

A Joint Marginalized Multilevel Model for Longitudinal Outcomes

Samuel Iddi¹ and Geert Molenberghs^{2,1},

Interuniversity Institute for Biostatistics and statistical Bioinformatics

¹ Katholieke Universiteit Leuven, Kapucijnenvoer 35, 3000 Leuven, Belgium

² Universiteit Hasselt, Agoralaan 1, 3590 Diepenbeek, Belgium

Abstract

The shared-parameter model and its so-called hierarchical or random-effects extension are widely used joint modeling approaches for a combination of longitudinal continuous, binary, count, missing, and survival outcomes that naturally occurs in many clinical and other studies. A random effect is introduced and shared or allowed to differ between two or more repeated measures or longitudinal outcomes, thereby acting as a vehicle to capture association between the outcomes in these joint models. It is generally known that parameter estimates in a linear mixed model (LMM) for continuous repeated measures or longitudinal outcomes allow for a marginal interpretation, even though a hierarchical formulation is employed. This is not the case for the generalized linear mixed model (GLMM), i.e., for non-Gaussian outcomes. The aforementioned joint models formulated for continuous and binary or two longitudinal binomial outcomes, using the LMM and GLMM will naturally have marginal interpretation for parameters associated to the continuous outcome but a subject-specific interpretation for the fixed effects parameters relating covariates to binary outcomes. To derive marginally meaningful parameters for the binary models in a joint model, we adopt the marginal multilevel model due to Heagerty (1999; Heagerty and Zeger 2000) and formulate a joint marginal multilevel model for two longitudinal responses. This enables to (1) capture association between the two responses and (2) obtain parameter estimates that have a population-averaged interpretation for both outcomes. The model is applied to two sets of data. Results are compared with existing approaches such as generalized estimating equations (GEE), GLMM, and the model of Heagerty (1999). Estimates were found to be very close to those from single analysis per outcome but the joint model yields higher precision and allows for quantifying the association between outcomes. Parameters were estimated using maximum likelihood. The model is easy to fit using available tools such as the SAS NL MIXED procedure.

Keywords: Generalized estimating equation; Joint model; Marginal multilevel model; Maximum likelihood estimation; Random effects model; Shared-parameter model.

1 Introduction

Joint modeling has received massive attention in recent years, owing to researchers' desire for more insight into their data with a single statistical model. The reason to find this type of analysis is because commonly researchers simultaneously record several kinds of outcomes in their studies. These outcomes are often of a mixed nature. Prevalent examples are situations where a combination of continuous, binary, ordinal, survival, and missing outcomes occurs. Continuous and binary outcomes often appear in longitudinal studies where one observes follow up measurements on patients. For time-to-event outcomes, patients are followed up until they experience an event of interest. Oftentimes, because studies are conducted in humans, data are incomplete owing to dropout or other reasons for

missingness. These issues requires careful attention.

Earlier, emphasis has been on separate analyses per outcome. Even in situations where joint analysis of outcomes were required, relatively naive approaches had been used. For a single longitudinal or survival outcome, we may be interested in finding out if a particular treatment has an effect on the outcome. We may also want to know if the longitudinal evolution of an outcome differs for example between males and females. Such questions could be addressed with well-known methods, such as the Cox model for time-to-event outcomes or mixed-effect and marginal models for longitudinal outcomes. This type of single analyses are limited in that they do not provide answers to questions that take several or all outcomes simultaneously into account. Conducting a joint analysis allows addressing additional scientifically relevant questions. For example, when one is interested in knowing whether a new treatment could improve all outcomes simultaneously or in the measurement of the association between the various responses and how this association evolves over time, a joint model is advisable. Also, joint models are popular owing to the fact that they ensure unbiased statistical inferences (Tsiatis, DeGruttola, and Wulfsohn 1995, Wulfsohn and Tsiatis 1997) in a variety of settings. As mentioned earlier, though scientific interest may be in the longitudinal or survival outcomes, we may also be required to accommodate incompleteness of measurements. Failure to do so can lead to biased parameter and precision estimates (Prentice 1988). Another situation that requires joint modeling techniques is when there is interest in improving predictions. For instance, to improve the prediction of the probability of survival, one could usefully include all available outcomes.

There are several instances in the statistical literature where joint models have been adopted. For example, in HIV/AIDS clinical trials, quality of life studies, cancer trials, etc., joint modeling of survival and longitudinal data are commonly found (DeGruttola and Tu 1994; Clayton 1992; Tsiatis, DeGruttola, and Wulfsohn 1995; Wulfsohn and Tsiatis 1997; Faucett and Thomas 1996) to assess the association between both outcomes. An excellent review of various joint modeling approaches for longitudinal and time to event data can be found in Tsiatis and Davidian (2004). These techniques have been made flexible so that they can handle different types of longitudinal outcomes. Horrocks and van den Heuvel (2009) used a joint longitudinal and GLM model, developed by Wang *et al* (2000), to predict pregnancy in a group of women undergoing treatment for infertility, based on longitudinal adhesion measurements. Li *et al* (2007) considered a semi-parametric joint model to study the association between bone status in peri-menopausal women, being the primary endpoint, and longitudinal hormone levels. Molenberghs and Verbeke (2005) discuss a number of techniques that jointly model continuous and discrete outcomes. Other studies have employed joint models to investigate how early measurements of a trial or an alternative outcome can be used as a surrogate for the primary endpoint (Buyse and Molenberghs 1998; Burzykowski, Molenberghs, and Buyse 2005).

Consider a scientific study where the research interest is not only in the association between the outcomes but also in the effect of some covariates on the outcomes and their interpretation. The random effects approach to joint modeling uses random effects to capture the correlation between

the outcomes, as well as within each longitudinal outcome (Laird and Ware 1982). It is generally known that the parameter estimates from a linear mixed model have a marginal interpretation even though a hierarchical formulation is employed. Such is not the case for the generalized linear mixed model for non-Gaussian outcomes, where the generalized linear mixed model parameters have a subject-specific interpretation. When continuous and binary or count longitudinal outcomes are modeled jointly, it is obvious that parameters associated with the continuous outcome will have a marginal interpretation while those for the non-Gaussian outcome will have their usual conditional interpretation. It is therefore of interest to have a model where population-averaged interpretations of the parameters are also available for the discrete outcomes. Heagerty (1999) proposed a now well-established marginalized multilevel model by formulating a separate model for the marginal means for a single binary outcome, which can depend on covariates only and another one for the conditional means which is related to a term that links the marginal with conditional means and random effects. In so doing, the model enjoys the various strengths of marginal and conditional modeling techniques. Particularly, effects of covariates will have a direct marginal interpretation.

We followed the modeling concepts of Heagerty (1999) and formulated a joint model for two longitudinal outcomes. The generalized linear mixed model component in a shared-parameter model and its so-called hierarchical extensions was replaced by the model of Heagerty (1999). The proposed model therefore has a marginal interpretation for fixed parameters of known covariates that are associated to the binary outcome, it also enables inferences about the association between the outcomes. We offered a brief review of models for correlated continuous and binary data. Full maximum likelihood estimation with iterative numerical quadrature methods are adopted to obtain parameter estimates.

The remainder of the paper is organized as follows. We devote Section 2 to the introduction of two datasets; these are analyzed in Section 5. Methodology for existing marginal and hierarchical models for continuous and binary longitudinal outcomes are reviewed, and our proposed Joint Marginal Multilevel Model (JOMMM) technique is presented in Section 3. The estimation method is the subject of Section 4. Section 6 reports a simulation study undertaken to explore the performance of the proposed method.

2 Case Studies

2.1 Serological HCV and HIV Data from Italy

This dataset is taken from an annual serological survey conducted between 1998 and 2006 from 20 regions in Italy. The study was conducted by the European Monitoring Center for Drugs and Drug Addition (Mathei *et al* 2006). The data consist of hepatitis C virus (HCV) and human immunodeficiency virus (HIV) status and risk factors of drug users who made an appointment in the drug treating center. Individual drug users data were not available but rather the total number of

drug users who visited the facility each year and the number tested positive for HCV and HIV per region. The resulting dataset is a sequence of binary outcomes for each of the 20 Italian regions over the years. Not all regions are present at all years. Consider, for example, the HCV data. The maximum number of respondents is 15,401 (average 3866.61), and the maximum of positive tests is 10,875 (average 2578.12). The aim is to investigate the change in HCV over time, i.e., whether year-effects are present in the profiles. Table 1 lists the overall HCV prevalence for each of the 20 regions. From these aggregated serological data, we can consider: (1) the evolution of each infection, i.e., whether year-effects are present in the profiles, both at the regional and population levels; (2) for each infection, the association between measurements taken over time; and (3) the association between both infections at regional and population levels. Observed prevalence profiles are shown in Figure 1. Unsurprisingly, the prevalence of HCV is seen to be higher than that for HIV. For both infections, there apparently is much more between variability as opposed to within variability. This is an indication of possible correlation between prevalence over time from the same region.

2.2 The Age Related Macular Degeneration Trial

The Age Related Macular Degeneration (ARMD) has been presented and studied by Buyse and Molenberghs (1998) and Molenberghs and Verbeke (2005). Primary trial results have been reported elsewhere (Pharmacological Therapy for Macular Degeneration Study Group 1997). The data resulted as a product of a randomized multi-centric clinical trial for patients with ARMD, a condition associated with progressive loss of vision in the elderly. The aim of the trial was to compare experimental interferon- α to placebo. The outcome of the trial was the patients' visual acuity, which was measured at 4 follow-up visits (4, 12, 24, and 52 weeks). During each visit, patients were made to read lines of letters on standardized vision charts and the total number of letters that were correctly read was recorded as the patients' visual acuity. Buyse and Molenberghs (1998) studied if patients' performance at 6 months could be used as a surrogate for their performance at 1 year with respect to the effect of the experimental treatment. They investigated whether the loss of vision of at least 2 lines at 6 months could serve as a surrogate for the loss of vision of at least 3 lines at 1 year. The full longitudinal profile was subjected to analysis in Molenberghs and Verbeke (2005).

For our purposes, two different versions of visual acuity will be considered, namely (1) change in visual acuity at the different time points, assumed to be normally distributed, after onset of treatment and (2) a binary variable indicating whether or not there is loss of vision at the various visit periods compared to when treatment started. While one outcome derives from the other, the setting is relevant because the continuous sequence obviously is the directly obtained one, while the derived binary sequence is often the one requested for regulatory purposes. It is then useful to assess the connection between the results derived from both, in particular the association between them and the correspondence between the treatment effect. Indeed, while there is a deterministic rule in deriving the second from the first sequence, it is not clear whether this will automatically result in

a treatment effect of the same magnitude; neither is it clear *a priori* what the correlation between the two sequences will be. These issues can be addressed satisfactorily only in the context of a joint model.

3 Methodology

3.1 Models for a Single Longitudinal Continuous Response

The hierarchical linear mixed model (LMM) and its corresponding marginal models are appropriate statistical models for continuous hierarchical data, given that they duly acknowledge dependence between observations within subjects, through the use of random effects. Assume that there are $i = 1, \dots, N$ subjects and $j = 1, 2, \dots, n_i$ follow-up continuous measurements Y_{ij} taken for each subject. The LMM is specified as:

$$Y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i + \varepsilon_{ij}. \quad (1)$$

This model involves two set of covariates \mathbf{x}_{ij} and \mathbf{z}_{ij} . The $n_i \times p$ covariates \mathbf{x}_{ij} are associated with a p -dimensional vector of fixed-effects parameters $\boldsymbol{\beta}$ and the $n_i \times q$ set of covariates \mathbf{z}_{ij} associated with the random effects $\mathbf{b}_i \sim N(0, \mathbf{D})$. In addition, $\varepsilon_{ij} \sim N(0, \mathbf{R}_i)$ represents the residual of the j th observation on the i th subject. Given the random effects \mathbf{b}_i , the residuals ε_{ij} are often (but not always) assumed independent. The variance-covariance matrix \mathbf{D} indicates the degree of heterogeneity of subjects. When all dependent residuals are considered, a variety of covariance structures are then possible for both \mathbf{D} and \mathbf{R}_i , such as unstructured, compound symmetry, and first-order autoregressive matrices. Note that $E(Y_{ij}) = E[E(Y_{ij}|\mathbf{b}_i)] = \mathbf{x}'_{ij}\boldsymbol{\beta}$ and so marginal and conditional parameters are equal.

Alternatively, one can postulate the following marginal model:

$$Y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \varepsilon_{ij}^*, \quad (2)$$

with correlated residuals $\varepsilon_{ij}^* \sim N(0, \mathbf{V}_i^*)$. The marginal distribution of the response is then $Y_{ij} \sim N(\mathbf{x}'_{ij}\boldsymbol{\beta}, \mathbf{V}_i^*)$. In this case, correlation is taken into account through covariance parameters in \mathbf{V}_i^* . Again, different specifications of the covariance structure can be imposed for the covariance \mathbf{V}_i^* as mentioned above for \mathbf{D} and \mathbf{R}_i . It is well known that the marginal model resulting from (1) is a special case of (2). The hierarchical linear mixed model therefore implies a specific marginal model with $\varepsilon_{ij}^* \sim N(0, \mathbf{V}_i^*)$ where $\mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}'_i + \mathbf{R}_i$. A very important fact is that the implied marginal model removes the positive definiteness restrictions on the \mathbf{D} and \mathbf{R}_i matrices, merely requiring that \mathbf{V}_i be positive definite; thus, weaker restrictions on the covariance parameters apply (Verbeke and Molenberghs 2003). Maximum likelihood (ML) and restricted maximum likelihood (REML) are popular estimation methods for the mixed-model parameters (Laird and Ware 1982, Verbeke and Molenberghs 2000). The latter is used to obtain less biased estimates of the covariance parameters

(Patterson and Thompson 1971, Harville 1997, Cooper and Thompson 1977, Verbyla 1990). The EM algorithm (Dempster, Laird and Rubin 1977) and Newton-Raphson-based procedures (Lindstrom and Bates 1988) are common updating methods. These have been implemented in statistical packages such as SAS and R. Bayesian alternatives are also available for estimation (Gelman *et al* 1995).

3.2 Models for a Single Longitudinal Binary Response

For correlated non-Gaussian outcomes, the generalized linear mixed model (GLMM) and generalized estimating equations (GEE) are widely used statistical tools. The former is an extension to the linear mixed model and the (univariate) generalized linear model (GLM). The GLMM is based upon specifying a exponential-family distribution for the outcome, a link function, and a random-effect structure. The model formulation for binary data will be discussed in Section 3.3. Unlike for their continuous counterparts, the marginal and conditional means are not equal. For example, using a logit link to relate covariates and random effects to the expectation of Y_{ij} ,

$$E(Y_{ij}) = E[E(Y_{ij}|\mathbf{b}_i)] = E \left[\frac{\exp(\mathbf{x}'_{ij}\boldsymbol{\alpha} + \mathbf{z}'_{ij}\mathbf{b}_i)}{1 + \exp(\mathbf{x}'_{ij}\boldsymbol{\alpha} + \mathbf{z}'_{ij}\mathbf{b}_i)} \right] \neq \frac{\exp(\mathbf{x}'_{ij}\boldsymbol{\alpha})}{1 + \exp(\mathbf{x}'_{ij}\boldsymbol{\alpha})}.$$

An alternative route is offered by a direct marginal specification. GEE was first introduced by Liang and Zeger (1986; Zeger and Liang 1986). The approach extends GLM by allowing for correlation within subject through a so-called working correlation matrix. Score equations are formulated and solved. Commonly used working structures are independence, exchangeability, autoregressive, and unstructured. As opposed to GLMM, GEE is a marginal model and so parameters have a marginal interpretation. The main advantage of GEE is that even when the working correlation structure is misspecified, parameter estimators are consistent and asymptotically normal. Empirically corrected standard errors are unbiased. The methodology has been implemented in various packages. One of its weaknesses lies in the fact that it is only valid under missing completely at random (MCAR) and not when data are missing at random (MAR) (Diggle *et al* 2002). Also, it performs well only with relatively large sample sizes, particularly for binary data. Evidently, it lacks a likelihood basis. Extensions, based on inverse probability weighting, allow for MAR missingness (Fitzmaurice *et al* 2009). Details about GEE and its extensions can be found in Molenberghs and Verbeke (2005).

Heagerty (1999) and Heagerty and Zeger (2000) proposed the concept of a marginally specified logistic-normal model for longitudinal binary data. This model brings together the strength of both GEE and GLMM. For a binary outcome Y_{ij} , the model formulation is given by

$$\begin{aligned} \eta(\mu_{ij}^m) &= \mathbf{x}'_{ij}\boldsymbol{\alpha}^m, \\ \eta(\mu_{ij}^c) &= \delta_{ij} + \mathbf{z}'_{ij}\mathbf{b}_i, \\ \mathbf{b}_i &\sim F_b(0, \mathbf{D}), \\ Y_{ij} &\sim F_Y(\mu_{ij}^c, v). \end{aligned}$$

A separate model for the marginal mean $\mu_{ij}^m = E(Y_{ij})$ and the conditional mean $\mu_{ij}^c = E(Y_{ij}|\mathbf{b}_i)$ are specified and linked together by δ_{ij} , which is allowed to depend on covariates, marginal parameters, and random-effect components. The term δ_{ij} is obtained from the integral equation:

$$\mu_{ij}^m = g(\mathbf{x}'_{ij}\boldsymbol{\alpha}^m) = \int_b g(\delta_{ij} + \mathbf{z}'_{ij}\mathbf{b}_i)dF_b, \quad (3)$$

where $g(\cdot) = \eta^{-1}(\cdot)$ is an inverse link function. When, for example, the link function is logit and the distribution of the random effect is normal, then δ_{ij} is obtained from

$$\text{expit}(\mathbf{x}'_{ij}\boldsymbol{\alpha}^m) = \int_b \text{expit}(\delta_{ij} + \mathbf{z}'_{ij}\mathbf{b}_i)f(\mathbf{b}_i)d\mathbf{b}_i.$$

where $\mathbf{b}_i \sim N(0, \mathbf{D})$. It is known from logistic-normal regression models that this integral has no closed form solution. Therefore, an iterative numerical approximation may be required to execute the integration. This is less than straightforward because another numerical integration is required to obtain the marginal density from the conditional density. Griswold and Zeger (2004) extended this model to allow for different link functions for the marginal and conditional model specification. Therefore, instead of a logit-logit-normal model, we may specify a logit link for the marginal model and a probit link for the conditional model. In this case, we will retain the odds ratio interpretation of the marginal parameters while taking advantage of the computational ease emanating from the probit-normal relationship. We then obtain a closed form expression for δ_{ij} :

$$\delta_{ij} = \sqrt{1 + \mathbf{z}'_{ij}\mathbf{D}\mathbf{z}_{ij}} \cdot \Phi^{-1}[\text{expit}(\mathbf{x}'_{ij}\boldsymbol{\alpha}^m)].$$

with $\Phi^{-1}(\cdot)$ the probit link function. Expressions for δ_{ij} for other data types and their convenient link functions can be found in Griswold and Zeger (2004). These authors also discuss maximum likelihood estimation. Apart from combining both marginal and conditional parameter interpretation, the models enjoys a likelihood basis and allows derivation of the the full probability distribution for the response (Fitzmaurice and Laird 1993; Molenberghs and Lesaffre 1994). Heagerty and Zeger (2000) demonstrated through a simulation study that parameters from this marginal multilevel model (MMM) are less sensitive to random-effects assumptions than the conditional GLMM model. Furthermore, the model produces valid inferences when data are missing at random (MAR).

3.3 Models for Joint Longitudinal Responses

Consider two longitudinal outcomes Y_{1ij} and Y_{2ik} , denoting the j th and k th measurement on the i th subject for continuous and binary type outcomes, respectively, ($i = 1, \dots, N$, $j = 1, 2, \dots, n_{1i}$, and $k = 1, 2, \dots, n_{2i}$). This means that we need to develop an appropriate model for the joint distribution $f(\mathbf{Y}_{1i}, \mathbf{Y}_{2i})$ of the continuous vector \mathbf{Y}_{1i} and binary vector \mathbf{Y}_{2i} . An attractive joint modeling technique is the shared-parameter model (Molenberghs and Verbeke 2005), where an unobserved random variable is introduced, given which, the two outcomes are further assumed independent. In

other words, the random effects are solely responsible for generating the association between the outcomes.

We begin by formulating a linear mixed model for the continuous outcomes. This means that \mathbf{Y}_{1i} follows a linear mixed model of the type (1). Next to this, a generalized linear mixed model is specified for the binary outcome. For $Y_{2ik} \sim \text{Bernoulli}(\mu_{ik})$, the expectation of Y_{2ik} is related to covariates and random effects through a known link function $\eta(\cdot)$ in the following way:

$$\eta(\mu_{ik}) = \eta[E(Y_{2ik}|\mathbf{b}_i)] = \mathbf{x}'_{2ik}\boldsymbol{\alpha} + \Lambda\mathbf{z}'_{2ik}\mathbf{b}_i. \quad (4)$$

In our general notation, we will subscript the covariate vectors with '2' for the binary sequence and '1' for the continuous sequence. The covariates can but do not have to have components in common. The fixed-effects parameters $\boldsymbol{\beta}$ for the continuous outcome and $\boldsymbol{\alpha}$ for the binary outcome are kept separate because the type of response variable is different in the first place. Nevertheless, the random effects \mathbf{b}_i have to be considered common because they are employed to generate both within- and between-sequence association. Because of the aforementioned difference of scale, a (usually diagonal) re-scaling matrix Λ is included in the GLMM; equivalently, it could be introduced into (1). The parameter Λ allows us to have the random effects \mathbf{b}_i shared between the two outcomes, even though the two outcome sequences are of a different type and can also have different measurement units. Supposing the two outcomes are of binomial nature, two GLMM's will be used with the scale parameter Λ introduced in any of the models to relax the assumption of common variance between the two outcomes. Because of the aforementioned conditional independence, the conditional joint density of the two outcomes becomes the product of the density of the individual outcomes conditional on the latent variable. Integrating out the random effect gives the marginal joint density. This shared-parameter modeling approach is flexible and convenient. For instance, it is straightforward to have different types of outcomes and very easy to postulate the model. Also, parameters in the joint model will still have the same interpretation as in their corresponding univariate versions. Extensions to very high dimensions are very straightforward as well, although it is important to choose sufficiently efficient updating algorithms when implementing a parameter estimation routine.

The shared-parameter joint model however is known to suffer from some drawbacks. Notable among them is that it imposes a specific kind of association between the responses. For example, in a random intercept shared-parameter model for two continuous longitudinal outcomes, the random intercept is used to capture the association in one outcome and at the same time in the other outcome and so a particular association is imposed between the two outcomes. This can easily be dealt with by what is known as the hierarchical or random effect approach where different random effects are used for the different longitudinal profiles. For example, using a single random effect for the individual response models, they can be expressed for two binomial outcomes as follows:

$$\eta(\mu_{1ij}) = \mathbf{x}'_{1ij}\boldsymbol{\beta} + \mathbf{z}'_{1ij}b_{1i}, \quad (5)$$

$$\eta(\mu_{2ik}) = \mathbf{x}'_{2ik}\boldsymbol{\alpha} + \mathbf{z}'_{2ik}b_{2i}, \quad (6)$$

and further

$$\mathbf{b}_i = (b_{1i}, b_{2i})' \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix} \right]. \quad (7)$$

This implies that the random intercepts in (5) and (6) not only have their individual variances d_{11} and d_{22} , respectively; they are also connected through the covariance d_{12} . The conditional means $\mu_{1ij} = E(Y_{1ij}|b_{1i})$ and $\mu_{2ik} = E(Y_{2ik}|b_{2i})$ are related to linear fixed and random factors through a link function $\eta(\cdot)$. Similar to the shared-parameter model, given these random effects, the outcomes are assumed to be independent of each other. If the random effects are assumed to be uncorrelated, then the resulting model is equivalent to modeling the two outcomes separately using a generalized linear mixed model. The joint marginal density can also be obtained by integrating out the two random effects from the conditional independence model.

In our proposal, we replace (4) by Heagerty's (1999) proposal. This yields what we refer to as a *joint marginalized multilevel model* (JOMMM). The logit-probit normal version is adopted so that analytical expressions can be derived for the joint distribution of the two responses. The new model is completely spelled out for continuous and binary data as (1) combined with $\text{logit}(\mu_{2ik}^m) = \mathbf{x}'_{2ik} \boldsymbol{\alpha}^m$ and $\Phi^{-1}(\mu_{2ik}^c) = \delta_{ik} + \mathbf{w}'_{ik} \mathbf{b}_i$ for a continuous and binary outcome.

Here, \mathbf{w}'_{ik} is a vector of scale parameters and covariates (i.e., $\mathbf{w}'_{ik} = \Lambda \mathbf{z}'_{2ik}$), $\mu_{2ik}^m = E(Y_{2ik} = 1)$, $\mu_{2ik}^c = E(Y_{2ik} = 1|\mathbf{b}_i)$ and $\delta_{ik} = \sqrt{1 + \mathbf{w}'_{ik} \mathbf{D} \mathbf{w}_{ik}} \cdot \Phi^{-1} \{ \text{expit}(\mathbf{x}'_{2ik} \boldsymbol{\alpha}^m) \}$. We can derive the joint marginal distribution by integrating out the random effect. Thus, the contribution of the i th subject to the likelihood is given by

$$\begin{aligned} f_i(\mathbf{Y}_{1i} = \mathbf{y}_i, \mathbf{Y}_{2i} = 1) &= \frac{1}{(2\pi)^{\frac{n_i}{2}} |\mathbf{R}_i|^{\frac{1}{2}} |\mathbf{D}|^{\frac{1}{2}}} \times e^{-\frac{1}{2} [(\mathbf{y}_i - \mathbf{X}_{1i} \boldsymbol{\beta})' \mathbf{R}^{-1} (\mathbf{y}_i - \mathbf{X}_{1i} \boldsymbol{\beta})]} \\ &\times \prod_k | \mathbf{D}^{-1} + \mathbf{Z}'_i \mathbf{R}_i \mathbf{Z}_{1i} + \mathbf{w}_{ik} \mathbf{w}'_{ik} |^{-\frac{1}{2}} \left[\frac{e^{\frac{1}{2} (\frac{r_2^2}{4r_1} - r_3)}}{r_1^{\frac{1}{2}}} \right] \\ &\times \Phi \left[\left(\delta_{ik} + \frac{r_2}{2r_1} \right) r_1^{\frac{1}{2}} \right], \end{aligned} \quad (8)$$

where

$$\begin{aligned} r_1 = r_{1(k)} &= I - \mathbf{w}'_{ik} \left[\left(\mathbf{D}^{-1} + \mathbf{Z}'_i \mathbf{R}_i \mathbf{Z}_{1i} + \mathbf{w}_{ik} \mathbf{w}'_{ik} \right)^{-1} \right]' \mathbf{w}_{ik}, \\ r_2 = r_{2(k)} &= \mathbf{w}'_{ik} \left[\left(\mathbf{D}^{-1} + \mathbf{Z}'_i \mathbf{R}_i \mathbf{Z}_{1i} + \mathbf{w}_{ik} \mathbf{w}'_{ik} \right)^{-1} \right]' \mathbf{Q}', \\ r_3 = r_{3(k)} &= -\frac{1}{4} \mathbf{Q} \left[\left(\mathbf{D}^{-1} + \mathbf{Z}'_i \mathbf{R}_i \mathbf{Z}_{1i} + \mathbf{w}_{ik} \mathbf{w}'_{ik} \right)^{-1} \right]' \mathbf{Q}', \\ \mathbf{Q} &= \left[(\mathbf{y}_i - \mathbf{X}_{1i} \boldsymbol{\beta})' \mathbf{R}_i^{-1} \mathbf{Z}_{1i} + (\mathbf{y}_i - \mathbf{X}_{1i} \boldsymbol{\beta})' (\mathbf{R}_i^{-1})' \mathbf{Z}_{1i} \right]. \end{aligned}$$

For two longitudinal binary or binomial responses, the JOMMM is formulated by replacing each of the two conditional models (5) and (6) in a joint model by the marginal and conditional model of the MMM, where the same or different random effects are introduced in the conditional models for the two outcomes.

4 Estimation

Parameters in the joint model are estimated using maximum likelihood, based on

$$L(\boldsymbol{\beta}, \boldsymbol{\alpha}^m, D) = \prod_{i=1}^N f_i(\mathbf{Y}_{1i} = \mathbf{y}_i, \mathbf{Y}_{2i} = 1).$$

Even though this analytical joint marginal likelihood can be maximized, it is cumbersome to manipulate. It is therefore more convenient to maximize the likelihood after employing numerical techniques, rather than to integrate out the random-effects distribution. Gaussian and adaptive Gaussian quadrature are designed for such purpose, up to a pre-specified level of accuracy (Pineiro and Bates 1995, 2000). The standard errors of the parameter estimates are computed from the inverse Hessian matrix (second derivatives) at the estimates obtained numerically. Major statistical tools, such as the SAS procedure NLMIXED, are readily available for fitting the models specified in this paper. Other estimation techniques are discussed in Molenberghs and Verbeke (2005) for discrete and in Verbeke and Molenberghs (2000) for continuous outcomes.

5 Analysis of Case Studies

5.1 Analysis of the HCV and HIV Data

We begin by analyzing the serological survey data which collates the HCV and HIV status of drug users. The prevalence sequence of these two infections will be modeled jointly to capture association between them, as well as the association within a sequence over time. Suppose Y_{1ij} and Y_{2ij} are the number of positive cases of HCV and HIV, each assumed to follow a binomial distribution with probability of success μ_{1ij} and μ_{2ij} , respectively, and with n_{1ij} and n_{2ij} trials, for the i th region at the j th year. We will begin by fitting to each outcome separately a GLMM model, GEE with a compound symmetry working assumption, and a MMM. The purpose of these models is to make comparisons between estimates from the hierarchical and the marginal models, the latter of which has a population-averaged interpretation for the fixed effects. The mean structure of all models is of the form: $\text{logit}(\mu_{1ij}) = \alpha_0 + \alpha_j T_{ij}$ and $\text{logit}(\mu_{2ij}) = \beta_0 + \beta_j T_{ij}$, where T_{ij} is a year indicator defined as $T_{ij} = 1$ if the measurement was taken in the j th year ($j = 1, 2, \dots, 8$) and 0 otherwise. Year 9 is taken as the reference year. The α and β parameters are the effect of time associated with the prevalence of HCV and HIV, respectively. To capture the correlation inherent in the data, a random

intercept is added to the conditional mean models. Thus,

$$\eta(\mu_{1ij}) = \alpha_0 + \alpha_j T_{ij} + b_{1i}, \quad \eta(\mu_{2ij}) = \beta_0 + \beta_j T_{ij} + b_{2i},$$

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

Apart from separate models, GLMM and MMM, a shared parameter version of the joint hierarchical model was formulated and fitted with the scale factor (λ) inserted into the HCV model.

As is expected, we observe from Table 2 that parameter estimates from the GLMM model are higher in magnitude than those of the marginal models (Molenberghs and Verbeke 2005), precisely the reason why MMM type models are considered. This is not surprising in the case of the HCV data given that the random effect variance is small in the GLMM. Also for HCV, the GEE and MMM model have similar estimates, albeit not in terms of their standard errors. The full likelihood MMM model enjoys higher precision of parameter estimates. However, for HIV, due to the relatively high random-effect variance, the GLMM and GEE estimates are quite more different, with the estimates from the MMM lying in between.

To turn to the joint models, results of the various joint models are presented in Table 3. We observed from these results that parameters associated with HCV and HIV in the joint models are similar to those from the separate GLMM models with the former superior in terms of precision. Also, the joint MMM models estimates are close in magnitude to the GEE and MMM models. These models have a population average interpretation for the effect of time on the prevalence of HCV and HIV. The magnitudes of the estimates from the joint hierarchical models again are higher than their marginal counterparts. The correlated version of the joint model is fitting better than the shared parameter models based on their log-likelihood or Akaike's Information Criterion derived thereof.

Note that both the hierarchical as well as the marginalized model are of relevance. The former can be used to derive regional prevalence, the latter corresponds to national prevalence. We observe that the association is stronger within the HIV sequence than within the HCV sequence. Also, there is a high correlation between the prevalence of both infections.

5.2 Analysis of the Age Related Macular Degeneration Trial

Results of fitting the various approaches to the longitudinal continuous visual acuity and the binary vision-loss outcome in the ARMD data are presented. The covariate structure is maintained across models, for ease of comparison. The models have an intercept and the effect of treatment at each time point for both outcomes. Precisely, we assume the predictors to be: $\mu_{1ij} = \beta_{0j} + \beta_{1j} T_i$ and $\text{logit}(\mu_{2ij}) = \alpha_{0j} + \alpha_{1j} T_i$, where $j = 1, 2, 3, 4$ and T_i is treatment allocation. In the corresponding conditional models, a random intercept, $b_i \sim N(0, d)$, was used. The single marginal models were fitted with a compound symmetry (exchangeable) variance or correlation structure. Also, the joint

models were fitted using the shared-parameter model where the inflation factor (λ) was introduced in the model for the continuous sequence. Results of all models fitted are presented in Table 4. We observe from these results that parameter estimates and standard errors associated with the continuous sequence are all similar. Therefore, whether or not a marginal or hierarchical formulation is employed, parameters retain both their interpretation as well as their magnitude. In the few instances where the standard errors are not exactly the same, the joint models are much more precise. On the other hand, the story for the binary sequence is different. Observe that the GEE, MMM, and the joint MMM models have marginal parameters that are numerically similar to each other, but that a higher precision is observed for the estimates from the joint MMM model. Also note that the parameters in the conditional models are not similar to those in the marginal models. This is not surprising given that in the GLMM, the random effect variance was very high. These observations are also a reflection of the fact that so-called marginally meaningful models are different from their hierarchical counterparts in terms of parameter interpretation. Furthermore, the conditionally interpreted GLMM has less similar parameters than those obtained from the joint hierarchical model. Due to these differences in estimates, the resulting standard errors seem to suggest that there is higher precision in the GLMM than the joint hierarchical model. This is, however, unwarranted. A close examination of the relative precision defined as the ratio of the standard error of the estimates and the value of the estimates reveals that the joint model yields higher relative precision than the single GLMM. This is consistent with the marginal models, where relative precision also tends to favor the joint marginalized multilevel model. In summary, our proposed joint marginal multilevel model, produces parameter estimates similar to those of GEE and the MMM which gives an indication that the JOMMM model parameters will indeed have a marginal interpretation. Finally, we again see the benefit of joint modeling against separate analyses for each outcome in their precision estimates. Those in the joint models tend to yield higher precision than those of the separate analyses.

6 Simulation Study

Further to the results from both case studies, a small simulation study was carried out. In the first part of the simulation, data were generated from a joint marginalized multilevel model with correlated random effects. The generating model for two Bernoulli outcomes was as follows:

$$\begin{aligned}\text{logit}(\pi_{1ij}^m) &= \beta_0 + \beta_1 T_i \times t_{ij}, & \Phi^{-1}(\pi_{1ij}^c) &= \delta_{1ij} + b_{1i}, \\ \text{logit}(\pi_{2ij}^m) &= \alpha_0 + \alpha_1 T_i \times t_{ij}, & \Phi^{-1}(\pi_{2ij}^c) &= \delta_{2ij} + b_{2i},\end{aligned}$$

with the random effects assumed to follow a normal distribution as in (7). T_i is a binary indicator for treatment allocation and t_{ij} is the time at which the j th measurement for subject i is taken. The true parameters were assumed to be $(\beta_0, \beta_1, \alpha_0, \alpha_1, d_{11}, d_{12}, d_{22}) = (3.8, 1.2, 2.4, 0.8, 1.22, 0.57, 1.56)$. Three marginally specified models were fitted to each of the 500 generated datasets under varying

number of subjects and measurement point per subject. Also, two treatment levels were used, in the sense that two different treatments (e.g., active treatment and placebo) are compared. The JOMMM with shared random effect, MMM, and GEE for each outcome are compared. From the results presented in Table 5, we observed the semblance between the three models in terms of their parameter estimates. In terms of relative precision, the effect of the shared joint MMM model is not extremely pronounced when compared with the MMM model. However, the added advantage to fitting such a joint model is the prospect of capturing the association arising from a joint collection of outcomes in practice. The GEE with exchangeable working assumption was also found to have parameter estimates rather close to the true.

In the second part of the simulation study, we investigated the performance of the shared JOMMM model and the MMM model when a common treatment evolution is assumed (i.e., $\beta_1 = \alpha_1$) for the two longitudinal outcomes. The same simulation model above was used but with a common treatment effect and different true model parameters. Each of the two models was fitted to 500 simulated datasets. Results are presented in Table 6. The true values employed are: $\beta_0 = 1.8$, $\beta_1 = -3.2$, $\alpha_0 = 2.4$, $\alpha_1 = -3.2$, $d_{11} = 1.22$, and $d_{22} = 1.56$. One can observe from these results that the shared joint MMM model reduced bias and increased precision when compared to the single MMM models. Of course, the single MMM is only modestly biased in the first place. It is therefore apparent to conclude that when the true treatment effect is common to the two outcomes, a single analysis per outcome may fail to capture this effect accurately and precisely.

7 Concluding Remarks

We have shown that a joint longitudinal model can be formulated where all parameters enjoy a marginal interpretation. This was achieved by incorporating the model of Heagerty (1999) into the shared-parameter and hierarchical joint models used to jointly model two longitudinal outcomes. Two analyses have been performed to show how these models are fitted. The resulting model at the same time captures association between the two responses, and yields parameter estimates that have a population-averaged interpretation for both outcomes. Estimates were found to be close to those from single-outcome analyses but provided higher precision. The difference in precision could affect inferences. Thus, it is important to make use of such joint modeling approaches, which tend to provide unbiased and more precise estimates. Note that in both real data examples, only a random intercept was used in the sub-models. This is not to say the model is merely restricted to one-dimensional random effect. It is indeed possible to add as many random effects as it is practicable. The user ought to be reminded that even for single response models, adding more random effects increases the complexity of the model and thus difficult or impossible to fit in some cases. In terms of implementation, the proposed method allows to efficiently make use of available resources, such as the SAS procedure NLMIXED. Little additional coding effort is required; see the Appendix for the

SAS code to the most complex of the models used to analyze the HCV and HIV data. The code is available at the authors' web pages.

Acknowledgment

The authors gratefully acknowledge the financial support from the IAP research Network P6/03 of the Belgian Government (Belgian Science Policy).

References

- Burzykowski, T., Molenberghs, G., and Buyse, M. (2005) *The Evaluation of Surrogate Endpoints*. New York: Springer.
- Buyse, M., and Molenberghs, G. (1998) The validation of surrogate endpoints in randomized experiments. *Biometrics*, **54**, 1014–1029.
- Clayton, D.G. (1992) Models for the analysis of cohort and case-control studies with inaccurately measured exposures. In *Statistical Models for Longitudinal Studies of Health*, eds. J.H. Dwyer, P. Lippert, M. Feinleib, and H. Hoffmeister, New York: Oxford University Press.
- Cooper, D.M., and Thompson, R. (1977) A note on the estimation of the parameters of the autoregressive-moving average process. *Biometrika*, **64**, 625.
- DeGruttola, V., and Tu, X.M. (1994) Modeling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics*, **50**, 1003–1014.
- Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society*, **39**, 1–38.
- Diggle, P.J., Heagerty, P.J., Liang, K.-Y., and Zeger, S.L. (2002) *Analysis of Longitudinal Data (2nd ed.)*, Oxford Science Publications. Oxford: Clarendon Press.
- Faucett, C.J., and Thomas, D.C. (1996) Simultaneously modeling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Statistics in Medicine*, **15**, 1663–1685.
- Fitzmaurice, G.M., Davidian, M., Molenberghs, G., and Verbeke, G. (2009) *Longitudinal Data Analysis*. Handbooks of Modern Statistical Methods. New York: Chapman & Hall/CRC.
- Fitzmaurice, G. M. and Laird, N. M. (1993) A likelihood based method for analysing longitudinal binary responses. *Biometrika*. **80**, 141–151.
- Gelman, A., Carlin, J.B, Stern, H.S., and Rubin, D.B (1995) *Bayesian Data Analysis*, Texts in Statistical Science. London: Chapman& Hall.

- Griswold, M.E. and Zeger, S.L. On marginalized multilevel models and their computation (November 2004). *Johns Hopkins University, Dept. of Biostatistics Working Papers. Working Paper 99*.
- Harville, D.A. (1977) Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, **72**, 320.
- Heagerty, P.J. (1999) Marginally specified logistic-normal models for longitudinal binary data. *Biometrics*, **55**, 688–698.
- Heagerty, P.J. and Zeger, S.L. (2000) Marginalized multilevel models and likelihood inference (with comments and a rejoinder by the authors). *Statistical Science*, **15**, 1–26.
- Horrocks, J., and van den Heuvel, M.J. (2009) Prediction of pregnancy: A joint model for longitudinal and binary data. *Bayesian Analysis*, **4**, 523–538.
- Laird, N.M., and Ware, J.H. (1982) Random effects models for longitudinal data. *Biometrics*, **38**, 963–974.
- Li, E., Wang, N., and Wang, S. (2007) Joint models for a primary endpoint and multiple longitudinal covariate processes. *Biometrics*, **63**, 1068–1078.
- Liang, K.-Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.
- Lindstrom, M.J., and Bates, D.M. (1988) Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, **83**, 1014.
- Mathei, C., Shkedy, Z., Denis, B., Kabali, C., Aerts, M., Molenberghs, G., Van Damme, P., and Buntinx, F. (2006). Evidence for a substantial role of sharing of injection paraphernalia other than syringes/needles to the spread of hepatitis C among injecting drug users. *Journal of Viral Hepatitis*, **13**, 560–570.
- Molenberghs, G. and Lesaffre, E. (1994) Marginal modelling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association*, **89**, 633–644.
- Molenberghs, G. and Verbeke, G. (2005) *Models for Discrete Longitudinal Data*. New York: Springer.
- Patterson, H.D., and Thompson, R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**, 545.

- Pharmacological Therapy for Macular Degeneration Study Group (1997) Interferon α -IIA is ineffective for patients with choroidal neovascularization secondary to age-related macular degeneration. Results of a prospective randomized placebo-controlled clinical trial. *Archives of Ophthalmology* **115**, 865–872.
- Pinheiro, J.C., and Bates, D.M. (1995) Approximations to the log-likelihood function in the non-linear mixed-effects model. *Journal of Computation and Graphical Statistics*. **4**, 12–35.
- Pinheiro, J.C., and Bates, D.M. (2000) *Mixed Effects Models in S and S-Plus*. New York: Springer-Verlag.
- Prentice, R.L. (1988) Correlated binary regression with covariates specific to each binary observation. *Biometrics*, **47**, 825–839.
- Tsiatis, A.A., and Davidian, M. (2004) A joint modeling of longitudinal and time-to-event data: An overview. *Stat Sin*, **14**, 809–834.
- Tsiatis, A.A., DeGruttola, V., and Wulfsohn, M.L. (1995) Modeling the relationship of survival and longitudinal data measured with error, Application to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association*, **90**, 27–37.
- Verbeke, G. and Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.
- Verbeke, G. and Molenberghs, G. (2003) The use of score tests for inference on variance components. *Biometrics*, **59**, 254–262.
- Verbyla, A.P. (1990) A conditional derivation of residual maximum likelihood. *The Australian Journal of Statistics*, **32**, 227.
- Wang, C.Y., Wang, N., and Wang, S. (2000) Regression analysis when covariates are regression parameters of a random effects model. *Biometrics*, **56**, 487–495.
- Wulfsohn, M.S., and Tsiatis, A.A. (1997) A joint model for survival and longitudinal data measured with error. *Biometrics*, **53**, 330–339.
- Zeger, S.L. and Liang, K.-Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42**, 121–130.
- Zeger, S.L., Liang, K.-Y., and Albert, P.S. (1988) Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, **44**, 1049–1060.

Table 1: HCV data. Overall prevalences of HCV per Italian region, 1998–2006.

Region	Prevalence	Region	Prevalence
Abruzzo	0.56	Molise	0.67
Basilicata	0.66	Piemonte	0.73
Calabria	0.53	Puglia	0.59
Campania	0.44	Sardegna	0.80
Emilia Romagna	0.84	Sicilia	0.61
Friuli Venezia Giulia	0.75	Toscana	0.68
Lazio	0.64	Trentino Alto Adige	0.86
Liguria	0.77	Umbria	0.63
Lombardia	0.68	Valle d'Aosta	0.48
Marche	0.62	Veneto	0.66

Table 2: Independent models for HIV and HCV prevalence in Italy. Parameter estimates (standard errors). RE: random effect.

Effect	Par.	Hierarchical models		Marginal models			
		GLMM		GEE		MMM	
		HCV	HIV	HCV	HIV	HCV	HIV
Intercept	α_0, β_0	0.5921(0.1124)	-2.1293(0.1839)	0.5088(0.1487)	-2.1474(0.2278)	0.5586(0.1049)	-1.9018(0.1716)
Time 1	α_1, β_1	0.2226(0.0111)	0.0197(0.0156)	0.2268(0.0796)	0.1180(0.1429)	0.2074(0.0110)	0.0283(0.0148)
Time 2	α_2, β_2	0.2091(0.0111)	-0.0283(0.0159)	0.2132(0.0685)	0.0519(0.1536)	0.1976(0.0110)	-0.0110(0.0151)
Time 3	α_3, β_3	0.2878(0.0111)	0.0432(0.0159)	0.2864(0.0629)	0.1064(0.1372)	0.2748(0.0114)	0.0417(0.0151)
Time 4	α_4, β_4	0.1792(0.0109)	-0.0035(0.0159)	0.1800(0.0574)	0.0507(0.1179)	0.1694(0.0107)	0.0021(0.0150)
Time 5	α_5, β_5	0.1060(0.0108)	-0.0081(0.0160)	0.1143(0.0531)	0.0410(0.1407)	0.1013(0.0104)	-0.0078(0.0151)
Time 6	α_6, β_6	0.1138(0.0107)	-0.0423(0.0161)	0.1199(0.0521)	-0.0006(0.1270)	0.1091(0.0103)	-0.0402(0.0153)
Time 7	α_7, β_7	0.0722(0.0107)	-0.0443(0.0163)	0.0692(0.0545)	-0.0268(0.1072)	0.0704(0.0102)	-0.0433(0.0155)
Time 8	α_8, β_8	-0.0369(0.0108)	-0.0776(0.0164)	-0.0276(0.0314)	-0.0494(0.0885)	-0.0350(0.0103)	-0.0742(0.0158)
RE Var.	d	0.2514(0.0796)	0.6710(0.2149)			0.0902(0.0286)	0.1802(0.05743)
Corr.	ρ			0.8377	0.9066		
-2 log-likelihood		9973				9987	

Table 3: Joint models of HIV and HCV prevalence in Italy. Parameter estimates (standard errors).
RE: random effect.

Effect	Par.	Joint hierarchical models				Joint marginalized multilevel model			
		Shared RE		Correlated RE		Shared RE		Correlated RE	
		HCV	HIV	HCV	HIV	HCV	HIV	HCV	HIV
Intercept	α_0, β_0	0.5033 (0.0902)	-1.9939 (0.1741)	0.5909 (0.1125)	-2.1249 (0.1823)	0.4839 (0.0895)	-1.8833 (0.1684)	0.5574 (0.1050)	-1.9008 (0.1708)
Time 1	α_1, β_1	0.2044 (0.0110)	0.0193 (0.0155)	0.2226 (0.0111)	0.0110 (0.0156)	0.1943 (0.0108)	0.0337 (0.0154)	0.2073 (0.0110)	0.0284 (0.0148)
Time 2	α_2, β_2	0.1909 (0.0110)	-0.0226 (0.0159)	0.2091 (0.0111)	-0.0283 (0.0159)	0.1831 (0.0108)	-0.0121 (0.0158)	0.1976 (0.0110)	-0.0199 (0.0151)
Time 3	α_3, β_3	0.2734 (0.0110)	0.0477 (0.0159)	0.2878 (0.0111)	0.0433 (0.0159)	0.2629 (0.0110)	0.0496 (0.0158)	0.2747 (0.0114)	0.0417 (0.0151)
Time 4	α_4, β_4	0.1669 (0.0108)	-0.0007 (0.0158)	0.1792 (0.0109)	-0.0034 (0.0159)	0.1592 (0.0105)	0.0077 (0.0157)	0.1693 (0.0107)	0.0022 (0.0150)
Time 5	α_5, β_5	0.1008 (0.0107)	-0.0069 (0.0159)	0.1060 (0.0108)	-0.0080 (0.0160)	0.0974 (0.0103)	-0.0042 (0.0158)	0.1013 (0.0104)	-0.0077 (0.0151)
Time 6	α_6, β_6	0.1086 (0.0107)	-0.0390 (0.0160)	0.1138 (0.0107)	-0.0422 (0.0161)	0.1048 (0.0103)	-0.0361 (0.0160)	0.1091 (0.0103)	-0.0401 (0.0153)
Time 7	α_7, β_7	0.0685 (0.0106)	-0.0417 (0.0163)	0.0722 (0.0107)	-0.0443 (0.0163)	0.0673 (0.0102)	-0.0406 (0.0162)	0.0704 (0.0102)	-0.0433 (0.0155)
Time 8	α_8, β_8	-0.0426 (0.0108)	-0.0750 (0.0164)	-0.0369 (0.0108)	-0.0776 (0.0164)	-0.0403 (0.0103)	-0.0737 (0.0164)	-0.0350 (0.0103)	-0.0743 (0.0158)
Scale	λ	1.9338 (0.0176)				1.6222 (0.0137)			
RE Var.	d	0.1613 (0.0511)		0.2517 (0.0798)	0.6589 (0.2107)	0.0646 (0.0205)		0.0903 (0.0286)	0.1784 (0.0568)
Corr.	ρ			0.7868 (0.0862)				0.7906 (0.0846)	
-2 log-likelihood		17181		9954		17221		9967	

Table 4: *The Age Related Macular Degeneration Trial: Comparison of joint and separate models for continuous and binary visual acuity sequences. Parameter estimates (standard errors). RE: random effect.*

Effect	Par.	Continuous sequence				Binary sequence				
		(1) Marg. Corr.	(2) Sep. hier.	(6) Joint hier.	(7) Joint MMM	(3) GEE	(4) GLMM	(5) MMM	(6) Joint hier.	(7) Joint MMM
Int. 4	β_{04}, α_{04}	-3.2353 (0.8074)	-3.2526 (1.2968)	-3.2675 (1.2964)	-3.2682 (1.2964)	1.0151 (0.2442)	1.7424 (0.4204)	0.9965 (0.2375)	2.0111 (0.4625)	1.0346 (0.2326)
Int. 12	$\beta_{0,12}, \alpha_{0,12}$	-4.6207 (1.0707)	-4.6207 (1.2922)	-4.6207 (1.2922)	-4.6207 (1.2922)	0.9083 (0.2369)	1.5610 (0.4108)	0.9103 (0.2372)	1.8151 (0.4535)	0.9280 (0.2307)
Int. 24	$\beta_{0,24}, \alpha_{0,24}$	-8.3678 (1.2634)	-8.3678 (1.2922)	-8.3678 (1.2922)	-8.3678 (1.2922)	1.1451 (0.2505)	1.9534 (0.4294)	1.1496 (0.2515)	2.2375 (0.4707)	1.1648 (0.2454)
Int. 52	$\beta_{0,52}, \alpha_{0,52}$	-15.1609 (1.6414)	-15.1609 (1.2922)	-15.1609 (1.2922)	-15.1609 (1.2922)	1.6514 (0.2918)	2.7633 (0.4805)	1.6308 (0.2867)	3.1073 (0.5200)	1.6296 (0.2792)
Trt. 4	$\beta_{0,4}, \alpha_{0,4}$	2.3323 (1.0551)	2.3356 (1.7609)	2.3864 (1.7602)	2.3858 (1.7603)	-0.4187 (0.3198)	-0.6724 (0.5395)	-0.3732 (0.3143)	-0.6879 (0.5922)	-0.3568 (0.3045)
Trt. 12	$\beta_{0,12}, \alpha_{0,12}$	2.3391 (1.5220)	2.3391 (1.7551)	2.3391 (1.7551)	2.3391 (1.7551)	-0.5351 (0.3104)	-0.8769 (0.5305)	-0.5120 (0.3114)	-0.9260 (0.5851)	-0.4774 (0.3036)
Trt. 24	$\beta_{0,24}, \alpha_{0,24}$	2.8338 (1.8422)	2.8338 (1.7551)	2.8338 (1.7551)	2.8338 (1.7551)	-0.5240 (0.3248)	-0.8392 (0.5442)	-0.5013 (0.3260)	-0.8757 (0.5956)	-0.4686 (0.3176)
Trt. 52	$\beta_{0,52}, \alpha_{0,52}$	4.1221 (2.3078)	4.1221 (1.7551)	4.1221 (1.7551)	4.1221 (1.7551)	-0.4049 (0.3756)	-0.6077 (0.5922)	-0.3651 (0.3689)	-0.6214 (0.6376)	-0.3382 (0.3556)
Res. Var.	σ^2	-	70.9225 (4.2076)	67.4893 (3.8163)	67.7113 (3.8292)	-	-	-	-	-
RE Var.	d	-	74.3544 (9.5060)	-	-	-	4.7951 (1.1643)	1.5858 (0.3712)	-	-
Inflation	λ	-	-	-3.3122 (0.3379)	-5.8694 (0.5697)	-	-	-	-	-
-2 log-likelihood		5480.73	5168.5	-	-	-	773.4	773.4	-	-
Common parameters in joint models										
RE Var.	d			7.0905 (1.5485)	2.2515 (0.4695)					
-2log-likelihood				5744.7	5744.6					

Table 5: *Simulation Study. Results comparing shared JOMMM, MMM, and GEE, based on 500 simulations. (Relative precision: ratio of standard error to estimate; Relative bias: ratio of bias to parameter value.)*

N	Par.	Shared JOMMM			MMM			GEE		
		Estimates	Rel. Bias	Rel. Precision	Estimates	Rel. Bias	Rel. Precision	Estimates	Rel. Bias	Rel. Precision
For 5 time points										
200	β_0	3.7513	-0.01	0.109	3.8112	0.00	0.105	3.7880	-0.00	0.102
	β_1	1.0879	-0.09	0.605	1.0570	-0.12	0.614	1.0663	-0.11	0.465
	α_0	2.3944	-0.00	0.102	2.4037	0.00	0.102	2.3955	-0.00	0.101
	α_1	0.9084	0.14	0.331	0.9331	0.17	0.328	0.9368	0.17	0.304
	d_{11}	1.3285	0.09	0.423	1.0953	-0.10	0.502			
	d_{22}				1.5643	0.00	0.214			
	λ	1.3986		0.474						
500	β_0	3.8182	0.01	0.069	3.8083	0.00	0.069	3.8183	0.01	0.068
	β_1	1.3073	0.09	0.407	1.3142	0.10	0.405	1.2912	0.08	0.374
	α_0	2.3996	-0.00	0.064	2.4004	0.00	0.064	2.4054	0.00	0.064
	α_1	0.8502	0.06	0.204	0.845	0.06	0.204	0.8328	0.04	0.207
	d_{11}	1.2514	0.03	0.251	1.2511	0.03	0.251			
	d_{22}				1.5618	0.00	0.131			
	λ	1.3216		0.283						
For 15 time points										
200	β_0	3.8146	0.00	0.093	3.7905	-0.00	0.091	3.7569	-0.01	0.090
	β_1	1.0437	-0.13	0.558	1.0954	-0.09	0.561	1.1025	-0.08	0.446
	α_0	2.4076	0.00	0.093	2.4035	0.00	0.093	2.4106	0.00	0.092
	α_1	0.8845	0.11	0.299	0.9097	0.14	0.300	0.8239	0.03	0.301
	d_{11}	1.2455	0.02	0.255	1.2233	0.00	0.261			
	d_{22}				1.554	-0.00	0.152			
	λ	1.353		0.298						
500	β_0	3.8222	0.01	0.058	3.8075	0.00	0.058	3.8336	0.01	0.057
	β_1	1.253	0.04	0.383	1.3351	0.11	0.391	1.104	-0.08	0.342
	α_0	2.4031	0.00	0.059	2.3992	-0.00	0.059	2.3745	-0.01	0.060
	α_1	0.8515	0.06	0.188	0.8494	0.06	0.188	0.8497	0.06	0.208
	d_{11}	1.2056	-0.01	0.153	1.2242	0.00	0.153			
	d_{22}				1.5747	0.01	0.095			
	λ	1.3293		0.181						

Table 6: *Simulation Study. Results comparing shared JOMMM and MMM when a common treatment evolution is assumed based on 500 simulations. (Relative precision: ratio of standard error to estimate; Relative bias: ratio of bias to parameter value.)*

N	Par.	Shared JOMMM			MMM		
		Estimates	Rel. Bias	Rel. Precision	Estimates	Rel. Bias	Rel. Precision
For 5 time points							
200	β_0	1.8030	0.00	0.098	1.8415	0.02	0.107
	β_1	-3.2247	0.01	-0.066	-3.2623	0.02	-0.092
	α_0	2.4121	0.01	0.088	2.4448	0.02	0.102
	α_1				-3.2475	0.02	-0.092
	σ_{11}	1.2336	0.01	0.178	1.2234	0.00	0.180
	σ_{22}				1.5367	-0.02	0.224
	λ	1.3538		0.278			
500	β_0	1.8132	0.01	0.0615	1.8015	0.00	0.068
	β_1	-3.2157	0.01	-0.0414	-3.2164	0.01	-0.059
	α_0	2.4103	0.00	0.0553	2.4205	0.01	0.065
	α_1				-3.2381	0.01	-0.059
	σ_{11}	1.2272	0.01	0.1106	1.2354	0.01	0.111
	σ_{22}				1.5686	0.01	0.137
	λ	1.2923		0.1735			
For 20 time points							
200	β_0	1.8165	0.01	0.088	1.8111	0.01	0.096
	β_1	-3.2331	0.01	-0.063	-3.2388	0.01	-0.089
	α_0	2.4199	0.01	0.081	2.4363	0.02	0.093
	α_1				-3.2474	0.02	-0.088
	σ_{11}	1.2172	-0.00	0.114	1.2145	-0.01	0.116
	σ_{22}				1.5599	0.00	0.147
	λ	1.3004		0.181			
500	β_0	1.8029	0.00	0.056	1.8241	0.01	0.060
	β_1	-3.2042	0.00	-0.040	-3.2417	0.01	-0.056
	α_0	2.4008	0.00	0.052	2.4202	0.01	0.059
	α_1				-3.2284	0.01	-0.056
	σ_{11}	1.2127	-0.01	0.072	1.2116	-0.01	0.073
	σ_{22}				1.565	0.00	0.093
	λ	1.2938		0.113			

Figure 1: *HCV and HIV data. Region-specific prevalence profiles, 1998–2006.*

Appendix

A SAS Code

A.1 Joint Hierarchical Model with Correlated Random Effects for the HCV and HIV Data

```
proc nlmixed data=italy.nlmldata qpoints=10 maxiter=100 technique=newrap;
title 'Unstructured means - correlated';
parms alpha0=0.59 alpha1=0.22 alpha2=0.21 alpha3=0.29 alpha4=0.18 alpha5=0.11
      alpha6=0.11 alpha7=0.07 alpha8=-0.04
      beta0=-2.39 beta1=0.27 beta2=0.13 beta3=0.15 beta4=0.08 beta5=0.04
      beta6=0.01 beta7=-0.07 beta8=-0.02 var1=0.25
var2=0.57 cov12=0.3;
if hcv=1 then do;
eta = alpha0+alpha1*time1+alpha2*time2+alpha3*time3+alpha4*time4+alpha5*time5
      +alpha6*time6+alpha7*time7+alpha8*time8+a;
p = exp(eta)/(1+exp(eta));
end;
if hiv=1 then do;
eta = beta0+beta1*time1+beta2*time2+beta3*time3+beta4*time4+beta5*time5+beta6*time6
+beta7*time7+beta8*time8+b;
p = exp(eta)/(1+exp(eta));
end;
model events ~ binomial(trials,p);
random a b ~ normal([0,0],[var1,cov12,var2]) subject=geoarea;
estimate 'rho' cov12/(sqrt(var1)*sqrt(var2));
run;
```

A.2 Joint Marginal Multilevel Model with Correlated Random Effects for the HCV and HIV Data

```
proc nlmixed data=italy.nlmldata qpoints=10 maxiter=100 technique=newrap;
title 'Unstructured means - fixed effects';
parms alpha0=0.59 alpha1=0.22 alpha2=0.21 alpha3=0.29 alpha4=0.18 alpha5=0.11
      alpha6=0.11 alpha7=0.07 alpha8=-0.04
      beta0=-2.39 beta1=0.27 beta2=0.13 beta3=0.15 beta4=0.08 beta5=0.04
      beta6=0.01 beta7=-0.07 beta8=-0.02 var1=0.25
```

```

var2=0.57 cov12=0.3;
if hcv=1 then do;
eta = alpha0+alpha1*time1+alpha2*time2+alpha3*time3+alpha4*time4+alpha5*time5
      +alpha6*time6+alpha7*time7+alpha8*time8;
pim = exp(eta)/(1+exp(eta));
delta = sqrt(1+(var1)) * probit(pim);
pic = probnorm(delta+a);
ll=lgamma(trials+1)-lgamma(trials-events+1)-lgamma(events+1)+events*log(pic)
    +(trials-events)*log(1-pic) ;
end;
if hiv=1 then do;
eta = beta0+beta1*time1+beta2*time2+beta3*time3+beta4*time4+beta5*time5
      +beta6*time6+beta7*time7+beta8*time8;
pim = exp(eta)/(1+exp(eta));
delta = sqrt(1+(var2)) * probit(pim);
pic = probnorm(delta+b);
ll=lgamma(trials+1)-lgamma(trials-events+1)-lgamma(events+1)+events*log(pic)
    +(trials-events)*log(1-pic) ;
end;
model events~general(ll);
random a b ~ normal([0,0],[var1,cov12,var2]) subject=geoarea;
estimate 'rho' cov12/(sqrt(var1)*sqrt(var2));
run;

```