



HAL
open science

Multivariate Analysis of Longitudinal Ordinal Data with Mixed Effects Models, with Application to Clinical Outcomes in Osteoarthritis

Céline M. Laffont, Marc Vandemeulebroecke, Didier Concordet

► **To cite this version:**

Céline M. Laffont, Marc Vandemeulebroecke, Didier Concordet. Multivariate Analysis of Longitudinal Ordinal Data with Mixed Effects Models, with Application to Clinical Outcomes in Osteoarthritis. 2013. hal-01003741

HAL Id: hal-01003741

<https://hal.science/hal-01003741>

Preprint submitted on 10 Jun 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multivariate Analysis of Longitudinal Ordinal Data
with Mixed Effects Models, with Application to
Clinical Outcomes in Osteoarthritis

Celine Marielle Laffont^{1,2}, Marc Vandemeulebroecke³, Didier Concordet^{1,2*}

¹INRA, UMR 1331, Toxalim, F-31027 Toulouse, France

²Universite de Toulouse, INPT, ENVT, UPS, EIP, F-31076 Toulouse, France

³Novartis Pharma AG, Basel, Switzerland

*Corresponding author: d.concordet@envt.fr

April 2, 2014

Abstract

Our objective was to evaluate the efficacy of robenacoxib in osteoarthritic dogs using four ordinal responses measured repeatedly over time. We propose a multivariate probit mixed effects model to describe the joint evolution of endpoints and to evidence the intrinsic correlations between responses that are not due to treatment effect. Maximum likelihood computation is intractable within reasonable time frames. We therefore use a pairwise modeling approach in combination with a stochastic EM algorithm.

Multidimensional ordinal responses with longitudinal measurements are a common feature in clinical trials. However, the standard methods for data analysis use unidimensional models, resulting in a loss of information. Our methodology provides substantially greater insight than these methods for the evaluation of treatment effects and shows a good performance at low computational cost. We thus believe that it could be used in routine practice to optimize the evaluation of treatment efficacy.

KEYWORDS: categorical data, clinical scores, pairwise fitting, pseudolikelihood.

1 Introduction

Osteoarthritis is a chronic disease characterized by articular cartilage lesions, bone remodeling with osteophytes, inflammation and pain. It affects not only humans but also companion animals and constitutes a common disorder in dogs. Robenacoxib is a non-steroidal anti-inflammatory drug that has been developed for the treatment of osteoarthritis in dogs to reduce pain and inflammation (Reymond et al., 2012). In clinical trials, changes in animal behavior, locomotion and demeanor patterns are considered the most relevant end points for the evaluation of treatment effects. Because the objective evaluation of these signs is difficult, scoring systems consisting of multiple ordinal responses are used. Typically, four ordinal outcomes are measured in clinical trials, including the posture of a dog while standing, lameness while walking, lameness while trotting and pain at the palpation/mobilization of the affected limb (Table 1). Figure 1 shows the results of a clinical trial in which 125 osteoarthritic dogs received robenacoxib once daily over 84 days. All outcomes clearly improved over time, with an increasing percentage of subjects in the lowest (normal) category and a decreasing percentage of subjects in the highest (most severe) category until a plateau was reached. Several questions of clinical interest are addressed in such clinical trials, i.e., what are the percentages of subjects with no symptoms (clinical cure), no or mild symptoms (acceptable clinical status), and clinical improvement (improving on all scales by at least one grade) and what is the time necessary until clinical improvement, etc. To avoid a lengthy paper, we deliberately chose to address only the first two questions.

The questions raised are multidimensional in nature and pose a challenge for the data analyst. A routinely used method consists of computing the sum of outcomes and analyzing this sum as if it were a continuous variable. Obviously, this method is not optimal. The first reason is that the multidimensional nature of the data is lost; the second, that it ignores the ordinal nature of the data and uses purely arbitrary coding (0, 1, 2, 3...) as a metric. A classical alternative is to analyze each ordinal response separately (so-called univariate analysis). The most popular models in this case apply a link function to cumu-

Posture at a stand	Lameness at walk	Lameness at trot	Pain at palpation
0 - normal	0 - normal	0 - normal	0 - none
1 - slightly abnormal	1 - mild	1 - mild	1 - mild
2 - markedly abnormal	2 - obvious	2 - obvious	2 - moderate
3 - severely abnormal	3 - marked	3 - marked	3 - severe

Table 1: Ordinal outcomes measured in clinical trials for the evaluation of the effects of robenacoxib in dogs with chronic osteoarthritis

lative probabilities, generally a logit or a probit (for a review, see Liu and Agresti, 2005). It is possible to interpret these models as a categorization of an underlying continuous unobserved variable (latent variable interpretation). As an example, an ordinal response for pain Y can be viewed as the categorization of a continuous underlying 'pain' variable Y^* , which is the true variable of interest for the evaluation of drug efficacy.

The univariate analysis strategy shows, however, some limitations. In particular, this approach does not document how the various responses jointly evolve in one subject. In fact, a subject might improve on one outcome but not another, and such information is critical to gauging the overall efficacy of the drug. To derive a joint distribution from separate univariate analyses (estimating marginal distributions), one has to assume that the different outcomes are independent. This assumption is quite restrictive and not very plausible because the multiplicity of outcomes in clinical trials results precisely from the decision to account for all aspects of the disease, despite any overlapping information. Thus, it can be anticipated that the outcomes are more or less correlated with the existence of potential redundancies. Obviously, a multivariate analysis adapted to longitudinal ordinal data would be of great value to address these issues.

Different approaches have been proposed in the literature to evaluate the association between several (K) observed variables. A popular approach is the so-called "factor analysis", which aims at finding a set of K independent latent variables (factors) that is smaller in number than the observed variables ($p < K$) but contains essentially the same information. An extreme case is when a single factor is used to model the correlations among all observed variables (Sammel et al., 1997, 1999; Teixeira-Pinto and Normand, 2009). Factor analysis is appealing, as it provides a simple framework to model correla-

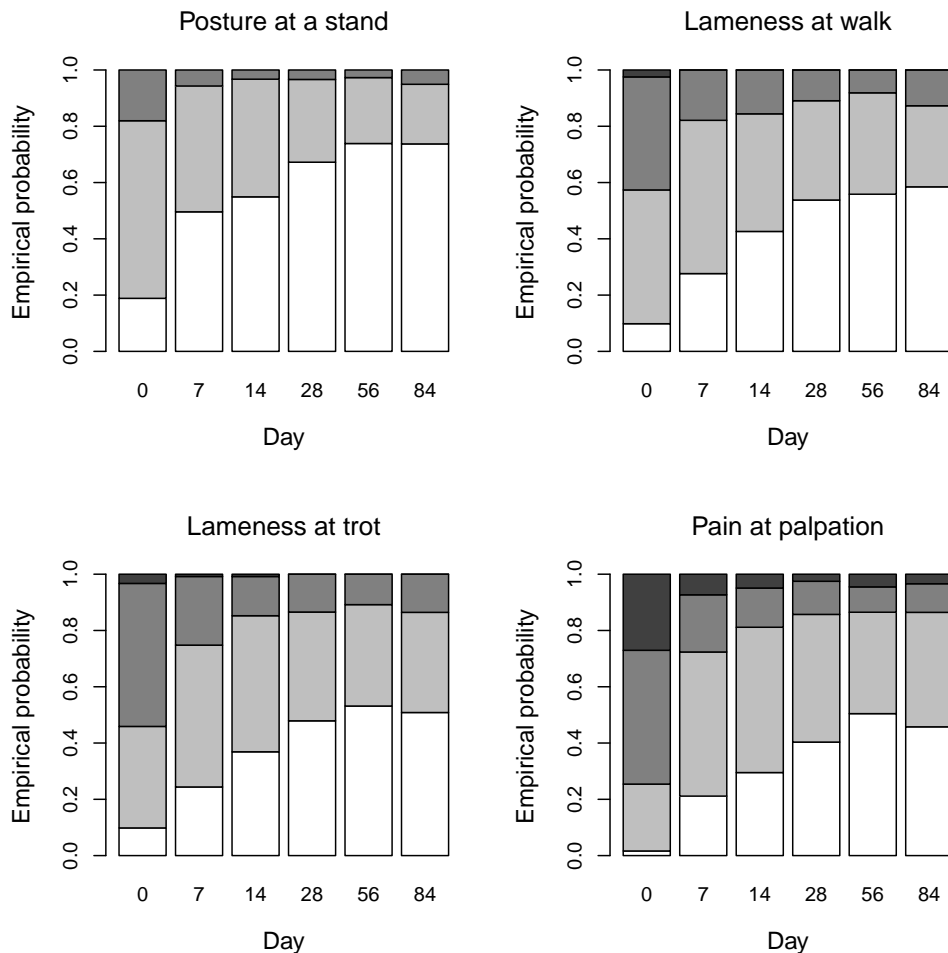


Figure 1: Empirical probability to observe a given category for posture, lameness at walk, lameness at trot and pain at palpation in a clinical trial in which 125 osteoarthritic dogs received robenacoxib once per day over 84 days at the oral dose of 1-2 mg/kg. Clinical examinations were performed at seven subsequent visits during the treatment period: at baseline and at days 7, 14, 28, 56 and 84. The variations in gray are used to represent the severity of symptoms, from no symptoms (white) to the most severe ones (dark gray). Note that the most severe category is either not or rarely observed for posture and lameness at walk/trot.

tions between the observed variables of different natures, including ordinal data, and to detect potential redundancies among these data (Shi and Lee, 2000; Dunson, 2003; Lee and Song, 2004; Katsikatsou et al., 2012). On the other hand, this approach must specify *a priori* the number of factors p . To select a reasonable value for p , it is necessary to fit and compare several models, which may be a considerably time-consuming process, especially when K is large. Another approach is to leave the correlations between the observed variables free of any structure and to estimate these correlations based on the data. This approach is more exploratory and provides the rationale for further factor analysis based on correlation estimates.

There is a large literature on the multivariate analysis of ordinal data. However, most of the published papers are restricted to the analysis of cross-sectional multivariate outcomes (Dale, 1986; Sammel et al., 1997; Shi and Lee, 2000; Lee and Song, 2004; Qaqish and Ivanova, 2006; Teixeira-Pinto and Normand, 2009; Katsikatsou et al., 2012) or to the analysis of repeated measurements over time for a single ordinal response (Glonck and McCullagh, 1995) or both (Molenberghs and Lesaffre, 1994; Chib and Greenberg, 1998; Molenberghs and Lesaffre, 1999). By comparison, little work has been performed on the analysis of multivariate longitudinal ordinal data, which requires the researcher first to take into account the dependence of observations in one subject, as well as to model the cross-sectional associations among multiple outcomes. A common way to accommodate repeated measurements in time is to apply mixed effects models where subject-specific random effects are used to tie together the observations from a same subject. The difference between the models then lies in the manner by which the contemporaneous associations between outcomes are modeled conditionally on subject-specific random effects and by the way serial correlations are addressed. Ten Have and Morabia (1999) proposed a model for bivariate binary outcomes with univariate logit components for the marginal distributions and log odds ratio components for the association of outcomes at given time points. Todem et al. (2007) suggested the use of probit mixed effects models based on the concept of continuous latent variables. In their work, Todem et al. (2007) assumed a multivariate Gaussian distribution for the latent variables conditionally on subject-specific random

effects. Other authors have assumed the conditional independence of the latent variables, assessing solely the correlations between subject-specific random effects (Liu and Hedeker, 2006; Fieuws et al., 2006). The complexity and flexibility of the model can be greatly increased; in fact, Dunson (2003) proposed flexible correlation structures between outcomes and times (serial correlation). However, as discussed by the author himself, the costs of this versatility are the accompanying potential identifiability and estimability issues, which favor the use of a Bayesian framework, as implemented in his article.

In the present paper, the challenge was to identify a good compromise between the complexity (and hence flexibility) of a model and its computation time. The ultimate objective was to propose a method that could be routinely used in practice for the analysis of multivariate longitudinal ordinal data in clinical trials. The model developed by Todem et al. (2007) offered sufficient flexibility and required a limited number of parameters to describe the associations among outcomes. Because Todem's work was restricted to the bivariate case, we propose an extension of this model to accommodate $K > 2$ outcomes ($K = 4$ in robenacoxib case study). Although this extension may appear conceptually modest, it requires a very substantial revision in the estimation procedure because the existing methods in the bivariate setting become intractable in even the trivariate setting. Here, we propose an estimation method that yields unbiased estimates within acceptable time frames. This method combines a pairwise approach (Varin et al., 2011) with the use of a stochastic EM algorithm (Delyon et al., 1999) adapted from Kuhn and Lavielle (2005), in the same spirit as Booth and Hobert (1999).

The article is organized as follows. In section 2, we describe the general model and likelihood used to model the longitudinal ordinal data in the robenacoxib case study. We discuss our fitting strategy in Section 3 and apply it to the robenacoxib data in Section 4. The objective of the robenacoxib analysis was twofold. The first objective was to estimate the joint evolution of ordinal responses over time, assessing the consequences of assuming independent outcomes for the evaluation of robenacoxib efficacy. The second objective was to identify possible redundancies between responses based on the estimation of intrinsic correlations. Finally, we conducted a Monte Carlo investigation of performance

for the proposed methodology and report the results in Section 5.

2 General model and likelihood

Suppose that K ordinal responses are measured repeatedly over time in N subjects. Each response k ($k = 1, \dots, K$) takes values in the range $0, \dots, c_k$ and $c_k + 1$ is the total number of categories. Let $\mathbf{Y}_{ij} = (Y_{ij}^{(1)}, \dots, Y_{ij}^{(K)})^T$ be the vector of observations for the K ordinal responses in individual i at time t_{ij} ($i = 1, \dots, N$, $j = 1, \dots, n_i$) and $\mathbf{Y}_i = (\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{i, n_i})$ the matrix obtained by an horizontal concatenation of \mathbf{Y}_{ij} . Then, assuming that each ordinal response $Y_{ij}^{(k)}$ comes from the categorization of an underlying latent variable $Y_{ij}^{(k)*}$ and that this categorization is achieved by using a vector of cut-points with monotonically increasing values $\mathbf{a}^{(k)} = (a_1^{(k)}, \dots, a_{c_k}^{(k)})^T$, we obtain the following relationship in K dimensions

$$\begin{aligned} (Y_{ij}^{(1)} = u; \dots; Y_{ij}^{(K)} = v) &\Leftrightarrow (a_u^{(1)} \leq Y_{ij}^{(1)*} < a_{u+1}^{(1)}; \dots; a_v^{(K)} \leq Y_{ij}^{(K)*} < a_{v+1}^{(K)}) \\ &u = 0, \dots, c_1 \text{ and } v = 0, \dots, c_K \end{aligned} \quad (2.1)$$

where $a_0^{(k)} = -\infty$ and $a_{c_k+1}^{(k)} = +\infty$ for all $k = 1, \dots, K$. Note that in the case of binary outcomes, only one cut-point is necessary.

At a second level, mixed effects models are used for the latent variables, with the following expression for the k^{th} latent variable in individual i at time t_{ij}

$$\begin{aligned} Y_{ij}^{(k)*} &= g^{(k)}(t_{ij}, \mathbf{x}_i, \boldsymbol{\beta}) + b_i^{(k)} + e_{ij}^{(k)}, \\ \mathbf{e}_{ij} &\stackrel{iid}{\sim} N(0, \boldsymbol{\Sigma}), \\ \mathbf{b}_i &\stackrel{iid}{\sim} N(0, \boldsymbol{\Omega}), \end{aligned} \quad (2.2)$$

$$i = 1, \dots, N, \quad j = 1, \dots, n_i, \quad k = 1, \dots, K,$$

where $g^{(k)}(\cdot)$ is a known real function, $\mathbf{x}_i = (x_1, \dots, x_n)^T$ is a vector of time-independent covariates for subject i (e.g., treatment dosing information), $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ is a vector of unknown parameters common to all subjects, $\mathbf{b}_i = (b_i^{(1)}, \dots, b_i^{(K)})^T$ is a vector of random effects specific to subject i , and $\mathbf{e}_{ij} = (e_{ij}^{(1)}, \dots, e_{ij}^{(K)})^T$ is a vector of residual random effects and is assumed independent from \mathbf{b}_i . The matrix $\boldsymbol{\Omega}$ is a $(K \times K)$ variance-covariance matrix and $\boldsymbol{\Sigma}$ a $(K \times K)$ correlation matrix (the variances of \mathbf{e}_{ij} components cannot be estimated due to a lack of identifiability and are set to 1 with no loss of generality, see Liu and Agresti, 2005). The random effects \mathbf{b}_i account for the longitudinal association of data from the same individual across time. The diagonal elements of $\boldsymbol{\Omega}$ quantify between-subject variability, and the off-diagonal elements of $\boldsymbol{\Omega}$ measure the overall association between outcomes. The residual random error \mathbf{e}_{ij} accounts for the variations with time conditionally to the subject-specific random effects \mathbf{b}_i , reflecting within-subject variability. The correlation matrix $\boldsymbol{\Sigma}$ models the contemporaneous association between the outcomes, given \mathbf{b}_i . Here, we assume that $\boldsymbol{\Sigma}$ does not change with time or with covariates, but we could extend this model to accommodate more general situations, in which the correlation structures in $\boldsymbol{\Sigma}$ are more complex, possibly depending on time and covariates, as shown in Todem et al. (2007) for the bivariate case. We denote $\boldsymbol{\theta}^* = (\mathbf{a}, \boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{\Sigma})$ the parameters to be estimated, where \mathbf{a} is a vector obtained by the concatenation of vectors $\mathbf{a}^{(k)}$.

Because the subject-specific random effects \mathbf{b}_i are unobserved, the maximum likelihood estimation is based on the marginal likelihood of the multivariate response

$$L_{\boldsymbol{\theta}^*}(\mathbf{y}) = \prod_{i=1}^N L_{\boldsymbol{\theta}^*}(\mathbf{y}_i) = \prod_{i=1}^N \int L_{\boldsymbol{\theta}^*}(\mathbf{y}_i | \mathbf{b}_i) P_{\boldsymbol{\theta}^*}(\mathbf{b}_i) d\mathbf{b}_i \quad (2.3)$$

where \mathbf{y}_i is the observed response matrix, $L_{\boldsymbol{\theta}^*}(\mathbf{y}_i | \mathbf{b}_i)$ is the conditional density of \mathbf{y}_i given \mathbf{b}_i and $P_{\boldsymbol{\theta}^*}(\mathbf{b}_i)$ is the density of \mathbf{b}_i . Although the latent variable model is linear in the random effects, the multivariate ordinal response model is not linear; thus, there is no analytical solution for the marginal likelihood. Assuming the independence of \mathbf{e}_{ij} across

time points, $L_{\theta^*}(\mathbf{y}_i|\mathbf{b}_i)$ is calculated as the product of the conditional probabilities at all time points, namely

$$L_{\theta^*}(\mathbf{y}_i|\mathbf{b}_i) = \prod_{j=1}^{n_i} \prod_{u=0}^{c_1} \dots \prod_{v=0}^{c_K} [P_{\theta^*}(Y_{ij}^{(1)} = u; \dots; Y_{ij}^{(K)} = v|\mathbf{b}_i)]^{I(Y_{ij}^{(1)}=u) \times \dots \times I(Y_{ij}^{(K)}=v)} \quad (2.4)$$

where $I(\cdot)$ is the indicator function. Then, based on the relations described in equation (2.1) and the latent variable models specified in (2.2), we get

$$\begin{aligned} P_{\theta^*}(Y_{ij}^{(1)} = u; \dots; Y_{ij}^{(K)} = v|\mathbf{b}_i) &= P_{\theta^*}(a_u^{(1)} \leq Y_{ij}^{(1)*} < a_{u+1}^{(1)}; \dots; a_v^{(K)} \leq Y_{ij}^{(K)*} < a_{v+1}^{(K)}|\mathbf{b}_i) \\ &= P_{\theta^*}(a_u^{(1)} - f_{ij}^{(1)} \leq e_{ij}^{(1)} < a_{u+1}^{(1)} - f_{ij}^{(1)}; \\ &\quad \dots; a_v^{(K)} - f_{ij}^{(K)} \leq e_{ij}^{(K)} < a_{v+1}^{(K)} - f_{ij}^{(K)}|\mathbf{b}_i) \end{aligned} \quad (2.5)$$

where $f_{ij}^{(k)} = g^{(k)}(t_{ij}, \mathbf{x}_i, \boldsymbol{\beta}) + b_i^{(k)} = E(Y_{ij}^{(k)*}|\mathbf{b}_i)$ for all $k = 1, \dots, K$.

We write $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\rho})$, where $\boldsymbol{\rho}$ denotes the collection of all distinct parameters in $\boldsymbol{\Sigma}$. Because $\mathbf{e}_{ij} = (e_{ij}^{(1)}, \dots, e_{ij}^{(K)})^T$ follows a multivariate standard normal distribution with cumulative distribution function $\Phi_{K,\boldsymbol{\rho}}$, the conditional joint probability specified in equation (2.5) is calculated by integrating the multivariate normal density on hypercubes of \mathbb{R}^K . When $K = 2$ (bivariate case), the equation (2.5) reduces to the expression

$$\begin{aligned} P_{\theta^*}(Y_{ij}^{(1)} = u, Y_{ij}^{(2)} = v|\mathbf{b}_i) &= \Phi_{2,\boldsymbol{\rho}}(a_{u+1}^{(1)} - f_{ij}^{(1)}, a_{v+1}^{(2)} - f_{ij}^{(2)}) - \Phi_{2,\boldsymbol{\rho}}(a_u^{(1)} - f_{ij}^{(1)}, a_{v+1}^{(2)} - f_{ij}^{(2)}) \\ &\quad - \Phi_{2,\boldsymbol{\rho}}(a_{u+1}^{(1)} - f_{ij}^{(1)}, a_v^{(2)} - f_{ij}^{(2)}) + \Phi_{2,\boldsymbol{\rho}}(a_u^{(1)} - f_{ij}^{(1)}, a_v^{(2)} - f_{ij}^{(2)}) \end{aligned} \quad (2.6)$$

where $\boldsymbol{\rho} = (\rho_{12})$ is the correlation coefficient between $e_{ij}^{(1)}$ and $e_{ij}^{(2)}$ and $\Phi_{2,\boldsymbol{\rho}}$ is the bivariate normal cumulative distribution function. For $K > 2$, the expression of the conditional joint probability of \mathbf{y}_{ij} gains rapidly in complexity as the number of dimensions increases.

3 Fitting strategy and estimation algorithm

Clearly, the maximum likelihood estimator is desirable, as it shows nice asymptotic properties. Unfortunately, this estimator cannot be calculated within a reasonable time-frame in our situation because the computation of $L_{\theta^*}(\mathbf{y}_i|\mathbf{b}_i)$, as defined in equation (2.4), involves the multivariate cumulative distribution function $\Phi_{K,\rho}$. As will be discussed below, the computation of this function is a tedious and rate-limiting task when $K \geq 3$. This is the reason why we chose to implement a pairwise modeling approach in which pseudo-likelihoods are optimized instead of the likelihood. Although the pairwise approach circumvents the problem of computing $\Phi_{K,\rho}$, the pseudo-likelihoods are in the same functional form as the likelihood in equation (2.3) and have no closed-form expression. We thus used a stochastic EM algorithm to optimize these pseudo-likelihoods in a short time-frame.

As in the univariate case, there is no analytical expression for the multivariate cumulative distribution function $\Phi_{K,\rho}$, but its numerical computation is much more complex due to the 'curse' of dimensionality. Much research has been devoted to the computation of multivariate normal distribution functions in recent decades, and reliable and accurate numerical algorithms are now available (for a review, see Genz and Bretz, 2009). However, while computations are very rapid and accurate in the bivariate case, higher dimensions increase considerably the complexity and amount of calculations needed to maintain acceptable numerical integration errors. In our situation, additional limitations need to be taken into consideration. Indeed, as illustrated for the simple bivariate case in (2.6), the evaluation of the joint probability generally requires the integration of the Gaussian density on a hypercube, inducing multiple intermediate computations, the number of which increases with the number of dimensions. Furthermore, these computations must be repeated many times during the iterative estimation process, i.e., at each evaluation of the conditional likelihood of \mathbf{y}_{ij} for all $i = 1, \dots, N$ and all $j = 1, \dots, n_i$. This leads to a major computational burden, even for moderate K values (e.g., $K = 4$ in our robenacoxib case study).

To reduce the computational complexity when more than two outcomes are analyzed, the dimensionality of the problem needs to be reduced. As stated above, we chose to implement a pairwise modelling approach. The pairwise approach dates at least back to Besag (1974) and represents an attractive solution to decrease the dimensionality of the problem. This approach has been successfully applied to the analysis of multivariate longitudinal continuous or discrete data (Molenberghs and Verbeke, 2005; Fieuws and Verbeke, 2006; Fieuws et al., 2006) and to non-longitudinal multivariate ordinal data (Katsikatsou et al., 2012). The pairwise approach consists of fitting all pairs of ordinal responses separately. In other words, instead of maximizing the likelihood (2.3), we analyze all pairs of outcomes independently and maximize the likelihood for each specific pair (r, s) , which is called a pseudo-likelihood and takes the following form:

$$\prod_{i=1}^N L_{\theta_{r,s}}(\mathbf{y}_i^{(r)}, \mathbf{y}_i^{(s)}) = \prod_{i=1}^N \int L_{\theta_{r,s}}(\mathbf{y}_i^{(r)}, \mathbf{y}_i^{(s)} | \mathbf{b}_i) P_{\theta_{r,s}}(\mathbf{b}_i) d\mathbf{b}_i \quad (3.1)$$

Here, $\mathbf{y}_i^{(r)}$ and $\mathbf{y}_i^{(s)}$ are the observed response vectors for outcomes r and s , respectively, and $\theta_{r,s}$ represents the vector of all parameters in the bivariate model corresponding to pair (r, s) . For the sake of simplicity, let us denote by θ_p the vector $\theta_{r,s}$ where p corresponds to pair (r, s) with $p = 1, \dots, P$ and $P = K(K - 1)/2$ the total number of possible pairs. Similarly, let us denote by \mathbf{y}_i^p the observation matrix containing the observed response vectors $\mathbf{y}_i^{(r)}$ and $\mathbf{y}_i^{(s)}$, and by L_{ip} the individual likelihood $L_{\theta_p}(\mathbf{y}_i^p) = L_{\theta_{r,s}}(\mathbf{y}_i^{(r)}, \mathbf{y}_i^{(s)})$.

Let $\theta = (\theta_1, \dots, \theta_P)$ be the vector resulting from the concatenation of all pair-specific parameter vectors θ_p . Note that the vectors $\theta^* = (\mathbf{a}, \boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{\Sigma})$ and θ do not have the same length; some parameters in θ^* have a single counterpart in θ (for example, any element of $\boldsymbol{\rho}$), whereas other parameters in θ^* have multiple counterparts in θ (for example, variances of subject-specific random effects). In the latter case, a single estimate is obtained for θ^* by averaging all corresponding pair-specific estimates in $\hat{\theta}$. More precisely, $\hat{\theta}^*$ is obtained as $\mathbf{A}\hat{\theta}$, where \mathbf{A} is a matrix containing the appropriate coefficients to calculate the averages over all pairs. Because the estimates stored in $\hat{\theta}$ are obtained by maximum

likelihood estimation within each pair, they show classical asymptotic properties, including consistency and asymptotic normality, and any linear combination of these estimates will share the same asymptotic properties. This process results in a normal asymptotic distribution of $\hat{\boldsymbol{\theta}}^*$ with mean $\boldsymbol{\theta}^*$. Regarding the uncertainty in parameter estimates, the standard errors of $\hat{\boldsymbol{\theta}}^*$ cannot be obtained simply by averaging the standard errors of the estimates in $\hat{\boldsymbol{\theta}}$, as we need to take into consideration the variability among pair-specific estimates. Furthermore, these estimates are expected to be correlated, as they are derived from datasets with overlapping information because a single outcome is analyzed several times across pairs. Under asymptotic conditions, $\hat{\boldsymbol{\theta}}$ follows a normal distribution (Fieuw and Verbeke, 2006)

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim N(0, \mathbf{J}^{-1} \mathbf{K} \mathbf{J}^{-1}) \quad (3.2)$$

where \mathbf{J} is a block-diagonal matrix with diagonal blocks \mathbf{J}_{pp} , and \mathbf{K} is a symmetric matrix containing blocks \mathbf{K}_{pq} . These blocks are given by

$$\mathbf{J}_{pp} = -\frac{1}{N} \sum_{i=1}^N E \left(\frac{\partial^2 l_{ip}}{\partial \boldsymbol{\theta}_p \partial \boldsymbol{\theta}_p^T} \right)$$

$$\mathbf{K}_{pq} = \frac{1}{N} \sum_{i=1}^N E \left(\frac{\partial l_{ip}}{\partial \boldsymbol{\theta}_p} \frac{\partial l_{iq}}{\partial \boldsymbol{\theta}_q^T} \right), \quad p, q = 1, \dots, P,$$

with $l_{ip} = \log(L_{ip})$ being the logarithm of the marginal individual pseudo-likelihood in subject i . As in Fieuw and Verbeke (2006), the blocks are calculated by dropping the expectations and replacing the unknown parameters with their estimates. The variance of $\hat{\boldsymbol{\theta}}^*$ is then obtained as $\mathbf{A} \boldsymbol{\Gamma}(\hat{\boldsymbol{\theta}}) \mathbf{A}^T$ where $\boldsymbol{\Gamma}(\hat{\boldsymbol{\theta}})$ is the variance of $\hat{\boldsymbol{\theta}}$ obtained in equation (3.2). Both \mathbf{J} and \mathbf{K} matrices are approximated using numerical derivatives.

As stated at the beginning of this section, the estimation of $\boldsymbol{\theta}_p$ for each pair is achieved by maximizing the pseudo-likelihood in equation (3.1), which has the same functional form as the likelihood (2.3) but involves Φ_{2, ρ_p} instead of $\Phi_{K, \rho}$. The correlation coefficient $\rho_p = \rho_{rs}$ is the correlation between two latent variables $Y_{ij}^{(r)*}$ and $Y_{ij}^{(s)*}$ conditionally to \mathbf{b}_i .

For bivariate Gaussian integration, we have chosen to implement a standard numerical algorithm developed by Alan Genz for Matlab software based on a previously described method (Drezner and Wesolowsky, 1990), which provides rapid and accurate results.

The maximization of the pseudo-likelihood is performed using a stochastic version of the EM algorithm. The EM algorithm (Dempster et al., 1977) has become extremely popular in recent decades, as it can be easily implemented and applied to a wide variety of problems. This algorithm iterates between an E-step and an M-step. The E-step computes, at each iteration t , the expectation of a complete data log-likelihood, given the observations and the current model parameter estimates obtained at iteration $(t-1)$. The M-step finds new parameter estimates that maximize this expected log-likelihood. With the pairwise approach, the expectation has the following expression:

$$Q(\boldsymbol{\theta}_p, \hat{\boldsymbol{\theta}}_p^{t-1}) = \sum_{i=1}^N E_{\hat{\boldsymbol{\theta}}_p^{t-1}} [\log L_{\boldsymbol{\theta}_p}(\mathbf{y}_i^p | \mathbf{b}_i) + \log P_{\boldsymbol{\theta}_p}(\mathbf{b}_i | \mathbf{y}_i^p)] \quad (3.3)$$

where \mathbf{y}_i^p is the total observation matrix corresponding to pair p for subject i . As we now address bivariate outcomes, the logarithm of the conditional individual pseudo-likelihood takes a simple form:

$$\log L_{\boldsymbol{\theta}_p}(\mathbf{y}_i^p | \mathbf{b}_i) = \sum_{j=1}^{n_i} \sum_{u=0}^{c_r} \sum_{v=0}^{c_s} \log [P_{\boldsymbol{\theta}_p}(Y_{ij}^{(r)} = u; Y_{ij}^{(s)} = v | \mathbf{b}_i)]^{I_{(Y_{ij}^{(r)}=u)} \times I_{(Y_{ij}^{(s)}=v)}}$$

where the variable $Y_{ij}^{(r)}$ takes values in $\{0, \dots, c_r\}$, the variable $Y_{ij}^{(s)}$ takes values in $\{0, \dots, c_s\}$; and $I(\cdot)$ is the indicator function.

The expectation in equation (3.3) is then approximated by a Monte Carlo sum, in which \mathbf{b}_i are simulated under the conditional density $P_{\hat{\boldsymbol{\theta}}_p^{t-1}}(\cdot | \mathbf{y}_i^p)$:

$$Q(\boldsymbol{\theta}_p, \hat{\boldsymbol{\theta}}_p^{t-1}) \cong \frac{1}{Z} \sum_{z=1}^Z \sum_{i=1}^N [\log L_{\boldsymbol{\theta}_p}(\mathbf{y}_i^p | \mathbf{b}_{i,z}) + \log P_{\boldsymbol{\theta}_p}(\mathbf{b}_{i,z})] \quad (3.4)$$

where Z is the number of Monte Carlo simulations performed and $\mathbf{b}_{i,z}$ is the vector gen-

erated at simulation z . Note that in the present case, $P_{\hat{\boldsymbol{\theta}}_p^{t-1}}(\cdot|\mathbf{y}_i^p)$ restricts to $P_{\hat{\boldsymbol{\Omega}}_p^{t-1}}(\cdot|\mathbf{y}_i^p)$. The \mathbf{b}_i cannot be exactly drawn from this conditional distribution but, based on the work of Kuhn & Lavielle (2004), we use the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970) to converge to the target distribution. The Metropolis-Hastings algorithm consists in the running of a Markov chain, the stationary distribution of which is the target distribution. This stationarity is reached after the chain has run a number of times, which is referred to as a burn-in period. In the present method, we use two successive transition mechanisms that differ only by their proposal distribution $\Pi(x, y)$, proposing y when the chain is in x . The first proposal distribution at the $t-1$ iteration of the EM algorithm is $N(0, \hat{\boldsymbol{\Omega}}_p^{t-1})$, and the second proposal distribution is $N(x, 0.3^2 \times \mathbf{Diag})$ where \mathbf{Diag} is diagonal matrix with components equal to the variance terms in $\hat{\boldsymbol{\Omega}}_p^{t-1}$. We run the Markov Chain $M = M_1 + M_2 + Z$ times, where M_1 and M_2 are the burn-in periods for the two transition mechanisms. Recall that Z is the number of terms used to approximate the EM function Q in (3.4). At iteration m of the chain, we simulate $\mathbf{b}_{i,m}$ with the transition probability

$$P(\mathbf{b}_i|\mathbf{b}_{i,m-1}) = \min \left\{ \frac{P(\mathbf{b}_i|\mathbf{y}_i^p) \times \Pi(\mathbf{b}_i, \mathbf{b}_{i,m-1})}{P(\mathbf{b}_{i,m-1}|\mathbf{y}_i^p) \times \Pi(\mathbf{b}_{i,m-1}, \mathbf{b}_i)}; 1 \right\}.$$

In our example, we found that $M_1 = 3$, $M_2 = 2$ and $Z = 5$ were sufficient for the two successive Markov chains. These numbers may appear rather low for the Metropolis-Hastings algorithm to reach its stationary distribution. However, when the convergence of EM is nearly obtained, the transition probability does not vary much between two successive iterations of the EM algorithm. As a result, this inhomogeneous Markov chain becomes more and more homogeneous along the EM iteration, allowing a stationary distribution to be reached.

Then, based on equation (3.4), the Q function is maximized, and new parameter estimates are produced (M step):

$$\hat{\boldsymbol{\theta}}_p = \arg \max_{\boldsymbol{\theta}_p} Q(\boldsymbol{\theta}_p, \hat{\boldsymbol{\theta}}_p^{t-1}). \quad (3.5)$$

The estimate of the variance matrix for the subject-specific random effects in pair p is computed empirically from $\mathbf{b}_{i,z}$, $z = 1, \dots, Z$. Other model parameter estimates are updated using standard methods, either the Gauss-Newton method or the classical Gradient method, when numerical difficulties are met with the inversion of the Hessian matrix.

To improve computation time and to stabilize the estimation process, the stochastic EM algorithm is run in a stepwise manner. First, optimization is performed on the marginal models, assuming independence between outcomes ($\mathbf{\Omega}$ and $\mathbf{\Sigma}$ are diagonal); then, on the joint model, assuming the independence of outcomes conditionally to the subject-specific random effects (only $\mathbf{\Sigma}$ is diagonal); and finally, on the joint model, estimating all correlations. The computations are very rapid in the first two steps, allowing reasonable estimates for (conditional) marginal models to be achieved in a short period of time. Correlations are then estimated quite rapidly within a limited number of iterations (50 iterations are usually sufficient). The convergence of the algorithm is checked graphically and is considered to be reached when the successive values of the estimates oscillate around a plateau. When such a plateau is reached, the parameter estimates are obtained by averaging the successive values of the Markov Chain (Kuhn and Lavielle, 2004).

4 Application to the evaluation of robenacoxib efficacy in osteoarthritis in dogs

Efficacy data were available from three multicenter, prospective, randomized clinical trials where robenacoxib was administered to dogs with chronic osteoarthritis. For a dog to be included in the trials, osteoarthritis had to be diagnosed on one or more joints and had to be present for at least three weeks. Other inclusion/exclusion criteria are given in details in Reymond et al. (2012) and were similar across studies. Study 1 was the largest trial performed, with 125 dogs receiving robenacoxib over 12 weeks (84 days). In the two other studies, robenacoxib was administered for a shorter period of time (28 days) and to a

smaller number of subjects (51 dogs in study 2; 61 dogs in study 3). In all trials, the drug was given as oral tablets with once-daily administrations in studies 1 and 2 and twice-daily administrations in study 3. A single dosage of 1-2 mg/kg/day was investigated in study 1, while in studies 2 and 3, three different dosages of 0.5-1, 1-2 and 2-4 mg/kg/day were administered to parallel groups of equal size (17/17/17 in study 2 and 20/21/20 in study 3). Clinical examinations were performed at various time points during the treatment: at days 0, 7, 14, 28, 56 and 84 for study 1; at days 0, 2, 7, 14 and 28 for study 2; and at days 0, 7, 14 and 28 for study 3. At each visit, the four ordinal outcomes listed in Table 1 were documented, as follows: posture at a stand, lameness at walk, lameness at trot and pain at palpation (see Reymond et al. (2012) for a more detailed description). Because the most severe category was poorly represented for three outcomes (less than 0.2, 0.6 and 1.3 % of the data for posture, lameness at walk and lameness at trot), we decided to pool the data with the adjacent categories to have sufficient observations at each level. The new coding was as follows: (0): normal, (1): slightly abnormal, (2): markedly/severely abnormal, for posture and (0): normal, (1): mild, (2): obvious/marked for lameness at walk and lameness at trot. The coding for pain at palpation was left unchanged. As it was not the purpose of this work to address missing data, any time point with one or more missing data in a single subject was excluded from the analysis. This concerned 1.2% of the observations and, in the end, a total of 4764 observations were measured at 1191 time points in 232 subjects who were available to build the model.

Three models with increasing complexity were successively developed. First, the data were explored to yield a mixed effects model that adequately described the marginal distribution of the different outcomes over time. More precisely, the data were fitted jointly but assuming the independence between outcomes (Model 1). In a second step, a model assuming the independence between outcomes conditional to subject-specific random effects was run (Model 2). Finally, the model estimating all correlations between outcomes was run using the methodology detailed in Section 3 (Model 3). Note that no pairwise approach was used for the estimation of Model 1 or Model 2. In these models, the independence of $e_{ij}^{(k)}$ allows one to decompose the multidimensional Gaussian cumulative

distribution function into a product of one-dimensional Gaussian cumulative distribution functions, which dramatically reduces computation time so that the pairwise approach is not more necessary after the implementation of the stochastic EM algorithm. For Model 1 and Model 2, the Fisher Information matrix was computed by stochastic approximations to generate the standard errors. All estimation algorithms were implemented in Microsoft Visual C++ and are available from the authors on request.

Figure 1 summarizes the data obtained for each outcome marginally in the study with the largest sample size and the longest treatment period (study 1). We can see that the daily administration of robenacoxib led to a gradual improvement of subjects in all outcomes and that the percentage of subjects in the various categories tended to stabilize after eight weeks, suggesting the existence of a plateau for treatment effect. These trends were supported by the data of the two other studies (not shown), and no difference between the dosing regimens was observed within and across studies. A simple model was selected for the latent variables to account for this evolution of over time:

$$Y_{ij}^{(k)*} = -\beta_{max}^{(k)} * (1 - \exp(-\beta_{eq} * t_{ij})) + b_i^{(k)} + e_{ij}^{(k)}, \quad (4.1)$$

where $\beta_{max}^{(k)}$ is the maximal treatment effect observed at the plateau, and an exponential function of time with coefficient β_{eq} was used to model the rate at which the plateau was reached. This type of model is widely used in clinical pharmacology when a delay in drug response is observed and the measurement of drug effect is quantitative (β_{eq} ; it is typically called the "equilibration rate constant"). Note that in equation (4.1), we used a single parameter for the equilibration rate constant common to all outcomes. A model with different rate constants, coded as β_{eq} for the first outcome and $\beta_{eq} \times \beta_k$ for the others, was tested but did not show any statistical superiority (p=0.37) over the reduced model, as confirmed by the Wald test $(\beta_2, \beta_3, \beta_4) = (1, 1, 1)$. In any case, the simplest model was retained as Model 1, and its parameter estimates are displayed in Table 2.

Model simulations were performed to evaluate the ability of Model 1 to describe the time-course of the effects on each outcome marginally. More precisely, Model 1 was used

to simulate $M = 500$ replicates of the dataset following the design of studies 1, 2 and 3 and to derive, for each outcome, the median and 95% confidence interval for the marginal probability to belong to a single category at a given time point. These were compared to the empirical probabilities calculated from the observations. Overall, the consistency between empirical and predicted probabilities was very good (Figure 2).

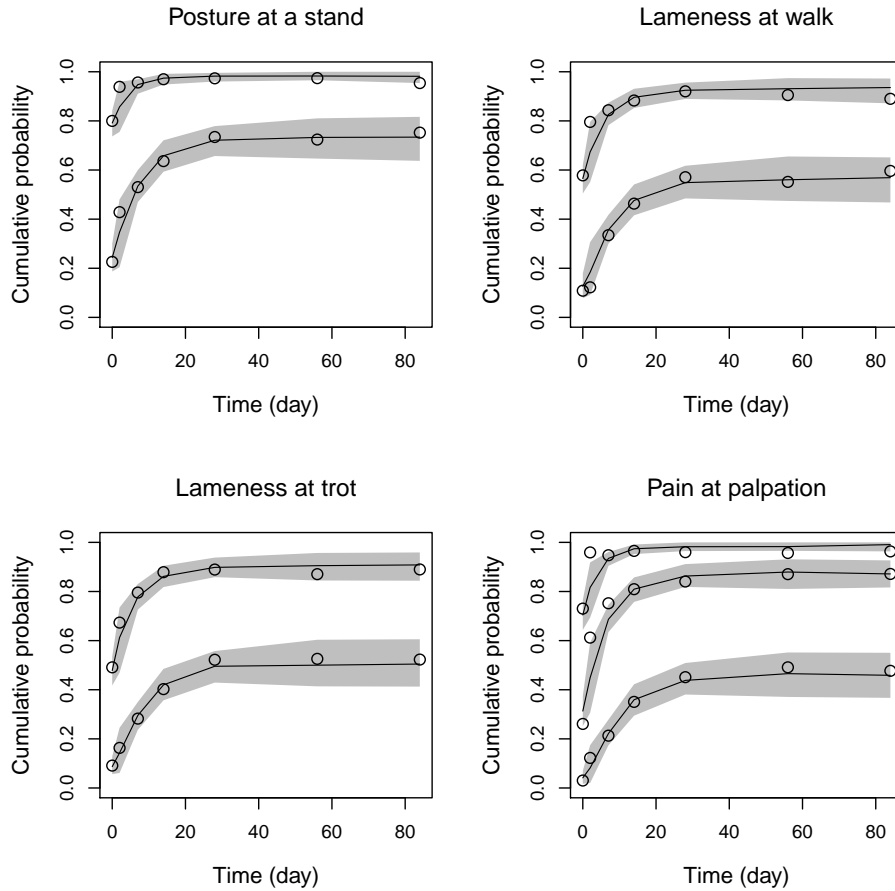


Figure 2: Time-course of cumulative probabilities for each outcome investigated marginally. The dots represent the empirical cumulative probability that the corresponding outcomes $\leq m$, with $m \in \{0, 1\}$ for posture, lameness at walk, lameness at trot, and $m \in \{0, 1, 2\}$ for pain at palpation. These values are compared to the median probability (solid line) and 95% confidence intervals (shaded area) computed from simulations under the tested model at each time point.

Thus, assuming that the marginal models specified in Model 1 were adequate, Model 2 and Model 3 were run on the data. Note that the difference in Models 1 and 2 lies solely in the estimation of the covariances in $\mathbf{\Omega} = (\omega_{mn})_{mn}$ and that the difference in Models 2

Parameters	Model 1	Model 2	Model 3
Cut-points			
$a_1^{(1)}$	-1.375 (0.155)	-1.373 (0.158)	-1.379 (0.150)
$a_2^{(1)}$	1.580 (0.163)	1.546 (0.181)	1.513 (0.164)
$a_1^{(2)}$	-2.070 (0.150)	-2.128 (0.164)	-2.029 (0.152)
$a_2^{(2)}$	0.286 (0.132)	0.246 (0.150)	0.345 (0.128)
$a_1^{(3)}$	-2.556 (0.175)	-2.586 (0.191)	-2.398 (0.145)
$a_2^{(3)}$	-0.110 (0.149)	-0.120 (0.170)	-0.006 (0.102)
$a_1^{(4)}$	-2.858 (0.146)	-2.740 (0.145)	-2.786 (0.149)
$a_2^{(4)}$	-0.813 (0.118)	-0.717 (0.119)	-0.747 (0.113)
$a_3^{(4)}$	0.884 (0.122)	0.945 (0.123)	0.916 (0.124)
Treatment effect			
$\beta_{max}^{(1)}$	2.581 (0.165)	2.520 (0.166)	2.534 (0.174)
$\beta_{max}^{(2)}$	2.350 (0.139)	2.370 (0.133)	2.376 (0.168)
$\beta_{max}^{(3)}$	2.582 (0.149)	2.613 (0.147)	2.505 (0.159)
$\beta_{max}^{(4)}$	2.688 (0.128)	2.639 (0.125)	2.639 (0.145)
β_{eq}	0.128 (0.008)	0.128 (0.008)	0.128 (0.013)
Random effects, variance terms Ω			
ω_{11}	2.840 (0.383)	2.659 (0.390)	2.811 (0.482)
ω_{22}	2.241 (0.330)	2.373 (0.404)	2.235 (0.362)
ω_{33}	2.580 (0.357)	2.926 (0.548)	2.455 (0.381)
ω_{44}	1.703 (0.219)	1.607 (0.202)	1.634 (0.226)
ω_{12}		2.037 (0.361)	1.715 (0.318)
ω_{13}		2.147 (0.444)	2.169 (0.442)
ω_{14}		0.681 (0.146)	0.685 (0.185)
ω_{23}		2.492 (0.445)	1.991 (0.305)
ω_{24}		0.640 (0.167)	0.580 (0.182)
ω_{34}		0.674 (0.226)	0.535 (0.184)
Random effects, variance terms Σ			
ρ_{12}			0.746 (0.046)
ρ_{13}			0.630 (0.057)
ρ_{14}			0.318 (0.062)
ρ_{23}			0.776 (0.039)
ρ_{24}			0.344 (0.053)
ρ_{34}			0.359 (0.060)
-2logLikelihood	7393.247	6933.052	6494.719^a
AIC	7429.247	6981.052	6554.719

^amarginal likelihood using Gauss-Legendre quadratures (8 fixed nodes) for the computation of $\Phi_{4,\rho}$

Table 2: Parameter estimates and standard errors (within parentheses) under the mixed effects model, assuming the independence between outcomes (Model 1), the independence between outcomes conditionally to subject-specific random effects (Model 2), or all correlations between outcomes (Model 3). The outcomes investigated were the posture of the dog at a stand ($k = 1$), the lameness at walk ($k = 2$), the lameness at trot ($k = 3$), and the pain at palpation ($k = 4$).

and 3 lies solely in the estimation of the correlations in Σ . The following notations were used for the components of Σ .

$$\Sigma = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{14} \\ \rho_{12} & 1 & \cdots & \rho_{24} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{14} & \rho_{24} & \cdots & 1 \end{pmatrix}$$

The parameter estimates of Models 2 and 3 are listed in Table 2. As expected, the parameter estimates characterizing marginal distributions were very similar across models, and the small differences observed were essentially attributed to the stochastic nature of the estimation algorithm. However, the estimation of the covariances in Ω (Model 2 or 3) and of Σ components (Model 3) revealed very strong correlations between some of the outcomes. To summarize and simplify the interpretation of these correlations, a Principal Component Analysis was conducted on both Ω and Σ . The two first principal components explained 86-89% of the variability in the data, and correlation circles were drawn for the first principal component versus the second principal component to summarize the relationships between outcomes. These correlation circles (displayed in Figure 3) reveal very unexpected results, namely, a very similar behavior for posture, lameness at walk and lameness at trot, while pain at palpation appeared to behave very differently. The results even suggest the independence of pain from the other outcomes, as what was measured actually looks more like two sub-scores with very different weights, at 3:1. Such findings could not have been deduced from the marginal analyses (Model 1), which rather show similarities between the outcomes regarding treatment effect parameters. In short, the marginal analyses do not document the intrinsic correlations between the ordinal responses, and these intrinsic correlations cannot be read directly from contingency tables.

As for Model 1, simulations were performed with Models 2 and 3 to assess their ability to predict joint probabilities, such as the probability that all outcomes equal zero

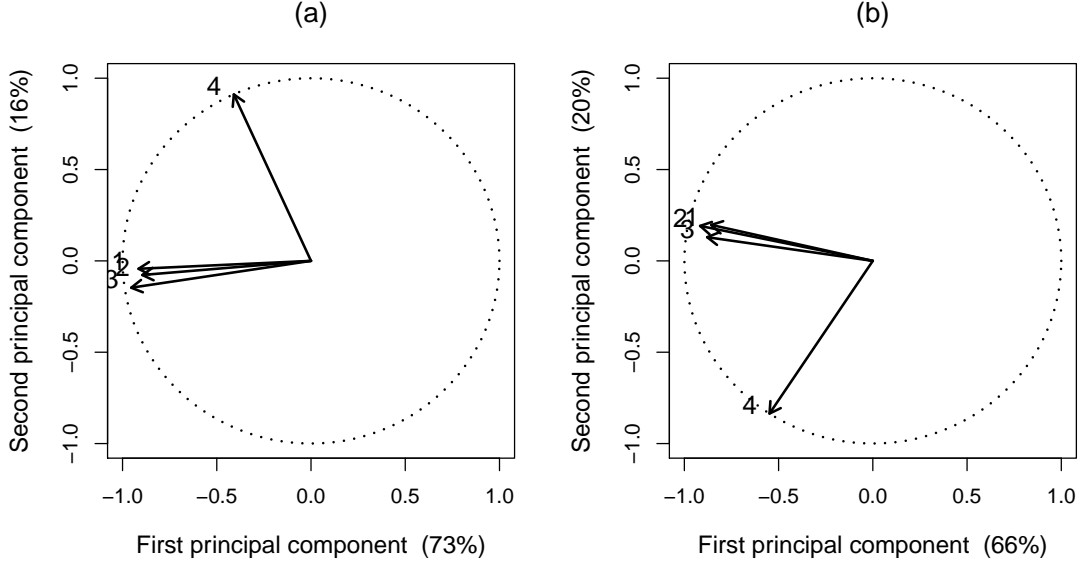


Figure 3: Results (correlation circle for the first and second principal components) of a principal components analysis on (a) the 4×4 variance matrix of the subject-specific random effects Ω and (b) the 4×4 correlation matrix of the residual random effects Σ (1, posture of the dog at a stand; 2, lameness at walk; 3, lameness at trot; 4, pain at palpation).

(meaning that the dog shows no symptoms) or the probability that all outcomes are below 1 (meaning that the dog shows no or mild symptoms). Such probabilities are very relevant from a clinical point of view. As before, $M = 500$ simulations of the dataset were generated to compute the median and 95% confidence interval for (i) the marginal probability to belong to a single category at a given time point, (ii) the joint probability that all outcomes equal zero, and (iii) the joint probability that all outcomes are below 1. As could be expected, very good predictions were obtained for the marginal distributions, with plots very similar to Figure 2. However, the models were not equivalent with respect to the prediction of joint probabilities. Major departures from the data were seen for Model 1 (Figure 4), and although Model 2 performed much better than Model 1, some departures from the data were also observed. The model that best predicted the joint data was Model 3, in line with AIC values, which is an argument in favor of a joint approach for estimating all correlations. The plots in Figure 4 indicate that clinical symptoms were suppressed in approximately 25% of the dogs and that approximately 80% of the dogs

had no or mild symptoms when the asymptote effect was reached.

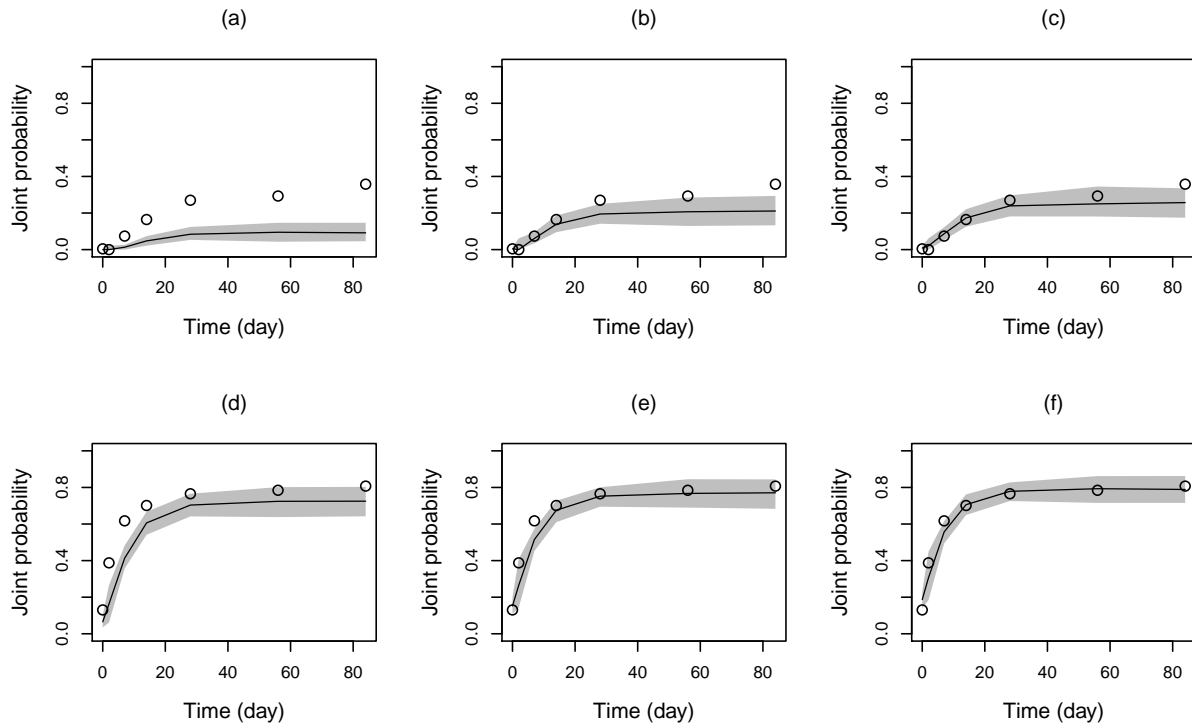


Figure 4: The upper panel shows the time-course of the joint probability that all outcomes are equal to zero (no symptoms) under (a) Model 1, assuming the independence between outcomes; under (b) Model 2, assuming the conditional independence between outcomes, given the subject-specific random effects; and under (c) Model 3, estimating all correlations. The lower panel shows the time-course of the joint probability that all outcomes are below 1 (no or mild symptoms) under (d) Model 1, (e) Model 2 and (f) Model 3. Dots represent the empirical probabilities computed from the observations. They are compared to the median probability (solid line) and 95% confidence interval (shaded area) computed from simulations under the tested model at each time point.

5 Monte Carlo investigation of performance

In this section, a simulation study was conducted to evaluate the performance of the proposed methodology. We used the final model (Model 3) that had been developed for the robenacoxib data in Section 4 for simulation and followed the design of study 1, which had the longest duration (see Section 4 for more details). As in the original robenacoxib dataset, four ordinal outcomes (posture of the dog, lameness at walk, lameness at trot,

pain at palpation) were simulated at the different visits for 232 subjects. Posture, lameness at walk and lameness at trot were graded on 3-level scales, whereas pain at palpation was graded on a 4-level scale. A total of 100 Monte Carlo datasets were generated and analyzed with Model 3 using the pairwise approach in combination with the stochastic EM algorithm. Reasonable values were used for the initial estimates, although these values were different from the true values to test the ability of the proposed method to identify proper estimates: $(-0.5; 0.5)$ for $\mathbf{a}^{(k)}$ when $k = 1, \dots, 3$, $(-0.5; 0; 0.5)$ for $\mathbf{a}^{(k)}$ when $k = 4$, 1 for all components in $\boldsymbol{\beta}$ and variances in $\boldsymbol{\Omega}$, and zero for covariances in $\boldsymbol{\Omega}$ and correlations in $\boldsymbol{\Sigma}$. The total number of iterations used for estimation was fixed at 200 for all samples. This value was checked in randomly chosen datasets so that these values were sufficiently large to ensure that convergence was reached in most cases. Additionally, to avoid lengthy computation times, the size of the Monte Carlo simulations for the computation of standard errors was set to 1000 in the simulation study, although it was equal to 5000 in the analysis of the robenacoxib data in Section 4. Under these conditions, the computation time was only 2.8 ± 0.06 hours, which is very reasonable given the size of the dataset (232 subjects, 1392 observations, 4 outcomes) and the complexity of the model. This nevertheless hampered intensive simulations and explains why only 100 Monte Carlo datasets were analyzed.

The results of the simulation study are summarized in Table 3. The pairwise estimator did not show any bias at this sample size, and the standard errors of the estimator were consistent with the standard deviations of the sampling distribution. It must be kept in mind, however, that the blocks in the \mathbf{J} and \mathbf{K} matrices used to calculate the variance-covariance matrix of the estimator (equation (3.2)) were computed by dropping the expectations and plugging in the final model parameter estimates. Although it has been shown that this plug-in approach leads to satisfactory results (Louis, 1982), this could explain the small differences seen between the Monte Carlo standard deviations and the means of the estimated standard errors derived from the pairwise approach.

Parameters	True value	MC mean	MC SD	SE mean
Cut-points				
$a_1^{(1)}$	-1.379	-1.385	0.166	0.167
$a_2^{(1)}$	1.513	1.508	0.141	0.177
$a_1^{(2)}$	-2.029	-2.028	0.152	0.176
$a_2^{(2)}$	0.345	0.335	0.120	0.163
$a_1^{(3)}$	-2.398	-2.400	0.172	0.161
$a_2^{(3)}$	-0.006	-0.012	0.141	0.114
$a_1^{(4)}$	-2.786	-2.814	0.141	0.165
$a_2^{(4)}$	-0.747	-0.773	0.116	0.147
$a_3^{(4)}$	0.916	0.894	0.138	0.153
Treatment effect				
$\beta_{max}^{(1)}$	2.534	2.538	0.168	0.148
$\beta_{max}^{(2)}$	2.376	2.376	0.116	0.135
$\beta_{max}^{(3)}$	2.505	2.500	0.143	0.135
$\beta_{max}^{(4)}$	2.639	2.661	0.114	0.122
β_{eq}	0.128	0.129	0.008	0.009
Random effects, variance terms Ω				
ω_{11}	2.811	2.834	0.442	0.473
ω_{22}	2.235	2.195	0.293	0.339
ω_{33}	2.455	2.465	0.333	0.370
ω_{44}	1.634	1.666	0.190	0.252
ω_{12}	1.715	1.693	0.273	0.324
ω_{13}	2.169	2.193	0.302	0.365
ω_{14}	0.685	0.696	0.165	0.252
ω_{23}	1.991	1.975	0.275	0.305
ω_{24}	0.580	0.605	0.163	0.275
ω_{34}	0.535	0.548	0.162	0.263
Random effects, correlation terms Σ				
ρ_{12}	0.746	0.754	0.042	0.039
ρ_{13}	0.630	0.635	0.047	0.047
ρ_{14}	0.318	0.323	0.052	0.055
ρ_{23}	0.776	0.785	0.033	0.031
ρ_{24}	0.344	0.348	0.043	0.053
ρ_{34}	0.359	0.360	0.052	0.052

Table 3: Results of the simulation study with the Monte Carlo sampling distribution of parameter estimates (from 100 simulated samples of size $N = 232$ subjects and a total of 1392 observations). MC mean and MC SD are the Monte Carlo mean and standard deviation. SE mean is the mean of the estimated standard errors for the parameter estimate.

6 Discussion

The objective of the paper was to analyze robenacoxib efficacy data consisting of four ordinal outcomes measured repeatedly over time in 232 osteoarthritic dogs. In particular, we wanted to evaluate the clinical improvement of dogs under robenacoxib treatment and to increase our understanding of the scoring system used by identifying possible redundancies between outcomes. Clearly, some of the questions addressed in this paper could have been treated by reducing the information to a single dichotomous variable. However, different clinical criteria (cure/no cure; improvement/no improvement; etc.) would require different dichotomizations, and any information on more gradual responses would be lost. The proposed model allows us to retain maximal information while answering several questions simultaneously.

Models with increasing complexity were tested to analyze the four ordinal outcomes simultaneously over time. Model 1 assumed the independence between outcomes; Model 2 assumed the independence between outcomes conditionally on subject-specific random effects \mathbf{b}_i ; and Model 3 assumed the full correlation of \mathbf{b}_i components and of \mathbf{e}_{ij} components. Although Models 2 and 3 obviously performed much better than Model 1, Model 3 did not show much improvement over Model 2 with respect to the percentage of subjects with no symptoms, despite a large difference in AIC. When assessing other, secondary criteria (such as the time before all scores are improved by at least one grade), a difference in fit could be seen between Model 2 and Model 3, supporting that Model 3 was better than Model 2, in agreement with the AIC values.

Overall, Model 3 provided a very good description of the data and showed that all symptoms of pain and inflammation were suppressed in 1 dog out of 4 after 4 weeks of treatment (4 dogs out of 5 having no or mild symptoms). By comparison, Model 1, assuming the independence between outcomes, largely underestimated the efficacy of the treatment, predicting no symptoms in only 1 dog out of 20. Such results raise a number of concerns regarding the interpretation of univariate analyses performed on each outcome separately. Additional information provided by Model 3 was the evaluation of

possible redundancies between outcomes. In the present case, three of the four outcomes were strongly correlated (posture, lameness at walk, lameness at trot); thus, what was measured looks more like two sub-scores with very different weights, at 3:1. Thus, the standard method analyzing the sum of outcomes would have counted the same information three times.

Because the pain response appeared to be largely independent of the highly correlated posture and lameness measures, the next step would be to conduct a factor analysis to validate this observation. In additional work (not shown), we performed a factor analysis with two independent factors, as suggested by the results of the present multivariate analysis. We obtained similar likelihood and goodness-of-fit plots as for the reference model with the four factors (i.e., four latent variables), thereby confirming the previous findings. This illustrates how the two approaches can be applied in a stepwise manner and are complementary.

Multiple ordinal responses are a common feature in clinical trials. Many diseases involve complex pathophysiological processes, and this complexity generally cannot be described using a single ordinal outcome. Numerous examples of multivariate ordinal outcomes can be found in the literature, e.g., the Glasgow Coma Scale for head trauma (3 dimensions), the Positive/Negative Syndrome Scale in schizophrenia (7 dimensions each) and the Hamilton Rating Scale for depression (21 dimensions). Although computational performance has greatly improved during the last decade, rendering multivariate analysis more tractable, computation time is a limiting factor for the use of these models in routine practice. This is the reason why we saved time at each computational step, using a stochastic EM algorithm for estimation with low values of M_1 , M_2 and Z (see Section 3). A main advantage of this algorithm is that the estimation of subject-specific random effect variance components is obtained directly by sampling from the posterior distribution without requiring any optimization process. This is a major gain in computation time when the dimension increases.

Indubitably, the largest gain in computation time comes from the use of the pairwise approach. It is clear that a joint model in 21 dimensions (Hamilton Rating Scale) is

intractable with classical maximum likelihood methods; therefore, the pairwise approach is very appealing. As an indication, a single evaluation of the likelihood in 4 dimensions took approximately 9 hours in our example, whereas the estimation of parameters, pseudo-likelihood and standard errors for the same data took 2.5 hours. It is noteworthy that the computation of the variance matrix of the estimator is the most time-demanding step of the estimation process (30-60% of the total time, depending on the desired precision). The counterpart of the pairwise approach is a possible efficiency loss because we do not compute the maximum likelihood. From what we see in the literature, it appears that this efficiency loss is rather low (Geys et al., 1999; Fieuws and Verbeke, 2006; Fieuws et al., 2006). In our example with 4 dimensions, we were not able to compute the maximum likelihood variance matrix in reasonable time. However, the standard errors of the parameter estimates obtained with the pairwise approach were of acceptable sizes.

Our strategy was to use a probit mixed effects model with a latent variable interpretation. Although there is no requirement that the latent variable exists (this can be viewed as an abstraction to motivate the model), the latent variable concept is very appealing in the field of clinical pharmacology and facilitates the interpretation of results from a biological background. In particular, many of the classical models developed from quantitative measurements of drug effect can be applied to continuous latent variables. It is clear that the probit model offers less flexibility than other (logistic) models. However, this approach worked well on our data, requiring only a limited number of parameters to be estimated. Extensions of the model can be considered to accommodate more complex situations. For example, the correlations in Σ might depend on covariates (see Todem et al., 2007, for an illustration in the bivariate case), or serial correlations within subjects can be modeled (Li and Schafer, 2008) (in the present work, the residual random effects were considered independent across time). Additionally, the proper handling of missing data would be useful. Finally, we used simple mixed effects models for the latent variables that were linear in the subject-specific random effects. Alternatively, non-linear models could be used in line with typical models in clinical pharmacology.

7 Acknowledgements

We would like to thank the Editor, the Associate Editor and the two anonymous Referees for their constructive comments and suggestions that greatly improved the quality of the paper.

References

- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236.
- Booth, J. G. and Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated monte carlo em algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(1):265–285.
- Chib, S. and Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, 85(2):347–361.
- Dale, J. (1986). Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics*, pages 909–917.
- Delyon, B., Lavielle, M., and Moulines, E. (1999). Convergence of a stochastic approximation version of the em algorithm. *Annals of Statistics*, pages 94–128.
- Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38.
- Drezner, Z. and Wesolowsky, G. (1990). On the computation of the bivariate normal integral. *Journal of Statistical Computation and Simulation*, 35(1-2):101–107.
- Dunson, D. B. (2003). Dynamic latent trait models for multidimensional longitudinal data. *Journal of the American Statistical Association*, 98(463):555–563.

- Fieuws, S. and Verbeke, G. (2006). Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics*, 62(2):424–431.
- Fieuws, S., Verbeke, G., Boen, F., and Delecluse, C. (2006). High dimensional multivariate mixed models for binary questionnaire data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55(4):449–460.
- Genz, A. and Bretz, F. (2009). *Computation of multivariate normal and t probabilities*. Springer Publishing Company, Incorporated.
- Geys, H., Molenberghs, G., and Ryan, L. (1999). Pseudolikelihood modeling of multivariate outcomes in developmental toxicology. *Journal of the American Statistical Association*, pages 734–745.
- Glonek, G. and McCullagh, P. (1995). Multivariate logistic models. *Journal of the royal statistical society. Series B (Methodological)*, pages 533–546.
- Hastings, W. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.
- Katsikatsou, M., Moustaki, I., Yang-Wallentin, F., and Jöreskog, K. G. (2012). Pairwise likelihood estimation for factor analysis models with ordinal data. *Computational Statistics & Data Analysis*.
- Kuhn, E. and Lavielle, M. (2004). Coupling a stochastic approximation version of em with an mcmc procedure. *ESAIM: Probability and Statistics*, 8:115–131.
- Kuhn, E. and Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Computational statistics & data analysis*, 49(4):1020–1038.
- Lee, S.-Y. and Song, X.-Y. (2004). Maximum likelihood analysis of a general latent variable model with hierarchically mixed data. *Biometrics*, 60(3):624–636.

- Li, Y. and Schafer, D. (2008). Likelihood analysis of the multivariate ordinal probit regression model for repeated ordinal responses. *Computational Statistics & Data Analysis*, 52(7):3474–3492.
- Liu, I. and Agresti, A. (2005). The analysis of ordered categorical data: An overview and a survey of recent developments. *Test*, 14(1):1–73.
- Liu, L. and Hedeker, D. (2006). A mixed-effects regression model for longitudinal multivariate ordinal data. *Biometrics*, 62(1):261–268.
- Louis, T. (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 226–233.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21:1087.
- Molenberghs, G. and Lesaffre, E. (1994). Marginal modeling of correlated ordinal data using a multivariate plackett distribution. *Journal of the American Statistical Association*, pages 633–644.
- Molenberghs, G. and Lesaffre, E. (1999). Marginal modelling of multivariate categorical data. *Statistics in Medicine*, 18(17-18):2237–55.
- Molenberghs, G. and Verbeke, G. (2005). *Models for discrete longitudinal data*. Springer.
- Qaqish, B. and Ivanova, A. (2006). Multivariate logistic models. *Biometrika*, 93(4):1011–1017.
- Reymond, N., Speranza, C., Gruet, P., Seewald, W., and King, J. (2012). Robenacoxib vs. carprofen for the treatment of canine osteoarthritis; a randomized, noninferiority clinical trial. *Journal of veterinary pharmacology and therapeutics*, 35(2):175–183.
- Sammel, M., Lin, X., and Ryan, L. (1999). Multivariate linear mixed models for multiple outcomes. *Statistics in medicine*, 18(17-18):2479–2492.

- Sammel, M. D., Ryan, L. M., and Legler, J. M. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(3):667–678.
- Shi, J.-Q. and Lee, S.-Y. (2000). Latent variable models with mixed continuous and polytomous data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1):77–87.
- Teixeira-Pinto, A. and Normand, S.-L. T. (2009). Correlated bivariate continuous and binary outcomes: issues and applications. *Statistics in medicine*, 28(13):1753–1773.
- Ten Have, T. and Morabia, A. (1999). Mixed effects models with bivariate and univariate association parameters for longitudinal bivariate binary response data. *Biometrics*, 55(1):85–93.
- Todem, D., Kim, K., and Lesaffre, E. (2007). Latent-variable models for longitudinal data with bivariate ordinal outcomes. *Statistics in medicine*, 26(5):1034–1054.
- Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42.