beta1_02+eta1)) -
      exp(beta1_01+eta1)/(1+exp(beta1_01+eta1));
end;
if outcome = 3 then do;
lik = 1 - exp(beta1_02+eta1)/(1+exp(beta1_02+eta1));
end;
ll = log(lik);
end;
if repeat in (5,6,7,8) then do;
eta2 = beta2_11*week4 + beta2_12*week8 + beta2_13*week12 + beta2_2*anteced +
      beta2_3*age1 + beta2_4*duration1 +beta2_5*severit0 + b;
if outcome = 0 then do;
lik = exp(beta2_00+eta2)/(1+exp(beta2_00+eta2));
end;
```

```

if outcome = 1 then do;
lik = exp(beta2_01+eta2)/(1+exp(beta2_01+eta2)) -
      exp(beta2_00+eta2)/(1+exp(beta2_00+eta2));
end;
if outcome = 2 then do;
lik = exp(beta2_02+eta2)/(1+exp(beta2_02+eta2)) -
      exp(beta2_01+eta2)/(1+exp(beta2_01+eta2));
end;
if outcome = 3 then do;
lik = 1 - exp(beta2_02+eta2)/(1+exp(beta2_02+eta2));
end;
ll = log(lik);
end;
model outcome ~ general(ll);
random a b ~ normal([0,0], [tau1*tau1, cov12, tau2*tau2])
      subject = patient;
run;

```

Next, we present the syntax for a model with shared random intercept for ordinal-ordinal responses:

```

proc nlmixed data=fluvo QPOINTS=20 maxiter=100;
parms beta1_00=-2.18 beta1_01=1.00 beta1_02=3.39 beta1_11=2.07 beta1_12=3.61
beta1_13=4.45 beta1_2=-0.32 beta1_3=-0.17 beta1_4=-0.82 beta1_5=-0.27
beta2_00=-1.70 beta2_01=3.17 beta2_02=5.45 beta2_11=0.73 beta2_12=1.59
beta2_13=1.66 beta2_2=-0.08 beta2_3=-1.52 beta2_4=-3.86 beta2_5=0.72 tau=3.21
lambda=1;
title "Model 2: Shared RI and Scale Parameter Lambda";
if repeat in (1,2,3,4) then do;
eta1 = beta1_11*week4 + beta1_12*week8 + beta1_13*week12+ beta1_2*anteced +
      beta1_3*age1 + beta1_4*duration1 + beta1_5*severit0 + lambda*b;
if outcome = 0 then do;
lik = exp(beta1_00+eta1)/(1+exp(beta1_00+eta1));
end;
if outcome = 1 then do;
lik = exp(beta1_01+eta1)/(1+exp(beta1_01+eta1)) -
      exp(beta1_00+eta1)/(1+exp(beta1_00+eta1));
end;
if outcome = 2 then do;
lik = exp(beta1_02+eta1)/(1+exp(beta1_02+eta1)) -
      exp(beta1_01+eta1)/(1+exp(beta1_01+eta1));
end;
if outcome = 3 then do;
lik = 1 - exp(beta1_02+eta1)/(1+exp(beta1_02+eta1));
end;
ll = log(lik);
end;
if repeat in (5,6,7,8) then do;

```

```

eta2 = beta2_11*week4 + beta2_12*week8 + beta2_13*week12 + beta2_2*anteced +
      beta2_3*age1 + beta2_4*duration1 +beta2_5*severit0 + b;
if outcome = 0 then do;
lik = exp(beta2_00+eta2)/(1+exp(beta2_00+eta2));
end;
if outcome = 1 then do;
lik = exp(beta2_01+eta2)/(1+exp(beta2_01+eta2)) -
      exp(beta2_00+eta2)/(1+exp(beta2_00+eta2));
end;
if outcome = 2 then do;
lik = exp(beta2_02+eta2)/(1+exp(beta2_02+eta2)) -
      exp(beta2_01+eta2)/(1+exp(beta2_01+eta2));
end;
if outcome = 3 then do;
lik = 1 - exp(beta2_02+eta2)/(1+exp(beta2_02+eta2));
end;
ll = log(lik);
end;
model outcome ~ general(ll);
random b ~ normal(0, tau*tau)
subject = patient;
run;

```

Given the implementation of the joint mixed model for the ordinal-ordinal case, the number of required  $Q$  points and maximum iterations should be increased to 'Q=20' and 'maxiter=100', to obtain convergence and an acceptable level of accuracy. Also, the starting values for the parameters should be very carefully selected to ensure that the global (and not local) maximum of the likelihood function is reached. This can be done, for example, by starting from univariate longitudinal models, and assembling the resulting parameter estimates into a starting-value vector for the joint model.