JOURNAL OF BIOPHARMACEUTICAL STATISTICS, 11(4), 253–273 (2001)

A RANDOM-EFFECTS MIXTURE MODEL FOR CLASSIFYING TREATMENT RESPONSE IN LONGITUDINAL CLINICAL TRIALS

Weichun Xu[1,*] and Donald Hedeker[2]

[1]Pfizer, Inc., Ann Arbor, Michigan
[2]University of Illinois at Chicago, Chicago, Illinois

ABSTRACT

A random-effects regression model that allows the random coefficients to have a multivariate normal mixture distribution is described for classifying treatment response in longitudinal clinical trials. The proposed model is capable of dealing with longitudinal data from unknown heterogeneous populations. As applied to longitudinal clinical trials, for example, the model can distinguish subgroups of treatment response. Use of the proposed model is illustrated by analyzing data from two psychiatric clinical trials. The first includes depressed patients assigned to drug treatment who are repeatedly measured in terms of their level of depression. The second trial examined schizophrenic patients longitudinally who were assigned to either a drug or placebo condition. For both, the random-effects mixture model allows an assessment of whether patients comprise distinct populations in terms of their treatment response. Based on parameter estimates of the mixture model, ample evidence for a mixture of response to treatment is observed for both datasets.

Key Words: Longitudinal data; Mixed-effects models; Mixture distributions; Maximum marginal likelihood; Fisher scoring solution

* Corresponding author. Weichun Xu, Ph.D., c/o Donald Hedeker, Ph.D., Division of Epidemiology and Biostatistics (MC 923), School of Public Health, University of Illinois at Chicago, 1603 West Taylor Street, Room 955, Chicago, IL 60612-4336. Fax: (312) 996-0064; E-mail: hedeker@uic.edu

1. INTRODUCTION

Random-effects regression models (1–5) have become increasingly used for analysis of longitudinal data. In these models, it is generally assumed that the random subject effects are distributed as a multivariate normal in the population of subjects. Because the multivariate normal is a single-component distribution, this assumption implies that individuals belong to a homogenous population with a single mean vector and variance-covariance matrix. However, particularly in large scale clinical trials, the assumption of a single-component distribution may be unreasonable for a number of reasons: subpopulations of patients may exist that consist of treatment responders and nonresponders, measurement may vary from site to site, or unknown contaminating factors may be present (See, for example, Refs. 6 and 7). Ignoring the possibility that individuals come from two or more different subpopulations with unknown proportion(s) can produce biased estimates of the random-effects and their associated population variance terms (8,9). Also, in the case of response to treatment, the possibility of heterogeneity in treatment response may be of clinical interest by itself. This is especially true in psychiatric research where the identification and classification of patients treated with medications as ''responders'' and ''nonresponders'' is a matter of great interest and discussion (10–12).

A more general model for longitudinal data allows the individual random effects to be distributed as a mixture of normal components with different proportions and means. This type of representation of the population distribution is capable of dealing with data sampled from mixture populations. This general model also applies to the situation where a subject-level grouping variable is thought to influence the outcome, and thus would normally enter the model as a covariate, however, information about actual group membership is not available.

In the past decade, much has been written about mixture distributions, including important texts by Titterington, Smith, and Makov (13), Everitt and Hand (14), McLachlan and Basford (15), and McLachlan and Peel (16). These monographs offer systematic approaches to the structure of finite mixture distributions under a variety of situations. In addition, further discussion of research issues and topics for mixture distributions can be found in the review article by Titterington (17).

Incorporating mixture distributions into random-effects models has also recently been proposed. Belin and Rubin (18) described a method for analyzing repeated-measures data on schizophrenic reaction times by modeling the distribution of response times for and within each schizophrenic individual as a two-component mixture. Xu (8) and Verbecke and Lesaffre (9) described a mixed-effects regression model allowing the random effects to be distributed from a mixture distribution. The latter authors applied this model to search for growth curve clusters in the heights of girls, and to identify potential prostate-cancer patients based on repeated assessments of blood levels of prostate-specific antigen. For the case of complete data across time, Muthén and Shedden (19) described

a growth mixture model for longitudinal substance abuse data to classify trajectories of alcohol use across time. More recently, Muthén and associates (20–23) have greatly extended their approach to growth mixture modeling in several important ways, including allowing missing data across time, allowing the random-effects variances and covariances to vary across the mixture subgroups, allowing covariates to have varying effects across the mixture subgroups, and allowing the ability to model the probability of subgroup membership. All of these advances, and others, have recently been made available in the Mplus (version 2) software program (24).

Our aim in this article is to describe the random-effects mixture model as a way of examining whether treatment response can be considered homogeneous or heterogeneous. Specifically, the model can be used to address the important question concerning whether distinct subgroups exist in terms of treatment response. Also, whereas Verbecke and Lesaffre (9) and Muthén and Shedden (19) estimate parameters using only the EM algorithm, we describe (in the appendix) a solution using both EM and Fisher scoring algorithms. This sequential use of EM followed by Fisher scoring iterations is potentially useful in overcoming drawbacks encountered when using either algorithm alone. For instance, as noted by Aitkin, ''the EM algorithm can be slow to converge in mixture models and does not provide the correct (asymptotic) information-based standard errors for the maximum-likelihood estimates without additional computation'' (25). And while the Fisher scoring solution is often faster than EM and does yield standard errors at convergence, it can be sensitive to the choice of starting values (13). Finally, whereas the Mplus software program relies upon subjects being measured at the same fixed timepoints (though subjects are not assumed measured at all of these fixed timepoints), we present our model, like Verbecke and Lesaffre (9), allowing time to be a continuous predictor variable that is capable of taking on different values across subjects. In the next section, the random-effects mixture model and resulting log-likelihood function are described. Data from two psychiatric clinical trials are then analyzed using the proposed model to illustrate its use for classification of treatment response. Discussion and comments are provided in the last section, and the appendix contains details on parameter estimation.

2. MODEL DESCRIPTION

For a given individual i, the mixed-effects regression model can be written as:

$$\boldsymbol{y}_i = \boldsymbol{Z}_i\alpha + \boldsymbol{X}_i\boldsymbol{\beta}_i + \varepsilon_i \quad i = 1, 2, \ldots, N \tag{1}$$

where $\boldsymbol{y}_i$ is the $n_i \times 1$ vector of responses for subject i, $\boldsymbol{Z}_i$ is a known $n_i \times p$ matrix containing explanatory variables, $\boldsymbol{\alpha}$ is a $p \times 1$ vector of unknown regression parameters, $\boldsymbol{X}_i$ is a known $n_i \times q$ design matrix, $\boldsymbol{\beta}_i$ is the $q \times 1$ vector of unknown individual effects distributed as $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, independent of each other and

of the ε_i, and ε_i is the $n_i \times 1$ vector of random errors distributed independently as $\mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_i)$. Occasionally, one has a model with both random effects ($\boldsymbol{\beta}_i$) and serial correlated errors (26). For this, $\sigma^2\mathbf{I}_i$ would be replaced by an AR(1) correlation matrix, for example. The methods described in this article would still hold, however for simplicity we will assume independent errors in what follows.

In the above representation, both subject-varying and time-varying covariates are included in $\boldsymbol{Z}_i$. An alternative representation includes only time-varying covariates in $\boldsymbol{Z}_i$, and specifies the $\boldsymbol{\beta}_i$ as being distributed with mean vectors ($\boldsymbol{\mu}$) that are linear functions of the subject-varying covariates. For example, for a model with a single random (subject) effect and a subject-varying grouping variable, $\mu = \gamma_0 + w_i\gamma_1$ (where w_i is a dummy-coded grouping variable indicating two subgroups). This representation suggests that when the information regarding group membership is not available, it is sensible to allow the mean vector $\boldsymbol{\mu}$ to be an unknown random vector that takes a finite vector with some proportion. In this case, the random-effects mixture model can be written as in model (1), except that $\boldsymbol{\beta}_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where the $\boldsymbol{\mu}$ itself is a random vector following the distribution

$$P\{\boldsymbol{\mu} = \boldsymbol{\mu}_k\} = \pi_k, \quad k = 1, \ldots, c \tag{2}$$

where $\pi_k \geq 0$ ($k = 1, \ldots, c$) and $\sum_{k=1}^{c}\pi_k = 1$. In other words, $\boldsymbol{\beta}_i$ is regarded as arising from a population P that is a mixture of a finite number, say c, of subpopulations $P_1, \ldots, P_c$ with probabilities π_k. Thus, β_i has the *finite mixture* distribution:

$$g(\boldsymbol{\beta}) = \sum_{k=1}^{c} \pi_k g_k(\boldsymbol{\beta}) \tag{3}$$

where $g_k(\beta) = (2\pi)^{-q/2}|\boldsymbol{\Sigma}|^{-1/2}\exp[-1/2(\boldsymbol{\beta} - \boldsymbol{\mu}_k)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_k)]$. Here, it's important to note that c is assumed known and must be specified a priori. Unfortunately, determination of c is far from trivial. For instance, comparing models based on likelihood ratio tests is problematic due to boundary problems (13,34). Thus, more ad-hoc approaches are often used. One approach, noted by Verbecke and Molenberghs (34), is to increase c until some of the subgroups get very small in terms of π_k or their means $\boldsymbol{\mu}_k$ coincide. Also, Verbecke and Lesaffre (9) advocate some goodness-of-fit checks that can be used as aids in deciding on the number of components. Clearly, more work is required in this area.

Under these assumptions, given $\boldsymbol{\mu}_k$, the observations $\boldsymbol{y}_i$ and random effects $\boldsymbol{\beta}$ have the joint multivariate normal distribution:

$$\begin{bmatrix} \boldsymbol{y}_i \\ \boldsymbol{\beta}_i \end{bmatrix} \Bigg|_{\boldsymbol{\mu}_k} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{Z}_i\boldsymbol{\alpha} + \boldsymbol{X}_i\boldsymbol{\mu}_k \\ \boldsymbol{\mu}_k \end{bmatrix}, \begin{bmatrix} \boldsymbol{X}_i\boldsymbol{\Sigma}\boldsymbol{X}_i' + \sigma^2\boldsymbol{I}_i & \boldsymbol{X}_i\boldsymbol{\Sigma} \\ \boldsymbol{\Sigma}\boldsymbol{X}_i' & \boldsymbol{\Sigma} \end{bmatrix}\right) \tag{4}$$

The mean of the posterior distribution of $\boldsymbol{\beta}_i$, given $\boldsymbol{y}_i$ and $\boldsymbol{\mu}_k$, yields the following Empirical Bayes or Expected A Posterior (EAP) estimator (2) of the individual trend parameters:

$$\hat{\boldsymbol{\beta}}_{ik} \equiv \hat{\boldsymbol{\beta}}_i|\boldsymbol{y}_i, \boldsymbol{\mu}_k = [\boldsymbol{X}_i'(\sigma^2\boldsymbol{I}_i)^{-1}\boldsymbol{X}_i + \boldsymbol{\Sigma}^{-1}]^{-1}\boldsymbol{X}_i'(\sigma^2\boldsymbol{I}_i)^{-1}(\boldsymbol{y}_i - \boldsymbol{Z}_i\boldsymbol{\alpha} - \boldsymbol{X}_i\boldsymbol{\mu}_k) + \boldsymbol{\mu}_k \quad (5)$$

with corresponding covariance matrix

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}_i|\boldsymbol{y}_i,\boldsymbol{\mu}_k} = [\boldsymbol{X}_i'(\sigma^2\boldsymbol{I}_i)^{-1}\boldsymbol{X}_i + \boldsymbol{\Sigma}^{-1}]^{-1} \quad (6)$$

The marginal density of $\boldsymbol{y}_i$, for fixed $\boldsymbol{\mu}_k$, can be obtained by integrating over $\boldsymbol{\beta}_i$ the product of the densities of $\boldsymbol{\beta}_i$ and $\boldsymbol{y}_i|\boldsymbol{\beta}_i$ [$\sim \mathcal{N}(\boldsymbol{Z}_i\boldsymbol{\alpha} + \boldsymbol{X}_i\boldsymbol{\mu}_k, \boldsymbol{X}_i\boldsymbol{\Sigma}\boldsymbol{X}_i' + \sigma^2\boldsymbol{I}_i)$], that is, $f_k(\boldsymbol{y}_i) = \int\boldsymbol{\beta}_i h_{ik}g_k d\boldsymbol{\beta}_i$, where $h_{ik} = (2\pi)^{-n_i/2}|\sigma^2\boldsymbol{I}_i|^{-1/2}\exp[-1/2\boldsymbol{e}_{ik}'(\sigma^2\boldsymbol{I}_i)^{-1}\boldsymbol{e}_{ik}]$, $\boldsymbol{e}_{ik} = \boldsymbol{y}_i - \boldsymbol{Z}_i\boldsymbol{\alpha} - \boldsymbol{X}_i\boldsymbol{\beta}_i$, and the prior distribution is given by $g_k \equiv g_k(\boldsymbol{\beta}_i)$. The dependence of $\boldsymbol{e}_{ik}$ on the index k is obvious by noting $\boldsymbol{\beta}_i \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma})$ for a given $\boldsymbol{\mu}_k$. If we let $\boldsymbol{\theta}_k = (\boldsymbol{\mu}_k', \text{vech}'\boldsymbol{\Sigma}, \boldsymbol{\alpha}', \sigma^2)'$, then $f_k(\boldsymbol{y}_i)$ can be written as $f_k(\boldsymbol{y}_i|\boldsymbol{\theta}_k)$. The vec-operator represents a matrix as a vector by stacking the columns of the matrix one below the other. The vech-operator transforms a $n^2 \times 1$ vector, say vec($\boldsymbol{\Sigma}$) into a $(1/2)n(n+1) \times 1$ vector vech ($\boldsymbol{\Sigma}$) by eliminating all supradiagonal elements of $\boldsymbol{\Sigma}$ (27).

Because the data from the mixture distribution can be regarded as a special case of incomplete data, the *complete data* log-likelihood function (28), in terms of all unknown parameters $\boldsymbol{\varphi}$, is given by

$$\ell(\boldsymbol{\varphi}) = \sum_{i=1}^{N} \boldsymbol{d}_i'\boldsymbol{U}(\boldsymbol{\pi}) + \sum_{i=1}^{N} \boldsymbol{d}_i'\boldsymbol{V}_i(\boldsymbol{\theta}) \quad (7)$$

where $\boldsymbol{U}(\boldsymbol{\pi}) = [\log \pi_1, \ldots, \log \pi_c]'$, $\boldsymbol{V}_i(\boldsymbol{\theta}) = [\log f_1(\boldsymbol{y}_i|\boldsymbol{\theta}_1), \ldots, \log f_c(\boldsymbol{y}_i|\boldsymbol{\theta}_c)]'$ and each $\boldsymbol{d}_i$ is a vector of length c with all elements equal to 0 except one (which equals 1) that indicates the appropriate subpopulation to which subject i belongs. The preference of Eq. (7) over the ordinary log-likelihood function $\tilde{\ell}(\boldsymbol{\varphi}) = \sum_{i=1}^{N}\log[\sum_{k=1}^{c}\pi_k f_k(\boldsymbol{y}_i|\boldsymbol{\theta}_k)]$ is because it readily lends itself to a solution by the general class of iterative procedures known as EM algorithms. Thus, in their description of the mixed-effects mixture model, Verbecke and Lesaffre (9) advocate use of EM to estimate the model parameters. For this, it can be shown (see Ref. 8) that the E-Step consists of Eqs. (5), (6), and

$$w_{ik}(\boldsymbol{\varphi}^{(m)}) = \pi_k^{(m)} f_k(\boldsymbol{y}_i|\boldsymbol{\theta}_k^{(m)})/P(\boldsymbol{y}_i|\boldsymbol{\varphi}^{(m)}) \quad (8)$$

where $\boldsymbol{w}_i(\boldsymbol{\varphi}^{(m)}) = E(\boldsymbol{d}_i|\boldsymbol{y}_i, \boldsymbol{\varphi}^{(m)})$ with $w_{ik}(\boldsymbol{\varphi}^{(m)})$ as the kth element of this vector, and $P(\boldsymbol{y}_i) = \sum_{k=1}^{c}\pi_k f_k(\boldsymbol{y}_i|\boldsymbol{\theta}_k)$. Notice that these weights (w_{ij}) are the probabilities of category membership for the ith subject, conditional on $\boldsymbol{y}_i$ (and given parameter vector $\boldsymbol{\varphi}^{(m)}$).

The M-Step consists of the following maximum likelihood solutions:

$$\hat{\boldsymbol{\mu}}_k = \frac{1}{n_k}\sum_i w_{ik}\hat{\boldsymbol{\beta}}_{ik} \quad (9)$$

$$\hat{\boldsymbol{\alpha}} = \left[\sum_i \boldsymbol{Z}_i'\boldsymbol{Z}_i\right]^{-1}\left[\sum_{i=1}^{N}\sum_{k=1}^{c} w_{ik}\boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_{ik})\right] \tag{10}$$

$$\hat{\sigma}^2 = \left(\sum_i n_i\right)^{-1}\sum_{i=1}^{N}\sum_{k=1}^{c} w_{ik}\,\mathrm{tr}(\boldsymbol{u}_{ik}\boldsymbol{u}_{ik}' + \boldsymbol{X}_i\boldsymbol{\Sigma}_{\boldsymbol{\beta}_i|\mathbf{y}_i,\boldsymbol{\mu}_k}\boldsymbol{X}_i') \tag{11}$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N}\left[\sum_{i=1}^{N}\sum_{k=1}^{c} w_{ik}(\boldsymbol{\Sigma}_{\boldsymbol{\beta}_i|\mathbf{y}_i,\boldsymbol{\mu}_k} + (\hat{\boldsymbol{\beta}}_{ik} - \hat{\boldsymbol{\mu}}_k)(\hat{\boldsymbol{\beta}}_{ik} - \hat{\boldsymbol{\mu}}_k)')\right] \tag{12}$$

$$\hat{n}_k = \sum_i w_{ik} \tag{13}$$

$$\hat{\pi}_k = \frac{1}{N}\sum_{i=1}^{N} w_{ik}(\boldsymbol{\varphi}), \quad k = 1, \ldots, c \tag{14}$$

where $\boldsymbol{u}_{ik} = \boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_{ik} - \boldsymbol{Z}_i\hat{\boldsymbol{\alpha}}$. The EM solutions then consist of iterating between E-step equations (5), (6), and (8) and M-step solutions (9)–(14) until convergence. It is easy to see that when $c = 1$, i.e., only one category exists, $w_{i1} = 1$ and therefore $\boldsymbol{\mu}_k = \boldsymbol{\mu}$; in this case, the above solution is the same as that obtained by Laird and Ware (1) for the ordinary homogeneous model. In carrying out the EM solution for the mixture model, the initial E-Step requires starting values for parameters $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$, σ^2, $\boldsymbol{\Sigma}$, and π_k. For this, estimates from the ordinary homogeneous model can be used, additionally specifying $\pi_k = 1/c$.

As many have noted, the EM solution can be slow to converge. Thus, as pointed out by Bock (29), it can be beneficial to switch to a Newton algorithm at some point in the iterative process. For this purpose, we have derived a Fisher-scoring solution (details in the appendix). The Fisher-scoring solution provides an estimate of the information matrix, which, at the solution point, is inverted to provide the large-sample variances and covariances of the maximum marginal likelihood estimators. Thus, confidence intervals and hypothesis testing of the model parameters can be easily obtained.

3. EXAMPLES

As noted in the introduction, in psychiatric research the notion that there are distinct subgroups in terms of treatment response is often of interest. Simply put, the idea is that while many patients do respond to medication, others do not. This has been observed both for major depressive disorder (10,11) and schizophrenia (12). In the examples below, we will use the random-effects mixture model to statistically examine this issue in two datasets. The first is a dataset of depressed patients, while the second consists of schizophrenics. For both, we will address the question of whether treatment response is better characterized as being unimodal (i.e., a matter of degree) or bimodal (distinct groups of ''responders'' and

''nonresponders''). Although the model, as written above, allows for more than two components, here we focus on contrasting the two-component model to the usual unimodel model because we are interested in illustrating the utility of the model for identification of treatment ''responders'' versus ''nonresponders.''

Example 1: Severity of Depression Across Time

The first dataset to be analyzed consists of data from a psychiatric study of Riesby et al. (30) In that study, 66 depressed inpatients were given the antidepressant drug imipramine for four treatment weeks after completing a one-week drug washout period. The Hamilton Depression Rating Scale (HDRS) was used to measure clinical response to depression. HDRS ratings were completed for each patient by study psychiatrists at the start of the study, after the drug washout week, and at the end of each of the four treatment weeks. Not all subjects were measured at each timepoint: the six timepoint sample sizes were 61, 63, 65, 65, 64, and 58. An important aspect of the study was the classification of subjects in terms of their treatment response to imipramine. Despite the fact that data were collected repeatedly over several timepoints for all subjects, Riesby and associates simply defined treatment response as a trichotomy (response, partial response, or nonresponse) based soley on the final HDRS score of each patient. Here, the random-effects mixture model will be used to classify treatment response using all available data from all subjects.

Our first model includes two individual-varying terms: an intercept β_{0i} representing the initial level of severity (at week 0) as measured by the HDRS, and a linear term β_{1i} reflecting the weekly change in the HDRS across the six timepoints of the study (where Week = 0, 1, 2, 3, 4, or 5). This model for y_{ij} (the HDRS score for subject i at timepoint j) can be written as

$$y_{ij} = \beta_{0i} + \beta_{1i}\text{Week}_j + \varepsilon_{ij} \tag{15}$$

In terms of a multilevel (31) or hierarchical (32) structure, the model can be partitioned into the within-subjects (or level-1) model that is specified in Eq. (15), and the between-subjects (or level-2) model given by

$$\begin{aligned} \beta_{0i} &= \mu_{\beta_0} + v_{0i} \\ \beta_{1i} &= \mu_{\beta_1} + v_{1i} \end{aligned} \tag{16}$$

The between-subjects model is sometimes referred to as a ''slopes as outcomes'' model (33). The multilevel representation shows that just as within-subjects (level-1) covariates are included in the model to explain variation in level-1 outcomes (y_{ij}), between-subjects (level-2) covariates are included to explain variation in level-2 outcomes (the subject's intercept β_{0i} and slope β_{1i}). Notice also that whereas $\boldsymbol{\beta}_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the corresponding random subject deviations $\boldsymbol{v}_i$ are distributed in the population as $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, i.e., they are random deviations centered around zero.

The results for the one-component model are listed in the left-hand column of Table 1. Both estimates of the population mean intercept (μ_{β_0}) and slope (μ_{β_1}) are found to be highly significant, suggesting that patients start with a nonzero HDRS score at baseline and have a highly significant overall weekly rate of improvement (about 2.4 units per week). The estimates of the intercept variance ($\sigma^2_{\beta_0} = \text{Var}[\beta_{0i}]$) and slope variance ($\sigma^2_{\beta_1} = \text{Var}[\beta_{1i}]$) easily exceed twice their standard error estimates, reflecting the wide variation in intercepts and slopes across individuals. It should be noted that use of the Wald test for hypothesis testing of variance terms has been called into question [see Bryk and Raudenbush (32), page 55], especially for small variances. Thus, we use the ratio of a variance estimate to its standard error as an approximate index. The negative covariance between the intercept and the slope is less than 1.5 times its standard error, expressed as a correlation a moderate value of -0.28 is obtained, suggesting a modest association between higher initial severity levels and greater improvement across time.

Marginal histograms of the empirical Bayes estimates of the random individual intercepts and slopes from the one component model are given in Figures 1a and 1b. A visual inspection of these histograms suggests that the marginal distributions do not seem to be entirely consistent with a one-component bivariate normal distribution. This is interesting given that Verbeke and Molenberghs (34) have noted that histograms of the empirical Bayes estimates often resemble unimodal normal distributions even when simulated from a bimodal distribution. Thus, to examine the possibility of bimodality more formally, the model is reestimated allowing the intercepts and slopes to follow a two-component bivari-

Table 1. Riesby Dataset: HDRS Across Time ($N = 66$) Parameter Estimates (Est.), Standard Errors (SE), and p Values

	One Component Model			Two Component Model		
	Est.	SE	$p<$	Est.	SE	$p<$
Intercept $\mu_{1\beta_0}$	23.58	0.55	0.0001	27.63	0.90	0.0001
Slope $\mu_{1\beta_1}$	−2.38	0.21	0.0001	−1.56	0.39	0.0001
Intercept $\mu_{2\beta_0}$				22.14	0.53	0.0001
Slope $\mu_{2\beta_1}$				−2.65	0.23	0.0001
Proportion 1 π_1				0.26	0.05	0.0001
Proportion 2 π_2				0.74	0.05	0.0001
Intercept variance $\sigma^2_{\beta_0}$	12.63	3.47		6.79	2.46	
Intercept-slope covariance $\sigma_{\beta_0\beta_1}$	−1.42	1.03		−2.53	0.94	
Slope variance $\sigma^2_{\beta_1}$	2.08	0.50		1.84	0.46	
Error variance σ^2	12.22	1.11		12.23	1.11	
$-2 \log L$	2219.0			2207.8		

Note that p-values are not given for variance and covariance terms [see Bryk and Raudenbush (32), page 55].

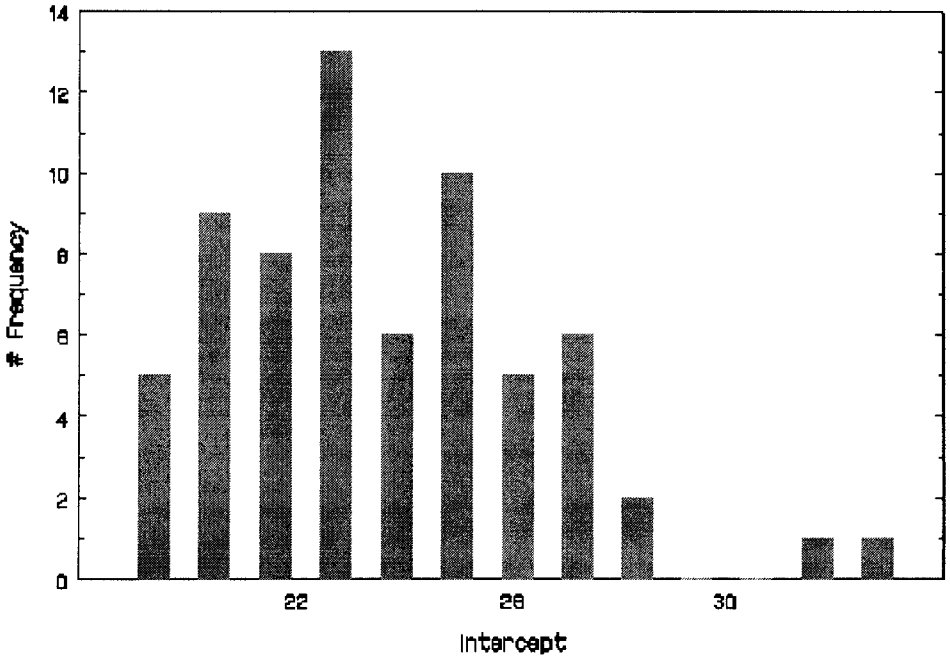

Figure 1a. Histogram of estimated random effect intercept in the one-component model.

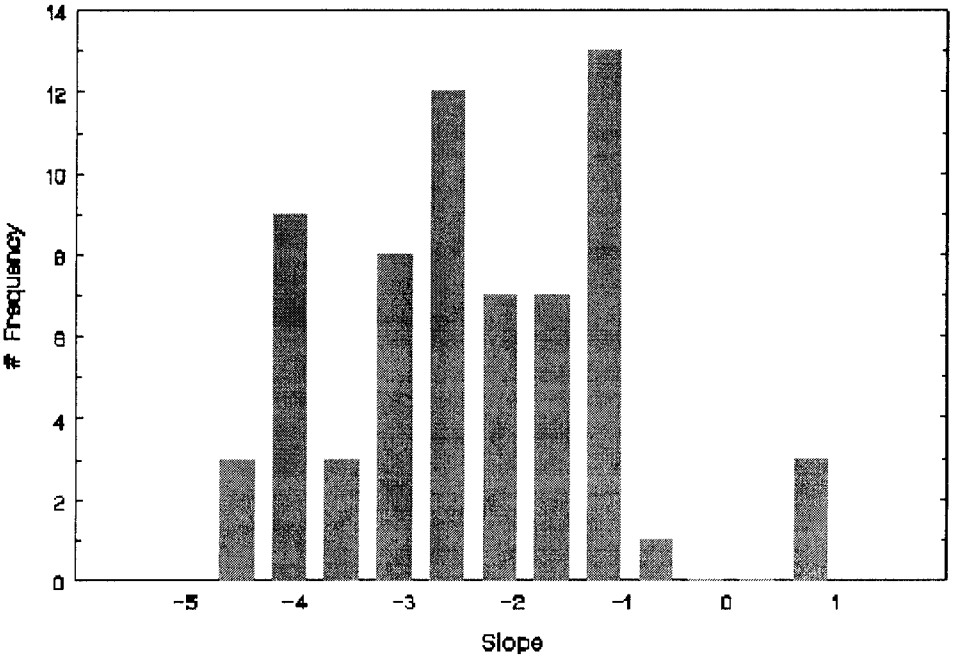

Figure 1b. Histogram of estimated random effect slope in the one-component model.

ate normal distribution. The right-hand side of Table 1 lists the results obtained from the proposed (two-component) model.

In the two-component model, the proportion is estimated as 0.26, indicating that about one-fourth of the patients are classified into the first group and three-fourths into the second group. The estimates of the means for these two groups ($\hat{\boldsymbol{\mu}}_1 = [27.63, -1.56]$ and $\hat{\boldsymbol{\mu}}_2 = [22.14, -2.65]$) indicate that the patients in the first group are initially more severely ill (high HDRS) and improve more slowly across time than the second group of patients who have a relatively less severe illness (lower HDRS) initially and improve more dramatically across time. As the rate of improvement is much larger for the second group than the first, the mixture model suggests that these two groups correspond to ''nonresponse'' and ''response'' groups, respectively. Turning to the estimates of the variances for the intercept and slope, we see that these have been reduced in the mixture model, especially the intercept variance. Thus, part of the subject heterogeneity in intercepts and slopes of the one-component model is now being explained by the two-component model. The negative covariance between the intercept and the slope is now in excess of twice its standard error, suggesting a very strong negative association (correlation $= -0.72$) between the initial severity of illness and the rate of improvement across the time.

Notice that, in terms of the multilevel representation, the resulting two-component mixture model can be represented as the same level-1 model given above in Eq. (15) and the following augmented between-subjects (or level-2) model:

$$\begin{aligned} \beta_{0i} &= \mu_{1\beta_0} G_{1i} + \mu_{2\beta_0} G_{2i} + v_{0i} \\ \beta_{1i} &= \mu_{1\beta_1} G_{1i} + \mu_{2\beta_1} G_{2i} + v_{1i} \end{aligned} \tag{17}$$

where the indicator variables G_{1i} and G_{2i} designate whether individual i belongs to the first or second groups, respectively (i.e., either G_{1i} or $G_{2i} = 1$, the other equals 0). Although these group membership indicator variables are unobserved, they can be estimated based on the results of the mixture model. Specifically, these are obtained using the weights w_{ik} (defined in the Appendix), which are the probabilities of group membership for the ith patient, conditional on $\boldsymbol{y}_i$ and estimates of the model parameters. Group membership is then estimated based on the value of w_{ik}. In these data, the ith individual was classified into the first group ($G_{1i} = 1$) if $w_{i1} > 0.5$, otherwise they were classified as belonging to the second group ($G_{2i} = 1$). Using this procedure, 16 out of 66 patients belong to the first group, and the remaining 50 patients belong to the second group. Using these groups, the observed and estimated group means are plotted in Figure 2. As can be seen, the model fits the observed means well. Figure 2 clearly illustrates the lower initial depression level and greater improvement rate of the second group, relative to the first.

It is interesting to compare these response subgroups of the mixture model with those obtained by Riesby et al. (30) where patients were classified based solely on their final HDRS score into three groups: ''full response'' (final HDRS

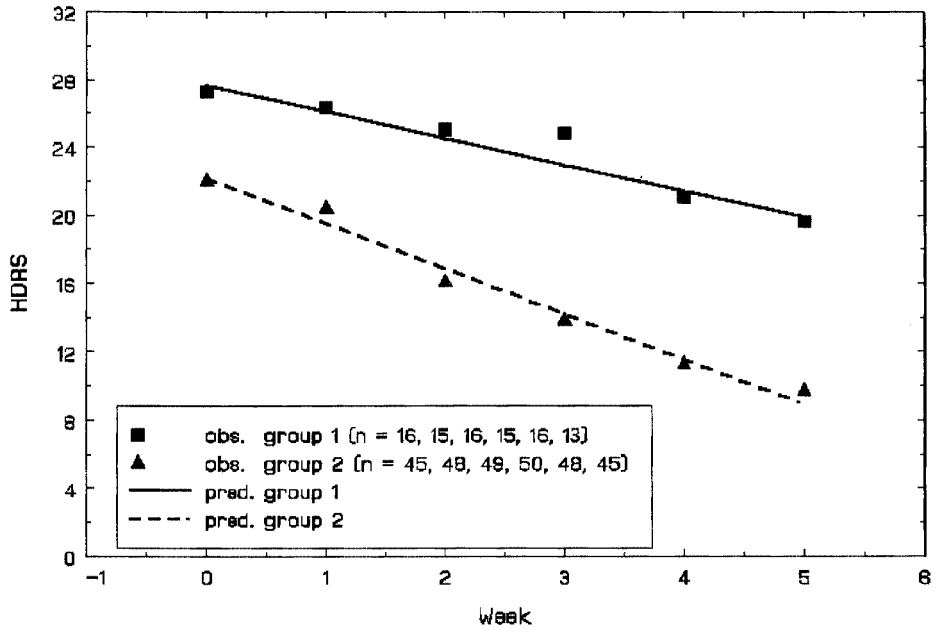

Figure 2. Riesby data: Hamilton depression scores across time mixture model observed and predicted group means.

≤ 7; $n = 21$), ''partial response'' (final HDRS = 8–15; $n = 23$), and ''nonresponse'' (final HDRS ≥ 16; $n = 22$). Of the 50 patients classified by the mixture model into the second group, Riesby et al. classified 20, 21, and 9 patients as full, partial, and nonresponse, respectively. Similarly, of the 16 patients in the first group of the mixture model, Riesby et al. classified 1 (response), 2 (partial response), and 13 (nonresponse). Thus, the first mixture group consists almost exclusively of ''nonresponders'' (13 of 16), and the second mixture group is predominantly ''full'' and ''partial'' responders (41 of 50). The use of the mixture model to classify subjects is preferred because it uses all the data from a subject across time, rather than simply the last timepoint. Additionally, the mixture model provides a more rigorous and less ad-hoc approach to determining the classification.

Example 2: Severity of Schizophrenic Symptomatology Across Time

The second dataset to be analyzed is from the NIMH Schizophrenia Collaborative Study on treatment related changes in overall severity. Specifically, we examined Item 79 of the clinician-rated Inpatient Multidimensional Psychiatric Scale [IMPS; Lorr and Klett (35)] and its change across time. Item 79, ''Severity of Illness,'' was scored as: 1 = normal, not at all ill; 2 = borderline mentally ill; 3 = mildly ill; 4 = moderately ill; 5 = markedly ill; 6 = severely ill; and 7 = among the most extremely ill. These data have been previously analyzed using

an ordinary one-component random-effects regression model (37). In this study, patients were randomly assigned to receive one of four medications: placebo, chlorpromazine, fluphenazine, or thioridazine. The latter three all belong to the class of antipsychotic drugs known as phenothiazines, which are commonly prescribed in the treatment of schizophrenia (36). Since previous analyses revealed similar effects for the three antipsychotic drug groups, they were combined in the present analysis. The experimental design and corresponding sample sizes are listed in Table 2.

As can be seen from Table 2, most of the measurement occurred at weeks 0, 1, 3, and 6. However, not all subjects were observed at these four primary timepoints, and so there were some intermittant measurements at weeks 2, 4, and 5 as well.

Based on previous analysis, the square root of week was used in the level-1 model instead of week (coded 0 to 6) to linearize the relationship between symptom severity and time. Also, because there was a fair amount of dropout in this study, which varied by treatment group, Hedeker and Gibbons (38) proposed a pattern-mixture random-effects regression model for analysis. Here, we will concentrate on a different issue, namely, the degree to which heterogeneity in treatment response exists while concurrently estimating the drug effect on response. For this, we use the same within-subjects (level-1) model in (15), albeit with the square root of week replacing week, and compare the following one-component between-subjects (level-2) model:

$$\begin{aligned} \beta_{0i} &= \mu_{\beta_0} + \alpha_0 \text{Drug}_i + v_{0i} \\ \beta_{1i} &= \mu_{\beta_1} + \alpha_1 \text{Drug}_i + v_{1i} \end{aligned} \tag{18}$$

to the two-component version:

$$\begin{aligned} \beta_{0i} &= \mu_{1\beta_0} G_{1i} + \mu_{2\beta_0} G_{2i} + \alpha_0 \text{Drug}_i + v_{0i} \\ \beta_{1i} &= \mu_{1\beta_1} G_{1i} + \mu_{2\beta_1} G_{2i} + \alpha_1 \text{Drug}_i + v_{1i} \end{aligned} \tag{19}$$

Drug_1 equals 0 and 1 for placebo and drug patients, respectively, and G_{1i} and G_{2i} are the unobserved indicators of whether the patient belongs to the first or second group of the mixture distribution.

Table 2. Experimental Design and Weekly Sample Sizes

	Sample Size at Week:						
Group	0	1	2	3	4	5	6
Placebo (n = 108)	107	105	5	87	2	2	70
Drug (n = 329)	327	321	9	287	9	7	265

Drug = chlorpromazine, fluphenazine, or thioridazine.

Table 3. Schizophrenia Dataset: IMPS79 Across Time (N = 437) Parameter Estimates (Est.), Standard Errors (SE), and p Values

	One Component Model			Two Component Model		
	Est.	SE	$p<$	Est.	SE	$p<$
Intercept $\mu_{1\beta_0}$	5.56	0.09	0.0001	5.36	0.07	0.0001
Slope $\mu_{1\beta_1}$	−0.32	0.07	0.0001	−0.01	0.04	0.74
Intercept $\mu_{2\beta_0}$				5.32	0.09	0.0001
Slope $\mu_{2\beta_1}$				−0.95	0.05	0.0001
Drug α_0	0.05	0.11	0.63	0.05	0.08	0.55
Drug × slope α_1	−0.65	0.08	0.0001	−0.52	0.04	0.0001
Proportion 1 π_1				0.56	0.02	0.0001
Proportion 2 π_2				0.44	0.02	0.0001
Intercept variance $\sigma^2_{\beta_0}$	0.40	0.07		0.36	0.03	
Intercept-slope covariance $\sigma_{\beta_0\beta_1}$	0.01	0.04		0.02	0.02	
Slope variance $\sigma^2_{\beta_1}$	0.25	0.03		0.01	0.03	
Error variance σ^2	0.64	0.03		0.59	0.06	
$-2 \log L$	2391.4			2314.6		

Note that p-values are not given for variance and covariance estimates [see Bryk and Raudenbush (32), page 55].

Table 3 presents results of both the one-component and the two-component models. The one-component model indicates that placebo patients improve across time (estimated slope = −0.32), but that drug patients improve more dramatically across time (estimated slope interaction = −0.65). The drug effect is nonsignificant which, in this model that includes the drug by time interaction, simply means that no differences are observed between the groups at week 0. There is a large degree of subject heterogeneity in terms of both the random intercept and slope parameters, with no appreciable association between these terms.

The two-component model indicates an approximate 50:50 split in terms of the subjects with estimates for the first group as $\hat{\boldsymbol{\mu}}_1 = [5.36, -0.01]$ and for the second group as $\hat{\boldsymbol{\mu}}_2 = [5.32, -0.95]$. Thus, there is little difference in patients in terms of the intercept, but a great deal of difference in terms of the slopes. We interpret the first group (with a slope equal to approximately zero) to represent nonresponders and the second group (with a large negative slope) to represent responders. The estimates of the random-effects variance terms reveal that the two-component model explains a great deal of the heterogeneity in subject change across time, as the slope variance is dramatically reduced in the two-component relative to the one-component model. In terms of the the drug effect, while there is a non-significant difference for the intercept, a significant negative drug by slope interaction is observed. It's interesting to note that the mixture component shift in terms of the slope ($-0.94 = -0.95 + 0.01$) is more pronounced than the drug effect on the slope (−0.52). This suggests that there are patients in both the placebo and drug groups that respond to treatment. This agrees with the notion in psychiatry of the existence of placebo responders.

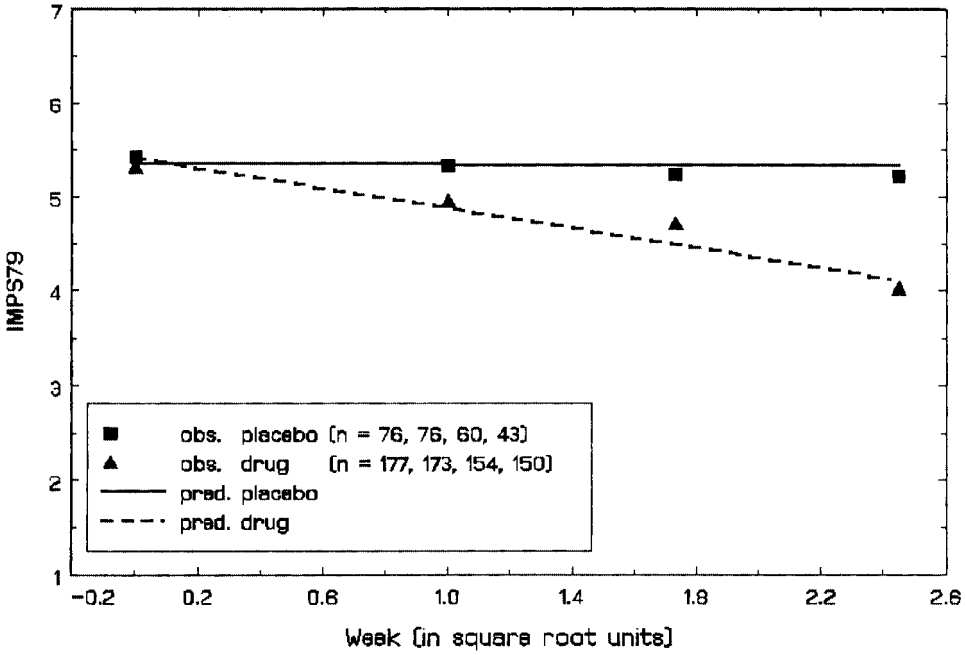

Figure 3a. Schizophrenia data: IMPS79 scores across time by treatment mixture model observed and predicted group means: Component 1.

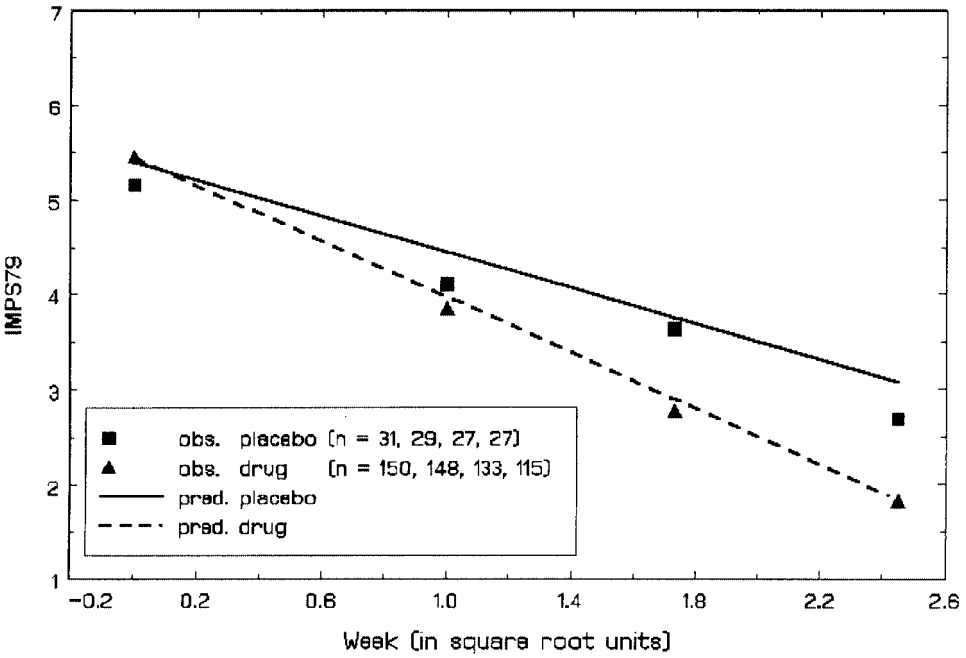

Figure 3b. Schizophrenia data: IMPS79 scores across time by treatment mixture model observed and predicted group means: Component 2.

Using the aforementioned procedure to determine mixture group membership, patients can be classified as belonging to either mixture group 1 (''nonresponders'') or 2 (''responders''). Figures 3a and 3b depict the observed and estimated drug and placebo patient group means for the first and second mixture component groups, respectively. The plots show good model fit and clearly illustrate the marked difference in slopes between responders (Figure 3b) and nonresponders (Figure 3a). It is reassuring that the model fits the observed means equally well for both responders and nonresponders because an assumption of the model is that the drug effect is the same among these two mixture groups. These figures lend support to this model assumption. Finally, it is interesting to note that 29% of the placebo patients (31 of 108) are classified as responders, whereas 46% of the drug patients are responders (151 of 329). This difference is statistically significant ($\chi^2_1 = 9.9$, $p < 0.002$), and so, while there is evidence that some patients do respond to placebo, the proportion of drug responders is clearly greater.

4. DISCUSSION

The random-effects regression model presented in this article, which allows the coefficients of the random terms to have a mixture multivariate normal distribution, provides a generalization that is able to accommodate data sampled from heterogeneous populations. The model is specified for a multicomponent multivariate mixture distribution, of which the one-component case represents the usual random-effects regression model described by Laird and Ware (1) and others. Maximum marginal likelihood methods are used to estimate the model parameters. The likelihood solution of the mixture random-effects model is presented both in terms of the EM algorithm and the Fisher scoring solution.

An issue in use of mixture random-effects regression models is model testing. Unfortunately, the traditional chi-square approximation to the likelihood ratio test for a multicomponent versus single-component model is not straightforward for the mixture case because the standard regularity conditions on which the asymptotic theory is based may fail. Specifically, if $\boldsymbol{\Theta}$ (π_1, in the mixture case) denotes the parameter for H_a, then under H_0(: $\pi_1 = 1$), the true $\boldsymbol{\Theta}_o$ ($\pi_1 = 1$) lies on the boundary of the parameter space, as demonstrated by Titterington et al. [(13), section 5.4]. If the mixture population parameters π_k can be specified a priori, then the corresponding negative two times log ratio function will, under H_0, be asymptotically chi-square with dim($\boldsymbol{\theta}_1$) degrees of freedom (i.e., the difference between the number of parameters in the mixture model and a single-component model). However, in practice, it is fairly difficult to specify π_k value in the population a priori. For this and other reasons, Verbecke and colleagues (9,34) recommend that goodness-of-fit tests and likelihood-ratio tests lack the proper mathematical foundation for use in mixed-effects mixture models. Alternatively, results of Self and Liang (40) on approximate tests for nonstandard testing situations

could be examined in this context. This would be similar to the approach taken by Stram and Lee (41) in their use of Self and Liang's results for likelihood-ratio testing of variance components in the longitudinal mixed-effects model. Clearly, more work is needed in this area.

It is important to note that although we have illustrated the model in terms of two-component mixture distributions for models with two random effects, in fact, the general model could allow multiple components and multiple random effects. In some cases, the amount of data may limit the numbers of components and/or random effects that can be estimated. For the examples presented here, two components provided reasonable results and helped to identify patients who were treatment responders versus nonresponders.

While the usefulness of the proposed model has been illustrated, further research is clearly needed to further extend this model. For example, one important extension would be to incorporate mixtures in random-effects models for categorical outcomes. Because categorical outcomes are common in many areas of applied research, this extension would have some practical merit. Such work is currently underway.

APPENDIX: FISHER SCORING SOLUTION

In order to derive the Fisher scoring procedure for the mixture case, both the first derivatives and the Fisher information matrix need to be obtained. For the first derivatives, denoting φ as the vector containing all unknown parameters and $P(\boldsymbol{y}_i, \varphi) = \sum_{k=1}^{c} \pi_k f_k(\boldsymbol{y}_i|\boldsymbol{\theta}_k)$, the log-likelihood function for incomplete-data

$$\tilde{\ell}(\varphi) = \sum_{i=1}^{N} \log P(\boldsymbol{y}_i, \varphi) \tag{20}$$

is differentiated with respect to the unknown parameters. For the mixture probabilities π_l ($l = 1, \ldots, c - 1$) the first derivative equals

$$\frac{\partial \tilde{\ell}(\varphi)}{\partial \pi_l} = \sum_{i=1}^{N} \frac{1}{P(\boldsymbol{y}_i, \varphi)} [f_l(\boldsymbol{y}_i|\boldsymbol{\theta}_l)] - f_c(\boldsymbol{y}_i|\boldsymbol{\theta}_c)]. \tag{21}$$

Similarly, we obtain

$$\frac{\partial \tilde{\ell}(\varphi)}{\partial \sigma^2} = \frac{1}{2\sigma^4} \left[\sum_{i=1}^{N} \sum_{k=1}^{c} w_{ik} \mathrm{tr}(\boldsymbol{X}_i \Sigma_{\boldsymbol{\beta}_i|\boldsymbol{y}_i, \boldsymbol{\mu}_k} \boldsymbol{X}_i' + \boldsymbol{u}_{ik} \boldsymbol{u}_{ik}') - \sum_i n_i \sigma^2 \right] \tag{22}$$

$$\frac{\partial \tilde{\ell}(\varphi)}{\partial \alpha} = \frac{1}{\sigma^2} \sum_{i=1}^{N} \sum_{k=1}^{c} w_{ik} \boldsymbol{Z}_i' (\boldsymbol{y}_i - \boldsymbol{Z}_i \boldsymbol{\alpha} - \boldsymbol{X}_i \boldsymbol{\beta}_{ik}) \tag{23}$$

$$\frac{\partial \tilde{\ell}(\varphi)}{\partial(\text{vech}\boldsymbol{\Sigma})'} = \frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{c} w_{ik}\text{vec}'[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}_{\boldsymbol{\beta}_i|\mathbf{y}_i,\boldsymbol{\mu}_k} + (\hat{\boldsymbol{\beta}}_{ik} - \boldsymbol{\mu}_k)(\hat{\boldsymbol{\beta}}_{ik} - \boldsymbol{\mu}_k)' - \boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}]\boldsymbol{D}_q \tag{24}$$

$$\frac{\partial \tilde{\ell}(\varphi)}{\partial \boldsymbol{\mu}_k} = \sum_{i} w_{ik}\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\beta}}_{ik} - \boldsymbol{\mu}_k) \tag{25}$$

Here, the duplication matrix $\boldsymbol{D}_q$ in Eq. (24) is the $q^2 \times (1/2)q(q+1)$ matrix with the property that $\boldsymbol{D}_q\,\text{vech}(\mathbf{A}) = \text{vec}\mathbf{A}$ for every symmetric $q \times q$ matrix $\mathbf{A}$ (39).

For the information matrix, $I_c(\varphi)$ is defined as the Fisher information matrix for *complete* data, namely,

$$I_c(\varphi) = E[\mathcal{D}_\varphi(\ell(\varphi))\mathcal{D}_\varphi(\ell(\varphi))'] \tag{26}$$

where $\ell(\varphi)$ denotes the log-likelihood function in Eq. (7), and $\mathcal{D}_\varphi(\cdot)$ denotes the first derivative operator with respect to φ. $I_c(\varphi)$ is used rather than $I(\varphi)$ (information matrix for *incomplete* data) because they yield near-identical results and $I_c(\varphi)$ is usually much easier to evaluate and invert than $I(\varphi)$ [Titterington et al. (13), pp. 208–211]. For convenience, let $\ell(\varphi) = \ell_1(\varphi) + \ell_2(\varphi)$, where $\ell_1(\varphi) = \sum_{i=1}^{N} \boldsymbol{d}_i'\boldsymbol{U}(\boldsymbol{\pi})$ and $\ell_2(\varphi) = \sum_{i=1}^{N} \boldsymbol{d}_i'\boldsymbol{V}_i(\theta)$. Here, $\boldsymbol{d}_i$ is a $c \times 1$ indicator vector with all zero elements except one; the sole element equal to 1 indicates the appropriate subpopulation to which observation i belongs. Assuming independence of subjects, we have

$$E[\mathcal{D}_\varphi(\ell_1)\mathcal{D}_\varphi(\ell_1)'] = -E[\mathcal{D}_\varphi^2(\ell_1)] = -\sum_{i=1}^{N}\sum_{k=1}^{c} E[d_{ik}\mathcal{D}_\varphi^2(\log \pi_k)] \tag{27}$$

and

$$E[\mathcal{D}_\varphi(\ell_2)\mathcal{D}_\varphi(\ell_2)'] = \sum_{i=1}^{N}\sum_{k=1}^{c} \pi_k E[\mathcal{D}_\varphi(\log f_k(\mathbf{y}_i|\boldsymbol{\theta}_k))\mathcal{D}_\varphi(\log f_k(\mathbf{y}_i|\boldsymbol{\theta}_k))']. \tag{28}$$

Noticing that $\mathcal{D}_\varphi(\log \pi_k)$ is not a random variable, we further obtain

$$E[\mathcal{D}_\varphi(\ell_1)\mathcal{D}_\varphi(\ell_2)'] = \sum_{i=1}^{N}\sum_{k=1}^{c} \pi_k \mathcal{D}_\varphi(\log \pi_k) E[\mathcal{D}_\varphi(\log f_k(\mathbf{y}_i|\boldsymbol{\theta}_k))']. \tag{29}$$

Using these results, the information matrix is derived and formed as:

$$I_c = \left[\begin{array}{c|cc|cc} I(\boldsymbol{\pi}) & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & I(\boldsymbol{\mu}) & I(\boldsymbol{\mu},\boldsymbol{\alpha}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I(\boldsymbol{\alpha},\boldsymbol{\mu}) & I(\boldsymbol{\alpha}) & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & I(\boldsymbol{\Sigma}) & I(\boldsymbol{\Sigma},\sigma^2) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & I(\sigma^2,\boldsymbol{\Sigma}) & I(\sigma^2) \end{array}\right] \tag{30}$$

where the $I(\pi_l)$ and $I(\pi_k, \pi_l)$ are the llth diagonal elements and the klth off-diagonal elements of $I(\boldsymbol{\pi})$, respectively, and $I(\boldsymbol{\mu})$ is a block-diagonal matrix with the kth block $I(\boldsymbol{\mu}_k)$. Notice that since the off-diagonal blocks of the information matrix are zero, it can be partitioned into three submatrices that can be inverted separately. For the nonzero partitions of the information matrix, we get

$$I(\pi_l) = -\sum_{i=1}^{N} E\left[\mathcal{D}_{\pi_l}\left(\frac{d_{il}}{\pi_l} - \frac{d_{ic}}{\pi_c}\right)\right] = N\left(\frac{1}{\pi_l} + \frac{1}{\pi_c}\right) \quad (l < c), \tag{31}$$

$$I(\pi_k, \pi_l) = -\sum_{i=1}^{N} E\left[\mathcal{D}_{\pi_l}\left(\frac{d_{il}}{\pi_l} - \frac{d_{ic}}{\pi_c}\right)\right] = \frac{N}{\pi_c} \quad (k \neq l \text{ and } k, l < c), \tag{32}$$

$$I(\boldsymbol{\mu}_j, \boldsymbol{\mu}_k) = \sum_{i=1}^{N}\sum_{l=1}^{c} \pi_l \Sigma^{-1} E[\hat{\beta}_{il} - \mu_l)\delta_{lj}(\hat{\beta}_{il} - \mu_l)'\delta_{lk}]\Sigma^{-1}, \tag{33}$$

where $\delta_{lk} = 1$ if $l = k$, and 0 otherwise. Denoting $\otimes$ as the Kronecker (direct) product, we also get

$$I(\boldsymbol{\Sigma}) = \frac{1}{2}\boldsymbol{D}_q'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\sum_{i=1}^{N}\sum_{k=1}^{c} \pi_k(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{\boldsymbol{\beta}_i|\mathbf{y}_i,\boldsymbol{\mu}_k}) \otimes (\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{\boldsymbol{\beta}_i|\mathbf{y}_i,\boldsymbol{\mu}_k})(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\boldsymbol{D}_q \tag{34}$$

$$I(\sigma^2) = \frac{1}{2}\sigma^{-8}\sum_{i=1}^{N}\sum_{k=1}^{c} \pi_k \mathrm{tr}(\sigma^2 \boldsymbol{I}_i - \boldsymbol{X}_i \boldsymbol{\Sigma}_{\boldsymbol{\beta}_i|\mathbf{y}_i,\boldsymbol{\mu}_k}\boldsymbol{X}_i')^2 \tag{35}$$

$$I(\boldsymbol{\alpha}) = \sigma^{-4}\sum_{i=1}^{N}\sum_{k=1}^{c} \pi_k(\boldsymbol{Z}_i'(\sigma^2 \boldsymbol{I}_i - \boldsymbol{X}_i \boldsymbol{\Sigma}_{\boldsymbol{\beta}_i|\mathbf{y}_i,\boldsymbol{\mu}_k}\boldsymbol{X}_i')\boldsymbol{Z}_i') \tag{36}$$

$$I(\boldsymbol{\alpha}, \boldsymbol{\mu}_k) = \frac{1}{\sigma^2}\sum_{i=1}^{N} \pi_k \boldsymbol{Z}_i'\boldsymbol{X}_i \boldsymbol{\Sigma}_{\boldsymbol{\beta}_i|\mathbf{y}_i,\boldsymbol{\mu}_k}\boldsymbol{\Sigma}^{-1} \tag{37}$$

$$I(\sigma^2, \boldsymbol{\Sigma}) = \frac{1}{2}\sigma^{-4}\boldsymbol{D}_q'\sum_{i=1}^{N}\sum_{k=1}^{c} \pi_k \mathrm{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}_{\boldsymbol{\beta}_i|\mathbf{y}_i,\boldsymbol{\mu}_k}\boldsymbol{X}_i'\boldsymbol{X}_i\boldsymbol{\Sigma}_{\boldsymbol{\beta}_i|\mathbf{y}_i,\boldsymbol{\mu}_k})\boldsymbol{\Sigma}^{-1} \tag{38}$$

REFERENCES

1. Laird, N.M.; Ware, J.H. Random effects models for longitudinal data. Biometrics **1982**, *38*, 963–974.
2. Bock, R.D. The discrete Bayesian. In *Modern Advances in Psychometric Research*, H. Wainer and S. Messick, Eds. Earlbaum: Hillsdale, N.J., 1983.
3. Jennrich, R.I.; Schluchter, M.D. Unbalanced repeated-measures models with structured covariance matrices. Biometrics **1986**, *42*, 805–820.

4. Longford, N.T.A. 'A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested effects', Biometrika **1987**, *74*, 817–827.
5. Lindstrom, M.J.; Bates, D.M. Newton-Raphson and EM algorithms for linear mixed models with for repeated-measures data. Journal of the American Statistical Association **1988**, *83*, 1014–1022.
6. Younossi, Z.M.; Mullen, K.D.; Zakko, W.; Hodnick, S.; Brand, E.; Barnes, D.S.; Carey, W.D.; McCullough, A.C.; Easley, K.; Boparai, N.; Gramlich, T. A randomized, double-blind controlled trial of interferon alpha-2b and ribavirin vs. interferon alpha-2b and amantadine for treatment of chronic hepatitis C non-responder to interferon monotherapy. Journal of Hepatology **2001**, *34* (1), 128–133.
7. Guerret, S.; Desmouliere, A.; Chossegros, P.; Costa, A.M.; Badid C.; Trepo, C.; Grimaud, J.A.; Chevallier, M. Long-term administration of interferon-alpha in non-responder patients with chronic hepatitis C: follow-up of liver fibrosis over 5 years. J. Viral Hepat. **1999**, *6* (2), 125–133.
8. Xu, W. Mixtures in Random-Effects Regression Models; Unpublished dissertation, 1995.
9. Verbeke, G.; Lesaffre, E. A Linear Mixed-Effects Model With Heterogeneity in the Random-Effects Population. Journal of the American Statistical Association **1996**, *91*, 217–221.
10. Frank, E.; Prien, R.F.; Jarrett, R.B.; Keller, M.B.; Kupfer, D.J.; Lavori, P.W.; Rush, A.J.; Weissman, M.M. Conceptualization and rationale for consensus definitions of terms in major depressive disorder—remission, recovery, relapse, and recurrence. Archives of General Psychiatry **1991**, *48*, 851–855.
11. Thase, M.E. Defining remission in patients treated with antidepressants. Journal of Clinical Psychiatry **1999**, *60* (Suppl 22), 3–6.
12. Cramer, J.; Rosenheck, R.; Xu, W.; Henderson, W.; Thomas, J.; Charney, D. Detecting improvement in quality of life and symptomatology in schizophrenia. Schizophrenia Bulletin **2001**, *27*, 227–234.
13. Titterington, D.M.; Smith, A.F.M.; Makov, U.E. *Statistical Analysis of Finite Mixture Distributions*. John Wiley & Sons: Chichester, 1985.
14. Everitt, B.S.; Hand, D.J. *Finite Mixture Distributions*. Chapman and Hall: London, 1981.
15. McLachlan, G.J.; Basford, K.E. *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker, Inc. New York, 1988.
16. McLachlan, G.J.; Peel, D. *Finite Mixture Models*. John Wiley & Sons, New York, 2000.
17. Titterington, D.M. Some Recent Research in the Analysis of Mixture Distributions. Statistics **1990**, *21* (4), 619–641.
18. Belin, T.R.; Rubin, D.B. The analysis of repeated-measures data on schizophrenic reaction times using mixture models. Statistics in Medicine **1995**, *14*, 747–768.
19. Muthén, B.; Shedden, K. Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics **1999**, *55*, 463–469.
20. Muthén, B. Methodological issues in random coefficient growth modeling using a latent variable framework: Applications to the development of heavy drinking. In *Multivariate Applications in Substance Use Research*, Rose, J.; Chassin, L.; Presson, C.; and Sherman, J.; Eds., Erlbaum: Hillsdale, N.J., 2000; 113–140.
21. Muthén, B. Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent

growth modeling. In *New Methods for the Analysis of Change*, Collins, L.M., and Sayer, A., Eds.; APA: Washington, D.C., 2001; 291–322.
22. Muthén, B.; Muthén, L. Integrating person-centered and variable-centered analysis: growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, **2000**, *24*, 882–891.
23. Muthén, B.; Brown, C.H.; Masyn, K.; Jo, B.; Khoo, S.T.; Yang, C.C.; Wang, C.P.; Kellam, S.; Carlin, J.; Liao, J. General growth mixture modeling for randomized preventive interventions. Biostatistics; *in press.*
24. Muthén, B.; Muthén, L. *Mplus User's Guide, Version 2*. Muthén & Muthén: Los Angeles, Calif., 2000.
25. Aitkin, M. A general maximum likelihood analysis of variance components in generalized linear models. Biometrics **1999**, *55*, 117–128.
26. Chi, E.M.; Reinsel, G.C. Models for longitudinal data with random effects and AR(1) errors. Journal of the American Statistical Association **1989**, *84*, 452–459.
27. McCulloch, C.E. Symmetric matrix derivations with application. JASA **1982**, *77*, 679–682.
28. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm (with discussion). J. R. Statist. Soc. B **1977**, *39*, 1–38.
29. Bock, R.D. Measurement of human variation: A two-stage model. In *Multilevel Analysis of Educational Data*; R.D. Bock, Ed.; Academic Press: New York, 1989.
30. Riesby N.; Gram, L.F.; Bech, P.; Nagy, A.; Petersen, G.O.; Ortmann, J.; Ibsen, I.; Dencker, S.J.; Jacobsen, O.; Kruautwald, O.; Sondergaard, I.; Christiansen, J. Imipramine: Clinical effects and pharmacokinetic variability. Psychopharmacology **1977**, *54*, 263–272.
31. Goldstein, H. *Multilevel Statistical Models*, 2nd Ed.; Halstead Press: New York, 1995.
32. Bryk, A.S.; Raudenbush, S.W. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publication: Newbury Park, Calif., 1992.
33. Burstein, L.; Linn, R.L.; Capell, I. Analyzing multi-level data in the presence of heterogeneous within-class regressions. Journal of Educational Statistics **1978**, *4*, 347–389.
34. Verbeke, G.; Molenberghs, M. *Linear Mixed Models for Longitudinal Data*. Springer: New York, 2000.
35. Lorr, M., Klett, C.J. *Inpatient Multidimensional Psychiatric Scale: Manual*. Consulting Psychologists Press: Palo Alto, Calif., 1966.
36. Goodman, L.S.; Gilman, A., Eds., *The Pharmacological Basis of Therapeutics*, 5th Ed., Macmillan: New York, 1975.
37. Gibbons, R.D.; Hedeker, D.; Waternaux, C.; Davis, J.M. Random regression models: a comprehensive approach to the analysis of longitudinal psychiatric data. Psychopharmacology Bulletin **1988**, *24*, 438–443.
38. Hedeker, D.R; Gibbons, R.D. Application of random-effects pattern-mixture models for missing data in longitudinal studies. Psychological Methods **1997**, *2*, 64–78.
39. Magnus, J.R. *Linear Structures*. Charles Griffin: London, 1988.
40. Self, S.G.; Liang, K.-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. Journal of the American Statistical Association, **1987**, *82*, 605–610.

41. Stram, D.O.; Lee, J.W. Variance components testing in the longitudinal mixed effects model. Biometrics **1994**, *50*, 1171–1177.

Received January 2001
Revised September 2001
Accepted October 2001