Conditional versus Marginal Covariance Representation for Linear and Nonlinear Models

José C. Pinheiro

Dept. of Biostatistics, Novartis Pharmaceuticals, East Hanover, USA

Abstract: Grouped data, such as repeated measures and longitudinal data, are increasingly collected in different areas of application, as varied as clinical trials, epidemiological studies, and educational testing. It is often of interest, for these data, to explore possible relationships between one or more response variables and available covariates. Because of the within-group correlation typically present with this type of data, special regression models that allow the joint estimation of mean and covariance parameters need to be used. Two main approaches have been proposed to represent the covariance structure of the data with these models: (i) via the use of random effects, the so-called conditional model and (ii) through direct representation of the covariance structure of the responses, known as the marginal approach. Here we discuss and compare these two approaches in the context of linear and non-linear regression models with additive Gaussian errors, using a real data example to motivate and illustrate the discussion.

Zusammenfassung: Gruppierte Daten, wie wiederholte Messungen und Longitudinaldaten, werden häufig in den verschiedensten Anwendungsgebieten, etwa bei klinischen und epidemiologischen Studien und in den Erziehungswissenschaften erhoben. Mögliche Zusammenhänge zwischen einer oder mehreren Responsevariablen und vorhandenen Kovariaten sind oft das Ziel der Untersuchung. Um die Korrelation innerhalb der Gruppe zu berücksichtigen, sind Regressionsmodelle nötig, die eine gemeinsame Schätzung von Mittelwert und Kovarianzparametern erlauben. Zwei Ansätze wurden vorgeschlagen, um die Kovarianzstruktur der Daten zu berücksichtigen: (i) durch die Verwendung von zufälligen Effekten, das sogenannte konditionale Modell und (ii) durch die direkte Repräsentation der Kovarianzstruktur der Responses, bekannt als das marginale Modell. Hier vergleichen und diskutieren wir beide Ansätze im Kontext von linearen und nicht-linearen Regressionsmodellen mit additiven Gaußschen Fehlern anhand eines realen Beispiels.

Keywords: Correlated Data, Clustered Data, Mixed Effects, Random Effects, Repeated Measures.

1 Introduction

Mixed-effects models are useful in describing relationships between a response variable and covariates in data that are grouped according to one or more classification factors. Examples of such *grouped data* include longitudinal, repeated measures, and multilevel data, which frequently arise in many areas of application, such as clinical trials, epidemiological studies, and educational testing. Mixed-effects models assume that the form of

the intra-group model relating the response variable to the covariates is common to all groups, but some of the parameters that define the model are allowed to vary with group through the incorporation of *random effects*. By associating common random effects to observations in the same group, mixed-effects models flexibly represent the covariance structure induced by the grouping of the data.

Alternatively, the covariance matrix of the responses may be directly modelled via the covariance model adopted for the within-group error term (assumed to enter the model linearly). Of course, both approaches can be used in combination, giving mixed-effects models considerable flexibility in representing the covariance structure present in grouped data.

In practice, one is usually interested in jointly estimating the regression model parameters (representing the relationship between the response and covariates of interest) and the covariance parameters, determining the correlation and variance structure of the response. The approach based on the use of random effects is called the *conditional model*, while the one relying on direct modelling of the within-group error covariance structure is known as the *marginal model*.

This paper describes and contrasts the two approaches in the context of linear mixed-effects (LME) models and nonlinear mixed-effects (NLME) models. A motivating example, the Dialyzer data, which will be used for illustration of both the LME and NLME models, is described in Section 2. In Section 3 it is shown that the two approaches are equivalent in the context of LME models, in the sense that both can be used to directly model the covariance matrix of the responses. Section 4 investigates extensions to the NLME model, in which the effect of the random effects model on the covariance structure of the response can only be determined approximately. Our conclusions are included in Section 5.

2   An Example: High-Flux Hemodialyzer

Vonesh and Carter (1992) describe and analyze data measured on high-flux hemodialyzers to assess their *in vivo* ultrafiltration characteristics. The ultrafiltration rates (in ml/hr) of 20 high-flux dialyzers were measured at 7 ascending transmembrane pressures (in dmHg). The *in vitro* evaluation of the dialyzers used bovine blood at flow rates of either 200 dl/min or 300 dl/min. These data are available in the `NLME` library (Pinheiro and Bates, 2000) in S-PLUS and R.

The plots of the ultrafiltration rates versus transmembrane pressure by bovine blood flow rate, displayed in Figure 1, reveal that, as expected, the ultrafiltration rate increases with transmembrane pressure up to a maximum, and that higher ultrafiltration rates are attained with the 300 dl/min blood flow dialyzers. These plots show a clear correlation in the measurements made on the same dialyzer and also indicate that the variability in the ultrafiltration rates increases with transmembrane pressure.

In their original paper, Vonesh and Carter (1992) use a nonlinear model to represent the relationship between ultrafiltration rate and transmembrane pressure. An alternative analysis, based on a linear polynomial model, is presented in Littell et al. (1996). Both analyses use random effects to account for the within-dialyzer correlation, with the latter
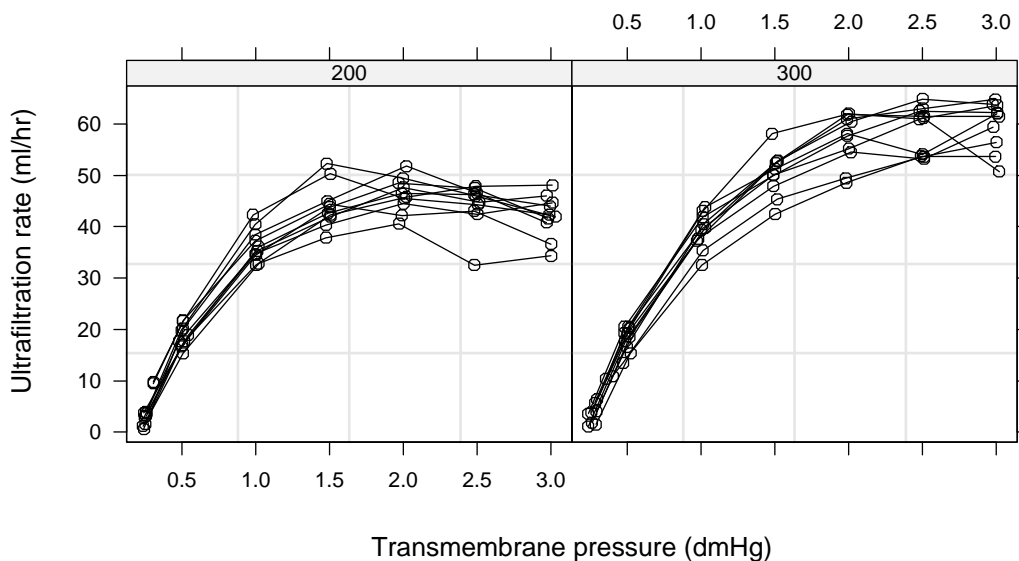
Figure 1: Hemodialyzer ultrafiltration rates (in ml/hr) measured at 7 different transmembrane pressures (in dmHg) on 20 high-flux dialyzers. *In vitro* evaluation of dialyzers based on bovine blood flow rates of 200 dl/min and 300 dl/min.

also investigating the use of marginal covariance models based on extended linear regression models. We use this example illustrate and compare the conditional and marginal model approaches for both the LME model and the NLME model.

3   Linear Mixed-Effects Model

The linear mixed-effects model for a normally distributed response grouped according to a single factor with M levels, proposed by Laird and Ware (1982), is expressed as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i \,, \tag{1}$$

where i is the group index, $\mathbf{y}_i$ is an n_i-dimensional vector of observed responses, $\mathbf{X}_i$ and $\mathbf{Z}_i$ are known $n_i \times p$ and $n_i \times q$ regression matrices corresponding to the p-dimensional *fixed effects* vector $\boldsymbol{\beta}$ and the q-dimensional *random effects* vector $\mathbf{b}_i$ respectively, and $\boldsymbol{\epsilon}_i$ is an n_i-dimensional vector of within-group errors.

The $\mathbf{b}_i$ are assumed to be independently distributed as $\mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$ and the $\boldsymbol{\epsilon}_i$ are assumed to be independently distributed as $\mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_i)$, independent of the $\mathbf{b}_i$. The $\boldsymbol{\Psi}$ covariance matrix may be unstructured or structured – e.g. diagonal (Jennrich and Schluchter, 1986), being parameterized by a set of parameters $\boldsymbol{\theta}$. The $\boldsymbol{\Lambda}_i$ matrices are typically assumed to depend on i only through their dimensions, being parameterized by a fixed, generally small, set of parameters $\boldsymbol{\lambda}$ – e.g. an AR(1) structure (Box et al., 1994).

Several methods of parameter estimation have been proposed for LME models and we consider here the two most widely used and available in statistical software: maximum likelihood (ML) and restricted maximum likelihood (REML). Descriptions and comparisons of the various estimation methods used for LME models can be found, for example,

in Searle et al. (1992) and Vonesh and Chinchilli (1997). For a Bayesian perspective see Wakefield et al. (1994).

Even though the random effects are useful and intuitive quantities to represent between-group differences in the coefficients, they are not observable in practice. Therefore, likelihood estimation and inference generally rely on the *marginal* distribution of the observed response vectors $\mathbf{y}_i$. Because of the linearity of the random effects in the LME model (1), the assumptions on the random effects and the within-group errors, and the properties of the multivariate normal distribution, it can be shown that the $\mathbf{y}_i$ are marginally distributed as independent $\mathcal{N}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i)$ random vectors, where the marginal covariance matrix is given by:

$$\operatorname{var}(\mathbf{y}_i) = \boldsymbol{\Sigma}_i = \underbrace{\mathbf{Z}_i\boldsymbol{\Psi}\mathbf{Z}_i'}_{\text{random effects}} + \underbrace{\boldsymbol{\Lambda}_i}_{\text{within-group error}} . \qquad (2)$$

There are two ways in which the LME model (1) can account for within-group correlation and heteroscedasticity (non-constant variance): through the random effects $\mathbf{b}_i$ and through the within-group errors $\boldsymbol{\epsilon}_i$. Because the random effects $\mathbf{b}_i$ are fixed by group, not varying with observation, the within-group observations share the same random effects and are, therefore, correlated. This is represented by the $\mathbf{Z}_i\boldsymbol{\Psi}\mathbf{Z}_i'$ component of $\boldsymbol{\Sigma}_i$. Note, also, that the diagonal elements of $\mathbf{Z}_i\boldsymbol{\Psi}\mathbf{Z}_i'$ need not be constant, so that it can also accommodate heteroscedasticity. This component of the marginal covariance matrix $\boldsymbol{\Sigma}_i$ is the one favored in the conditional model approach, with $\boldsymbol{\Lambda}_i$ assuming a simple form (typically $\boldsymbol{\Lambda}_i = \sigma^2\mathbf{I}_i$, with $\mathbf{I}_i$ denoting the identity matrix of order n_i.)

The within-group error contribution to the marginal covariance matrix is given directly by $\boldsymbol{\Lambda}_i$, which can be non-diagonal (correlation) and have different diagonal elements (heteroscedasticity). In the marginal model approach one sets $\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i$, so that the entire covariance structure is determined by the within-group error.

Of course, in practice, one may use both components in (2) when modelling $\boldsymbol{\Sigma}_i$. For example, one may use the random effects component to account mostly for the correlation and the within-group component to account for heteroscedasticity via the use of a variance function (Pinheiro and Bates, 2000). Some care should be exercised when using both components in a model, as they may very well compete with each other in explaining the marginal covariance and lead to nearly or fully overparameterized models.

As a simple example, consider the case of an LME model with a single random intercept, that is, with random effects model given by $\mathbf{1}_i b_i$, where b_i is a scalar. The corresponding random effect component of $\boldsymbol{\Sigma}_i$ is then equal to $\psi\mathbf{1}_i\mathbf{1}_i' = \psi\mathbf{J}_i$, with $\mathbf{J}_i$ representing an $n_i \times n_i$ matrix of ones. If we assume a compound symmetry structure for the within-group covariance, that is, $\boldsymbol{\Lambda}_i = \sigma^2((1 - \rho)\mathbf{I}_i + \rho\mathbf{J}_i)$, the resulting marginal covariance would have diagonal terms equal to $\sigma^2 + \psi$ and off-diagonal terms $\sigma^2\rho + \psi$, that is, an overparameterized compound symmetry structure. Another example of overparameterization would result from the use of an unstructured (general) covariance matrix $\boldsymbol{\Lambda}_i$ together with any random effects model.

Conditional on the parameters that determine $\boldsymbol{\Sigma}_i$, the (RE)ML estimate of the fixed effects $\boldsymbol{\beta}$ and its covariance matrix are given by

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{M} \mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\right)^{-1} \sum_{i=1}^{M} \mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{y}_i \quad \text{and} \quad \operatorname{var}\left(\widehat{\boldsymbol{\beta}}\right) = \left(\sum_{i=1}^{M} \mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\right)^{-1}. \qquad (3)$$

The (RE)ML estimates of the parameters determining Σ_i can not be expressed in closed form, except in trivial cases, and numerical optimization of the (restricted) likelihood function must be employed. The MLE of $\boldsymbol{\beta}$ is then obtained by replacing Σ_i in (3) with their corresponding estimates.

Note that both $\widehat{\boldsymbol{\beta}}$ and its estimated variance depend on the covariance model through the marginal covariance matrices Σ_i only. Therefore, methods that lead to similar estimates of the Σ_i will also lead to similar inferences on the fixed effects, including confidence intervals and tests of hypothesis. Therefore, if the main questions of interest associated with the LME model are related to inferences about the fixed effects (like would be the case in clinical trials, for example), then the conditional and marginal approaches may lead to equivalent conclusions, provided similar estimates of Σ_i can be obtained with both methods. However, if one is interested in issues related to the covariance model, like inter-subject variability or spatial correlation, it may be that only one of the approaches can be used to address the specific question of interest. This will, of course, be application dependent, so no general recommendations can be made about the greater adequacy of either approach.

3.1   Dialyzer Example: Linear Model Version

As an empirical model to describe the relationship between the ultrafiltration rates and the transmembrane pressure in the Dialyzer example of Section 2, Littell et al. (1996) proposed the use of a fourth-order linear mixed-effects polynomial allowing for differences in the fixed effects between the dialyzers with in vitro ultrafiltration rates of 200 dl/min and 300 dl/min.

After several model building steps to determine which parameters should vary with in vitro ultrafiltration class and which should be assigned random effects to accommodate the inter-subject variation, the following LME model was chosen to represent the ultrafiltration rate y_{ij} at the jth transmembrane pressure x_{ij} for the ith subject.

$$y_{ij} = (\beta_0 + \gamma_0 Q_i + b_{0i}) + (\beta_1 + \gamma_1 Q_i + b_{1i})\, x_{ij} + (\beta_2 + b_{2i})\, x_{ij}^2 + \beta_3 x_{ij}^3 + \beta_4 x_{ij}^4 + \epsilon_{ij} \,, \quad (4)$$

$$\mathbf{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{bmatrix} \overset{\text{ind}}{\sim} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Psi}\right) \,, \quad \epsilon_{ij} \overset{\text{ind}}{\sim} \mathcal{N}\left(0, \sigma^2 x_{ij}^{\delta}\right) \,,$$

where Q_i is a binary variable taking values 0 for 200 dl/min hemodialyzers and 1 for 300 dl/min hemodialyzers; β_0, β_1, β_2, β_3, and β_4 are, respectively, the intercept, linear, quadratic, cubic, and quartic fixed effects corresponding to 200 dl/min dialyzers; γ_k is the blood flow effect associated with the fixed effect β_k, $k = 0, 1$; $\mathbf{b}_i$ is the vector of random effects, assumed independent for different i; and ϵ_{ij} is the within-group error, assumed independent for different i, j and independent of the random effects. Different structures can be used to represent the random effects covariance matrix $\boldsymbol{\Psi}$, as illustrated further below. The variance of the ϵ_{ij} is allowed to change with the transmembrane pressure according to a power model with parameter δ (estimated together with the other parameters in the model.) This was needed due to heteroscedastic behavior observed in the residuals from the LME fit with homocedastic within-subject errors.

The LME model in (4) uses random effects to account for the correlation among within-subject measurements and some of the heteroscedasticity in the response as well. An additional variance function model is used for the within-subject error to properly accommodate the observed heteroscedasticity, but the approach can be considered primarily a conditional model, as discussed previously.

An alternative marginal model to represent the data using the empirical linear polynomial model is given below.

$$y_{ij} = (\beta_0 + \gamma_0 Q_i) + (\beta_1 + \gamma_1 Q_i) x_{ij} + \beta_2 x_{ij}^2 + \beta_3 x_{ij}^3 + \beta_4 x_{ij}^4 + \epsilon_{ij}, \tag{5}$$
$$\epsilon_{ij} \sim \mathcal{N}\left(0, \sigma^2 x_{ij}^\delta\right) \quad \text{corr}(\epsilon_{ij}, \epsilon_{ik}) = \rho^{|j-k|},$$

that is, the same fixed effects and variance function as in (4) are used, but the within-subject correlation is now modeled by an AR(1) structure in the within-subject errors. We also considered marginal models similar to (5), but with within-subject correlation structure given by a Gaussian spatial correlation model with a nugget effect (Pinheiro and Bates, 2000), an ARMA(1,3) model, and an unstructured model (that is, with 21 different correlations allowed for the within-subject measurement pairs).

The conditional and marginal models described above for the Dialyzer data were fitted using, respectively, the functions `lme` and `gls` in the `NLME` library in S-PLUS, with REML estimation in all cases. Figure 2 shows the scatter plots of the conditional model estimates for variance and covariance parameters in Σ_i versus the corresponding estimates using the different marginal models.
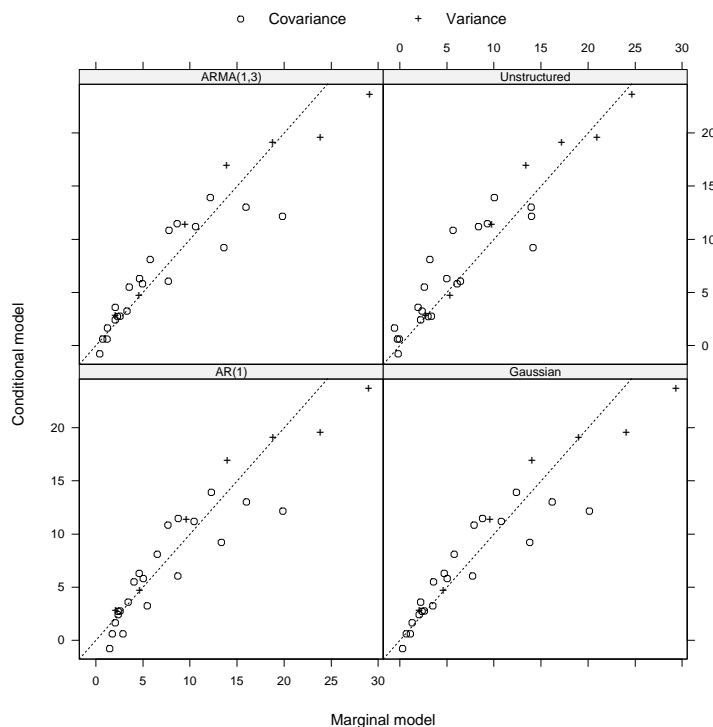

Figure 2: Conditional versus marginal estimates for variance and covariance parameters in the Dialyzer linear models. Dotted line represents $y = x$ reference line.

There is generally good agreement between the estimates from the conditional and marginal models, with greater discrepancies observed for larger parameter values. The unstructured correlation marginal model has the best agreement with the conditional model with regard to the variances, but not the covariances. It is instructive to examine the estimation results for the fixed effects in each of the models, given below.

```
## LME fit (conditional model)
                Value Std.Error  DF t-value p-value
  (Intercept) -16.393    0.8273 115 -19.814  <.0001
     pressure  90.616    3.2660 115  27.745  <.0001
           QB  -2.777    0.8278  18  -3.355  0.0035
I(pressure^2) -49.692    4.2295 115 -11.749  <.0001
I(pressure^3)  12.474    2.1509 115   5.799  <.0001
I(pressure^4)  -1.277    0.3575 115  -3.573  0.0005
  pressure:QB   7.031    0.7856 115   8.949  <.0001
```

```
## AR(1) marginal model
                Value Std.Error t-value p-value
  (Intercept) -16.405    0.8453 -19.406  <.0001
     pressure  90.240    3.8747  23.290  <.0001
           QB  -2.615    0.7399  -3.535  0.0006
I(pressure^2) -48.592    5.0571  -9.609  <.0001
I(pressure^3)  11.802    2.4749   4.769  <.0001
I(pressure^4)  -1.153    0.3980  -2.897  0.0044
  pressure:QB   6.615    0.8105   8.161  <.0001
```

```
## Spatial Gaussian marginal model
                Value Std.Error t-value p-value
  (Intercept) -16.357    0.7930 -20.626  <.0001
     pressure  89.759    3.5459  25.314  <.0001
           QB  -2.391    0.7441  -3.214  0.0016
I(pressure^2) -48.095    4.5640 -10.538  <.0001
I(pressure^3)  11.663    2.2276   5.236  <.0001
I(pressure^4)  -1.144    0.3588  -3.188  0.0018
  pressure:QB   6.419    0.8396   7.645  <.0001
```

```
## ARMA(1,3) marginal model
                Value Std.Error t-value p-value
  (Intercept) -16.297    0.7878 -20.686  <.0001
     pressure  89.489    3.5166  25.448  <.0001
           QB  -2.385    0.7351  -3.244  0.0015
I(pressure^2) -47.780    4.5177 -10.576  <.0001
I(pressure^3)  11.534    2.2032   5.235  <.0001
I(pressure^4)  -1.127    0.3549  -3.174  0.0019
  pressure:QB   6.408    0.8285   7.734  <.0001
```

```
## Unstructured marginal model
               Value Std.Error t-value p-value
  (Intercept) -15.901    0.7827 -20.316  <.0001
     pressure  87.770    3.7826  23.203  <.0001
           QB  -2.027    0.6547  -3.097  0.0024
I(pressure^2) -45.890    5.1765  -8.865  <.0001
I(pressure^3)  10.611    2.6619   3.986  0.0001
I(pressure^4)  -0.975    0.4447  -2.192  0.0302
  pressure:QB   6.309    0.6930   9.103  <.0001
```

The fixed effects estimates and respective standard errors obtained in the LME fit are similar to the estimates obtained in the marginal model fits. The conclusions about the significance of the model parameters are the same in all five models. The greater discrepancies are observed for the unstructured correlation model with respect to the other models, but those differences are not pronounced and within the estimated precisions (overlapping confidence intervals). The main conclusion from the exercise above is that the conditional and marginal approaches lead to similar inferences about the model parameters and similar estimated covariance matrices $\widehat{\Sigma}_i$.

4   Nonlinear Mixed-Effects Model

Nonlinear mixed-effects (NLME) models are mixed-effects models in which the response function is nonlinear in at least some of the underlying parameters. Several different nonlinear mixed effects models have been proposed in the literature (Sheiner and Beal, 1980; Mallet et al., 1988; Lindstrom and Bates, 1990; Vonesh and Carter, 1992; Davidian and Gallant, 1992; Wakefield et al., 1994). We adopt here the NLME model proposed by Lindstrom and Bates (1990), which can be viewed as a hierarchical model that generalizes both the linear mixed-effects model of Section 3 and the usual nonlinear regression model for independent data (Bates and Watts, 1988). In the first stage, the jth observation on the ith group is described as

$$y_{ij} = f(\boldsymbol{\phi}_{ij}, \boldsymbol{x}_{ij}) + \epsilon_{ij}, \quad i = 1, \ldots, M, \; j = 1, \ldots, n_i, \tag{6}$$

where f is a nonlinear function of a group-specific vector of parameters $\boldsymbol{\phi}_{ij}$ and the vector of covariates $\boldsymbol{x}_{ij}$, M is the total number of groups, and n_i is the number of observations in the ith group. Like in the LME model (1), the within-group error vectors $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})'$ are assumed to be independently distributed as $\mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_i)$, with the $\boldsymbol{\Lambda}_i$ parameterized by a fixed parameter vector $\boldsymbol{\lambda}$. In the second stage the group-specific parameters are modelled as

$$\boldsymbol{\phi}_{ij} = \boldsymbol{A}_{ij}\boldsymbol{\beta} + \boldsymbol{B}_{ij}\mathbf{b}_i. \tag{7}$$

Like in the LME model, $\boldsymbol{\beta}$ represents the fixed effects and $\mathbf{b}_i$ the random effects (varying with i but not with j), which are assumed to be independently distributed as $\mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$. $\boldsymbol{A}_{ij}$ and $\boldsymbol{B}_{ij}$ are design matrices for the fixed and random effects respectively, which may depend on the values of some covariates at the jth observation. It is further assumed that the $\mathbf{b}_i$ are independent of the ϵ_{ij}.

Different methods have been proposed to estimate the parameters in the NLME model (Ramos and Pantula, 1995; Davidian and Giltinan, 1995; Vonesh and Chinchilli, 1997); we concentrate here on methods based on the likelihood function. As in the LME case, because the random effects $\mathbf{b}_i$ are unobserved quantities, maximum likelihood estimation in NLME models is based on the marginal density of the responses $\mathbf{y}$, which is calculated as

$$\ell(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\lambda}|\mathbf{y}) = \int p(\mathbf{y}|\mathbf{b}, \boldsymbol{\beta}, \boldsymbol{\lambda}) \, p(\mathbf{b}|\boldsymbol{\Psi}) \, d\mathbf{b} \,. \tag{8}$$

Because the model function f in (6) can be nonlinear in the random effects, the integral in (8) generally does not have a closed-form expression. To make the numerical optimization of the likelihood function a tractable problem, different approximations to (8) have been proposed. Some of these methods consist of taking a first-order Taylor expansion of the model function f around the expected value of the random effects (*i.e.*, $\mathbf{0}$) (Sheiner and Beal, 1980; Vonesh and Carter, 1992), or around the conditional modes of the random effects ($\widehat{\mathbf{b}}_i$) (Lindstrom and Bates, 1990). We adopt here the approximation suggested by Lindstrom and Bates (1990), which is implemented via an alternating algorithm comprising a linear mixed-effects (LME) step and a penalized nonlinear least squares (PNLS) step. This is the algorithm used in the `nlme` function of the NLME library. Inferences on the model parameters, including hypothesis testing, are based on asymptotic results for the linear mixed-effects log-likelihood used in the LME step of the alternating algorithm (Pinheiro and Bates, 2000).

Like for the LME model, the marginal covariance matrix of the responses in the NLME model can be expressed as the sum of two components, one associated with the random effects and another with within-group errors. Letting $\boldsymbol{f}(\boldsymbol{\beta}, \mathbf{b}_i, \mathbf{X}_i) = [f(\boldsymbol{\phi}_{i1}, \boldsymbol{x}_{i1}), \ldots, f(\boldsymbol{\phi}_{in_i}, \boldsymbol{x}_{in_i}))]'$ it follows

$$\mathrm{var}\,(\mathbf{y}_i) = \boldsymbol{\Sigma}_i = \underbrace{\mathrm{var}\,(\boldsymbol{f}(\boldsymbol{\beta}, \mathbf{b}_i, \mathbf{X}_i))}_{\text{random effects}} + \underbrace{\boldsymbol{\Lambda}_i}_{\text{within-group error}} \,. \tag{9}$$

Note that $\mathrm{var}\,(\boldsymbol{f}(\boldsymbol{\beta}, \mathbf{b}_i, \mathbf{X}_i))$ depends not only on $\boldsymbol{\Psi}$, but also on $\boldsymbol{\beta}$ and $\mathbf{X}_i$. One can obtain an estimate for this component of the variance by simulating random effects vectors $\mathbf{b}_1^*, \ldots, \mathbf{b}_N^* \overset{\text{ind}}{\sim} \mathcal{N}(\mathbf{0}, \widehat{\boldsymbol{\Psi}})$, calculating the corresponding vectors $\boldsymbol{f}(\widehat{\boldsymbol{\beta}}, \mathbf{b}_k^*, \mathbf{X}_i)$, $k = 1, \ldots, N$, and deriving the associated sample covariance matrix. Alternatively, a first-order Taylor expansion of $\boldsymbol{f}(\boldsymbol{\beta}, \cdot, \mathbf{X}_i)$ around $\mathbf{b}_i = \mathbf{0}$ can be used to obtain an LME-like approximation

$$\mathrm{var}\,(\boldsymbol{f}(\boldsymbol{\beta}, \mathbf{b}_i, \mathbf{X}_i)) \simeq \mathbf{Z}_i(\boldsymbol{\beta}, \mathbf{X}_i)\boldsymbol{\Psi}\mathbf{Z}_i'(\boldsymbol{\beta}, \mathbf{X}_i), \quad \text{where } \mathbf{Z}_i(\boldsymbol{\beta}, \mathbf{X}_i) = \left.\frac{\partial \boldsymbol{f}(\boldsymbol{\beta}, \mathbf{b}_i, \mathbf{X}_i)}{\partial \mathbf{b}_i}\right|_{\mathbf{b}_i=\mathbf{0}} \,.$$

The same comments and recommendations discussed in Section 3 for the LME model also apply here. The conditional approach attempts to explain the covariance structure of the response mostly via the random effects component of the marginal covariance, while the marginal approach focuses on the within-group component. Care should be exercised in not making both components too complex, to avoid overparameterization problems.

Unlike the LME case, however, the marginal covariance in the NLME model also depends on the fixed effects, which makes the problem more complex. In addition, because

of the need to approximate the likelihood function of the NLME model to make the estimation problem feasible, the relationship between the marginal covariance matrix in (9) and estimation results for the fixed effects is not as clear cut as in the LME model case. Another important difference worth pointing out between the conditional and marginal models in the nonlinear case is in the interpretation of the parameters $\boldsymbol{\beta}$. In the conditional model, the fixed effects are associated with the response of *typical* individual (*i.e.*, with $\mathbf{b}_i = \mathbf{0}$), while in the marginal model, these parameters are associated with the average response across the individuals in the population.

4.1   Dialyzer Example: Nonlinear Model Version

An empirical fourth-order linear polynomial model proposed by Littell et al. (1996) was used in Section 3.1 to represent the Dialyzer data. The model originally proposed for these data by Vonesh and Carter (1992) is an asymptotic regression model with an offset, which expresses the expected ultrafiltration rate y at transmembrane pressure x as

$$\mathrm{E}\left(y\right) = \phi_1 \left\{1 - \exp\left[-\exp(\phi_2)\left(x - \phi_3\right)\right]\right\} . \tag{10}$$

Unlike the parameters in the empirical linear model of Section 3.1, the parameters in model (10) have a physiological interpretation: ϕ_1 is the maximum ultrafiltration rate that can be attained, ϕ_2 is the logarithm of the hydraulic permeability transport rate, and ϕ_3 is the transmembrane pressure required to offset the oncotic pressure.

Vonesh and Carter (1992) suggest using different parameters in (10) for each blood flow rate level. After several model building steps, the following NLME version of (10) was selected to represent the ultrafiltration rate y_{ij} at the jth transmembrane pressure x_{ij} for the ith subject

$$y_{ij} = (\beta_1 + \gamma_1 Q_i + b_{1i}) \left\{1 - \exp\left[-\exp\left(\beta_2 + \gamma_2 Q_i + b_{2i}\right)\left(x_{ij} - \beta_3\right)\right]\right\} + \epsilon_{ij} , \tag{11}$$

$$\mathbf{b}_i = \begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix} \overset{\mathrm{ind}}{\sim} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Psi}\right) , \quad \epsilon_{ij} \overset{\mathrm{ind}}{\sim} \mathcal{N}\left(0, \sigma^2 x_{ij}^\delta\right) ,$$

where, as in Section 3.1, Q_i is a binary indicator taking values 0 for 200 dl/min hemodialyzers and 1 for 300 dl/min hemodialyzers; β_1, β_2, and β_3 are, respectively, the asymptotic ultrafiltration rate, the log-transport rate, and the transmembrane pressure offset fixed effects corresponding to 200 dl/min dialyzers; γ_k is the blood flow effect associated with the fixed effect β_k, $k = 1, 2$; $\mathbf{b}_i$ is the vector of random effects, assumed independent for different i; and ϵ_{ij} is the within-group error, assumed independent for different i, j and independent of the random effects. The variance of the ϵ_{ij} is allowed to change with the transmembrane pressure according to a power model with parameter δ (estimated together with the other parameters in the model.)

Random effects are used in the NLME model (11) to account for within-subject correlation, as well as some of the heteroscedasticity in the response. A variance function model is used in the within-subject covariance matrix $\boldsymbol{\Lambda}_i$ to accommodate the remaining heteroscedasticity in the data. Overall, this approach falls within the conditional model category.

Alternatively, a marginal nonlinear regression model, without random effects, can be used to represent the data.

$$y_{ij} = (\beta_1 + \gamma_1 Q_i) \{1 - \exp\left[-\exp\left(\beta_2 + \gamma_2 Q_i\right)(x_{ij} - \beta_3)\right]\} + \epsilon_{ij}, \quad (12)$$

$$\epsilon_{ij} \overset{\text{ind}}{\sim} \mathcal{N}\left(0, \sigma^2 x_{ij}^\delta\right), \quad \text{corr}(\epsilon_{ij}, \epsilon_{ik}) = \rho^{|j-k|}.$$

This model has the same fixed effects and variance function as (11), but the correlation among measurements is now modelled by an AR(1) structure in the within-subject errors. As in the linear model analysis of Section 3.1, three additional marginal nonlinear models similar to (5) were considered, with within-subject correlation structures: Gaussian spatial correlation with a nugget effect, ARMA(1,3), and unstructured.

The conditional and marginal models described above were fitted using, respectively, the functions `nlme` and `gnls` in the `NLME` library in S-PLUS, with approximate ML estimation for the NLME model and exact ML estimation for the marginal models. The simulation approach outlined previously (based on sampling random effects from their estimated distribution) was used to estimate the marginal covariance matrices in the NLME model. Note that, because of the dependency of Σ_i on $\boldsymbol{\beta}$, different matrices are obtained for 200 dl/min and 300 dl/min dialyzers. For simplicity, we consider here the 200 dl/min dialyzer estimates.
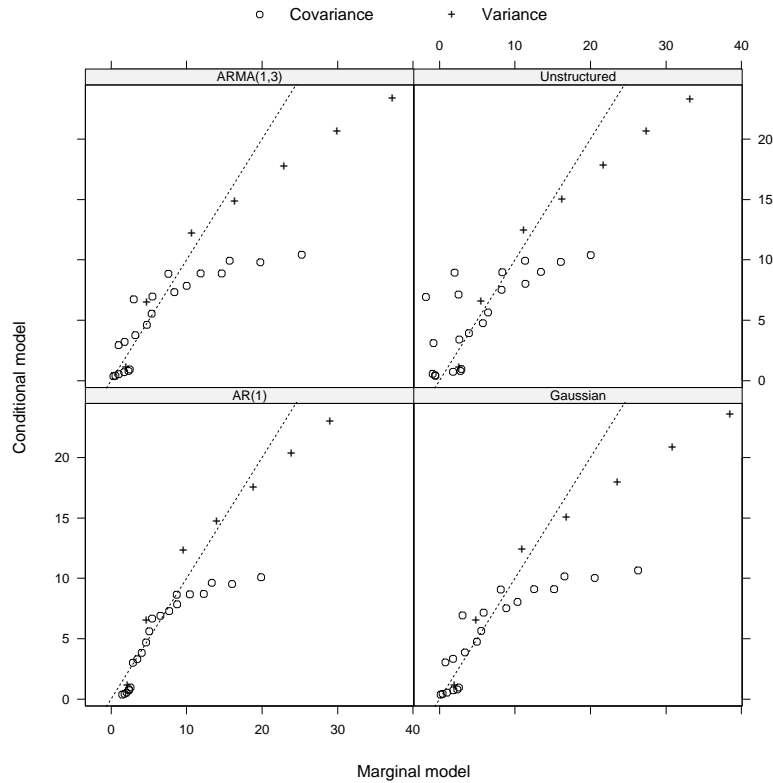

Figure 3: Conditional versus marginal estimates for variance and covariance parameters in the Dialyzer nonlinear models. Conditional model estimates correspond to 200 dl/min dialyzers. Dotted line represents $y = x$ reference line.

Figure 3 shows the scatter plots of the conditional model estimates for variance and covariance parameters in Σ_i associated with 200 dl/min dialyzers versus the corresponding estimates using the different marginal models.

There is now considerably less agreement between the estimates from the conditional and marginal models than in the linear model case depicted in Figure 2. This is consistent with the fact that, for nonlinear models, the two approaches are not equivalent, leading to different representations of the marginal covariance structures. As in the linear case, it is instructive to investigate the impact of the different models in the estimation of the fixed effects.

```
## NLME fit (conditional model)
                 Value Std.Error  DF t-value p-value
Asym.(Intercept) 46.133    1.2359 116  37.326  <.0001
        Asym.QB  16.974    1.8689 116   9.082  <.0001
 lrc.(Intercept)  0.658    0.0598 116  10.992  <.0001
         lrc.QB  -0.454    0.0796 116  -5.702  <.0001
             c0   0.217    0.0035 116  62.971  <.0001


## AR(1) marginal model
                 Value Std.Error t-value p-value
Asym.(Intercept) 46.911    1.5343  30.576  <.0001
        Asym.QB  16.400    2.4104   6.804  <.0001
 lrc.(Intercept)  0.542    0.0538  10.069  <.0001
         lrc.QB  -0.339    0.0748  -4.538  <.0001
             c0   0.215    0.0043  50.527  <.0001


## Spatial Gaussian marginal model
                 Value Std.Error t-value p-value
Asym.(Intercept) 47.122    1.6343  28.833  <.0001
        Asym.QB  15.877    2.6108   6.081  <.0001
 lrc.(Intercept)  0.502    0.0572   8.775  <.0001
         lrc.QB  -0.301    0.0807  -3.731   3e-04
             c0   0.214    0.0042  50.653  <.0001


## ARMA(1,3) marginal model
                 Value Std.Error t-value p-value
Asym.(Intercept) 47.037    1.5952  29.487  <.0001
        Asym.QB  15.955    2.5515   6.253  <.0001
 lrc.(Intercept)  0.510    0.0568   8.976  <.0001
         lrc.QB  -0.308    0.0802  -3.835   2e-04
             c0   0.214    0.0042  50.476  <.0001


## Unstructured marginal model
                 Value Std.Error t-value p-value
Asym.(Intercept) 47.229    1.3340  35.405  <.0001
        Asym.QB  15.938    2.1380   7.455  <.0001
 lrc.(Intercept)  0.451    0.0485   9.289  <.0001
         lrc.QB  -0.274    0.0683  -4.017   1e-04
             c0   0.209    0.0041  50.672  <.0001
```

Even though there are noticeable differences in the estimates of the marginal covariance matrices, the fixed effects estimates obtained via the marginal and conditional approaches are fairly similar and lead to the same inferences about the differences between the two types of dialyzers. The greater discrepancies are observed between the unstructured marginal model and the NLME conditional model, but even these are not of considerable magnitude.

5 Conclusions

This paper describes and contrasts the conditional and marginal modelling approaches in the context of linear and nonlinear regression models. In the linear case, the marginal covariance matrix corresponding to the responses in a linear mixed-effects model is expressed as a sum of two components: one determined solely by the random effects model (associated with the conditional approach) and another determined by the within-group error covariance (associated with the marginal approach.) The two approaches are equivalent in the sense that they directly model the marginal covariance of the responses and that both can use the same estimation method (maximum likelihood or restricted maximum likelihood.)

In the nonlinear case, however, the association between random effects model and the marginal covariance matrix is less direct, depending also on the fixed effects model. In addition, because the random effects typically occur nonlinearly in an NLME model, for computational reasons exact likelihood methods are seldom used. Likelihood methods, on the other hand, can be easily implemented with the marginal approach. As a consequence, the two approaches can not be considered as equivalent, even though similar inferences on the fixed effects are usually obtained.

Perhaps the most important differences between the two approaches have to do with the questions that they are intended to answer. If the primary questions of interest refer to the fixed effects only, then, depending on the choice of random effects and within-group covariance, both methods could be equivalent in the linear case, and generally nearly equivalent in the nonlinear case. However, if one is interested in using predicted random effects to explore the impact of covariates on inter-subject variation, like is often done in pharmacokinetics/pharmacodynamics modeling (Davidian and Giltinan, 1995), then the conditional approach would be the natural choice.

References

Bates, D. M., and Watts, D. G. (1988). *Nonlinear regression analysis and its applications*. New York: Wiley.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time series analysis: Forecasting and control* (3rd ed.). San Francisco: Holden-Day.

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 Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data*. London: Chapman & Hall.

Jennrich, R. I., and Schluchter, M. D. (1986). Unbalanced repeated measures models with structural covariance matrices. *Biometrics*, *42*(4), 805–820.

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.

Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. (1996). *Sas system for mixed models*. Cary, NC: SAS Institute Inc.

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

Pinheiro, J. C., and Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer-Verlag.

Ramos, R. Q., and Pantula, S. G. (1995). Estimation of nonlinear random coefficient models. *Statistics & Probability Letters*, *24*, 49–56.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance components*. New York: Wiley.

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.

Vonesh, E. F., and Carter, R. L. (1992). Mixed-effects nonlinear regression for unbalanced repeated measures. *Biometrics*, *48*, 1–18.

Vonesh, E. F., and Chinchilli, V. M. (1997). *Linear and nonlinear models for the analysis of repeated measures*. New York: Marcel Dekker.

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*, *43*, 201–221.

Authors' address:

José C. Pinheiro
Department of Biostatistics
Novartis Pharmaceuticals
One Health Plaza
East Hanover, NJ 07934
U.S.A.

E-Mail: `jose.pinheiro@novartis.com`