

A high-dimensional joint model for longitudinal outcomes of different nature

Peer-reviewed author version

FAES, Christel; AERTS, Marc; MOLENBERGHS, Geert; GEYS, Helena; TEUNS, Greet & BIJNENS, Luc (2008) A high-dimensional joint model for longitudinal outcomes of different nature. In: STATISTICS IN MEDICINE, 27(22). p. 4408-4427.

DOI: 10.1002/sim.3314

Handle: <http://hdl.handle.net/1942/8541>

# A High-Dimensional Joint Model for Longitudinal Outcomes of Different Nature

**Christel Faes, Marc Aerts, Geert Molenberghs**

Center for Statistics, Hasselt University, Agoralaan, Diepenbeek, Belgium

*Email:* christel.faes@uhasselt.be

**Helena Geys, Greet Teuns, Luc Bijns**

Johnson and Johnson, PRD Biometrics and Clinical Informatics, Beerse, Belgium

March 21, 2008

## **Abstract**

In repeated dose-toxicity studies, many outcomes are repeatedly measured on the same animal, to study the toxicity of a compound of interest. This is only one example in which one is confronted with the analysis of many outcomes, possibly of a different type. Probably the most common situation is that of an amalgamation of continuous and categorical outcomes. A possible approach towards the joint analysis of two longitudinal outcomes of a different nature is the use of random-effects models [1]. Whereas a random-effects model can easily be extended to jointly model many outcomes of a different nature, computational problems arise as the number of outcomes increases. To avoid maximization of the full likelihood expression, Fieuws and Verbeke [2] proposed a pairwise modeling strategy in which all possible pairs are modeled separately, using a mixed model, yielding several different estimates for the same parameters. These latter estimates are then combined into a single set of estimates. Also inference, based on pseudo-likelihood principles, is indirectly derived from the separate analyses. In this paper, we extend the approach of Fieuws and Verbeke [2] in two ways: the method is applied to different types of outcomes and the full pseudo-likelihood expression is maximized at once, leading directly to unique estimates as well as direct application of pseudo-likelihood inference. This is very appealing when interested in hypothesis testing. The method is applied to data from a repeated dose-toxicity study designed for the evaluation of the neurofunctional effects of a psychotropic drug. The relative merits of both methods are discussed.

**Keywords:** Mixed outcomes, High-dimensional joint model, Pseudo-likelihood, Longitudinal data

# 1 Introduction

When a pharmaceutical sponsor aims to bring a new medicinal product to market, it has to be ensured that the product is safe for intended and accidental use alike. To properly assess the toxicity of the substance of interest, animals are evaluated on many different outcomes and exposed during several days to a wide variety of testing conditions, including repro-toxicity and carcinogenicity. Based on such experiments, one can then establish the dose-response relationship and test whether the substance causes adverse effects. Since such repeated dose-toxicity studies involve considerable amounts of time and money, as well as experimental animals, the most appropriate and efficient statistical models, accounting for all possible correlations in the data, should be used for analysis. This raises some challenges. First, the number of outcomes in safety experiments is typically large, implying the necessity of a flexible, high-dimensional joint model. Second, the outcomes in these experiments are often of a mixed continuous and discrete nature. For example, both the malformation status of a live fetus, generally recorded as a binary outcome, and the birth weight, measured on a continuous scale, are important variables in the context of teratogenesis. Perhaps the most common heterogeneous situation is that of a continuous, often assumed normally distributed, and a binary or ordinal outcome. Finally, since measurements are routinely recorded repeatedly over time, the association stemming from the longitudinal nature has to be accounted for. We are thus confronted with multivariate repeated measures, potentially of a different type.

While multivariate methods for the analysis of continuous outcomes are well established [3], methods that jointly analyze discrete and continuous outcomes and adequately account for the correlation structure in the data are less widespread. Broadly there are two approaches [4]. A first approach is based on a conditioning argument that allows the joint distribution to be factored into a marginal component and a conditional one, where the conditioning can be done either on the discrete or on the continuous outcome. Conditional models have been discussed in several contexts [5-10]. A drawback of mixed outcome models based on factorization is that

they may be difficult to apply for quantitative risk assessment because there is no direct access to the marginal distributions [11]. Also, factorization models do not easily extend to the setting of three or more outcomes. Also, the correlation among the two outcomes itself cannot be directly estimated. A second approach directly formulates a joint model for both outcomes. In this context, one often starts from a bivariate continuous variable, one component of which is explicitly observed and the other one observed in dichotomized, or generally discretized, version only. Effectively, this comes down to assuming a continuous latent variable to be underlying the discrete explicit variable. The binary event is then assumed to occur if the latent variable exceeds some threshold value. Catalano and Ryan [12] note that latent-variable models provide a useful and intuitive way to motivate the distribution of the discrete outcome. A common method assumes an unobservable normally-distributed random variable underlying the binary outcome, resulting in a probit-type model. Alternatively, Geys et al. [13] presented a model based on a Plackett-Dale approach, where a bivariate Plackett distribution is assumed. Instead of using a latent variable, one can directly specify the joint distribution for both outcomes via a mixed model, by specification of the marginal distributions, conditional on a correlated random effect [1]. In this paper, the latter approach is considered. The bivariate joint distribution of a continuous and binary outcome is specified via a mixed model, by assuming a general form of a mixed model where the residual error structure and the link function are allowed to change with the nature of the various outcomes. The correlation among the two outcomes is specified through the random-effects structure. An advantage of the mixed model approach is that additional correlation structures in the data, such as a longitudinal data structure, can be modeled within the same framework.

A difficulty arises when a really large number outcomes are of interest, as in safety studies. While (generalized) linear mixed models are very flexible when modeling multivariate data, computational issues arise with increasing numbers of outcomes. Fieuws and Verbeke [2] note that the higher the number of outcomes, the higher the dimension of the random-effects vector when modeling the correlation between the different outcomes via a random effect,

and the more likely computational problems will arise during the estimation process. Diggle et al. [14] indicate that numerical maximum likelihood methods are only feasible for models with a limited number of random effects, say,  $\leq 5$ . This is especially true when the outcomes are of different data types, because switching to heterogeneous settings will increase the complexity of the likelihood. In this paper, a method is proposed to fit a high-dimensional joint model, when full likelihood methods fail due to computational complexity. Fieuws and Verbeke [2] presented a pairwise model-fitting procedure to circumvent the computational complexities in the setting of many continuous outcomes, replacing the maximization of the full likelihood distribution by maximization of each pairwise density separately. This method is based on pseudo-likelihood ideas [15], and is very general and flexible. Pseudo-likelihood methodology is dotted around many fields of application. It has been very advantageously used in a spatial data context, where the full likelihood function is typically cumbersome. But the methodology can be found elsewhere, too [11,16]. The pairwise estimation method achieves important computational economies and, at the same time, it yields unbiased estimates and valid standard errors [2]. Two extensions of the pairwise model-fitting procedure as proposed by Fieuws and Verbeke [2] are considered in this paper. First, a pseudo-likelihood approach is used for jointly analyzing many outcomes of a mixed type. Second, an alternative estimation procedure is used, maximizing the pseudo-likelihood function at once, rather than maximizing all pairwise likelihoods separately.

In this paper, focus is on the risk assessment of a chemical, based on a repeated dose-toxicity study. A description of the data is given in Section 2. In Section 3, it is explained how mixed models can be used to jointly describe one continuous and one binary outcome. The extension towards many continuous and many binary outcomes is given in Section 4. In Section 5, the results are presented.

## 2 Irwin's Toxicity Study

The data considered here come from a three-day repeated dose-toxicity study for the evaluation of the neurofunctional effects of a psychotropic drug. The purpose of this study is to determine and assess the effects of the chemical on general activity and behavior of rats, based on Irwin's method [17]. This method includes a series of non-invasive observational and interactive measurements to assess the behavioral and neurofunctional integrity of the rat.

Male rats were dosed during 3 consecutive days by gavage, and several behavioral observations were recorded. During three consecutive days, 15 rats in the dosed group received 40 mg/kg/day of a chemical substance. There were five rats in the vehicle group (0 mg/kg/day). On days 0, 1 and 2, at 2, 4, 6, 8, and 24 hours after daily oral administration of the chemical substance, all rats were examined for possible neurotoxic effects. Three animals from the exposed group were killed on day 1 because of severe signs of toxicity, and are ignored in the analysis.

Eight variables that were measured during the experiment are described in Table 1, and presented in Figure 1. Each panel shows the average profile in the control group (dashed line) and treatment group (full line) for the specified response. The flags represent the corresponding standard error. The possible effects of a psychotropic compound when dosed repeatedly at a high concentration are e.g. hypothermia, i.e., a subnormal body temperature, and an increased pupil diameter. The side effect of the compound on pupil size seems to be large on the first day, but decreases thereafter. Such autonomic side-effects are indeed most often very rapidly corrected or adjusted by the body itself. From the figure, the adaptive process is clearly visible after repeated dosing. The effects seen in pinna reflex and vocalisation might be related to the sedative effect of the compound. The increase of the toe pinch, locomotor activity and positional passivity in the dosed group might be related to an increased sensitivity of the sensory nerves, i.e., pain sensation.

Interest is in testing whether there is a treatment effect on the outcomes. Also, the association

between the motor-affective and the autonomic outcomes is of interest.

### 3 A Bivariate Model for Longitudinal outcomes

Assume we have two sequences of  $n_i$  outcomes for subject  $i$  ( $i = 1, \dots, N$ ). We denote the sequences for subject  $i$  by  $\mathbf{Y}_{i1} = (Y_{i11}, Y_{i12}, \dots, Y_{i1n_i})'$  and  $\mathbf{Y}_{i2} = (Y_{i21}, Y_{i22}, \dots, Y_{i2n_i})'$ . All outcomes  $Y_{ikj}$  from the same sequence are either continuous ( $k = 1$ , say) or binary, and the  $j$ th outcomes of both sequences are recorded at the same time.

Random-effects models are, arguably, the most frequently used models to analyze longitudinal and multivariate data. A mixed model also allows incorporation of different types of outcomes of different nature in a uniform and natural way [1]. Let us first define the model in the general setting of two outcomes (possibly of different type). For the bivariate response vector  $\mathbf{Y}_i = (\mathbf{Y}'_{i1}, \mathbf{Y}'_{i2})'$  we can assume a generalized linear mixed model of the form

$$\mathbf{Y}_i = \boldsymbol{\mu}_i(\boldsymbol{\eta}_i) + \boldsymbol{\varepsilon}_i = \mathbf{h}(X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i) + \boldsymbol{\varepsilon}_i, \quad (1)$$

where  $\boldsymbol{\mu}_i$  is thus specified in terms of fixed and random effects and  $\boldsymbol{\varepsilon}_i$  is the residual error structure. The model is written in its most general form, as a decomposition in terms of the mean and an appropriate error term, where both mean and error term are allowed to change with the nature of the outcomes. The components of the residual error structure  $\boldsymbol{\varepsilon}_i$  have the appropriate distribution with variance depending on the mean-variance relationship of the various outcomes, and can contain in addition a correlation matrix  $R_i$  and an overdispersion parameter. The components of the inverse link function  $\mathbf{h}(\cdot)$  depend on the nature of the outcomes in  $\mathbf{Y}_i$ . For example, one can choose the identity link for a continuous component, and the logit link for a binary component.  $X_i$  and  $Z_i$  are  $(2n_i \times p)$  and  $(2n_i \times q)$ -dimensional matrices of known covariate values corresponding to subject  $i$ , and  $\boldsymbol{\beta}$  a  $p$ -dimensional vector of unknown fixed regression coefficients. Furthermore,  $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$  are the  $q$ -dimensional random effects.

Because of interest in the correlation structure of the data, a general first-order approximate expression for the variance-covariance matrix of  $\mathbf{Y}_i$  is derived [1]

$$V_i = \text{Var}(\mathbf{Y}_i) \simeq \Delta_i Z_i D Z_i' \Delta_i' + \Sigma_i, \quad (2)$$

with

$$\Delta_i = \left( \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i} \right) \Big|_{\mathbf{b}_i = \mathbf{0}},$$

and

$$\Sigma_i \simeq \Xi_i^{1/2} A_i^{1/2} R_i(\alpha) A_i^{1/2} \Xi_i^{1/2},$$

where  $A_i$  is a diagonal matrix containing the variances following from the generalized linear model specification of  $\mathbf{Y}_{ik}$  ( $k = 1, 2$ ), given the random effects  $\mathbf{b}_i = \mathbf{0}$ , i.e., with diagonal elements  $v(\mu_{ik} | \mathbf{b}_i = \mathbf{0})$ . Likewise,  $\Xi_i$  is a diagonal matrix with the overdispersion parameters along the diagonal. The first term at the right hand side of 2 corresponds to the random-effects structure  $\mathbf{h}(X_i \boldsymbol{\beta} + Z_i \mathbf{b}_i)$ ; the second term at the right hand side of (2) captures the variance-covariances in the residual error  $\boldsymbol{\varepsilon}_i$ .

From (2) it is clear that the correlation between the outcomes can be modeled either using the residual variance of  $\mathbf{Y}_i$  or through specification of the random-effects structure  $Z_i \mathbf{b}_i$ . When there are no random effects in (1), a marginal model is obtained. When there are no residual correlations in  $R_i(\alpha)$  this results in a so-called conditional independence model or purely random-effects model.

More specifically, in what follows, we formulate a possible joint model for continuous and binary outcomes, while accounting for the longitudinal structure of the outcomes, using a conditional independence random-intercepts model with a general variance-covariance matrix  $D$  and residual correlation matrix  $R_i(\alpha) = I$ . Also, the special cases of two continuous and two binary responses are discussed.



### 3.1 One continuous and one binary response

As a first case, denote the sequence of continuous outcomes for subject  $i$  by  $\mathbf{Y}_{i1}$  and its binary counterpart as  $\mathbf{Y}_{i2}$ .  $Y_{i1j}$  and  $Y_{i2j}$  represent, respectively, the  $j$ th continuous and binary outcome for subject  $i$ . The model can be written in the following form:

$$\begin{pmatrix} Y_{i1j} \\ Y_{i2j} \end{pmatrix} = \begin{pmatrix} \alpha_0 + \alpha_1 X_{ij} + b_{i1} \\ \frac{\exp(\beta_0 + \beta_1 X_{ij} + b_{i2})}{1 + \exp(\beta_0 + \beta_1 X_{ij} + b_{i2})} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1j} \\ \varepsilon_{i2j} \end{pmatrix}, \quad (3)$$

where the random effects  $b_{i1}$  and  $b_{i2}$  are normally distributed as

$$\begin{pmatrix} b_{i1} \\ b_{i2} \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_1\tau_2 & \tau_2^2 \end{pmatrix} \right\} \quad (4)$$

and where  $\varepsilon_{i1j}$  and  $\varepsilon_{i2j}$  are independent. The random effects  $b_{i1}$  and  $b_{i2}$  are used to accommodate the longitudinal structure in the data. Because of interest in inference for the correlation, the variance-covariance matrix of the continuous and binary endpoints needs to be calculated.

The variances of  $Y_{i1j}$  and  $Y_{i2j}$  can be derived from (2), in which

$$\begin{aligned} Z_{ij} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \Delta_{ij} = A_i = \begin{pmatrix} 1 & 0 \\ 0 & v_{i2j} \end{pmatrix}, \\ D &= \begin{pmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_1\tau_2 & \tau_2^2 \end{pmatrix}, \Xi_{ij} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 1 \end{pmatrix}, \end{aligned}$$

and where  $v_{i2j} = \pi_{i2j}(b_{i2} = 0)[1 - \pi_{i2j}(b_{i2} = 0)]$ , with  $\pi_{i2j} = \exp(\beta_0 + \beta_1 X_{ij})/[1 + \exp(\beta_0 + \beta_1 X_{ij})]$ . As a result, the approximate variance-covariance matrix of the two measurements for subject  $i$  at time point  $j$  is equal to:

$$V_{ij} = \begin{pmatrix} \tau_1^2 + \sigma^2 & \rho\tau_1\tau_2 v_{i2j} \\ \rho\tau_1\tau_2 v_{i2j} & v_{i2j}^2 \tau_2^2 + v_{i2j} \end{pmatrix}.$$

The correlation  $\rho_{Y_1 Y_2}$  among the continuous and binary outcomes is induced by the incorporation of a correlation  $\rho$  among the two random effects, and is approximately equal to

$$\rho_{Y_1 Y_2} = \frac{\rho\tau_1\tau_2 v_{i2j}}{\sqrt{\tau_1^2 + \sigma^2} \sqrt{v_{i2j}^2 \tau_2^2 + v_{i2j}}}.$$

This could be, at first sight, further approximated by

$$\rho_{Y_1 Y_2} = \frac{\rho \tau_1 \tau_2}{\sqrt{\tau_1^2 + \sigma^2} \sqrt{\tau_2^2 + \pi^2/3}},$$

where  $\pi^2/3$  is the variance of the standard logistic density. However, the basis for this latter approximation is not very strong and caution is needed with this [18]. Consequently, we will not attribute great importance to it. Note that this correlation is always smaller or equal to the correlation  $\rho$  among the two random effects. In the case of conditional independence ( $\rho = 0$ ), the approximate marginal correlation function  $\rho_{Y_1 Y_2}$  also equals zero. In case  $\rho = 1$ , this model reduces to a shared-parameter model, with scale parameter  $\lambda$  equal to  $\tau_1/\tau_2$ . Standard software, such as the SAS procedure NL MIXED, can be used to obtain parameter estimates for this bivariate model.

### 3.2 Two binary responses

Similarly, when both sequences of outcomes are binary, a generalized linear mixed model (GLMM) can be assumed with correlated random effects:

$$\begin{pmatrix} Y_{i2j} \\ Y_{i1j} \end{pmatrix} = \begin{pmatrix} \frac{\exp(\alpha_0 + \alpha_1 X_{ij} + b_{i1})}{1 + \exp(\alpha_0 + \alpha_1 X_{ij} + b_{i1})} \\ \frac{\exp(\beta_0 + \beta_1 X_{ij} + b_{i2})}{1 + \exp(\beta_0 + \beta_1 X_{ij} + b_{i2})} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i2j} \\ \varepsilon_{i1j} \end{pmatrix}, \quad (5)$$

where the random effect  $b_{i1}$  and  $b_{i2}$  are normally distributed as in (4), and where  $\varepsilon_{i1j}$  and  $\varepsilon_{i2j}$  are independent. It is assumed that  $\text{Var}(\varepsilon_{i1j}) = v_{i1j} = \pi_{i1j}(b_{i1} = 0)[1 - \pi_{i1j}(b_{i1} = 0)]$  and  $\text{Var}(\varepsilon_{i2j}) = v_{i2j} = \pi_{i2j}(b_{i2} = 0)[1 - \pi_{i2j}(b_{i2} = 0)]$ . The approximate variance-covariance matrix of the two binary measurements for subject  $i$  at time point  $j$  is equal to:

$$V_{ij} = \begin{pmatrix} v_{i1j}^2 \tau_1^2 + v_{i1j} & \rho \tau_1 \tau_2 v_{i1j} v_{i2j} \\ \rho \tau_1 \tau_2 v_{i1j} v_{i2j} & v_{i2j}^2 \tau_2^2 + v_{i2j} \end{pmatrix},$$

and the correlation between the two outcomes in this case is approximately equal to:

$$\rho_{Y_1 Y_2} = \frac{\rho \tau_1 \tau_2 v_{i1j} v_{i2j}}{\sqrt{v_{i1j}^2 \tau_1^2 + v_{i1j}} \sqrt{v_{i2j}^2 \tau_2^2 + v_{i2j}}}.$$

Note, again, that in most cases no constant correlation would be obtained.

### 3.3 Two continuous responses

When both outcomes are continuous, a linear mixed model can be assumed with correlated random effects. The correlation between the two outcomes in this case is equal to:

$$\rho_{Y_1 Y_2} = \frac{\rho\tau_1\tau_2}{\sqrt{\tau_1^2 + \sigma_1^2}\sqrt{\tau_2^2 + \sigma_2^2}}.$$

So far, focus was on a joint model for two longitudinal sequences. In the next section, the perspective will be broadened to the joint analysis of many, potentially a high number of, outcomes. To avoid computational problems when the number of responses increase, a pseudo-likelihood method is proposed.

## 4 Extension to High-dimensional Data

Assume we have  $m$  sequences  $\mathbf{Y}_{ik} = (Y_{ik1}, Y_{ik2}, \dots, Y_{ikn_i})$ ,  $k = 1, \dots, m$ , of  $n_i$  outcomes for individual  $i$ . The sequences  $\mathbf{Y}_{ik}$  can be either continuous or binary. The  $m$  sequences can then be simultaneously modeled by specifying a joint distribution for the random effects, in analogy with Section 3, but with an  $m \times q$ -dimensional random effects vector  $\mathbf{b}_i$ . The marginal likelihood contribution for subject  $i$  then becomes

$$L_i(\Theta | \mathbf{Y}_{i1}, \mathbf{Y}_{i2}, \dots, \mathbf{Y}_{im}) = \int_{\mathbb{R}^{mq}} \prod_{j=1}^{n_i} f_{ij}(y_{i1j}, y_{i2j}, \dots, y_{imj} | \mathbf{b}_i, \Theta) f(\mathbf{b}_i | D) d\mathbf{b}_i,$$

with  $\Theta = (\beta, \alpha, \mathbf{D})$ . However, computational problems often arise when  $m$  increases, owing to the  $m \times q$ -dimensional integral, especially when outcomes are of a different type. In this case, rather than considering the full likelihood contribution for each subject  $i$ , one can avoid the computational complexity by using a pseudo-likelihood approach, similar to the pairwise modeling approach proposed by [2]. The full likelihood contribution for subject  $i$  is replaced by the pseudo-likelihood function

$$PL_i = \prod_{k=1}^{m-1} \prod_{l=k+1}^m L_{ikl}(\Theta | \mathbf{Y}_{ik}, \mathbf{Y}_{il}) = \prod_{k=1}^{m-1} \prod_{l=k+1}^m \int_{\mathbb{R}^{2q}} \prod_{j=1}^{n_i} f_{ij}(y_{ikj}, y_{ilj} | \mathbf{b}_i^{kl}, \Theta) f(\mathbf{b}_i^{kl} | D) d\mathbf{b}_i^{kl},$$

where each contribution  $L_{ikl}$  is equal to the bivariate likelihood function for outcomes  $k$  and  $l$ . In this way, we have simplified the  $m \times q$ -dimensional integration problem to  $2 \times q$ -dimensional integrations. In practice, this is achieved by restructuring the data in all possible pairs of outcomes, and assuming, as a working assumption, that, conditional on the random effects, all combinations of pair  $(k, l)$  and subject  $i$  are independent. A SAS program is given in the appendix. Inference for  $\Theta$  follows from pseudo-likelihood theory, and is based on a sandwich-type robust variance estimator [15]. The asymptotic multivariate normal distribution for  $\hat{\Theta}$  is given by

$$\sqrt{N}(\hat{\Theta} - \Theta) \sim N(\mathbf{0}, J(\Theta)^{-1}K(\Theta)J(\Theta)^{-1}), \quad (6)$$

where  $J = J(\Theta)$  is a matrix with elements defined by

$$J_{pq} = - \sum_{k=1}^{m-1} \sum_{l=k+1}^m E \left( \frac{\partial^2 \ln L_{ikl}(\Theta | \mathbf{Y}_{ik}, \mathbf{Y}_{il})}{\partial \theta_p \partial \theta_q} \right),$$

and  $K = K(\Theta)$  is a symmetric matrix with elements

$$K_{pq} = - \sum_{k=1}^{m-1} \sum_{l=k+1}^m E \left( \frac{\partial \ln L_{ikl}(\Theta | \mathbf{Y}_{ik}, \mathbf{Y}_{il})}{\partial \theta_p} \frac{\partial \ln L_{ikl}(\Theta | \mathbf{Y}_{ik}, \mathbf{Y}_{il})}{\partial \theta_q} \right).$$

An important advantage of the pseudo-likelihood approach is the close connection with likelihood, which enabled Geys et al. [19] to construct pseudo-likelihood ratio test statistics that have easy-to-compute expressions and intuitively appealing limiting distributions. Since it is well-known that Wald tests can yield erroneous results, especially when a variable has a large effect in the model [19], the pseudo-likelihood ratio test statistic is preferable in this situation. Suppose we are interested in testing the null hypothesis  $H_0 : \gamma = \gamma_0$ , where  $\gamma$  is an  $r$ -dimensional subvector of the  $p$ -dimensional vector of regression parameters  $\beta$  and write  $\beta$  as  $(\gamma^T, \delta^T)^T$ . Then, the pseudo-likelihood ratio test statistic, defined by

$$G_a^{*2} = 2[PL(\hat{\beta}_N) - PL(\gamma_0, \hat{\delta}(\gamma_0))]/\bar{\lambda}, \quad (7)$$

is approximately  $\chi_r^2$  distributed. In (7),  $\hat{\beta}_N$  is the pseudo-likelihood parameter estimate of  $\beta$  and  $\hat{\delta}(\gamma_0)$  denotes the maximum pseudo-likelihood estimator in the subspace where  $\gamma = \gamma_0$ .

Furthermore,  $\bar{\lambda}$  is the mean of the eigenvalues of  $(J^{\gamma\gamma})^{-1}\Sigma_{\gamma\gamma}$ , where  $J^{\gamma\gamma}$  is the  $r \times r$  submatrix of the inverse of  $J$  and  $\Sigma_{\gamma\gamma}$  is the submatrix of  $\Sigma = J^{-1}KJ^{-1}$ .

In what comes next, the proposed methodology is compared with the approach as proposed by Fieuws and Verbeke [2]. They use the pseudo-likelihood function

$$PL_i = \prod_{k=1}^{m-1} \prod_{l=k+1}^m L_{ikl}(\Theta_{k,l} | \mathbf{Y}_{ik}, \mathbf{Y}_{il}), \quad (8)$$

where  $\Theta_{k,l}$  represents the vector of parameters in the bivariate model for the pair of outcomes  $(k, l)$ . While this method was proposed for outcomes of the same type, it is not restricted to this setting, and can handle simultaneously continuous and binary outcomes. With this pseudo-likelihood function, the bivariate likelihood functions can be maximized separately, resulting in different parameter estimates for different pairs of outcomes  $\Theta^* = (\Theta_{1,2}, \Theta_{1,3}, \dots, \Theta_{m-1,m})$ . The estimates of the parameters  $\Theta$  of the joint model are then derived by taking averages over estimates obtained from the different bivariate models, or  $\widehat{\Theta} = A\widehat{\Theta}^*$  (for an appropriate weight matrix  $A$ ). The covariance matrix of  $\Theta$  is given by  $A\Sigma(\Theta^*)A'$ , where  $\Sigma(\Theta^*)$  is the covariance matrix of  $\Theta^*$ . This method further simplifies the computationally very challenging problem, making the approach quite attractive. A possible disadvantage is that no (pseudo)-likelihood is directly available in terms of the parameter of interest,  $\Theta$ , since all pairwise GLMMs are fitted separately. Also, one might lose some efficiency as a consequence of repeatedly estimating the same parameters. Especially when estimating the model under the hypothesis that the effect in different responses is the same, the loss in efficiency could be large when estimating all pairwise models separately.

In conclusion, both the proposed method and the pair-wise method have important advantages. While the first one is slightly more principled, avoids post hoc combination of various estimates of the same parameter, and ensures that resulting covariance matrices are positive definite, it is also computationally considerably more complex, both in terms of computation time requirements, as well as with regard to convergence. For a typical dataset, the proposed method would take about 5 times longer than the pair-wise technique. As such, the two

methods can be seen as complementary, rather than in competition.

## 5 Analysis of Irwin's Toxicity Study

In this section, the data introduced in Section 2 are analyzed. Recall that the purpose of this study is to assess the effects of the investigational chemical on general activity and behavior. Eight outcomes, as described in Table 1, were selected and a joint model for them was estimated.

As before, we denote by  $Y_{ikj}$  the  $j$ th outcome of the  $k$ th response for subject  $i$ . The model for each response  $k$  is specified as

$$h_k^{-1}(\mu_{ij}) = \eta_{ikj} = \beta_{0k} + \beta_{1k}g_i + \beta_{2k}t_{ij} + \beta_{3k}d_{ij} + \beta_{4k}t_{ij}d_{ij} + \beta_{5k}g_it_{ij} + \beta_{6k}g_id_{ij} + b_{ik},$$

where  $h_k^{-1}$  is the identity link in case of a continuous outcome ( $k = 5, \dots, 8$ ) and the logit link in case of a binary outcome ( $k = 1, \dots, 4$ ),  $g_i$  is an indicator variable taking value 1 for the rats in the treatment group and 0 otherwise,  $t_{ij}$  is the time after exposure within a day, and  $d_{ij}$  is the day of exposure. The model is sufficiently general to adapt to the diverse responses and to answer the research questions of interest. The time effect describes the effect of the drug some time after exposure while the day effect describes the effect of repeated exposures. The interaction term of time and day allows for a systematic increase or decrease in the slope of the time-effect at consecutive days. Correlations between the responses at different time points are modeled by way of a general variance-covariance matrix for the random intercepts  $b_{ik}$ :

$$\mathbf{b}_i = \begin{pmatrix} b_{i1} \\ b_{i2} \\ \vdots \\ b_{i8} \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_1^2 & \rho_{12}\tau_1\tau_2 & \cdots & \rho_{18}\tau_1\tau_8 \\ \rho_{12}\tau_1\tau_2 & \tau_2^2 & \cdots & \rho_{28}\tau_2\tau_8 \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{18}\tau_1\tau_8 & \rho_{28}\tau_2\tau_8 & \cdots & \tau_8^2 \end{pmatrix} \right\}.$$

Owing to the computational complexity, the full likelihood is replaced by a pseudo-likelihood function, as described in Section 4.

The parameter estimates and standard errors, based on the proposed pseudo-likelihood function (6) and based on the pairwise estimation method (8), are displayed in Table 2. Based on the pseudo-likelihood estimates, it can be seen that almost all responses show a significant effect of treatment (Treat). Only for the autonomic variables 'pupil size' and 'temperature,' does the treatment effect change significantly from day to day. For the locomotor activity and pinna reflex, the time trend is different among the treatment and vehicle group. Comparing top and bottom panels of Table 2, it can be seen that the parameter estimates and the standard errors of the fixed effects are very similar for the two fitting procedures. Observe that there are slightly larger differences between the two fitting procedures for the parameter estimates corresponding with the binary response, as compared with the continuous responses. It is also for these parameter that the parameter estimates among the different pairwise models show more variability. Figure 2 shows the estimated (marginal) profiles, based on the proposed pseudo-likelihood estimates. For most responses, the model fits the data quite well.

Table 3 shows the estimated variance-covariance matrix of the random effects, based on, respectively, the pseudo-likelihood and pairwise estimation methods. These correlations can be interpreted as the association between the individual deviations from the overall profile. Values on the diagonal are the (approximate) intra-class correlations corresponding to each outcome. Several of the random effects are highly correlated, such as pupil size and temperature, as well as grip strength and vocalization. The estimates based on the two estimation procedures are similar, although larger differences as compared with the fixed-effects parameters are seen. Also, the pseudo-likelihood methods seems to result more often in a significant correlation as compared with the pairwise fitting procedure. Based on the estimated variance-covariance matrix of the random effects, one can estimate the correlation between the different outcomes. Results are displayed in Table 4. As indicated before, we see that these correlations are much smaller when compared with the correlation between the random effects.

Finally, it is of interest to test some specific effects, such as: (*i*) whether there is a treatment

effect for each response separately; (ii) whether there is an overall effect (overall responses); (iii) whether there is a treatment effect for groups of variables (such as the motor-affective responses); (iv) whether there is a difference among groups of variables (such as a difference among the sensoro-motor responses pinna reflex and toe pinch).

Results are displayed in Table 5. We will illustrate the pseudo-likelihood ratio test for the first null hypothesis. The log-pseudo-likelihood for the full model is equal to  $-11759$ . To test for the effect of treatment on all responses, this log-pseudo-likelihood needs to be compared with a reduced model without treatment-effect and without interactions of treatment with time and day. The log-pseudo-likelihood for this reduced model equals  $-12554$ . The corresponding likelihood-ratio test is equal to  $G_a^{*2} = 1590/7.96 = 200$ , compared with a  $\chi_{24}^2$  this results in a  $p\text{-value} < 0.0001$ , rejecting the reduced model of no treatment effect. The second panel shows in Table 5 there is a significant treatment effect on locomotor activity, pinna reflex, toe pinch, grip strength, pupil size, temperature and vocalization. The third panel indicates that all effects are important. For the motor-affective responses, the treatment seems to have no effect on the time and day effect. Finally, there is no significant difference among the variable pinna reflex and toe pinch.

Extending the model with a random slope is difficult in this setting from a computational perspective, in the sense of both computation time and severe convergence problems. Also with the estimation method proposed by [2], reaching convergence for an extended model, including random intercept and random slope, is a less than trivial task in this data example.

## 6 Discussion

In this paper, a method to jointly analyze many outcomes that are measured at several time points on the same individual and that are possibly of a different type is proposed. The method is an extension of the proposal made by Fieuws and Verbeke [2]. The basis of the method is to use a generalized linear mixed model, allowing the link function and residual error structure



to differ for the various outcomes and assuming correlated random effects to account for the correlation between the outcomes. When many outcomes are of interest, this results in a high-dimensional integration problem, which is computationally no longer feasible. Therefore, a pseudo-likelihood function is used instead of the full likelihood. The principal idea of the pseudo-likelihood methodology is to replace the computationally cumbersome likelihood by a function set up as the product of easy-to-compute functions, but which is no longer equal to the full likelihood. In our setting, the pseudo-likelihood function is defined as the product of all pairwise bivariate models. This pseudo-likelihood yields unbiased estimates and valid standard errors. An important advantage of the proposed method is that it directly yields unique parameter estimates of the joint model and that it allows to directly use the pseudo-likelihood ratio statistic, which is very appealing for inference.

We put particular emphasis on the joint analysis of several continuous and several binary outcomes, but the method is in no way restricted to this setting, and can be used for arbitrary combinations of outcome types.

In the Irwin study, three animals from the exposed group were killed on day 1 because of severe signs of toxicity, and were further ignored in the analysis. While it is tempting to assume that the missing observations are MAR owing to the reason of observed toxicity, one has to be careful since it is impossible to definitively make assumptions about the proper missing data mechanism. Therefore, it is prudent to consider ways of conducting sensitivity analysis [21]. To this end, we supplement our analysis with one where the incomplete animals are added, too. The results are shown in Table 6; they indicate that there are no substantial changes in the conclusions, and therefore that we can be rather confident about the results. In general, because the pseudo-likelihood's behavior under the assumption of MAR is presently not fully understood from a methodological standpoint, one should be careful with the analysis of incomplete data. Indeed, even if the missing data are MAR, it remains to be seen whether the mechanism is ignorable. While ignorability would follow under likelihood inference [22], this is

not generally true for non-likelihood methods such as generalized estimating equations (GEE, [23]) and pseudo-likelihood. However, while the GEE case received considerable attention [24,25], the case of pseudo-likelihood is open and is currently being investigated. For all of these reasons, proper treatment of the missing-data issues is important but evidently outside the scope of this paper.

## Acknowledgments

We gratefully acknowledge support from the Institute for the Promotion of Innovation by Science and Technology (IWT) in Flanders, Belgium and from the IAP research network nr P5/24 of the Belgian Government (Belgian Science Policy).

## References

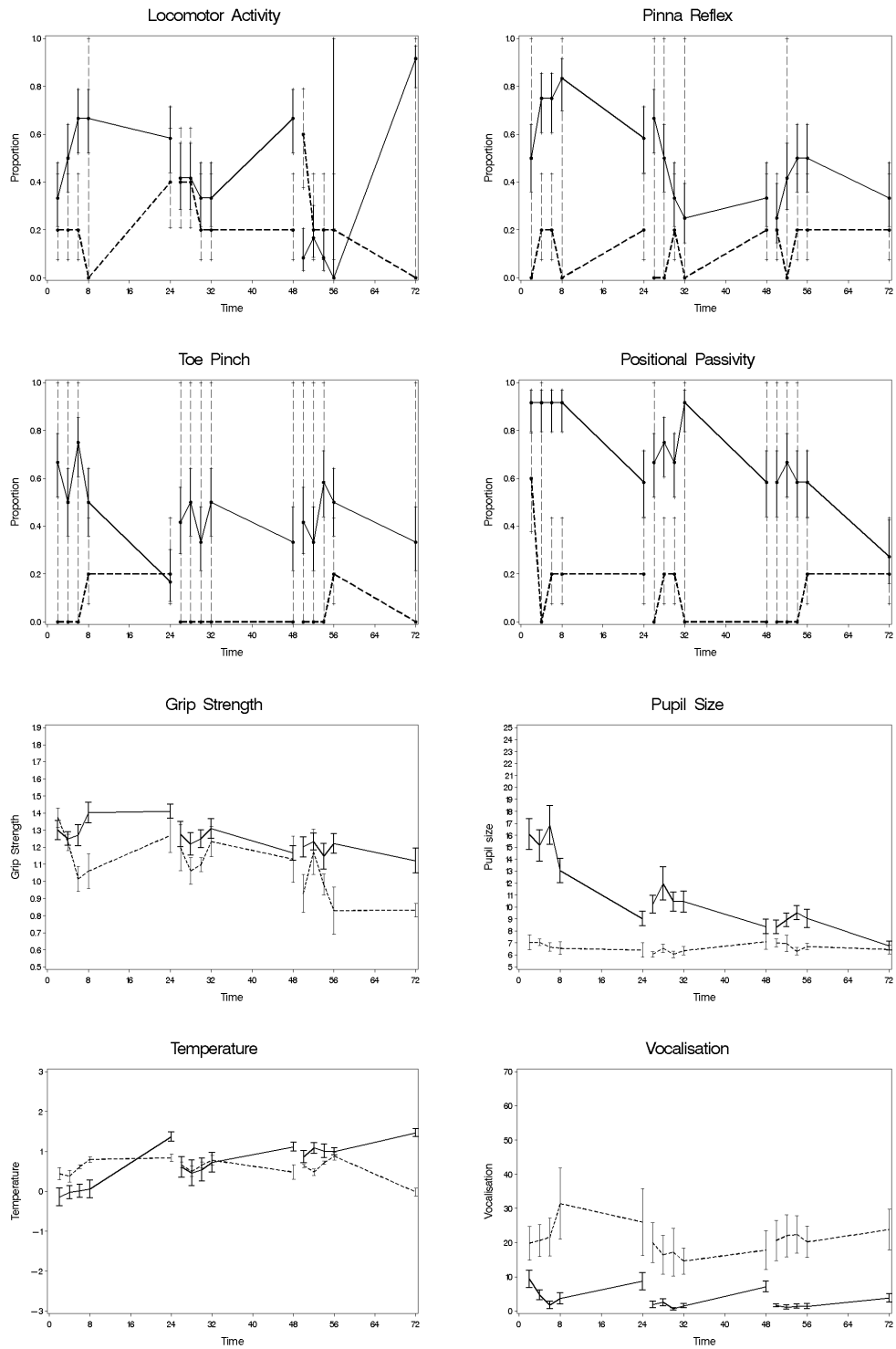
1. Molenberghs G, Verbeke G. *Models for Discrete Longitudinal Data*. Springer Series in Statistics. New York: Springer, 2005.
2. Fieuws S, Verbeke G. Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics*, 2006; **62**: 424–431.
3. Johnson RA, Wichern DW. *Applied Multivariate Statistical Analysis*, 5th Edn. New Jersey: Prentice Hall, 2002.
4. Aerts M, Geys H, Molenberghs G, Ryan L. *Topics in Modelling of Clustered Data*. London: Chapman and Hall, 2002.
5. Tate RF. Correlation between a discrete and a continuous variable. *Annals of Mathematical Statistics*, 1954; **25**: 603–607.
6. Olkin I, Tate RF. Multivariate correlation models with mixed discrete and continuous variables. *Annals of Mathematical Statistics*, 1961; **32**: 448–465 (with correction in **36**, 343–344).

7. Little RJA, Schluchter MD. Maximum likelihood estimation for mixed continuous and categorical data with missing values. *Biometrika*, 1985; **72**: 497–512.
8. Krzanowski WJ. *Principles of Multivariate Analysis*. Oxford: Clarendon Press, 1988.
9. Cox DR, Wermuth N. Response models for mixed binary and quantitative variables. *Biometrika*, 1992; **79**: 441–461.
10. Cox DR, Wermuth N. A Note on the Quadratic Exponential Binary Distribution. *Biometrika*, 1994; **81**: 403–408.
11. Geys H, Regan M, Catalano P, Molenberghs G. Two latent variable risk assessment approaches for mixed continuous and discrete outcomes from developmental toxicity data. *Journal of Agricultural Biological and Environment Statistics*, 2001; **6**: 340–355.
12. Catalano PJ, Ryan LM. Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association*, 1992; **87**: 651–658.
13. Molenberghs G, Geys H, Buyse M. Evaluations of surrogate endpoints in randomized experiments with mixed discrete and continuous outcomes. *Statistics in Medicine*, 2001; **20**: 3023–3038.
14. Diggle PJ, Heagerty P, Liang KY, Zeger SL. *Analysis of longitudinal data*, Clarendon Press, Oxford, 2002.
15. Arnold BC, Strauss D. Pseudolikelihood estimation: Some examples. *Sankhya B*, 1991; **53**: 233–243.
16. Faes C, Geys H, Aerts M, Molenberghs G, Catalano P. Modelling combined continuous and ordinal outcomes in a clustered setting. *Journal of Agricultural, Biological and Environmental Statistics*, 2004; **9**: 515–530.

17. Irwin S. Comprehensive Observational Assessment. Ia. A Systematic, Quantitative Procedure for Assessing the Behavioral and Physiologic State of the Mouse. *Physiopharmacologia*, 1964; **13**: 222–257.
18. Laenen A, Vangeneugden T, Geys H, Molenberghs G. Generalized reliability estimation using repeated measurements. *British Journal of Mathematical and Statistical Psychology*, 2006; **59**: 113–131
19. Geys H, Molenberghs G, Ryan L. Pseudo-likelihood inference for clustered binary data. *Communications in Statistics: Theory and Methods*, 1997; **26**: 2743–2767.
20. Hauck WW, Donner A. Walds test as applied to hypotheses in logit analysis, *Journal of the American Statistical Association*, 1977; **72**: 851–853.
21. Molenberghs G, Kenward MG. *Missing Data in Clinical Studies*, 2007. Chichester: Wiley.
22. Rubin DB. Inference and missing data (with discussion). *Biometrika*, 1976, **63**, 581-592.
23. Liang KY, Zeger S. Longitudinal data analysis using generalized linear models. *Biometrika*, 1986; **73**, 13-22.
24. Robins JM, Rotnitzky A, Scharfstein DO. Semiparametric regression for repeated outcomes with nonignorable nonresponse *Journal of the American Statistical Association*, 1998; **93**: 1321–1339.
25. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data *Journal of the American Statistical Association*, 1995; **90**: 106–121.

**Table 1:** Irwin's Toxicity Study. Description and classification of the 8 responses of interest, of which 4 are binary and 4 are continuous.

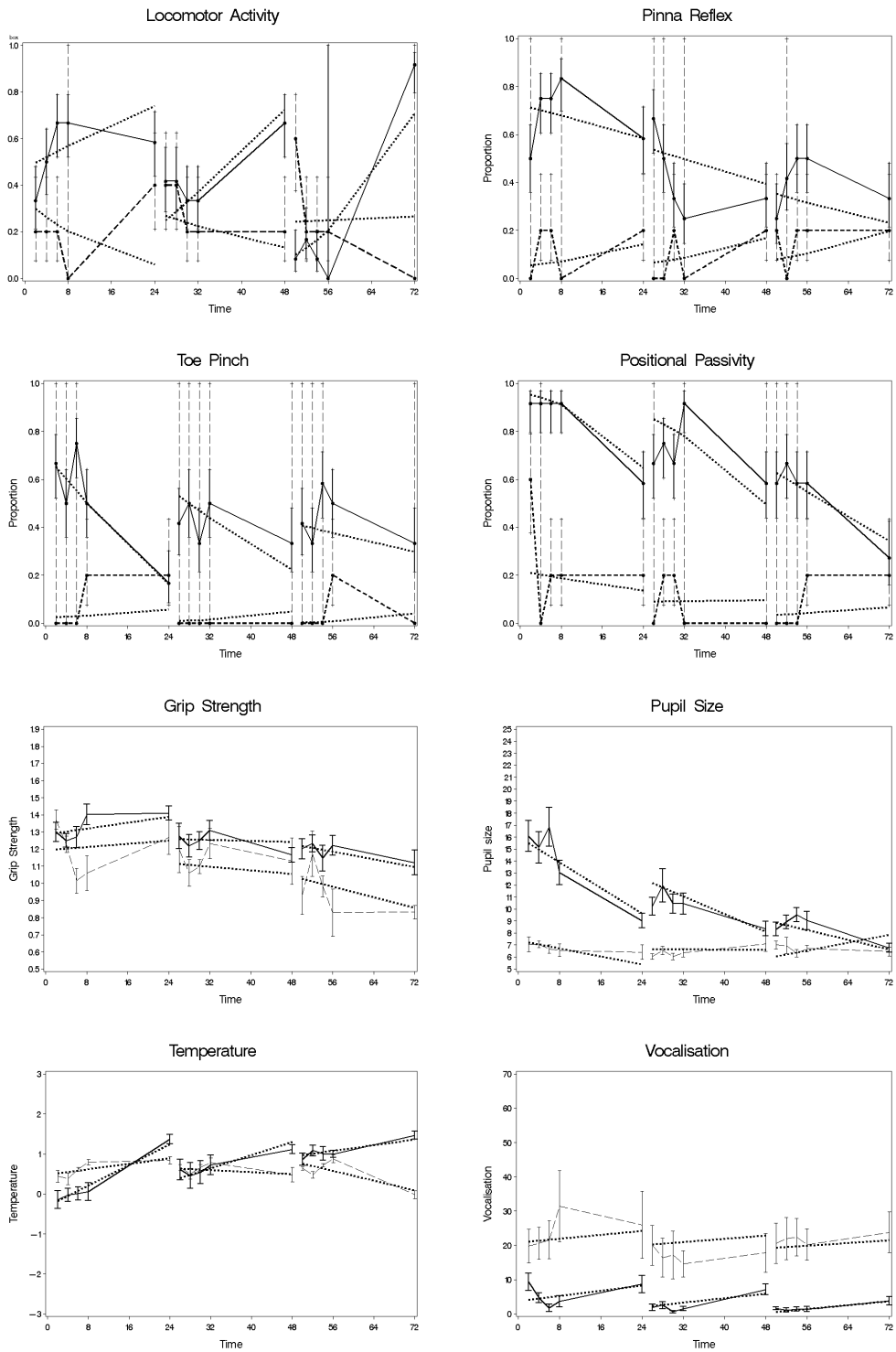
	Variable	Type	Description	Classification
1	Locomotor Activity	Binary	Characterized by abnormal biting, restlessness, writhing	Behavioral; spontaneous activity
2	Pinna Reflex	Binary	Animal's twitch reflex upon being touched on the auricle	Behavioral; sensoro-motor response
3	Toe Pinch	Binary	Animal's response to pain upon having the toe squeezed	Behavioral; sensoro-motor response
4	Positional Passivity	Binary	Animal's struggle response to sequential handling	Behavioral; motor-affective response
5	Grip Strength	Continuous	Animal's fore limb grip strength	Neurological; muscle tone
6	Pupil Size	Continuous	Animal's pupil diameter	Autonomic
7	Temperature	Continuous	Animal's body temperature	Autonomic
8	Vocalization	Continuous	Animal's incidence of squeaking during manipulation	Behavioral; motor-affective response



**Figure 1:** Irwin's Toxicity Study. Four binary and four continuous outcomes of interest, are presented. The full and dashed lines correspond to mean profiles of, respectively, the treatment and vehicle group. The flags represent the corresponding standard error.

**Table 2:** Irwin’s Toxicity Study. Parameter estimates (standard errors). The top panel corresponds to pseudo-likelihood estimation (PL Method). The bottom panel corresponds with the pairwise estimation technique (Pairwise Method [2]). Significant values at the 0.01 significance level are indicated with a ‘\*’, significant values at the 0.05 significance level are indicated with a ‘+’.

Parameter	Response							
	1	2	3	4	5	6	7	8
<b>PL Method</b>	<b>Locom Act</b>	<b>Pinna Reflex</b>	<b>Toe Pinch</b>	<b>Vert Hind</b>	<b>Grip Strength</b>	<b>Pupil Size</b>	<b>Temperature</b>	<b>Vocalization</b>
$\beta_{0k}$ Intercept	-0.686(0.494)	-2.966(0.935)*	-3.758(1.099)*	-1.273(0.544)+	1.193(0.043)*	7.380(0.313)*	0.480(0.098)*	20.781(4.715)*
$\beta_{1k}$ Treat	0.571(0.677)	3.923(0.882)*	4.588(1.202)*	4.480(1.289)*	0.091(0.059)	8.669(1.120)*	-0.786(0.233)*	-17.151(5.069)*
$\beta_{2k}$ Time	-0.086(0.067)	0.048(0.042)	0.039(0.087)	-0.024(0.044)	0.002(0.003)	-0.084(0.021)*	0.017(0.007)+	0.140(0.145)
$\beta_{3k}$ Day	-0.231(0.505)	0.211(0.741)	-1.138(0.941)	-1.058(0.365)*	-0.076(0.032)+	-0.758(0.193)*	0.169(0.053)*	-0.850(1.885)
$\beta_{4k}$ Time*Day	0.046(0.033)	0.000(0.034)	0.040(0.026)	0.028(0.039)	-0.005(0.001)*	0.083(0.018)*	-0.024(0.004)*	-0.022(0.054)
$\beta_{5k}$ Treat*Time	0.135(0.056)+	-0.074(0.021)*	-0.141(0.087)	-0.084(0.052)	0.002(0.002)	-0.184(0.029)*	0.048(0.008)*	0.048(0.110)
$\beta_{6k}$ Treat*Day	-0.945(0.532)	-0.968(0.617)	0.558(0.946)	-0.236(0.519)	0.049(0.036)	-2.729(0.368)*	0.454(0.083)*	-0.864(1.941)
$\tau_k^2$ Variance RI	0.268(0.139)	1.314(0.783)	2.024(1.062)	1.019(0.702)	0.011(0.003)*	3.148(1.155)+	0.122(0.072)	35.655(13.478)*
$\sigma_k^2$ Res. Variance					0.032(0.015)*	4.783(1.424)*	0.207(0.057)*	32.119(10.117)*
<b>Pairwise Method</b>	<b>Locom Act</b>	<b>Pinna Reflex</b>	<b>Toe Pinch</b>	<b>Vert Hind</b>	<b>Grip Strength</b>	<b>Pupil Size</b>	<b>Temperature</b>	<b>Vocalization</b>
$\beta_{0k}$ Intercept	-0.683(0.493)	-2.963(0.934)*	-3.763(1.100)*	-1.270(0.545)+	1.193(0.043)*	7.386(0.313)*	0.480(0.098)*	20.781(4.717)*
$\beta_{1k}$ Treat	0.566(0.679)	3.918(0.880)*	4.594(1.201)*	4.488(1.293)*	0.091(0.060)	8.662(1.123)*	-0.787(0.234)*	-17.153(5.068)*
$\beta_{2k}$ Time	-0.087(0.067)	0.048(0.042)	0.039(0.087)	-0.025(0.044)	0.002(0.003)	-0.084(0.021)*	0.017(0.007)+	0.139(0.145)
$\beta_{3k}$ Day	-0.234(0.506)	0.205(0.740)	-1.141(0.940)	-1.059(0.366)+	-0.075(0.032)+	-0.763(0.192)*	0.169(0.053)*	-0.859(1.884)
$\beta_{4k}$ Time*Day	0.046(0.033)	0.000(0.034)	0.040(0.026)	0.028(0.039)	-0.005(0.001)*	0.083(0.018)*	-0.024(0.004)*	-0.021(0.054)
$\beta_{5k}$ Treat*Time	0.135(0.057)+	-0.074(0.021)*	-0.142(0.087)	-0.084(0.052)	0.002(0.002)	-0.183(0.029)*	0.048(0.008)*	0.049(0.110)
$\beta_{6k}$ Treat*Day	-0.945(0.534)	-0.967(0.617)	0.562(0.946)	-0.241(0.521)	0.049(0.036)	-2.726(0.369)*	0.454(0.083)*	-0.860(1.939)
$\tau_k^2$ Variance RI	0.273(0.142)	1.350(0.807)	2.068(1.096)	1.033(0.712)	0.011(0.003)*	3.159(1.162)+	0.122(0.072)	35.480(13.378)*
$\sigma_k^2$ Res. Variance					0.032(0.003)*	4.788(0.955)*	0.207(0.057)*	32.063(10.085)*



**Figure 2:** Irwin's Toxicity Study. Estimated (marginal) profiles for the four binary and four continuous outcomes of interest. The full and dashed lines correspond to observed mean profiles of, respectively, the treatment and vehicle group. The flags represent the corresponding standard error. The dotted lines correspond to the estimated mean profiles.



**Table 3:** Irwin’s Toxicity Study. Estimated correlation matrix of the random effects, based on the pseudo-likelihood approach (top panel) and on the pairwise estimation method (bottom panel). Values on the diagonal are the (approximate) intra-class correlations corresponding to each outcome. Correlations significantly different from zero, according to a Wald test, at the 0.01 significance level are indicated with a ‘\*’, significant correlations at the 0.05 significance level are indicated with a ‘+’.

<b>PL Method</b>	1	2	3	4	5	6	7	8
1 Locom Act	0.08*							
2 Pinna Reflex	0.36	0.29*						
3 Toe Pinch	-0.34	-0.58*	0.38*					
4 Vertical Hind	0.11	0.15	-0.56 <sup>+</sup>	0.24 <sup>+</sup>				
5 Grip Strength	0.39	-0.23	0.76*	-0.09	0.25*			
6 Pupil Size	0.50	0.25	-0.37	0.40	-0.21	0.40*		
7 Temperature	-0.42	-0.43	0.36	-0.55 <sup>+</sup>	0.12	-0.82*	0.37*	
8 Vocalization	-0.69*	-0.32	-0.34	-0.08	-0.60*	-0.03	0.24*	0.53*

<b>Pairwise Method</b>	1	2	3	4	5	6	7	8
1 Locom Act	0.08*							
2 Pinna Reflex	0.41	0.23*						
3 Toe Pinch	-0.37	-0.56 <sup>+</sup>	0.39*					
4 Vertical Hind	0.12	0.13	-0.53	0.24 <sup>+</sup>				
5 Grip Strength	0.42	-0.22	0.75*	-0.09	0.25*			
6 Pupil Size	0.52	0.24	-0.39	0.39	-0.21	0.40*		
7 Temperature	-0.43	-0.41	0.37	-0.54 <sup>+</sup>	0.11	-0.82*	0.37*	
8 Vocalization	-0.68 <sup>+</sup>	-0.31	-0.34	-0.08	-0.61*	-0.03	0.24*	0.53*

**Table 4:** Irwin's Toxicity Study. Estimated correlation matrix of the outcomes, based on the pseudo-likelihood approach.

	1	2	3	4	5	6	7	8
1 Locom Act	1							
2 Pinna Reflex	0.02	1						
3 Toe Pinch	-0.02	-0.03	1					
4 Vertical Hind	0.01	0.01	-0.05	1				
5 Grip Strength	0.05	-0.03	0.08	-0.02	1			
6 Pupil Size	0.07	0.04	-0.05	0.10	-0.07	1		
7 Temperature	-0.06	-0.07	0.05	-0.13	0.04	-0.31	1	
8 Vocalization	-0.12	-0.06	-0.05	-0.02	-0.22	-0.01	0.11	1

**Table 5:** Irwin's Toxicity Study. Results of hypothesis testing.

Test	Response	Null Hypothesis	$G_a^{*2}$	df	p-value
Treatment effect	Responses 1-8	$\beta_{1i} = \beta_{5i} = \beta_{6i} = 0, \forall i$	199.75	24	< 0.0001
Treatment effect	Response 1	$\beta_{11} = \beta_{51} = \beta_{61} = 0$	13.60	3	0.003
Treatment effect	Response 2	$\beta_{12} = \beta_{52} = \beta_{62} = 0$	18.69	3	<0.001
Treatment effect	Response 3	$\beta_{13} = \beta_{53} = \beta_{63} = 0$	16.51	3	<0.001
Treatment effect	Response 4	$\beta_{14} = \beta_{54} = \beta_{64} = 0$	7.23	3	0.065
Treatment effect	Response 5	$\beta_{15} = \beta_{55} = \beta_{65} = 0$	72.02	3	<0.001
Treatment effect	Response 6	$\beta_{16} = \beta_{56} = \beta_{66} = 0$	115.51	3	<0.001
Treatment effect	Response 7	$\beta_{17} = \beta_{57} = \beta_{67} = 0$	72.02	3	<0.001
Treatment effect	Response 8	$\beta_{18} = \beta_{58} = \beta_{68} = 0$	9.77	3	0.021
Treat effect	All	$\beta_{11} = \beta_{12} = \dots = \beta_{18} = 0$	78.61	8	<0.001
Time effect	All	$\beta_{21} = \beta_{22} = \dots = \beta_{28} = 0$	15.27	8	0.054
Day effect	All	$\beta_{31} = \beta_{32} = \dots = \beta_{38} = 0$	16.71	8	0.033
Time*day effect	All	$\beta_{41} = \beta_{42} = \dots = \beta_{48} = 0$	44.42	8	<0.001
Treat*time effect	All	$\beta_{51} = \beta_{52} = \dots = \beta_{58} = 0$	80.56	8	<0.001
Treat*day effect	All	$\beta_{61} = \beta_{62} = \dots = \beta_{68} = 0$	59.78	8	<0.001
Treat effect	Motor-affective	$\beta_{14} = \beta_{18} = 0$	18.37	2	<0.001
Time effect	Motor-affective	$\beta_{24} = \beta_{28} = 0$	1.58	2	0.455
Day effect	Motor-affective	$\beta_{34} = \beta_{38} = 0$	2.14	2	0.342
Time*day effect	Motor-affective	$\beta_{44} = \beta_{48} = 0$	0.78	2	0.677
Treat*time effect	Motor-affective	$\beta_{54} = \beta_{58} = 0$	0.53	2	0.767
Treat*day effect	Motor-affective	$\beta_{64} = \beta_{68} = 0$	0.44	2	0.804
Difference	Sensoro-motor	$\beta_{i2} = \beta_{i3}, \forall i$	15.58	7	0.029

**Table 6:** Irwin's Toxicity Study. Pseudo-likelihood parameter estimates (standard errors) from an analysis of all rats, including those which drop out during the experiment. Significant values at the 0.01 significance level are indicated with a '\*', significant values at the 0.05 significance level are indicated with a '+'. .

Parameter	Response							
	1	2	3	4	5	6	7	8
<b>PL Method</b>	<b>Locom Act</b>	<b>Pinna Reflex</b>	<b>Toe Pinch</b>	<b>Vert Hind</b>	<b>Grip Strength</b>	<b>Pupil Size</b>	<b>Temperature</b>	<b>Vocalization</b>
$\beta_{0k}$ Intercept	-0.686(0.494)	-2.966(0.935)*	-3.758(1.099)*	-1.273(0.544)+	1.193(0.043)*	7.380(0.313)*	0.480(0.098)*	20.781(4.715)*
$\beta_{1k}$ Treat	0.571(0.677)	3.923(0.882)*	4.588(1.202)*	4.480(1.289)*	0.091(0.059)	8.669(1.120)*	-0.786(0.233)*	-17.151(5.069)*
$\beta_{2k}$ Time	-0.086(0.067)	0.048(0.042)	0.039(0.087)	-0.024(0.044)	0.002(0.003)	-0.084(0.021)*	0.017(0.007)+	0.140(0.145)
$\beta_{3k}$ Day	-0.231(0.505)	0.211(0.741)	-1.138(0.941)	-1.058(0.365)*	-0.076(0.032)+	-0.758(0.193)*	0.169(0.053)*	-0.850(1.885)
$\beta_{4k}$ Time*Day	0.046(0.033)	0.000(0.034)	0.040(0.026)	0.028(0.039)	-0.005(0.001)*	0.083(0.018)*	-0.024(0.004)*	-0.022(0.054)
$\beta_{5k}$ Treat*Time	0.135(0.056)+	-0.074(0.021)*	-0.141(0.087)	-0.084(0.052)	0.002(0.002)	-0.184(0.029)*	0.048(0.008)*	0.048(0.110)
$\beta_{6k}$ Treat*Day	-0.945(0.532)	-0.968(0.617)	0.558(0.946)	-0.236(0.519)	0.049(0.036)	-2.729(0.368)*	0.454(0.083)*	-0.864(1.941)
$\tau_k^2$ Variance RI	0.268(0.139)	1.314(0.783)	2.024(1.062)	1.019(0.702)	0.011(0.003)*	3.148(1.155)+	0.122(0.072)	35.655(13.478)*
$\sigma_k^2$ Res. Variance					0.032(0.015)*	4.783(1.424)*	0.207(0.057)*	32.119(10.117)*
<b>PL Method</b>	<b>Locom Act</b>	<b>Pinna Reflex</b>	<b>Toe Pinch</b>	<b>Vert Hind</b>	<b>Grip Strength</b>	<b>Pupil Size</b>	<b>Temperature</b>	<b>Vocalization</b>
$\beta_{0k}$ Intercept	-0.558(0.515)	-3.005 (0.912)*	-3.879 (1.088)*	-1.343 (0.538)+	1.204 (0.044)*	7.267 (0.329)*	0.5759 (0.110)*	20.979 (4.693)*
$\beta_{1k}$ Treat	0.658 (0.708)	4.343 (0.885)*	4.356 (1.228)*	4.426 (1.244)*	0.072 (0.058)	9.001 (1.016)*	-1.006 (0.242)*	-17.275 (5.066)*
$\beta_{2k}$ Time	-0.106 (0.067)	0.048 (0.040)	0.037 (0.087)	-0.020 (0.042)	0.001 (0.003)	-0.070 (0.024)*	0.006 (0.009)	0.117 (0.145)
$\beta_{3k}$ Day	-0.338 (0.505)	0.207 (0.722)	-1.222 (0.936)	-1.006 (0.336)*	-0.087 (0.033)*	-0.640 (0.217)*	0.071 (0.074)	-1.051 (1.858)
$\beta_{4k}$ Time*Day	0.060 (0.033)	0.001 (0.032)	0.047 (0.025)	0.021 (0.035)	-0.004 (0.002)+	0.069 (0.021)*	-0.0124 (0.007)	0.001 (0.056)
$\beta_{5k}$ Treat*Time	0.135 (0.056)+	-0.074 (0.021)*	-0.148 (0.087)	-0.079 (0.049)	0.002 (0.002)	-0.173 (0.029)*	0.0404 (0.008)*	0.031 (0.112)
$\beta_{6k}$ Treat*Day	-0.945 (0.540)	-0.935 (0.618)	0.559 (0.941)	-0.142 (0.515)	0.035 (0.037)	-2.644 (0.365)*	0.388 (0.084)*	-0.914 (1.939)
$\tau_k^2$ Variance RI	0.268 (0.701)	1.883 (0.910)+	2.674 (0.912)*	1.455 (0.717)+	0.014 (0.004)*	5.983 (2.310)*	1.681 (0.293)*	31.192 (13.871)+
$\sigma_k^2$ Res. Variance					0.036 (0.003)*	6.141 (0.973)*	0.577 (0.148)*	32.382 (9.362)*

## Appendix: SAS code

We show how models of a mixed type can be analyzed using standard statistical software. We focus on the SAS procedure PROC NL MIXED, and show how this programs can be used to jointly analyze two continuous and two binary longitudinal outcomes.

Assume we have data of the following form:

Obs	id	time	name	resp
1	1	0	LocActivity	1.000
2	1	0	PinnaReflex	1.000
3	1	0	GripStrength	0.976
4	1	0	Pupil Size	0.675
5	1	1	LocActivity	1.000
6	1	1	PinnaReflex	0.000
7	1	1	GripStrength	0.846
8	1	1	Pupil Size	0.698
9	2	0	LocActivity	0.000
10	2	0	PinnaReflex	1.000
11	2	0	GripStrength	0.778
12	2	0	Pupil Size	0.701
...				

In this data set, the variable 'id' is the subject's indicator and 'time' is time at which the experiment is conducted. The binary responses (Locomotor Activity and Pinna Reflex) and continuous responses (Grip Strenght and Pupil Size) are summarized in one variable ('resp').

This dataset is first restructured in pairs of outcomes for each subject and at each time point:

Obs	id	time	name1	resp1	name2	resp2	pair
1	1	0	LocActivity	1.000	PinnaReflex	1.000	1
2	1	0	LocActivity	1.000	GripStrenght	0.976	2
3	1	0	LocActivity	1.000	PupilSize	0.675	3
4	1	0	PinnaReflex	1.000	GripStrenght	0.976	4
5	1	0	PinnaReflex	1.000	PupilSize	0.675	5
6	1	0	GripStrength	0.976	PupilSize	0.675	6
7	1	1	LocActivity	1.000	PinnaReflex	0.000	1
8	1	1	LocActivity	1.000	GripStrength	0.846	2
9	1	1	LocActivity	1.000	Pupil Size	0.698	3
10	1	1	PinnaReflex	0.000	GripStrength	0.846	4
11	1	1	PinnaReflex	0.000	Pupil Size	0.698	5
12	1	1	GripStrength	0.846	Pupil Size	0.698	6
...							

Finally, we create indicator variables 'id11', 'id12', ... indicating whether the first response in the considered pair is the 1th, 2d, ... response in the dataset (LocActivity, PinaReflex, GripStrength, PupilSize), and the same for the second response in the considered pair. Also an indicator variable 'idpair1', 'idpair2', ... is made, indicating whether the observation refers to pair 1, 2, ... Using the latter dataset, we analyze the data using the following program:

```

proc nlmixed hess;

/*Specify LL1*/
if name1="LocActivity" then do;
    eta1 = beta11 + beta21*time + u1;
    p1 = exp(eta1)/(1+exp(eta1));
    ll1 = (resp1)*log(p1) + (1- resp1)*log(1-p1);
end;
if name1="PinnaReflex" then do;
    eta1 = beta12 + beta22*time + u1;
    p1 = exp(eta1)/(1+exp(eta1));
    ll1 = (resp1)*log(p1) + (1- resp1)*log(1-p1);
end;
if name1="GripStrenght" then do;
    mean1 = beta13 + beta23*time + u1;
    ll1 = -0.5*log(3.14159265358) -log(sigma3) -0.5*(resp1-mean1)**2/(sigma3**2);
end;
if name1="Pupil Size" then do;
    mean1 = beta14 + beta24*time + u1;
    ll1 = -0.5*log(3.14159265358) -log(sigma4) -0.5*(resp1-mean1)**2/(sigma4**2);
end;

/*Specify LL2*/
if name2="LocActivity" then do;
    eta2 = beta11 + beta21*time + u2;
    p2 = exp(eta2)/(1+exp(eta2));
    ll2 = (resp2)*log(p2) + (1- resp2)*log(1-p2);
end;
if name2="PinnaReflex" then do;
    eta2 = beta12 + beta22*time + u2;
    p2 = exp(eta2)/(1+exp(eta2));
    ll2 = (resp2)*log(p2) + (1- resp2)*log(1-p2);
end;
if name2="GripStrenght" then do;
    mean2 = beta13 + beta23*time + u2;
    ll2 = -0.5*log(3.14159265358) -log(sigma3) -0.5*(resp2-mean2)**2/(sigma3**2);
end;
if name2="Pupil Size" then do;
    mean2 = beta14 + beta24*time + u2;
    ll2 = -0.5*log(3.14159265358) -log(sigma4) -0.5*(resp2-mean2)**2/(sigma4**2);
end;

```

```

ll=ll1+ll2;

sigmab1=(id11*tau1+id12*tau2+id13*tau3+id14*tau4);
sigmab2=(id21*tau1+id22*tau2+id23*tau3+id24*tau4);
rhob=(idpair1*rho12+idpair2*rho13+idpair3*rho14
      +idpair4*rho23+idpair5*rho24+idpair6*rho34);

model resp1 ~ general(ll);
random u1 u2 ~normal([0,0],[sigmab1**2,rhob*sigmab1*sigmab2,sigmab2**2])
           subject=id*pair;
ods output parameterestimates=parmest;
run;

```

It should be noted that for each line in the dataset a contribution to the likelihood function is calculated, based on the two components in the pair of outcomes, where each component comes from a normal distribution or from a Bernoulli distribution. The correlation among the two outcomes is modeled via bivariate normally distributed random effects. Due to the nonlinear parameter estimation procedure, a careful selection of starting values may be required.

The parameter estimates are written to the output-file 'parmest'. To calculate the robust variance estimator corresponding to these parameter estimates, the matrices  $J(\Theta)$  and  $K(\Theta)$  have to be estimated. These are obtained from the program

```

proc nlmixed hess tech=none ;
parms /data=parmest;

```

Specificaton of likelihood...

```

by id;
ods output parameterestimates=gradientid hessian=hessianid;
run;

```

The file 'gradientid' contains the gradient-vector, and is used to calculate the pairwise-products matrix  $K(\Theta)$ . The uncorrected Hessian-matrix, stored into the output-file 'hessian', refers to the matrix  $J(\Theta)$  in (6). The program can be provided upon request.