

Harvard University

Harvard University Biostatistics Working Paper Series

Year 2006

Paper 37

A Diagnostic Test for the Mixing Distribution in a Generalised Linear Mixed Model

Eric J. Tchetgen*

Brent A. Coull†

*Harvard University, etchetge@hsph.harvard.edu

†Harvard University, bcoull@hsph.harvard.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://biostats.bepress.com/harvardbiostat/paper37>

Copyright ©2006 by the authors.

A Diagnostic Test for the Mixing Distribution in a Generalised Linear Mixed Model

Eric J. Tchetgen and Brent A. Coull

Abstract

We introduce a diagnostic test for the mixing distribution in a generalised linear mixed model. The test is based on the difference between the marginal maximum likelihood and conditional maximum likelihood estimates of a subset of the fixed effects in the model. We derive the asymptotic variance of this difference, and propose a test statistic that has a limiting chi-square distribution under the null hypothesis that the mixing distribution is correctly specified. For the important special case of the logistic regression model with random intercepts, we evaluate via simulation the power of the test in finite samples under several alternative distributional forms for the mixing distribution. We illustrate the method by applying it to data from a clinical trial investigating the effects of hormonal contraceptives in women.

A Diagnostic Test for the Mixing Distribution in a Generalised Linear Mixed Model

Eric J. Tchetgen and Brent A. Coull
Department of Biostatistics
Harvard School of Public Health
655 Huntington Avenue
Boston, MA 02115

`etchetge@hsph.harvard.edu`, `bcoull@hsph.harvard.edu`

SUMMARY

We introduce a diagnostic test for the mixing distribution in a generalised linear mixed model. The test is based on the difference between the marginal maximum likelihood and conditional maximum likelihood estimates of a subset of the fixed effects in the model. We derive the asymptotic variance of this difference, and propose a test statistic that has a limiting chi-square distribution under the null hypothesis that the mixing distribution is correctly specified. For the important special case of the logistic regression model with random intercepts, we evaluate via simulation the power of the test in finite samples under several alternative distributional forms for the mixing distribution. We illustrate the method by applying it to data from a clinical trial investigating the effects of hormonal contraceptives in women.

Some key words: Clustered binary data; Conditional maximum likelihood; Marginal maximum likelihood; Specification test.



1 Introduction

Generalised linear mixed models (Breslow & Clayton 1993) have become a popular approach to modeling correlated discrete data. The models account for correlation among clustered observations by including random effects in the linear predictor component of the model. Although generalised linear mixed model fitting is typically complex, standard random intercept and random intercept/random slope models with normally distributed random effects can now be routinely implemented in such commercial software packages as SAS and Stata.

A natural question that arises with use of this class of models is the sensitivity of the estimated regression coefficients to distributional assumptions for the random effects. In the normal linear mixed model, incorrect choice of the covariance structure for correlated responses does not bias the coefficients, due to the fact that the fixed effects represent marginal effects of the covariates. In a nonlinear mixed model, however, coefficients represent within-subject effects. Zeger, Liang & Albert (1988) and Heagerty & Zeger (2000) noted that a within-subject effect is a function of both the marginal effect and the assumed distribution of the random effects. Thus, if this random effects distribution is misspecified, there is the potential for the estimates in a nonlinear mixed effects model to be biased.

Early work suggested that the effects of misspecifying the mixing distribution are small (Neuhaus, Hauck & Kalbfleisch 1992). However, more recent work suggests that violations of these assumptions can adversely affect inference. Agresti, Caffo & Ohman-Strickland (2002) showed that severe misspecification of this distribution, such as assuming normality when the true distribution is a two-point mixture, can arise in a considerable loss of efficiency. Moreover, Heagerty & Kurland (2001) showed that substantial bias in the regression coefficient estimates can result in simple random intercept models when either the variance of the random effects depends on a between-cluster covariate or when the random effects follow an autoregressive structure. A common strategy for guarding against such misspecification is to build more flexible distributional assumptions for the random effects into the model. For instance, Aitkin (1996, 1999) proposed estimating this distribution nonparametrically as a finite number of mass points and corresponding probabilities. Magder and Zeger (1996), Verbeke & Lesaffre (1996), and Chen, Zhang & Davidian (2002) constructed alternative nonparametric estimates based on mixtures of Gaussian distributions. These authors showed that such strategies can model a wide variety of shapes, including skewed

and multimodal forms, for the distribution of the random effects. Heagerty (1999) and Heagerty & Zeger (2000) proposed estimating effects in a marginally specified model, and Heagerty & Kurland (2001) showed that such an approach yields fixed effect estimates that are more robust to misspecification of the random effect distribution.

A disadvantage of these alternative strategies is that they are typically computationally intensive. That is, with the exception of the nonparametric approach of Aitkin (1996, 1999), for which there exist GLIM macros, we are not aware of any commercial software or macros that easily implement these more general models. Before investing time programming one of these more robust approaches, one would like to be able to diagnose whether such methods are required. Ideally, such diagnostics would be based on quantities from the fits of the simpler models that assume normality. Lange and Ryan (1989) proposed the weighted normal probability plot for assessing the normality of the random effects in the linear mixed model setting, and Houseman, Coull, & Ryan (2004) extended this approach to construct global hypothesis tests for this aspect of fit. Unfortunately, such methods are currently unavailable in the generalised case. Thus, there is a need for computationally simple, yet effective, diagnostics for model misspecification in generalised linear mixed models.

In this article, we propose a diagnostic test for the assumptions on the distribution of the random effects in a generalised linear mixed model. The test statistic is the suitably standardized difference between the estimates of a subset of the regression coefficients from the mixed model to those from a model that conditions out the random effects. When the distributional assumptions are correct, then asymptotically this difference should be zero, as both the mixed-effect and conditional estimators are consistent. When this distribution is misspecified, this difference has the potential to be large, since in this case the conditional estimator is consistent but the mixed model estimator is not. We derive the correct null and alternative variances for this difference, propose a test statistic for goodness-of-fit of a generalised linear mixed model, and investigate the performance of the test both in example data and via simulation. This strategy essentially uses an idea presented by Hausman (1978), who considered analogous tests for the linear mixed model. An important advantage of the methods outlined here is that the resulting diagnostic test is easily implemented in commercial software.

The remainder of this paper is as follows. In § 2 we outline the generalised linear mixed model and propose two test statistics for the diagnostic test. In § 3 we describe how one

can easily implement the method in commercial software. In § 4 we report the results of a simulation study, and in § 5 we illustrate the method by applying it to data from a clinical trial investigating the effects of hormonal contraceptives in women. In § 6 we include further discussion.

2 The Generalised Linear Mixed Model

Let Y_{ij} be observation j , $j = 1, \dots, n_i$ from cluster i , $i = 1, \dots, N$, and let X_i be a known $n_i \times p$ matrix of covariates constructed so that the j th row X_{ij} is the vector of covariates corresponding to Y_{ij} . The generalised linear mixed model specifies two components of the model: a model for the vector of outcome variables $Y_i = (Y_{i1}, \dots, Y_{in_i})'$ conditional on X_i and unobserved, cluster-specific random effects b_i , and distributional assumptions on b_i . Specifically, we assume Y_{ij} given X_i and b_i are independent random variables having a distribution in the natural exponential family,

$$f_{y|b}(y_{ij}|X_i, b_i) = \exp \left[\frac{w_{ij}}{\phi} \{y_{ij}\vartheta_{ij} - b(\vartheta_{ij})\} + c \left(y_{ij}, \frac{\phi}{w_{ij}} \right) \right], \quad (1)$$

where ϑ and ϕ are unknown parameters, b and c are known functions, and w_{ij} are known weights. Denote $\mu_{ij} = E(Y_{ij}|X_{ij}, b_i) = b'(\vartheta_{ij})$. A generalised linear mixed model for Y_{ij} is

$$g(\mu_{ij}) = X_{ij}\beta + Z_{ij}b_i, \quad (2)$$

where $g(x)$ is a monotonically increasing link function, β is a $p \times 1$ vector of unknown fixed parameters, and Z_{ij} is a known covariate matrix with columns typically a subset of those in X_{ij} . The results in this article apply to models using the canonical link function $g(\mu_{ij}) = \vartheta_{ij}$ for a given natural exponential family, although in § 6 we discuss how this restriction may be relaxed. For canonical link models, the existence of sufficient statistics for the random effects b_i make conditional maximum likelihood estimation possible. The mixture component of the likelihood is defined by $b_i \sim f_b(b_i|\psi)$, where $f_b(\cdot|\psi)$ is a regular joint density of the vector of random effects such that the usual regularity conditions for the standard maximum likelihood theory hold (Cox & Hinkley 1974). This density is parameterized by variance components ψ . For concreteness, we consider testing the assumption

$$b_i \stackrel{iid}{\sim} N \{0, \Sigma_b(\psi)\}, \quad (3)$$

where $\Sigma_b(\psi)$ denotes the fact that the variance covariance matrix Σ_b is parameterized as a known function of a small number of variance components ψ . Although we focus

on this normality assumption, which is the most common within the generalised linear mixed model class, the test statistic proposed in § 2.3 is applicable for testing any assumed mixing distribution. Throughout the next two sections, we implicitly assume that the fixed component of the model $X_{ij}\beta$ is correctly specified, although we also address how this type of misspecification relates to the distributional assumptions for the random effects in § 6.

The marginal likelihood is constructed under the additional assumption that the elements of Y_i are conditionally independent given b_i . Write $\theta = (\beta', \psi')'$. Under model (2)-(3), the resulting marginal likelihood function for θ is

$$\begin{aligned} L_M = L_M(\theta; Y) &= \prod_{i=1}^N \int \left\{ \prod_{j=1}^{n_i} f_{y|b}(Y_{ij} | X_{ij}, b_i) \right\} f_b(b | \psi) db, \\ &= \prod_{i=1}^N L_{M,i}(\theta; Y_i) \end{aligned}$$

In the next two sections, interest focuses on the elements of β that correspond to the covariates that vary within a cluster. Accordingly, we partition the covariates and fixed effects into $X_{ij} = (X_{ij}^W, X_i^B)$, $\beta = \{(\beta^W)', (\beta^B)'\}'$, respectively, where W and B denote within and between-cluster covariates, and we define $\eta = \{(\beta^B)', \psi'\}'$. We now consider two separate likelihood-based estimation methods for the within-cluster fixed effects β^W .

2.1 Marginal Maximum Likelihood Estimation

The asymptotic properties of the marginal maximum likelihood estimator under a possibly misspecified model are well-established. Define θ_0 as the true value of the parameter vector θ , and let $\hat{\theta}_M$ be the marginal maximum likelihood estimate under the generalised linear mixed model (2)-(3). We have

$$\hat{\theta}_M \xrightarrow{P} \theta_M,$$

where θ_M satisfies

$$\lim_{N \rightarrow \infty} E_{\theta_0} \frac{1}{N} \sum_{i=1}^N \left(\frac{\partial \log L_{M,i}}{\partial \theta} \right)_{\theta=\theta_M} = 0.$$

That is, θ_M minimizes the Kullback-Leibler information criterion (White 1982; Heagerty & Kurland 2001). Moreover,

$$\sqrt{N}(\hat{\theta}_M - \theta_M) \xrightarrow{D} MVN(0, \Sigma_M),$$

where

$$\Sigma_M = I_M(\theta_M)^{-1} \Omega_M(\theta_M) I_M(\theta_M)^{-1},$$

with

$$I_M(\theta_M) = \lim_{N \rightarrow \infty} \left(-\frac{1}{N} \frac{\partial^2 \log L_M}{\partial \theta^2} \right)_{\theta=\theta_M},$$

and

$$\Omega_M(\theta_M) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \text{cov} \left\{ \left(\frac{\partial \log L_{M,i}}{\partial \theta} \right)_{\theta=\theta_M} \right\}.$$

When the marginal likelihood is correctly specified, $\theta_M = \theta_0$ and $\Sigma_M = I_M(\theta_0)^{-1}$.

For reasons that will soon be clear, our interest focuses on the asymptotic distribution of the within-cluster fixed effects β^W . Let $\hat{\beta}_M^W$ denote the marginal maximum likelihood estimate of β^W . We partition the information matrix and its inverse from the mixed model as $I_M = \begin{pmatrix} I_M^{W,W} & I_M^{W,\eta} \\ I_M^{W,\eta} & I_M^{\eta,\eta} \end{pmatrix}$ and $I_M^{-1} = \begin{pmatrix} I_M^{W,W-} & I_M^{W,\eta-} \\ I_M^{W,\eta-} & I_M^{\eta,\eta-} \end{pmatrix}$, respectively. Note that $I_M^{W,W-} \neq (I_M^{W,W})^{-1}$. One can easily show that

$$\sqrt{N}(\hat{\beta}_M^W - \beta_M^W) \xrightarrow{D} MVN(0, \Sigma_M^W),$$

where

$$\Sigma_M^W = I_M^{W,W-}(\theta_M) \Omega_M^{W,W}(\theta_M) I_M^{W,W-}(\theta_M)$$

and

$$\Omega_M^{W,W}(\theta_M) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \text{cov} \left\{ \left(\frac{\partial \log L_{M,i}}{\partial \beta^W} \right)_{\theta=\theta_M} - I_M^{W,\eta} (I_M^{\eta,\eta})^{-1} \left(\frac{\partial \log L_{M,i}}{\partial \eta} \right)_{\theta=\theta_M} \right\}.$$

The first term within the brackets is the usual score component for the parameters of interest, from which we subtract the second term to account for the estimation of nuisance parameters η .

2.2 Conditional Maximum Likelihood Estimation

Define $T_1 = \sum_{i=1}^N \sum_{j=1}^{n_i} X_{ij}^W Y_{ij}$ and $T_{0,i} = \sum_{j=1}^{n_i} Z_{ij} Y_{ij}$ as the sufficient statistics for β^W and b_i , $i = 1, \dots, N$, respectively. In this article, we focus primarily on the case of a binary response, and so focus on the case in which Y_{ij} is discrete. Under generalised linear mixed model (2) for discrete responses, the conditional probability mass function of T_1 given $T_{0,i} = t_{0,i}$, $i = 1, \dots, N$, is

$$f_{T_1|T_0}(t_1|t_0) = \frac{C(t_1) \exp(t_1' \beta^W)}{\sum_{u \in \Psi} C(u) \exp(u' \beta^W)}, \quad (4)$$

where the reference set $\Psi = \{T_1 : \forall Y \text{ such that } T_{0,i} = t_{0,i}, i = 1, \dots, N\}$, $C(u) = \|\Psi\|_u$ denotes the number of elements in the set Ψ that are equal to u . As noted by Casella & Berger (2002, p. 272), for continuous responses a more sophisticated definition of conditional probability, such as that found in Lehmann (1986), would be required.

The conditional likelihood function $L_C(\beta^W; y)$, as a function of β^W , takes the form (4), and the asymptotic properties of the conditional maximum likelihood estimator have been well established (Cytel Software 2003). Specifically, assuming that the within-cluster portion of the model has been correctly specified, we have

$$\widehat{\beta}_C^W \xrightarrow{P} \beta_0^W.$$

Moreover,

$$\sqrt{N}(\widehat{\beta}_C^W - \beta_0^W) \xrightarrow{D} MVN(0, \Sigma_C),$$

where

$$\Sigma_C = I_C(\beta_0^W)^{-1}$$

with

$$I_C(\beta_0^W) = \lim_{N \rightarrow \infty} \left\{ -\frac{1}{N} \frac{\partial^2 \log L_C}{\partial (\beta^W)^2} \right\}_{\beta^W = \beta_0^W}.$$

2.3 A Test Statistic

The previous two sections described two likelihood-based estimation methods that yield consistent and asymptotically normal estimates of the within-cluster effects under generalised linear mixed model (2)-(3). Moreover, assuming that the distributional assumptions for the random effects are correct, the marginal maximum likelihood estimators are fully efficient in the sense that they achieve the Cramer-Rao efficiency bound for regular consistent and asymptotically normal estimators. However, misspecification of the mixing distribution may induce asymptotic bias in the point estimates of the fixed effect marginal maximum likelihood estimators. In contrast, the conditional maximum likelihood estimators are robust to any misspecification in the mixing distribution as the conditional likelihood does not involve the random effects. One possible drawback of the conditional approach is that one might experience some loss in efficiency as the conditional likelihood will only use informative clusters. An example of this loss of information is in the simple matched pair case, where the conditional maximum likelihood estimate is solely based on discordant pairs and

ignores all concordant pairs. Another drawback is that the conditional approach doesn't accommodate predictive inference at the cluster level, as all cluster-level information is conditioned out of the likelihood.

2.3.1 Formulation

Since both the marginal maximum likelihood estimator and the conditional maximum likelihood estimator are consistent when the generalised linear mixed model is correctly specified, a natural diagnostic test for a misspecified mixing distribution focuses on $\delta = \beta_0^W - \beta_M^W$, a q dimensional vector of component-wise differences between the marginal maximum likelihood estimator and conditional maximum likelihood estimator. Accordingly, we propose the test statistic

$$D = N\hat{\delta}^T \hat{\Sigma}_\delta^{-1} \hat{\delta}$$

for testing a misspecified mixing distribution, where $\hat{\delta} = \hat{\beta}_C^W - \hat{\beta}_M^W$ and Σ_δ is the asymptotic variance of the normalised difference estimator $\sqrt{N}\hat{\delta}$. Thus, D is the quadratic form of the standardized difference of two separate point estimates for the within-cluster fixed effect. Under the null hypothesis that the mixing distribution is correctly specified, D is asymptotically χ_q^2 .

Alternatively, if scientific interest focuses on a particular coefficient or subset of coefficients, we can construct analogous tests focused on these parameters of interest. Let A be a contrast matrix of rank a . Then we can consider tests $H_0 : A\delta = 0$ using the test statistic

$$D_A = (A\hat{\delta})' \hat{\Sigma}_A^{-1} (A\hat{\delta}),$$

where $\hat{\Sigma}_A = A\hat{\Sigma}_\delta A'$ is the variance-covariance matrix of $A\hat{\delta}$. Under the null hypothesis, this statistic has a χ_a^2 distribution. This generalisation allows us to assess the impact of the random effect assumptions on a subset of the within-subject parameter vector. For instance, the single-component test $D_i = \hat{\delta}_i^2 / \widehat{\text{var}}(\hat{\delta}_i)$ is a special case of D_A , having a chi-squared distribution with 1 degree of freedom.

In order to compute D or D_A , we need a consistent estimator of the variance Σ_δ . As shown in Appendix A, the asymptotic variance of $\hat{\delta}$ is

$$\Sigma_\delta = \Sigma_C + \Sigma_M - 2 \times \Sigma_C \left(\tilde{\Sigma}_C \right)^{-1} I_M^{W, W-},$$

where $\tilde{\Sigma}_C^2 = [\lim_{N \rightarrow \infty} N \times E \{ S_C(\beta_0^W) S_C^T(\beta_M^W) \}]^{-1}$. Here, S_C denotes the score corresponding to the conditional likelihood L_C . Under the null hypothesis that the distributional

assumptions for the mixing distribution hold, this asymptotic variance simplifies to

$$\Sigma_\delta = \Sigma_C - \Sigma_M, \tag{5}$$

which is guaranteed to be positive semi-definite under the null hypothesis since $\widehat{\beta}_M$ achieves the Cramer-Rao lower bound for consistent and asymptotically normal estimators. In Section 3, we describe how standard commercial software can be used to construct the proposed test statistic.

2.3.2 Consistency and Power Properties

Holly (1982) and Newey (1983, 1985) studied the consistency and the theoretical power properties of Hausman specification tests like the one considered in this article. We restrict our discussion of their findings to the simplest parametric case. Suppose we have a family of models indexed by two vectors β and η of t and s parameters with $t \leq s$ such that, suppressing the dependence on N , the likelihood based on a sample of size N is $L(\beta, \eta)$. In our setting, β consists of the within-subject regression coefficients, and η indexes the mixing distribution and between-subject coefficients. Therefore, in conducting a specification test for the mixing distribution, the vector η is of primary interest, and the vector β denotes a set of nuisance parameters used only to construct a Hausman type of statistic for the null hypothesis $H_0 : \eta = \eta^0$. We consider the distribution of the Hausman test statistic on the sequence of ‘‘Pitman’’ alternatives $\eta_N^0 = \eta^0 + N^{-1/2}\omega$. Note that $\omega = 0$ corresponds to the null hypothesis. Then, we can define the constrained and unconstrained estimators of (β, η) as

$$\begin{aligned} \widehat{\beta}^0 &= \max_{\eta=\eta^0, \beta \in \Gamma_\beta} L(\beta, \eta) \\ (\widehat{\eta}', \widehat{\beta}')' &= \max_{\eta \in \Gamma_\eta, \beta \in \Gamma_\beta} L(\beta, \eta), \end{aligned}$$

where Γ_β and Γ_η are the parameter spaces of β and η , respectively. Standard Taylor expansions give us the asymptotic approximation to the distribution of $\widehat{\beta} - \widehat{\beta}^0$ and the Hausman test statistic

$$N(\widehat{\beta} - \widehat{\beta}^0)' \left\{ (I_{\beta\beta} - I_{\beta\eta} I_{\eta\eta}^{-1} I_{\eta\beta})^{-1} - I_{\beta\beta}^{-1} \right\}^{-1} (\widehat{\beta} - \widehat{\beta}^0)$$

converges to the noncentral chi-squared distribution with t degrees of freedom and noncentrality parameter μ^2 defined as

$$\mu^2 = \omega' I_{\eta\beta} \left\{ I_{\beta\eta} \left(I_{\eta\eta} - I_{\eta\beta} I_{\beta\beta}^{-1} I_{\beta\eta} \right)^{-1} I_{\eta\beta} \right\}^{-1} I_{\beta\eta} \omega,$$

where $I_{\eta\eta}$, $I_{\eta\beta}$, $I_{\beta\eta}$ and $I_{\beta\beta}$ are the submatrices of the information matrix of $(\widehat{\beta}, \widehat{\eta})$. In view of this result, two comments can be made:

- The Hausman statistic derived here requires that $(I_{\beta\beta} - I_{\beta\eta}I_{\eta\eta}^{-1}I_{\eta\beta})^{-1} - I_{\beta\beta}^{-1}$ is a positive definite matrix. This is the case if and only if the t by s matrix $I_{\beta\eta}$ has rank t ; that is, $t \leq s$ as assumed.
- Since $t \leq s$ and the rank of $I_{\beta\eta}$ is t , its null space is of dimension $s - t$. Therefore, for any vector ω which lies in the null space of $I_{\beta\eta}$, the noncentrality parameter μ^2 is equal to zero. This implies that the power function of Hausman's test statistic in this example is flat in the directions of the null space of $I_{\beta\eta}$.

This simple parametric example shows that Hausman tests are not omnibus tests of misspecification; that is, if the dimension of the alternative space is larger than the degrees of freedom of the test, there will exist directions such that the noncentrality parameter of the test statistic is zero. These directions correspond to alternatives such that the test fails to reject with probability approaching one as the sample size grows. In settings more general than this simple parametric example, Newey (1985) argued that the actual set of failure points of Hausman specification tests has Lebesgue measure zero; consequently, the issue of consistency may not be of great importance in applications, although the power of the test can be quite low for a set of points in the misspecification parameter space.

3 Implementation

As noted in the Introduction, ideally one would like model diagnostics that are available from simple model fits. Here, we outline a strategy for calculating D or D_A from conditional maximum likelihood and mixed model software in SAS.

We focus on the random intercept logistic regression model for clustered binary responses. SAS allows straightforward calculation of $\widehat{\delta}$. We use PROC LogXact 5, a SAS procedure available from Cytel Software (Cytel Software 2003), to obtain $\widehat{\beta}_C^W$. This software uses the network algorithms developed by Mehta, Patel & Senchaudhuri (2000) to enumerate the reference set Ψ and hence obtain the conditional maximum likelihood estimate of β^W . We use SAS PROC NLMIXED (SAS Institute, 1999) to obtain marginal maximum likelihood estimates $\widehat{\beta}_M^W$. In theory, we can obtain $\widehat{\Sigma}_M$ and $\widehat{\Sigma}_C$ from the variance-covariance matrices obtained from these two fits. However, under the null in small samples,

an estimate of the difference (5) can yield a non-positive definite matrix, and under the alternative there's no guarantee that even the asymptotic limit of $\widehat{\Sigma}_\delta$ is positive definite. Thus, in such situations, we compute the influence functions corresponding to the marginal and conditional fits for each cluster, computed at the estimated values $\widehat{\beta}_C^W$ and $\widehat{\theta}_M = (\widehat{\beta}_M^W, \widehat{\eta})$, respectively, and take the empirical variance of the difference of these quantities. See Appendix A, equation (8) for details. This results in test statistic $\widetilde{D} = N\widehat{\delta}'\widetilde{\Sigma}_\delta\widehat{\delta}$, where $\widetilde{\Sigma}_\delta$ is the estimate of Σ_δ based on influence functions. Here, $\widetilde{\Sigma}_\delta$ is a consistent estimator of Σ_δ regardless of whether the null hypothesis holds, whereas $\widehat{\Sigma}_\delta$ is consistent under H_0 only. Appendix B shows simple SAS code for computing the test statistic using the simple variance formula (5). SAS code that implements the influence function estimator, which is somewhat longer, is available from the authors upon request.

4 A Simulation Study

Here we present the results of a simulation study that investigated the ability of the proposed test statistic \widetilde{D} to detect model misspecification in a generalised linear mixed model in finite samples. We present Monte Carlo estimates of the size and power of the test based on \widetilde{D} under different models. The mean model is motivated by the amenorrhea example in § 6, in which randomization makes inclusion of a main effect of the between-cluster covariate unnecessary.

Specifically, we simulated data from the model

$$\text{logit}\{P(Y_{ij} = 1|b_i)\} = \beta_0 + \beta_1 \mathbf{time}_{ij} + \beta_2 X_i^B \times \mathbf{time}_{ij} + b_i,$$

for $i = 1, \dots, N$ and $j = 1, \dots, n$.

The model contains a single random intercept, a within-cluster covariate $\mathbf{time}_{ij} = j$, and the interaction of this within-cluster covariate with between-cluster covariate X_i^B . We specify the between-cluster covariate as $X_i^B \sim \text{Bin}(1, .5)$, $i = 1, \dots, N$. We present results for $\beta = (\beta_0, \beta_1, \beta_2)' = (-0.5, 0.2, 0.5)'$. Here, $\beta^W = (\beta_1, \beta_2)'$.

We simulated the power of the test statistic \widetilde{D} to detect misspecification of the random effects under different scenarios defined by varying the cluster size (n), the number of clusters (N), and the true distribution of the random effects. Specifically, we simulated the power of the test statistic by simulating 250 data sets for each combination of the settings $N = (150, 300, 600)$ and $n = (3, 5)$, under seven different distributions for the random

effects. The seven distributions considered for b_i , and their shorthand labels for use in Table 2, are

1. $b_i \sim N(0, 1)$. (Standard Normal)
2. $b_i \sim 2 * W_i$, where $W_i \sim \text{Gamma}(1, 1)$. (Gamma 1)
3. $b_i \sim 3 * W_i$, where $W_i \sim \text{Gamma}(0.5, 1)$. (Gamma 0.5)
4. $b_i \sim N \{0, (1 + 2X_i^B)^2\}$. (Heterogeneous Variance 1)
5. $b_i \sim N \{0, (0.5 + 2.5X_i^B)^2\}$. (Heterogeneous Variance 0.5)
6. $b_i \sim \gamma * N(2.0, 0.5) + (1 - \gamma) * N(-0.86, 0.5)$, where $\gamma \sim \text{Bin}(0.3)$. (Mixture 0.5)
7. $b_i \sim \gamma * N(2.0, 0.25) + (1 - \gamma) * N(-0.86, 0.25)$, where $\gamma \sim \text{Bin}(0.3)$. (Mixture 0.25)

The first distribution corresponds to the null hypothesis of normality. The second and third distributions reflect skewed mixing distributions. Scenarios 4 and 5 correspond to distributions in which the variance of the random effects varies according to the between-cluster covariate, and are two cases in which the bias of the fixed effect estimates can be large (Heagerty & Kurland 2001). The last two cases correspond to mixtures of normal distributions, which result in bimodal distributions.

Table 1 reports the simulated power of the 0.05 level χ_2^2 test based on \tilde{D} . The table indicates that the size of the test is close to the nominal level for all sample scenarios, with the power increasing for cases of model misspecification. The power is lowest for scenario 2, which corresponds to the less skewed Gamma distribution, and scenario 6, which corresponds to the mixture distribution having normal components with relatively large variances. This is somewhat expected since, out of all the alternatives considered by Heagerty & Kurland (2001), mistakenly assuming normality when the true mixing distribution is gamma induces the smallest amount of asymptotic relative bias in the fixed effect estimates, on the order of 10% or lower. However, for relatively large samples and moderately-sized clusters ($n_i = 5$), the test statistic is able to detect departures even for this relatively mild form of misspecification. In contrast, the simulated power of the test is highest for the case Heagerty & Kurland (2001) singled out as the one that induces the largest asymptotic relative bias in the fixed effects estimators. Specifically, the test has the highest power for the case in which we assume a constant variance component for the random effects when

in fact this variance depends on the between-cluster covariate. The tests are able to detect this type of misspecification even for very small clusters ($n_i = 3$). This is reassuring since Heagerty & Kurland (2001) showed that the asymptotic relative bias in this case can reach as much as 70-80%. The power increases as we move from clusters of size 3 to clusters of size 5.

5 Application

Fitzmaurice, Laird, & Ware (2004) used a logistic generalised linear mixed model with random intercepts to analyze data from a longitudinal clinical trial examining the effects of hormonal contraceptives in women. In the trial, contracepting women received four successive injections of either 100 mg or 150 mg of depot-medroxyprogesterone acetate at 0, 90, 180, and 270 days after randomization, with this dosage remaining constant for each subject over the course of the study. There was also a final follow-up visit one year after the first injection. The analysis, which was based on $N = 1151$ women, focused on the within-subject effects of time on the binary outcome of whether a woman experienced amenorrhea in the four successive three-month intervals, and whether this trend in risk varied according to dosage.

Let $Y_{ij} = 1$ if woman i , $i = 1, \dots, 1151$, experienced amenorrhea in the j^{th} injection interval, $j = 1, \dots, 4$, and $Y_{ij} = 0$ otherwise. Fitzmaurice, Laird & Ware (2004) considered the model

$$\text{logit} \{E(Y_{ij}|b_i)\} = \beta_1 + \beta_2 \mathbf{time}_{ij} + \beta_3 \mathbf{time}_{ij}^2 + \beta_4 \mathbf{dose}_i \times \mathbf{time}_{ij} + \beta_5 \mathbf{dose}_i \times \mathbf{time}_{ij}^2 + b_i, \quad (6)$$

where $\mathbf{time}_{ij} = 1, 2, 3, 4$ for the four consecutive 90-day injection intervals and $\mathbf{dose}_i = 1$ if subject i is randomized to 150mg of depot-medroxyprogesterone acetate and $\mathbf{dose}_i = 0$ otherwise. The model specifies a quadratic within-subject effect of time, with this trend differing according to the dosage received. Because of randomization, the model does not include a main effect of drug, which corresponds to assuming that no differences exist between the two drug groups at baseline. In this model, \mathbf{dose}_i is a between-subject effect and \mathbf{time}_{ij} is a within-subject effect. Thus, $\beta^W = (\beta_2, \beta_3, \beta_4, \beta_5)'$.

Fitzmaurice, Laird & Ware (2004) completed the specification of the model by assuming $b_i \stackrel{iid}{\sim} N(0, \sigma_b^2)$. Table 2 presents the parameter estimates for θ and corresponding standard errors from the fit of this model. Results suggest that there is a significant effect of dose on

the trend for the risk of amenorrhea, and that there is a large amount of heterogeneity in the baseline risk among subjects. Table 2 also presents the results from the conditional fit, the differences in the estimates, δ , from the two methods of fitting, and their standard errors. Results suggest that the distributional assumptions on the random effects play a large role in the resulting values of the parameter estimates, with the differences in elements of $\hat{\beta}^W$ between the two fits generally being 50% of the generalised linear mixed model estimates. The proposed hypothesis test of the mixing distribution based on (5) yields $D = 14.29$ and $p = .0064$. Results based on the influence function variance estimator are similar, with $\tilde{D} = 18.84$ and $p = .0008$. Thus, these diagnostics strongly suggest that the normal assumption does not hold for these data. Figure 1 displays this violation graphically by showing a histogram of the Empirical Bayes predictions \hat{b}_i of the random effects, obtained from NLMIXED. This plot suggests that there are multiple modes in the distribution, with an unusually large right tail. Thus, there appears to be multiple modes in the distribution of amenorrhea risk, with a large number of women exhibiting a relatively high baseline risk of amenorrhea. The magnitude of the elements in $\hat{\delta}$ suggest that this non-normality can noticeably affect inference on the effect of depot-medroxyprogesterone acetate dosage on a women's change in risk over time. The single component tests, which can be inferred by comparing the estimates of each difference to its standard error, are not significant. However, Figure 2, which shows the estimated trends over time for the two dose groups from both model fits, suggests that when taken together these assumptions play a large role on the resulting overall inference. The plot on the left shows the estimated trends based on the random effects model, whereas the plot on the right shows the estimates from the conditional fit, using the marginal maximum likelihood estimate of the intercept. We see that, in this case, the model assuming normal random effects underestimates the difference in these trends by approximately 50%.

6 Discussion

We have proposed a simple method for testing the distributional assumptions on the random effects distribution in a generalised linear mixed model. We focused on a diagnostic test for the random effects distribution that is valid and computationally simple; that is, analysts could conduct the test with minimal effort using available output from SAS. This led us to focus on the test formed by comparing the marginal maximum likelihood estimate to the

conditional maximum likelihood estimate, and we used relatively straightforward algebraic arguments presented in the appendix to derive the asymptotic variance of the test statistic based on this difference.

This simplicity of the method, which we believe to be a strong selling point, comes at the expense of two restrictions: the test is only applicable to models using the canonical link and to models containing at least one within-subject effect. Assuming computational simplicity is not paramount, one can overcome these restrictions by comparing the marginal maximum likelihood estimate to an alternative estimate derived under nonparametric assumptions for the mixing distribution. For instance, one could construct a difference estimator using estimates of the fixed effects obtained using the smooth nonparametric estimator of Chen, Zhang & Davidian (2001) for the fixed effects. This alternative difference estimator has the advantage that it is applicable both when the model contains only between-subject covariates or for non-canonical link functions, thus relaxing restrictions encountered by the conditional maximum likelihood estimator. However, it has the disadvantage that it requires complex algorithms, such as Monte Carlo EM, for model fitting. Although the algebraic arguments in the appendix do not extend naturally to this more general difference estimator, one can apply the powerful theory of Newey (1985) to obtain a null variance estimator for this alternative estimator. Such a variance estimator has form analogous to equation (5) in the paper, with the variance covariance matrix for the conditional maximum likelihood estimator replaced by the variance covariance matrix for the nonparametric maximum likelihood estimator of the fixed effects. Following Newey, this result holds since the marginal maximum likelihood estimator is fully efficient under the null hypothesis.

Even though the above extension makes specification tests applicable to models containing only between-subject effects, this is not necessarily desirable. This is because the fixed effect coefficients in a random effects model relate to within-subject effects that are neither directly of interest nor directly observable from the data in purely between-subject designs. Such considerations led Neuhaus (2000) to conclude that “...one should not report [fixed effect estimates from random effects models] for [between-subject] covariates in the first place.” Thus, for such designs, one would ordinarily prefer a marginal regression model (either likelihood-based or one based on generalised estimating equations) for the data analysis, and so the proposed tests of a mixing distribution in a mixed model are of less relevance in this situation. Thus, the restriction that the test requires at least one

within-subject covariate does not seem to be a severe shortcoming of the method.

Another standard method for assessing normality of the random effects is simple visual inspection of the predictions of the random effects. However, non-normality may not always be apparent from this informal check since these estimates inherently shrink the estimates back towards the assumed distribution (Shen & Louis 1998). Moreover, any observed deviations from normality may be attributable to sampling variability. Thus, as demonstrated by the amenorrhea example, the proposed test can often serve as a useful complement to less formal methods.

Strictly speaking, because the conditional approach conditions out all between cluster information, the test is focused on the goodness of fit of the entire between-cluster component, $X_{ij}^B \beta^B + Z_{ij} b_i$, and not solely on the random effects distribution of b_i . However, the fixed effects in the between-cluster component can be viewed as structure placed on the mean of the cluster-specific random effects. For instance, suppose we omit an important dichotomous cluster-specific covariate from the model. Then, the random effects will appear to arise from a bimodal distribution. Using the distributional diagnostic proposed, this “violation of the normality assumption” will be apparent. Thus, such between-cluster effects and the random effects distribution are intrinsically linked, and it is appropriate to test for them jointly and to label this test a test for the mixing distribution of a generalised linear mixed model.

We have framed the problem as one that distinguishes between between-subject and within-subject covariates. Situations arise, however, in which a covariate has both a between and a within-cluster component. As noted by Neuhaus and Kalbfleisch (1998), for marginal models that specify a common effect for both of these components, there is no reason to expect the marginal maximum likelihood and conditional maximum likelihood should agree, even when assumptions about the mixing distribution are met. Thus, a test based on this difference is not a test of the mixing distribution. However, in such cases one can follow the recommendation of Neuhaus and Kalbfleisch (1998) and partition the between- and within-cluster components of such an effect. The proposed specification test applied to the within-cluster coefficient in this model is then valid.

Throughout this article, we focus our attention on the simple random intercepts case. Also focusing on this case, Heagerty & Kurland (2001) showed that model misspecification in the form of the variance of the random effects depending on a measured covariate or the

random effects following an autoregressive structure can result in severe bias for the fixed effects estimates. This first scenario is a type of misspecification for which no effective, simple diagnostic test exists, and we show via simulation that our method is particularly sensitive to this form of violation. Our diagnostic test does not effectively detect departures of the second type. This is because, in the autoregressive case, the usual conditional maximum likelihood estimator is also biased for the fixed effects, and the difference between the two available estimators is not a consistent estimate of the bias incurred by the generalised linear mixed model estimator. This insensitivity is not a serious drawback of the method, however, since methods already exist for assessing serial correlation among random intercepts. For instance, Diggle et al. (2002) proposed fitting the autoregressive model using a fully Bayesian approach via Markov Chain Monte Carlo, and performing inference on the autoregressive correlation parameter. Such a test can be implemented using the commercial MCMC software WinBUGS (Spiegelhalter, Thomas & Best 2000).

ACKNOWLEDGEMENT

This research was supported in part by grants from the U. S. National Institute of Environmental Health Sciences. The authors thank two referees for insightful comments that significantly improved the manuscript.

APPENDIX

Appendix A: Asymptotic Variance of $\hat{\delta}$

We derive the asymptotic distribution of D . It is useful to write the marginal likelihood score for β^W , which we denote as $S_M(\beta^W, \eta)$, as a function of the conditional likelihood score $S_C(\beta^W)$.

Recall that by the sufficiency of T_1 for β^W and of $T_{0,i}$ for b_i , $i = 1, \dots, N$, the likelihood



function factors as (Tjur 1982; Satori & Severini 2004)

$$\begin{aligned}
 L_M &= \prod_{i=1}^N \int \exp \left[t'_{1,i} \beta^W + t'_{0,i} b + y'_i X_i^B \beta^B - \sum_{j=1}^{n_i} \log \{1 + \exp (Z_{ij} b + X_{ij} \beta)\}' \right] f(b) db \\
 &= \frac{C(t_1) \exp (t'_1 \beta^W)}{\sum_{u \in \Psi} C(u) \exp (u' \beta^W)} \times \\
 &\quad \prod_{i=1}^N \int \exp \left[t'_{0,i} b + y'_i X_i^B \beta^B - \sum_{j=1}^{n_i} \log \{1 + \exp (Z_{ij} b + X_{ij} \beta)\}' \right. \\
 &\quad \left. + \log \left\{ \sum_{u' \in \Psi} C(u') \exp (u' \beta^W) \right\} \right] f(b) db \\
 &= L_C L_R.
 \end{aligned}$$

The scores $S_M(\beta^W, \eta)$ and $S_C(\beta^W)$ are given by differentiating the respective log-likelihoods with respect to β^W , yielding

$$\begin{aligned}
 S_M(\theta_M) &= S_C(\beta^W) + \\
 &\quad \sum_{i=1}^N \left[\frac{\sum_{u \in \Psi} u \times \frac{C(u) \exp (u' \beta^W)}{\sum_{u \in \Psi} C(u) \exp (u' \beta^W)} - \int \left\{ \sum_{j=1}^{n_i} X_{ij} \frac{\exp (Z_{ij} b + X_{ij} \beta)}{1 + \exp (Z_{ij} b + X_{ij} \beta)} \right\}}{\exp [t'_{0,i} b + y'_i X_i^B \beta^B - \sum_{j=1}^{n_i} \log \{1 + \exp (Z_{ij} b + X_{ij} \beta)\}]} \right. \\
 &\quad \left. \times \frac{f(b) db}{\int \exp [t'_{0,i} b + y'_i X_i^B \beta^B - \sum_{j=1}^{n_i} \log \{1 + \exp (Z_{ij} b + X_{ij} \beta)\}] f(b) db} \right] \\
 &= S_C(\beta^W) + \sum_{i=1}^N \{E(T_{1i}|T_{0,i}) - E_b E(T_{1,i}|b)\} \\
 &= S_C(\beta^W) + S_R(\beta^W).
 \end{aligned}$$

Here, for a given value of β^W , $E(T_{1,i}|T_{0,i})$ denotes the expectation with respect to the conditional distribution of $T_{1,i}$ given $T_{0,i} = t_{0,i}$ and E_b denotes the expectation with respect to the posterior distribution $h(b|y)$ given the data, where

$$h(b|y) = \frac{\exp \left[t'_{0,i} b + y'_i X_i^B \beta^B - \sum_{j=1}^{n_i} \log \{1 + \exp (Z_{ij} b + X_{ij} \beta)\}' \right] f(b)}{\int \exp \left[t'_{0,i} b + y'_i X_i^B \beta^B - \sum_{j=1}^{n_i} \log \{1 + \exp (Z_{ij} b + X_{ij} \beta)\}' \right] f(b) db}$$

Thus, the score for β^W based on marginal maximum likelihood estimation of a generalised linear mixed model is a sum of two components: the conditional score and a term that is the discrepancy between the conditional (on $T_{0,i}$) and marginal expectations of the sufficient statistic for β^W . Note that $S_C(\beta^W)$ depends on the data only through T_1 whereas $S_R(\theta_M)$ and $\frac{d \log(L_M(\theta_M))}{d \eta}$ depend on the data through T_0 and $\left\{ (X_i^B)', (X_{ij}^W)' \right\}'$. Therefore, $\sqrt{N} S_C(\beta_0^W)$ and $\left\{ \sqrt{N} S_R(\theta_M), \frac{d \log L_M(\theta_M)}{d \eta} \right\}$ are uncorrelated regardless of whether the mixing distribution is correctly specified.

Therefore,

$$\begin{aligned}
\text{var} \left(\lim_{N \rightarrow \infty} \sqrt{N} \widehat{\delta} \right) &= \text{var} \left(\lim_{N \rightarrow \infty} \sqrt{N} \widehat{\beta}_M^W \right) + \text{var} \left(\lim_{N \rightarrow \infty} \sqrt{N} \widehat{\beta}_C^W \right) - 2 \times \text{cov} \left(\lim_{N \rightarrow \infty} \sqrt{N} \widehat{\beta}_C^W, \lim_{N \rightarrow \infty} \sqrt{N} \widehat{\beta}_M^W \right) \\
&= \Sigma_M + \Sigma_C - 2 \times \Sigma_C \\
&\quad \times E \left[\left\{ \lim_{N \rightarrow \infty} \sqrt{N} S_C(\beta_0^W) \right\} \left[\lim_{N \rightarrow \infty} \left[\begin{array}{c} \sqrt{N} S_M(\theta_M) - \\ I_M^{W\eta} (I_M^{\eta\eta})^{-1} \left\{ \frac{1}{\sqrt{N}} \frac{d \log L_M(\theta_M)}{d\eta} \right\} \end{array} \right] \right]' \right] \\
&\quad \times I_M^{W, W-} \\
&= \Sigma_M + \Sigma_C - 2 \times \Sigma_C \times E \left[\left\{ \lim_{N \rightarrow \infty} \sqrt{N} S_C(\beta_0^W) \right\} \left\{ \lim_{n \rightarrow \infty} \sqrt{N} S_C(\beta_M^W) \right\}' \right] \times I_M^{W, W-} \\
&= \Sigma_C + \Sigma_M - 2 \times \Sigma_C \times \left(\widetilde{\Sigma}_C \right)^{-1} \times I_M^{W, W-}.
\end{aligned}$$

Here, for notational simplicity, we have suppressed the dependence of the submatrices of I and I^{-1} on θ . Under the null hypothesis, we have $\widetilde{\Sigma}_C = \Sigma_C$ and $\Sigma_M = I_M^{W, W-}$, leading to the simplified expression

$$\text{var} \left(\lim_{N \rightarrow \infty} \sqrt{N} \widehat{\delta} \right) = \Sigma_C - \Sigma_M. \quad (7)$$

In some small to moderate sample settings, closed-form expression (7) yields a non-positive definite matrix. In such settings one may use the influence functions directly to obtain a consistent estimate of Σ_δ (Van der Vaart 1998, p. 292). Specifically, we use the representation

$$\begin{aligned}
\lim_{N \rightarrow \infty} \sqrt{N} \begin{pmatrix} \widehat{\beta}_C^W - \beta_0^W \\ \widehat{\beta}_M^W - \beta_M^W \end{pmatrix} &\stackrel{D}{=} \lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ \begin{array}{c} I_C^{-1} S_{C,i}(\beta_0^W) \\ I_M^{W, W-} S_{M,i}(\theta_M) - I_M^{W\eta} (I_M^{\eta\eta})^{-1} \frac{d \log L_{M,i}(\theta_M)}{d\eta} \end{array} \right\} \\
&\equiv \lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \sum_{i=1}^N \begin{pmatrix} \text{IF}_{C,i} \\ \text{IF}_{M,i} \end{pmatrix}. \quad (8)
\end{aligned}$$

Based on (8),

$$\widetilde{\Sigma}_\delta = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left(\widehat{\text{IF}}_{C,i} - \widehat{\text{IF}}_{M,i} \right) \left(\widehat{\text{IF}}_{C,i} - \widehat{\text{IF}}_{M,i} \right)',$$

where $\widehat{\text{IF}}_{C,i}$ and $\widehat{\text{IF}}_{M,i}$ are estimates of $\text{IF}_{C,i}$ and $\text{IF}_{M,i}$, respectively, obtained by replacing the information matrices by their empirical versions and substituting estimates $\widehat{\beta}_C^W$ and $\widehat{\theta}_M$ for β_0^W and θ_M , respectively.

Appendix B: SAS Code

```
proc nlmixed data=new qpoints=50 FDHESSIAN=CENTRAL cov;
  parms b1=0.7 b2=.3 b3=-0.5 b4=0.5  b0=-2  s1=2;
  lp = b0 + b1*time  +b2*trt*time + b3*time2 + b4*trt*time2 + u*s1;
  p=(exp(lp))/(1+exp(lp));
  model y ~ binary(p);
  random u~normal(0,1)  subject=id;
  ods output ParameterEstimates=Parms(where=(parameter
                                         in  "b1","b2","b3","b4")));
  ods output CovMatParmEst=ml_cov0(keep= b1 b2 b3 b4);
run;
proc logxact  data =new  ;
  STRATUM ID;
  model y=time trttime time2 trttime2  ;
  ES/AS POSTFIT  ESTIMATEFILE=fit time trttime time2 trttime2  ;
run;
proc iml;
  use ml_cov0;
  read  all var {b1 b2  b3 b4 } into ml_cov;
  use parms;
  read  all var {estimate} into parms;
  use fit ;
  read all var {time trttime time2 trttime2 } into cond_cov;
  read all var {beta} into cparms;
  delta=parms-cparms;
  test_var=cond_cov-ml_cov[1:4,];
  test=t(delta)*inv(test_var)*(delta);
  pvalue=1-probchi(test,4);
  print test pvalue;
quit;
```

REFERENCES

- Aitkin M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models *Statist. Comput.* **6**, 251–62.
- Aitkin M. (1999). A general maximum likelihood analysis of variance components in generalized linear models *Biometrics* **55**, 117–28.
- Agresti, A., Caffo, B. & Ohman-Strickland, P. (2003). Examples in which misspecification of a random effects distribution reduces efficiency. Technical Report, Department of Statistics, University of Florida.
- Breslow, N. E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Am. Statist. Assoc.* **88**, 9–25.
- Casella, G. and Berger, R. L. (2002). *Statistical Inference*, 2nd Ed. Pacific Grove, CA: Duxbury Press.
- Chen J. L., Zhang D. W. & Davidian M. (2002). A Monte Carlo EM algorithm for generalized linear mixed models with flexible random effects distribution. *Biostatist.* **3**, 347-60.
- Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman & Hall.
- Cytel Software, Inc. (2003). *Proc-LogXact 5 For SAS Users. Logistic Regression Software Featuring Exact Methods. User Manual*. Cambridge, MA: Cytel Software, Inc.
- Diggle, P. J., Heagerty, P., Liang, K.-Y. & Zeger, S. L. (2002). *Analysis of Longitudinal Data* (2nd ed). Oxford: Oxford University Press.
- Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied Longitudinal Analysis*. New York: Wiley.
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica* **46**, 1251–72.
- Heagerty P. J. (1999). Marginally specified logistic-normal models for longitudinal binary data *Biometrics* **55**, 688–98.

- Heagerty, P. J. & Kurland, B. F. (2001). Misspecified maximum likelihood estimates and generalized linear mixed models. *Biometrika* **88**, 973-85.
- Heagerty P. J. & Zeger S. L. (2000). Marginalized multilevel models and likelihood inference *Statist. Sci.* **15**, 1-19.
- Holly, A. (1982). A remark on Hausman's specification test. *Econometrica*, **50** 749-59.
- Houseman, E. A., Coull, B. A. & Ryan, L. M. (2004). A functional-based distribution diagnostic for a linear model with correlated outcomes. Technical Report, Department of Biostatistics, Harvard School of Public Health.
- Lange, N. and Ryan, L. (1989). Assessing normality in random effects models. *Ann. Statist.* **17**, 624-42.
- Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, 2nd edition. New York: Wiley.
- Magder, L. S. & Zeger, S. L. (1996). A smooth nonparametric estimate of a mixing distribution using mixtures of Gaussians. *J. Am. Statist. Assoc.* **91**, 1141-51.
- Mehta C. R., Patel N. R. & Senchaudhuri P. (2000). Efficient Monte Carlo methods for conditional logistic regression. *J. Am. Statist. Assoc.* **95** 99-108.
- Neuhaus, J. M. (2000). Comment on "Marginalized multilevel models and likelihood inference". *Statist. Sci.* **15**, 21-22.
- Neuhaus, J. M., Hauck, W. W. & Kalbfleisch, J. D. (1992). The effects of mixture distribution misspecification when fitting mixed-effects logistic-models. *Biometrika* **79**, 755-62.
- Neuhaus, J. M. & Kalbfleisch, J. D. (1998). Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* **54**, 638-45.
- Newey, W. K. (1983). Specification testing and estimation using a generalized method of moments. Ph.D. thesis Department of Economics. MIT.
- Newey, W. K. (1985). Generalized Method of Moments Specification Testing. *J. Econ.* **29**, 229-56 .

- SAS Institute, Inc. (1999). SAS/STAT User's Guide, Version 8. Cary, NC: SAS Institute, Inc. pp. 2421–504.
- Sartori, N. & Severini, T. A. (2004). Conditional likelihood inference in generalized linear mixed models. *Statist. Sin.* **14**, 349–60.
- Shen W. & Louis T. A. (1998). Triple-goal estimates in two-stage hierarchical models *J. Roy. Statist. Soc., B* **60**, 455–71.
- Spiegelhalter, D., Thomas, A. & Best, N. (2000). *WinBUGS Version 1.3 User Manual*. MRC Biostatistics Unit.
- Tjur, T. (1982). A connection between Rasch's item analysis model and a multiplicative Poisson model. *Scand. J. Statist.* **9**, 23–30.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.
- Verbeke, G. & Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random effects population. *J. Am. Statist. Assoc.* **91**, 217–21.
- White, H. (1982). Maximum-likelihood estimation of mis-specified models. *Econometrica* **50**, 1-25.
- Zeger, S. L., Liang, K.-Y. & Albert, P. A. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics* **44**, 1049–60.

Table 1. Simulated Power of the Test Based on \tilde{D} to Detect Misspecification of the Random Effects Distribution.

Mixing Distribution	N =	n=3			n=5		
		150	300	600	150	300	600
1 Standard Normal		3.0	3.6	6.0	5.2	4.4	5.8
2 Gamma 1		8.0	11.6	24.0	18.2	40.2	68.0
3 Gamma 0.5		10.0	18.0	40.4	34.2	68.0	93.4
4 Heterogeneous Variance 1		11.0	58.8	85.0	33.2	83.8	99.8
5 Heterogeneous Variance 0.5		17.0	59.0	98.0	55.8	99.0	100
6 Mixture 0.5		4.8	15.4	25.0	18.6	37.6	64.0
7 Mixture 0.25		7.4	16.2	36.2	29.0	50.0	77.8

Table 2. Marginal and Conditional MLE's from logistic model (6) for the Amenorrhea Data.

Variable	MMLE (SE)	CMLE (SE)	$\hat{\delta}$	$s.e(\hat{\delta})$
Intercept	-3.8057 (0.3050)	—	—	
time_{ij}	1.1332 (0.2682)	0.7587 (0.3766)	-0.3745	0.26
time_{ij}^2	-0.0419 (0.0548)	0.03668 (0.07422)	0.0786	0.05
$\text{dose}_i \times \text{time}_{ij}$	0.5644 (0.1922)	1.2107 (0.5324)	0.6463	0.50
$\text{dose}_i \times \text{time}_{ij}^2$	-0.1095 (0.0496)	-0.2235 (0.1040)	-0.1139	0.09
σ_b^2	5.0646 (0.5840)	—	—	
$D = 14.29, p = .0064$				
$\tilde{D} = 18.84, p = .0008$				

Figure 1: Histogram of Empirical Bayes predictions \hat{b}_i of the random effects from Model (6) Applied to the Amenorrhea Data.

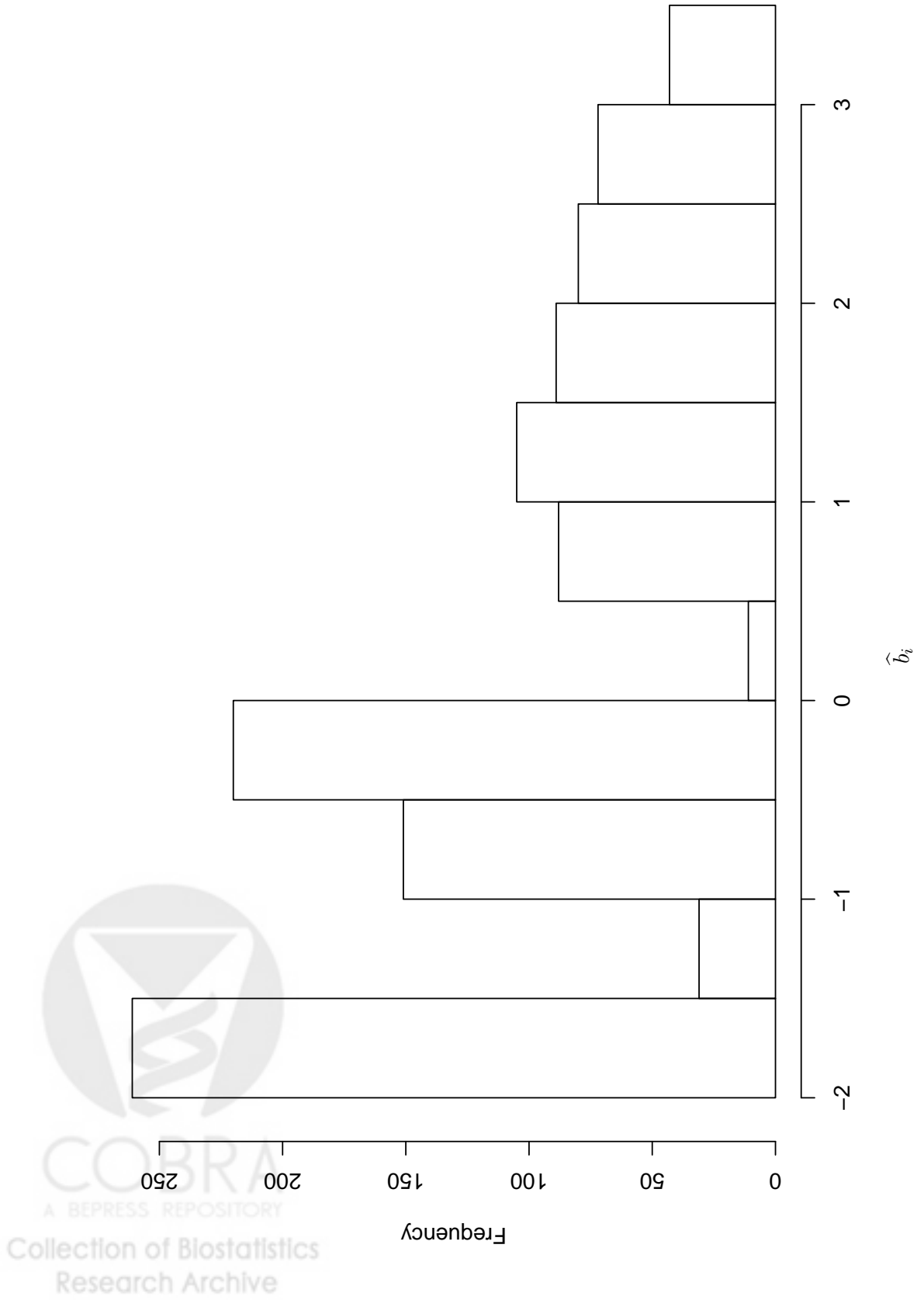


Figure 2: Fitted Probabilities from Model (6) for the Amenorrhea Data, Based on Marginal Maximum Likelihood Estimation Assuming Normal Random Intercepts and Conditional Maximum Likelihood Estimation, Using the MML Estimate of the Intercept.

