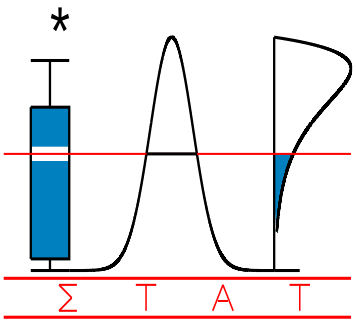
T E C H N I C A L

R E P O R T

0657

GENERALIZED LINEAR MIXED MODELS WITH NON-GAUSSIAN RANDOM EFFECTS

LITIERE S., ALONSO A., and G. MOLENBERGHS



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Generalized linear mixed models with non-Gaussian random effects

Saskia Litière[*1], **Ariel Alonso**[1], and **Geert Molenberghs**[1]

[1] Hasselt University, Center for Statistics, Diepenbeek, Belgium

Summary

Recent research is showing that generalized linear mixed models (GLMM) may not always be robust against certain model misspecifications. In this work we focus on misspecifying the random-effects distribution and its impact on maximum likelihood estimation. We propose to deal with possible misspecification by way of sensitivity analysis, considering several random-effects distributions. First, we analyze a case study using the heterogeneity model, i.e., a GLMM where the normal random-effects distribution is replaced by a finite mixture of normals. It is shown through simulations that this model performs slightly better in the presence of misspecification. However, the model can be very unstable and convergence is sometimes hard to obtain. We therefore complete the sensitivity analysis of the case study with a Bayesian approach, where we fit logit models with different distributions for the random effects. Here the Deviance Information Criterion (DIC) can be used as a criterion to select the most appropriate model.

Key words: Hierarchical models, Misspecification, Random-effects, Maximum likelihood, Heterogeneity model, Bayesian modeling.

1 Introduction

When dealing with non-Gaussian longitudinal measurements, observations on a subject are unlikely to be independent. This dependence can be taken into account using subject-specific parameters. A popular

* Corresponding author: e-mail: saskia.litiere@uhasselt.be, Phone: +32(0)11 268282, Fax: +32(0)11 268299

approach to handle this type of correlated data is the generalized linear mixed model (GLMM; Agresti, 2002; Diggle et al., 2002; Fahrmeir and Tutz, 2001; Molenberghs and Verbeke, 2005). This model has been widely used in different areas like, e.g., toxicology (Molenberghs and Verbeke, 2005), epidemiology (Kleinman, Lazarus and Platt, 2004), dairy science (Tempelman, 1998), etc., and is easy to apply using software tools such as the SAS procedures NLMIXED and GLIMMIX. Conditional on the random effects $\boldsymbol{b}_i$, the outcome variable $\boldsymbol{Y_i}$ for a subject i follows a pre-specified distribution $F_i(\boldsymbol{y}_i|\boldsymbol{\varphi}, \boldsymbol{b}_i)$, parameterized through a vector $\boldsymbol{\varphi}$ of unknown parameters common to all subjects. The subject-specific effects $\boldsymbol{b}_i$ are assumed to come from a distribution $G(\boldsymbol{b}_i|\boldsymbol{\delta})$, which may depend on a vector $\boldsymbol{\delta}$ of unknown parameters. Estimation is usually based on maximum likelihood, assuming that the underlying probability model is correctly specified.

A wide range of software tools are available for fitting these models. However, the analysis is often limited to the setting of Gaussian random effects. Since random effects are not observed, diagnostic tools to study the random-effects distribution are not straightforward. Indeed, one should be careful in using empirical Bayes estimates of the random effects to detect departures from normality. Even when the random effects are coming from a normal distribution, the empirical Bayes estimates will usually not be normally distributed. Therefore, it is relevant to assess the robustness of the parameter estimates with respect to this type of model misspecification.

Verbeke and Lesaffre (1997) showed that the maximum likelihood estimators for fixed effects and variance components in linear mixed models, obtained under the assumption of normally distributed random effects, are consistent and asymptotically normally distributed, even when the random-effects distribution is not normal. Nevertheless, it could be argued that the work presented by Verbeke and Lesaffre (1997) provides only asymptotic results whereas in practice one has to deal with limited sample sizes, so the discussion of the properties of estimators under finite sample sizes is also a very important one. Relevant results in this area include Pinheiro et al. (2001), which discusses an alternative robust approach by implementing linear mixed models with multivariate t-distributions for the random-effects and the within-subject

errors. Rosa, Gianola and Padovani (2004) extended this approach by using normal/independent distributions, such as the Student-t, the slash, and contaminated normal distributions, as robust alternatives to the Gaussian assumption for both the residuals and the random effects. These authors adopted a Bayesian framework with Markov chain Monte Carlo methods to carry out the posterior analysis.

In the present work our main focus will be on the robustness of generalized linear mixed models with respect to the misspecification of the random-effects distribution. Some work has been done to study the impact on the parameter estimates of these models. Neuhaus et al. (1992) examined the performance of mixed-effects logistic regression models with misspecified random-effects distributions. They showed that the maximum likelihood estimators of the model parameters are inconsistent but that the magnitude of the bias is typically small. More recently, Heagerty and Kurland (2001) studied the impact of the misspecification of the random-effects distribution on the maximum likelihood estimators of the regression coefficients in logistic regression models for clustered binary response data. These authors found that the marginal regression parameters are much less susceptible to the misspecification than the parameters of the corresponding hierarchical model. In Agresti et al. (2004) the choice of the random-effects distribution seems to have little effect on the maximum likelihood estimators. A similar message can be found in some very well known texts for the analysis of categorical data like Agresti (2002), where it is stated that usually, assuming normality does not hurt when the true distribution is not normal.

It should be noted that most of these studies used random-effect distributions with small variances. For example, in Agresti et al. (2004) the largest random-effects variance used for simulations was equal to 1. As we will illustrate with our case study in Section 2, these small values may not always be realistic. When dealing with larger random-effects variances, Litière et al. (2005a) have found that misspecification of the random-effects distribution can introduce a relative bias of up to 30% for the mean structure parameters and up to 70% for the variance components. Additionally, these authors have shown that the type I and II errors can be seriously affected, depending on the shape of the true random-effects distribution.

This overview illustrates the wide range of opinions that exist in the literature with respect to the impact of misspecifying the random-effects distribution. Nevertheless, it becomes clear that more research is needed to find models that are robust against this type of misspecification. Some alternative robust approaches have been suggested. Butler and Louis (1992) proposed to replace the normal random-effects distribution by a non-parametric distribution. However, Agresti et al. (2004) reported that there can be some loss of efficiency, when using a non-parametric approach, compared to a parametric assumption close to the real distribution. Additionally, model comparison can be difficult as standard asymptotic theory does not apply. Chen, Zhang and Davidian (2002) suggested a semi-parametric random-effects distribution, allowing the random-effects density to be skewed, multi-modal, fat- or thin-tailed and including the normal as a special case. These authors then used a Monte Carlo EM algorithm with a rejection sampling scheme to obtain estimates of the model parameters and the semi-parametric random-effects distribution.

Furthermore, Kizilkaya et al. (2003) discussed a hierarchical threshold mixed model based on a cumulative t-link specification for the analysis of ordinal data. Additionally, Kizilkaya and Tempelman (2005) proposed a general Bayesian approach to model heteroskedastic error in GLMM, in which linked functions of conditional means and residual variances were specified as separate linear combinations of fixed and random effects. Finally, the heterogeneity model, where the normal random-effects distribution is replaced by a finite mixture of normals, has been suggested as another plausible alternative (Fieuws, Spiessens and Draney, 2004; Molenberghs and Verbeke, 2005). This allows one to cover a wide range of shapes for the random-effects distribution. In the present work we will study the performance of this alternative using both real and simulated data.

On the other hand, Bayesian models are very flexible in the choice of the random-effects distribution and have been used in the past to deal with possible random-effects misspecification (Pinheiro et al., 2001; Kizilkaya et al., 2003; Rosa et al., 2004). Therefore, in this case a Bayesian approach might be a good alternative to the classic frequentist methods.

We will start by analyzing in Section 2 the motivating case study using a GLMM with normal random effects. In Section 3, we will introduce and apply the heterogeneity model to the example. Next, in Section 4, various aspects of the parameter estimates resulting from the heterogeneity model are investigated through extensive simulations. Finally, in Section 5, we will consider a Bayesian approach to fitting GLMM with non-Gaussian random effects and perform a sensitivity analysis for our case study within the Bayesian framework.

2 Case study: the schizophrenia data

The study consists of individual patient data from a randomized clinical trial, comparing the effect of risperidone to conventional antipsychotic agents for the treatment of chronic schizophrenia (Alonso et al., 2004). The response variable Y is a dichotomous version of the Clinical Global Impression (CGI) scale and equals 1 for patients classified as normal to mildly ill, and 0 for patients classified as moderately to severely ill. The treatment variable, Z, is set to 0 for the control group and 1 for the risperidone group. Treatment was administered for 8 weeks and the outcome was measured at 6 fixed time points: at the beginning of the study and after 1, 2, 4, 6 and 8 weeks. One hundred twenty-eight patients were included in the study.

The following random-intercept logistic model was fitted, using maximum likelihood:

$$\text{logit}\{P(Y_{ij}=1|b_i)\} = \beta_0 + \beta_1 Z_i + \beta_2 t_j + b_i, \tag{1}$$

where Y_{ij} denotes the response for the i-th patient at time t_j and b_i denotes a random intercept assumed to follow a mean zero normal distribution with variance σ_b^2. The results are shown in Table 1 and they give evidence of certain treatment effect.

Taking into account the problems described in Section 1, it becomes clear that these results can be influenced by possible model misspecifications, like the choice of the random-effects distribution. Naturally, we are concerned with the impact of such model violations on the parameter estimates. Let us first note that

Table 1 Estimates and standard errors for the parameters in Model (1). The SAS procedure NLMIXED with adaptive Gaussian quadrature and 20 quadrature points was used.

	Parameter	Estimate	S.E.	p-value
Regression Coefficients	β_0	-7.369	1.177	< 0.0001
	β_1	2.144	1.079	0.0490
	β_2	0.650	0.096	< 0.0001
Variance Component	σ_b^2	21.01	6.808	

in Litière, Alonso and Molenberghs (2005b) a theoretical result was introduced which states the conditions under which the type I error is robust to misspecification of the random-effects distribution. Applied to our case study, this theorem implies that the type I error corresponding to the treatment effect will not be affected by the choice of the random-effects distribution. Therefore, we can be fairly confident about the presence of a (borderline) significant treatment effect. However, we should be careful when interpreting the estimated size of the effect due to the bias that can be introduced by misspecification.

A plausible alternative approach to the GLMM could consist in replacing the Gaussian random-effects distribution by a finite mixture of normals. Note that mixtures of normal distributions are a very flexible class of distributions as they can cover unimodal as well as multimodal, and symmetric as well as very skewed distributions. Additionally, mixtures can be used to model unobserved heterogeneity that can appear when an important categorical variable has been left out of the model. In the next section we will study the heterogeneity model in more detail.

3 The heterogeneity model

In the homogeneity model considered so far, the random-effects $\boldsymbol{b}_i$ were assumed to be sampled from a normal distribution. The heterogeneity model is an extension of this model, obtained by sampling the random effects from a mixture of k normal distributions with mean vectors $\boldsymbol{\mu}_r$ and covariance matrix $\boldsymbol{D}$, i.e., $\boldsymbol{b}_i \sim \sum_{r=1}^{k} \pi_r N(\boldsymbol{\mu}_r, \boldsymbol{D})$. The probability for a subject to belong to component r is π_r, with $\sum_{r=1}^{k} \pi_r = 1$. Note that each component has the same covariance matrix $\boldsymbol{D}$. This constraint is necessary to avoid unbounded likelihoods (Böhning, 1999).

Let $\boldsymbol{\pi}' = (\pi_1, ..., \pi_k)$ and $\boldsymbol{\gamma}$ be the vector containing the remaining parameters, i.e., the vector $\boldsymbol{\varphi}$ of unknown parameters common to all subjects, as well as all parameters in $\boldsymbol{\mu}_r$ and $\boldsymbol{D}$. The joint density function of $\boldsymbol{y}_i$ can then be written as $f_i(\boldsymbol{y}_i) = \sum_{r=1}^{k} \pi_r f_{ir}(\boldsymbol{y}_i|\boldsymbol{\gamma})$ where

$$f_{ir}(\boldsymbol{y}_i|\boldsymbol{\gamma}) = \int f_i(\boldsymbol{y}_i|\boldsymbol{\varphi}, \boldsymbol{b}_i)\phi_r(\boldsymbol{b}_i)d\boldsymbol{b}_i. \tag{2}$$

Note that $\phi_r(\boldsymbol{b}_i)$ refers to the multivariate normal with mean $\boldsymbol{\mu}_r$ and covariance matrix $\boldsymbol{D}$. Estimation is now based on the maximization of

$$\ell(\boldsymbol{\theta}|\boldsymbol{y}) = \sum_{i=1}^{N} \ln \left\{ \sum_{r=1}^{k} \pi_r f_{ir}(\boldsymbol{y}_i|\boldsymbol{\gamma}) \right\}, \tag{3}$$

where $\boldsymbol{\theta}' = (\boldsymbol{\gamma}', \boldsymbol{\pi}')$, using the Expectation-Maximization (EM) algorithm described in Laird (1978). Initially, the EM algorithm was developed for missing data problems. It has however also been used a lot for estimation in the linear mixed models. For our purpose, the algorithm is very useful if we treat the component membership indicator z_{ir}, defined as

$$z_{ir} = \begin{cases} 1 & \text{if } \boldsymbol{b}_i \text{ is sampled from the } r\text{th component in the mixture} \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

as missing. Using these indicators, the log-likelihood function can be rewritten as

$$\ell(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{z}) = \sum_{i=1}^{N} \sum_{r=1}^{k} z_{ir}[\ln \pi_r + \ln f_{ir}(\boldsymbol{y}_i|\boldsymbol{\gamma})], \tag{5}$$

where $\boldsymbol{z}$ is the vector of all unobserved z_{ir}. This function is easier to maximize, however maximizing $\ell(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{z})$ with respect to $\boldsymbol{\theta}$ will lead to estimates of $\boldsymbol{\theta}$ which depend on the unobserved indicators z_{ir}. To avoid this, it has been suggested to use the EM algorithm so that the expected value of (5) rather than $\ell(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{z})$ itself, will be maximized with respect to $\boldsymbol{\theta}$ (with the expectation taken over all unobserved z_{ir}). More specifically, in the E step (expectation) the conditional expectation of $\ell(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{z})$, given the observed data $\boldsymbol{y}$, is determined. In the M step (maximization), the so-obtained expected log-likelihood function is maximized with respect to $\boldsymbol{\theta}$, providing an updated estimate for $\boldsymbol{\theta}$. The algorithm is repeated until the difference between two successive loglikelihood evaluations is small enough.

To fit heterogeneity models in practice, a SAS macro (described in Fieuws et al., 2004) has been implemented, based on the SAS procedure NLMIXED and the EM algorithm. The macro can be used for fitting nonlinear and generalized linear mixed models with finite normal mixtures as random-effects distributions. In the next subsection we will use this macro to re-analyze the case study.

3.1 Case study - the heterogeneity model

To fit a heterogeneity model to the case study, some small changes have to be made to the model formulation. For example, since there are no restrictions on the $\boldsymbol{\mu}_r$, the expected value of the random effects is no longer fixed at zero. Therefore, to avoid overparameterization, we will not include an intercept:

$$\text{logit}\{P(Y_{it} = 1|b_i)\} = \beta_1 Z_i + \beta_2 t + b_i. \tag{6}$$

Note that now β_0 equals $\sum_{r=1}^{k} \pi_r \mu_r$. We will consider random-effects distributions with two components

$$b_i \sim \pi_1 N(\mu_1, d) + (1 - \pi_1) N(\mu_2, d), \tag{7}$$

as well as with three components

$$b_i \sim \pi_1 N(\mu_1, d) + \pi_2 N(\mu_2, d) + (1 - \pi_1 - \pi_2) N(\mu_3, d). \tag{8}$$

Table 2 Estimates and standard errors for the parameters of model (6): mixture with 2 components - mixture with 3 components.

		2 components			3 components		
		Estimate	S.E.	p-value	Estimate	S.E.	p-value
Regression Coefficients	β_0	-7.882	1.227	< 0.0001	-7.774	4.283	0.0695
	β_1	1.992	0.937	0.0336	2.703	0.854	0.0015
	β_2	0.665	0.096	< 0.0001	0.678	0.094	< 0.0001
	μ_1	-9.189	1.333	< 0.0001	-10.757	5.374	0.0453
	μ_2	-5.306	0.953	< 0.0001	-6.510	1.969	0.0010
	μ_3				-2.951	2.031	0.1461
	π_1	0.664	0.007		0.506	0.453	
	π_2				0.245	0.575	
Variance Component	d	24.664	7.658		9.525	11.883	
	AIC			395.1			396.0

For these mixtures the overall variance of the random-effects corresponds to

$$\sigma_b^2 = \sum_{j=1}^{k} \pi_j \mu_j^2 - \left(\sum_{j=1}^{k} \pi_j \mu_j \right)^2 + d, \tag{9}$$

with $k = 2$ or 3, depending on the number of components in the mixture. The parameter estimates for these two models are shown in Table 2. The estimates of the fixed effects for both models are very

similar, and they are also close to the results shown in Table 1. The overall variance of the random effects can be calculated using (9) and the estimates given in Table 2. This leads to $\hat{\sigma}_b^2 = 28.03$ for the two-component model, and $\hat{\sigma}_b^2 = 20.21$ for the three-component model. These results are in the same order of magnitude of $\hat{\sigma}_b^2 = 21.01$ reported in Table 1. Additionally, there is not so much difference between the AIC of these models and the AIC corresponding to the homogeneity model, given by 391.0. Therefore, the homogeneity model seems to be most appropriate for the data at hand. Evidently, this result increases the level of confidence in our previous findings.

As illustrated here, the heterogeneity model could be a plausible alternative or extension of the classic GLMM. To explore the performance of this model under misspecification of the random-effects distribution, we have carried out a simulation study that will be presented in the next section.

4 Simulations

In this simulation study, binary data were generated using Expression (1). For the fixed effects, values close to the estimates in Table 1 were chosen: $\beta_0 = -8$, $\beta_1 = 2$ and $\beta_2 = 1$, whereas the variance of the correct random-effects distribution was fixed at $\sigma_b^2 = 32$, to be in the same order of magnitude as the variance estimate in Table 1.

Further, we considered the following distributions for the random effects: the mean zero normal distribution with variance equal to 32; the uniform distribution with support between -9.8 and 9.8; the exponential distribution with $\lambda = 0.177$; the lognormal distribution with scale parameter 0 and shape parameter 1.35; the power function distribution with shape parameter 464 and scale parameter 80; and the asymmetric mixture $0.231 \times N(-10, 1.407^2) + 0.769 \times N(3, 1.407^2)$. If necessary, the distributions were transformed to satisfy the mean zero condition of the random effects. These distributions cover a wide range of shapes varying from very symmetric to very skewed; with potentially very heavy tails. We considered 3 different sample sizes: 50, 100 and 200, and for each of these settings 100 data sets were generated. Model (6) was then fitted to the generated data assuming a mixture of two normals for the random-effects distribution.

Consistency was studied through the evolution of the relative distance between the estimates and their real value, over increasing sample size. Let $\boldsymbol{\gamma}_0 = (-8, 2, 1, 32)'$ represent the vector of true parameter values and $\hat{\boldsymbol{\gamma}}_n = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}_b^2)'$ the corresponding vector of maximum likelihood estimates. We can then define the relative distance between $\boldsymbol{\gamma}_0$ and $\hat{\boldsymbol{\gamma}}_n$ as

$$d_\gamma = \frac{||\hat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_0||}{||\boldsymbol{\gamma}_0||}, \tag{10}$$

where $||.||$ denotes the Euclidean distance. The relative distance between the treatment effect estimate and its real value is similarly given by

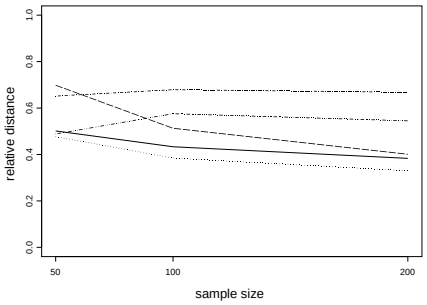
$$d_{\beta_1} = \left| \frac{\hat{\beta}_1 - \beta_1^0}{\beta_1^0} \right|. \tag{11}$$

If the estimators remain consistent after misspecifying the model then these relative distances should go to zero with increasing sample sizes.

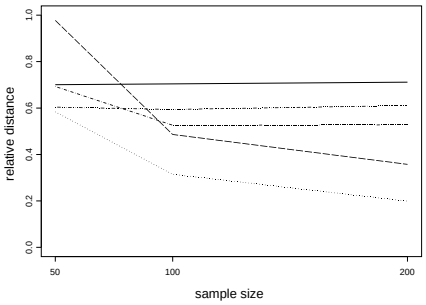
4.1 Results of the simulations

Figure 1(a) shows the evolution of the relative distance between $\boldsymbol{\gamma}_0$ and $\hat{\boldsymbol{\gamma}}_n$. The smallest relative bias is observed for the correctly specified models, i.e., when the random-effects distribution is normal or an asymmetric mixture of normals. However, due to misspecification, the relative bias of $\hat{\boldsymbol{\gamma}}_n$ can be as high as 67%, as was the case for the lognormal random effects.

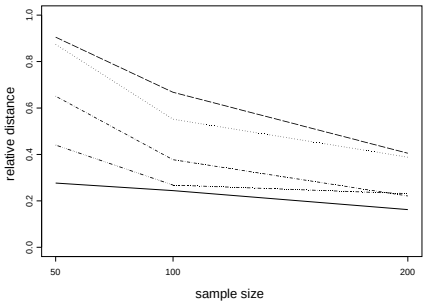
To be able to compare these results with the results from fitting the homogeneity, or one-component model, the simulations were repeated. We considered the same settings as before and generated 500 data sets for each setting (see also Litière et al., 2005a). Model (1) was then fitted to the generated data using maximum likelihood and assuming normal random effects (using the SAS procedure NLMIXED with gaussian quadrature and 50 quadrature points). The corresponding relative distances d_γ are displayed in Figure 1(b). Comparing this graph with Figure 1(a), it is clear that the heterogeneity model performs better than the homogeneity model, when the random-effects distribution is an asymmetric mixture of normals. Additionally, the heterogeneity model seems to perform better for smaller sample sizes. Figure 1(a) clearly



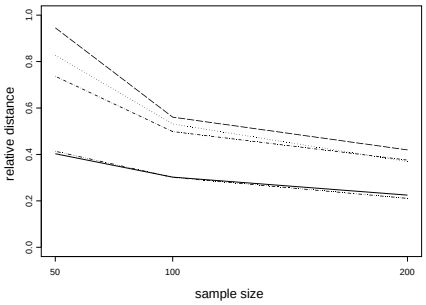
(a) Heterogeneity model - d_γ



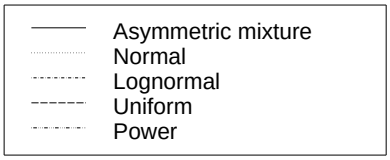
(b) Homogeneity model - d_γ



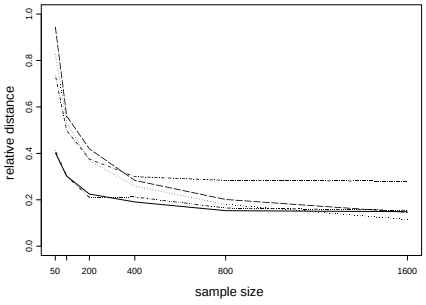
(c) Heterogeneity model - d_{β_1}



(d) Homogeneity model - d_{β_1}



(e) Legend



(f) Homogeneity model - evolution d_{β_1} for large sample sizes

Fig. 1 Consistency of the parameter estimates under homo- and heterogeneity models: the behavior of the relative distance d_γ and d_{β_1} for each distribution over increasing sample size.

illustrates that when the sample size is 50, the overall bias exceeds 60% only when the random effects come from a lognormal or uniform distribution. On the other hand, Figure 1(b) shows that for the homogeneity model the overall bias is above 60% for all the misspecified random-effects distributions and it is near 100% when the random effects come from a lognormal distribution. The differences between both models are less dramatic for larger sample sizes.

Since interest often lies in the observed treatment effect, we studied in more detail how its estimate can be influenced by the misspecification. Figures 1(c) and (d) show the relative distance between the treatment effect and its estimate from fitting a heterogeneity and a homogeneity model respectively. Here again, the heterogeneity model seems to be more efficient in the estimation of this parameter, especially when the sample size is small. Therefore, there seems to be some benefit (in terms of robustness) when using the heterogeneity model for estimating the mean structure parameters.

Additionally, the simulations showed (results not included here) that both models are very robust to the random-effects misspecification when estimating the time effect. However, there was some substantial bias when estimating the variance of the random-effects (up to 73% for the homogeneity model, when the random-effects distribution was an asymmetric mixture, and 69% for the heterogeneity model with lognormal random effects, both for sample size 200). Using the heterogeneity model considerably improved the bias in the case of the asymmetric mixture of normals, and in general the model seems to perform better again for small sample sizes.

Markedly Figure 1(d) shows large relative distances for β_1, even under the correctly specified model with normal random effects. This result is counterintuitive since, under these conditions, the maximum likelihood method is expected to provide consistent estimates. However, it should be emphasized that consistency is an asymptotic result. To explore this issue in more detail we have extended the simulations with the homogeneity model to include larger sample sizes like 400, 800 and 1600. The corresponding results for the relative distances d_{β_1} are shown in Figure 1(f). From this graph it is clear that with increasing

sample sizes, the curve corresponding to the correctly specified (normal) model slowly decreases, while the other curves stabilize.

We can conclude that, in general, the heterogeneity model seems to perform somewhat better than the homogeneity model. Therefore in any practical situation, this model is worthy of consideration. However, reaching convergence can be very difficult. In comparison, convergence of the homogeneity model was 100% in all the considered scenarios. Additionally, the heterogeneity model is also very sensitive to the choice of starting values. These issues limit the models' applicability in practice, reducing also their value as a feasible alternative to the GLMM.

One might also be interested in the performance of the probit link versus the logit link. Suffice it to say that in most conditions performances of both are very similar, due to the approximate conversion factor between parameters from both models: $\pi/\sqrt{3}$ (Molenberghs and Verbeke, 2005). The choice for the logit rather than the probit link in our study is inspired by the omnipresence of the normal random-effect logistic model.

5 A Bayesian approach

An alternative sensible approach to dealing with possible random-effects misspecification could be to perform a sensitivity analysis, considering different random-effects distributions. The Bayesian paradigm provides a natural framework for this type of analysis. The Markov Chain Monte Carlo (MCMC) algorithm allows considerable flexibility and has made Bayesian models a popular and efficient tool in the analysis of hierarchical data (Gilks et al., 1996, and Gelman et al., 2005).

The Bayesian analyst needs to explicitly state prior information $p(\boldsymbol{\gamma})$ on the model parameters $\boldsymbol{\gamma} = (\boldsymbol{\varphi}, \boldsymbol{\delta})$ of the generalized linear mixed model, independent from the evidence given by the data. Additionally, he needs to specify the support for different values of the parameter effects based on the available data, i.e., the likelihood $p(\boldsymbol{y}|\boldsymbol{\gamma})$. Together, these two sources of information are combined into a posterior distribution $p(\boldsymbol{\gamma}|\boldsymbol{y})$ for the parameters of interest. Whereas maximum likelihood estimation leads to a

point estimate for the fixed effects and the variance components of the random-effects distribution, a full Bayesian approach will quantify our uncertainty about the unknown parameters through a whole posterior distribution. Samples drawn from the joint posterior distribution then allow to estimate characteristics of the joint and marginal posterior distribution like, e.g. posterior means of the parameters of interest.

A popular method for sampling from the posterior distribution is the Gibbs sampler (Zeger and Karim, 1991, and Natarajan and Kass, 2000). This algorithm proceeds by iteratively sampling $\boldsymbol{\gamma}$ from its full-conditional distribution until convergence is achieved. It is currently implemented in a software called WinBUGS, an acronym for "Bayesian inference Using Gibbs Sampling" on a Windows platform.

For a detailed exploration and application of Bayesian model formulation, model parametrization, choice of prior distribution, diagnosing convergence, comparison between models and model adequacy for binary longitudinal data, we refer to Albert and Jais (1998). A Bayesian formulation of the model mentioned in Section 1 requires specification of prior distributions for the parameters $\boldsymbol{\varphi}$ and $\boldsymbol{\delta}$. Whether estimation of the random effects is of interest or not, the prior information of $\boldsymbol{\delta}$ is necessary for inference about the regression coefficients. Usually, when no prior information is available, vague priors are assumed. In the next subsection we will continue the analysis of the case study by expanding the sensitivity analysis to a wide range of non-normal random-effects distributions.

5.1 Case study - the Bayesian approach

In this section we will revisit the case study introduced in Section 2. We will use Bayesian methods as implemented in WinBUGS 1.3 to fit GLMM, considering different random-effects distributions. We will include a mean zero normal with precision τ (note that $\tau = 1/\sigma^2$), an exponential with parameter λ, a chi-square with k degrees of freedom, a lognormal with scale parameter 0 and precision parameter τ, a uniform with support between $-b$ and b and a discrete distribution with unequal probability π at two support points X_i, $i = 1, 2$. Furthermore, we will choose vague priors for the parameters of the random-effects distributions. For example, a gamma distribution was chosen for τ, k, λ and b; the X_i

are assumed to be sampled from a mean zero normal distribution; and the prior for the probability π is a uniform distribution with support between 0 and 1. For each setting, the Gibbs sampler had 3 chains with 10000 iterations as the burn-in period, plus 100000 additional iterations with a thinning interval of 10 for the normal, exponential and lognormal, 30 for the uniform, 60 for the chi-square and 75 for the discrete distributions.

Note that the considered random-effects distributions do not necessarily have mean zero. It has been suggested that this parametrization can improve convergency and stability of the samples by reducing the autocorrelations of the Gibbs chains (Albert and Jais, 1998). Also note that keeping the intercept in the model (in contrast to the approach in Section 3) improved the convergence of the models. To be able to compare the new estimates with the previously obtained results, we will use the following model

$$\text{logit}\{P(Y_{ij} = 1|b_i)\} = \alpha + \beta_1 Z_i + \beta_2 t_j + b_i, \tag{12}$$

such that $\beta_0 = \alpha + E(b)$.

Convergency of the models was studied through the trace plots of the sample values of the main parameters. As the number of iterations increases the trace plots should stabilize, varying randomly around a mean value. Additionally we have studied the Gelman-Rubin diagnostic tool (Gelman et al., 1996), which uses several parallel chains with widely dispersed starting values with respect to the true posterior distribution, to check convergence. This diagnostic tool compares the variability between- and within-chains by estimating a scale reduction factor. If the variance between the different chains is not larger than the variance within each individual chain, then approximate convergence can be diagnosed.

The parameter estimates of fitting Model (12) are shown in Table 3. Noticeably, a lot of variability can be observed in the estimation of β_0 and σ_b^2. As one would intuitively expect, the variance components estimates seem to be very sensitive to the choice of the random-effects distribution. However, the estimates for the treatment and time effects are similar in all considered settings, and they are also similar to the results from the homogeneity and the heterogeneity model. Note that Table 3 also contains the Monte

Table 3 Parameter estimates (standard error and MC error between parenthesis) for model 1 and the corresponding deviance information criterion (DIC) for the different random-effects distributions using MCMC in Winbugs.

Distribution	β_0 (S.E., MC Error)	β_1 (S.E., MC Error)	β_2 (S.E., MC Error)	σ_b^2 (S.E., MC Error)	DIC
Normal	-7.905 (1.325, 0.023)	2.322 (1.208, 0.018)	0.680 (0.100, 0.001)	26.28 (9.58, 0.131)	275.72
Exponential	-6.448 (1.160, 0.023)	2.098 (1.203, 0.021)	0.660 (0.098, 0.001)	19.09 (7.72, 0.141)	275.27
Chi-square	-12.800 (3.081, 0.191)	2.390 (1.251, 0.036)	0.683 (0.099, 0.003)	10.86 (4.16, 0.261)	276.60
Uniform	-7.828 (1.427, 0.079)	1.507 (0.994, 0.014)	0.672 (0.103, 0.003)	27.04 (9.966, 0.612)	276.20
Lognormal	-4.305 (1.160, 0.009)	1.377 (0.823, 0.008)	0.5764 (0.085, 0.0008)	247.4 (1786.0, 11.24)	290.38
Discrete	-4.846 (0.599, 0.011)	1.246 (0.507, 0.009)	0.528 (0.077, 0.001)	60.45 (67.86, 4.244)	344.14

Carlo error, which can be used to assess the accuracy of the posterior estimates. This error decreases as the sample size used for posterior inference increases.

A useful tool to select the model that fits our data best, is given by the Deviance Information Criterion (DIC; Spiegelhalter et al., 2002). It is similar to the Akaike Information Criterium (AIC), i.e., a compromise between the deviance and the number of parameters in the model. Smaller values are better. The DIC values of the models with the different random-effects assumptions are shown in the last column of Table 3. Here, the models that assume a normal and exponential distribution for the random effects seem to perform best and produce very similar estimates for the treatment effect. Therefore, we can still be very confident about the results obtained from the homogeneity model for the treatment effect.

6 Discussion

In contrast to the conventional wisdom amongst data analysts, recent research is showing that the choice of the random-effects distribution can be crucial to the quality of inference about regression coefficients. Indeed, unlike for the linear mixed model, misspecifying the random-effects distribution in GLMM leads

to inconsistent estimators for both the mean and the covariance structure. At the present time fully robust alternatives are not available for the analyst and therefore we strongly suggest exploring the impact of such misspecifications using a sensitivity analysis. Along with these ideas, we have focused in the current work on two possible approaches to do a sensitivity analysis. The heterogeneity model could be, in certain circumstances, a plausible choice, especially when dealing with small sample sizes. However, our simulations clearly showed that the model can be unstable and convergency can heavily depend on initial values. Even though the heterogeneity model performed better than the GLMM, we still observed serious bias under certain model misspecifications. Additionally, considering more than two components in the mixture can become computationally unfeasible in most practical situations.

The second alternative considered in this work was a Bayesian approach to model hierarchical data. Bayesian models have become easy to apply in practice with the implementation of the MCMC algorithm in free software like WinBUGS. With its flexibility in the choice of the random-effects distribution and the implementation of the DIC to choose the most appropriate model, this approach offers a natural way of implementing a sensitivity analysis, as illustrated in Section 5. Note that for this analysis we did not use the latest available version of WinBUGS. Our models did not run in version 1.4 due to some changes in the update order of the parameters. The authors of the software are aware of the fact that some models which run in version 1.3 are running slowly or not at all in version 1.4. They are currently trying to deal with this issue. Since WinBUGS 1.4 allows the automation of routine analysis, it should therefore be conceivable in the near future to study the effectiveness of an analysis with non-Gaussian random effects in more detail via simulations.

In this work we have confined attention to the impact of misspecifying the random-effects distribution. However, misspecifications of other model aspects deserve a great deal of research attention too. It is becoming clear that there probably will not be a general easy answer on how to deal with model misspecification. Perhaps in some specific situations, good alternative models can be found by using e.g., random-effects distributions conjugate to the distribution of the outcome (Lee and Nelder, 1996). Still, an

important topic for future research will be the development of diagnostic tools for detecting the lack of consistency and therefore the need for alternative models. These tools, together with the ability to consider several random-effects distributions, would allow for a useful and, arguably, necessary sensitivity analysis.

Acknowledgements Financial support from the IAP research network nr P5/24 of the Belgian Government (Belgian Science Policy) is gratefully acknowledged.

References

Agresti, A. (2002) *Categorical Data Analysis*, 2nd edn., Hoboken, N.J.: Wiley.

Agresti, A., Caffo, B., and Ohman-Strickland, P. (2004) Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Computational Statistics and Data Analysis* **47** (3), 639-653.

Albert, I. and Jais, J-P. (1998) Gibbs sampler for the logistic model in the analysis of longitudinal binary data *Statistics in Medicine*, **17**, 2905-2921.

Alonso, A., Geys, H., Molenberghs, G., Kenward, M.G., Vangeneugden, T. (2004) Validation of surrogate markers in multiple randomized clinical trials with repeated measurements: Canonical correlation approach. *Biometrics*, **60** (4), 845-853.

Böhning, D. (1999) *Computer-assisted analysis of mixtures and applications: meta-analysis, disease mapping and others*, Number 81 in Monographs on Statistics and Applied Probability, Chapman & Hall/CRC.

Breslow, N.E. and Lin, X. (1995) Bias correction in generalized linear mixed models with a single component of dispersion. *Biometrika*, **82**, 81-91.

Butler, S.M. and Louis, T.A. (1992) Random effects models with non-parametric priors. *Statistics in Medicine*, 11, 1981-2000.

Chen, J., Zhang, D., and Davidian, M. (2002) A Monte Carlo EM algorithm for generalized linear mixed models with flexible random effects distribution. *Biostatistics*, **3**, 347-360.

Diggle, P.J., Heagerty, P., Liang, K-Y., and Zeger, S.L. (2002) *Analysis of Longitudinal Data*, New York: Oxford University Press.

Fahrmeir, L. and Tutz, G. (2001) *Multivariate Statistical Modelling Based on Generalized Linear Models*, Heidelberg: Springer-Verlag.

Fieuws, S., Spiessens, B., and Draney, K. (2004) Mixture models,in P. De Boeck and M. Wilson, editors, *Explanatory item response models: A generalized linear and nonlinear approach*, Statistics for Social Science and Public Policy, chapter 11, 317–340, Springer-Verlag, New York.

Gelman, A. , Carlin, J.B., Stern, H.S., and Rubin, D.B. (1995) *Bayesian data analysis*, Chapman and Hall, London.

Gelfand, A.E., and Smith, A.F.M. (1990) Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 389-409.

Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. (1996) *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London.

Heagerty, P.J. and Kurland, B.F. (2001) Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika*, **88**, 973-985.

Kizilkaya, K., Carnier, P., Albera, A., Bittante, G., and Tempelman, R.J. (2003) Cumulative t-link threshold models for the genetic analysis of calving ease scores. *Genetics Selection Evolution*, **35**, 489-512.

Kizilkaya, K. and Tempelman R.J. (2005) A general approach to mixed effects modeling of residual variances in generalized linear mixed models. *Genetics Selection Evolution*, textbf37, 31-56.

Kleinman, K., Lazarus, R., and Platt, R. (2004) A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism. *American Journal of Epidemiology*, **159**, 217-224.

Laird, N.M. (1978) Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, **73**, 805-811.

Lee, Y. and Nelder, J.A. (1996) Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society, Series B*, **58**, 619-678.

Litière, S., Alonso, A., Molenberghs, G. and Geys, H. (2005a) The impact of a misspecified random-effects distribution on maximum likelihood estimation in generalized linear mixed models. Manuscript submitted for publication.

Litière, S., Alonso, A. and Molenberghs, G. (2005b) Type I and type II error under random-effects misspecification in generalized linear mixed models. Manuscript submitted for publication.

Molenberghs, G. and Verbeke, G. (2005) *Models for discrete longitudinal data*, Springer, New York..

Natarajan, R. and Kass, R.E. (2000) Reference Bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association*, **95**, 227-237.

Neuhaus, J.M., Hauck, W.W., and Kalbfleisch, J.D. (1992) The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika*, **79**, 755-762.

Pinheiro, J.C., Liu, C.H., and Wu, Y.N. (2001) Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t-distribution. *Journal of Computational and Graphical Statistics*, **10**, 249-276.

Rosa, G.J.M, Gianola, D., and Padovani, C.R. (2004) Bayesian longitudinal data analysis with mixed models and thick-tailed distributions using MCMC. *Journal of Applied Statistics*, **31**, 855-873

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, **64**, 583-639.

Tempelman, R.J. (1998) Generalized linear mixed models in dairy cattle breeding. *Journal of Dairy Science*, **81**(5), 1428-1444.

Verbeke, G. and Lesaffre, E. (1997) The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics & Data Analysis*, **53**, 541-556.

Zeger, S.L. and Karim M.R. (1991) Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American Statistical Association*, **86**, 79-86.