

Type I and type II error under random-effects misspecification in  
generalized linear mixed models

Peer-reviewed author version

LITIERE, Saskia; ALONSO ABAD, Ariel & MOLENBERGHS, Geert (2007) Type I and type II error under random-effects misspecification in generalized linear mixed models. In: *BIOMETRICS*, 63(4). p. 1038-1044.

DOI: 10.1111/j.1541-0420.2007.00782.x

Handle: <http://hdl.handle.net/1942/1956>

# Type I and Type II Error Under Random-effects Misspecification in Generalized Linear Mixed Models

Saskia LITIÈRE\*, Ariel ALONSO and Geert MOLENBERGHS

Center for Statistics, Hasselt University,  
Agoralaan, Building D, B3590 Diepenbeek, Belgium.

\* *email*: saskia.litiere@uhasselt.be

## Abstract

Generalized linear mixed models (GLMM) have become a frequently used tool for the analysis of non-Gaussian longitudinal data. Estimation is based on maximum likelihood theory which assumes that the underlying probability model is correctly specified. Recent research is showing that the results obtained from these models are not always robust against departures from the assumptions on which these models are based. In the present work we have used simulations with a logistic random-intercept model to study the impact of misspecifying the random-effects distribution on the type I and II errors of the tests for the mean structure in GLMM. We found that the misspecification can either increase or decrease the power of the tests, depending on the shape of the underlying random-effects distribution, and it can considerably inflate the type I error rate. Additionally, we have found a theoretical result which states that, whenever a subset of fixed-effects parameters, not included in the random-effects structure, equals zero, the corresponding maximum likelihood estimator will consistently estimate zero. This implies that under certain conditions a significant effect could be considered as a reliable result, even if the random-effects distribution is misspecified.

**KEY WORDS:** Logistic random-intercept model; Maximum likelihood; Robustness; Sandwich estimator.

# 1 Introduction

Generalized linear mixed models (GLMM; Agresti, 2002; Diggle et al., 2002; Fahrmeir and Tutz, 2001; Molenberghs and Verbeke, 2005) have become a frequently used tool in the analysis of hierarchical data and have been widely applied in different areas like, e.g., toxicology (Molenberghs and Verbeke, 2005), epidemiology (Kleinman, Lazarus and Platt, 2004), dairy science (Tempelman, 1998), etc. These models are an extension of generalized linear models for non-Gaussian data with multiple sources of variation, and they are easy to apply using software tools like the SAS procedures NLMIXED and GLIMMIX.

Thus far, limited research has been done to study the behavior of the parameter estimators and the performance of the inferential procedures under certain model misspecifications. For instance, due to software limitations, random effects are often assumed to be normally distributed. However, since random effects are not observed, the accuracy of this assumption is difficult to check. So naturally we are concerned with the impact of misspecifying the random-effects distribution on the maximum likelihood estimators. 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 normal distributed, even when the random-effects distribution is not normal. Nevertheless, results obtained in recent years show that this does not hold for generalized linear mixed models (Agresti, Caffo, and Ohman-Strickland, 2004; Litière et al., 2006; Neuhaus, Hauck, and Kalbfleisch, 1992).

In the present work we address the impact of the misspecification of the random-effects distribution on the type I and type II errors of the tests for the mean structure in GLMM. First, the case study that motivated this work will be introduced and analyzed in Section 2. Next, an overview of some results that have appeared in the literature, investigating model

misspecifications in GLMM, will be presented in Section 3. In Section 4, we will study via simulations, the impact of misspecifying the random-effects distribution on the type I and II errors, and we will introduce a theoretical result that gives the conditions under which the type I error is robust against misspecification. Section 5 contains some concluding remarks.

## 2 The Case Study

The case 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). Several measures can be used to assess a patient's global condition. The Clinical Global Impression (CGI) is generally accepted as a subjective clinical measure of change. It is a 7-grade scale used to characterize a subject's mental condition. Our binary response variable  $y$  is a dichotomous version of the CGI scale which equals 1 for patients classified as normal to mildly ill, and 0 for patients classified as moderately to severely ill. Since it has been established that risperidone is most effective at doses ranging from 4 to 6 mg/day, we included only those patients receiving either these doses of risperidone (i.e., the treatment group  $z = 1$ ) or an active control (i.e., the control group  $z = 0$ ). 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 trial. Figure 1 summarizes the probability of being classified as normal to mildly ill ( $P(Y = 1)$ ) by time point and treatment group.

### 2.1 Generalized Linear Mixed Models

Let us denote by  $y_{ij}$  the response of subject  $i$  at time point  $j$ . Further, we assume that, conditional on a vector of individual random effects  $\mathbf{b}_i$  all the outcome variables  $y_{ij}$  are

independent and have density functions belonging to the exponential family

$$f(y_{ij}|\theta_{ij}, \phi) = \exp[\phi^{-1}\{y_{ij}\theta_{ij} - \psi(\theta_{ij})\} + c(y_{ij}, \phi)], \quad (1)$$

where  $\theta_{ij} = \eta(\beta_0 + \mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i)$ ,  $\eta(\cdot)$  denotes a known link function,  $\beta_0$  is an intercept,  $\mathbf{x}_{ij}$  and  $\mathbf{z}_{ij}$  are vectors of covariates,  $\boldsymbol{\beta}$  is a vector of unknown fixed regression coefficients,  $\phi$  is a scale parameter and  $c(\cdot)$  is a function only depending on  $y_{ij}$  and  $\phi$ . Furthermore,  $\psi(\cdot)$  is a function satisfying  $E(y_{ij}) = \psi'(\theta_{ij})$  and  $Var(y_{ij}) = \phi\psi''(\theta_{ij})$ .

The subject-specific effects  $\mathbf{b}_i$  are generally assumed to be normally distributed with mean zero and variance-covariance matrix  $\mathbf{D}$ . The parameters are then estimated by maximizing the marginal likelihood, obtained by integrating out the random effects. In the next section we apply these models to analyze the case study. All the analyses and simulations in this paper were carried out using the SAS procedure NLMIXED, choosing adaptive Gaussian quadrature with 50 quadrature points to approximate the likelihood.

## 2.2 Analysis of the Case Study

We analyzed the data introduced in Section 2 using a random intercept model and considering different link functions and mean structures. The random intercept was assumed to follow a normal distribution with mean zero and variance  $\sigma_b^2$ . In the model building a total of nine models were fitted. These models were constructed as combinations of three link functions (i.e., the logit, log-log and probit link) and three different mean structures. For the mean structure we considered i) intercept, treatment, time and treatment by time interaction, ii) like in (i) but without treatment by time interaction and iii) like in (i) but without treatment. The AIC criterion was used to select the best fitting model. Our final model had the form:

$$\text{logit}\{P(y_{ij} = 1|b_i)\} = \beta_0 + \beta_1 z_i + \beta_2 t_j + b_i, \quad (2)$$

where  $z_i = 1$  (0) denotes the treatment (control) group,  $t_j$  denotes the occasion of measurement and  $b_i$  denotes a random intercept. The maximum likelihood estimates of the fixed effects are given in Table 1, and the variance of the random effects was estimated as 21.01 (S.E. = 6.81).

Figure 1 displays the plot of the fitted values obtained from our final model against the observed probability of being classified as normal to mildly ill ( $P(Y = 1)$ ) by time point and treatment group. The fitted probabilities are calculated by numerically integrating out the random effect for each subject. Until week 4 there seems to be a reasonable agreement between the fitted and the observed values. Nevertheless, some discrepancy is observed in the last two measurement occasions. It is important to point out that the proportion of dropouts is significantly high for these two measurements, specially in week 8 for the control group (50%). In presence of missing data such a discrepancy is not necessarily an evidence of lack of fit (Molenberghs and Verbeke, 2005). However, we do not want to enter here in a full discussion of the missing data problem so we will assume that the missing data generating mechanism is MAR making our likelihood approach a valid option.

Note that, even though the model given by (2) emerged as the best fitting model among all the ones considered in the model building exercise, it produces relatively extreme estimates for the intercept and the variance component. We believe this is the result of some extreme response pattern in the data. For example, in the control group a high proportion of the patients (75%) have a response pattern of nothing but zeros whereas in the treatment group a more variable pattern of responses is observed. There, only 56% of the patients have a response pattern of all zeros.

Therefore, the large estimate for the variance of the random component could be explained by the high inter-subject correlation that these data seem to suggest. Arguably, these circumstances could render the assumption of a normal distribution for the random effects

questionable. However, this situation should not be considered exceptional or infrequent. Indeed, in a typical placebo controlled clinical trial such an extreme pattern of all zeros could be expected in the placebo control group, whereas a more variable pattern should be expected in the responses of the treated group.

The problem is aggravated by the fact that random effects are unobserved latent variables and it can be difficult to evaluate their distributional assumptions. Diagnostic tools to analyze the random-effects distribution are not straightforward, for instance, one should be careful in using empirical Bayes estimates of the random effects to detect departures from normality. Indeed, it can be shown that in GLMM, the empirical Bayes estimates no longer follow a normal distribution, even when the random-effects distribution is correctly specified as normal. Nevertheless, the conventional wisdom among data analysts seems to be that the choice of the random-effects distribution is not crucial to the quality of the inference about regression coefficients, even though, as can be seen from Agresti et al. (2004), this does not always hold. Therefore, the study of the impact of this misspecification on our inferences is of utmost importance. Are the maximum likelihood estimators still consistent? How are the power and the type I error of the tests for the mean structure parameters affected by possible misspecification of the random-effects distribution? The following sections try to shed light on these topics.

### **3 The Effect of Misspecification on the Maximum Likelihood Estimators**

Let us consider a random variable  $\mathbf{y}$  with density function  $g$ , and a parametric family of density functions  $f = \{f(\mathbf{y}; \boldsymbol{\xi}) : \boldsymbol{\xi} \in \Gamma\}$ . If there exists a  $\boldsymbol{\xi}_0 \in \Gamma$  such that  $g(\mathbf{y}) = f(\mathbf{y}, \boldsymbol{\xi}_0)$  then the maximum likelihood estimator  $\hat{\boldsymbol{\xi}}_n$  of  $\boldsymbol{\xi}_0$  is consistent and asymptotically normal. However, since in practice  $g$  is unknown, it can be difficult to check whether  $f$  contains  $g$  or

not. White (1982) found that under general regularity conditions, the maximum likelihood estimator  $\hat{\boldsymbol{\xi}}_n$  will (strongly) converge to the value of  $\boldsymbol{\xi}$ , denoted by  $\boldsymbol{\xi}^*$ , which minimizes the Kullback-Leibler Information Criterion (KLIC)

$$I(g : f, \boldsymbol{\xi}) = E \left\{ \log \frac{g(\mathbf{y})}{f(\mathbf{y}, \boldsymbol{\xi})} \right\}, \quad (3)$$

where the expectation is taken with respect to the true distribution.

Note that if the model for  $\mathbf{y}$  is correctly specified then the information criterion attains its unique minimum at  $\boldsymbol{\xi}^* = \boldsymbol{\xi}_0$ . In this case,  $\hat{\boldsymbol{\xi}}_n$  is a consistent estimator for  $\boldsymbol{\xi}_0$ . Nevertheless, the maximum likelihood estimator can fail to be consistent due to misspecified distributional assumptions. Additionally, Kent (1982) studied the asymptotic distribution of the likelihood ratio statistic when the data generating mechanism differs from the assumed parametric model. In general, the likelihood ratio statistic no longer follows an asymptotic chi-squared distribution and an alternative approach is needed.

In the context of GLMM, Neuhaus et al. (1992) examined the performance of a random-intercept logistic regression model with misspecified random-effect distributions. They showed that the maximum likelihood estimators of the model parameters are inconsistent but that the magnitude of the bias is typically small. Simulations by Chen, Zhang and Davidian (2002) indicate that the estimation of the regression coefficients may be subject to only negligible bias under misspecification of the random-effects distribution. 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. They found that the marginal regression parameters are much less sensitive to misspecification than the parameters of the corresponding hierarchical model. According to Agresti et al. (2004) the choice of the random-effects distribution seems to have, in most situations, little effect on the maximum likelihood estimators. However, when there is a severe polarization of subjects, e.g., by omitting an influential binary covariate,



this can affect the quality of prediction of characteristics involving the random effects as well as the fixed effects. Litière et al. (2006) found that, in some circumstances, misspecification of the random-effects distribution can introduce substantial relative bias, especially in the estimation of the variance-covariance parameters. However, the mean structure parameters can also be affected.

This literature overview illustrates the wide variety of opinions about the impact of the random-effects misspecification on GLMM. However, up to now, not much research has been carried out to directly evaluate the performance, under misspecification, of the most frequently used statistical tests in GLMM. In the following section we will investigate this issue in more detail.

## 4 The Effect of Random-Effects Misspecification on Statistical Inference

In many situations, data analysts consider test statistics and corresponding  $p$ -values to evaluate, e.g., whether or not a drug has a significant influence. Therefore, the impact of misspecifying the random-effects distribution on the type I and the type II error is very important from a practical point of view. To explore this effect, we designed the following simulation study. A binary response was generated using the logistic random-intercept model given by (2). This model includes a treatment indicator  $z$  (0 or 1) and a within-cluster time variable  $t_j$ , with values 0, 1, 2, 4, 6 and 8. For the mean structure we chose  $\beta_0 = -8$  and  $\beta_2 = 1$ , whereas five different values for the treatment effect  $\beta_1$  were considered: 0, 0.5, 1, 2 and 5.

Further, 4 different random-effects distributions, each with variances  $\sigma_b^2 = 1, 4, 16,$  and  $32$  were included in the study. These distributions were a mean zero normal density, a power function distribution, a discrete distribution with equal probability at two support points

and an asymmetric mixture of two normal densities. If necessary, they were transformed to satisfy the mean zero condition of the random effects. Note that on the one hand, the variances  $\sigma_b^2 = 16$  and  $32$  of the random effects will help us to investigate relatively more extreme scenarios with variances in the same order of magnitude as the one observed in the case study. On the other hand, the smaller values considered for  $\sigma_b^2$  should allow us to study the performance of the tests in less extreme settings. In this way we expect to cover in our study a large range of practically relevant situations.

The simulations were performed for three different sample sizes, namely 25, 100, and 400. For each setting, 500 data sets were generated and the model given by (2) was fitted to these data under the assumption of normally distributed random effects. Further, we determined the proportion of cases in which a treatment effect different from zero (at a 5% significance level) was detected. When  $\beta_1 = 0$ , this proportion corresponds to the type I error; otherwise, it represents the power of the test. The results of these simulations are displayed in Figure 2.

It is clear from these graphs that misspecification can severely affect the power of the analysis, depending on the shape and the variance of the real random-effects distribution. Actually, the power can be seriously affected even in settings where the random intercept accounts for a small variability. For example, let us consider in Figure 2 the graphs corresponding to a sample of 100 patients, when  $\beta_1 = 1$ . Even with  $\sigma_b^2 = 1$ , the power to detect a significant treatment effect can drop as low as 20% for the power function distribution, whereas for the correctly specified model, we observed a value around 70%. This makes it very difficult to interpret negative results, i.e., it would be difficult to determine if we fail to reject the null hypothesis because there is no real treatment effect or because of a lack of power due to misspecification.

Interestingly, the type I error rate rarely exceeded the specified 5% level of significance in all the scenarios displayed in Figure 2. This finding confirms the results of Neuhaus et al.

(1992). Indeed, these authors showed for a similar logistic random-intercept model that when  $\beta_1 = 0$ , the corresponding maximum likelihood estimator consistently estimates zero. It is possible to prove that in this situation a significant treatment effect could be considered as a reliable result even though caution may be needed in the interpretation of the point estimates. The question remains whether this is true only for this specific model or whether the type I error also remains unaffected in more general situations. In the following theorem we try to answer this question.

**Theorem 1** *Let  $y_{ij}$  denote the  $j$ th measurement for the  $i$ th subject, with  $i = 1, \dots, n$  and  $j = 1, \dots, r$ . Conditional on a vector  $\mathbf{b}_i$  of individual random effects for subject  $i$ , it is assumed that all responses  $y_{ij}$  are independent with density*

$$f(y_{ij}|\theta_{ij}, \phi) = \exp[\phi^{-1}\{y_{ij}\theta_{ij} - \psi(\theta_{ij})\} + c(y_{ij}, \phi)], \quad (4)$$

where  $\theta_{ij}$  is modeled as

$$\theta_{ij} = \eta(\beta_0 + \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i), \quad (5)$$

and  $\eta(\cdot)$  denotes a known link function,  $\beta_0$  is an intercept,  $\mathbf{x}_{ij} = (\mathbf{x}_{ij}^F, \mathbf{x}_{ij}^R)$  denotes a  $p$ -dimensional vector of covariates with  $\mathbf{x}_{ij}^F \cap \mathbf{x}_{ij}^R = \emptyset$ ,  $\mathbf{z}_{ij} = \mathbf{x}_{ij}^R$  is a  $q$ -dimensional vector,  $\boldsymbol{\beta} = (\boldsymbol{\beta}^F, \boldsymbol{\beta}^R)$  is a vector of fixed parameters and  $\mathbf{b}_i$  is a vector of random effects assumed to follow a distribution  $F$  with  $E(\mathbf{b}_i) = \mathbf{0}$ . Without loss of generality, the covariates are assumed to be centered around zero, i.e.  $E(\mathbf{x}_{ij}) = \mathbf{0}$ .

If  $G$  represents the true random-effects distribution ( $G \neq F$ ) and if for any subset  $\mathbf{x}_{S_{ij}}^F$  of  $\mathbf{x}_{ij}^F$ ,  $\boldsymbol{\beta}_S^F = \mathbf{0}$ , then under the assumptions A1-A3 given by White (1982),  $\boldsymbol{\beta}_S^{F*}$ , which minimizes the KLIC, is also zero. Therefore, the maximum likelihood estimator  $\hat{\boldsymbol{\beta}}_{S_n}^F$ , based on a model with a misspecified random-effects distribution, satisfies

$$\hat{\boldsymbol{\beta}}_{S_n}^F \xrightarrow{n} \mathbf{0}. \quad (6)$$

A general idea of the proof can be seen in Appendix 1 whereas a full detailed proof can be found in Web Appendix A. This theorem implies that if the parameters associated with a subset of variables, which are not included in the random-effect structure, equal zero then the corresponding maximum likelihood estimator will consistently estimate zero. The main implication of this theorem is stated in the following corollary.

**Corollary 1** *Consider the hypothesis testing problem*

$$H_0 : \boldsymbol{\beta}_S^F = 0 \quad \text{vs} \quad H_1 : \boldsymbol{\beta}_S^F \neq 0, \quad (7)$$

and the corresponding Wald test statistic  $W = (\hat{\boldsymbol{\beta}}_{S_n}^F)^T V_n^{-1}(\hat{\boldsymbol{\beta}}_{S_n}^F) (\hat{\boldsymbol{\beta}}_{S_n}^F)$ , where  $V_n(\hat{\boldsymbol{\beta}}_{S_n}^F)$  is the so-called sandwich estimator of the asymptotic covariance matrix corresponding to  $\hat{\boldsymbol{\beta}}_S^F$  (White, 1982). Let  $p_0$  denote the dimension of  $\boldsymbol{\beta}_S^F$ . Then, given the assumptions stated in Theorem 1, the type I error rate associated with the critical region  $W > \chi_{p_0, 1-\frac{\alpha}{2}}^2$  is preserved even under misspecification of the random-effects distribution, i.e.,

$$P(W > \chi_{p_0, 1-\frac{\alpha}{2}}^2 \mid \boldsymbol{\beta}_S^F = 0) \leq \alpha \quad (8)$$

The proof of this result can be found in Web Appendix B. This corollary implies that the type I error will be maintained, even under a misspecified random-effects distribution, provided that the corresponding subset of covariates is not included in the random-effects structure. Both theorems can play a relevant role in studies where randomization is used. For example, in a clinical trial, where patients are randomized, the treatment variable will usually not be included in the random-effect structure and therefore, if a significant treatment effect is observed, one could generally be confident about this result.

Notice that the two previous theorems are valid asymptotically. However, the practical scope and reliability of an asymptotic result will heavily depend on its performance for small sample sizes. To study this further, we have displayed in Table 2 the results for the type I

error, obtained from the previous simulation study. Recall that in this study we evaluated the performance of the Wald test associated with the treatment effect parameter  $\beta_1$ . The treatment variable was not included in the random effect structure and therefore in this case, according to Corollary 1, the type I error should be maintained. Note that so far, only the inverse of the Fisher information matrix had been used in all our analyses and simulations to obtain standard errors for the estimates of the model parameters. However, now we used both uncorrected and corrected standard errors, obtained using the sandwich estimator, to calculate the test statistic. Clearly, the type I error was maintained in almost all the settings even with relatively small samples sizes.

In a second simulation study the binary response variable was generated using the model given by (2), with  $\beta_0 = 0$  (and  $\beta_1 = 2$ ,  $\beta_2 = 1$ ). The results of this study can be seen in Table 3. Here, we study the performance of the Wald test associated with the intercept parameter  $\beta_0$ . Unlike the treatment variable the intercept was included in the random-effect structure. The results clearly illustrate that the Type I error is severely affected by the misspecification of the random-effects distribution. Even when the variance of the random intercept is relatively small, e.g., when  $\sigma_b^2 = 1$ , the type I error rate can be dramatically inflated, up to 16% in some scenarios.

Moreover, in the second study the sandwich correction for the variance-covariance matrix of the parameter estimates did not always lead to better results. Actually, in some settings the use of the correction seems to inflate even further the type I error. Additional simulations (not shown) using more complicated models also seem to confirm this erratic behavior of the sandwich estimator. Other problems with the standard errors have been recently described in the literature in the generalized estimating equation framework (Pan and Wall, 2002; Mancl and deRouen, 2001) and therefore some care should be taken when this correction is used. However, a full discussion and study of this issue supersedes the scope and objectives

of the present work and may require additional research.

## 5 Discussion

Generalized linear mixed models have become a powerful parametric tool in the analysis of non-Gaussian longitudinal data. Importantly, recent research is showing that problems can arise due to certain model misspecifications. In the present paper we have focused on the impact of misspecifying the random-effects distribution on the inferences in GLMM. The simulation results presented in Section 4 clearly show that the power can be seriously affected, depending on the shape and the variance of the underlying random-effects distribution. However, our simulations also showed that the type I error rate seems to be maintained under the 5% level in all the considered scenarios. These findings have led us to a theoretical result showing that the type I error associated with a test for a covariate's effect will not be asymptotically affected, as far as this variable is not included in the random-effects structure. It is important to note that all these results are conditional on the correct specification of any other aspect of the model, like the link function, the mean structure and so on.

The results reported in this work clearly show the need for alternative approaches to analyze non-Gaussian longitudinal data. A plausible choice would be to replace the normal random-effects distribution by a nonparametric distribution (Butler and Louis, 1992; Aitkin, 1999), a semi-parametric distribution (Chen et al., 2002), or a finite mixture of normals (i.e., a heterogeneity model; Fieuws, Spiessens, and Draney, 2004). However, Agresti et al. (2004) reported that there can be some loss of efficiency, when using a nonparametric approach, compared to a parametric assumption close to the real distribution. Additionally, model comparison can be difficult in this setting as standard asymptotic theory does not apply. On the other hand, modelling the random-effects distribution through a finite mixture of Gaussian densities can be difficult from a computational point of view and especially sensitive

to the choice of the initial values. As a general guideline we suggest to consider different random-effects distributions as a way of sensitivity analysis to evaluate the robustness of our conclusions. Obviously a lot of research is still necessary in this area, however, the results obtained in the present work can allow us to feel relatively confident about the presence of a treatment effect of risperidone on the CGI scores of schizophrenic patients. Clearly, a valuable clinical finding.

## Supplementary Materials

The Web Appendices referenced in Section 4 are available under the Paper Information link at the Biometrics website <http://www.tibs.org/biometrics>.

## Acknowledgment

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 edition. 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**, 639-653.
- Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics* **55**, 117-128.

- Alonso, A., Geys, H., Molenberghs, G., Kenward, M.G., and Vangeneugden, T. (2004). Validation of surrogate markers in multiple randomized clinical trials with repeated measurements: canonical correlation approach. *Biometrics* **60**, 845-853.
- 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 *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*, P. De Boeck and M. Wilson (eds), 317-340. Statistics for Social Science and Public Policy, New York: Springer-Verlag.
- Heagerty, P.J. and Kurland, B.F. (2001). misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika* **88**, 973-985.
- 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.
- Kent, J. T. (1982). Robust properties of likelihood ratio tests. *Biometrika* **69**, 19-27.
- Litière, S., Alonso, A., Molenberghs, G., and Geys, H. (2006). The impact of a misspecified random-effects distribution on maximum likelihood estimation in generalized linear



mixed models. *Manuscript submitted for publication.*

Mancl, L.A. and deRouen, T.A. (2001). A covariance estimator for GEE with improved small sample properties. *Biometrics* **57**, 126-134.

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

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.

Pan, W. and Wall, M.M. (2002). Small-sample adjustments in using the sandwich variance estimator in generalized estimating equations. *Statistics in Medicine* **21**, 1429-41.

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

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1-25.

## Appendix 1

The general idea of the proof of Theorem 1 is as follows. For simplicity of notation we will work out the proof for  $\mathbf{x}_{S_{ij}}^F = \mathbf{x}_{ij}^F$ . The proof for any other subset  $\mathbf{x}_{S_{ij}}^F$  can be obtained in a similar way.

First note that there always exists a lower triangular matrix  $\mathbf{U}$ , so that  $\mathbf{b}_i = \mathbf{U}\mathbf{a}_i$  with  $E(\mathbf{a}_i) = \mathbf{0}$  and  $V(\mathbf{a}_i) = \mathbf{I}$ . This allows to write (5) as  $\theta_{ij} = \eta(\beta_0 + \mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{U}\mathbf{a}_i)$ . Let us further denote by  $G$  and  $F$  the true and the assumed distribution of the random effects. According to White (1982), the maximum likelihood estimator of  $\boldsymbol{\xi} = (\beta_0, \boldsymbol{\beta}, \mathbf{U})$  converges to the unique value  $\boldsymbol{\xi}^* = (\beta_0^*, \boldsymbol{\beta}^*, \mathbf{U}^*)$  which minimizes the KLIC (3), i.e.  $\boldsymbol{\xi}^*$  minimizes

$$I(G : F, \boldsymbol{\xi}) = E_{\mathbf{x}} E_{\mathbf{y}|\mathbf{x}} \log \left\{ \frac{f_G(\mathbf{y}|\boldsymbol{\xi}, \mathbf{x}, \mathbf{z})}{f_F(\mathbf{y}|\boldsymbol{\xi}^*, \mathbf{x}, \mathbf{z})} \right\}, \quad (9)$$

where the expectation is taken with respect to the true model. In the previous expression,

$$\begin{aligned} f_G(\mathbf{y}|\boldsymbol{\xi}, \mathbf{x}, \mathbf{z}) &= \int \prod_j \exp[\phi^{-1}\{y_j\theta_j - \psi(\theta_j)\} + c(y_j, \phi)] dG(\mathbf{a}), \\ f_F(\mathbf{y}|\boldsymbol{\xi}^*, \mathbf{x}, \mathbf{z}) &= \int \prod_j \exp[\phi^{-1}\{y_j\theta_j^* - \psi(\theta_j^*)\} + c(y_j, \phi)] dF(\mathbf{a}), \end{aligned}$$

with

$$\begin{aligned} \theta_j &= \eta(\beta_0 + \mathbf{x}_j^T\boldsymbol{\beta} + \mathbf{z}_j^T\mathbf{U}\mathbf{a}), \\ \theta_j^* &= \eta(\beta_0^* + \mathbf{x}_j^T\boldsymbol{\beta}^* + \mathbf{z}_j^T\mathbf{U}^*\mathbf{a}). \end{aligned}$$

For simplicity of notation, the index  $i$  has been omitted from the previous equations. To find  $\boldsymbol{\xi}^*$  we have to differentiate (9) with respect to  $\beta_0^*$ ,  $\boldsymbol{\beta}^*$  and  $\mathbf{U}^*$ . This leads to the following system of simultaneous equations

$$E_{\mathbf{x}} \left[ \int \varphi(\mathbf{y}|\mathbf{x}, \mathbf{z}) \left\{ \int \sum_i h_i(\mathbf{x}, \mathbf{a}) dF(\mathbf{a}) \right\} dy \right] = 0, \quad (10)$$

$$E_{\mathbf{x}} \left[ \int \varphi(\mathbf{y}|\mathbf{x}, \mathbf{z}) \left\{ \int \sum_i (\mathbf{x}_i^F)^T h_i(\mathbf{x}, \mathbf{a}) dF(\mathbf{a}) \right\} dy \right] = 0, \quad (11)$$

$$E_{\mathbf{x}} \left[ \int \varphi(\mathbf{y}|\mathbf{x}, \mathbf{z}) \left\{ \int \sum_i (\mathbf{x}_i^R)^T h_i(\mathbf{x}, \mathbf{a}) dF(\mathbf{a}) \right\} dy \right] = 0, \quad (12)$$

$$E_{\mathbf{x}} \left[ \int \varphi(\mathbf{y}|\mathbf{x}, \mathbf{z}) \left\{ \int \sum_i \mathbf{M}_i h_i(\mathbf{x}, \mathbf{a}) dF(\mathbf{a}) \right\} dy \right] = 0, \quad (13)$$

where  $\varphi(\mathbf{y}|\mathbf{x}, \mathbf{z}) = f_G(\mathbf{y}|\boldsymbol{\xi}, \mathbf{x}, \mathbf{z})/f_F(\mathbf{y}|\boldsymbol{\xi}^*, \mathbf{x}, \mathbf{z})$  and

$$h_i(\mathbf{x}, \mathbf{a}) = \eta'(\boldsymbol{\xi}^*) \phi^{-1}\{y_i - \psi'(\theta_i^*)\} \prod_j (\exp[\phi^{-1}\{y_j\theta_j^* - \psi(\theta_j^*)\} + c(y_j, \phi)]).$$

If  $\boldsymbol{\beta}^F = \boldsymbol{\beta}^{F*} = \mathbf{0}$  then  $\varphi(\mathbf{y}|\mathbf{x}, \mathbf{z})$ ,  $\eta(\beta_0^* + \mathbf{x}_i^T \boldsymbol{\beta}^* + \mathbf{z}_i^T \mathbf{U}^* \mathbf{a})$ ,  $\phi(\theta_i^*)$ ,  $\phi'(\theta_j^*)$  and  $h_i(\mathbf{x}, \mathbf{a})$  are functions which are independent of  $\mathbf{x}_i^F$ . Therefore, the left-hand side of (11) is zero for all  $\beta_0$ ,  $\boldsymbol{\beta}^R$ ,  $\mathbf{U}$ ,  $\beta_0^*$ ,  $\boldsymbol{\beta}^{R*}$  and  $\mathbf{U}^*$ . Note now that equations (10), (12) and (13) determine  $\beta_0^*$ ,  $\boldsymbol{\beta}^{R*}$  and  $\mathbf{U}^*$  in terms of  $\beta_0$ ,  $\boldsymbol{\beta}^R$  and  $\mathbf{U}$ . Thus, when  $\boldsymbol{\beta}^F = \boldsymbol{\beta}^{F*} = \mathbf{0}$ , we have found the unique solution for  $\boldsymbol{\xi}^*$ . Therefore, when  $\boldsymbol{\beta}^F = \mathbf{0}$ , the maximum likelihood estimate  $\hat{\boldsymbol{\beta}}_n^F$  consistently estimates zero.

Table 1: Parameter estimates and standard errors of the fixed effects in the logistic random-intercept model given by (2) for the schizophrenia data.

Parameter	Estimate	S.E.	<i>p</i> -value
$\beta_0$	-7.37	1.18	< 0.0001
$\beta_1$	2.14	1.08	0.049
$\beta_2$	0.65	0.10	< 0.0001

Table 2: Type I error for detecting a significant treatment effect when  $\beta_1 = 0$  in the logistic random-intercept model given by (2), using both uncorrected and corrected (between parenthesis) standard errors. Values for which the lower bound of the corresponding 95% confidence interval was larger than 0.05 are highlighted

Distribution	Sample size	$\sigma_b^2 = 1$		$\sigma_b^2 = 4$		$\sigma_b^2 = 16$		$\sigma_b^2 = 32$	
Normal	$n = 25$	0.012	(0.015)	0.025	(0.029)	0.029	(0.047)	0.025	(0.041)
	$n = 100$	0.041	(0.036)	0.052	(0.052)	0.050	(0.050)	0.026	(0.028)
	$n = 400$	0.050	(0.052)	0.046	(0.044)	0.052	(0.054)	0.058	(0.058)
Power function	$n = 25$	0.008	(0.031)	0.023	(0.030)	0.036	(0.056)	0.016	(0.028)
	$n = 100$	0.041	(0.038)	0.040	(0.038)	0.050	(0.050)	0.028	(0.026)
	$n = 400$	0.046	(0.046)	0.064	(0.060)	<b>0.076</b>	<b>(0.074)</b>	0.050	(0.050)
Discrete	$n = 25$	0.023	(0.051)	0.012	(0.032)	0.014	(0.026)	0.004	(0.022)
	$n = 100$	0.032	(0.034)	0.016	(0.018)	<b>0.084</b>	<b>(0.100)</b>	0.018	(0.018)
	$n = 400$	0.048	(0.050)	<b>0.080</b>	<b>(0.078)</b>	0.024	(0.024)	<b>0.088</b>	<b>(0.088)</b>
Asymmetric mixture	$n = 25$	0.014	(0.027)	0.014	(0.032)	0.018	(0.034)	0.038	(0.040)
	$n = 100$	0.053	(0.048)	0.066	(0.064)	0.036	(0.032)	0.038	(0.036)
	$n = 400$	0.053	(0.053)	0.057	(0.057)	0.036	(0.036)	0.032	(0.032)

Table 3: Type I error for detecting a significant intercept when  $\beta_0 = 0$  in the logistic random-intercept model given by (2), using both uncorrected and corrected (between parenthesis) standard errors. Values for which the lower bound of the corresponding 95% confidence interval was larger than 0.05 are highlighted

Distribution	Sample size	$\sigma_b^2 = 1$		$\sigma_b^2 = 4$		$\sigma_b^2 = 16$		$\sigma_b^2 = 32$	
Normal	$n = 25$	0.014	(0.020)	0.035	(0.043)	0.016	(0.022)	0.023	(0.021)
	$n = 100$	0.042	(0.044)	0.048	(0.052)	0.040	(0.038)	0.034	(0.036)
	$n = 400$	0.060	(0.060)	0.046	(0.046)	0.054	(0.054)	0.050	(0.050)
Power function	$n = 25$	0.019	(0.024)	0.031	(0.046)	0.028	(0.049)	0.022	(0.065)
	$n = 100$	0.043	(0.045)	<b>0.164</b>	<b>(0.174)</b>	<b>0.320</b>	<b>(0.308)</b>	<b>0.370</b>	<b>(0.366)</b>
	$n = 400$	<b>0.158</b>	<b>(0.174)</b>	<b>0.682</b>	<b>(0.708)</b>	<b>0.946</b>	<b>(0.952)</b>	<b>0.962</b>	<b>(0.966)</b>
Discrete	$n = 25$	0.021	(0.029)	0.046	(0.054)	<b>0.087</b>	(0.064)	0.073	<b>(0.082)</b>
	$n = 100$	0.040	(0.040)	0.060	(0.052)	<b>0.136</b>	<b>(0.098)</b>	<b>0.156</b>	<b>(0.098)</b>
	$n = 400$	<b>0.080</b>	<b>(0.074)</b>	<b>0.252</b>	<b>(0.212)</b>	<b>0.594</b>	<b>(0.494)</b>	<b>0.604</b>	<b>(0.510)</b>
Asymmetric mixture	$n = 25$	0.015	(0.025)	0.025	(0.042)	0.011	(0.071)	0.045	<b>(0.249)</b>
	$n = 100$	0.030	(0.028)	<b>0.328</b>	<b>(0.354)</b>	<b>0.408</b>	<b>(0.428)</b>	<b>0.886</b>	<b>(0.892)</b>
	$n = 400$	<b>0.076</b>	<b>(0.080)</b>	<b>0.924</b>	<b>(0.938)</b>	<b>0.986</b>	<b>(0.992)</b>	<b>1.000</b>	<b>(0.998)</b>

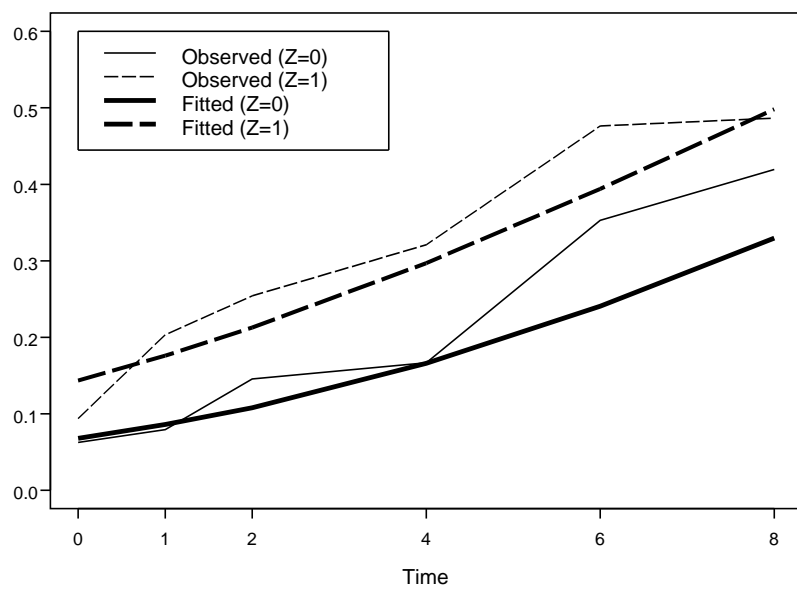


Figure 1: Evolution of the observed and fitted (using Model (2)) probabilities to be classified as a normal to mildly ill patient by treatment group. Here  $Z = 1$  (0) denotes the treatment (control) group.

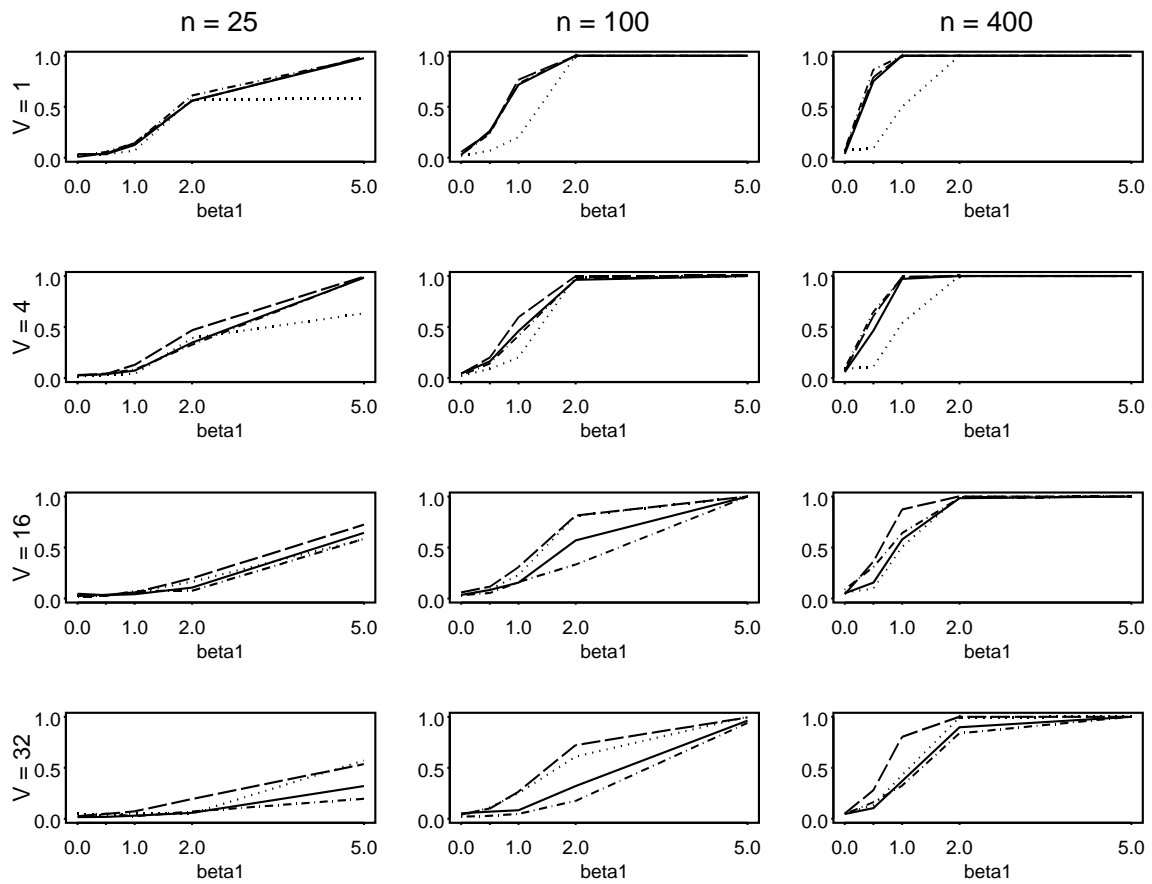


Figure 2: Power of the analysis of the logistic random-effects model given by (2) to detect a significant treatment effect over a range of possible  $\beta_1$  values, for the 4 considered random-effects distributions: normal (solid line), power function (dotted line), discrete (dash-dotted line) and asymmetric mixture (dashed line).