GEERT MOLENBERGHS
DIDIER RENARD
GEERT VERBEKE

A review of generalized linear mixed models

A REVIEW OF GENERALIZED LINEAR MIXED MODELS

Geert MOLENBERGHS *, Didier RENARD *,
Geert VERBEKE **

ABSTRACT

A general framework for modeling repeated categorical data is pictured, with three main model families : marginal, conditional, and subject-specific. The primary focus is on subject-specific or random-effects model, with some emphasis on the generalized linear mixed model. Estimation and optimization algorithms are discussed, together with available software. Advantages and disadvantages are pointed out. These tools have been exemplified using a simple but illustrative analysis. Similarities and differences between linear mixed models and generalized linear mixed models are discussed in detail.

Keywords : Conditional Model, Longitudinal data, Linear Mixed Models, Marginal Model, Random Effects, Repeated Measures, Subject-specific Parameter.

RÉSUMÉ

Un cadre général pour la modélisation de données catégorielles répétées est présenté. Trois familles principales de modèles sont introduites : les modèles marginaux, conditionnels et spécifiques au sujet. On traite principalement des modèles spécifiques au sujet, ou à effets aléatoires, en insistant notamment sur le modèle mixte linéaire généralisé. On discute l'estimation et les algorithmes d'optimisation, ainsi que les logiciels disponibles. Les avantages et désavantages sont également indiqués. On utilise ces outils pour présenter une analyse simple mais illustrative. Les similarités et les différences entre les modèles mixtes linéaires et les modèles mixtes linéaires généralisés sont discutées en détail.

Mots clés :

1. Introduction

In applied sciences, one is often confronted with the collection of *correlated data* or otherwise hierarchical data. This generic term embraces a multitude of data structures, such as multivariate observations, clustered data, repeated measurements, longitudinal data, and spatially correlated data. In particular, studies are often designed to investigate changes in a specific parameter which

* Center for Statistics, Limburgs Universitair Centrum, Universitaire Campus, B-3590 Diepenbeek ; Email : geert.molenberghs@luc.ac.be
** Biostatistical Centre, Catholic University of Leuven, U.Z. St.-Rafaël, Kapucijnenvoer 35, B-3000 Leuven.

is measured repeatedly over time in the participating subjects. This is in contrast to cross-sectional studies where the response of interest is measured only once for each individual. Longitudinal studies are most appropriate for the investigation of such changes, together with the evolution of relevant covariates.

A very important characteristic of data to be analyzed is the type of outcome. Methods for continuous longitudinal data form no doubt the best developed and most advanced body of research (Verbeke and Molenberghs, 2000); the same is true for software implementation. This is natural, since the special status and the elegant properties of the normal distribution simplify model building and ease software development. A number of software tools, such as the SAS procedure MIXED, the S-PLUS function lme and MLwiN have been developed in this area. However, also categorical (nominal, ordinal, and binary) and discrete outcomes are very prominent in statistical practice. For example, quality-of-life outcomes are often scored on binary or ordinal scales. In many surveys, all or part of the information is recorded on a categorical scale.

Two fairly different views can be adopted. The first one, supported by large-sample results, states that normal theory should be applied as much as possible, even to non-normal data such as ordinal scores and counts. A different view is that each type of outcome should be analyzed using instruments that exploit the nature of the data. Extensions of generalized linear models to the longitudinal case are discussed in Diggle, Liang, and Zeger (1994), where the main emphasis is on generalized estimating equations (Liang and Zeger, 1986). Generalized linear mixed models have been proposed and/or studied by, for example, Stiratelli, Laird, and Ware (1984), Wolfinger and O'Connell (1993), and Breslow and Clayton (1993). Fahrmeir and Tutz (1994) devote an entire book to generalized linear models for multivariate settings. Subscribing to the second point of view, this review will present, discuss, and illustrate methodology specific to the case of non-continuous data.

In longitudinal settings, each unit typically has a *vector* Y of responses. This leads to several, generally nonequivalent, extensions of univariate models. In a *marginal model*, marginal distributions are used to describe the outcome vector Y, given a set X of predictor variables. The correlation among the components of Y can then be captured either by adopting a fully parametric approach or by means of working assumptions, such as in the semiparametric approach of Liang and Zeger (1986). Alternatively, in a *random-effects model*, the predictor variables X are supplemented with a vector b of random (or random effects) effects, conditional upon which the components of Y are usually assumed to be independent. This does not preclude that more elaborate models are possible if residual dependence is detected (Longford 1993). Finally, a *conditional model* describes the distribution of the components of Y, conditional on X but also conditional on (a subset of) the other components of Y. Well-known members of this class of models are log-linear models (Gilula and Haberman 1994).

A more elaborate sketch of the different model families is provided in Section 2. Random-effects models, and in particular the generalized linear mixed model, are discussed in Section 3. Specific attention is devoted to fitting algorithms (Section 4), as well as to software tools, illustrated by a case study (Section 5).

2. Model Families

Among the clustered data settings, longitudinal data perhaps require the most elaborate modeling of the random variability. Diggle, Liang, and Zeger (1994) and Verbeke and Molenberghs (2000) distinguish among three components of variability. The first one groups traditional random effects (as in a random-effects ANOVA model) and random coefficients (Longford 1993). It stems from interindividual variability (*i.e.*, heterogeneity between individual profiles). The second component, serial association, is present when residuals close to each other in space or time are more similar than residuals further apart. This notion is well known in the spatial and time-series literatures (Ripley 1981, Diggle 1983, Cressie 1991; Diggle 1990). Finally, in addition to the other two components, there is potentially also measurement error. This results from the fact that, for delicate measurements (e.g., laboratory assays), even immediate replication will not be able to avoid a certain level of variation. In longitudinal data, these three components of variability can be distinguished by virtue of both *replication* as well as a clear *distance* concept (time), one of which is lacking in classical spatial and time-series analysis and in clustered data.

This structure form the basis for the construction of the three modeling families referred to earlier : conditionally specified models, marginal models, and random effects models. The distinction between these is much less important for normally distributed outcomes than it is in this context. Indeed, for normally distributed data, marginal models can easily be fitted, for example, with the SAS procedure MIXED, the S-PLUS function lme, or within the MLwiN package. For such data, integrating a mixed-effects model over the random effects produces a marginal model, in which the regression parameters retain their meaning and the random effects contribute in a simple way to the variance-covariance structure. For example, the marginal model corresponding to a random-intercepts model is a compound symmetry model that can be fitted without explicitly acknowledging the random-intercepts structure. A compound symmetry model is defined by a covariance matrix with common variance σ^2 and common covariance $\rho\sigma^2$. In the same vein, certain types of transition models induce simple marginal covariance structures. For example, some first-order stationary autoregressive models imply an exponential or AR(1) covariance structure. As a consequence, many marginal models derived from random-effects and transition models can be fitted with mixed-models software. For more details, see Verbeke and Molenberghs (2000); see also Section 3.2.

The above elegant properties of normal models do not extend to the general case of non-normally distributed repeated measures. For example, opting for a marginal model for repeated binary data precludes the researcher from

answering conditional and transitional questions in terms of simple model parameters. This implies that each model family requires its own specific analysis and, consequently, software tools. In many cases, standard maximum likelihood analyses are prohibitive in terms of computational requirements. Therefore, specific methods such as generalized estimating equations (Liang and Zeger 1986) and pseudo-likelihood (Aerts *et al.* 2002) have been developed. Both apply to marginal models, whereas pseudo-likelihood methodology can be used in the context of conditional models as well. In case random-effects models are used, the likelihood function involves integration over the random-effects distribution for which generally no closed forms are available. Estimation methods then either employ approximations to the likelihood or score functions, or resort to numerical integration techniques (see Section 4).

These considerations imply it is important to reflect on which model family is going to be selected for analysis. In conditionally-specified models the probability of a positive response for one member of the cluster is modeled conditionally upon other outcomes for the same subject, while marginal models relate the covariates directly to the marginal probabilities. Random effects models differ from the two previous models by the inclusion of parameters which are specific to the cluster. What method is used to fit the model, should not only depend on the assumptions the investigator is willing to make, but also (to some extent) on the availability of computational algorithms. In the remainder of this chapter, we will briefly describe the marginal and conditional families. Subsequent chapters are devoted to random effects models.

2.1. Marginal Models

In marginal models, the parameters characterize the marginal probabilities of a subset of the outcomes, without conditioning on the other outcomes. Advantages and disadvantages of conditional and marginal modeling have been discussed in Diggle, Liang and Zeger (1994), and Fahrmeir and Tutz (1994).

Bahadur (1961) proposed a marginal model, accounting for the association via marginal correlations. This model has also been studied by Cox (1972), Kupper and Haseman (1978) and Altham (1978). The general form of the Bahadur model requires the specification of a number of parameters, exponential in the number of measurements per subject, often prohibiting its use.

Let us specify the Bahadur model in some detail. Assume the binary response Y_{ij} indicate if member j of unit i has the response under investigation. The marginal distribution of Y_{ij} is Bernoulli with $E(Y_{ij}) = P(Y_{ij} = 1) \equiv \pi_{ij}$. Next, to describe the association between binary outcomes, the pairwise probability $P(Y_{ij} = 1, Y_{ik} = 1) = E(Y_{ij}Y_{ik}) \equiv \pi_{ijk}$ has to be characterized. This joint probability of two members of the same unit can be modeled in terms of the two marginal probabilities π_{ij} and π_{ik}, as well as a marginal correlation coefficient :

$$\text{Corr}(Y_{ij}, Y_{ik}) \equiv \rho_{ijk} = \frac{\pi_{ijk} - \pi_{ij}\pi_{ik}}{[\pi_{ij}(1 - \pi_{ij})\pi_{ik}(1 - \pi_{ik})]^{1/2}}.$$

In terms of this association parameter, the joint probability π_{ijk} can then be written as

$$\pi_{ijk} = \pi_{ij}\pi_{ik} + \rho_{ijk}[\pi_{ij}(1 - \pi_{ij})\pi_{ik}(1 - \pi_{ik})]^{1/2}.$$

Hence, given the marginal correlation coefficient ρ_{ijk} and the univariate probabilities π_{ij} and π_{ik}, the pairwise probability π_{ijk} can easily be calculated.

The first and second moments of the distribution have been specified. However, a likelihood-based approach requires the complete representation of the joint probabilities of the vector of binary responses in each litter. The full joint distribution $f(\boldsymbol{y})$ of $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^t$ is multinomial with a 2^{n_i} probability vector. Let

$$\varepsilon_{ij} = \frac{Y_{ij} - \pi_{ij}}{\sqrt{\pi_{ij}(1 - \pi_{ij})}} \quad \text{and} \quad e_{ij} = \frac{y_{ij} - \pi_{ij}}{\sqrt{\pi_{ij}(1 - \pi_{ij})}},$$

where y_{ij} is an actual value of the binary response variable Y_{ij}. Further, let $\rho_{ijk} = E(\varepsilon_{ij}\varepsilon_{ik})$, $\rho_{ijk\ell} = E(\varepsilon_{ij}\varepsilon_{ik}\varepsilon_{i\ell})$, ..., $\rho_{i12\ldots n_i} = E(\varepsilon_{i1}\varepsilon_{i2}\ldots\varepsilon_{in_i})$. Then, the general Bahadur model can be represented by the expression $f(\boldsymbol{y}_i) = f_1(\boldsymbol{y}_i)c(\boldsymbol{y}_i)$, where

$$f_1(\boldsymbol{y}_i) = \prod_{j=1}^{n_i} \pi_{ij}^{y_{ij}}(1 - \pi_{ij})^{1-y_{ij}}$$

and

$$c(\boldsymbol{y}_i) = 1 + \sum_{j<k} \rho_{ijk}e_{ij}e_{ik} + \sum_{j<k<\ell} \rho_{ijk\ell}e_{ij}e_{ik}e_{i\ell} + \ldots + \rho_{i12\ldots n_i}e_{i1}e_{i2}\ldots e_{in_i}.$$

Thus, the probability mass function is the product of the independence model $f_1(\boldsymbol{y}_i)$ and the correction factor $c(\boldsymbol{y}_i)$. The factor $c(\boldsymbol{y}_i)$ can be viewed as a model for overdispersion.

A drawback of the Bahadur approach is the existence of severe constraints on the correlation parameter space. A general study of this phenomenon is given in Declerck, Aerts and Molenberghs (1998). Molenberghs and Lesaffre (1994) and Lang and Agresti (1994) have proposed models which parameterize the association in terms of marginal odds ratios. Dale (1986) defined the bivariate global odds ratio model, based on a bivariate Plackett distribution (Plackett 1965). Molenberghs and Lesaffre (1994, 1999) extended this model to multivariate ordinal outcomes. They generalize the bivariate Plackett distribution in order to establish the multivariate cell probabilities. Their 1994 method involves solving polynomials of high degree and computing the derivatives thereof, while in 1999 generalized linear models theory is exploited, together with the use of an adaption of the iterative proportional fitting algorithm. Lang and Agresti (1994) exploit the equivalence between direct modeling and imposing restrictions on the multinomial probabilities, using undetermined Lagrange multipliers. Alternatively, the cell probabilities can be fitted using a Newton iteration scheme, as suggested by Glonek and McCullagh (1995).

However, even though a variety of flexible models exist, maximum likelihood can be unattractive due to excessive computational requirements, especially

when high dimensional vectors of correlated data arise. As a consequence, alternative methods have been in demand. Liang and Zeger (1986) proposed so-called *generalized estimating equations* (GEE) which require only the correct specification of the univariate marginal distributions provided one is willing to adopt "working" assumptions about the association structure. Le Cessie and Van Houwelingen (1994) suggested to approximate the true likelihood by means of a pseudo-likelihood (PL) function that is easier to evaluate and to maximize. Both GEE and PL yield consistent and asymptotically normal estimators, provided an empirically corrected variance estimator, often referred to as the sandwich estimator, is used. However, GEE is typically geared towards marginal models, whereas PL can be used with both marginal (Le Cessie and Van Houwelingen 1994, Geys, Molenberghs and Lipsitz 1998) and conditional models (Geys, Molenberghs and Ryan 1997, 1999). Alternative marginal models include the correlated binomial models of Altham (1978) and the double binomial model of Efron (1986).

2.2. Conditional Models

In a conditional model, the parameters describe a feature (probability, odds, logit, ...) of (a set of) outcomes, given values for the other outcomes (Cox 1972). The best known example is undoubtedly the log-linear model. Rosner (1984) described a conditional logistic model. Due to the popularity of marginal (especially generalized estimating equations) and random-effects models for correlated binary data, conditional models have received relatively little attention, especially in the context of multivariate clustered data. Diggle, Liang and Zeger (1994, pp. 147–148) criticized the conditional approach because the interpretation of the dose effect on the risk of one outcome is conditional on the responses of other outcomes for the same individual, outcomes of other individuals and the litter size. Molenberghs, Declerck and Aerts (1998) and Aerts, Declerck and Molenberghs (1997) have compared marginal, conditional and random-effects models for univariate clustered data. Their results are encouraging for the conditional model, since they are competitive for the dose effect testing and for benchmark dose estimation, and because they are computationally fast and stable. Molenberghs and Ryan (1999) discuss, in the specific context of exchangeable binary data, the advantages of conditional models and show how, with appropriate care, the disadvantages can be overcome. An advantage of such a likelihood-based approach is that, under correct model specification, efficiency can be gained over other procedures such as generalized estimating equations (GEE) methods.

3. Random-Effects Models

Models with subject-specific parameters are differentiated from population-averaged models by the inclusion of parameters which are specific to the cluster. Unlike for correlated Gaussian outcomes, the parameters of the random effects and population-averaged models for correlated binary data

describe different types of effects of the covariates on the response probabilities (Neuhaus 1992).

The choice between population-averaged and random effects strategies may heavily depend on the scientific goals. Population-averaged models evaluate the overall risk as a function of covariates. With a subject-specific approach, the response rates are modeled as a function of covariates and parameters, specific to a subject. In such models, interpretation of fixed-effect parameters is conditional on a constant level of the random-effects parameter. Population-averaged comparisons, on the other hand, make no use of within cluster comparisons for cluster varying covariates and are therefore not useful to assess within-subject effects (Neuhaus, Kalbfleisch, and Hauck 1991).

Subject-specific parameters can be dealt with in essentially three ways : (1) as fixed effects, (2) as random-effects, and (3) by conditioning upon them. The first approach is seemingly simplest but in many cases flawed since the number of parameters then increases with a rate proportional to the sample size, thereby invalidating most standard inferential results. The second approach is very popular. There are two routes to introduce randomness into the model parameters. Stiratelli, Laird, and Ware (1984) assume the parameter vector to be normally distributed. This idea has been carried further in the work on so-called *generalized linear mixed models* (Breslow and Clayton 1993) which is closely related to linear and non-linear mixed models. Alternatively, Skellam (1948) introduced the beta-binomial model, in which the adverse event probability of any response of a particular subject comes from a beta distribution. Hence, this model can also be viewed as a random effects model. The third approach is well known in epidemiology, more precisely in the context of matched case-control studies. In particular, conditional logistic regression is then often considered (Breslow and Day 1987). In general, with so-called conditional likelihood methods, one conditions on the sufficient statistics for the random effects (Ten Have, Landis, and Weaver 1995, Conaway 1989). Note that the conditioning considered here is different from the one considered in Section 2.2, since here we condition on random effects, rather than on outcomes. In the remainder of this section we will consider the beta-binomial model and classical mixed-effects models.

3.1. The Beta-binomial Model

Rather than modeling marginal functions directly, a popular approach is to assume a random effects model in which each litter has a random parameter (vector). Skellam (1948) and Kleinman (1973) assume the malformation probability P_i of any fetus in litter i to come from a beta distribution with parameters α_i and β_i :

$$\frac{p^{\alpha_i - 1}(1 - p)^{\beta_i - 1}}{B(\alpha_i, \beta_i)}, \qquad 0 \leq p \leq 1,$$

where $B(.,.)$ denotes the beta function. Conditional on P_i, the number of malformations Z_i in the ith cluster follows a binomial distribution. This leads to the well-known beta-binomial model.

The mean of this distribution is

$$\mu_i = n_i \pi_i = n_i \frac{\alpha_i}{\alpha_i + \beta_i} \tag{1}$$

and the variance is $\sigma_i^2 = n_i \pi_i (1 - \pi_i) \frac{1 + n_i \theta_i}{1 + \theta_i}$ with $\theta_i = 1/(\alpha_i + \beta_i)$. It can be shown that the intra-litter correlation can be expressed as

$$\rho_i = \frac{1}{\alpha_i + \beta_i + 1}. \tag{2}$$

In a litter of size n_i, the probability mass function of Z_i can be expressed directly in terms of the mean and correlation parameters (1) and (2) :

$$f(z_i \mid \pi_i, \rho_i, n_i) = \binom{n_i}{z_i} \frac{B(\pi_i(\rho_i^{-1} - 1) + z_i, (1 - \pi_i)(\rho_i^{-1} - 1) + n_i - z_i)}{B(\pi_i(\rho_i^{-1} - 1), (1 - \pi_i)(\rho_i^{-1} - 1))}, \tag{3}$$

where $B(.,.)$ denotes the beta function. The only association parameter of this model is ρ_i, which is the correlation between two binary responses of litter i. The higher order correlations of the beta-binomial model can be expressed as a function of the mean malformation probability π_i and ρ_i. The association in both the beta-binomial and the Bahadur model is expressed by means of the intraclass correlation. It turns out that both models have the same first and second moments. As a consequence, the parameter ρ_i of the beta-binomial model equals $\rho_{i(2)}$ of the Bahadur model. The parameters π_i and ρ_i of the beta-binomial model have a marginal interpretation and therefore, they are the parameters in the derived marginal model as well. This results in similarities between the beta-binomial and marginal models, such as the Bahadur model.

It can be shown (Williams 1975) that the contribution of the ith cluster to the log-likelihood, $\ln f(z_i|\pi_i, \rho_i, n_i) \equiv \ell_i$, can be written as

$$\ell_i = \sum_{r=0}^{z_i - 1} \ln \left(\pi_i + \frac{r \rho_i}{1 - \rho_i} \right) + \sum_{r=0}^{n_i - z_i - 1} \ln \left(1 - \pi_i + \frac{r \rho_i}{1 - \rho_i} \right)$$
$$- \sum_{r=0}^{n_i - 1} \ln \left(1 + \frac{r \rho_i}{1 - \rho_i} \right), \tag{4}$$

with $i = 1, ..., N$. It follows from (4) that if the association parameter ρ_i equals zero, then the beta-binomial model reduces to the logistic regression model. Generalized linear model ideas can be applied to model the mean parameter π_i (e.g., using a logit link) and the correlation parameter ρ_i (e.g., using Fisher's z transform). Kupper and Haseman (1978) compare the Bahadur model to the beta-binomial model. They conclude that the models perform similarly in three clustered data experiments, whereas they both outperform the (naive) binomial model.

3.2. Mixed Models

Perhaps the most commonly encountered subject-specific (or random effects model) is the generalized linear mixed model. It is best to first introduce linear mixed models and non-linear mixed models as a basis for the introduction of generalized linear mixed models. To emphasize they fit within a single common framework, we first give a general formulation.

3.2.1. General Formulation

Let y_{ij} denote the jth measurement available for the ith unit, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$, and let $\boldsymbol{y_i}$ denote the vector of all measurements for the ith unit, i.e., $\boldsymbol{y_i}^T = (y_{i1}, \ldots, y_{in_i})$. Our general model assumes that $\boldsymbol{y_i}$ (possibly appropriately transformed) satisfies

$$\boldsymbol{y_i} | \boldsymbol{b_i} \sim F_i(\boldsymbol{\theta}, \boldsymbol{b_i}), \tag{5}$$

i.e., conditional on $\boldsymbol{b_i}$, $\boldsymbol{y_i}$ follows a pre-specified distribution F_i, possibly depending on covariates, and parameterized through a vector $\boldsymbol{\theta}$ of unknown parameters, common to all subjects. Further, $\boldsymbol{b_i}$ is a q-dimensional vector of subject-specific parameters, called random effects, assumed to follow a so-called mixing distribution G which may depend on a vector $\boldsymbol{\psi}$ of unknown parameters, i.e., $\boldsymbol{b_i} \sim G(\boldsymbol{\psi})$. The $\boldsymbol{b_i}$ reflect the between-unit heterogeneity in the population with respect to the distribution of $\boldsymbol{y_i}$. Different specifications of F_i will lead to different models. For example, considering the factors made up of the outcomes y_{ij} given its predecessors $(y_{i1}, \ldots, y_{i,j-1})^T$ leads to a so-called transitional model. A model without any random effects $\boldsymbol{b_i}$ is called a marginal model for the response vector $\boldsymbol{y_i}$. In the presence of random effects, conditional independence is often assumed, under which the components y_{ij} in $\boldsymbol{y_i}$ are independent, conditional on $\boldsymbol{b_i}$. The distribution function F_i in (5) then becomes a product over the n_i independent elements in $\boldsymbol{y_i}$.

In general, unless a fully Bayesian approach is followed, inference is based on the marginalized model for $\boldsymbol{y_i}$ which is obtained from integrating out the random effects, over their distribution $G(\boldsymbol{\psi})$. Let $f_i(\boldsymbol{y_i}|\boldsymbol{b_i})$ and $g(\boldsymbol{b_i})$ denote the density functions corresponding to the distributions F_i and G, respectively, we have that the marginal density function of $\boldsymbol{y_i}$ equals

$$f_i(\boldsymbol{y_i}) = \int f_i(\boldsymbol{y_i}|\boldsymbol{b_i})g(\boldsymbol{b_i})d\boldsymbol{b_i}, \tag{6}$$

which depends on the unknown parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$. Assuming independence of the units, estimates of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\psi}}$ can be obtained from maximizing the likelihood function built from (6), and inferences immediately follow from classical maximum likelihood theory.

Obviously, the random-effects distribution G is crucial in the calculation of the marginal model (6). One approach is to leave G completely unspecified and to use non-parametric maximum likelihood (NPML) estimation, which maximizes the likelihood over all possible distributions G. The resulting estimate $\widehat{G}$ is then always discrete with finite support. Depending on the

context, this may or may not be a realistic reflection of the true heterogeneity between units. One therefore often assumes G to be of a specific parametric form, such as a (multivariate) normal. Depending on F_i and G, the integration in (6) may or may not be possible analytically. Proposed solutions are based on Taylor series expansions of $f_i(y_i|b_i)$, or on numerical approximations of the integral, such as (adaptive) Gaussian quadrature, or on the EM algorithm (Dempster, Laird, and Rubin 1997).

Although in practice one is usually primarily interested in estimating the parameters in the marginal model, it is often useful to calculate estimates for the random effects b_i as well. They reflect between-subject variability, which makes them helpful for detecting special profiles (i.e., outlying individuals) or groups of individuals evolving differently in time. Also, estimates for the random effects are needed whenever interest is in prediction of subject-specific evolutions. Inference for the random effects is often based on their so-called posterior distribution $f_i(b_i|y_i)$, given by

$$f_i(b_i|y_i) = \frac{f_i(y_i|b_i)\ g(b_i)}{\displaystyle\int f_i(y_i|b_i)\ g(b_i)\ db_i}, \tag{7}$$

in which the unknown parameters θ and ψ are replaced by their estimates obtained earlier from maximizing the marginal likelihood. The mean or mode corresponding to (7) can be used as point estimates for b_i, yielding empirical Bayes (EB) estimates.

3.2.2. Linear Mixed Models

When continuous (normally distributed) hierarchical data are considered (repeated measures, clustered data, geographical data, longitudinal data,...), a general, and very flexible, class of parametric covariance models is obtained from introducing random effects b_i in the multivariate linear regression model. Linear mixed models assume the outcome vector y_i follows a multivariate normal distribution, with mean vector $X_i\beta + Z_i b_i$ and some covariance matrix Σ_i, and assume that the random effects b_i also follow a (multivariate) normal distribution, i.e., it is assumed that the n_i-dimensional vector y_i satisfies

$$y_i|b_i \sim N(X_i\beta + Z_i b_i, \Sigma_i), \tag{8}$$

$$b_i \sim N(0, D), \tag{9}$$

where X_i and Z_i are $(n_i \times p)$ and $(n_i \times q)$ dimensional matrices of known covariates, β is a p-dimensional vector of regression parameters, called the fixed effects, D is a general $(q \times q)$ covariance matrix, and Σ_i is a $(n_i \times n_i)$ covariance matrix which depends on i only through its dimension n_i, i.e., the set of unknown parameters in Σ_i will not depend upon i. This formulation corresponds to making Gaussian assumptions about F_i and G, introduced in Section 3.2.1, and a linear mean model.

The above model can be interpreted as a linear regression model for the vector y_i of repeated measurements for each unit separately, where some of the

regression parameters are specific (random effects, b_i), while others are not (fixed effects, β). The distributional assumptions in (9) with respect to the random effects can be motivated as follows. First, $\mathrm{E}(b_i) = 0$ implies that the mean of y_i still equals $X_i\beta$, such that the fixed effects in the random-effects model (8) can also be interpreted marginally. Not only do they reflect the effect of changing covariates within specific units, they also measure the marginal effect in the population of changing the same covariates. As will be discussed further, this important property only holds for very specific random-effects models, one of which is the linear mixed model considered here. Second, the normality assumption immediately implies that, marginally, y_i also follows a normal distribution with mean vector $X_i\beta$ and with covariance matrix $V_i = Z_i D Z_i^T + \Sigma_i$. Hence, no numerical approximation to the integral in (6) is needed. Apart from this mathematical convenience, the normality assumption for the b_i is further supported by noticing that the b_i express how unit-specific trends deviate from the population-averaged trends, which suggests that they can be interpreted as residuals.

Note that the random effects in (8) implicitly imply the marginal covariance matrix V_i of y_i to be of the very specific form $V_i = Z_i D Z_i^T + \Sigma_i$. Let us consider two examples under the assumption of conditional independence, i.e., assuming $\Sigma_i = \sigma^2 I_{n_i}$. First, consider the case where the random effects are univariate and represent unit-specific intercepts. This corresponds to covariates Z_i which are n_i-dimensional vectors containing only ones. The implied covariance matrix can then easily be shown to have the compound symmetry structure which makes the strong assumption that the variance remains constant over all repeated measures and that the correlation between any two measures within a specific unit is also constant. Second, for longitudinal data, suppose that the b_i represent unit-specific intercepts as well as linear time effects. The corresponding Z_i are then of the form

$$Z_i = \begin{pmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix},$$

where t_{ij} is the time point at which the jth measurement was taken for the ith subject. Denoting the (k, l) element in D as d_{kl}, we have that the covariance between two repeated measures within a single unit is given by

$$\mathrm{Cov}(y_{ik}, y_{il}) = (1 \; t_{ik}) \, D \begin{pmatrix} 1 \\ t_{il} \end{pmatrix} \; + \; \sigma^2$$
$$= d_{22} \, t_{ik} \, t_{il} + d_{12}(t_{ik} + t_{il}) + d_{11} + \sigma^2.$$

Note how the model now implies the variance function of the response to be quadratic over time, with positive curvature d_{22}.

The marginal model implied by expressions (8) and (9) is

$$y_i \; \sim \; N(X_i\beta, V_i), \quad V_i = Z_i D Z_i^T + \Sigma_i$$

which can be viewed as another multivariate linear regression model, with a very particular parameterization of the covariance matrix V_i. Hence, our earlier remarks with respect to the fitting of the marginal model remain valid. The vector $\boldsymbol{\alpha}$ of variance components then consists of the variances and covariances in D as well as all unknown parameters in Σ_i.

With respect to the estimation of unit-specific parameters $\boldsymbol{b}_i$, the posterior distribution of $\boldsymbol{b}_i$ given the observed data $\boldsymbol{y}_i$ can be shown to be (multivariate) normal with mean vector equal to $DZ_i^T V_i^{-1}(\boldsymbol{\alpha})(\boldsymbol{y}_i - X_i \boldsymbol{\beta})$. Replacing $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ by their maximum likelihood estimates, we obtain the EB estimates $\widehat{\boldsymbol{b}_i}$ for the $\boldsymbol{b}_i$ (Verbeke and Molenberghs 2000).

3.2.3. Non-linear Mixed Models

An extension of model (8) which allows for non-linear relationships between the responses in $\boldsymbol{y}_i$ and the covariates in X_i and/or Z_i is

$$\boldsymbol{y}_i | \boldsymbol{b}_i \ \sim \ N(h(X_i, Z_i, \boldsymbol{\beta}, \boldsymbol{b}_i), \Sigma_i) \tag{10}$$

for some known inverse link function h. The definition of X_i, Z_i, $\boldsymbol{\beta}$, and $\boldsymbol{b}_i$ remains unchanged, the random effects $\boldsymbol{b}_i$ are again assumed to be normally distributed with mean vector $\boldsymbol{0}$ and covariance matrix D and inference can proceed as explained for the general model.

There are at least two major differences in comparison to the linear mixed model discussed in the previous section. First, the marginal distribution of $\boldsymbol{y}_i$ can no longer be calculated analytically, such that numerical approximations to the marginal density (6) come into play, seriously complicating the computation of the maximum likelihood estimates of the parameters in the marginal model, i.e., $\boldsymbol{\beta}$, D, and the parameters in all Σ_i. A consequence is that the marginal covariance structure does not immediately follow from the model formulation, such that it is not always clear in practice what assumptions a specific model implies with respect to the underlying variance function and the underlying correlation structure in the data.

A second important difference is with respect to the interpretation of the fixed effects $\boldsymbol{\beta}$. Under the linear model (8), we have that $E(\boldsymbol{y}_i)$ equals $X_i \boldsymbol{\beta}$, such that the fixed effects have a subject-specific as well as a population-averaged interpretation. Indeed, the elements in $\boldsymbol{\beta}$ reflect the effect of specific covariates, conditionally on the random effects $\boldsymbol{b}_i$, as well as marginalized over these random effects. Under non-linear mixed models, however, this does no longer hold in general. The fixed effects now only reflect the conditional effect of covariates, and the marginal effect is not easily obtained anymore as $E(\boldsymbol{y}_i)$ is given by

$$E(\boldsymbol{y}_i) = \int \boldsymbol{y}_i \int f_i(\boldsymbol{y}_i | \boldsymbol{b}_i) g(\boldsymbol{b}_i) d\boldsymbol{b}_i d\boldsymbol{y}_i,$$

which, in general, is not of the form $h(X_i, Z_i, \boldsymbol{\beta}, \boldsymbol{0})$, but rather integrates the random effects from $h(X_i, Z_i, \boldsymbol{\beta}, \boldsymbol{b}_i)$.

Only for very particular models, (some of) the fixed effects can still be interpreted as marginal covariate effects. For example, consider the model

where, apart from an exponential link function, the mean is linear in the covariates, and the only random effects in the model are intercepts. More specifically, this corresponds to the model with $h(X_i, Z_i, \boldsymbol{\beta}, \boldsymbol{b_i}) = \exp(X_i\boldsymbol{\beta} + Z_i b_i)$, in which Z_i is now a vector containing only ones. The expectation of $\boldsymbol{y_i}$ is now given by

$$\mathrm{E}(\boldsymbol{y_i}) = \mathrm{E}\left[\exp(X_i\boldsymbol{\beta} + Z_i b_i)\right]$$
$$= \exp(X_i\boldsymbol{\beta})\; Z_i E\left[\exp(b_i)\right], \tag{11}$$

which shows that, except for the intercept, all parameters in $\boldsymbol{\beta}$ have a marginal interpretation.

3.2.4. The Generalized Linear Mixed Model

The generalized linear mixed model is the most frequently used random-effects model for discrete outcomes. A general formulation is as follows. Conditionally on random effects $\boldsymbol{b_i}$, it assumes that the elements y_{ij} of $\boldsymbol{y_i}$ are independent, with density function of the form

$$f_i(y_{ij}|\boldsymbol{b_i}) = \exp\left[(y_{ij}\eta_{ij} - a(\eta_{ij}))/\phi + c(y_{ij}, \phi)\right],$$

with mean $\mathrm{E}(y_{ij}|\boldsymbol{b_i}) = a'(\eta_{ij}) = \mu_{ij}(\boldsymbol{b_i})$ and variance $\mathrm{Var}(y_{ij}|\boldsymbol{b_i}) = \phi a''(\eta_{ij})$, and where, apart from a link function h, a linear regression model with parameters $\boldsymbol{\beta}$ and $\boldsymbol{b_i}$ is used for the mean, i.e., $h(\boldsymbol{\mu_i}(\boldsymbol{b_i})) = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i}$. Note that the linear mixed model is a special case, with identity link function. The random effects $\boldsymbol{b_i}$ are again assumed to be sampled from a (multivariate) normal distribution with mean $\boldsymbol{0}$ and covariance matrix D. Usually, the canonical link function is used, i.e., $h = {a'}^{-1}$, such that $\boldsymbol{\eta_i} = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i}$.

The non-linear nature of the model again implies that the marginal distribution of $\boldsymbol{y_i}$ is, in general, not easily obtained, such that model fitting requires approximation of the marginal density function. An exception to this occurs when the probit link is used (Liang, Zeger, and Qaqish 1992). Further, as was also the case for non-linear mixed models, the parameters $\boldsymbol{\beta}$ have no marginal interpretation, except for some very particular models. An example where the marginal interpretation does hold is the Poisson model for count data, for which the logarithm is the canonical link function. In case the model only includes random intercepts, it immediately follows from the calculations in (11) that the only element in $\boldsymbol{\beta}$ which has no marginal interpretation is the intercept.

As another example, consider the binomial model for binary data, with the logit canonical link function, and where the only random effects are intercepts b_i. It can then be shown that the marginal mean $\boldsymbol{\mu_i} = \mathrm{E}(y_{ij})$ satisfies $h(\boldsymbol{\mu_i}) \approx X_i\boldsymbol{\beta}^*$ with $\boldsymbol{\beta}^* = [c^2\mathrm{Var}(b_i) + 1]^{-1/2}\boldsymbol{\beta}$, in which c equals $16\sqrt{3}/15\pi$. Hence, although the parameters $\boldsymbol{\beta}$ in the generalized linear mixed model have no marginal interpretation, they do show a strong relation to their marginal counterparts. Note that, as a consequence of this relation, larger covariate effects are obtained under the random-effects model in comparison to the marginal model.

4. Fitting Generalized Linear Mixed Models

As indicated earlier, likelihood inference for generalized linear mixed models requires evaluation of integrals (Breslow and Clayton 1993), where the integral's dimension is equal to the number of random effects. In this section, we will discuss ways to handle this issue. Broadly, we can distinguish between approximations to the marginal likelihood, thereby avoiding the integration all together, and methods based upon numerical integration. We will discuss each of these approaches in turn. A useful reference on estimation methods is Lavergne and Trottier (2000).

4.1. Approximate Methods

Zeger and Karim (1991) avoid the need for numerical integration by casting the generalized linear random-effects model in a Bayesian framework and by resorting to the Gibbs sampler. Breslow and Clayton (1993) exploit the penalized quasi-likelihood (PQL) estimator by applying Laplace's method for integral approximation. They also consider marginal quasi-likelihood (MQL), a name they give to a procedure previously proposed by Goldstein (1991). These two approaches entail iterative fitting of linear models based on first-order Taylor expansions of the mean function about the current estimated fixed part predictor (MQL) or the current predicted value (PQL). EM-type methods have been reviewed in Santner and Duffy (1989) and Fahrmeir and Tutz (1994). The method proposed by Gilmour, Anderson, and Rae (1985) has seen some use as well. Wolfinger and O'Connell (1993) circumvent numerical integration by using pseudo-likelihood (and restricted pseudo-likelihood) procedures.

The approach proposed by Wolfinger and O'Connell (1993), is based on an extension of the method of Nelder and Wedderburn (1972) (see also McCullagh and Nelder 1989) to fit fixed-effects generalized linear models. Let us briefly recall this procedure. Dropping the subject-specific index i, the basic form of a generalized linear model is $\boldsymbol{\eta} = X\boldsymbol{\beta}$, where $\boldsymbol{\eta} = g(\boldsymbol{\mu})$, $\boldsymbol{\mu} = E(\boldsymbol{Y})$ and g is an appropriate link function. Nelder and Wedderburn (1972) showed that maximum likelihood estimates for $\boldsymbol{\beta}$ can be obtained by iteratively solving

$$X^T W X \boldsymbol{\beta} = X^T W \boldsymbol{y}^*, \qquad (12)$$

where $W = Q\Sigma^{-1}Q$, $\boldsymbol{y}^* = \hat{\eta} + (\boldsymbol{y} - \hat{\mu})Q^{-1}$, $Q = (\partial\mu/\partial\eta)$, and $\Sigma = \Sigma_\mu^{1/2} A \Sigma_\mu^{1/2}$. Here, Σ_μ is a diagonal matrix of variances and A is a correlation matrix. McCullagh and Nelder (1989) note that the "working" dependent variable in these estimating equations is not $\boldsymbol{y}$ but $\boldsymbol{y}^*$, a linearized version of $\boldsymbol{y}$.

Wolfinger and O'Connell's method (see also Schall 1991 and Breslow and Clayton 1993), implemented in the SAS macro GLIMMIX, is essentially a random-effects extension of (12). The GLIMMIX macro is known to have some drawbacks such as, for example, downward biases in fixed-effects and covariance parameters. This issue will be taken up in the next section. The estimation procedure is based on iterating between the computation of working dependent variables on the one hand and fitting a *linear* mixed model

to it on the other hand. The explicit steps of the algorithm can be found in, for example, Aerts *et al* (2002).

4.2. Numerical Integration

An obvious way to deal with the integral in (6) is to apply numerical integration. Of course, a wide toolkit of numerical integration tools, available from the optimization literature, can be applied. Several of those have been used in such software tools and the NLMIXED procedure in SAS and the MIXOR program. We refer to the next section for explicit software-related details.

A general class of quadrature rules selects a set of abscissas and constructs a weighted sum of function evaluations over those. In the particular context of random-effects distributions, so-called *adaptive* quadrature rules can be used (Pinheiro and Bates 2000), were the numerical integration is centered around the EB estimates of the random effects, and the number of quadrature points is then selected in terms of the desired accuracy.

An alternative to numerical integration is based on Monte Carlo Markov Chain methods (MCMC; Gilks and Wild 1992).

5. Case Study and Software

To illustrate the fitting of random-effects models to longitudinal data, we will consider data from a single-arm trial with 530 patients recruited (491 selected for analysis), given analgesic treatment for pain caused by chronic nonmalignant disease. Treatment was to be administered for 12 months and assessed by means of a Global Satisfaction Assessment (GSA), rated on a five-point scale ranging from 1 (very good) to 5 (very bad). For our analysis, we will focus on a dichotomized version ("success" if the original outcome is 3 or less, and "failure" otherwise). GSA was rated by each subject 4 times during the trial, at months 3, 6, 9, and 12. A number of subjects have incomplete measurements. These will be taken into account for analysis. A thorough treatment of missingness is beyond the scope of this paper (Verbeke and Molenberghs 2000).

To facilitate comparison, we will fit the same model using a variety of software tools :

$$\text{logit}(\text{GSAB}_{ij}) = \beta_0 + b_i + \beta_1 t_j + \beta_2 t_j^2 + \beta_3 \text{BASE}_i. \qquad (13)$$

Here GSAB_{ij} is the dichotomized version of GSA at time j for patient i and BASE_i is a baseline evaluation of severity of the disease; t_j is the time of evaluation j (common to all subjects and hence the subscript i is dropped). Finally, note that a random intercept b_i has been included into the linear predictor.

5.1. SAS macro GLIMMIX

The GLIMMIX macro in SAS is based upon the algorithm proposed by Wolfinger and O'Connell (1993). Technically, it combines elements from generalized linear mixed models (e.g., the specification of a binomial error

structure) with a linear mixed models core, in line with the algorithm outlined in Section 4.1. A sample program, used to specified the above model, is :

```
%glimmix(data = gsa,
      procopt = %str(method=ml noclprint covtest),
      stmts=%str(
            class patid timecls ;
            model gsab = time|time base / solution ;
            random intercept / subject=patid type=un ;
                  ),
      error = binomial) ;
```

An advantage of the procedure is that it allows for the full flexibility of the MIXED procedure. Not only random effects, also serial correlation (by means of the REPEATED statement) is allowed. Empirically corrected standard errors, in addition to purely model-based standard errors, can be requested by means of the EMPIRICAL option.

5.2. SAS procedure NLMIXED

This procedure allows fitting a wide class of linear, generalized linear, and non-linear mixed models. It is available in SAS Version 7.0 and higher, and it relies on numerical integration. Not only are different integral approximations available, the principal one being (adaptive) Gaussian quadrature, it also includes a number of optimization algorithms. The procedure can perform Gaussian quadrature by use of the options NOAD (requesting non-adaptive Gaussian quadrature, i.e., the quadrature points are centered at zero for each of the random effects and the current random-effects covariance matrix is used as the scale matrix) and NOADSCALE (requesting non-adaptive scaling for adaptive Gaussian quadrature, i.e., the quadrature points are centered at the empirical Bayes estimates for the random effects, and the current random-effects covariance matrix is used as the scale matrix). The number of quadrature points can be specified with the option QPOINTS=m. The procedure can maximize the marginal likelihood using the Newton-Raphson algorithm by specifying the option TECHNIQUE=NEWRAP. Constraints are allowed in the optimization process. Importantly, the conditional distribution of the response, given the random effects, can be specified as normal, binomial, Poisson or, very importantly, as *any distribution*, in which case you can specify the likelihood by programming statements. EB estimates of the random effects can be obtained. The procedure allows to specify a single RANDOM statement (allowing for several random effects though, but all within a 2-level hierarchy). Only a normal random-effects distribution is allowed.

A program to fit model (13) is :

```
proc nlmixed data=gsa npoints=20 noad noadscale tech=newrap ;
      parms beta0=3 beta1=-0.8 beta2=0.2 beta3=-0.2 su=1 ;
      eta = beta0 + beta1*time + beta2*time2 + beta3*base + u ;
```

```
      expeta = exp(eta) ;
      p = expeta/(1+expeta) ;
      model gsab ~ binary(p) ;
      random u ~ normal(0,su**2) subject=patid ;
      estimate 'ICC' su**2/(arcos(-1)**2/3 + su**2) ;
run ;
```

Clearly, there is a lot of flexibility in setting up a model thanks to the use of programming statements. For example, the distribution of the random intercept is specified through the standard deviation. Alternatively, one could choose to use the variance instead. The RANDOM statement of the above program is then changed to

```
      random u ~ normal(0,s2u) subject=patid ;
run ;
```

The final statement has been added to calculate the intraclass correlation :

$$\rho = \frac{\sigma^2}{\pi^2/3 + \sigma^2}.$$

Here, σ^2 denotes the random-intercept variance and $\pi^2/3$ is the logistic variance.

In case one is interested in the mean (population-averaged profiles), then one needs to calculate

$$\int_{-\infty}^{+\infty} \frac{\exp(x_{ij}^T\beta + b)}{1 + \exp(x_{ij}^T\beta + b)} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-b^2/(2\sigma^2)} db.$$

When there are covariates, it is customary to take mean values (or several important values) for the covariates. For our model, this is exemplified in Figures 1 and 2. In the first of these, a model without the BASE covariate was considered. Clearly, this results in precision loss. Further, it follows from the second figure that the fitted line somewhat "overestimates" the observed means. This is due to missingness. Indeed, the fitted line applies to all subjects, also those with incomplete measurements, whereas the observed means plotted are based on the available observations only. Now, typically those with a favorable pain score stay in the study longer and vice versa.

5.3. MIXOR

The MIXOR program is in the public domain and can be downloaded from

<div align="center">http ://www.uic.edu/ hedeker/mixreg.html.</div>

It is developed for mixed-effects ordinal regression analysis, and hence in particular in the binary case, and has been documented extensively in Hedeker and Gibbons (1993, 1994, 1996). It performs numerical integration (Gaussian
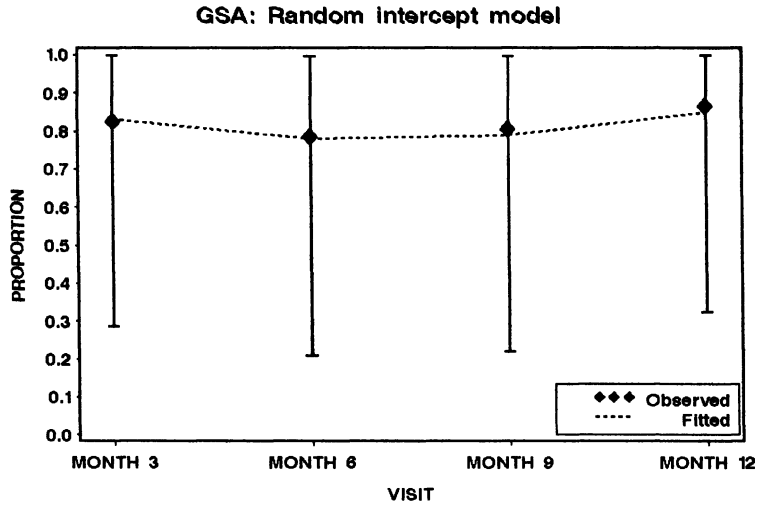
GSA: Random Intercept model

FIG. 1. – Fitted profiles for model without BASE covariate. Estimates have been obtained using PROC NLMIXED.

GSA: Random Intercept model
Adjusted for current PCA
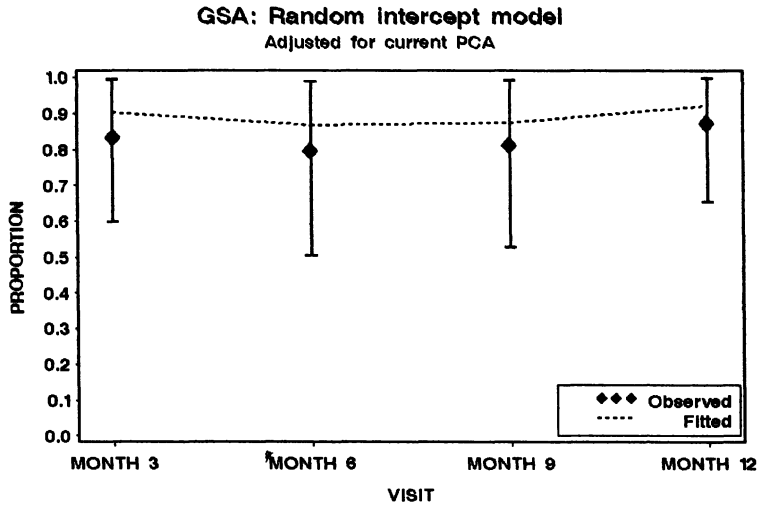
FIG. 2. – Fitted profiles for model with BASE covariate. Estimates have been obtained using PROC NLMIXED.

quadrature) and uses Newton-Raphson algorithm to maximize the marginal likelihood. Technically, MIXOR is most directly comparable to NLMIXED. This is reflected in the parameter estimates (Table 1) but not entirely in standard errors, because MIXOR uses an approximation to the (empirical) information matrix, whereas NLMIXED uses numerical derivatives.

5.4. MLwiN

MLwiN is the successor of an earlier DOS incarnation MLN, and is the implementation of the *multilevel modeling* approach, proposed in Bryk and Raudenbush (1992), Longford (1993) and Goldstein (1995). Kreft and de Leeuw (1998) provide a more informal and introductory approach to the subject. This modeling approach for hierarchical data (and hence in particular longitudinal data) is primarily used and known in the social sciences environment. While the "language" typically used to describe the model is somewhat different from the linear and generalized linear mixed model formalisms described above, it is very similar and a wide class of mixed models can be considered within the multilevel paradigm as well.

Parameter estimation in multilevel linear regression models can be carried out by maximizing the likelihood function. To this end, direct maximization, using Newton-Raphson or the EM algorithm (Little and Rubin 1987) for instance, can be performed. An equivalent procedure, called *iterative generalized least squares* (IGLS), was proposed by Goldstein (1986). His algorithm simply iterates between the estimation of the fixed and random parameters obtained by standard generalized least squares formulae, hence its name. IGLS is an attractive procedure as it tends to be quite efficient with large data sets typically encountered in the multilevel modeling framework. Note that the IGLS algorithm can be slightly modified (RIGLS) to perform similarly to residual (or restricted) maximum likelihood estimation, which yields unbiased estimates for variance components in random-effects models (Verbeke and Molenberghs 2000).

MLwiN also allows for non-normal (e.g., binary) outcomes. As before, to maximize the marginal likelihood function, random effects need to be integrated out. As Rodriguez and Goldman (1995) demonstrate, the approximate procedures PQL and MQL, proposed by Breslow and Clayton (1993), may be seriously biased when applied to binary response data. Their simulations reveal that both fixed effects and variance components may suffer from substantial, if not severe, attenuation bias in certain situations. Goldstein and Rasbash (1996) show that including a second-order term in the PQL expansion greatly reduces the bias described by Rodriguez and Goldman. Other authors have advised the introduction of bias-correction terms (Lin and Breslow 1996) or the use of iterative bootstrap (Kuk 1995) among other things.

We will illustrate the use of both PQL and PQL2, using MLwiN. The software is menu driven and a typical menu is displayed in Figure 3.

5.5. Parameter Estimates

Parameter estimates, obtained with all software tools discussed are summarized in Table 1. For the sake of comparison, logistic regression and generalized estimating equation based estimates have been added. The first one is model based (also called "naive"), the second one is empirically corrected (also called "robust"). Unstructured working assumptions have been used.

$$gsabin_{ij} \sim \text{Binomial}(denom_{ij}, \pi_{ij}) \Bigg\}$$
$$gsabin_{ij} = \pi_{ij} + e_{4ij}bint^{*}$$

$$logit(\pi_{ij}) = \beta_{0j}intcpt + -1.172(0.477)time_{ij} + 0.245(0.098)time2_{ij} + -0.309(0.150)pca0_{j}$$
$$\beta_{0j} = 4.067(0.703) + u_{0j}$$

$$\left[u_{0j} \right] \sim N(0, \Omega_{u}) : \Omega_{u} = \left[2.594(0.473) \right]$$

$$bint^{*} = bint[\pi_{ij}(1 - \pi_{ij})/denom_{ij}]^{0.5}$$

$$\left[e_{4ij} \right] \sim (0, \Omega_{e}) : \Omega_{e} = \left[1.000(0.000) \right]$$
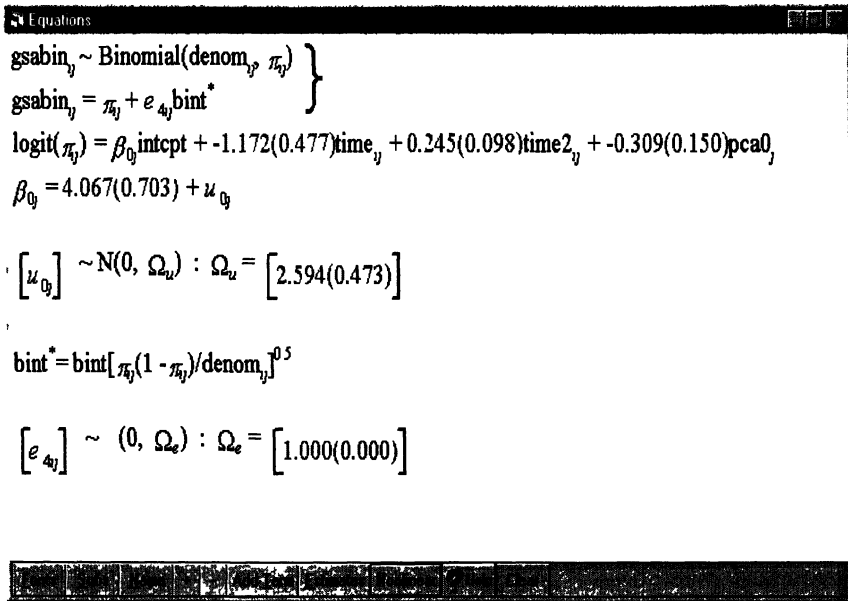
FIG 3. – Fitted profiles for model with BASE covariate. Estimates have been obtained using PROC NLMIXED.

There is a clear distinction between the marginal model estimates (logistic and GEE) and the random-effects model estimates. This is in line with earlier observations that there is a fundamental interpretational difference between both. In the first case, parameters describe the average evolution of the population, in the second case, they describe the evolution of a population *with a given level of the random effect*. These two have a different form, except in special cases such as the linear mixed model. Using the logit link, the population averaged mean is

$$P(Y_{ij} = 1 | X_{i}, \hat{\beta}) = \frac{\exp(X_{i}\hat{\beta})}{1 + \exp(X_{i}\hat{\beta})},$$

while the mean, conditional on the random effects, is

$$P(Y_{ij} = 1 | X_{i}, Z_{i}, \hat{\beta}, \hat{b}_{i}) = \frac{\exp(X_{i}\hat{\beta} + Z_{i}\hat{b}_{i})}{1 + \exp(X_{i}\hat{\beta} + Z_{i}\hat{b}_{i})}.$$

Now, the average of the latter over the random effects is *not* of a logistic form any more. In general,

$$h^{-1}(X_{i}\hat{\beta}) \neq E\left[h^{-1}(X_{i}\hat{\beta} + Z_{i}\hat{b}_{i}) \right],$$

for an arbitrary link function $h(\cdot)$, unless one chooses a linear (linear mixed model), probit, or log link, as stated in Section 3.2. As discussed by Neuhaus,

TABLE 1. – Overview of parameter estimates for Model (13), using software tools for generalized linear mixed models discussed in this section. In addition, logistic regression and GEE parameters have been added. The first line of each entry is the parameter estimate. (The second line of each entry is the standard error.) For GEE two standard errors are reported : (model based ; empirically corrected).

Parameter		logistic	GEE	GLIM-MIX	NL-MIXED	MIXOR	MLWIN (PQL1)	(PQL2)
Interc.	β_0	2.802	2.873	4.029	4.047	4.047	3.021	4.067
		(0.490)	(0.484;0.459)	(0.548)	(0.710)	(0.713)	(0.547)	(0.703)
Time	β_1	-0.786	-0.778	-1.279	-1.160	-1.160	-0.868	-1.172
		(0.387)	(0.328;0.323)	(0.334)	(0.466)	(0.475)	(0.405)	(0.477)
Time^2	β_2	0.177	0.167	0.259	0.245	0.244	0.189	0.245
		(0.079)	(0.067;0.066)	(0.068)	(0.095)	(0.097)	(0.083)	(0.098)
Base	β_3	-0.206	-0.228	-0.292	-0.300	-0.300	-0.223	-0.309
		(0.086)	(0.103;0.096)	(0.130)	(0.143)	(0.154)	(0.108)	(0.150)
R.I. var.	σ^2			3.165	2.533		1.019	2.594
				(0.391)	(0.676)		(0.248)	(0.473)
R.I. std.	σ				1.591	1.591		
					(0.213)	(0.206)		
resid.var.				0.495				
				(0.025)				
ICC					0.435	0.435		
					(0.066)			

Abbreviations : R.I. var. : random-intercept variance ; R.I. std. : random-intercept standard deviation ; resid.var. : residual variance ; ICC : intraclass correlation.

Kalbfleisch, and Hauck (1991), the marginal-model estimates are typically smaller in absolute value than their random-effects counterparts, in case a logistic link is issues. However, one should not refer to this phenomenon as *bias* since the two sets of parameter target at different scientific questions.

Within the group of random-effect estimates, NLMIXED can be considered the standard since it employs the most exact solution of all (numerical integration for the likelihood and numerical derivation for the information matrix), even though Lesaffre and Spiessens (2001) discuss some issues with the quadrature algorithms used by NLMIXED. Note that the MIXOR estimates are virtually the same at the level of precision reported, with slightly different standard error. Further, PQL2 within MLwiN gives, due to the second-order approximation, reasonably accurate results. PQL1 is unacceptably biased (note the bias is towards the marginal-model estimates), with some level of bias also observed with GLIMMIX. Some packages report the random-intercept variance, while others report the random-intercept standard devia-

tion. While these two pieces of information are equivalent, one should carefully check which one is reported. Due to the flexibility of the NLMIXED coding, one can choose to implement either, as illustrated before.

6. Concluding Remarks

In this paper, a general framework for modeling repeated categorical data has been sketched, with three main model families : marginal, conditional, and subject-specific. We have focused mainly on subject-specific or random-effects model, with some emphasis on the generalized linear mixed model. Estimation and optimization algorithms have been discussed, together with available software. These tools have been exemplified using a simple but illustrative analysis.

While the similarities between linear and generalized linear mixed models are often pointed out, perhaps rightly so, one should be very aware of and careful with the differences.

First, there is a close connection between parameter estimates from all model families when the responses are normally distributed. This in not true in the generalized linear case. Indeed, in marginal modeling, the regression parameters are unequivocally population parameters; they determine the effect of explanatory variables on the population mean response. In conditional (e.g., transition) and random-effects modeling, the regression parameters are still population parameters, in the sense that they operate on all subjects, but they determine the effects of explanatory variables on the mean response of an individual subject, *conditional* on that subject's measurement history (transition model), or on the subject's own random characteristics (random effects model). No easy conversion between them is possible, and therefore the researcher has to reflect carefully on the scientific question, before selecting a model.

Second, the random components in both types of mixed models interact completely differently. In a linear mixed model, the random effects are part of the linear predictor, and this is shared with the generalized linear mixed model case. However, the measurement error (residual error) are *also* within the linear predictor in the linear mixed model. There is no such thing in the generalized case. Above all, this is due to the mean-variance link, typical for the generalized linear mixed model (e.g., the variance of a Bernoulli outcome is the product of success and failure probabilities; the mean and variance are equal for Poisson data). Since the mean is transformed by means of the link function, so is the residual variance. Hence, both components of variability are not part of the same linear function in the generalized case.

Third, and related to the previous issue, the presence of a link function generally prohibits the existence of a closed form expression for the score-equation contribution, i.e., integrals cannot be avoided. This renders parameter estimation more complicated and a wide class of algorithms have been proposed, with an associated class of software tools. The existence of various tools with different numerical properties and performance makes this field rather tricky

and a thorough understanding of the issues is necessary before a satisfactory analysis can be undertaken. Arguably, knowledge of several (software) tools with a good understanding of the approximations on which they are based can greatly enhance insight.

REFERENCES

AERTS M., DECLERCK L., MOLENBERGHS G. (1997), Likelihood misspecification and safe dose determination for clustered binary data, *Environmetrics*, 8, 613-627.

AERTS M., GEYS H., MOLENBERGHS G., RYAN L.M. (2002), *Topics in Modelling of Clustered Binary Data*, London : Chapman & Hall.

ALTHAM P.M.E. (1978), Two generalizations of the binomial distribution, *Applied Statistics*, 27, 162-167.

BAHADUR R.R. (1961), A representation of the joint distribution of responses of n dichotomous items, In : *Studies in item analysis and prediction*, H. Solomon (ed.), Stanford Mathematical Studies in the Social Sciences VI. Stanford, California, Stanford University Press.

BRESLOW N.E., CLAYTON D.G. (1993), Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association*, 88, 9-25.

BRESLOW N.E., DAY N.E. (1987), *Statistical Methods in Cancer Research, Volume II*, Oxford : Oxford University Press.

BRYK A.S., RAUDENBUSH S.W. (1992), *Hierarchical Linear Models : Applications and Data Analysis Methods*, Newbury Park : Sage Publications.

CONAWAY M. (1989), Analysis of repeated categorical measurements with conditional likelihood methods, *Journal of the American Statistical Association*, 84, 53-62.

COX D.R. (1972), The analysis of multivariate binary data, *Applied Statistics*, 21, 113-120.

COX N.R. (1974), Estimation of the correlation between a continuous and a discrete variable, *Biometrics*, 30, 171-178.

CRESSIE N.A.C. (1991), *Statistics for Spatial Data*, New York : John Wiley.

DALE J.R. (1986), Global cross-ratio models for bivariate, discrete, ordered responses, *Biometrics*, 42, 909-917.

DECLERCK L., AERTS M., MOLENBERGHS G. (1998), Behaviour of the likelihood ratio test statistic under a Bahadur model for exchangeable binary data, *Journal of Statistical Computations and Simulations*, 61, 15-38.

DEMPSTER A.P., LAIRD N.M., RUBIN D. B. (1977), Maximum likelihood from incomplete data via the EM algorithm (with discussion), *Journal of the Royal Statistical Society, Series B*, 39, 1-38.

DIGGLE P.J. (1983), *Statistical Analysis of Spatial Point Patterns*, Mathematics in Biology, London : Academic Press.

DIGGLE P.J. (1990), *Time Series : A Biostatistical Introduction*, Oxford : Oxford University Press.

DIGGLE P.J., LIANG K-Y., ZEGER S.L. (1994), *Analysis of Longitudinal Data*, New York : Oxford University Press.

EFRON B. (1986), Double exponential families and their use in generalized linear regression, *Journal of the American Statistical Association*, 81, 709-721.

FAHRMEIR L., TUTZ G. (1994), *Multivariate Statistical Modelling Based on Generalized Linear Models*, Heidelberg : Springer-Verlag.

GEYS H., MOLENBERGHS G., LIPSITZ S.R. (1998), A note on the comparison of pseudo-likelihood and generalized estimating equations for marginal odds ratio models, *Journal of Statistical Computation and Simulation*, 62, 45-72.

GEYS H., MOLENBERGHS G., RYAN L.M. (1997), Pseudo-likelihood inference for clustered binary data, *Communications in Statistics : Theory and Methods*, 26, 2743-2767.

GEYS H., MOLENBERGHS G., RYAN L. (1999), Pseudo-likelihood modelling of multivariate outcomes in developmental toxicology, *Journal of the American Statistical Association*, 94, 34-745.

GILKS W.R., WILD P. (1992), Adaptive rejection sampling for Gibbs sampling, *Applied Statistics*, 4, 337-348.

GILMOUR A.R., ANDERSON R.D., RAE A.L. (1985), The analysis of binomial data by a generalized linear mixed model, *Biometrika*, 72, 593-599.

GILULA Z., HABERMAN S. (1994), Conditional log-linear models for analyzing categorical panel data, *Journal of the American Statistical Association*, 89, 645-656.

GLONEK G.F.V., McCULLAGH P. (1995), Multivariate logistic models, *Journal of the Royal Statistical Society, Series B*, 57, 533-546.

GOLDSTEIN H. (1986), Multilevel mixed linear model analysis using iterative generalized least squares, *Biometrika*, 73, 43-56.

GOLDSTEIN H. (1991), Nonlinear multilevel models, with an application to discrete response data, *Biometrika*, 73, 43-56.

GOLDSTEIN H. (1995), *Multilevel Statistical Models* (2nd edition), London : Edward Arnold.

GOLDSTEIN H., RASBASH J. (1996), Improved approximations for multilevel models with binary responses, *Journal of the Royal Statistical Society A*, 159, 505-513.

HEDEKER D., GIBBONS R.D. (1993), *MIXOR : a Computer Program for Mixed Effects Ordinal Probit and Logistic Regression*, Chicago : University of Illinois.

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

HEDEKER D., GIBBONS R.D. (1996), MIXOR : A computer program for mixed-effects ordinal regression analysis, *Computer Methods and Programs in Biomedicine*, 49, 157-176.

KLEINMAN J. (1973), Proportions with extraneous variance : single and independent samples, *Journal of the American Statistical Association*, 68, 46-54.

KREFT I., DE LEEUW J. (1998), *Introducing Multilevel Modeling*, London : Sage Publications.

KUK A.Y.C. (1995), Asymptotically unbiased estimation in generalised linear models with random effects, *Journal of the Royal Statistical Society B*, 57, 395-407.

KUPPER L.L., HASEMAN J.K. (1978), The use of a correlated binomial model for the analysis of certain toxicology experiments, *Biometrics*, 34, 69-76.

LANG J.B., AGRESTI A. (1994), Simultaneously modeling joint and marginal distributions of multivariate categorical responses, *Journal of the American Statistical Association*, 89, 625-632.

LAVERGNE C., TROTTIER C. (2000), Sur l'estimation dans les modèles linéaires généralisés à effets aléatoires, *Revue de Statistique Appliquée*, 48, 49-67.

LE CESSIE S., VAN HOUWELINGEN J.C. (1994), Logistic regression for correlated binary data, *Applied Statistics*, 43, 95-108.

LESAFFRE E., SPIESSENS B. (2001), On the effect of the number of quadrature points in a logistic random-effects model : an example, *Applied Statistics*, 50, 325-336.

LIANG K.-Y., ZEGER S.L. (1986), Longitudinal data analysis using generalized linear models, *Biometrika*, 73, 13-22.

LIANG K.Y., ZEGER S.L., QAQISH B. (1992), Multivariate regression analyses for categorical data, *Journal of the Royal Statistical Society, Series B*, 54, 3-40.

LIN X., BRESLOW N.E. (1996), Bias correction in generalized linear mixed models with multiple components of dispersion, *Journal of the American Statistical Association*, 91, 1007-1016.

LITTLE R.J.A., RUBIN D.B. (1987), *Statistical Analysis with Missing Data*, New York : John Wiley.

LONGFORD N.T. (1993), *Random Coefficient Models*, London : Oxford University Press.

MCCULLAGH P., NELDER J.A. (1989), *Generalized Linear Models*, London : Chapman & Hall.

MOLENBERGHS G., DECLERCK L., AERTS M. (1998), Misspecifying the likelihood for clustered binary data, *Computational Statistics and Data Analysis*, 26, 327-349.

MOLENBERGHS G., 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., LESAFFRE E. (1999), Marginal modelling of multivariate categorical data, *Statistics in Medicine*, 18, 2237-2255.

MOLENBERGHS G., RYAN L.M. (1999), Likelihood inference for clustered multivariate binary data, *Environmetrics*, 10, 279-300.

NELDER J.A., WEDDERBURN R.W.M. (1972), Generalized linear models, *Journal of the Royal Statistical Society, Series B*, 135, 370-384.

NEUHAUS J.M. (1992), Statistical methods for longitudinal and clustered designs with binary responses, *Statistical Methods in Medical Research*, 1, 249-273.

NEUHAUS J.M., KALBFLEISCH J.D., HAUCK W.W. (1991), A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data, *International Statistical Review*, 59, 25-35.

OCHI Y., PRENTICE R.L. (1984), Likelihood inference in a correlated probit regression model, *Biometrika*, 71, 531-543.

OLKIN I., TATE R.F. (1961), Multivariate correlation models with mixed discrete and continuous variables, *Annals of Mathematical Statistics*, 32, 448-465 (with correction in 36, 343-344).

PINHEIRO J.C., BATES D.M. (2000), *Mixed effects models in S and S-Plus*, New-York : Springer-Verlag.

PLACKETT R.L. (1965), A class of bivariate distributions, *Journal of the American Statistical Association*, 60, 516-522.

RIPLEY B.D. (1981), *Spatial Statistics*, New York : John Wiley.

RODRÍGUEZ G., GOLDMAN N. (1995), An assessment of estimation procedures for multilevel models with binary responses, *Journal of the Royal Statistical Society A*, 158, 73-89.

ROSNER B. (1984), Multivariate methods in opthalmology with applications to other paired-data, *Biometrics*, 40, 1025-1035.

SANTNER T.J., DUFFY D.E. (1989), *The Statistical Analysis of Discrete Data*, New York : Springer-Verlag.

SCHALL R. (1991), Estimation in generalised linear models with random effects, *Biometrika*, 78, 719-727.

SKELLAM J.G. (1948), A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials, *Journal of the Royal Statistical Society, Series B*, 10, 257-261.

STIRATELLI R., LAIRD N., WARE J.H. (1984), Random-effects model for serial observations with binary response, *Biometrics*, 40, 961-971.

TEN HAVE T.R., LANDIS R., WEAVER S.L. (1995), Association models for periodontal disease progression : A comparison of methods for clustered binary data, *Statistics in Medicine*, 14, 413-429.

VERBEKE G., MOLENBERGHS G. (2000), *Linear Mixed Models for Longitudinal Data*, New York : Springer-Verlag.

WILLIAMS D.A. (1975), The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity, *Biometrics*, 31, 949-952.

WOLFINGER R., O'CONNELL M. (1993), Generalized linear mixed models : A pseudo-likelihood approach, *Journal of Statistical Computing and Simulation* 48, 233-243.

ZEGER S.L., KARIM M.R. (1991), Generalised linear models with random effects : a Gibbs sampling approach, *Journal of the American Statistical Association*, 86, 79-102.