

# A Linear Mixed-Effects Model With Heterogeneity in the Random-Effects Population

Geert VERBEKE and Emmanuel LESAFFRE

This article investigates the impact of the normality assumption for random effects on their estimates in the linear mixed-effects model. It shows that if the distribution of random effects is a finite mixture of normal distributions, then the random effects may be badly estimated if normality is assumed, and the current methods for inspecting the appropriateness of the model assumptions are not sound. Further, it is argued that a better way to detect the components of the mixture is to build this assumption in the model and then "compare" the fitted model with the Gaussian model. All of this is illustrated on two practical examples.

KEY WORDS: Empirical Bayes; Goodness-of-fit test; Longitudinal model; Mixture model; Normality assumption.

## 1. INTRODUCTION

A frequently used model for describing longitudinal continuous data is the linear mixed-effects model where random effects serve to model the between-individual correlation structure. Often, the normality assumption for the random effects and the error structure is almost automatically taken for granted, and little attention has been devoted to the question of what impact this assumption has on the estimation of the different parts of the model. Although Butler and Louis (1992) have recently shown that the normality assumption has little effect on the fixed-effects estimates, this assumption's effect on the random-effects estimates has not yet been investigated. Further, tools seem to be lacking to verify this assumption.

In this article we concentrate on the detection of a mixture in the distribution of the random effects. This is of importance in (linear) longitudinal models where the systematic part has been misspecified due to the omission of a categorical variable. For instance, studies on the evolution of the blood pressure of patients treated with an antihypertensive drug often report "responders" and "nonresponders" to medication.

After defining the usual linear mixed-effects model in Section 2, we show in Section 3 that the normality assumption for the random effects can seriously influence their estimates and is very difficult to check. We thus extend the model to the so-called "heterogeneity model" in Section 4. In Section 5 we use two medical data sets to illustrate how this model can yield additional insight in longitudinal data.

## 2. THE HOMOGENEITY MODEL

In the linear mixed-effects model, it is assumed that the $n_i \times 1$ vector $\mathbf{y}_i$ of responses for the $i$th individual can be modeled as $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\alpha} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i, i = 1, \ldots, N$, with $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D}^*)$ and $\mathbf{e}_i \sim N(\mathbf{0}, \sigma^2\mathbf{I}_{n_i})$. $\mathbf{X}_i$ and $\mathbf{Z}_i$ are $n_i \times p$ and $n_i \times q$ full-rank covariate matrices, $\boldsymbol{\alpha}$ is a $p \times 1$ vector of unknown parameters describing the population mean, and $\mathbf{b}_i$ is a vector of subject-specific regression coefficients, assumed to be independently distributed from the error terms $\mathbf{e}_i$. Marginally, the vector $\mathbf{y}_i$ is normally distributed with mean $\mathbf{X}_i\boldsymbol{\alpha}$ and covariance matrix $\text{var}(\mathbf{y}_i) = \mathbf{Z}_i\mathbf{D}^*\mathbf{Z}_i' + \sigma^2\mathbf{I} = \mathbf{V}_i^* = \mathbf{W}_i^{*-1}$. The empirical Bayes estimate for $\mathbf{b}_i$ equals $\hat{\mathbf{b}}_i = \mathbf{D}^*\mathbf{Z}_i'\mathbf{W}_i^*(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\alpha})$, and is calculated replacing all parameters by their maximum likelihood or restricted maximum likelihood estimates. Because this model can be seen as a hierarchical Bayes model where, given $\boldsymbol{\mu}, \mathbf{b}_i \sim N(\boldsymbol{\mu}, \mathbf{D}^*), \boldsymbol{\mu}$ equals zero with probability 1, this model is called the *homogeneity model.* The effect of this homogeneity assumption is investigated in the next section. Due to Butler and Louis' result (1992), it's assumed that the fixed effects $\boldsymbol{\alpha}$ are known.

## 3. THE IMPACT OF THE NORMALITY ASSUMPTION ON THE ESTIMATES OF THE RANDOM EFFECTS

Suppose that the $\mathbf{b}_i$ are distributed according to the mixture $pN(\boldsymbol{\mu}_1, \mathbf{D}) + (1-p)N(\boldsymbol{\mu}_2, \mathbf{D})$. Despite this deviation from normality, one could still estimate $\mathbf{b}_i$ by $\hat{\mathbf{b}}_i = \mathbf{D}^*\mathbf{Z}_i'\mathbf{W}_i^*(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\alpha})$, in which $\mathbf{D}^*$ now equals $\text{var}(\mathbf{b}_i) = \mathbf{D} + p\boldsymbol{\mu}_1\boldsymbol{\mu}_1' + (1-p)\boldsymbol{\mu}_2\boldsymbol{\mu}_2'$. The vector $\hat{\mathbf{b}}_i$ then follows a mixture of two normals, with proportions $p$ and $(1-p)$, and with mean and covariance depending on the covariates $\mathbf{Z}_i$. Using results from Titterington, Smith, and Makov (1985, sec. 5.5), it can be shown that this new mixture is unimodal if the eigenvalues of $\sigma^2(\mathbf{Z}_i'\mathbf{Z}_i)^{-1}$ are sufficiently large, independently of the modality of the original mixture for $\mathbf{b}_i$. This is also in agreement with Strenio, Weisberg, and Bryk (1983), who showed that the $\hat{\mathbf{b}}_i$ are shrunk toward the population mean, and that this shrinkage is more severe in cases where measurements are not very precise (large $\sigma^2$).

The consequences of this can be best illustrated with a small-scale simulation study. Random effects $\mathbf{b}_i$ were first simulated from the mixture $1/2N(-2, 1) + 1/2N(2, 1)$ and were then used to simulate 1,000 profiles with 5 repeated measurements each. This was repeated for the $\mathbf{b}_i$ representing intercepts ($\mathbf{Z}_i = (1\,1\,1\,1\,1)'$) as well as slopes

Geert Verbeke is Research Assistant and Emmanuel Lesaffre is Department Head, Department of Biostatistics and Epidemiology, Biostatistical Centre for Clinical Trials, Catholic University of Leuven, B-3000 Leuven, Belgium. The authors are grateful to Kung-Yee Liang and Larry Brant for suggesting the problem, to Dankmar Böhning for helpful comments on the mixture problems, and to the associate editor and the referees whose comments improved the paper considerably. They also wish to thank Larry Brant and Jay Pearson, who provided the prostate cancer data from the Baltimore Longitudinal Study of Aging.


Journal of the American Statistical Association
March 1996, Vol. 91, No. 433, Theory and Methods

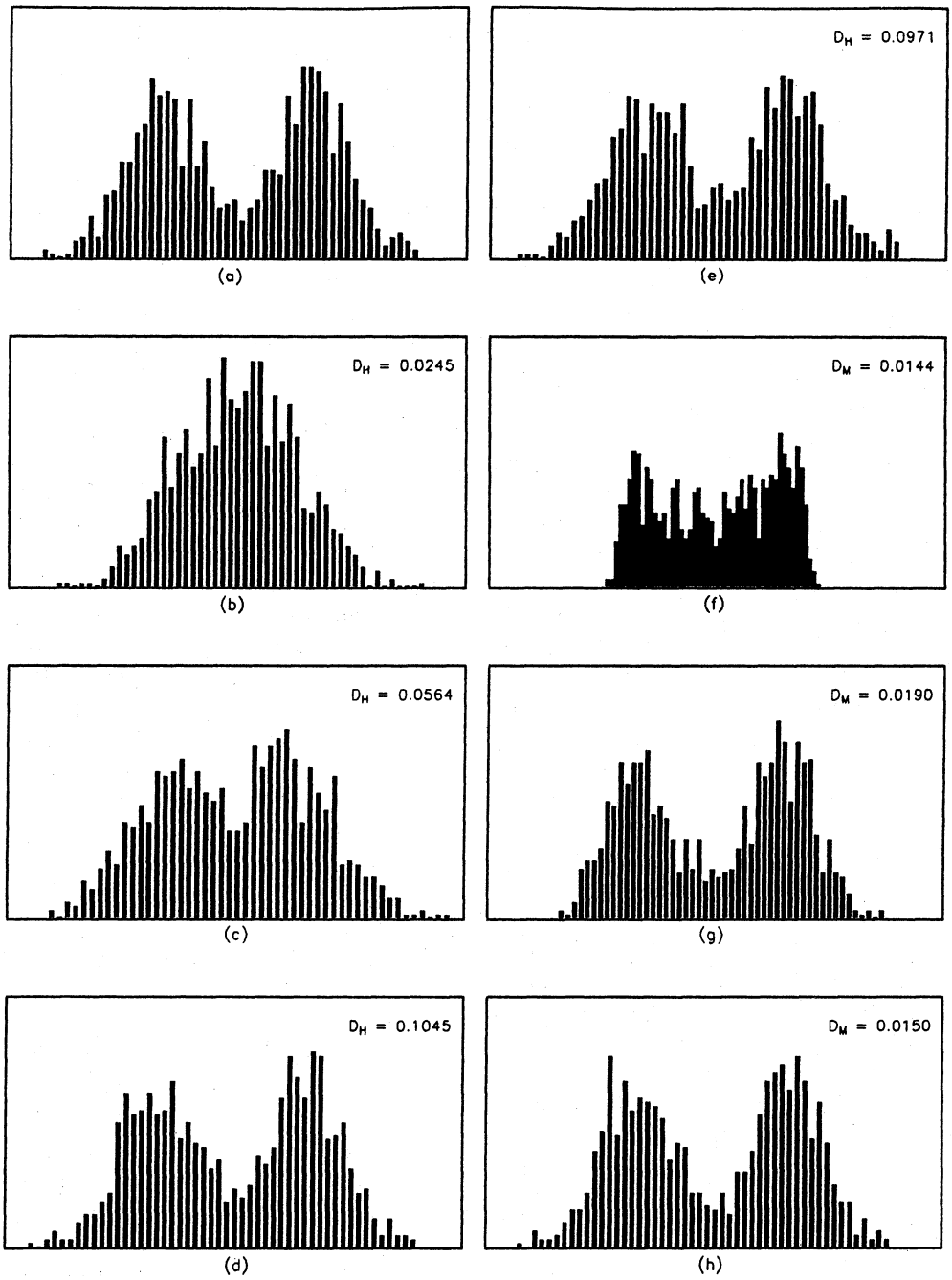


*Figure 1. Histograms (Range [−6, 6]) of Simulated Random Intercepts and Slopes Together With Their Estimates, Under the Homogeneity Model and Under the Two-Component Heterogeneity Model. $D_H$ and $D_M$ are the Kolmogorov—Smirnov statistics for testing goodness of fit of the homogeneity model and the two-component heterogeneity model. Their 5% critical value is .0428. (a) Real random effects; (b) intercept estimates under homogeneity, $\sigma^2 = 30$; (c) intercept estimates under homogeneity, $\sigma^2 = 5$; (d) intercept estimates under homogeneity, $\sigma^2 = .5$; (e) slope estimates under homogeneity, $\sigma^2 = 30$; (f) intercept estimates under heterogeneity, $\sigma^2 = 30$; (g) intercept estimates under heterogeneity, $\sigma^2 = 5$; (h) intercept estimates under heterogeneity, $\sigma^2 = .5$.*

($\mathbf{Z}_i = (1\,3\,5\,6\,10)'$), and for varying values of $\sigma^2$ ($\sigma^2 = .5, 5, 30$). Each time, the random effects were then estimated assuming normality for the $\mathbf{b}_i$. Some of the histograms of these estimates are shown in Figure 1, b–e and must be compared with the histogram of the $\mathbf{b}_i$, given in Figure 1(a). We conclude that in cases with large error variances $\sigma^2$ and covariates $\mathbf{Z}_i$ with small range, the estimates $\hat{\mathbf{b}}_i$ do not reflect the heterogeneity in the random-effects population.

In practice, the estimates $\hat{\mathbf{b}}_i$ are frequently used to highlight special profiles or to look for (groups of) individuals evolving differently in time. Examples have been provided by Waternaux, Laird, and Ware (1989) and by De Gruttola, Lange, and Dafni (1991). Thus it is very important to investigate to what extent the $\hat{\mathbf{b}}_i$ distribution represents the true random-effects distribution. First, histograms of the $\hat{\mathbf{b}}_i$ are only fully interpretable when the random-effects covariates $\mathbf{Z}_i$ are the same for all individuals; otherwise, the estimates $\hat{\mathbf{b}}_i$ are no longer identically distributed. Further, because $\hat{\mathbf{b}}_i$ depends on $\mathbf{b}_i$ as well as on $\mathbf{e}_i$, the weighted Q-Q plots introduced by Lange and Ryan (1989) cannot, strictly speaking, distinguish between wrong distributional assumptions for the random effects or for the error terms.

## 4. THE HETEROGENEITY MODEL

### 4.1 The Model

To accommodate clustered $\mathbf{b}_i$'s, the model described in Section 2 is extended by assuming that the random effects are sampled from a mixture of $g$ normal distributions with means $\boldsymbol{\mu}_j$ and covariance matrix $\mathbf{D}$. Each component of the mixture then represents a cluster containing a proportion $p_j$ from the population, $\sum_{j=1}^{g} p_j = 1$. To assure that $E(\mathbf{y}_i) = \mathbf{X}_i\boldsymbol{\alpha}$, the additional constraint $E(\mathbf{b}_i) = \sum_{j=1}^{g} p_j\boldsymbol{\mu}_j = \mathbf{0}$ is needed. The marginal distribution of the measurements $\mathbf{y}_i$ is then given by

$$\mathbf{y}_i \sim \sum_{j=1}^{g} p_j N(\mathbf{X}_i\boldsymbol{\alpha} + \mathbf{Z}_i\boldsymbol{\mu}_j, \mathbf{V}_i), \tag{1}$$

with $\mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \sigma^2 I_{n_i} = \mathbf{W}_i^{-1}$. The EM algorithm can now be used to calculate the maximum likelihood estimates for all parameters in (1). More details about fitting finite-mixture models have been provided by Redner and Walker (1984).

Note how model (1) can be seen as a hierarchical Bayes model where, given $\boldsymbol{\mu}$, $\mathbf{b}_i \sim N(\boldsymbol{\mu}, \mathbf{D})$. Because the distribution of $\boldsymbol{\mu}$ is now assumed to be discrete with probabilities $p_j$ at support points $\boldsymbol{\mu}_j, j = 1, \ldots, g$, (1) will be called the heterogeneity model. Further, a reviewer pointed out that model (1) shows some resemblance to the theory of George (1986), where the mean of a multivariate normal distribution is estimated by a weighted average of multiple shrinkage estimators, based on a set of prespecified prior weights. But our heterogeneity model combines parametric models instead of estimators and also allows to estimate the prior weights $p_j, j = 1, \ldots, g$. Also, our objective is to model heterogeneity in the random-effects distribution, whereas George's objective was to reduce the expected squared error loss.

### 4.2 Empirical Bayes Estimation and Classification

Let $\boldsymbol{\pi}$ denote the vector of component probabilities, $\boldsymbol{\pi}' = (p_1, \ldots, p_g)$, and let $\boldsymbol{\theta}$ be the vector containing the remaining parameters $\boldsymbol{\alpha}, \sigma, \mathbf{D}$, and all $\boldsymbol{\mu}_j$'s. Further, suppose that $\boldsymbol{\Psi}' = (\boldsymbol{\pi}', \boldsymbol{\theta}')$. Let $p_{ij}$ denote the posterior probability for the $i$th individual to belong to the $j$th component of the mixture, defined as

$$p_{ij} = p_{ij}(\boldsymbol{\Psi}) = \frac{p_j f_j(\mathbf{y}_i|\boldsymbol{\theta})}{\sum_{k=1}^{g} p_k f_k(\mathbf{y}_i|\boldsymbol{\theta})},$$

where $f_k(\mathbf{y}_i|\boldsymbol{\theta})$ denotes the density function of a multivariate normal distribution with mean $\mathbf{X}_i\boldsymbol{\alpha} + \mathbf{Z}_i\boldsymbol{\mu}_k$ and variance–covariance matrix $\mathbf{V}_i$. The empirical Bayes estimates for the random effects $\mathbf{b}_i$ can then be seen to equal

$$\hat{\mathbf{b}}_i = E(\mathbf{b}_i|\mathbf{y}_i, \boldsymbol{\Psi}) = \mathbf{D}\mathbf{Z}_i'\mathbf{W}_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\alpha}) + \mathbf{A}_i \sum_{j=1}^{g} p_{ij}(\boldsymbol{\Psi})\boldsymbol{\mu}_j, \tag{2}$$

where $\mathbf{A}_i = I - \mathbf{D}\mathbf{Z}_i'\mathbf{W}_i\mathbf{Z}_i$. The first component of (2) has the same expression as the estimate for $\mathbf{b}_i$ that we obtained in Section 2 under the normality assumption. But the overall covariance matrix $\mathbf{D}^*$ of the $\mathbf{b}_i$ is now replaced by the within-cluster covariance matrix $\mathbf{D}$. The second component of (2) can be viewed as a correction term toward the component means, proportional to the posterior probability of belonging to each of these components. In the case of univariate random effects, $A_i$ can be easily seen to be an increasing function of $\sigma^2(\mathbf{Z}_i'\mathbf{Z}_i)^{-1}$ satisfying $0 < A_i < 1$. Hence the correction term will receive more weight in those cases for which the random effect is poorly estimated under the homogeneity model.

The empirical Bayes estimates in expression (2) were calculated for the three simulated data sets of Section 3 with random intercepts from a two-component normal mixture model. The histograms of these estimates under the correct model reflect the correct distribution (much) better than do the corresponding estimates under the homogeneity model; see Figure 1, f–h. The better performance of the heterogeneity estimators is best seen for large values of $\sigma^2$.

Finally, classification of profiles no longer must be based on the random-effects estimates. It is very common in mixture models to classify the $i$th case in the component to which it has the highest probability to belong; that is, in the $j(i)$th component defined by $p_{i,j(i)}(\hat{\boldsymbol{\Psi}}) = \max_{1 \le j \le g} p_{ij}(\hat{\boldsymbol{\Psi}})$.

### 4.3 Tests for Heterogeneity

One main issue in deciding on the correct random-effects distribution is the choice of the appropriate number $g$ of mixture components. We thus propose the use of a statistical test to decide on the degree of heterogeneity. We consider three tests—a likelihood ratio (LR) test and two omnibus goodness-of-fit tests—all conditional on the maximum likelihood estimates.

The LR test can be used to specifically test the hypothesis $H_0$: $g = g_0$ against the alternative hypothesis $H_a$: $g = g_a > g_0$. It is known from analogous but simpler settings that because of boundary problems, the null distribution of the LR statistic does not necessarily converge to a $\chi^2$ distribution, and if there is any convergence at all, it can be painfully slow. Thus one often uses parametric bootstrap methods to simulate the null distribution for a specific data set and for a particular test on the number of components. Although the concept of the bootstrapping procedure is quite simple, its implementation involves many practical problems, such as finding starting values for maximization under the alternative hypothesis. This is especially true for repeated-measurement models with a covariate structure. Therefore, the LR test is not considered any further. McLachlan and Basford (1988, sec. 1.10) gave an extensive overview of the literature on the use of the LR test in finite-mixture problems.

An alternative and less time-consuming way of checking whether the correct number of mixture components was used in the model is the use of an omnibus goodness-of-fit test. The mixture model with the smallest $g$ that fits the data well is then taken as the candidate model. To avoid the evaluation of multivariate distribution functions, we use linear combinations $\mathbf{a}_i'\mathbf{y}_i$ of the original measurements $\mathbf{y}_i$, for which the distribution function $F$ can be easily cal-

*Table 1. Results of the Mixture Analysis for the Growth Curve Example: The Homogeneous Model (One Component) and Two Heterogeneous Models (Two and Three Components)*

| Number of components | Component means and component probabilities | Covariance matrix D | Residual variance $\sigma^2$ | Mixture classification (child numbers) |
|---|---|---|---|---|
| 1 | $\mu_1 = \begin{pmatrix}82.48\\5.72\end{pmatrix}$ $p_1 = 1$ | $\mathbf{D} = \begin{pmatrix}6.71 & -.07\\-.07 & .27\end{pmatrix}$ | $\sigma^2 = .47$ | — |
| 2 | $\mu_1 = \begin{pmatrix}82.78\\5.39\end{pmatrix}$ $p_1 = .68$ | $\mathbf{D} = \begin{pmatrix}6.73 & .10\\.10 & .03\end{pmatrix}$ | $\sigma^2 = .47$ | 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 18 |
| | $\mu_2 = \begin{pmatrix}82.06\\6.42\end{pmatrix}$ $p_2 = .32$ | | | 9, 15, 16, 17, 19, 20 |
| 3 | $\mu_1 = \begin{pmatrix}79.46\\5.60\end{pmatrix}$ $p_1 = .20$ | $\mathbf{D} = \begin{pmatrix}3.64 & .32\\.32 & .03\end{pmatrix}$ | $\sigma^2 = .47$ | 2, 6, 13 |
| | $\mu_2 = \begin{pmatrix}84.21\\5.32\end{pmatrix}$ $p_2 = .50$ | | | 1, 3, 4, 5, 7, 8, 10, 11, 12, 14, 18 |
| | $\mu_3 = \begin{pmatrix}81.65\\6.47\end{pmatrix}$ $p_3 = .30$ | | | 9, 15, 16, 17, 19, 20 |

culated, replacing all parameters by their maximum likelihood estimates. Under the correct model, we then have that the stochastic variables $\mathcal{U}_i = F(\mathbf{a}_i'\mathbf{y}_i)$ are uniformly distributed. The Kolmogorov–Smirnov test, applied to the $\mathcal{U}_i$, or the Shapiro–Wilk test, applied to the $\Phi^{-1}(\mathcal{U}_i)$, can then be viewed as a goodness-of-fit test for the model at hand.

The goodness-of-fit test can be performed for any $\mathbf{a}_i$, but the choice of $\mathbf{a}_i$ with maximal variability in $\mathbf{a}_i'\mathbf{y}_i$ due to the random effects compared to the variability due to the error terms is of interest. This corresponds to an $\mathbf{a}_i$ such that $\widehat{\text{var}}(\mathbf{a}_i'\mathbf{Z}_i\mathbf{b}_i)/\widehat{\text{var}}(\mathbf{a}_i'\mathbf{e}_i)$ is maximal; that is, the eigenvector corresponding to the largest eigenvalue of $\mathbf{Z}_i\widehat{\text{var}}(\mathbf{b}_i)\mathbf{Z}_i'$. For univariate random effects, this simplifies to the choice $\mathbf{a}_i'\mathbf{y}_i = \mathbf{Z}_i'\mathbf{y}_i$, which is sufficient for $\mathbf{b}_i$.

As an example, the Kolmogorov–Smirnov statistics for all models fitted to the simulated data sets from Section 3 are also shown in Figure 1. $D_H$ and $D_M$ denote the test statistics for testing the homogeneity model and the two-component heterogeneity model. For $\sigma^2 = .5$, only the heterogeneity model was accepted as a good fit [histograms (d) and (h)], whereas for $\sigma^2 = 30$, both models were found to fit the data well [histograms (b) and (f)]. This illustrates the fact that the power for detecting heterogeneity using any of the above-described tests converges to zero when $\sigma^2$ becomes infinitely large, as was shown by Verbeke and Lesaffre (1994).

Unfortunately, but also inevitably, goodness-of-fit tests and LR tests are still lacking a fundamental mathematical foundation. But we do not foresee an improvement in the near future, certainly not in the context of general mixed-effects models, because the problems are already formidable in much simpler situations (see, e.g., McLaclan and Basford 1988).

## 5. EXAMPLES

### 5.1 Example 1

Growth curves of 20 girls with height measured on a yearly basis from age 6 to 10 were analyzed by Goldstein (1979, table 4.3, p. 101). The girls were classified according to the height of their mother (group A: $< 155$ cm, child #1–6; group B: 155–164 cm, child #7–13; group C: $> 164$ cm, child #14–20). A significant (at 5%) group, as well as a significant group by age effect, were found. Because the group structure was obtained by discretizing the height of the mother at arbitrary points, which is very artificial, it may be useful to search for growth curve clusters, neglecting this a priori group structure. Because this can be seen as an explorative cluster analysis, no tests for heterogeneity were performed. Three mixture models with random intercepts and slopes were fitted. The results are summarized in Table 1.

First, it follows from the one-component estimates that the average intercept is 82.48 and the average slope is 5.72. The two-component analysis divides the children into "slow" growers (68%) and "fast" growers (32%). Finally, the three-component analysis further subdivides the slow growers, but its effect is mainly a reduction of the intercept variability. Because our mixture approach only partially reconstructs the prior group structure of Goldstein, we conclude that the latter does not fully reflect the heterogeneity in the growth curves.

### 5.2 Example 2

Recent research (Pearson, Morrell, Landis, Carter, and Brant 1994) has suggested that blood levels of prostate-specific antigen (PSA) may be useful in detecting prostate cancer. PSA is an enzyme produced by both normal and cancerous prostate cells, and its level is related to the volume of prostate tissue. Pearson et al. (1994) fitted linear mixed-effects models to longitudinal case-control data from the Baltimore Longitudinal Study of Aging and showed that cancer cases can be characterized by an extremely high rate of change in PSA levels. Using the Pearson et al. (1994) data from 16 healthy patients (controls) and 18 cancer patients (cases), we have investigated how well repeated measures of PSA levels can correctly classify patients as a "normal" or a "cancer" patient. Retrospective data are available for each individual, but the number of measurements varies between 4 and 15, and the time span they cover is quite different; that is, between 9.4 and 25.3 years.

Pearson et al. (1994) discussed the use of a linear mixed-effects model with $\ln(1 + \text{PSA})$ as response, age at diagnosis as fixed effect, and with random intercepts, random

*Table 2. Results of a Three-Component Mixture Analysis for the Prostate Cancer Example*

| Fixed effect $\alpha$ | Component means and component probabilities | Covariance matrix D | Residual variance $\sigma^2$ |
|---|---|---|---|
| .009 | $\mu_1 = \begin{pmatrix} -.0202 \\ .0124 \\ .0012 \end{pmatrix}$ $p_1 = .72$ | $\mathbf{D} = \begin{pmatrix} .0306 & .0082 & -.0003 \\ .0082 & .0023 & -.0001 \\ -.0003 & -.0001 & .00001 \end{pmatrix}$ | $\sigma^2 = .027$ |
| | $\mu_2 = \begin{pmatrix} .5110 \\ -.0088 \\ .0045 \end{pmatrix}$ $p_2 = .19$ | | |
| | $\mu_3 = \begin{pmatrix} .2167 \\ -.0288 \\ .0207 \end{pmatrix}$ $p_3 = .09$ | | |

time effects and random time$^2$ effects, where time is measured as years before diagnosis. But if specific interest is in classifying the patients from a prospective study as early as possible, then we have to use age at first visit as fixed effect, and time must be measured as follow up time. The results of our analysis are summarized in Table 2.

Both goodness-of-fit tests reveal a lack of fit of the homogeneous model, which of course can be due to several deviations from the underlying assumptions of the model. But because we are particularly interested in classifying patients in subgroups according to their random effects, our heterogeneity model seems to be a reasonable alternative. The two-component mixture analysis yields a first component (79% of the patients) containing patients evolving mainly linearly and a second component (21% of the patients) of patients who evolve quadratically. Although this model is accepted by the Kolmogorov–Smirnov test ($D_M = .2113 < .2274 = 5\%$ critical value), it is not accepted by the Shapiro–Wilk test ($p < .0001$). This last test suggests ($p = .0797, D_M = .1061 < .2274$), a mixture model with three components, the parameter estimates of which are given in Table 2.

Although there is no a priori reason why posterior classification of individuals should be expected to exactly correspond to the predefined groups of cases and controls, it is still of interest to compare both classifications. Indeed, if our three-component model is to be used to detect prostate cancer in an early stage, then our classification of the patients should reflect the true group structure (case/control). Except for one patient, all controls were classified in the first component, together with eight cases for which the profiles show hardly any difference from many profiles in the control group. This component represents the profiles which increase mainly linearly, but very slowly. The third component contains three cases. All other patients are classified into the second component. The third and second components represent the profiles that increase quadratically, immediately after enrollment in the study or after a period of very small linear increase.

Detection of the correct diagnostic group was hampered by the different onsets of observation periods. This leads to the conclusion that early detection of prostate cancer cannot be based on the linear mixed-effects model, which is in agreement with the results of Pearson et al. (1994), who argued that a piecewise nonlinear mixed-effects model should be used to estimate the time when rapid increases in PSA were first observable.

## 6. CONCLUSION

It was shown that the detection of subgroups in the random-effects population should not be based on the empirical Bayes estimates of the random effects obtained under the Gaussian mixed model. The introduced heterogeneity model, which explicitly takes into account the presence of subgroups, can be helpful as an explorative cluster analysis, but also allows testing of the Gaussian mixture assumption. Further, it provides estimates for the subgroup-specific parameters and yields a mathematical rule for the classification of the subjects in the several mixture components. But it also should be emphasized that the mixture components may be due to effects other than those one might expect, as was illustrated in the second example.

*[Received November 1993. Revised May 1995.]*

## REFERENCES

Butler, S. M., and Louis, T. A. (1992), "Random Effects Models with Non-Parametric Priors," *Statistics in Medicine*, 11, 1981–2000.

De Gruttola, V., Lange, N., and Dafni, U. (1991), "Modeling the Progression of HIV Infection," *Journal of the American Statistical Association*, 86, 569–577.

George, E. I. (1986), "Minimax Multiple Shrinkage Estimation," *The Annals of Statistics*, 14, 188–205.

Goldstein, H. (1979), *The Design and Analysis of Longitudinal Studies*, London: Academic Press.

Lange, N., and Ryan, L. (1989), "Assessing Normality in Random Effects Models," *The Annals of Statistics*, 17, 624–642.

McLachlan, G. J., and Basford, K. E. (1988), *Mixture Models, Inference and Applications to Clustering*, New York: Marcel Dekker.

Pearson, J. D., Morrell, C. H., Landis, P. K., Carter, H. B., and Brant, L. J. (1994), "Mixed-Effects Regression Models for Studying the Natural History of Prostate Disease," *Statistics in Medicine*, 13, 587–601.

Redner, R. A., and Walker, H. F. (1984), "Mixture Densities, Maximum Likelihood, and the EM Algorithm," *SIAM Review*, 26, 2, 195–239.

Strenio, J. F., Weisberg, H. J., and Bryk, A. S. (1983), "Empirical Bayes Estimation of Individual Growth-Curve Parameters and Their Relationship to Covariates," *Biometrics*, 39, 71–86.

Titterington, D. M., Smith, A. F. M., and Makov, U. E. (1985), *Statistical Analysis of Finite Mixture Distributions*, Chichester, U.K.: John Wiley.

Verbeke, G., and Lesaffre, E. (1994), "A Linear Mixed Effects Model with Heterogeneity in the Random Effects Population," technical report, Biostatistical Centre for Clinical Trials, Catholic University of Leuven.

Waternaux, C., Laird, N. M., and Ware, J. H. (1989), "Methods for Analysis of Longitudinal Data: Blood Lead Concentrations and Cognitive Development," *Journal of the American Statistical Association*, 84, 33–41.