OXFORD

A joint normal-ordinal (probit) model for ordinal and continuous longitudinal data

Margaux Delporte[1,*], Geert Molenberghs[1,2], Steffen Fieuws[1],
Geert Verbeke[1,2]

[1]Department of Public Health & Primary Care, Leuven Biostatistics and Statistical Bioinformatics Centre,
Kapucijnenvoer 7 - box 7001, 3000 Leuven, Belgium
[2]Data Science Institute, Interuniversity Biostatistics and Statistical Bioinformatics Centre, Agoralaan Gebouw
D - B -3590 Diepenbeek, Belgium

*Corresponding author: Leuven Biostatistics and Statistical Bioinformatics Centre, Kapucijnenvoer 7 - box 7001, 3000
Leuven, Belgium. Email: margaux.delporte@kuleuven.be

SUMMARY

In biomedical studies, continuous and ordinal longitudinal variables are frequently encountered. In many of these studies it is of interest to estimate the effect of one of these longitudinal variables on the other. Time-dependent covariates have, however, several limitations; they can, for example, not be included when the data is not collected at fixed intervals. The issues can be circumvented by implementing joint models, where two or more longitudinal variables are treated as a response and modeled with a correlated random effect. Next, by conditioning on these response(s), we can study the effect of one or more longitudinal variables on another. We propose a normal-ordinal(probit) joint model. First, we derive closed-form formulas to estimate the model-based correlations between the responses on their original scale. In addition, we derive the marginal model, where the interpretation is no longer conditional on the random effects. As a consequence, we can make predictions for a subvector of one response conditional on the other response and potentially a subvector of the history of the response. Next, we extend the approach to a high-dimensional case with more than two ordinal and/or continuous longitudinal variables. The methodology is applied to a case study where, among others, a longitudinal ordinal response is predicted with a longitudinal continuous variable.

KEYWORDS: joint model; longitudinal data analysis; probit link; random effects model; time-dependent effects.

1. INTRODUCTION

In many clinical studies the same patients are repeatedly examined, which results in longitudinal data. This data can be analyzed with generalized linear mixed models. This family of models encompasses linear mixed models, which was introduced by Laird and Ware (1982). Later the approach was extended to noncontinuous data by Breslow and Clayton (1993), Wolfinger and O'Connell (1993) and Engel and Keen (1994). In these models, the correlation induced by the repeated measurements is captured with random effects. These effects explicitly model the variation between the subjects.

It can be of interest to model the association between multiple longitudinal responses. A first option is to treat one of the responses as a predictor, and use it as a time-dependent covariate.

However, this method has several pitfalls. First, the lag has to be correctly specified, as incorrect specification can lead to illogical results. An example given by Rizopoulos (2012) is a study where they found a positive, but insignificant, effect of smoking on the survival of patients with coronary artery disease (Cavender et al., 1992). However, there was no lagged effect in the model, and hence only the immediate effect of smoking on death was gauged. The explanation of the faulty conclusion is that most of the smokers had stopped smoking at the last time of follow up before their death. In the meantime, many patients that were still alive, were still smoking. A second drawback is the classification of a covariate into an exogenous or endogenous time-dependent covariate. If a response at time t predicts the value of the covariate at a time $s > t$, the covariate is endogenous (Diggle, 2002). Endogenous covariates have important modeling implications. Qian et al. (2020) have, for example, found that the marginal interpretation of the parameters in a linear mixed model does not hold anymore. Third, attention has to be given to missing data, which will likely occur in patient studies with follow-up. While ignorability holds for missing data of the response values under MAR under direct likelihood (Rubin, 1976), this is not the case for missing covariate values. Fourth, the time-dependent covariate can possibly be an intermediate variable. This means that the time-dependent covariate is in the causal pathway between another covariate and the response. As a consequence, including the time-dependent covariate will make the effect of the covariate on the response disappear. Fifth, when the responses, as well as time-dependent covariates, are not collected at fixed intervals, the utilization of time-dependent covariates with lags is not possible.

Joint models provide an alternative to time-dependent covariates. Several approaches for jointly modeling responses of a mixed nature exist. An overview can be found in Molenberghs and Verbeke (2005) in their Chapter 24. They outline three approaches that are applicable in both hierarchical and non-hierarchical settings. A first approach employs a bivariate Plackett-Dale distribution and postulates the existence of an unobserved continuous response that underlies the observed binary/ordinal response. The second approach, known as the probit-normal formulation, also assumes the presence of a latent response, with the added feature of errors being correlated to the continuous response. The third approach is the generalized linear random-effects model, which we will describe here in greater detail for the longitudinal case.

In joint generalized linear random-effects models, the relation between the responses is symmetric. Here, all longitudinal variables are treated as responses and are modeled with an appropriate random effects model. The random effects of the different models are allowed to be correlated to capture the associations. One of the advantages is that the effect of a covariate can be assessed on multiple outcomes simultaneously and that the association between the responses as well as the evolution of this association can be assessed. For example, Chakraborty et al. (2003) fitted a joint model for the continuous HIV-1 RNA concentration in both blood and semen. With the joint model, he could compare the correlation between both responses between the group with and without HIV treatment. Notably, the use of joint random-effects models is not limited to continuous responses. In Delporte et al. (2022) a joint model was developed for a longitudinal continuous and a longitudinal binary response. Not only the latent correlations between the random effects were scrutinized, but also the correlations between the responses on their original scale could be gauged. They derived a closed-form formula for the correlation function from the joint model, with the possibility to include covariates. Their case study focused on the relation between the occurrence of allergic bronchopulminary aspergilosis (ABPA) and FEV values. Based on the latent correlations, they found that a better FEV value than expected under the model resulted in higher probabilities of lung infection at baseline and that higher increase in lung function than expected under the model is positively related to a higher probability of absence of ABPA than expected at baseline on the probit scale. Still, when gauging the correlations on the original scale, the conclusions were far more clinically relevant. We found that the correlation, between the responses as observed, is slightly stronger for earlier measurements of the ABPA and later measurements of FEV. This suggests that ABPA at an early stage shows an overall frailty, which exhibits itself later in life. In addition, they proposed a prediction model where one of the responses and potentially the history of the predicted response are included as predictors in the model.

Table 1. Number of measurements of the Mini Mental State Exam (MMSE) and Activities of Daily Living (ADL) at each time point.

| Response	Day 1	Day 3	Day 5	Day 8	Day 12
MMSE	59	58	60	52	38
ADL	60	0	60	0	40

The discrepancy between the manifest and latent correlation in random-effects models is also discussed in several other papers. For example, Milanzi et al. (2015) caution against drawing misleading conclusions by using latent and manifest-based correlation reliability measures interchangeably in IRT models. They emphasize that latent correlation-based reliability measures consistently result in higher values than their manifest correlation-based counterparts. Moreover, Molenberghs and Verbeke (2005) compare in their Chapter 7 the associations found via the Bahadur, probit and Dale models. They found a strong downward bias in the marginal correlation estimates obtained from the Bahadur model in comparison to their probit model counterparts. Lastly, Fieuws et al. (2006) use joint random-effects models for analyzing binary questionnaire data. They stress that their interest is in the association between the (latent) concepts underlying the sets of items, in contrast to the the association between observed responses, for which they recommend other models.

Some work has been done on the joint model for a continuous and a ordinal response. Faes et al. (2004) used a Plackett-Dale approach to jointly model the birth weight (continuous) and the probabilities of degrees of malformation (ordinal) of a fetus, where they take into account the clustering induced by a common mother. Still, the model cannot be readily extended to a longitudinal setting where responses are measured at different time points. Ivanova et al. (2016) formulated a joint random-effects model in a case study of repeated measures of BMI and clinical targets of diabetes patients. "Clinical targets" was treated as an ordinal variable. The covariance between the random intercepts of the variables was examined in order to gauge the association between the responses. In this paper, we extend the approach of Ivanova et al. (2016) by deriving closed-form formulas to calculate the correlations between the responses on their original scale. In addition, a conditional model is derived in order to construct predictions of one response conditional on the other response(s). The outline of the paper is as follows: Section 2 presents the case study that serves as the foundation for the subsequent analysis in Section 5. Section 3 discusses the methodology. It commences with a review of the established methods for clustered continuous and clustered ordinal responses. Following that, we introduce the normal-ordinal (probit) model and our methodology based on the joint model. In Section 6 concluding remarks are offered.

2. CASE STUDY

The dataset contains information about the occurrence and progression of cognitive impairment in 60 elderly hip fracture patients from admission to the twelfth postoperative day (Milisen et al., 1998).We will focus on the connection between cognitive abilities and functional status and how the association between both varies over time. Throughout the study, neurocognitive status and the functional performance were assessed longitudinally; neurocognitive status was measured at day 1, 3, 5, 8, and 12, while functional status was recorded at day 1, 5, and 12. Table 1 provides an overview of the number of measurements taken of each response at each time point. Drop-out occurred because patients were discharged from the hospital before the twelfth post-operative day. Notably, while deaths were recorded, there were no reported mortalities throughout the duration of the study.

Neurocognitive status was assessed using the mini-mental state exam (MMSE), which includes subscales for memory, linguistic ability, concentration, and psychomotor executive skills. Cognitive status was classified as no impairment (MMSE≥24), moderate impairment ($18 \leq$ MMSE≤ 23), or
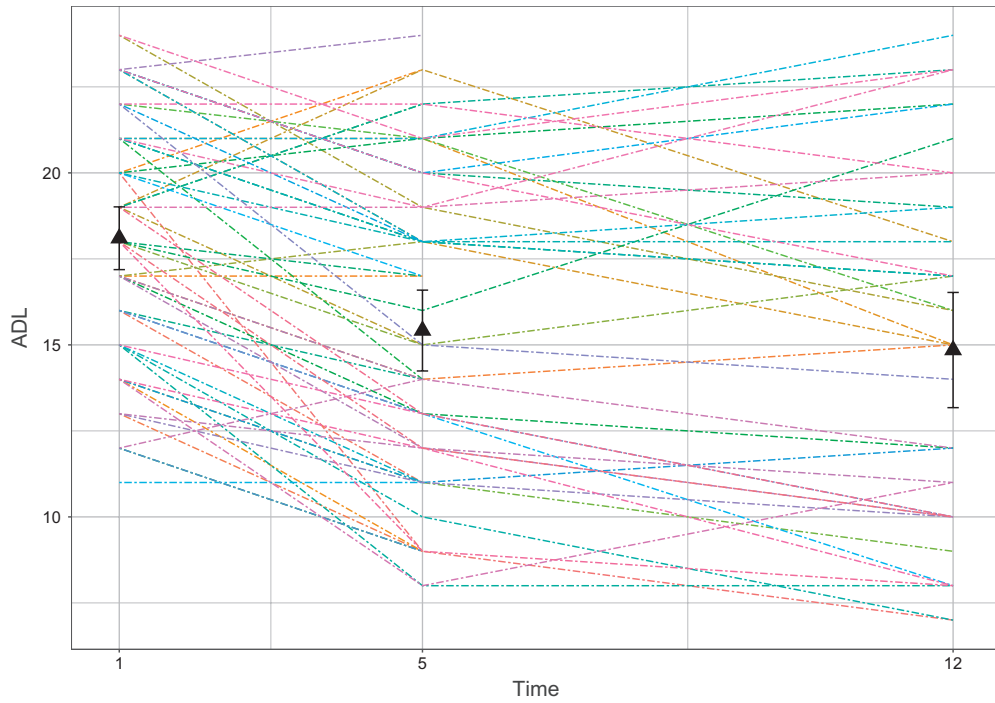
Figure 1. Observed average (with 95% confidence interval) of the activities of daily living scores on day 1, 5 and 12 (solid) and individual profiles of the 60 subjects (dashed).

severe impairment ($\text{MMSE} \leq 17$) (Milisen et al., 1998; Tombaugh and McIntyre, 1992). In addition, the functional status was measured using an adapted version of the Katz ADL-scale (ADL), which is treated as continuous. The mean ADL scores and individual profiles are presented at Fig. 1. A higher ADL value indicates more dependence on caretakers for activities of daily living, whereas a higher category of MMSE indicates a lower level of impairment. For exploratory purposes, the point-biserial correlations between the observed responses at several time points has been calculated (see Appendix H). It suggests that there exists a moderately strong relation between ADL and both the event of having severe impairment and the event of having impairment. This correlation seems to slightly increase over time. However, these correlations are not corrected for covariates and are only valid when the data would be missing completely at random, which is a very strict assumption.

3. METHODOLOGY
3.1. Model for a single longitudinal continuous response

One of the most popular models for longitudinal continuous variables is the linear mixed model. Suppose we have N subjects and the jth measurement for subject i is denoted by Y_{ij}. The vector $(Y_{i1}, ..Y_{in_i})$ of all n_i measurements of subject i is denoted by $\boldsymbol{Y_i}$. With this notation, we can write the model as

$$\boldsymbol{Y_i}|\boldsymbol{b_i} \quad \sim \quad N(\boldsymbol{X_i}\boldsymbol{\beta} + \boldsymbol{Z_i}\boldsymbol{b_i}, \boldsymbol{\Sigma_i}), \qquad (3.1)$$
$$\boldsymbol{b_i} \quad \sim \quad N(\boldsymbol{0}, \boldsymbol{D}),$$

where $\boldsymbol{X_i}$ and $\boldsymbol{Z_i}$ are, respectively, $(n_i \times k)$ and $(n_i \times q)$ dimensional matrices of known covariates of, respectively, the fixed effects $\boldsymbol{\beta}$ and the random effects $\boldsymbol{b_i}$. $\boldsymbol{\Sigma_i}$ denotes the $(n_i \times n_i)$ dimensional covariance matrix. Notably, i does not mean that the estimates of the variance depends

on the subject. It indicates that the dimensions of the residual matrix can depend on the subject (Verbeke and Molenberghs, 2000). We can simplify $\boldsymbol{\Sigma}_i$ to $\sigma^2 \boldsymbol{I}_i$ with the assumption that the random effects fully capture the correlation between the measurements within subjects. This conditional independence assumption is however not necessary. It can be relaxed by the inclusion of, for example, serial correlation.

A property of the linear mixed model is that the parameters of the conditional model and the marginal model are exactly equal. This holds since $E[\boldsymbol{Y}_{ij}] = E[E(\boldsymbol{Y}_{ij}|\boldsymbol{b}_i)] = \boldsymbol{x}'_{ij}\boldsymbol{\beta}$. Still, the marginal model is defined as

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i^*. \tag{3.2}$$

The residuals $\boldsymbol{\epsilon}_i^*$ are here by definition correlated and are normally distributed around $\boldsymbol{0}$ with variance $\boldsymbol{V}_i^*$. As a consequence, the distribution of the response is

$$\boldsymbol{Y}_i \sim N(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{V}_i^*), \tag{3.3}$$

with $\boldsymbol{V}_i^* = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}'_i + \boldsymbol{\Sigma}_i$. More information about linear mixed models can be found in Verbeke and Molenberghs (2000).

3.2. Model for a single longitudinal ordinal response

A random-effects ordinal regression model can be used for clustered or repeated measures of an ordinal response. A threshold concept is applied, which assumes that the observed ordered response categories are determined by the value of an underlying continuous response. A series of threshold values $\gamma_1, \gamma_2, ..\gamma_{d-1}$ are assumed for the d categories. A response is categorized as category c if the latent response Y_{ik}^* surpasses the threshold value γ_{c-1}, but not γ_c. For the measurement at time k of this latent response of subject i, $k = 1, ..., p_i$,

$$Y_{ik}^* = \boldsymbol{x}'_{ik}\boldsymbol{\beta} + \boldsymbol{z}'_{ik}\boldsymbol{b}_i + \epsilon_{ik},$$

is the hierarchical linear mixed model, where $\boldsymbol{x}_{ik}$ is the $r \times 1$ vector that contains values for the covariates of the r-dimensional fixed effects vector $\boldsymbol{\beta}$. Next, $\boldsymbol{z}_{ik}$ is the $q \times 1$ design vector for the q random effects $\boldsymbol{b}_i$. $\boldsymbol{b}_i$ is assumed to follow a normal distribution around 0 with the covariance matrix $\boldsymbol{D}$. ϵ_{ik} are the residuals and are assumed to be independently normally distributed with mean 0 and variance σ^2.

From the latter model for Y_{ik}^*, the probabilities of the response categories can be derived. The probability that a response at time k for subject i falls into category c equals

$$P(Y_{ik} = c) = \Phi\left(\frac{\gamma_c - \zeta_{ik}}{\sigma}\right) - \Phi\left(\frac{\gamma_{c-1} - \zeta_{ik}}{\sigma}\right),$$

where $\zeta_{ik} = \boldsymbol{x}'_{ik}\boldsymbol{\beta} + \boldsymbol{z}'_{ik}\boldsymbol{b}_i$ and $\Phi(.)$ equals the cumulative normal distribution. Similarly, the probability that a response k of subject i is less than or equal to category c equals

$$P(Y_{ik} \leq c) = \Phi\left(\frac{\gamma_c - \zeta_{ik}}{\sigma}\right).$$

The choice of the unit and the origin of ζ is arbitrary (Hedeker and Gibbons, 1994). Alternatively, the logit link function can be applied (Ivanova et al., 2016), but this leads to more cumbersome calculations and less closed forms can be derived than with the probit link.

In Supplementary Appendix A the marginal random-effects ordinal regression model is derived. In the latter model, the interpretation is no longer conditional on the random effects. Let $\boldsymbol{Z}_i$ denote

the $n_i \times q$ dimensional design matrix of the random effects and X_i denote the $n_i \times r$ dimensional design matrix of the fixed effects. The marginal model is the following:

$$P(\mathbf{y}_i \leq \mathbf{c}) = \Phi(\boldsymbol{\gamma}_c - X_i\boldsymbol{\beta}; L_i^{-1}), \tag{3.4}$$

where

$$L_i = I - Z_i(D^{-1} + Z_i'Z_i)^{-1}Z_i'.$$

3.3. Joint model

The joint mixed model employs a q-dimensional random effects vector $\boldsymbol{\xi}_i$ to encompass random effects linked the ordinal response as well as random effects linked with the continuous response. This vector follows a multivariate normal distribution with a mean of zero and a covariance matrix D. This matrix D accounts for the correlation between repeated measurements of the same response, as well as the correlations between (the vectors of) measurements for different responses. We assume that the responses are independent given the random effects, meaning that the random effects fully capture the correlation between the responses. Consequently, the joint density of the responses, given the random effects, is equivalent to the product of the conditional densities of the individual responses.

The joint marginal density can be obtained by integrating out the random effects out of the joint density, these calculations can be found in Supplementary Appendix B. Note that the primary purpose of this joint marginal density is to provide an intermediate result for future calculations. The joint marginal density is as follows:

$$f(\mathbf{y}_{1i}, \mathbf{y}_{2i} \leq \mathbf{c}) = \phi(X_{1i}\boldsymbol{\beta}; V_i)\Phi(\boldsymbol{\gamma}_c - X_{2i}\boldsymbol{\beta} - \boldsymbol{\alpha}_i; B_i), \tag{3.5}$$

where

$$
\begin{aligned}
V_i &= Z_{1i}DZ_{1i}' + \Sigma_i, \\
\boldsymbol{\alpha}_i &= H_i(\mathbf{y}_{1i} - X_{1i}\boldsymbol{\beta}), \\
H_i &= B_iZ_{2i}K_iZ_{1i}'\Sigma_i^{-1}, \\
K_i^{-1} &= D^{-1} + Z_{1i}'\Sigma_i^{-1}Z_{1i} + Z_{2i}'Z_{2i}, \\
B_i^{-1} &= I - Z_{2i}K_iZ_{2i}'.
\end{aligned}
$$

It is possible to extend (3.5) to the high-dimensional case, with multiple ordinal and/or continuous responses. Let Y_{ci} represent a vector containing all the measurements of a continuous responses: $Y_{ci} = (Y_{1i1}^c, \ldots, Y_{1in_{1i}}^c, Y_{2i1}^c, \ldots, Y_{2in_{2i}}^c, \ldots, Y_{ai1}^c, \ldots, Y_{ain_{ai}}^c)$. Notably, the number of measurements for each response does not have to be the same. Similarly, let Y_{bi} denote a vector containing all the measurements of the o ordinal responses: $Y_{bi} = (Y_{1i1}^b, \ldots, Y_{1ip_{1i}}^b, Y_{2i1}^b, \ldots, Y_{2ip_{2i}}^b, \ldots, Y_{oi1}^b, \ldots, Y_{op_{oi}}^b)$. Additionally, the matrices Z_{ci} and Z_{bi} consist of concatenated matrices of covariates for the random effects of continuous and ordinal responses, respectively. Specifically, Z_{ci} is formed by combining the matrices of covariates for the separate continuous responses: $Z_{ci} = [Z_{1i}^c, Z_{2i}^c, \ldots, Z_{ai}^c]'$. Similarly, Z_{bi} is formed by concatenating the matrices of covariates for the separate ordinal responses: $Z_{bi} = [Z_{1i}^b, Z_{2i}^b, \ldots, Z_{oi}^b]'$. The matrices of covariates for the fixed effects are defined in a similar manner: X_{ci} contains the concatenated matrices of covariates for the continuous responses: $[X_{1i}^c, X_{2i}^c, \ldots, X_{ai}^c]'$, while X_{bi} contains the concatenated matrices of

covariates for the ordinal responses: $[\boldsymbol{X}_{1i}^b, \boldsymbol{X}_{2i}^b, \ldots, \boldsymbol{X}_{oi}^b]'$. Further, $\boldsymbol{\Sigma}_i$ is a block diagonal matrix with as blocks the variance-covariance matrices of the continuous responses

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} \boldsymbol{\Sigma}_{1i} & 0 & \ldots & 0 \\ 0 & \boldsymbol{\Sigma}_{2i} & .. & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \boldsymbol{\Sigma}_{ai} \end{bmatrix}.$$

It is easy to see that the marginal hierarchical model is now

$$f(\boldsymbol{y}_{ci}, \boldsymbol{y}_{bi} \le \boldsymbol{c}) = \phi(\boldsymbol{X}_{ci}\boldsymbol{\beta}; \boldsymbol{V}_i)\Phi(\boldsymbol{\gamma}_c - \boldsymbol{X}_{bi}\boldsymbol{\beta} - \boldsymbol{\alpha}_i; \boldsymbol{B}_i). \tag{3.6}$$

To be as general as possible, we will use the above expressions for the remainder of the paper.

Due to potential computational difficulties associated with high-dimensional models, Fieuws and Verbeke (2006) introduced a pseudo-likelihood method to simplify the model fitting. This involves fitting a bivariate model for every pair of responses and then combining the results. Kundu (2011) offers a convenient guide to implementing this method in SAS NLMIXED.

3.4. Conditional models

Conditional models offer a practical approach for making predictions of one subset of measurements, conditional on another subset. This methodology proves particularly valuable in circumventing challenges related to time-dependent covariates. In analogy to Section 3.3, we define $\boldsymbol{Y}_{ci}$ as a vector composed of all the measurements of a continuous responses and $\boldsymbol{Y}_{bi}$ as a vector composed of all the measurements of o ordinal responses. Next, let $\widetilde{\boldsymbol{Y}}_{ci}$ denote a $\widetilde{n}_i$-dimensional subset of the continuous response vector $\boldsymbol{Y}_{ci}$, while $\widetilde{\boldsymbol{Y}}_{bi}$ represents a $\widetilde{p}_i$-dimensional subset of the ordinal response vector $\boldsymbol{Y}_{bi}$. Notably, $\widetilde{\boldsymbol{Y}}_{ci}$ and $\widetilde{\boldsymbol{Y}}_{bi}$ can contain measurements of different, respectively, continuous and ordinal responses. By analogy, $\widetilde{\boldsymbol{X}}_{ci}$ and $\widetilde{\boldsymbol{X}}_{bi}$ represent the $\widetilde{n}_i \times q$ and $\widetilde{p}_i \times q$ submatrices of $\boldsymbol{X}_{ci}$ and $\boldsymbol{X}_{bi}$. Leveraging the ratios of the marginal distributions, it becomes feasible to derive expected values and their corresponding prediction intervals. A first conditional expected value is a subset the continuous responses, given a subset of both continuous and ordinal responses. This specific n_a-dimensional subvector of predicted continuous response(s) is denoted as $\widetilde{\boldsymbol{Y}}_{ci}^a$. This prediction is conditional on, on the one hand, $\widetilde{\boldsymbol{Y}}_{ci}^b$, the subvector of length n_b of values of the continuous response vector and, on the other hand, $\widetilde{\boldsymbol{Y}}_{bi}$, the subvector of ordinal responses. The notation will be as follows: the superscript specifies the submatrices or subvectors; superscript a and b denote, respectively, the rows a_1 until a_{n_a} and b_1 to b_{n_b}. In addition, the superscript bb specifies the rows b_1 until b_{n_b} and columns b_1 until b_{n_b}. The superscript ab indicates row a_1 until a_{n_a} and column b_1 until b_{n_b}. The conditional expected value is as follows:

$$E[\widetilde{\boldsymbol{Y}}_{ci}^a | \widetilde{\boldsymbol{Y}}_{ci}^b = \widetilde{\boldsymbol{y}}_{ci}^b, \widetilde{\boldsymbol{y}}_{bi} \le \boldsymbol{c}] = \left((\boldsymbol{E}_i \boldsymbol{V}_i^{-1} \widetilde{\boldsymbol{X}}_{ci} \boldsymbol{\beta})^a + \boldsymbol{E}_i^{ab} (\boldsymbol{E}_i^{bb})^{-1} (\widetilde{\boldsymbol{y}}_{ci}^b - (\boldsymbol{E}_i \boldsymbol{V}_i^{-1} \widetilde{\boldsymbol{X}}_{ci} \boldsymbol{\beta})^b) \right)$$

$$+ \left((\boldsymbol{E}_i \boldsymbol{H}_i' \boldsymbol{B}_i^{-1})^a - \boldsymbol{E}_i^{ab} (\boldsymbol{E}_i^{bb})^{-1} (\boldsymbol{E}_i \boldsymbol{H}_i \boldsymbol{B}_i^{-1})^b \right) \kappa, \tag{3.7}$$

where κ equals the expected value of the truncated normal distribution with variance $\boldsymbol{T}_i$, mean $\boldsymbol{F}_i$ and limits $]-\infty; \boldsymbol{d}]$. This expression is implemented in standard statistical software, such as in the R package tmvtnorm. The analytical expression can be found in Manjunath and Wilhelm (2021).

In addition,

$$
\begin{aligned}
E_i^{-1} &= H_i' B_i^{-1} H_i + V_i^{-1}, \\
T_i^{-1} &= (E_i H_i' B_i^{-1})^{b'} (E_i^{bb})^{-1} (E_i H_i' B_i^{-1})^b + B_i^{-1} - (H_i' B_i^{-1})' E_i (H_i' B_i^{-1}), \\
F_i &= T_i \Big((E_i H_i' B_i^{-1})^{b'} (E_i^{bb})^{-1} (\widetilde{y}_{ci}^b - (E_i V_i^{-1} \widetilde{X}_{ci} \beta)^b) + (H_i' B_i^{-1})' E_i (V_i^{-1} \widetilde{X}_{ci} \beta) \Big), \\
d &= \gamma_c - \widetilde{X}_{bi} \beta + H_i \widetilde{X}_{ci} \beta.
\end{aligned}
$$

The corresponding prediction interval and the derivations can be retrieved in Supplementary Appendix C.

A special case of (3.7) is when the continuous response is modeled conditional on solely the ordinal response. In this case, the expression of the expected value simplifies as follows

$$
E[\widetilde{Y}_{ci} | \widetilde{y}_{bi} \le c] = E_i \Big(V_i^{-1} \widetilde{X}_{ci} \beta + H_i' B_i^{-1} \kappa \Big), \tag{3.8}
$$

where κ is again the expected value of the truncated normal distribution with variance T_i^*, mean F_i^* and limits $]-\infty; d]$. Further,

$$
\begin{aligned}
T_i^{*-1} &= B_i^{-1} - (H_i' B_i^{-1})' E_i (H_i' B_i^{-1}) \\
F_i^* &= T_i^* \cdot (H_i' B_i^{-1})' E_i (V_i^{-1} \widetilde{X}_{ci} \beta).
\end{aligned}
$$

The expressions for the prediction interval and the related details considering the calculations can be found in Supplementary Appendix D.

To make predictions for the ordinal response, we can derive conditional probabilities for a subset of the ordinal response conditional on a subset of the ordinal response and the continuous response. We denote the subset for which we calculate the probability of being in category c or lower as $\widetilde{Y}_{bi}^a$, and we calculate it conditional on a subset of the ordinal response $\widetilde{Y}_{bi}^b$, and a subset of the continuous response, $\widetilde{Y}_{ci}$. The use of the superscripts b, ab and bb is in analogy with (3.7). The conditional probability can be expressed as follows:

$$
f(\widetilde{y}_{bi}^a \le c | \widetilde{y}_{bi}^b \le c, \widetilde{y}_{ci}) = \frac{\Phi(\gamma_c - \widetilde{X}_{bi} \beta - H_i (\widetilde{y}_{ci} - \widetilde{X}_{ci} \beta); B_i)}{\Phi(\gamma_c^b - \widetilde{X}_{bi}^b \beta - H_i^b (\widetilde{y}_{ci} - \widetilde{X}_{ci} \beta); B_i^{bb})}. \tag{3.9}
$$

After applying the logit transformation to ensure that the boundaries are constrained to the unit interval, the corresponding confidence interval can be calculated using the delta method. The gradients of the parameters can be found in Supplementary Appendix E.

A special case is the conditional density of the ordinal response(s) conditional on solely a subvector of the continuous response vector. The expected probability is then simplified to

$$
f(\widetilde{y}_{bi} \le c | \widetilde{y}_{ci}) = \Phi(\gamma_c - \widetilde{X}_{bi} \beta - H_i (\widetilde{y}_{ci} - \widetilde{X}_{ci} \beta); B_i). \tag{3.10}
$$

Supplementary Appendix F contains the formulas related to the standard errors.

3.5. Correlation function

By using the property that the responses are independent conditional on the random effects, it is feasible to deduce a correlation function from the hierarchical joint model. This correlation captures the manifest correlation, denoted as $\rho_{Y_{1ij}, Y_{2ik} \le c}$. It quantifies the relationship between the continuous response Y_{1i} at time j and the event of an ordinal response Y_{2i} below category c at

time k. This model-based manifest correlation represents the correlation between the scores on the original scale, whereas the latent correlation quantifies the correlation between the underlying random effects. Although calculating the latent correlation is simpler, the scientific interest often focuses on the manifest correlation rather than the latent correlation. The formula for the manifest correlation function is as follows:

$$\rho_{Y_{1ij},Y_{2ik}\leq c} = \frac{-\frac{1}{L_i}z'_{1ij}M_i^{-1}z_{2ik}\phi(\gamma_c-x'_{2ik}\beta;L_i^{-1})}{\sqrt{\left(z'_{1ij}D^*z_{1ij}+\Sigma_{1ij}\right)\Phi(\gamma_c-x'_{2ik}\beta;L_i^{-1})(1-\Phi(\gamma_c-x'_{2ik}\beta;L_i^{-1}))}}, \tag{3.11}$$

where D^* denotes the submatrix of D relating to the variances and covariances of the random effects of the responses Y_{1i} and Y_{2i}. In addition, $M = (D^*)^{-1} + z'_{2ik}z_{2ik}$.

The details of the derivations and the formulas regarding the standard errors can be found in Supplementary Appendix G.

4. PARAMETER ESTIMATION

The parameters in the joint random effects model are estimated via maximum likelihood. The likelihood function of the joint random-effects model is constructed under the assumption that the responses are independent given the random effects. As a result the likelihood function for a joint model of the responses Y_{ci} and Y_{bi} equals

$$L(\theta) = \prod_{i=1}^{N} \int f_{1i}(y_{ci}|b_i)f_{2i}(y_{bi} \leq c|b_i)f(b_i|D)\,db_i, \tag{4.12}$$

in which the vector θ contains all parameters of the conditional distributions and the distribution of the random effects b_i. In most cases, numerical approximations are needed for the integral in 4.12. In this paper, adaptive Gaussian quadrature is used for the estimation, which is implemented in the SAS procedure `NLMIXED` (Pinheiro and Bates, 1995; Molenberghs and Verbeke, 2005). The code for fitting the joint model can be found in Appendix H or via Github.

5. DATA ANALYSIS

In this section, the relationship between the continuous functioning score (ADL) and the ordinal level of impairment (MMSE) is examined. First, a joint model is implemented, as discussed in Section 3.3. Here, we fit a linear mixed model for ADL (as discussed in Section 3.1) and a generalized linear mixed model with a probit link for MMSE (as discussed in Section 3.2). Next, we allow the random effects of the responses to correlate to create the joint model. Since Fig. 1 clearly indicates that the evolution of ADL is not linear, we will include time as a categorical covariate in the linear mixed model. In contrast, time since the operation is included as a continuous covariate in the generalized linear mixed model for impairment. The model can be written as

$$
\begin{aligned}
Y_{1ij} \quad &= \quad \beta_{1,0} + \beta_{1,1}I(\text{Time}_{ij} = 5) + \beta_{1,2}I(\text{Time}_{ij} = 12) + \beta_{1,3}I(\text{Sex}_i = F) \\
&\quad + \beta_{1,4}\text{Age}_{ij} + b_{10i} + b_{11i}\frac{\text{Time}_{ij}}{100} + \epsilon_{1ij}, \\
\Phi^{-1}(P(Y_{2ij} \leq c)) \quad &= \quad \gamma_c - \left(\beta_{2,1}\text{Time}_{ij} + \beta_{2,2}I(\text{Sex}_i = F) + \beta_{2,3}\text{Age}_{ij} + b_{20i} + b_{21i}\frac{\text{Time}_{ij}}{100}\right).
\end{aligned}
$$

The hierarchical models include several random effects: b_{10i} and b_{20i} are the random intercepts for, respectively, ADL and MMSE. Next, b_{11i} and b_{21i} are the random slopes for respectively ADL and MMSE. In order to account for the correlation between the responses, different assumptions can be made regarding the joint distribution of the random effects such as for example setting the correlations to 1 (i.e. shared random-effects model). However, we have chosen to make the distribution as flexible as possible, i.e., not to impose any restriction on the covariance matrix.

Table 2. Parameter estimates (standard errors) of *ADLTOT* and *MMSE*.

Effect	ADLTOT		MMSE	
Intercept	3.42	(4.50)	–	
γ_1	–		–19.17	(5.70)
γ_2	–		–16.61	(5.50)
Time	–		0.04	(0.04)
Time 5	–2.68	(0.35)	–	
Time 12	–3.62	(0.57)	–	
Sex: Female	–1.58	(1.02)	–0.37	(1.11)
Age	0.20	(0.05)	–0.22	(0.07)
σ^2	3.02	(0.60)	–	

Table 3. Latent correlations [CI] between the random effects of MMSE and ADL.

	b_{10i}	b_{11i}	b_{20i}	b_{21i}
b_{10i}	1			
b_{11i}	.12 $[-.44; .61]$	1		
b_{20i}	$-.70[-.89; -.31]$	$-.38 [-.77; .21]$	1	
b_{21i}	.38 $[-.80; .95]$	$-.07 [-.98; .98]$	$-.72 [-1; .95]$	1

Note that to ensure that the correlation is bounded between -1 and 1, the Fisher-Z-transformation is applied to compute the confidence intervals, after which the values are transformed back to the original scale.

We assumed that $[b_{10i}, b_{11i}, b_{20i}, b_{21i}] \sim MVN(\mathbf{0}, \mathbf{D})$ and $\epsilon_{1i} \sim MVN(\mathbf{0}, \sigma^2 \mathbf{I}_i)$. The full SAS code can be found in Supplementary Appendix H and in Github. Convergence was reached within 10 hours and 19 minutes on a regular laptop (CPU = Processor Intel(R) Core(TM) i7-8850H CPU @ 2.60GHz, 2592 Mhz, 6 Core(s), 12 Logical Processor(s), RAM = 24GB) and resulted in the parameter estimates shown in Table 2.

An association between the responses was indicated by a Wald test, by showing that the co-variances among the random effects of the different responses significantly differ from zero (H_0 : $d_{13} = d_{14} = d_{23} = d_{24} = 0$, $\chi^2_{df=4} = 12.11, p = 0.02$). The latent correlations between the random effect offer a first glimpse into the relationship between the two responses (Table 3). These values can be interpreted in terms of the latent random effects. For instance, the correlation between the random intercepts ($r = -.69$) shows that immediately after the operation, a lower starting value of ADL (better functioning) than expected based on the covariates, is related to a lower probability of having a more severe level of impairment than expected based on the covariates. Still, these are the correlations between the underlying (latent) random effects on the probit-scale. It can be of interest to also examine the manifest correlations, as discussed in Section 3.5, which are the model-based correlations between the responses on their original scale.

By the use of (3.11) the correlations between the responses on their original scale can be computed. Since these model-based correlations depend on the covariate values chosen, we computed them for a man of mean age (78). They are displayed in Table 4. Functional impairment is at each timepoint significantly correlated with the event of a severe cognitive impairment and the event of impairment. The correlation is quite constant over time, but better cognitive functioning is consistently related to a higher probability of having no impairment and a lower probability of having severe impairment.

Table 5 presents the predicted MMSE (Mini-Mental State Examination) statuses on days 8 and 12, conditional on ADL (Activities of Daily Living) scores recorded on days 1 and 8. The latter ADL scores were set to one standard deviation below, at, or one standard deviation above the respective day's mean. Additionally, sex was set to female, and age was set to 78 years. The predictions are derived from the parameter estimates obtained from the joint model, which were then substituted

Table 4. Correlations between ADL (higher: lower functioning) and MMSE (cognitive impairment) for a 78-year-old man.

Panel A: Manifest correlations between ADL and the event of having severe impairment.				
Time(Impairment)				
Time (ADL)	1	3	5	8	12
1	.44 [.28 ; .58]	.44 [.29 ; .57]	.44 [.29 ; .57]	.43 [.28 ; .57]	.42 [.26 ; .57]
5	.48 [.34 ; .60]	.48 [.34 ; .60]	.48 [.34 ; .60]	.48 [.34 ; .61]	.48 [.31 ; .61]
12	.47 [.26 ; .65]	.48 [.28 ; .64]	.48 [.29 ; .64]	.49 [.29 ; .64]	.49 [.28 ; .65]

Panel B: Manifest correlations between ADL and the event of having impairment.				
Time(Impairment)				
Time (ADL)	1	3	5	8	12
1	.47 [.31 ; .60]	.47 [.32 ; .60]	.47 [.33 ; .6]	.48 [.34 ; .60]	.48 [.32 ; .61]
5	.51 [.38 ; .62]	.52 [.39 ; .62]	.52 [.41 ; .62]	.53 [.41 ; .63]	.54 [.41 ; .65]
12	.50 [.30 ; .66]	.51 [.32 ; .66]	.52 [.34 ; .66]	.54 [.37 ; .67]	.55 [.37 ; .70]

Table 5. Prediction of cognitive impairment based on the history of ADL at time 1 and 5 for a female of 78 years.

Timepoint prediction	History ADL (day 1–day 5)	P(Impairment=Severe)	P(Impairment)
8	14.57–10.87	0.03 [0.00; 0.94]	0.22 [0.12; 0.35]
8	18.10–15.42	0.25 [0.17; 0.36]	0.68 [0.52; 0.80]
8	21.63–19.97	0.72 [0.53; 0.86]	0.96 [0.75; 0.99]
12	14.57–10.87	0.02 [0.00; 1.00]	0.19 [0.47; 0.84]
12	18.10–15.42	0.21 [0.12; 0.35]	0.66 [0.48; 0.80]
12	21.63–19.97	0.69 [0.47; 0.84]	0.95 [0.73; 0.99]

into the conditional model (3.10). The results in Table 5 shows that a history of strong reliance on a caregiver corresponds to a high probability of cognitive impairment both in the present and in the immediate future. Furthermore, the confidence intervals emphasize that a low ADL score, indicative of low caregiver dependence, holds limited predictive value for MMSE status. This is in contrast to moderate or high ADL scores, which demonstrate stronger association with MMSE outcomes.

6. CONCLUDING REMARKS

In this research, our primary focus has been on introducing two new methodologies that are built upon the foundation of joint models: The first methodology involves closed-form expressions to obtain the manifest correlations from the model between the responses on their original scale, providing an alternative to investigating latent correlations between the underlying random effects on the probit-scale. Our second methodology entails employing conditional joint models in lieu of time-dependent covariates to analyze the effect of a longitudinal predictor on the longitudinal response. This shift effectively sidesteps several complications associated with time-dependent covariates. Firstly, the need for specifying lags is obviated with the joint modeling approach. With manifest correlations, we can assess the effect of a predictor on the response at each time point. In addition, our conditional model allows for straightforward adjustments of lags without refitting the joint model. Secondly, due to the symmetric nature of the relationship in a joint model, challenges posed by endogeneity or intermediary variables are mitigated. Thirdly, the presence of missing data necessitates no additional steps, thanks to the principle of ignorability (Rubin, 1976).

Consequently, our methodology becomes highly suitable for unbalanced data, as it operates without the requirement of lags or additional methods for handling missing data.

Moreover, our paper extends the application of our conditional model to scenarios involving multiple longitudinal predictors. By consolidating all predictors into an elongated vector and adapting the design matrices accordingly, our methodology remains seamlessly applicable.

Illustrating the practical utility of our methodology, we added a case study investigating the association of two longitudinal responses: a continuous physical functioning score and an ordinal mental functioning score. We show that predictions concerning the ordinal response can be effectively derived from the historical trajectory of the continuous response. In addition, the missingness is assumed to be at random, and hence results in no additional steps in the data analysis due to ignorability (Rubin, 1976). Implementation of the joint model can be achieved through the NLMIXED procedure in SAS. Code is provided in Supplementary Appendix H for both transforming the data in the correct format and fitting the joint model.

However, it's worth noting that a limitation of our methodology is its reliance on the proportional odds assumption inherent in the ordinal regression model. Of course, extension to non-proportionality is possible by having (certain) covariate effects category-dependent. But then, as always, care needs to be taken to ensure non-negative probabilities ensue. A second drawback is the computational complexity of a joint model, especially when more than two responses are included. Various methodologies can be used, such as the pairwise fitting approach for high-dimensional data (Fieuws and Verbeke, 2006), the split-sample approach for large datasets (Molenberghs et al., 2011), or a combination of both (Ivanova et al., 2017). A third limitation arises from the dependence of correlations and confidence intervals on the selected random effect structure. Therefore, it is recommended to model the random effects with a high degree of flexibility, potentially incorporating splines. Importantly, even within this scenario, our methodology remains applicable.

Further research can be conducted on the bounds of the manifest correlation function. Research in the context of surrogate markers (Alonso and Molenberghs, 2007) and the Bahadur model (Molenberghs and Verbeke, 2005) have shown that respectively the bounds of the R^2 or the correlation between dichotomous responses can generally be smaller than one. In addition, it can be of interest to implement the methodology in a SAS macro to facilitate the usability. Secondly, other approaches can be explored, such as multiple imputation models, where the value of the one longitudinal variable is imputed via a random-effects model and then included as a time-dependent covariate in the other longitudinal model. Still, some issues of time-dependent covariates would persist, such as endogeneity, possibility of intermediate variables and the definition of lags.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Biostatistics Journal* online.

FUNDING

REFERENCES

ALONSO A, MOLENBERGHS G. Surrogate marker evaluation from an information theory perspective. *Biometrics.* 2007:63(1):180–186.

BRESLOW NE, CLAYTON DG. Approximate inference in generalized linear mixed models. *J Am Stat Assoc.* 1993:88(421):9–25.

CAVENDER JB, ROGERS WJ, FISHER LD, GERSH BJ, COGGIN CJ, MYERS WO. Effect of smoking on survival and morbidity in patients randomized to medical or surgical therapy in the coronary artery surgery study (CASS): 10-Year follow-up. *J Am College Cardiol.* 1992:20(2):287–294.

CHAKRABORTY H, HELMS RW, SEN PK, COHEN MS. Estimating correlation by using a general linear mixed model: evaluation of the relationship between the concentration of HIV-1 RNA in blood and semen. *Stat Med*. 2003:22(9):1457–1464.

DELPORTE M, FIEUWS S, MOLENBERGHS G, VERBEKE G, SITUMA WANYAMA S, HATZIAGOROU E, DE BOECK C. A joint normal-binary (probit) model. *Int Stat Rev*. 2022:90:S37–S51.

DIGGLE P. *Analysis of longitudinal data*. New York: Oxford Statistical Science Series; 2002. p. 245–281.

ENGEL B, KEEN A. A simple approach for the analysis of generalized linear mixed models. *Stat Neerlandica*. 1994:48(1):1–22.

FAES C, GEYS H, AERTS M, MOLENBERGHS G, CATALANO PJ. Modeling combined continuous and ordinal outcomes in a clustered setting. *J Agric Biol Environ Stat*. 2004:9(4):515–530.

FIEUWS S, VERBEKE G Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics*. 2006:62(2):424–431.

FIEUWS S, VERBEKE G, BOEN F, DELECLUSE C. High dimensional multivariate mixed models for binary questionnaire data. *J R Stat Soc Ser C (Appl Stat)*. 2006:55(4):449–460.

HEDEKER D, GIBBONS RD. A random-effects ordinal regression model for multilevel analysis. *Biometrics*. 1994:50(4):933–944.

IVANOVA A, MOLENBERGHS G, VERBEKE G. Mixed models approaches for joint modeling of different types of responses. *J Biopharm Stat*. 2016:26(4):601–618.

IVANOVA A, MOLENBERGHS G, VERBEKE G. Fast and highly efficient pseudo-likelihood methodology for large and complex ordinal data. *Stat Methods Med Res*. 2017:26(6):2758–2779.

KUNDU M. Implementation of pairwise fitting technique for analyzing multivariate longitudinal data in SAS. Nashville, TN, United States: PharmaSUG; 2011, p. 1–12.

LAIRD NM, WARE JH. Random-effects models for longitudinal data. *Biometrics*. 1982:38(4):963–974.

MANJUNATH BG, WILHELM S. Moments calculation for the doubly truncated multivariate normal density. *J Behav Data Sci*. 2021:1(1):13–33.

MILANZI E, MOLENBERGHS G, ALONSO A, VERBEKE G, DE BOECK P. Reliability measures in item response theory: manifest versus latent correlation functions. *Br J Math Stat Psychol*. 2015:68(1):43–64.

MILISEN K, ABRAHAM IL, BROOS PL. Postoperative variation in neurocognitive and functional status in elderly hip fracture patients. *J Adv Nursing*. 1998:27(1):59–67.

MOLENBERGHS G, VERBEKE G. *Models for discrete longitudinal data*. New York, NY: Springer; 2005.

MOLENBERGHS G, VERBEKE G, IDDI S. Pseudo-likelihood methodology for partitioned large and complex samples. *Stat Probab Lett*. 2011:81(7):892–901.

PINHEIRO JC, BATES DM. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *J Comput Graphical Stat*. 1995:4(1):12–35.

PODDAR A. *Analysis off dependent discrete choices using Gaussian Copula*. PhD thesis, Old Dominion University, 2016.

QIAN T, KLASNJA P, MURPHY SA. Linear mixed models with endogenous covariates: modeling sequential treatment effects with application to a mobile health study. *Stat Sci*. 2020:35(3):375–390.

RIZOPOULOS D. *Joint models for longitudinal and time-to-event data: With applications in R*. New York: Chapman and Hall/CRC; 2012, p. 100.

RUBIN DB. Inference and missing data. *Biometrika*. 1976:63(3):581–592.

TOMBAUGH T, MCINTYRE N. The mini-mental state examination: a comprehensive review. *J Am Geriatrics Soc*. 1992:40(9):922–935.

VERBEKE G, MOLENBERGHS G. *Linear mixed models for longitudinal data*. New York: Springer Series in Statistics; 2000; p. 24.

WOLFINGER R, O'CONNELL M. Generalized linear mixed models: a pseudo-likelihood approach. *J Stat Comput Simul*. 1993:48:233–243.