Published in final edited form as:

Biometrics. 2011 June ; 67(2): 654–660. doi:10.1111/j.1541-0420.2010.01474.x.

A note on Type II error under random effects misspecification in generalized linear mixed models

John M. Neuhaus*, Charles E. McCulloch, and Ross Boylan

Department of Epidemiology and Biostatistics University of California San Francisco, California 94143-0560, U.S.A.

Summary

Litière et al. (2007) presented the results of simulation studies that they claimed showed that misspecification of the shape of the random effects distribution can produce marked increases in Type II error (decreases in power) of tests based on fits of generalized linear mixed models. However, the paper contains a logical fallacy that invalidates this claim. We present logically correct simulation studies that demonstrate little increase in Type II error, consistent with the earlier work that shows little effect due to misspecification.

Keywords

Conditional likelihood; generalized linear mixed models; misspecified random effect distributions; power

1. Introduction

Generalized linear mixed models are useful methods for analyzing clustered and longitudinal data. Fitting generalized linear mixed models in practice typically requires parametric specification of the distribution of the random effects and there exists little information to guide this choice. While it is possible to fit models with non-parametric random effects distributions (Lesperance and Kalbfleisch, 1992), this requires large sample sizes. Furthermore, there is an extensive literature on the effects of misspecifying the shape of the random effects distribution when fitting random intercepts models and most of this work shows that misspecification produces very little bias in estimated regression coefficients (e.g., Neuhaus et al., 1992, 1994; Butler and Louis, 1992; Verbeke and Lesaffre, 1997; Heagerty and Kurland, 2001).

However, Litière et al. (2007) recently presented the results of simulation studies that they claimed showed that random effects misspecification can produce marked increases in Type II error (decreases in power). This note argues that Litière et al. (2007) contains a logical fallacy that invalidates this claim. We present logically correct simulation studies that demonstrate little increase in Type II error, consistent with the earlier work that shows little bias in estimates of covariate effects due to misspecification. In addition, the strongest evidence to support the conclusions of Litière et al. (2007) comes from simulations that we were unable to replicate, despite very careful programming.

* john@biostat.ucsf.edu.

Supplementary Materials

The Web Appendix referenced in Section 2.1 is available under the Paper Information link at the Biometrics website <http://www.tibs.org/biometrics>.

To demonstrate effects of misspecification one needs to vary the assumed, fitted distribution while holding constant the true distribution. Instead, Litière et al. (2007) base their claims on fitting a normal distribution but varying the true distribution. Differences observed when fitting the same assumed distribution to multiple true distributions may merely reflect the different efficiency that is achievable with different distributions, not the effects of misspecification. For example, in the simulations described in Section 2.1 for the case of $n = 100$ clusters, random effects variance, $\sigma_b^2 = 32$ and regression coefficient, $\beta_1 = 2$, mixed-effects models that correctly specified the random effects to follow a power function distribution rejected $H_0 : \beta_1 = 0$ in 91% of data sets while mixed-effects models that correctly specified the random effects to follow a normal distribution rejected $H_0 : \beta_1 = 0$ only 32% of the time. In the next section we describe the simulation design used by Litière et al. (2007), as well as their primary results, and the simulation design that we used.

2. Simulation studies

Litière et al. (2007) conducted a set of simulations where they generated clustered binary responses from mixed-effects logistic models with several different random effects distributions, including non-normal distributions such as the Power (Johnson et al., 1994). To these data they fit mixed-effects logistic models that assumed that the random effects followed a normal distribution. Specifically, letting y_{ij} denote the binary response of the j^{th} observation in the i^{th} cluster, $i = 1, \dots, m$, $j = 1, \dots, n$, z_i denote a binary cluster-level covariate, t_j denote the common, within-cluster occasion of measurement and b_i denote the random intercept of the i^{th} cluster, Litière et al. (2007) generated correlated binary responses from a mixed-effects logistic model of form

$$\text{logit} \{ \text{pr}(Y_{ij}=1 \mid b_i, z_i, t_j) \} = \beta_0 + b_i + \beta_1 z_i + \beta_2 t_j, \quad (1)$$

where $b_i \sim G$. The within-cluster covariate, t_j , assumed the fixed values 0, 1, 2, 4, 6 and 8 in each cluster. Litière et al. (2007) used several generating distributions G but we will focus on two: $G_1 = N(0, \sigma_b^2)$, and $G_2 = \text{Power}(\gamma, \theta)$ since these two distributions produced the biggest discrepancies in results. Litière et al. (2007) used a Power Function density of the form

$$g_2(b) = \frac{\gamma b^{\gamma-1}}{\theta^\gamma}, \quad (2)$$

with $\gamma = 80$ and θ equal to values that produced desired values of $\text{var}(b) = \sigma_b^2$ for the simulations. Litière et al. (2007) also shifted the distribution of the random effects so that it had mean 0. Figure 1 plots such a normalized Power density along with a standard normal distribution. It demonstrates that the two distributions are very different and that the Power density has an asymptote.

The simulations used true values $\beta_0 = -8$, $\beta_2 = 1$, five values for $\beta_1 : 0, 0.5, 1, 2, 5$ and four values for $\sigma_b^2 : 1, 4, 16, 32$. We note that the σ_b^2 values of 16 and 32 represent a very high degree of clustering and far exceed the magnitudes of variance components that we have encountered in practice across a wide array of datasets.

Litière et al. (2007) used the SAS procedure NLMIXED to fit mixed-effects logistic models of form (1), assuming that $b \sim N(0, \sigma_b^2)$. The principal results of Litière et al. (2007) focus on comparisons of rates of rejection of the hypothesis $H_0 : \beta_1 = 0$ based on models that

correctly and incorrectly assume $bN(0, \sigma_b^2)$. Their Figure 2 presents plots of the rejection rates as a function of β_1 , σ_b^2 and the number of clusters n . While the majority of the scenarios and distributions indicate little difference between the rejection rates resulting from different underlying random effects distribution, there are some discrepancies between fits to true normal and true Power random effects. For example, for $\beta_1 = 1$, $\sigma_b^2 = 1$ and $n = 100$ the rejection rate exceeded 0.7 when the true random effects were normal but the rate was less than 0.25 when the true random effects followed the Power family. However, as we noted previously, a discrepancy between the fits to data sets from two different true distributions does not prove sensitivity to misspecification.

2.1 Simulation studies to assess misspecification

We carried out a set of simulation studies using the model in equation (1) with true normal or Power family random effects and fit two models to each generated data set: 1) one model that assumed $bN(0, \sigma_b^2)$ and 2) a second model that assumed $b \sim \text{Power}$. We fit all models using the NLMIXED procedure in SAS. We used the same true parameter values as Litière et al. (2007) and recorded for each model fit whether or not a standard Wald test rejected the hypothesis $H_0: \beta_1 = 0$. We fixed the parameter γ in (2) to its true value when fitting models involving the Power family random effects for comparability to the Normal distribution (which has no unknown shape parameters).

We first sought to replicate the results of Figure 2 in Litière et al. (2007). Our Figure 2 presents the rejection rates (power) from our simulation studies (solid black lines with circles) where the true random effects followed a Power family but we assumed normal random effects. We added rejection rates from Litière et al. (2007) (dotted lines with circles) to Figure 2. Although most of the two sets of rejection rates in Figure 2 closely correspond, there are several settings with large discrepancies, despite very careful programming on our part. In our simulation we identified problematic data sets and deleted non-convergent β_1 estimates from the analysis. Overall our simulations resulted in high convergence rates; 99.2% of the simulations yielded convergent β_1 estimates. The worst convergence rates occurred when fitting models assuming Power family random effects to data generated with Power family random effects, but the convergence rates were still very high; the lowest convergence rate was 98.4% and occurred in the setting with the smallest number of clusters, $n = 25$, and the largest random effects variance, $\sigma_b^2 = 32$. Our simulation code is available as an online supplement. Litière et al. (2007) did not report convergence rates. We note that the settings with the largest discrepancy between our results and those of Litière et al. (2007) are also the settings of largest discrepancies in Figure 2 of Litière et al. (2007).

Figure 2 also presents the rejection rates from simulation studies that fit models that assumed Power family random effects (dashed black lines with diamonds) to data with random effects generated according to a Power distribution. Consistent with previous work that indicates little effect of misspecification of the random effects distribution (e.g., Neuhaus et al., 1992, 1994), our simulation studies showed close correspondence between the rejection rates from the models that fit normal and Power family random effects. That is, incorrectly assuming that the random effects are normal results in nearly identical rejection rates as the model that correctly assumes Power family random effects.

The rejection rates from Litière et al. (2007) depart from ours in several parameter configurations. These discrepancies arise for most of the cases where the results for the normal and Power function distributions are most different within the results of Litière et al. (2007). For example, the results of Litière et al. (2007) suggest that the power of a test of the hypothesis $H_0: \beta_1 = 0$ changes little when the true value of β_1 changes from 2.0 to 5.0, a

highly unusual finding. Our results indicate that the power of a test of $H_0 : \beta_1 = 0$ based on assumed normal random effects increases with increasing β_1 , as one would expect.

3. Discussion

We disagree with the claims of Litière et al. (2007) – namely that random effects misspecification can produce marked increases in Type II error (decreases in power). First, one cannot establish effects of misspecification through fits of models based on a single random effects distribution to several different underlying true distributions. Second, we simulated fits of correct and incorrect models to the same data set and showed little increase in Type II error due to misspecification of the random effects distribution.

The settings simulated by us and Litière et al. (2007) represent severe misspecification. The Power family densities are quite far from Normal and the larger random effects variances were very extreme. Even so, the results for the between cluster covariate showed only minimal effects due to misspecification. Not surprisingly, the within-cluster covariate coefficient, β_2 , showed even less sensitivity to misspecification. To summarize, the results in this paper demonstrate that misspecification of the shape of the random effects distribution when fitting mixed-effects logistic models leads to little increase in Type II error, consistent with earlier work.

Acknowledgments

A grant from the U.S. National Institutes of Health supported this research. We thank Dr. Saskia Litière of Hasselt University for providing unpublished details of her simulation studies.

References

- Butler SM, Louis TA. Random effects models with non-parametric priors (Disc: p2017-2023). *Statistics in Medicine*. 1992; 11:1981–2000. [PubMed: 1480884]
- Heagerty PJ, Kurland BF. Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika*. 2001; 88:973–986.
- Johnson, NL.; Kotz, S.; Balakrishnan, N. Continuous univariate distributions. Wiley; New York: 1994.
- Lesperance ML, Kalbfleisch JD. An algorithm for computing the nonparametric MLE of a mixing distribution. *Journal of the American Statistical Association*. 1992; 87:120–126.
- Litière S, Alonso A, Molenberghs G. Type I and type II error under random effects misspecification in generalized linear mixed models. *Biometrics*. 2007; 63:1038–1044. [PubMed: 17425642]
- Neuhaus JM, Hauck WW, Kalbfleisch JD. The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika*. 1992; 79:755–762.
- Neuhaus JM, Kalbfleisch JD, Hauck WW. Conditions for consistent estimation in mixed-effects models for binary matched-pairs data. *The Canadian Journal of Statistics*. 1994; 22:139–148.
- Verbeke G, Lesaffre E. The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics and Data Analysis*. 1997; 23:541–556.

Figure 1.
Normal and Power Family densities

Figure 2.

Power under a true power function distribution. By $\text{Var}(b)=1, 4, 16$ and 32 (rows) and no. of clusters= $25, 100, 400$ (columns)