STATISTICS IN MEDICINE
Statist. Med. 2003; **22**:3527–3541 (DOI: 10.1002/sim.1572)

Measuring explained variation in linear mixed effects models

Ronghui Xu*,†,

Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, U.S.A.

SUMMARY

We generalize the well-known R^2 measure for linear regression to linear mixed effects models. Our work was motivated by a cluster-randomized study conducted by the Eastern Cooperative Oncology Group, to compare two different versions of informed consent document. We quantify the variation in the response that is explained by the covariates under the linear mixed model, and study three types of measures to estimate such quantities. The first type of measures make direct use of the estimated variances; the second type of measures use residual sums of squares in analogy to the linear regression; the third type of measures are based on the Kullback–Leibler information gain. All the measures can be easily obtained from software programs that fit linear mixed models. We study the performance of the measures through Monte Carlo simulations, and illustrate the usefulness of the measures on data sets. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS: empirical Bayes; explained randomness; Kullback–Leibler information; predicted random effects; residual sum of squares

1. INTRODUCTION

The R^2 measure of explained variation is well-known for linear regression. Similar measures have also been developed for more complex regression models such as the logistic regression [1, 2] and the proportional hazards regression [3–5] for independent identically distributed (i.i.d.) data. Little has been done, however, to develop such measures for correlated data, in particular under mixed effects models that are often used to fit this type of data.

The work of this paper was motivated by an informed consent study (E1Z96) conducted by the Eastern Cooperative Oncology Group (ECOG) [6]. In E1Z96, two different versions of informed consent documentation, standard or easy-to-read, were given to cancer patients who might potentially participate in ECOG clinical trials. Written informed consent has been required by the Department of Health and Human Services for patients to participate in research studies sponsored by the agency. It serves a major role in providing the patient with the information necessary to make a truly informed decision as whether or not to participate

* Correspondence to: Ronghui Xu, Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, U.S.A.
† E-mail: rxu@jimmy.harvard.edu

Received October 2002
Accepted April 2003

in a clinical trial. The regulations also stipulate that an Institutional Review Board review and approve or disapprove all clinical trials conducted at an institution. Owing to the potential 'contamination' and logistical difficulties of using two different consent forms within a single institution, the patients on E1Z96 were randomized to one of the two versions of consent forms according to their institutions, that is, a cluster randomization design [7] was used. The study endpoints included informed consent comprehension, state anxiety, consent anxiety, satisfaction and decisional conflict. To account for the possible correlation among patients from the same institution, random effects models were used to analyse the outcome data for differences between the two versions of consent documentation. As with many studies in biomedicine, psychiatry and other social sciences, beyond the primary comparison between the two randomized arms, the investigators would also like to know further the association between the outcomes and the patient characteristics that were collected in the study, and how much variation in the outcomes was explained by these covariates.

There is perhaps little doubt that the R^2 measure is very popular in the applied fields of statistics. Unlike hypothesis testing, where the p-values can be heavily influence by sample sizes, an R^2 measure, when properly defined, consistently estimates the proportion of explained variation in the population. In assessing the importance of a covariate, it also overcomes the limitation of the regression coefficient which depends on the scale and, in multiple regression, it summarizes the effects of several covariates. As pointed out by reference [8], such measures are not concerned with fit but with predictability. We note that in linear regression, the usual R^2 measure has several equivalent forms, including the squared correlation coefficient, a ratio of sums of squares, and a transformation of the likelihood ratio statistic. From our previous experience in developing R^2 measures for the Cox proportional hazards regression model [4, 5, 9] we have learned that, in order to develop an R^2 measure with desired properties under a more complex model, it is important to understand the concept of explained variation. It is also important to understand the relationship among the different forms mentioned above, which are generally not algebraically equivalent outside the linear models but should reflect similar strengths of association.

In this paper we quantify the variation in the response that is explained by the covariates under the linear random effects model [10]. We investigate three measures aimed to estimate this population quantity. The first measure makes direct use of the estimated variances under the models. For the other two measures, we take an empirical Bayesian approach, where the random effects in the model are viewed as parameters to be estimated; they are a realized or sample value of the random effects given the data [11, 12]. There are many applications where the estimation of the random effects is of practical interest (see, for example, reference [13–15]); it is a strength of the random effects model over the marginal models to be able to estimate the random effects. Here we focus on prediction using the estimated fixed as well as random effects, so that the prediction is based on cluster level, rather than the population average.

In the next section we introduce the concept of explained variation under the linear mixed model and two types of measures for estimating such quantities. The measures are generalizations of the R^2 for linear regression. In Section 3 we introduce measures of explained randomness based on the Kullback–Leibler information gain, and show their relationship to the two types of measures of Section 2. In Section 4 through simulations we study the performance of the measures and their approximation to the population quantities. We return to the E1Z96 data in Section 5, and further illustrate the use of the measures on a clinical trial for the treatment of schizophrenia. Section 6 contains further discussion and the last section conclusion.

2. MEASURES OF EXPLAINED VARIATION

2.1. Explained variation

Suppose that we have clustered data. The linear mixed effects model [10] can be written

$$Y_{ij} = \boldsymbol{\beta}'\mathbf{Z}_{ij} + \mathbf{b}_i'\mathbf{W}_{ij} + \varepsilon_{ij} \tag{1}$$

where Y_{ij} is the observed response, $\mathbf{Z}_{ij}$ and $\mathbf{W}_{ij}$ are the covariate vectors corresponding to the fixed and the random effects, for subject j from cluster i ($i=1,\ldots,n,\ j=1,\ldots,n_i$), $\boldsymbol{\beta}$ is the vector of unknown fixed effects, and $\mathbf{b}_i$ is the vector of random effects for cluster i. In our E1Z96 data example, if $\mathbf{W}_{ij}=1$, then $\mathbf{b}_i$ is the random institutional effect on the outcome such as comprehension score (see also Section 5); if $\mathbf{W}_{ij}$ is the intervention assignment (standard versus easy-to-read), then $\mathbf{b}_i$ is the random institution by intervention interaction. The ε_{ij}'s are i.i.d. $\mathrm{N}(0,\sigma^2)$, and are independent of $\mathbf{Z}$, $\mathbf{W}$ and $\mathbf{b}$. Here the same notation without subscripts denotes the random variable of which data is the realization. Although $\mathbf{Z}$ and $\mathbf{W}$ are sometimes fixed by design, here for the purpose of studying the underlying population variation, it is helpful to treat them as random. We assume $\mathbf{W}$ to be a subvector of $\mathbf{Z}$, that is, if a random effect is included in model (1) the corresponding fixed effect is also included. We assume that $\mathbf{b}_i \sim \mathrm{N}(0,\Sigma)$ i.i.d.. Let $N=\sum_{i=1}^n n_i$.

We can write $\boldsymbol{\beta}=(\beta_0,\boldsymbol{\beta}_1')'$, $\mathbf{b}_i=(b_{i0},\mathbf{b}_{i1}')'$, where β_0 and b_{i0} are the fixed and random intercepts, $\mathbf{Z}_{ij}=(1,\mathbf{z}_{ij}')'$ and $\mathbf{W}_{ij}=(1,\mathbf{w}_{ij}')'$. Under model (1) the amount of variation in the response Y that is not explained by the covariates is given by the residual variance, $\mathrm{var}(Y|\mathbf{z},\mathbf{b})=\mathrm{var}(\varepsilon)=\sigma^2$. In order to measure the proportion of explained variation, we need to define the corresponding 'null' model. There are two possible null models here. In our E1Z96 informed consent study, the investigators were interested in the predictive power of the patient characteristics, while the clustering by institutions was not of interest. In this type of application the corresponding 'null' hypothesis is that all the covariates have no regression effects:

$$H_0:\ Y_{ij}=\beta_0^*+b_{i0}^*+\varepsilon_{ij}^* \tag{2}$$

For ease of notation we assume that a fixed and a random intercept are always included in model (1), but if one of them is not included in model (1) then it is not included in the null model. In other applications the clustering variable itself might also be of interest, and the null model no longer includes the random intercept:

$$H_0:\ Y_{ij}=\beta_{00}^*+\varepsilon_{0ij}^* \tag{3}$$

Models (2) and (3) are in fact nested. Notice that when the random intercept is not included in model (1), model (2) becomes model (3). In the following we will devote the major part of our investigation to null model (2), for two reasons: it was motivated by our application; null model (3) and its related quantities are relatively straightforward.

We now elaborate the relationship between models (1) and (2). Denote $\boldsymbol{\mu}_z=E(\mathbf{z})$ and $\boldsymbol{\mu}_w=E(\mathbf{w})$. Although we do not impose the following assumptions, when the vector of random effects $\mathbf{b}$ is uncorrelated with the covariates $\mathbf{z}$ (and therefore $\mathbf{w}$), and when b_0 is uncorrelated with $\mathbf{b}_1$, we can show that $\beta_0^* = \beta_0 + \boldsymbol{\beta}_1' \boldsymbol{\mu}_z$, $b_{i0}^* = b_{i0} + \mathbf{b}_{i1}' \boldsymbol{\mu}_w$, $\varepsilon_{ij}^* = \boldsymbol{\beta}_1'(\mathbf{z}_{ij} - \boldsymbol{\mu}_z) + \mathbf{b}_{i1}'(\mathbf{w}_{ij} - \boldsymbol{\mu}_w) + \varepsilon_{ij}$, and that ε_{ij}^* is uncorrelated with b_{i0}^*. That is, model (2) as a mixed model with normal errors holds under model (1). Under model (2) the variation in Y given only the clustering is $\sigma_0^2 = \text{var}(Y|b_0^*) = \text{var}(\varepsilon^*)$. Furthermore

$$\sigma_0^2 = \text{var}\{\boldsymbol{\beta}_1'(\mathbf{z}_{ij} - \boldsymbol{\mu}_z) + \mathbf{b}_{i1}'(\mathbf{w}_{ij} - \boldsymbol{\mu}_w)\} + \sigma^2 \tag{4}$$

Therefore, given the clustering, the proportion of variation in Y explained by the covariates $\mathbf{z}$ is

$$\Omega^2 = 1 - \frac{\text{var}(Y|\mathbf{z}, \mathbf{b})}{\text{var}(Y|b_0^*)} = 1 - \frac{\sigma^2}{\sigma_0^2} \tag{5}$$

From (4) we see that under the above conditions, Ω^2 does not involve the distribution of b_0 since it is 'conditioned out' when we consider null model (2).

For null model (3) we can similarly define the proportion of explained variation. Here the total variation is $\sigma_{00}^2 = \text{var}(\varepsilon_0^*) = \text{var}(Y)$. We define

$$\Omega_0^2 = 1 - \frac{\text{var}(Y|\mathbf{z}, \mathbf{b})}{\text{var}(Y)} = 1 - \frac{\sigma^2}{\sigma_{00}^2} \tag{6}$$

In fact Ω^2 defined in (5) may be seen as a partial coefficient if we consider (3) to be the null model. See Section 6 for more discussion on partial coefficients.

2.2. The measures

Inference under models (1) and (2) is usually carried out via the (restricted) maximum likelihood and the EM algorithm [16]. To keep focus as well as parallel with the classic linear regression, in the following we assume that the maximum likelihood (ML) method is used. The restricted maximum likelihood method is discussed in Section 6. We can directly estimate Ω^2 defined in (5) by

$$r^2 = 1 - \frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} \tag{7}$$

where $\hat{\sigma}^2$ and $\hat{\sigma}_0^2$ are the estimates of σ^2 and σ_0^2, respectively.

We can also define another measure of explained variation in analogy to ordinary linear regression, using the residuals under models (1) and (2). This is a commonly used approach to extend the R^2 measure outside linear regression. Let $\eta_{ij} = \boldsymbol{\beta}'\mathbf{Z}_{ij} + \mathbf{b}_i'\mathbf{W}_{ij}$ be the linear predictor in model (1). Following the estimation of the unknown parameters $\boldsymbol{\beta}$, Σ and σ^2 by $\hat{\boldsymbol{\beta}}$, $\hat{\Sigma}$ and $\hat{\sigma}^2$, respectively, the predicted random effect $\hat{\mathbf{b}}_i$ can be obtained by an empirical Bayes (EB) estimator, the posterior mean or mode of the random effect given the data. Denote the residual under the fitted model (1) by

$$r_{ij} = Y_{ij} - \hat{\eta}_{ij} \tag{8}$$

where

$$\hat{\eta}_{ij} = \hat{\boldsymbol{\beta}}'\mathbf{Z}_{ij} + \hat{\mathbf{b}}_i'\mathbf{W}_{ij} \tag{9}$$

Similarly denote $\hat{\beta}_0^*$ and $\hat{b}_{i0}^*$ the (EB) estimate of β_0^* and b_{i0}^* under model (2). The residual under model (2) is given by

$$r_{ij}^* = Y_{ij} - \hat{\eta}_{ij}^* \tag{10}$$

where $\hat{\eta}_{ij}^* = \hat{\beta}_0^* + \hat{b}_{i0}^*$. An R^2 measure of explained variation, or predictive power, is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \sum_{j=1}^{n_i} r_{ij}^2}{\sum_{i=1}^{n} \sum_{j=1}^{n_i} r_{ij}^{*2}} = 1 - \frac{\text{RSS}}{\text{RSS}_0} \tag{11}$$

where RSS and RSS_0 are the residual sums of squares under models (1) and (2), respectively.

Under model (1) RSS$/N$ estimates the residual variance σ^2, and under model (2) RSS_0/N estimates the variance σ_0^2. Therefore R^2 defined in (11) estimates Ω^2. Notice, however, in order for R^2 to accurately estimate Ω^2, we need reasonable cluster sizes because the predicted random effects are used. We discuss this further in Sections 3 and 6. In the simulations of Section 4 we will see that, when the cluster sizes are small, R^2 can overestimate the underlying Ω^2. The amount of overestimation in the simulations is generally under 10 per cent, even if the cluster sizes are as small as five observations and the model includes up to three random effects terms.

Alternatively suppose we want to estimate Ω_0^2. Under model (3) we have $\hat{\sigma}_{00}^2 = \text{RSS}_{00} = \sum_{i,j} (Y_{ij} - \bar{Y})^2$, where $\bar{Y} = \sum_{i,j} Y_{ij}/N$. We can then define $r_0^2 = 1 - \hat{\sigma}^2/\hat{\sigma}_{00}^2$ and $R_0^2 = 1 - \text{RSS}/\text{RSS}_{00}$.

For the rest of this section we elaborate on five simple cases of the linear random effects model with at most one covariate, in order to better understand the measures of explained variation. In the following $\mathbf{Z} = (1, z)'$:

(i) *Linear regression* (*fixed effects only*)

$$Y_{ij} = \boldsymbol{\beta}'\mathbf{Z}_{ij} + \varepsilon_{ij} \tag{12}$$

In this case $\Omega^2 = \Omega_0^2 = 1 - \sigma^2/\text{var}(Y)$. All the sample-based measures are equal to the squared sample correlation coefficient between Y_{ij} and z_{ij}, and that between Y_{ij} and $\hat{Y}_{ij}$.

(ii) *One-way random effects*

$$Y_{ij} = \beta_0 + b_{i0} + \varepsilon_{ij} \tag{13}$$

In this case only Ω_0^2 is relevant. Since $\text{var}(Y) = \tau_0^2 + \sigma^2$, where $\tau_0^2 = \text{var}(b_0)$, we have $\Omega_0^2 = \tau_0^2/(\tau_0^2 + \sigma^2)$. That is, the proportion of variation explained by the clustering increases with the ratio of the between- versus the within-cluster variances.

(iii) *Random intercept* (*variance components model*)

$$Y_{ij} = b_{i0} + \boldsymbol{\beta}'\mathbf{Z}_{ij} + \varepsilon_{ij} \tag{14}$$

When b_0 is uncorrelated with z, $b_{i0}^* = b_{i0}$, and $\Omega^2 = 1 - \sigma^2/\{\beta_1^2\sigma_z^2 + \sigma^2\}$ where $\sigma_z^2 = \text{var}(z)$. Notice that the same formula here also holds for the linear regression (i) above. Furthermore, $\Omega_0^2 = 1 - \sigma^2/(\tau_0^2 + \beta_1^2\sigma_z^2 + \sigma^2)$.

(iv) *Random slope* (*with no random intercept*)

$$Y_{ij} = \boldsymbol{\beta}'\mathbf{Z}_{ij} + b_{i1}z_{ij} + \varepsilon_{ij} \tag{15}$$

In this case $\Omega^2 = \Omega_0^2 = 1 - \sigma^2/\text{var}(Y)$, since the null models (2) and (3) are the same.

(v) *Random slope and random intercept*

$$Y_{ij} = b_{i0} + \boldsymbol{\beta}'\mathbf{Z}_{ij} + b_{i1}z_{ij} + \varepsilon_{ij} \tag{16}$$

Following (4) it can be verified that $\Omega^2 = 1 - \sigma^2/\{(\beta_1^2 + \tau_1^2)\sigma_z^2 + \sigma^2\}$ under the uncorrelative conditions, where $\tau_1^2 = \text{var}(b_1)$. This shows that Ω^2 increases with $|\beta_1|$, τ_1 and σ_z. It is known in ordinary linear regression that the population equivalent of R^2 increases with the strength of the regression effect as well as the spread of the covariate. Here in addition it increases with the variance of the random slope. This is sensible since the random slope also reflects the importance of a covariate in predicting the response. Similarly one can verify that $\Omega_0^2 = 1 - \sigma^2/\{\tau_0^2 + (\beta_1^2 + \tau_1^2)\sigma_z^2 + \tau_1^2\mu_z^2 + \sigma^2\}$. Notice that the formulae for this last case apply to all the previous special cases.

From the above we see that the proportions of explained variation defined in this section for linear mixed effects models are very similar to the well-known R^2 for linear regression, except that they count for the additional random covariate effects and the additional clustering effects (in the case of Ω_0^2).

3. A CONNECTION TO EXPLAINED RANDOMNESS

The notion of explained randomness was first given in reference [17]. The concept was applied to linear, generalized linear, as well as proportional hazards models to develop R^2 type measures of correlation [4, 18]. Here, following the empirical Bayesian viewpoint as in Section 2, we define a measure of explained randomness using the conditional likelihood of the observed data given the predicted random effects. We show the close connection between such a measure and the measures defined in Section 2.

The randomness of a random variable Y is defined as a monotonic transformation of its entropy, $\exp\{-2I(\theta)\}$, where $I(\theta) = E\{\log\ p(y;\theta)\}$ is the expected log-likelihood, also called the Fraser information [19]. Under a regression setting as given by model (1), we define the residual randomness to be $D(Y|\mathbf{z},\mathbf{b}) = \exp[-2E\{\log\ p(y|\mathbf{z},\mathbf{b})\}]$, and under the null model (2) the total randomness of Y given only the clustering is $D(Y|b_0^*) = \exp[-2E\{\log\ p(y|b_0^*)\}]$. Here we omit in the notation the dependence of the above quantities on different parameter values under the full and the null model; the expectations, however, are always taken under the 'true', that is, the full, model. The proportion of explained randomness [17] is defined by

$$1 - \frac{D(Y|\mathbf{z},\mathbf{b})}{D(Y|b_0^*)} = 1 - \mathrm{e}^{-\Gamma} \tag{17}$$

where $\Gamma = 2\{I(\theta) - I(\theta_0)\}$ is twice the Kullback–Leibler [20] information gain. Given the data, the expectations in the above can often be replaced by sample means, thus (17) can be estimated by a simple transformation of the likelihood ratio statistic; see also below.

Denote by θ and θ_0 the vector of unknown parameters under models (1) and (2), respectively. Under model (1), the conditional likelihood of the observed data given the random effects is

$$L(\theta) = \prod_{i=1}^{n} \prod_{j=1}^{n_i} p(Y_{ij}|\mathbf{b}_i)$$

$$= \frac{1}{(\sqrt{2\pi}\sigma)^N} \exp\left\{ -\frac{\sum_{i=1}^{n} \sum_{j=1}^{n_i} (Y_{ij} - \boldsymbol{\beta}'\mathbf{Z}_{ij} - \mathbf{b}_i'\mathbf{W}_{ij})^2}{2\sigma^2} \right\} \quad (18)$$

Notice that given the random effects, the observations are independent. Using the predicted random effects, the likelihood ratio statistic of model (1) versus (2) can be seen to be

$$N\hat{\Gamma} = 2\log\{L(\hat{\theta})/L(\hat{\theta}_0)\} = N\log\left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}\right) - \frac{\text{RSS}}{\hat{\sigma}^2} + \frac{\text{RSS}_0}{\hat{\sigma}_0^2} \quad (19)$$

We define the measure of explained randomness

$$\rho^2 = 1 - \mathrm{e}^{-\hat{\Gamma}} = 1 - \frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} \exp\left(\frac{\text{RSS}}{N\hat{\sigma}^2} - \frac{\text{RSS}_0}{N\hat{\sigma}_0^2}\right) \quad (20)$$

Note that when there are no random effects, the ρ^2 measure was given in reference [17], and is equal to the R^2 for linear regression.

Since RSS/N estimates σ^2, and RSS_0/N estimates σ_0^2, we see that ρ^2 should be close to r^2 and R^2 defined in the previous section. In fact, using the concept of effective degrees of freedom given by Hodges and Sargent [21], in an unpublished work Vaida and Blanchard verified that $\hat{\sigma}^2$ is approximately equal to RSS divided by the residual degrees of freedom of reference [21] under model (1). If we write $\hat{\sigma}^2 \approx \text{RSS}/(N - \text{df})$, and $\hat{\sigma}_0^2 \approx \text{RSS}_0/(N - \text{df}_0)$, using a first-order Taylor approximation to the exponential function in (20), we can show that $\rho^2 \approx 1 - \text{RSS}/\text{RSS}_0 = R^2$. Here we see that a difference between r^2 and the other two measures is that r^2 takes into account the different degrees of freedom under the full model and under the null model, while the other two measures do not. In this way r^2 corresponds to the adjusted R^2 on page 92 of Draper and Smith [22]. Furthermore, the Hodges and Sargent model degrees of freedom increases with the number of clusters. Therefore when the number of clusters are large compared to the overall sample size, we expect discrepancies between r^2 and the other two measures. This is confirmed in our simulation as well as the data examples below. See also Section 6 for more discussion.

The coefficient ρ_0^2 using null model (3) can be similarly defined as in (19) and (20). Here we have $\rho_0^2 = 1 - \exp(-\hat{\Gamma}_0)$, where $\hat{\Gamma}_0 = \log(\hat{\sigma}_{00}^2/\hat{\sigma}^2) - \text{RSS}/N\hat{\sigma}^2 + 1$. The relationship among r^2, R^2 and ρ^2 as described above also carries over to r_0^2, R_0^2 and ρ_0^2.

4. SIMULATION STUDIES

In this section we show some results of our simulation experiments to study the performance of the measures given in this paper. Although more simulations have been carried out, here

Table I. A Monte Carlo study of R^2, ρ^2, r^2 and Ω^2 (single covariate).

$\boldsymbol{\beta}_1$	τ_0	τ_1	$n=50$, $n_i=5$ R^2	ρ^2	r^2	$n=10$, $n_i=25$ R^2	ρ^2	r^2	Ω^2
0.7	0	0	0.49 (0.05)	0.49 (0.05)	0.49 (0.05)	0.50 (0.04)	0.50 (0.04)	0.50 (0.04)	0.49
	0	0.5	0.65 (0.06)	0.64 (0.06)	0.60 (0.06)	0.58 (0.09)	0.58 (0.09)	0.57 (0.10)	0.59
	0	1	0.78 (0.05)	0.77 (0.05)	0.73 (0.05)	0.71 (0.10)	0.71 (0.10)	0.70 (0.10)	0.75
	1	0	0.52 (0.05)	0.52 (0.05)	0.51 (0.05)	0.49 (0.04)	0.49 (0.04)	0.49 (0.04)	0.49
	1	0.5	0.64 (0.07)	0.63 (0.07)	0.58 (0.07)	0.60 (0.08)	0.60 (0.08)	0.59 (0.08)	0.59
	1	1	0.80 (0.04)	0.79 (0.05)	0.74 (0.05)	0.72 (0.11)	0.72 (0.11)	0.71 (0.11)	0.75
0.5	1	0	0.36 (0.06)	0.36 (0.06)	0.34 (0.06)	0.34 (0.05)	0.34 (0.05)	0.34 (0.05)	0.33
	1	0.5	0.55 (0.08)	0.53 (0.07)	0.47 (0.07)	0.50 (0.11)	0.50 (0.11)	0.48 (0.12)	0.50
	1	1	0.76 (0.05)	0.74 (0.06)	0.69 (0.06)	0.72 (0.09)	0.72 (0.09)	0.71 (0.09)	0.71

The measures are defined in (7), (11), (20) and (5), when the clustering variable is not of predictive interest.

we only show a proportion of the results which are sufficient to demonstrate the behaviour of the measures. We simulated data sets under model (1) for both a single covariate and two covariates. Data were generated with $\beta_0=0$ without loss of generality; in fitting the models a fixed effect intercept term was always included. We used the S-plus function lme( ) written by Pinheiro and Bates [23] to fit the random effects models. For each combination of the parameters, 100 simulations were carried out, and in the ( )'s are standard errors from the 100 simulation runs. We generated data with two different sample sizes: 10 clusters with 25 observations in each cluster, and 50 clusters with 5 observations in each cluster.

4.1. Single covariate

We first generated data according to $z\sim \mathrm{N}(0,1)$, $\beta_1=0.7$ and $\sigma^2=0.51$ under model (1). When there are no random effects, this corresponds to bivariate normal with a correlation coefficient of 0.7 and standard normal marginals, and the R^2 from ordinary linear regression estimates the squared correlation coefficient of 0.49. We then added independent random intercept $b_0\sim \mathrm{N}(0,\tau_0^2)$, independent random slope $b_1\sim \mathrm{N}(0,\tau_1^2)$, and also considered different values of β_1. The results on Ω^2 and the corresponding sample-based measures are shown in Table I.

From Table I we see that the measures can obtain a range of values between 0 and 1 for realistic strengths of the fixed and random regression effects. They are basically unaffected by the variance τ_0 of the random intercept, but increase with the variance τ_1 of the random slope. We have analytically shown these two properties in Section 2. The r^2 measure gives good estimates of Ω^2 in all cases. We also see that the values of R^2 and ρ^2 are very close for all cases. For the sample size of 10 clusters with 25 observations in each cluster, both R^2 and ρ^2 are very close to Ω^2. For the other sample size of 50 clusters with 5 observations in each cluster, there is slight overestimation in some cases, but the amount of overestimation does not exceed 0.05 for all cases.

We also computed the measures corresponding to Ω_0^2 under simulated settings; the results are shown in Table II. The first row of the table is under the one-way random effects model. Overall from the table we see that r_0^2 gives good estimates of Ω_0^2. The values of R_0^2 and ρ_0^2

Table II. A Monte Carlo study of R_0^2, ρ_0^2, r_0^2 and Ω_0^2 (single covariate).

β_1	τ_0	τ_1	$n=50,\ n_i=5$ R_0^2	ρ_0^2	r_0^2	$n=10,\ n_i=25$ R_0^2	ρ_0^2	r_0^2	Ω_0^2
0	0.5	0	0.45 (0.07)	0.44 (0.07)	0.35 (0.07)	0.33 (0.10)	0.34 (0.10)	0.31 (0.10)	0.33
	0.5	0.5	0.60 (0.07)	0.59 (0.07)	0.48 (0.07)	0.50 (0.08)	0.50 (0.08)	0.47 (0.08)	0.50
	1	0.5	0.78 (0.05)	0.77 (0.05)	0.70 (0.06)	0.70 (0.09)	0.70 (0.09)	0.67 (0.10)	0.71
0.5	1	0	0.77 (0.03)	0.76 (0.03)	0.71 (0.04)	0.68 (0.08)	0.68 (0.08)	0.67 (0.08)	0.71
	1	0.5	0.82 (0.03)	0.81 (0.03)	0.74 (0.04)	0.73 (0.08)	0.73 (0.08)	0.71 (0.09)	0.75
	1	1	0.87 (0.03)	0.86 (0.03)	0.81 (0.04)	0.81 (0.06)	0.81 (0.06)	0.80 (0.06)	0.82

The measures are similar to R^2, ρ^2, r^2 and Ω^2, when the clustering itself is also of predictive interest.

Table III. A Monte Carlo study of R^2, ρ^2, r^2 and Ω^2 (two covariates).

β_1	β_2	τ_1	τ_2	$n=50,\ n_i=5$ R^2	ρ^2	r^2	$n=10,\ n_i=25$ R^2	ρ^2	r^2	Ω^2
2	2	2	2	0.86 (0.03)	0.79 (0.04)	0.85 (0.03)	0.81 (0.05)	0.79 (0.06)	0.81 (0.05)	0.80
2	1	1	1	0.71 (0.05)	0.64 (0.05)	0.70 (0.05)	0.66 (0.06)	0.64 (0.06)	0.66 (0.06)	0.64
1	1	1	1	0.58 (0.07)	0.57 (0.07)	0.49 (0.07)	0.52 (0.08)	0.52 (0.08)	0.49 (0.08)	0.50
0.5	1	1	1	0.54 (0.08)	0.52 (0.08)	0.44 (0.08)	0.46 (0.10)	0.46 (0.10)	0.43 (0.10)	0.45
1	1	0.5	1	0.53 (0.08)	0.52 (0.08)	0.45 (0.07)	0.46 (0.08)	0.46 (0.08)	0.43 (0.08)	0.45
0.5	0.5	1	1	0.48 (0.10)	0.46 (0.09)	0.37 (0.08)	0.41 (0.09)	0.40 (0.09)	0.37 (0.09)	0.38
1	1	0.5	0.5	0.44 (0.08)	0.44 (0.07)	0.39 (0.07)	0.40 (0.07)	0.40 (0.07)	0.38 (0.07)	0.38
0.5	0.5	0.5	0.5	0.29 (0.09)	0.28 (0.09)	0.22 (0.07)	0.23 (0.08)	0.23 (0.08)	0.21 (0.07)	0.20
0.5	0.5	0	0.5	0.24 (0.07)	0.23 (0.07)	0.19 (0.06)	0.18 (0.07)	0.18 (0.06)	0.17 (0.06)	0.15

The measures are defined in (7), (11), (20) and (5), when the clustering variable is not of predictive interest.

are again very close, and they are close to Ω_0^2 for the case of 10 clusters with 25 observations in each cluster, but for the case of 50 clusters with 5 observations in each cluster, they overestimate Ω_0^2, and the amount of overestimation is greater than what is seen in Table I. This is because there is now an even larger difference in the degrees of freedom between the full and the null model. We can also compare the lower half of Table II to the last three rows of Table I, since the data are generated under the same models. In this case Ω^2 can be seen as a partial coefficient of Ω_0^2 (see Section 2.1), and we can easily compute the proportion of variation explained by clustering alone, which is $1-\sigma_0^2/\sigma_{00}^2 = 1-(1-\Omega_0^2)/(1-\Omega^2)$, to be 0.57, 0.50 and 0.38 for the three rows, respectively.

4.2. Two covariates

We also simulated data under model (1) with two independent covariates, both distributed as standard normal. We allowed different strengths of fixed regression effects β_1 and β_2, and different variances for the corresponding random effects, τ_1^2 and τ_2^2. Although not included in Table III, the variance of the random intercept again did not affect the values of the measures corresponding to Ω^2. For the results of Table III, $\tau_0^2 = \mathrm{var}(b_0) = 1$ and $\sigma^2 = \mathrm{var}(\varepsilon) = 4$.

Table III demonstrates once again that different values of the measures can be obtained as functions of different fixed and random regression effects. The model now contains three random effects terms. The r^2 measure is again very close to Ω^2 in all cases. The R^2 and ρ^2 values are also very close to each other. These two measures from the 10 clusters with 25 observations each have good agreement with Ω^2. For the 50 clusters with 5 observations each, R^2 and ρ^2 have overestimation of up to 10 per cent, with R^2 having a slightly larger amount of overestimation than ρ^2.

In summarizing the simulation results, the measures defined in this paper quantify the predictability of the variables as given by their fixed and random regression effects. For both small and large cluster sizes the r^2 and r_0^2 measures appear to give accurate estimates of the population Ω^2 and Ω_0^2, respectively. The R^2 and R_0^2, and ρ^2 and ρ_0^2 measures give good estimates with reasonably large cluster sizes, but overestimate Ω^2 and Ω_0^2 if the cluster sizes are too small, with the overestimation being more severe for R_0^2 and ρ_0^2.

5. EXAMPLES

5.1. E1Z96 consent trial

We now return to the E1Z96 informed consent study. The study enrolled 226 patients from 44 institutions. All of the outcomes considered here, including comprehension of the consent documentation, consent anxiety, consent satisfaction and decisional conflict, are continuous scores, obtained from averaging a number of individual item scores for a particular outcome. For example, comprehension is measured by the percentage correct in a patient's answer to 23 true-or-false and multiple-choice questions. More details are given in reference [6]. Preliminary analysis revealed no major departure of the outcome scores from the normality assumption. The patient characteristics collected in the study included the 'parent' study a patient was considering participating in (lung or breast cancer clinical trials), the Rapid Estimate of Adult Literacy in Medicine (REALM) score, education, age, race, sex, institution type, and the Monitor-Blunter Style Scale scores reflecting a patient's coping style. As mentioned in Section 1, beyond the primary comparisons of the outcomes between the two randomized arms, it is of interest to understand how the endpoints are affected by the patient characteristics; in particular, how much variation in the outcomes is explained by these covariates. In the following we will use the measures developed in this paper to quantify the amount of variation explained, without carrying out a comprehensive analysis of the whole data set. To be consistent with the original manuscript [6], all models will include fixed effects for the covariates and a random intercept for institutional effect on the outcomes. All the multiple regression models below were reported in reference [6].

For the two primary endpoints of the study, comprehension and consent anxiety, there was a significant difference in the scores between the randomized arms for consent anxiety (p-value $= 0.016$), but not for comprehension (p-value $= 0.21$). Both outcomes indicated the easy-to-read version to be advantageous over the standard version of the consent documents, that is, improved comprehension and lowered anxiety. Greater comprehension was also found to be associated with younger age, higher REALM score, higher level of education and lower monitoring, 'although coefficients are small' [6]. Using these four covariates, we have $R^2 = \rho^2 = 0.17$ and $r^2 = 0.19$; that is, conditional on the cluster randomization design, slightly under 20 per cent of the variation in the comprehension scores is explained by these four covariates. Although not of main interest in this study itself, we can also estimate Ω_0^2 which quantifies the proportion of variation explained by the four covariates plus clustering. We have $R_0^2 = \rho_0^2 = 0.25$ and $r_0^2 = 0.24$. This also shows that the proportion of variation in the comprehension scores explained by the clustering alone is approximately $1-(1-0.25)/(1-0.17) = 10$ per cent. As another illustration, when all the patient characteristics collected in the study are included in a mixed model, Ω^2 is estimated to be 0.23 by both R^2 and ρ^2, and 0.25 by r^2. This shows that the additional four patient characteristics would only increase the proportion of explained variation by about 5 per cent. The other primary endpoint, consent anxiety, was found to be associated with REALM score, education and monitoring score, in addition to the intervention. The proportion of variation explained by these four variables is estimated to be 10 per cent by R^2, ρ^2 and r^2. Note that for this data set only the random intercept is included in all the models, and the values of r^2 are close to that of R^2 and ρ^2.

Two other interesting endpoints are consent satisfaction and decisional conflict. The consent satisfaction score was very highly significantly (p-value $= 0.004$) associated with intervention, with the easy-to-read arm having higher satisfaction of an estimated 0.35 points (out of the total range of 1.0–4.0) over the standard arm. None the less, the estimated Ω^2 using intervention assignment is almost zero (< 0.01) by all three measures. This shows a phenomenon that we have encountered before in both linear and proportional hazards regression, where a covariate can be highly significant for the response, but yet has a very low R^2 value. Such a phenomenon is possible partially because 'everything becomes significant' when the sample size is large enough, while Ω^2 is a population quantity that does not depend on the sample size. It further demonstrates the usefulness of R^2 type measures, in that a highly significant covariate may not have high predictability. No other variables were found to be associated with the satisfaction score. The only variable that was significantly associated with the decisional conflict score was parent study, with lung cancer patients demonstrating greater decisional conflict than breast cancer patients. Both R^2 and ρ^2, in this case, turn out to have a quite high value of 0.65, and $r^2 = 0.66$; that is, a large proportion of variation in decisional conflict is explained by parent study alone.

5.2. Schizophrenia data

Next we illustrate the use of the measures in model selection. There are many methods for model selection in general, R^2 being a simple and often conveniently available one. Our example is a Canadian clinical trial for the treatment of schizophrenia [24]. Schizophrenia is a type of psychotic disorder, and the clinical trial was double-blinded with randomization among four treatment arms: low, medium and high dose of an experimental drug, and a control drug with known antipsychotic effects as well as known side-effects. The trial was conducted at 13 clinical centres, and the primary outcome was assessed using the Brief Psychiatric Rating Scale (BPRS) at baseline, 1, 2, 3, 4 and 6 weeks of treatment. Some patients had early discontinuation of treatment due to a perceived lack of effectiveness of treatment by their physicians. The data set was used as a case study in a tutorial on the general linear mixed model by Cnaan *et al.* [25]. Here, following the discussion in reference [25], we take another look at the model selection problem considered in the tutorial. The six models considered by Cnaan et al. [25] are given in Table IV. Since the models all contain the same random

Table IV. Schizophrenia data.

Model number	Covariates	R^2	ρ^2	r^2
1	Week, week^2	0.58	0.55	0.48
2	Treatment, week, week^2	0.58	0.55	0.48
3	Treatment, week, week^2, week $\times$ treatment	0.58	0.55	0.48
4	Treatment, week indicators	0.61	0.58	0.51
5	Treatment, week, week^2, week $\times$ centre	0.57	0.55	0.48
6	Treatment, week, week^2, status	0.64	0.62	0.56

All models include baseline BPRS and center as covariates, a random intercept, and random linear and quadratic effects of time.

effects terms, we will use the estimates of Ω^2. As an alternative one could also use those of Ω_0^2.

There were 233 patients in total. The response variables in this case consisted of all the evaluations done after the start of the treatment, while the baseline BPRS score was included as a covariate. Status in model 6 was defined to be 0 while a patient was on study and 1 at the last observation if the patient was discontinued due to lack of therapeutic effect, therefore the last observation reflected a status of being off-study. In model 4 the week numbers were only entered as indicators, reflecting no assumptions on the time trend. In all the other models, in modelling the time trend 3 weeks were subtracted from the week number according to reference [25]; this way the data for the linear and quadratic effects were nearly orthogonal to each other for subjects with all six measurements. The random effects in all the models consisted of a random intercept and random linear and quadratic effects of time. As the purpose of the study was to show that the experimental treatment was as efficacious as the active control, it was perhaps not surprising that the overall treatment effect was not significant in any of the models, but following reference [25] we kept the treatment variable in models 2–6. All of the other variables, except for week by treatment or week by centre interactions which are discussed below, were significant at 0.001 level.

Models 1–3 give the same value for each of the three measures. In particular, the additional week by treatment interaction in model 3 turns out not to be significant (p-value $= 0.26$). Since there are at most five observations in a cluster, and there are three random effects terms, it is consistent with the earlier simulation results that the R^2 values can be up to 10 per cent higher than r^2. Model 4 makes no assumption on the time trend, and has all the time indicators significant, but the values of all three measures are only increased by 0.03 at the cost of two more parameters as compared to model 2. Model 5, despite having the largest number of parameters of all the models, does not show any improvement in explaining the variation in the response. The overall week by centre interaction also turns out not to be significant (p-value $= 0.25$). Finally, model 6, with only one additional covariate than model 2, has the highest values for all the three measures, and is therefore the preferred model. All of the above agree with the findings of reference [25] where likelihood comparison was used for model selection.

This example illustrated the use of the measures in model selection, including non-nested models. Notice that although R^2 and ρ^2 appear to overestimate the population quantities, in this case, because the random effect terms are the same in all the models, they consistently reflect the relative predictability of the models for selection purposes.

6. DISCUSSION

Random effects models are more complex than the fixed-effects-only models. We have focused on the Laird–Ware type models for clustered data. We have also put more emphasis on Ω^2 and its sample-based measures, which our application has called upon. There are no doubt other applications where the clustering itself is of primary interest, in which case Ω_0^2 can be used. Our model notation does not include situations like hierarchical clustering, but the same ideas apply and the generalization is straightforward.

For comparing nested models it is useful to quantify the 'extra amount of variation explained' by using a partial coefficient. In fact, as mentioned earlier, Ω^2 is a partial coefficient of Ω_0^2. For all the measures in this paper, it is straightforward to define partial coefficients in general. More specifically, each of the measures can be written in the form of $1 - v/v_0$. For comparing two nested models, suppose a measure is calculated to be $1 - v_1/v_0$ and $1 - v_2/v_0$ under these two models where '2' indicates the larger model, the corresponding partial coefficient can then be defined as $1 - v_2/v_1$.

We have aimed to develop measures which quantify the amount of variation in a response variable that is explained by the covariates. In practice R^2 measures are also used by some to assess the fit of a model. In a recent book [26] on linear mixed models, they described such measures for overall fit based on residuals.

The restricted maximum likelihood (REML) estimates are sometimes used under model (1). In our simulation studies (not shown), the REML and ML estimates give similar results. This is perhaps not surprising as in our settings the numbers of fixed effects are not very large compared to the total sample size. For the more extreme cases where the REML and the ML estimates differ greatly, the resulting measures might be different. This deserves further study on its own and is beyond the scope of this paper. We note that in ordinary linear regression, if REML estimates are used, the r^2 given in (7) is the adjusted R^2 defined on page 92 of reference [22].

Asymptotic theory under linear mixed models is generally difficult and there are only relatively few results in the literature, including references [27] and [28] on the ML and reference [29] on the REML estimator. A few others studied the same estimators without the normality assumption, see, for example, references [30–33]. In particular, references [32, 33] showed that in order for the residual methods using the predicted random effects to work under the mixed model (via the condition that the empirical distribution of the estimated random effects converges to the true distribution of the random effects), we need $n_i \to \infty$. This is consistent with our requirement of reasonable cluster sizes for R^2 and ρ^2 to accurately estimate Ω^2. Following the discussion in Section 3, we may define an adjusted R^2 under model (1), using the residual sums of squares and the definition of degrees of freedom in Hodges and Sargent [21]. Notice, however, that the degrees of freedom do not have closed-form expressions in general, and a program is required for their calculation. On the other hand, the r^2 measure approximates the adjusted R^2 and is easily available after model fitting. The same also applies to the estimates of Ω_0^2.

Among many equivalent definitions of R^2 in linear regression, one is the squared correlation coefficient between the observed and the predicted response. In our simulations under model (1) (results not shown here), such a measure appears to estimate Ω_0^2, but not Ω^2. For example, for the settings given in Table I, the measure would be greater than 0.8 for most cases. Its estimation of Ω_0^2 is good when the cluster sizes are large; when the cluster sizes are small, it has a larger amount of overestimation than R_0^2 and ρ_0^2.

Finally one might attempt to use the marginal instead of the conditional likelihood of the observed data to define a measure of the explained randomness. The challenge here is to find the 'effective' sample size in order to quantify the information gain per observation. In the extreme cases, when the data are independent, the effective sample size is $N=\sum n_i$; when the data from the same cluster are perfectly correlated, the effective sample size is n. Otherwise it is also possible to compute the effective sample size in simple cases like balanced one-way random effects models, but it is not clear how to do so for the more general cases. In simulation studies (not shown here), the measures that we proposed in this paper fall between the ones using the marginal observed data likelihood divided by the two extreme sample sizes N and n.

7. CONCLUSION

In this paper we investigated three types of measures to estimate the proportions of explained variation under the linear mixed effects model. The measures extend the well-known R^2 for linear regression. They showed overall desirable performance in simulation studies. The data examples illustrated their usefulness in practice.

ACKNOWLEDGEMENTS

The author would like to thank Dr Peter Raich for permission to use the E1Z96 data and Dr Nan Laird for the schizophrenia data.

REFERENCES

1. Cox DR, Snell EJ. *Analysis of Binary Data*. 2nd edn. Chapman and Hall: London, 1989.
2. Cox DR, Wermuth N. A comment on the coefficient of determination for binary responses. *American Statistician* 1992; **46**:1–4.
3. Schemper M, Stare J. Explained variation in survival analysis. *Statistics in Medicine* 1996; **15**:1999–2012.
4. Xu R, O'Quigley J. A R^2 type of measure of dependence for proportional hazards models. *Nonparametric Statistics* 1999; **12**:83–107.
5. O'Quigley J, Xu R. Explained variation in proportional hazards regression. In *Handbook of Statistics in Clinical Oncology*. Marcel Dekker: New York, 2001.
6. Coyne C, Xu R, Raich P, Plomer K, Dignan M, Wenzel L, Fairclough D, Marcus A, Habermann T, Schnell L, Quella S, Cella D. A randomized controlled trial of an easy-to-read informed consent statement for clinical trial participation. *Journal of Clinical Oncology* 2003; **21**(5):836–842.
7. Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research*. Arnold: London, 2000.
8. Korn EL, Simon R. Explained residual variation, explained risk, and goodness of fit. *American Statistician* 1991; **45**:201–206.
9. Xu R. Inference for the proportional hazards model. PhD thesis, University of California, San Diego, 1996.
10. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**:963–974.
11. Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 1977; **72**:320–338.

12. Morris CN. Parametric Empirical Bayes inference: theory and applications (with discussion). *Journal of the American Statistical Association* 1983; **78**:47–65.
13. Henderson CR. Sire evaluation and genetic trends. In *Proceedings of the Animal Breeding and Genetics Symposium in Honor of Dr. Jay L. Lush.* American Society of Animal Science: Champaign, Illinois, 1973: 10–41.
14. Harris DL, Lofgren DL, Stewart TS, Schinckel AP. Adapting best linear unbiased prediction (BLUP) for timely genetic evaluation: I. Preogeny traits in a single contemporary group for each sex. *Journal of Animal Science* 1989; **67**:3209–3222.
15. Ghosh M, Rao JNK. Small area estimation: an appraisal. *Statistical Science* 1994; **9**:55–93.
16. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society*, *Series B* 1977; **39**:1–38.
17. Kent JT. Information gain and a general measure of correlation. *Biometrika* 1983; **70**(1):163–174.
18. Kent JT, O'Quigley J. Measure of dependence for censored survival data. *Biometrika* 1988; **75**:525–534.
19. Fraser DAS. On information in statistics. *Annals of Mathematical Statistics* 1965; **36**:890–896.
20. Kullback S, Leibler RA. On information and sufficiency. *Annals of Mathematical Statistics* 1951; **22**:79–86.
21. Hodges JS, Sargent DJ. Counting degrees of freedom in hierarchical and other richly-parametrized models. *Biometrika* 2001; **88**:367–379.
22. Draper N, Smith H. *Applied Regression*. 2nd edn. Wiley: New York, 1981.
23. Pinheiro JC, Bates DM. *Mixed-Effects Models in S and S-PLUS*. Springer: New York, 2000.
24. Lapierre YD, Nai NV, Chauinard G, Awad AG, Saxena B, James B, McClure DJ, Bakish D, Max P, Manchanda R, Beaudry P, Bloom D, Rostein E, Ancill R, Sandor P, Sladen-Dew N, Durand C, Chandrasena R, Horn E, Elliot D, Das M, Ravindra A, Matsos G. A controlled dose-ranging study of remoxipride and haloperidol in schizophrenia—a Canadian multicenter trial. *Acta Psychiatric Scandinavica* 1990; **82**:72–76.
25. Cnaan A, Laird NM, Slasor P. Using the general linear mixed model to analyze unbalanced repeated measures and longitudinal data. *Statistics in Medicine* 1997; **16**:2349–2380.
26. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. Springer: New York, 2000.
27. Hartley HO, Rao JNK. Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* 1967; **54**:93–108.
28. Miller JJ. Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Annals of Statistics* 1977; **5**:746–762.
29. Cressie N, Lahiri SN. The asymptotic distribution of REML estimators. *Journal of Multivariate Analysis* 1993; **45**:217–233.
30. Richardson AM, Welsh AH. Asymptotic properties of restricted maximum likelihood (REML) estimates for hierarchical mixed linear models. *Australian Journal of Statistics* 1994; **36**:31–43.
31. Jiang J. REML estimation: asymptotic behavior and related topics. *Annals of Statistics* 1996; **24**:255–286.
32. Jiang J. Asymptotic properties of the empirical BLUP and BLUE in mixed linear models. *Statistica Sinica* 1998; **8**:861–885.
33. Jiang J. Goodness-of-fit tests for mixed model diagnostics. *Annals of Statistics* 2001; **29**:1137–1164.

