Model Building for Nonlinear Mixed-Effects Models

José C. Pinheiro and **Douglas M. Bates**
Department of Biostatistics   Department of Statistics
University of Wisconsin — Madison

Nonlinear mixed-effects models involve both fixed effects and random effects. Model building for nonlinear mixed-effects models is the process of determining the characteristics of both the fixed and the random effects so as to give an adequate but parsimonious model. We describe procedures based on information criterion statistics for comparing different structures of the random effects component. These include procedures for determining which parameters in the model should be mixed effects and which should be purely fixed effects, as well as procedures for modeling the dependence of parameters on individual-specific covariates. We also consider procedures for assessing the variability in the parameter estimates, based on asymptotic theory and on likelihood profiling techniques. These methods are illustrated using the nonlinear mixed-effects methods and classes for S-plus and using data sets from forestry and pharmacokinetics studies.

KEY WORDS: Mixed effects, random effects, nonlinear models, maximum likelihood.

1. INTRODUCTION

Nonlinear mixed-effects models have recently received a great deal of attention in the statistical literature because they are flexible models for unbalanced repeated measures data that occurs in pharmacokinetics, economics, biology, and other areas.

Model building techniques for mixed-effects models involve questions that do not have a parallel with conventional, fixed-effects, linear and nonlinear models. Some of these issues are:

- determining which effects should have an associated random component and which should be purely fixed;
- using covariates to explain inter-individual parameter variability;
- using structured random effects variance-covariance matrices (e.g. diagonal matrices) to reduce the number of parameters in the model.

We consider here strategies for addressing these questions in the context of nonlinear mixed-effects models, though most of the techniques described are also applicable to linear mixed-effects models.

Any model building strategy is by nature iterative: a tentative model is initially fitted and modified to generate possibly better models (according to some goodness-of-fit criterion) then the process is repeated until no further improvement is possible. In comparing alternative models one must also analyze the residuals from the fit, checking for departures from the assumptions in the model. It is highly recommended that any model building analysis be done in conjunction with experts in the field of application of the model, to ensure the practicality of the chosen model.

Once a model has been chosen to represent the data, measures of variability for the estimates and confidence regions on the parameters in the model are usually needed for statistical inference purposes. We consider methods for assessing the variability in the parameter estimates based on maximum likelihood asymptotic theory and likelihood profiling (Bates and Watts, 1988).

2. A NONLINEAR MIXED-EFFECTS MODEL

Several different nonlinear mixed-effects models have been proposed (Sheiner and Beal, 1980; Mallet, Mentre, Steimer and Lokiek, 1988; Lindstrom and Bates, 1990; Vonesh and Carter, 1992; Davidian and Gallant, 1992; Wakefield, Smith, Racine-Poon and Gelfand, 1994). We consider here a slightly modified version of the model proposed in Lindstrom and Bates (1990). This model can be viewed as a two-stage model that in some ways generalizes both the linear mixed-effects model of Laird and Ware (1982) and the usual nonlinear model for independent data (Bates and Watts, 1988). In the first stage the jth observation on the ith individual is modeled as

$$y_{ij} = f(\boldsymbol{\phi}_{ij}, \boldsymbol{x}_{ij}) + \epsilon_{ij}, \quad i = 1, \ldots, M,\ j = 1, \ldots, n_i \tag{2.1}$$

where f is a nonlinear function of an individual-specific parameter vector $\boldsymbol{\phi}_{ij}$ and the predictor vector $\boldsymbol{x}_{ij}$, ϵ_{ij} is a normally distributed noise term, M is the total number of individuals, and n_i is the number of observations on the ith individual. In the second stage the individual-specific parameter vector is modeled as

$$\boldsymbol{\phi}_{ij} = \boldsymbol{A}_{ij}\boldsymbol{\beta} + \boldsymbol{B}_{ij}\boldsymbol{b}_i, \quad \boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{D}), \tag{2.2}$$

where $\boldsymbol{\beta}$ is a p-dimensional vector of fixed population parameters, $\boldsymbol{b}_i$ is a q-dimensional random effects vector associated with the ith individual (not varying with j), $\boldsymbol{A}_{ij}$ and $\boldsymbol{B}_{ij}$ are design matrices for the fixed and random effects respectively, and $\sigma^2\boldsymbol{D}$ is a (general) variance-covariance matrix. It is further assumed that observations made on different individuals are independent and that the $\boldsymbol{\epsilon}_i$ follow a $\mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ distribution and are independent of the $\boldsymbol{b}_i$.

The parameters in the model are estimated either by maximum likelihood, or by restricted maximum likelihood, based on the marginal density of $\boldsymbol{y}$. In general this density does not have a

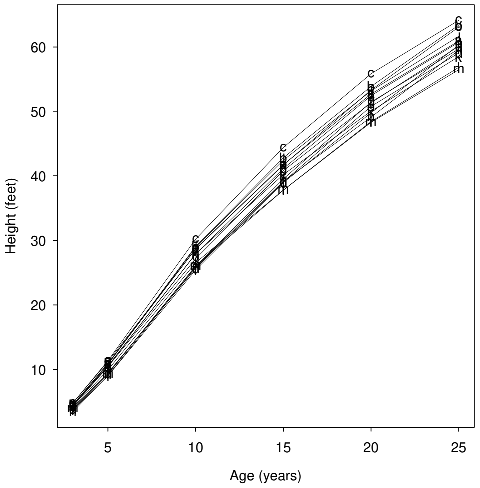

Figure 1: Height versus age for 14 Loblolly pine trees grown under similar conditions.

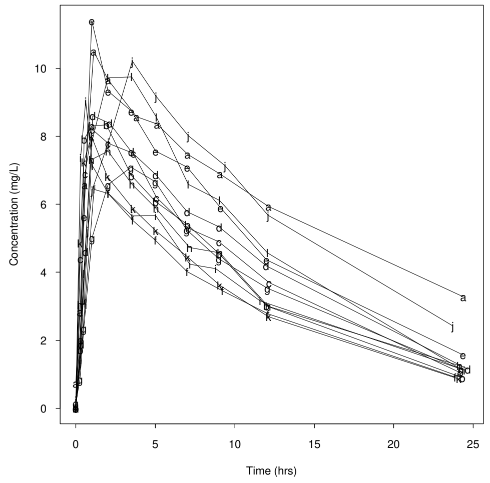

Figure 2: Serum concentrations of Theophylline in twelve volunteers after oral dose.

closed-form expression when the model function f is nonlinear in $\boldsymbol{b}$, so different approximations have been proposed for estimating it. We adopt here the approximation suggested in Lindstrom and Bates (1990). Pinheiro and Bates (1995a) analyze several approximations to the log-likelihood of nonlinear mixed-effects models and conclude that Lindstrom and Bates' approximation usually gives very accurate results.

3. EXAMPLES

We now introduce the data sets that will be used throughout this paper.

3.1 Pine Trees

Data on the growth of Loblolly pine trees were described in Kung (1986). The height of these 14 trees, planted in the southern United States, were measured at 3, 5, 10, 15, 20, and 25 years of age. Figure 1 shows a plot of these data.

Kung (1986) used a logistic curve to model the trees' growth, but an asymptotic regression model seems to explain the observed growth pattern better. We also tried the logistic model, the Gompertz model, the Morgan, Mercer, and Flodin model, and the Weibull type model (Ratkowsky, 1990), but the asymptotic regression gave the best overall fit. This model can be expressed as

$$f(t, \boldsymbol{\phi}) = \phi_1 - \phi_2 \exp(-\phi_3 t) \tag{3.1}$$

where t denotes the tree's age, ϕ_1 the asymptotic height, ϕ_2 the difference between ϕ_1 and the height at age zero, and ϕ_3 the growth rate.

3.2 Theophylline

These data are provided by Dr. Robert A. Upton of the University of California, San Francisco. Theophylline was administered orally to 12 individuals whose serum concentrations were measured at 11 times over the next 25 hours. The data, presented in Figure 2, are an example of a laboratory pharmacokinetic study characterized by many observations on each of a moderate number of individuals.

A common model for such data is a first order compartment model with absorption in a peripheral compartment. It can be written

$$C_t = \frac{DKk_a}{Cl(k_a - K)} [\exp(-Kt) - \exp(-k_a t)] \tag{3.2}$$

where C_t is the predicted concentration at time t (mg/L), t is the time (hr), D is the dose (mg), Cl is the clearance (L), K is the elimination rate constant (1/hr), and k_a is the absorption rate constant (1/hr). To ensure positivity of the rate constants and the clearance, the logarithms of these quantities can be used in (3.2), giving the reparametrized model

$$\begin{aligned} C_t = & \frac{D \exp(lk_a + lK - lCl)}{\exp(lk_a) - \exp(lK)} \\ & \times \{\exp[-\exp(lK)\, t] - \exp[-\exp(lk_a)\, t]\} \end{aligned} \tag{3.3}$$

where $lCl = \log(Cl)$, $lk_a = \log(k_a)$, and $lK = \log(K)$.

3.3 Quinidine

The last data set comes from a pharmacokinetics clinical study of the antiarrhytimic drug quinidine. A total of 361 quinidine concentration measurements were made on 136 hospitalized patients under varying dosage regimens. Additional data were collected on a set of nine covariates: age, height, weight, race, smoking status, ethanol abuse, congestive heart failure, creatinine clearance, and α-1-acid glycoprotein concentration. Some of these covariates, such as age, weight or creating clearance, were "time-varying". That is, the value of these covariates for a particular patient could change during the course of the study. Others, such as race, would necessarily be constant. One of the main objectives of the study was to investigate relationships between the individual pharmacokinetic parameters and the covariates. A full description of the data can be found in Verme, Ludden, Clementi and Harris (1992). Statistical analyses of these data using alternative modeling approaches are given in Davidian and Gallant (1993) and in Wakefield (1993).

The model that has been suggested for the quinidine data is the one-compartment open model with first-order absorption. This model can be defined recursively as follows.

Suppose that, at time t, the patient receives a dose d_t and prior to that time the last dose was given at time t'. The expected concentration in the serum compartment, C_t, and in the absorption compartment, Ca_t, are given by

$$\begin{aligned} C_t &= C_{t'} \exp\left[-K\left(t-t'\right)\right] + \frac{Ca_{t'} k_a}{k_a - K} \\ &\quad \times \left\{\exp\left[-K\left(t-t'\right)\right] - \exp\left[-k_a\left(t-t'\right)\right]\right\} \\ Ca_t &= Ca_{t'} \exp\left[-k_a\left(t-t'\right)\right] + \frac{d_t}{V} \end{aligned} \tag{3.4}$$

where V is the apparent volume of distribution, k_a is the absorption rate constant, and K is the elimination rate constant.

When a patient receives the same dose d at regular time intervals Δ, model (3.4) converges to the steady state model

$$\begin{aligned} C_t &= \frac{d\, k_a}{V\left(k_a - K\right)} \\ &\quad \times \left[\frac{1}{1-\exp\left(-K\Delta\right)} - \frac{1}{1-\exp\left(-k_a\Delta\right)}\right] \\ Ca_t &= \frac{d}{V\left[1-\exp\left(-k_a\Delta\right)\right]} \end{aligned} \tag{3.5}$$

Finally, for a between-dosages time t, the model for the expected concentration C_t, given that the last dose was received at time t', is identical to (3.4).

Using the fact that the elimination rate constant K is equal to the ratio between the clearance (Cl) and the volume of distribution (V), we can reparametrize models (3.4) and (3.5) in terms of V, k_a, and Cl.

To ensure that the estimates of V, k_a, and Cl are positive, we can rewrite models (3.4) and (3.5) in terms of $lV = \log(V)$, $lk_a = \log(k_a)$ and $lCl = \log(Cl)$.

The initial conditions for the recursive model (3.4) and (3.5) are $C_0 = 0$ and $Ca_0 = d_0/V$, with d_0 denoting the first dose received by the patient. It has been assumed throughout the model's definition that the bioavailability of the drug – the percentage of the administered dose that reaches the measurement compartment – is equal to one.

4. VARIANCE-COVARIANCE MODELING

In this section we consider the questions of determining which parameters in the model should have a random component and whether the scaled variance-covariance matrix of the random effects ($\boldsymbol{D}$) can be structured in a simpler form with fewer parameters.

The first question that should be addressed in the analysis is choosing which parameters should be random effects and which purely fixed effects. An intuitive approach is to obtain separate fits for each individual and assess the inter-individual parameter variability by considering the individual confidence intervals. This approach requires sufficient observations on each individual to give meaningful parameter estimates in the individual fits. This is often not the case in clinical studies.

A more general approach is to fit different prospective models and compare nested models using likelihood ratio tests or information criterion statistics, such as the Akaike Information Criterion (AIC) (Sakamoto, Ishiguro and Kitagawa, 1986). One of the problems with this approach is deciding which way to construct the nesting: from smaller to larger models, or the other way around. Starting with a model where all parameters have associated random effects and then removing unnecessary terms is probably the best strategy, but may not be possible to implement if the full model is badly overparametrized. In these cases the variance-covariance matrix of the random effects may become seriously ill-conditioned, making convergence difficult or impossible. The smaller to larger approach is an alternative in these cases, but has the disadvantage of the large number of models that may have to be fit before the desired one is found. There is yet another important aspect that is overlooked by the model nesting approach: sometimes it is a linear combination of random effects being treated as fixed that gives the best model reduction.

The strategy we suggest for choosing the random effects to be included in the model is to start with all parameters as mixed effects, whenever no prior information about the random effects variance-covariance structure is available and convergence is possible. Then we examine the eigenvalues of the estimated $\boldsymbol{D}$ matrix, checking if one, or more, are close to zero. The associated eigenvector(s) would then give an estimate of the linear combination of the parameters that could be taken as fixed. We used the AIC to decide between alternative models.

Small eigenvalues may arise when the relative magnitude of the scales of the parameters in the model are quite different, without necessarily implying overparametrization. Therefore we suggest using a normalized, scale invariant version of the variance-covariance matrix. There are different ways of normalizing $\boldsymbol{D}$, the most common being the correlation matrix. This is not a particularly good choice in the present context, since all random effects would then have *normalized* variance equal to one and we would not be able to identify those with relatively small dispersion (which would be natural candidates to be dropped from the model). Whenever the $\boldsymbol{A}$ and $\boldsymbol{B}$ matrices in (2.2) are incidence-like matrices (i.e. with just one nonzero entry per row), a more convenient choice of normalization is the *coefficient of variation (CV)* matrix $\boldsymbol{D}_{CV}$ with

$$[\boldsymbol{D}_{CV}]_{ij} = \frac{[\boldsymbol{D}]_{ij}}{\left|\beta_{k(i)}\beta_{k(j)}\right|}$$

where β_k represents the kth fixed effect and $k(i)$, $k(j)$ represent the indices of the fixed effects associated with the ith and the jth random effects. When the nonzero elements of the ith row of $\boldsymbol{A}$ and $\boldsymbol{B}$ are equal to one, the ith diagonal element of $\boldsymbol{D}_{CV}$ is equal to the square of the coefficient of variation of ϕ_i. We note that, in the majority of applications of model (2.1), $\boldsymbol{A}$ and $\boldsymbol{B}$ will be incidence-like matrices.

To illustrate the use of the variance-covariance modeling methods we consider the examples described in section 3. We do not include any analyses of residuals here, but in all cases they did not indicate violations of the assumptions in the model. All maximum likelihood calculations in the examples were done using the `nlme` function in S-plus (Pinheiro and Bates, 1995b).

4.1 Pine Trees

Even though there are only six observation on each tree, we can obtain confidence intervals on the parameters in model (3.1) based on individual fits. Figure 3 gives the approximate 95% confidence intervals on ϕ_1, ϕ_2, and ϕ_3 for each tree. These symmetric confidence intervals are constructed from the standard er-

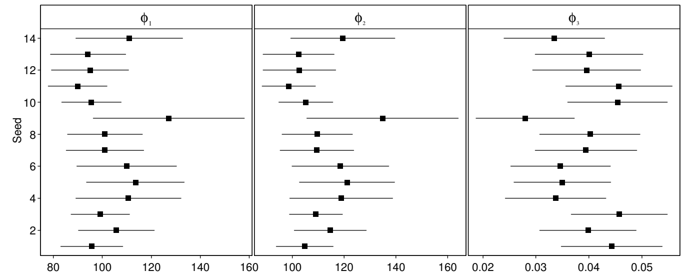

Figure 3: Approximate 95% confidence intervals on the parameters in the pine trees model (3.1), based on individual fits.

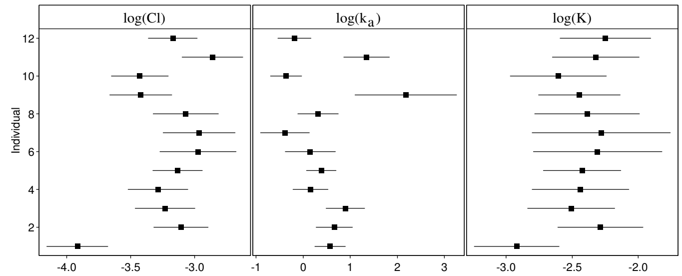

Figure 4: Approximate 95% confidence intervals for the parameters in the theophylline model (3.3), based on individual fits.

rors for the parameter estimates (Bates and Watts, 1988, Chapter 2).

We note that, for each parameter, all 14 confidence intervals overlap but these intervals are fairly wide due to the small number of observations available on each tree. In this case, the *individual fits* approach is not very helpful in indicating which parameters should be mixed effects and which should be purely fixed effects.

We consider now the second approach for choosing which parameters to include as mixed effects in the asymptotic regression model (3.1). We fit the model with each of ϕ_1, ϕ_2, and ϕ_3 a mixed effect, called model I. The resulting AIC is 243.62. The eigenvalues of $\widehat{\boldsymbol{D}}_{CV}$, $0.0419, 0.0026$, and 6.681×10^{-14}, give a clear indication of rank deficiency. The eigenvector corresponding to the smallest eigenvalue, converted back to the original scale of the random effects and normalized, is $(0.699, -0.711, -0.076)^T$, suggesting that the difference between the first two random effects can be considered nonrandom. This can be checked by reparametrizing model (3.1) as follows

$$y_{ij} = \phi_1' + (\phi_2' - \phi_1') \exp(-\phi_3' t_j) + e_{ij} \qquad (4.1)$$

where ϕ_1' and ϕ_3' continue to have the same interpretation as ϕ_1 and ϕ_3 in the previous parametrization, but $\phi_2' = \phi_1 - \phi_2$ now represents the height at age zero. Using this reformulation of model I, with only ϕ_1 and ϕ_3 as mixed effects, we get an AIC of 237.64, considerably better than before. This is model II.

The eigenvalues of the $\widehat{\boldsymbol{D}}_{CV}$ matrix corresponding to model II are 0.023 and 0.002, suggesting that no further reductions in the number of random effects can be attained. In fact, if we fit model (4.1) with only ϕ_1' or ϕ_3' as a mixed effect, we get AIC values of 239.49 and 240.30 respectively, both larger than that of model II. We also considered the model in which ϕ_1' and ϕ_3' were independent mixed effects (i.e. model II with a diagonal $\boldsymbol{D}$ matrix), but the resulting AIC, 238.94, indicated that model II was preferred.

4.2 Theophylline

The theophylline data give an example where there are enough observations on each individual to provide meaningful individual fits. Figure 4 gives the approximate 95% confidence intervals for the parameters in model (3.3), for each individual.

These confidence intervals indicate that the absorption rate constant k_a is the parameter with greatest inter-individual variability, followed by the clearance Cl. The elimination rate constant K does not appear to vary much with individual, as all the associated confidence intervals overlap. These results suggest that lk_a and lCl should be included as mixed effects in model (3.3), but lK should be included as a purely fixed effect.

We now consider the second, more general approach to choosing which parameters should be mixed effects. We were able to converge on parameter estimates for the model with all parameters as mixed effects. We call this set of estimates model I.

The AIC of model I is 366.64 and the $\widehat{\boldsymbol{D}}_{CV}$ matrix has eigenvalues $4.328, 0.019$, and 5.618×10^{-8} indicating that the model is probably overparametrized. The eigenvector corresponding to the smallest eigenvalue, converted back to the original scale of the random effects and normalized is $(0.464, 0.020, \ -0.886)^T$, suggesting that the lCl random effect is approximately equal to twice the lK random effect. Recalling that the volume of distribution (V) is equal to the ratio between Cl and K, we see that $lCl = 2lK$ implies that the ratio between V and K, that we will denote by R, is a fixed effect. The recommendation at this point would be to contact a pharmacologist and check the plausability of this finding, as well as the interpretability of the parameter R. Because this was not possible in our case, we decided to follow a more conservative approach and solve the overparametrization problem by removing one, or more, of the original random effects from the model.

The AIC of the models considering each of lCl, lk_a, and lK at a time as fixed effects are respectively $405.821, 436.789$, and 368.050. Even though these values are all larger than the AIC of model I, they give a clear indication that the elimination of the lK random effect has the smallest impact on the quality of the fit. This is consistent with the conclusions of the individual fits analysis discussed before.

The eigenvalues of the $\widehat{\boldsymbol{D}}_{CV}$ matrix for the model with lK as a purely fixed effect, called model II, are 5.000 and 0.013, indicating that the overparametrization is no longer present. The estimated correlation between the lCl and the lk_a random effects in model II is -0.110, suggesting that the two random effects are uncorrelated and a diagonal $\boldsymbol{D}$ can be used. The AIC of this model (III) is 366.05 indicating that it should be preferred over the previous model. No further reductions in the number of parameters in $\boldsymbol{D}$ can be obtained and we conclude that model III is preferred.

4.3 Quinidine

The quinidine data provide an example where there are insufficient observations on each individual to give meaningful individual fits. The data are characterized by few quinidine concentration measurements on many patients: for 46 patients (34%)

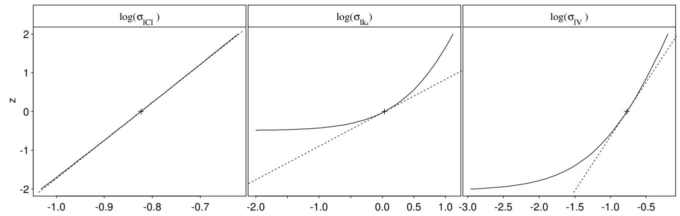

Figure 5: Likelihood profile z plots for the Quinidine model I.

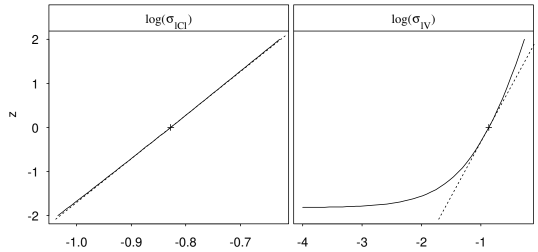

Figure 6: Likelihood profile z plots for the Quinidine model III.

there is only one observation and for 32 patients (24%) only two. The maximum number of observations on a single patient is 11 and only 6 patients (4.5%) have 7 or more observations.

Due to the sparsity of the concentration measurements across patients, the optimization of the log-likelihood for the model in which all parameters are mixed effects and the variance-covariance matrix of the random effects $\boldsymbol{D}$ is unstructured is a very ill-conditioned numerical problem. The optimizing algorithm alternates between equivalent solutions (in terms of the value of the log-likelihood) without ever converging.

Different strategies can be used to try to circumvent the non-convergence problem:

- try to achieve convergence using a diagonal $\boldsymbol{D}$ and examine the relative variability of the random effects, investigating the possibility of eliminating one, or more, of them from the model;
- force convergence (e.g. letting the algorithm run until a pre-established maximum number of iterations) and examine the corresponding $\widehat{\boldsymbol{D}}_{CV}$ matrix for rank deficiency;
- try to achieve convergence for models with fewer random effects.

We used the first of these strategies. The model with all parameters as mixed effects but with diagonal $\boldsymbol{D}$, called model I, converged and gave an AIC of 1006.97. The estimated standard deviations of the random effects in model I are $\widehat{\sigma}_{lCl} = 0.32$, $\widehat{\sigma}_{lk_a} = 0.67$, and $\widehat{\sigma}_{lV} = 0.34$. To investigate the possibility of eliminating random effects from the analysis, we must assess the variability in these estimates. We use, for this purpose, the likelihood profile z plots (Bates and Watts, 1988), included in Figure 5. The linear mixed-effects approximation to the log-likelihood function (Pinheiro and Bates, 1995a) is used in the profiling calculations.

The profile z plots in Figure 5 represent, with the solid line, a series of likelihood ratio tests converted to a standard normal z scale, for each parameter. If the parameter behaves according to the asymptotic approximation represented by the dashed lines, the two lines would coincide. Details on the calculation of the profile z are given in Appendix 2. The profile z plots in Figure 5 indicate that the fit becomes insensitive to arbitrarily small values of σ_{lk_a} and σ_{lV}. As $\sigma_{lk_a} \to 0$, $z\left(\sigma_{lk_a}\right) \to -0.5$, while as $\sigma_{lV} \to 0$, $z\left(\sigma_{lV}\right) \to -2$. This implies that as $\sigma_{lk_a} \to 0$, the p-value of the likelihood ratio test for this parameter approaches 0.62 and, as $\sigma_{lV} \to 0$, the p-value of the associated likelihood ratio test approaches 0.04. We conclude that lk_a should be treated as a purely fixed effect, but lV should be kept as a mixed effect.

The model with lk_a as a purely fixed effect and lCl and lV as mixed effects with unstructured variance-covariance matrix, called model II, converged with an AIC of 1005.34. The eigenvalues of the $\widehat{\boldsymbol{D}}_{CV}$ matrix, 0.031 and 0.006, do not indicate overparametrization problems. The estimated correlation coefficient between the two random effects in model II is $\widehat{\rho}_{lCl,lV} = 0.03$. This suggests that the random effects are not correlated and a diagonal matrix can be used to model $\boldsymbol{D}$. The AIC of this reduced model (III) is 1003.56, smaller than those of both models I and II. To investigate the possibility of eliminating either the lCl or the lV random effects from model III, we consider the likelihood profile z plots for the corresponding variance components, presented in Figure 6.

The profile z plot for $\log\left(\sigma_{lV}\right)$ indicates that the fit becomes insensitive to arbitrarily small values of this parameter. As $\sigma_{lV} \to 0$, $z\left(\sigma_{lV}\right) \to -1.81$, corresponding to a p-value for the associated likelihood ratio test of 0.07. This can be considered as borderline significant, making it unclear if the lV random effect should be removed from the model. This borderline significance is also indicated by an AIC of 1003.26 for the model with only lCl as a mixed effect. This is almost identical to the AIC of model III. We decided to follow a conservative approach and keep lV as a mixed effect in the covariate modeling analysis, presented next.

5. COVARIATE MODELING

In this section we consider the use of covariates to model random effects variability. This variability can either be related to *natural* inter-individual variation, or can be caused by differences in covariate values between and/or within individuals.

The first questions to be addressed in the covariate modeling process are the determination of which variables are potentially useful in explaining random effects variation and which random effects may have their variability explained by covariates. This is probably best achieved by analyzing plots of the random effects estimates versus the covariates, looking for trends and patterns. The conditional modes of the random effects (Lindstrom and Bates, 1990) were used here for this purpose.

After the candidate covariates have been chosen, a decision has to be made on how to test for their inclusion in the model. The number of extra parameters to be estimated tends to grow considerably with the inclusion of covariates and their associated random effects in the model. If the number of covariates/random effects combinations is large, we suggest using a *forward stepwise* approach in which covariates are included one at a time and the potential importance of the remaining covariates is graphically assessed at each step. The decision on whether or not to include a covariate can be based on the AIC of the fits with and without it.

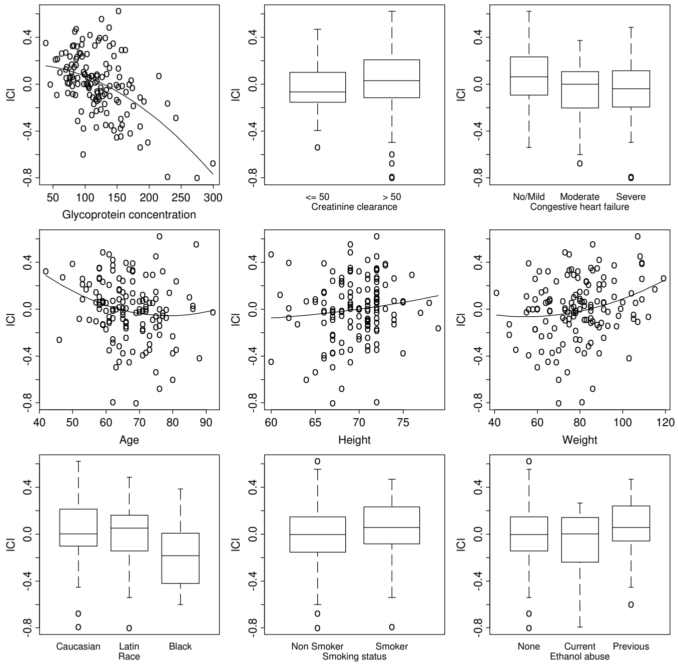

Figure 7: Conditional modes of the lCl random effect in model III versus available covariates.

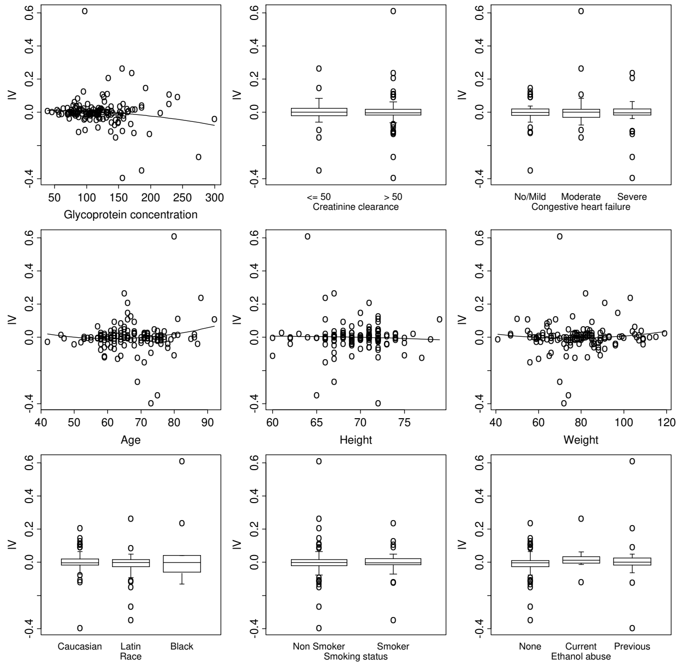

Figure 8: Conditional modes of the lV random effects in model III versus available covariates.

Another question that has to be addressed when including a covariate in the model, is which of the new parameters should be random or purely fixed. We suggest using an approach similar to the one described in section 4, for modeling the variance-covariance structure: whenever no prior information is available and convergence is possible, start with a saturated model (in which all new parameters are random) and, by examining the eigenstructure of the estimated $\boldsymbol{D}$ (or $\boldsymbol{D}_{CV}$) matrix, search for plausible structures with fewer parameters. We use the quinidine data to illustrate this model building approach. We reiterate that any model building strategy is not complete without a careful analysis of residuals and expert advice. In all examples considered here the residual analyses did not indicate departures from the model's assumptions.

5.1 Quinidine

Figure 7 presents the scatter plots of the conditional modes of the lCl random effect, based on model III of subsection 4.2, versus the available covariates. When the covariate value changed over time, the modal value was used to construct the plot. However, if the covariate is incorporated in the model, the parameter estimation uses the time-varying values. A *loess* smoother (Cleveland, Grosse and Shyu, 1992) is included on the plots of the continuous covariates to aid in visualizing possible trends.

Clearance appears to decrease with α-1-acid glycoprotein concentration and with age and to increase with weight and height. There is also some evidence that clearance decreases with severity of congestive heart failure and are smaller in Blacks than in both Caucasians and Latins. Clearly the α-1-acid glycoprotein concentration is the most important covariate for explaining the lCl inter-individual variation. A straight line seems adequate to model the observed relationship.

Figure 8 presents the scatter plots of the conditional modes of the lV random effect versus the available covariates. None of the covariates seems helpful in explaining the variability of this random effect and we did not pursue the modeling of its variability any further.

Initially, only the α-1-acid glycoprotein concentration was included in the model to explain the lCl random effect variation according to a linear model. In the notation of (2.2) this modification of models (3.4) and (3.5) is accomplished by writing

$$lCl_{ij} = (\beta_1 + b_{1i}) + (\beta_2 + b_{2i})\,\text{glycoprotein}_{ij}$$

This model IV treated all parameters as mixed effects, but the random effects associated with lCl were assumed independent of the lV random effect, preserving the covariance structure of model III. The AIC of this model was 879.27 indicating a considerable gain in goodness-of-fit when compared to model III (AIC of 1003.56). The analysis of the eigenstructure of the $\widehat{\boldsymbol{D}}_{CV}$ matrix did not indicate any overparametrization problems.

Figures 9 and 10 present the scatter plots of the conditional modes of the lCl random effects in model IV versus the available covariates. These plots indicate that the intercept random effect does not vary systematically with any of the covariates, but the slope random effect tends to increase with weight and height and is smaller among Blacks and patients with previous history of congestive heart failure. This suggests an interaction between the effects of these covariates and the α-1-acid glycoprotein on the quinidine clearance. At this point expert advice would be needed to clarify the plausability of this hypothesis. Since this was not possible here, we proceeded the model building analysis just for the purpose of illustrating the use of the proposed methodology.

Using the *forward stepwise* approach we included the interactions between α-1-acid glycoprotein and weight (as a linear predictor), race (as an indicator variable of black/not black status),

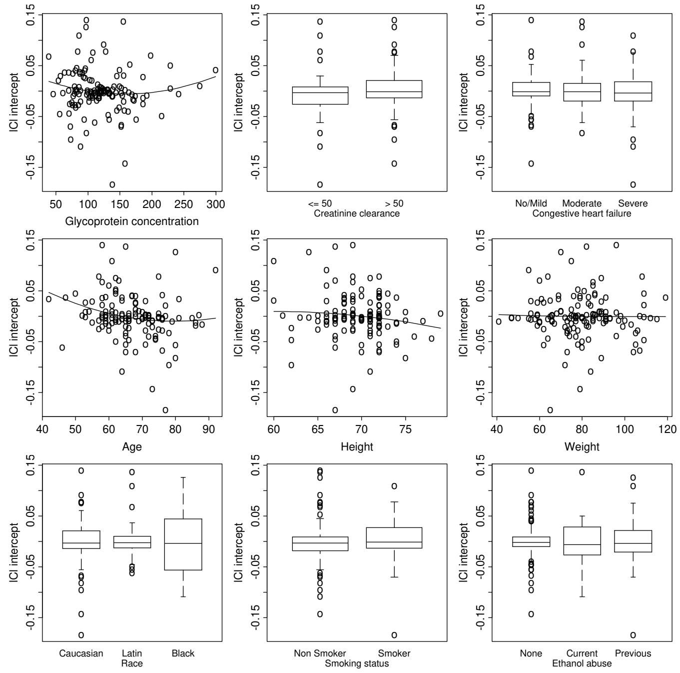

Figure 9: Conditional modes of the lCl intercept random effect in model V versus available covariates.

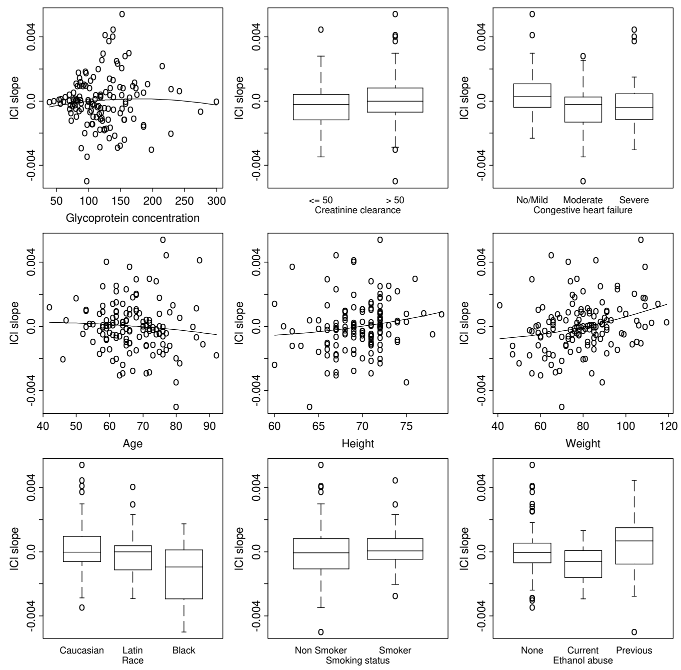

Figure 10: Conditional modes of lCl slope random effect in model V versus available covariates.

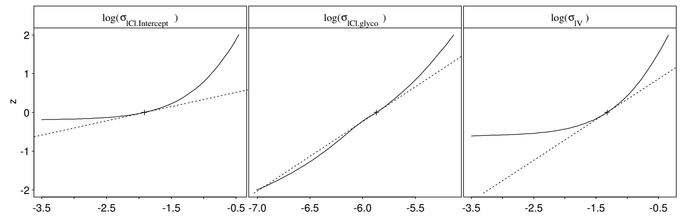

Figure 11: Likelihood profile z plots for the quinidine model V.

and congestive heart failure (as an indicator variable of previous/no previous history of congestive heart failure) in the model, in this order. The same random effect variance-covariance structure as in model V was used in all cases. The corresponding AIC values were respectively $875.59, 870.04$, and 865.29. In all three cases substantial reductions in the AIC values were observed. The random effects plots from the last model (with all three interactions with α-1-acid glycoprotein included) did not indicate any other candidate covariates to be included in the model.

After the inclusion of covariates in the model, part, or maybe all, of the inter-individual parameter variability will have been explained. Hence, it is important to check again if random effects may be eliminated from the model. Figure 11 presents the likelihood profile z plots for the variance components in model V.

The profile z plots of $\log(\sigma_{lCl.Intercept})$ and $\log(\sigma_{lV})$ indicate that the associated random effects should be eliminated from the model. As $\sigma_{lCl.Intercept} \rightarrow 0$, the p-value of the associated likelihood ratio test goes to 0.853, and as $\sigma_{lV} \rightarrow 0$, the p-value of the associated likelihood ratio test goes to 0.562. Note that, even after the inclusion of the covariates in the model, there remains a significant inter-individual variability in the α-1-glycoprotein slope. The AIC of the model (VI) with only the α-1-glycoprotein slope as a mixed effect is 859.52, substantially smaller than all the previous models. No further reduction in the number of parameters is possible and we conclude that model VI is preferred.

6. ASSESSING THE VARIABILITY IN THE ESTIMATES

Once a model has been chosen to represent the data, measures of variability for the estimates and confidence regions on the model's parameters are usually needed for statistical inference. We consider here two approaches for assessing the variability in the parameter estimates: a normal approximation based on (restricted) maximum likelihood asymptotic theory and likelihood profiling (Bates and Watts, 1988). Details of the calculations involved in the normal approximation and the likelihood profiling approaches are presented in Appendices 1 and 2.

One of the difficulties with implementing either of these approaches is that, except when the model is linear in the random effects, the log-likelihood of the nonlinear mixed-effects model does not have a closed form expression. As before, we use the linear mixed-effects (LME) approximation proposed by Lindstrom and Bates (1990) as the approximation to the log-likelihood. Pinheiro and Bates (1995a) compare several approximations to the nonlinear mixed-effects model log-likelihood and conclude that the LME approximation is usually suitable for assessing variability.

We use the normal approximation and the likelihood profiling techniques to assess the variability in the parameter estimates for the pine trees, theophylline, and quinidine final models, described in sections 4 and 5.

6.1 Pine Trees

Figure 12 presents the likelihood profile z plots, traces, and contours for all the parameters in model II of section 4.1. The

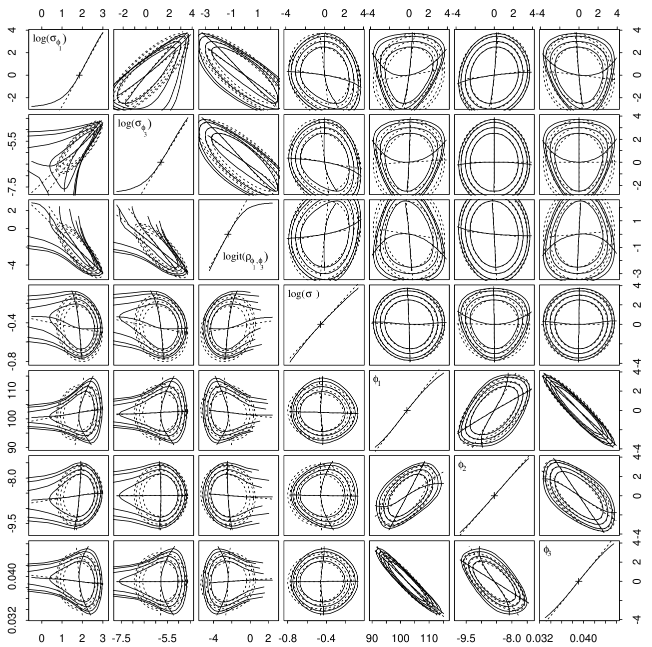

Figure 12: Likelihood profile z plots, traces, and contours (solid line) and profile z plots, traces, and contours based on the normal approximation (dashed line) for the pine trees model II. Plots below the diagonal are in the original scale and plots above the diagonal are in the profile z scale. Interpolated contours correspond approximately to joint confidence levels of 50%, 75%, 90%, and 95%.

profile z plots, traces, and contours based on the normal approximation are also included for comparison. An unrestricted parametrization of variance-covariance components is used to enhance the accuracy of the normal approximation.

We see that there is considerable curvature in the profile z plots and the profile traces corresponding to the random effects variance-covariance components, but the profile z plots and profile traces for the remaining parameters are close to linear. There is considerable correlation among the random effects variance-covariance components estimates and also among the fixed effects estimates, especially the asymptotic height ϕ_1 and the growth rate ϕ_3. The profile traces for the fixed effects and the variance-covariance components meet almost perpendicularly, indicating a low correlation between these estimates.

Table 1 gives the approximate 95% confidence intervals on the parameters in model II, based on the normal approximation and the profile traces.

The approximate confidence intervals indicate that the fixed effects (ϕ_1, ϕ_2, and ϕ_3) and the intra-individual standard error (σ) are estimated with considerably more relative precision than the random effects variance-covariance components (σ_{ϕ_1}, σ_{ϕ_3}, and ρ_{ϕ_1,ϕ_3}).

6.2 Theophylline

Figure 13 presents the likelihood profile z plots, traces, and contours for the parameters in model III of section 4.2. The profile z plots, traces, and contours based on the normal approximation are also included for comparison.

The normal approximation is reasonably accurate for all the estimates in the model. There is almost no correlation between the parameter estimates, with the exception of the fixed effects estimates of lCl and lV.

Table 2 gives the approximate 95% confidence intervals on the parameters in model III, based on the normal approximation and the profile z plots. These intervals are given in the original scale of the parameters.

As expected, the normal approximation confidence intervals are very similar to the profile z plots confidence intervals. We note that the absorption rate constant k_a is estimated with considerably less relative precision than the clearance Cl and the elimination rate constant K.

Table 1: Approximate 95% confidence intervals for the parameters in the pine trees model II, based on the normal approximation and the profile z plots.

Parameter	Normal Approximation	Profile z
σ_{ϕ_1}	(3.479, 11.566)	(2.773, 11.241)
σ_{ϕ_3}	(0.0017, 0.0057)	(0.0012, 0.0055)
ρ_{ϕ_1,ϕ_3}	(-0.964, -0.394)	(-0.960, -0.088)
σ	(0.528, 0.758)	(0.533, 0.766)
ϕ_1	(97.111, 107.393)	(97.117, 107.755)
ϕ_2	(-9.110, -8.038)	(-9.120, -8.031)
ϕ_3	(0.036, 0.042)	(0.036, 0.042)

Table 2: Approximate 95% confidence intervals for the parameters in the theophylline model III, based on the normal approximation and the profile z plots.

Parameter	Normal Approximation	Profile z
σ_{lCl}	(0.109, 0.256)	(0.113, 0.270)
σ_{lk_a}	(0.406, 1.021)	(0.416, 1.068)
σ	(0.620, 0.811)	(0.624, 0.816)
Cl	(0.035, 0.045)	(0.035, 0.045)
k_a	(1.083, 2.343)	(1.056, 2.448)
K	(0.078, 0.095)	(0.078, 0.095)

6.3 Quinidine

Figure 14 presents the likelihood profile z plots, traces, and contours for all the parameters in model VI of section 4.3. The profile z plots, traces, and contours based on the normal approximation are also included for comparison. An unrestricted parametrization is used for all the parameters in the model in order to enhance the accuracy of the normal approximation.

The normal approximation is virtually identical to the profile likelihood approximation in this example. The variance-covariance components profile traces and the fixed effects profile traces meet almost perpendicularly, indicating a very low correlation between the corresponding estimates. There is a strong negative correlation between the $lCl.glyco$ and the $lCl.glyco * weight$ fixed effects estimates and a moderately high positive correlation between the lk_a and the lV fixed effects estimates.

Table 3 gives the approximate 95% confidence intervals on the parameters in model VI, based on the normal approximation and the profile z plots.

As expected, the normal approximation confidence intervals

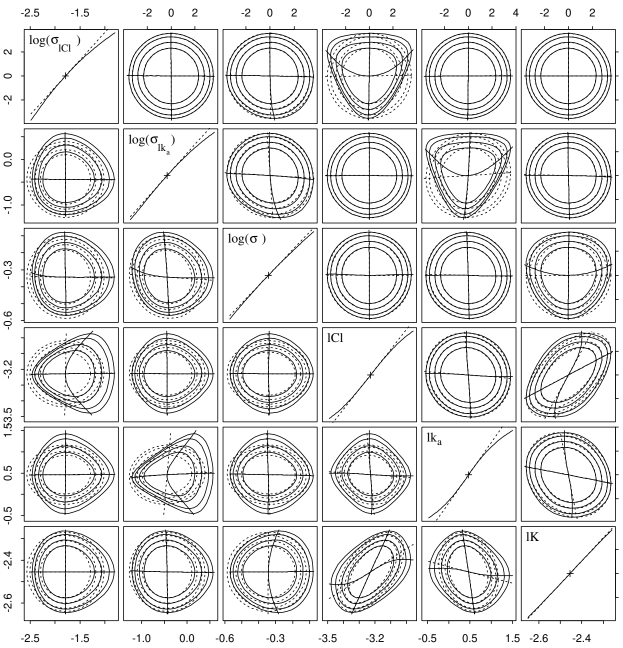

Figure 13: Likelihood profile z plots, traces, and contours (solid line) and profile z plots, traces, and contours based on the normal approximation (dashed line) for the theophylline model III parameters. Plots below the diagonal are in the original scale and plots above the diagonal are in the profile z scale. Interpolated contours correspond approximately to joint confidence levels of 50%, 75%, 90%, and 95%.

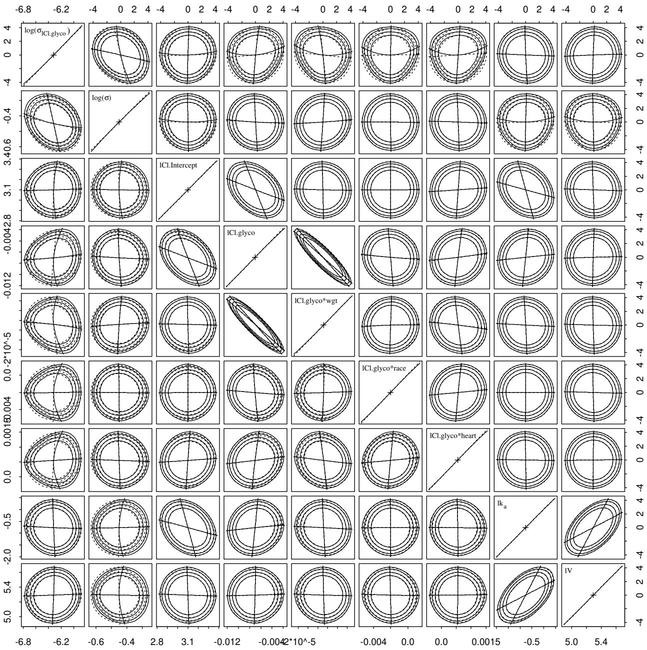

Figure 14: Likelihood profile z plots, traces, and contours (solid line) and profile z plots, traces, and contours based on the normal approximation (dashed line) for the quinidine model VI parameters. Plots below the diagonal are in the original scale and plots above the diagonal are in the profile z scale. Interpolated contours correspond approximately to joint confidence levels of 50%, 75%, 90%, and 95%.

are very similar to the profile z plots confidence intervals. We note that the absorption rate constant k_a is estimated with considerably less relative precision than the apparent volume of distribution V.

7. CONCLUSIONS

The analysis of the eigenstructure of the estimated variance-covariance matrix ($\boldsymbol{D}$) of the random effects, combined with profile z plots of the random effects variance-covariance components, is a useful tool to determine which terms in the model should be random and which should be purely fixed. The estimated $\boldsymbol{D}$ matrix also provides useful information to identify structured variance-covariance patterns. Information criterion statistics, such as the AIC, can be used as guidelines to model selection, but analysis of residuals and consultation with experts in the field of application of the model should also be used.

The goodness-of-fit and interpretability of a mixed effects model can be substantially enhanced through the inclusion of covariates to explain random effects variability. Information criterion statistics can again be used for model selection, together with analysis of residuals and expert advice.

We restricted ourselves here to mixed models in which the intra-individual errors, ϵ in model (2.1), are independent and identically distributed, but other covariance structures (e.g. autoregressive processes) can easily be incorporated into mixed-effects models (Chi and Reinsel, 1989; Lindstrom and Bates, 1990). There is usually a trade-off between the number of random effects incorporated in the model and the complexity of the intra-individual error covariance structure (Jones, 1990). Further research is needed in that area, especially from a model building perspective.

After a model has been chosen to represent the data, measures of variability for the estimates and confidence regions on the model's parameters are usually needed for inferential purposes. Maximum likelihood asymptotic theory and likelihood profiling are two of the possible approaches that can be used for this purpose. The second approach is more computer intensive than the first one, but usually gives more reliable results. Other (computationally intensive) approaches to assess the variability in the estimates are the bootstrap (Efron and Tibshirani, 1993), the Laplacian approximation to the marginal distribution of the estimators (Tierney and Kadane, 1986), and importance sampling (Tanner, 1993). More research is needed to determine which methods provide the most reliable statistical results and to compare their relative computational performance.

Table 3: Approximate 95% confidence intervals for the parameters in the quinidine model VI, based on the normal approximation and the profile z plots.

Parameter	Normal Approximation	Profile z
$\sigma_{lCl.glyco}$	(0.0015, 0.0022)	(0.0015, 0.0022)
σ	(0.584, 0.696)	(0.586, 0.698)
$lCl.Intercept$	(2.971, 3.227)	(2.970, 3.227)
$lCl.glyco$	(-0.0094, -0.0047)	(-0.0094, -0.0047)
$lCl.glyco * weight$	$(0.51, 5.85)10^{-5}$	$(0.47, 5.83)10^{-5}$
$lCl.glyco * race$	(-0.0035, -0.0005)	(-0.0035, -0.0005)
$lCl.glyco * heart$	(0.0001, 0.0010)	(0.0001, 0.0010)
k_a	(0.266, 0.814)	(0.266, 0.816)
V	(166.59, 237.32)	(166.45, 237.40)

REFERENCES

Bates, D. M., and Watts, D. G. (1988), *Nonlinear Regression Analysis and Its Applications*, New York: Wiley.

Chi, E. M., and Reinsel, G. C. (1989), "Models for Longitudinal Data With Random Effects and AR(1) Errors," *Journal of the American Statistical Association*, 84, 452–459.

Cleveland, W. S., Grosse, E., and Shyu, W. M. (1992), "Local Regression Models," *Statistical Models in* S, Wadsworth, Belmont, CA, chapter 8.

Davidian, M., and Gallant, A. R. (1992), "Smooth Nonparametric Maximum Likelihood Estimation for Population Pharmacokinetics, with Application to Quinidine," *Journal of Pharmacokinetics and Biopharmaceutics*, 20, 529–556.

Davidian, M., and Gallant, A. R. (1993), "The Nonlinear Mixed Effects Model with a Smooth Random Effects Density," *Biometrika*, 80, 475–488.

Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, New York: Chapman & Hall.

Jones, R. H. (1990), "Serial Correlation or Random Subject Effects," *Communications in Stat., Part B–Simulation and Comp.*, 19, 1105–1123.

Kung, F. H. (1986), "Fitting Logistic Growth Curve With Predetermined Carrying Capacity," *ASA Proceedings of the Statistical Computing Sect.*, pp. 340–343.

Laird, N. M., and Ware, J. H. (1982), "Random-effects Models for Longitudinal Data," *Biometrics*, 38, 963–974.

Lindstrom, M. J., and Bates, D. M. (1990), "Nonlinear Mixed Effects Models for Repeated Measures Data," *Biometrics*, 46, 673–687.

Mallet, A., Mentre, F., Steimer, J.-L., and Lokiek, F. (1988), "Nonparametric maximum likelihood estimation for population pharmacokinetics, with applications to Cyclosporine," *J. Pharmacokin. Biopharm.*, 16, 311–327.

Pinheiro, J. C. (1994), "Topics in Mixed Effects Models," PhD thesis, University of Wisconsin–Madison.

Pinheiro, J. C., and Bates, D. M. (1995a), "Approximations to the log-likelihood function in the nonlinear mixed-effects model," *Journal of Computational and Graphical Statistics*, 4(1), 12–35.

Pinheiro, J. C., and Bates, D. M. (1995b), "Mixed-Effects Methods and Classes for *S* and *Splus*," *Technical Report 89*, Department of Biostatistics, University of Wisconsin – Madison.

Pinheiro, J. C., and Bates, D. M. (1995c), "Unrestricted Parametrizations for Variance-Covariance Matrices," *Technical Report 90*, Department of Biostatistics, University of Wisconsin – Madison.

Ratkowsky, D. A. (1990), *Handbook of Nonlinear Regression Models*, New York: Marcel Dekker.

Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1986), *Akaike Information Criterion Statistics*, Holland: D. Reidel Publishing Company.

Sheiner, L. B., and Beal, S. L. (1980), "Evaluation of Methods for Estimating Population Pharmacokinetic Parameters. I. Michaelis-Menten Model: Routine Clinical Pharmacokinetic Data," *Journal of Pharmacokinetics and Biopharmaceutics*, 8(6), 553–571.

Tanner, M. (1993), *Tools for Statistical Inference. Methods for the Exploration of Posterior Distributions and Likelihood Functions (Second Edition)*, New York: Springer-Verlag.

Tierney, L., and Kadane, J. B. (1986), "Accurate approximations for posterior moments and densities," *Journal of the American Statistical Association*, 81(393), 82–86.

Verme, C. N., Ludden, T. M., Clementi, W. A., and Harris, S. C. (1992), "Pharmacokinetics of quinidine in male patients: A population analysis," *Clin. Pharmacokin.*, 22, 468–480.

Vonesh, E. F., and Carter, R. L. (1992), "Mixed-Effects Nonlinear Regression for Unbalanced Repeated Measures," *Biometrics*, 48, 1–18.

Wakefield, J. C. (1993), "The Bayesian Analysis of Population Pharmacokinetic Models," *Technical Report TR-93-11*, Statistics Section, Imperial College.

Wakefield, J. C., Smith, A. F. M., Racine-Poon, A., and Gelfand, A. E. (1994), "Bayesian Analysis of Linear and Nonlinear Population Models Using the Gibbs Sampler," *Applied Statistics*. Accepted for publication.

APPENDIX 1 NORMAL APPROXIMATION

The asymptotic normality of the (restricted) maximum likelihood estimators has not yet been proven for the nonlinear mixed-effects model, but has been established for the linear mixed-effects model (Pinheiro, 1994). Hence, at least as a *first-order-like* approximation, asymptotic standard errors and confidence regions based on the normal distribution can be used to assess the variability in the parameter estimates.

Let

$$\widehat{\boldsymbol{\Sigma}} = -\left[\left.\frac{\partial^2 \ell_L(\boldsymbol{\theta}|\boldsymbol{y})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right|_{\widehat{\boldsymbol{\theta}}}\right]^{-1}$$

where $\ell_L(\theta|\boldsymbol{y})$ denotes the LME approximation to the log-likelihood function and $\widehat{\boldsymbol{\theta}}$ denotes the (restricted) maximum likelihood estimates of the model's parameters $\boldsymbol{\theta}$. In practical applications, $\widehat{\boldsymbol{\Sigma}}$ is usually obtained using numerical approximations to the derivatives.

Approximate $1-\alpha$ confidence intervals on the individual parameters are given by

$$\widehat{\theta}_i \pm \Phi^{-1}(1-\alpha/2)\sqrt{\widehat{\boldsymbol{\Sigma}}_{ii}}.$$

where $\Phi^{-1}(1-\alpha/2)$ denotes the $(1-\alpha/2)\times 100$ percentile of the $\mathcal{N}(0,1)$ distribution.

An approximate $1-\alpha$ joint confidence region on $\boldsymbol{\theta}$ is

$$\left\{\boldsymbol{\theta} \in R^p : \left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}\right)^T \widehat{\boldsymbol{\Sigma}}^{-1}\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}\right) \le \chi_p^2(1-\alpha)\right\}$$

where p is the dimension of $\boldsymbol{\theta}$ and $\chi_p^2(1-\alpha)$ is the $(1-\alpha)\times 100$ percentile of the χ_p^2 distribution. By letting $\widehat{\boldsymbol{\theta}}(i,j) = \left(\widehat{\theta}_i, \widehat{\theta}_j\right)^T$ and letting $\widehat{\boldsymbol{\Sigma}}(i,j)$ denote the estimated variance-covariance matrix of $\widehat{\boldsymbol{\theta}}(i,j)$, approximate $1-\alpha$ confidence regions on $\boldsymbol{\theta}(i,j) = (\theta_i,\theta_j)^T$ are given the bivariate projections of the joint confidence region on $\boldsymbol{\theta}$

$$\left\{\boldsymbol{\theta}(i,j) \in R^2 : \left[\boldsymbol{\theta}(i,j)-\widehat{\boldsymbol{\theta}}(i,j)\right]^T\left[\widehat{\boldsymbol{\Sigma}}(i,j)\right]^{-1} \times \left[\boldsymbol{\theta}(i,j)-\widehat{\boldsymbol{\theta}}(i,j)\right] \le \chi_p^2(1-\alpha)\right\}$$

The normal approximation is usually more accurate when the estimation space is unrestricted. Hence, it is recommended that an unrestricted parametrization be used for $\boldsymbol{\theta}$, e.g. using the logarithm of a standard deviation, or the logit of a correlation coefficient ($\text{logit}(\rho) = \log[(1+\rho)/(1-\rho)]$). Pinheiro and Bates (1995c) consider different unrestricted parametrizations for variance-covariance matrices which can be applied to mixed-effects models estimation. The confidence intervals on the parameters may be obtained on the unrestricted scale and then converted back to the original scale.

APPENDIX 2 PROFILE z PLOTS, TRACES, AND CONTOURS

Likelihood profile z plots, traces, and contours provide an alternative way of deriving confidence intervals and bivariate confidence regions for the model's parameters. This approach is usually more accurate than the normal approximation and is invariant under reparametrizations of the model. Its basic disadvantage is that it is more computationally intensive than the normal approximation.

We define the approximate profile log-likelihood in the nonlinear mixed-effects model (2.1), corresponding to the ith parameter, as

$$\ell_L^{(i)}(\delta|\boldsymbol{y}) = \max_{\{\boldsymbol{\theta}:\theta_i=\delta\}} \ell_L(\boldsymbol{\theta}|\boldsymbol{y})$$

That is, $\ell_L^{(i)}(\delta|\boldsymbol{y})$ is the maximum value assumed by the log-likelihood when θ_i is kept fixed at δ. Now let

$$\widehat{\boldsymbol{\theta}}^{(i)}(\delta) = \arg\max_{\{\boldsymbol{\theta}:\theta_i=\delta\}} \ell_L(\boldsymbol{\theta}|\boldsymbol{y})$$

The likelihood profile traces corresponding to θ_i are obtained by plotting the other components of $\widehat{\theta}^{(i)}(\delta)$ against δ. If $\widehat{\boldsymbol{\theta}}$ were normally distributed, the profile traces would be linear.

The profile z function is defined as

$$z_i(\delta) = \text{sign}\left(\delta-\widehat{\theta}_i\right)\sqrt{2\left[\ell_L\left(\widehat{\boldsymbol{\theta}}|\boldsymbol{y}\right) - \ell_L^{(i)}(\delta|\boldsymbol{y})\right]}$$

and the profile z plot corresponding to θ_i is given by the curve $(\delta, z_i(\delta))$. If $\widehat{\theta}_i$ were normally distributed, the profile z plot would be a straight line. Hence, curvature in the profile z plots gives an indication of departures from normality. Approximate $1-\alpha$ confidence intervals on θ_i, based on the profile z function, are given by $\left(z_i^{-1}\left(\Phi^{-1}(\alpha/2)\right), z_i^{-1}\left(\Phi^{-1}(1-\alpha/2)\right)\right)$. Note that this is equivalent to the acceptance region of the α-level likelihood ratio test for θ_i.

The approximate profile log-likelihood corresponding to the pair (θ_i,θ_j) is defined as

$$\ell_L^{(ij)}(\delta_1,\delta_2|\boldsymbol{y}) = \max_{\{\boldsymbol{\theta}:\theta_i=\delta_1,\theta_j=\delta_2\}} \ell_L(\boldsymbol{\theta}|\boldsymbol{y})$$

The $1-\alpha$ profile contour corresponding to (θ_i,θ_j) is defined as

$$\left\{(\delta_1,\delta_2) : \ell_L^{(ij)}(\delta_1,\delta_2|\boldsymbol{y}) \quad \ge \quad \ell_L\left(\widehat{\boldsymbol{\theta}}|\boldsymbol{y}\right) - \chi_p^2(1-\alpha)/2\right\}$$

which is the bivariate projection of the acceptance region of the α-level likelihood ratio test for $\boldsymbol{\theta}$. These likelihood contours provide approximate bivariate confidence regions on pairs of parameters. Bates and Watts (1988, Appendix A.6) describe interpolation techniques that allow creating the profile contours from the profile traces.