STATISTICS IN MEDICINE
Statist. Med. 2008; **27**:435–448
DOI: 10.1002/sim.2925

Streamlined variance calculations for semiparametric mixed models

Andrew D. A. C. Smith[‡] and M. P. Wand[*,†]

School of Mathematics and Statistics, University of New South Wales, Sydney 2052, NSW, Australia

SUMMARY

Semiparametric mixed model analysis benefits from variability estimates such as standard errors of effect estimates and variability bars to accompany curve estimates. We show how the underlying variance calculations can be done extremely efficiently compared with the direct naïve approach. These streamlined calculations are linear in the number of subjects, representing a two orders of magnitude improvement. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: additive mixed models; longitudinal data analysis; penalized splines; semiparametric regression; subject-specific curves

1. INTRODUCTION

A current vibrant area of research is the use of non-parametric regression or smoothing techniques in the analysis of longitudinal data. Prominent examples include [1–5]. Summaries may be found in books such as [6–8].

Figure 1 shows an example of data that benefit from such methodology. It consists of longitudinal measurements on the spinal bone mineral density (SBMD) of a cohort of young female subjects (source: Reference [9]). One question of interest concerns differences in mean SBMD among the four ethnic groups after accounting for age. An appropriate model is the *additive mixed model*

$$\texttt{SBMD}_{ij}=f(\texttt{age}_{ij})+\beta_2\,\texttt{black}_i+\beta_3\,\texttt{hispanic}_i+\beta_4\,\texttt{white}_i+U_i+\varepsilon_{ij} \tag{1}$$

Here, $\texttt{SBMD}_{ij}$ denotes the jth $(1\leqslant j\leqslant n_i)$ SBMD measurement on subject i $(1\leqslant i\leqslant m)$, f is a smooth, but otherwise unspecified, function for the mean effect of age and $\texttt{black}_i$, $\texttt{hispanic}_i$

---

[*] Correspondence to: M. P. Wand, School of Mathematics and Applied Statistics, University of Wollongong, Wollongong 2522, Australia.
[†] E-mail: mwand@uow.edu.au
[‡] E-mail: andrew.smith@europe.com

Received 25 July 2006
Accepted 4 April 2007

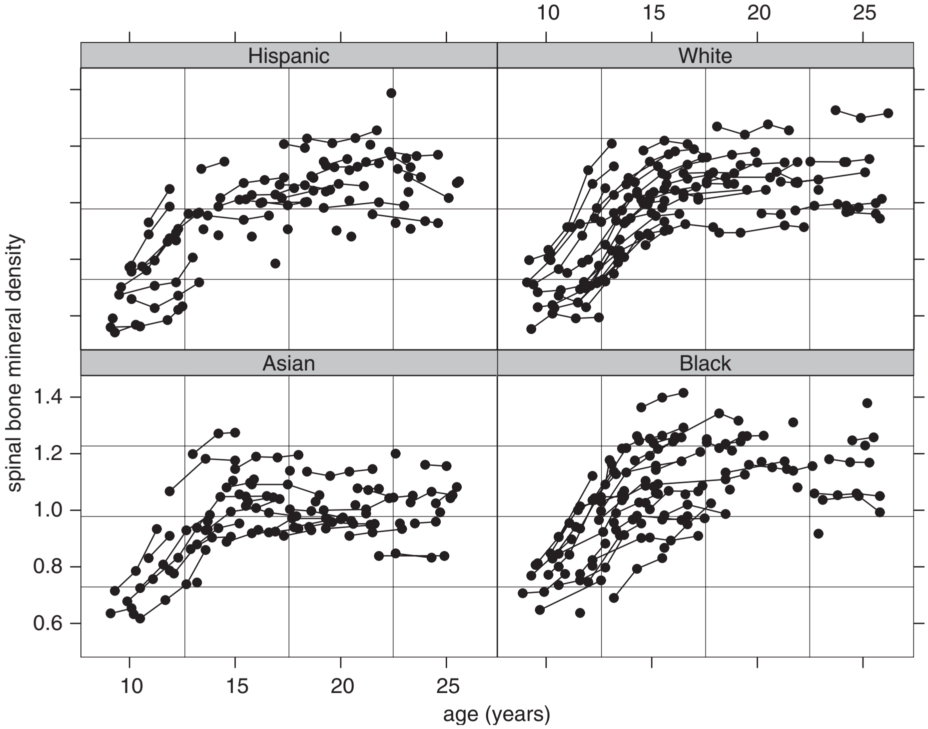

Figure 1. Spinal bone mineral density data broken down according to ethnicity of the subjects.

Table I. Summary of ethnicity effects for linear mixed model fit of (1).

Variable	Value	Standard error	z-Ratio
Black	0.106	0.0207	5.14
Hispanic	0.0131	0.0217	0.607
White	0.0260	0.0216	1.20

and $\texttt{white}_i$ are indicator variables for ethnicity. In addition, the U_i i.i.d. $N(0, \sigma_U^2)$ are random subject intercepts, and the ε_{ij} i.i.d. $N(0, \sigma_\varepsilon^2)$, independent of the U_i's, account for within-subject variability. If f is modelled using penalized splines, such as

$$f(\texttt{age}) = \beta_0 + \beta_1\,\texttt{age} + \sum_{k=1}^{K} u_k(\texttt{age} - \kappa_k)_+, \quad u_k \text{ i.i.d. } N(0, \sigma_u^2)$$

where $\kappa_1, \ldots, \kappa_K$ is a dense set of knots, then (1) reduces to a linear mixed model. The details are given in Section 2. Standard fitting leads to the summary table given in Table I showing, for example, a highly significant difference between black and Asian females in terms of mean SBMD.

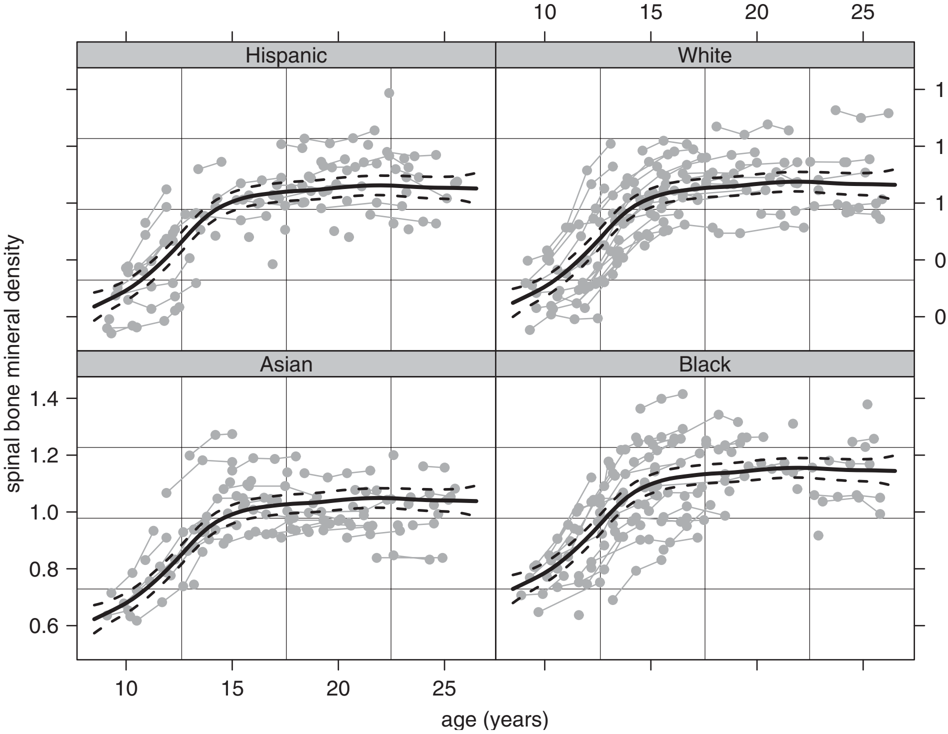

Figure 2. Data from Figure 1 with estimates of $f(\texttt{age})$ added. The dashed lines correspond to ±2 × the estimated standard error.

Figure 2 provides a summary of the estimate of f and its variability. The curves correspond to the function $\widehat{f}(\texttt{age})$ with vertical shifting according to $\widehat{\beta}_i$, $i=2,3,4$. The variability bars, shown as dashed lines, are ±2 × the estimated standard errors and correspond to approximate pointwise 95% confidence intervals of mean SBMD.

Underlying Table I and Figure 2 is the estimated covariance matrix of the coefficient estimates

$$[\widehat{\beta}_0, \ldots, \widehat{\beta}_4, \widehat{u}_1, \ldots, \widehat{u}_K, \widehat{U}_1, \ldots, \widehat{U}_m]^{\mathrm{T}}$$

This covariance matrix involves the inversion of a $(5+K+m)\times(5+K+m)$ matrix $\mathbf{M}$. Typically K is in the range 15–40 regardless sample size variables. However, the number of subjects m can be arbitrarily large. For the SBMD example, $m=230$ which does not pose serious problems. But other studies involve much larger m. For example, the *Six Cities Study of Air Pollution and Health*, described in [10, p. 210], has $m=13\,379$. Thus, for many situations, the number of subjects m is the dominant term in the dimension of $\mathbf{M}$. Henceforth, let K be fixed and small but allow m to be arbitrarily large. Then it is well known from numerical linear algebra (e.g. Reference [11]) that naïve computation of $\mathbf{M}^{-1}$ is $O(m^3)$. Hence, direct variance calculations can be very costly, or even prohibitive, in large longitudinal studies.

The purpose of this article is to show that variance calculations of interest, such as those required for the standard errors of Table I and the variability bars of Figure 2, can be done in $O(m)$ operations. The key is recognition that the contribution to $\mathbf{M}$ from the random-intercept component is an $m \times m$ diagonal matrix. Such streamlining essentially removes computational obstacles involving variances for models such as (1) for most practical values of m and, thus, greatly benefits semiparametric mixed model analysis.

Section 2 gives the details of our streamlined approach to variance calculations for models like (1). In Section 3, we describe the extension of the approach to subject-specific curve models. Closing remarks are made in Section 4.

2. ADDITIVE MIXED MODELS

In this section, we consider a more general version of (1):

$$y_{ij} = f(s_{ij}) + \mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_{\mathbf{x}} + U_i + \varepsilon_{ij} \tag{2}$$

Here, y_{ij} is the jth ($1 \leqslant j \leqslant n_i$) measurement of the response of the ith subject ($1 \leqslant i \leqslant m$), s_{ij} is a predictor with a possibly non-linear effect and $\mathbf{x}_{ij}$ is a $p \times 1$ vector of predictors with a linear effect, with corresponding coefficient vector $\boldsymbol{\beta}_{\mathbf{x}}$. Again U_i i.i.d. $\mathrm{N}(0, \sigma_U^2)$ is a random intercept and ε_{ij} i.i.d. $\mathrm{N}(0, \sigma_\varepsilon^2)$, independently of the U_i's. The smooth function f is modelled using penalized splines of the form

$$f(s) = \beta_0 + \beta_1 s + \sum_{k=1}^{K} u_k z_k(s)$$

where z_k, $1 \leqslant k \leqslant K$, is an appropriate spline basis. The simplest choice is $z_k(s) = (s - \kappa_k)_+$ for a dense set of knots $\kappa_1, \ldots, \kappa_K$ but many other options exist (e.g. Reference [6, Section 3.7]). If we let u_k i.i.d. $\mathrm{N}(0, \sigma_u^2)$ then (2) becomes a linear mixed model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$, where

$$\boldsymbol{\beta} = [\beta_0, \beta_1, \boldsymbol{\beta}_{\mathbf{x}}{}^{\mathrm{T}}]^{\mathrm{T}}, \quad \mathbf{Z} = [\mathbf{Z}_G \ \ \mathbf{Z}_R], \quad \mathbf{u} = [\mathbf{u}_G^{\mathrm{T}} \ \ \mathbf{u}_R^{\mathrm{T}}]^{\mathrm{T}}$$

$$\mathbf{X} = \begin{bmatrix} 1 & s_{11} & \mathbf{x}_{11}^{\mathrm{T}} \\ \vdots & \vdots & \vdots \\ 1 & s_{1n_1} & \mathbf{x}_{1n_1}^{\mathrm{T}} \\ \vdots & \vdots & \vdots \\ 1 & s_{m1} & \mathbf{x}_{m1}^{\mathrm{T}} \\ \vdots & \vdots & \vdots \\ 1 & s_{mn_m} & \mathbf{x}_{mn_m}^{\mathrm{T}} \end{bmatrix}, \quad \mathbf{Z}_G = \begin{bmatrix} z_1(s_{11}) & \cdots & z_K(s_{11}) \\ \vdots & & \vdots \\ z_1(s_{1n_1}) & \cdots & z_K(s_{1n_1}) \\ \vdots & & \vdots \\ z_1(s_{m1}) & \cdots & z_K(s_{m1}) \\ \vdots & & \vdots \\ z_1(s_{mn_m}) & \cdots & z_K(s_{mn_m}) \end{bmatrix} \tag{3}$$

$$\mathbf{Z}_R = \underset{1 \leqslant i \leqslant m}{\mathrm{blockdiag}}(\mathbf{1}_{n_i}), \quad \mathbf{u}_G = [u_1, \ldots, u_K]^{\mathrm{T}}, \quad \mathbf{u}_R = [U_1, \ldots, U_m]^{\mathrm{T}}$$

$$\mathbf{G}_G = \mathrm{cov}(\mathbf{u}_G) = \sigma_u^2 \mathbf{I}_K, \quad \mathbf{G}_R = \mathrm{cov}(\mathbf{u}_R) = \sigma_U^2 \mathbf{I}_m$$

and the vectors $\mathbf{y}$ and $\boldsymbol{\varepsilon}$ are defined analogously. Note that the random effects have been partitioned into spline coefficients (subscript G) and subject effects (subscript R).

Let $\widehat{\mathbf{G}}_G$ and $\widehat{\mathbf{G}}_R$ be the restricted maximum likelihood estimates of $\mathbf{G}_G$ and let $\mathbf{G}_R$ and let $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{u}}$ be the empirical best linear unbiased predictors of $\boldsymbol{\beta}$ and $\mathbf{u}$ (e.g. Reference [6, Section 4.5]). Then all variance calculations can be done using the estimated covariance matrix

$$\widehat{\text{cov}}\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ \widehat{\mathbf{u}}-\mathbf{u}\end{bmatrix}=\mathbf{M}^{-1}$$

where

$$\mathbf{M}=\widehat{\sigma}_\varepsilon^{-2}\mathbf{C}^{\mathrm{T}}\mathbf{C}+\widehat{\mathbf{B}},\quad \mathbf{C}=[\mathbf{X}\ \mathbf{Z}]\quad\text{and}\quad \widehat{\mathbf{B}}=\text{blockdiag}(\mathbf{0},\widehat{\mathbf{G}}_G^{-1},\widehat{\mathbf{G}}_R^{-1})$$

(e.g. Reference [6, Section 4.7]).

The direct approach to obtaining standard errors of the entries of $\widehat{\boldsymbol{\beta}}$ (as in Table I) and variability bars for the smooth function estimate (as in Figure 2) involves inversion of $\mathbf{M}$. As mentioned in Section 1, the matrix $\mathbf{M}$ will increase in dimension as the number of subjects, m grows, and for very large m it will become too computationally intensive to invert practically, since inversion is an $O(m^3)$ process. However, we do not need to find all the entries of the matrix. For example, some entries relate to the correlation between different subjects' responses, which would rarely be useful. We can also exploit the fact that, for large m, most of $\mathbf{M}$ is diagonal. Hence, we propose a streamlined approach based around a block decomposition of $\mathbf{M}$:

$$\mathbf{M}=\widehat{\sigma}_\varepsilon^{-2}\left[\begin{array}{cc|c}\mathbf{X}^{\mathrm{T}}\mathbf{X} & \mathbf{X}^{\mathrm{T}}\mathbf{Z}_G & \mathbf{X}^{\mathrm{T}}\mathbf{Z}_R\\ \mathbf{Z}_G^{\mathrm{T}}\mathbf{X} & \mathbf{Z}_G^{\mathrm{T}}\mathbf{Z}_G+\widehat{\sigma}_\varepsilon^2\widehat{\mathbf{G}}_G^{-1} & \mathbf{Z}_G^{\mathrm{T}}\mathbf{Z}_R\\ \hline \mathbf{Z}_R^{\mathrm{T}}\mathbf{X} & \mathbf{Z}_R^{\mathrm{T}}\mathbf{Z}_G & \mathbf{Z}_R^{\mathrm{T}}\mathbf{Z}_R+\widehat{\sigma}_\varepsilon^2\widehat{\mathbf{G}}_R^{-1}\end{array}\right]\equiv\widehat{\sigma}_\varepsilon^{-2}\begin{bmatrix}\mathbf{M}_{11} & \mathbf{M}_{12}\\ \mathbf{M}_{21} & \mathbf{M}_{22}\end{bmatrix}$$

Using standard results on the inverse of a block-partitioned matrix (e.g. Reference [12])

$$\mathbf{M}^{-1}=\widehat{\sigma}_\varepsilon^{2}\begin{bmatrix}\mathbf{M}^{11} & \mathbf{M}^{12}\\ \mathbf{M}^{21} & \mathbf{M}^{22}\end{bmatrix}$$

where

$$\begin{aligned}&\mathbf{M}^{11}=(\mathbf{M}_{11}-\mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{M}_{21})^{-1}\\ &\mathbf{M}^{12}=-\mathbf{M}^{11}\mathbf{M}_{12}\mathbf{M}_{22}^{-1},\quad \mathbf{M}^{21}=(\mathbf{M}^{12})^{\mathrm{T}}\end{aligned}\tag{4}$$

and

$$\mathbf{M}^{22}=\mathbf{M}_{22}^{-1}-\mathbf{M}_{22}^{-1}\mathbf{M}_{21}\mathbf{M}^{11}\mathbf{M}_{12}\mathbf{M}_{22}^{-1}$$

Straightforward matrix multiplication finds $\mathbf{X}^{\mathrm{T}}\mathbf{X}$, $\mathbf{X}^{\mathrm{T}}\mathbf{Z}_G$ and $\mathbf{Z}_G^{\mathrm{T}}\mathbf{Z}_G$ in $O(m)$ steps, but $\mathbf{Z}_R^{\mathrm{T}}\mathbf{X}$ and $\mathbf{Z}_R^{\mathrm{T}}\mathbf{Z}_G$ are more complicated since $\mathbf{Z}_R$ is at least $m\times m$. However, the special structure of $\mathbf{Z}_R$

DOI: 10.1002/sim

means that the matrix products amount to within-subject row sums, which are $O(m)$, since for $\mathbf{A}=\mathbf{X}$ or $\mathbf{A}=\mathbf{Z}_G$

$$\mathbf{Z}_R^{\mathrm{T}}\mathbf{A}=\begin{bmatrix} 1 & \cdots & 1 & & & & \mathbf{0} \\ & & & \ddots & & & \\ \mathbf{0} & & & & 1 & \cdots & 1 \end{bmatrix}\begin{bmatrix} \mathbf{a}_{11}^{\mathrm{T}} \\ \vdots \\ \mathbf{a}_{1n_1}^{\mathrm{T}} \\ \vdots \\ \mathbf{a}_{m1}^{\mathrm{T}} \\ \vdots \\ \mathbf{a}_{mn_m}^{\mathrm{T}} \end{bmatrix}=\begin{bmatrix} \sum_{j=1}^{n_1}\mathbf{a}_{1j}^{\mathrm{T}} \\ \vdots \\ \sum_{j=1}^{n_m}\mathbf{a}_{mj}^{\mathrm{T}} \end{bmatrix}$$

In addition, $\mathbf{M}_{22}$ is diagonal:

$$\mathbf{Z}_R^{\mathrm{T}}\mathbf{Z}_R+\widehat{\sigma}_\varepsilon^2\widehat{\sigma}_U^{-2}\mathbf{I}_m=\operatorname*{diag}_{1\leqslant i\leqslant m}(n_i+\widehat{\sigma}_\varepsilon^2/\widehat{\sigma}_U^2)$$

This feature is crucial. It enables us to find the biggest inverse required in (4) in $O(m)$ steps since $\mathbf{M}_{22}^{-1}=\operatorname{diag}_{1\leqslant i\leqslant m}(\widehat{\sigma}_U^2/(\widehat{\sigma}_U^2 n_i+\widehat{\sigma}_\varepsilon^2))$. Thus, we can write $\mathbf{M}$ as

$$\widehat{\sigma}_\varepsilon^{-2}\begin{bmatrix}\mathbf{M}_{11} & \mathbf{M}_{12}\\ \mathbf{M}_{21} & \mathbf{M}_{22}\end{bmatrix}=\widehat{\sigma}_\varepsilon^{-2}\left[\begin{array}{c|ccc} \mathbf{M}_{11} & \mathbf{h}_1 & \cdots & \mathbf{h}_m \\ \hline \mathbf{h}_1^{\mathrm{T}} & n_1+\dfrac{\widehat{\sigma}_\varepsilon^2}{\widehat{\sigma}_U^2} & & \mathbf{0} \\ \vdots & & \ddots & \\ \mathbf{h}_m^{\mathrm{T}} & \mathbf{0} & & n_m+\dfrac{\widehat{\sigma}_\varepsilon^2}{\widehat{\sigma}_U^2} \end{array}\right]$$

where $\mathbf{h}_i^{\mathrm{T}}=\sum_{j=1}^{n_i}[1\ \ s_{ij}\ \ \mathbf{x}_{ij}^{\mathrm{T}}\ \ z_1(s_{ij})\ \ \cdots\ \ z_K(s_{ij})]$. Then using (4) we see that

$$\mathbf{M}^{11}=\left(\mathbf{M}_{11}-\sum_{i=1}^{m}\frac{\widehat{\sigma}_U^2\mathbf{h}_i\mathbf{h}_i^{\mathrm{T}}}{\widehat{\sigma}_U^2 n_i+\widehat{\sigma}_\varepsilon^2}\right)^{-1}$$

This inverse is $(2+p+K)\times(2+p+K)$ and is relatively easy to compute. The summation term renders this whole process $O(m)$. This matrix is all we require to plot global or group error bars, or perform hypothesis tests as in Table I. However, to find the error covariance of the individual subjects' fitted responses we need $\mathbf{M}^{12}$ and $\mathbf{M}^{22}$, where

$$\mathbf{M}^{12}=\left[-\frac{\widehat{\sigma}_U^2}{\widehat{\sigma}_U^2 n_1+\widehat{\sigma}_\varepsilon^2}\mathbf{M}^{11}\mathbf{h}_1\quad\cdots\quad-\frac{\widehat{\sigma}_U^2}{\widehat{\sigma}_U^2 n_m+\widehat{\sigma}_\varepsilon^2}\mathbf{M}^{11}\mathbf{h}_m\right]$$

Each submatrix is a constant times $(2+p+K)\times(2+p+K)$ times $(2+p+K)\times 1$. However, we must calculate m of them, hence $\mathbf{M}^{12}$ (and consequently $\mathbf{M}^{21}$) requires $O(m)$ calculations.

Finally, we only need to find the diagonal entries of $\mathbf{M}^{22}$ since we would rarely be interested in the correlation between two subjects' fitted responses. These are given by

$$\mathbf{M}^{22}_{ii} = \frac{\widehat{\sigma}^2_U}{\widehat{\sigma}^2_U n_i + \widehat{\sigma}^2_\varepsilon}\left(1 + \frac{\widehat{\sigma}^2_U\,\mathbf{h}_i^{\mathrm{T}}\mathbf{M}^{11}\mathbf{h}_i}{\widehat{\sigma}^2_U n_i + \widehat{\sigma}^2_\varepsilon}\right), \quad 1\leqslant i\leqslant m$$

The relevant diagonal entries can be calculated in $O(m)$ steps. Therefore, the total asymptotic complexity of this process is $O(m)$, representing an improvement of order m^2 over the naïve approach to matrix inversion. As the number of subjects, m increases, the improvements due to streamlining become enormous.

We could alter the model in (2) to include more random subject effects, such as random slopes. This would only affect the final stage of the calculation as it would alter the structure of $\mathbf{M}_{22}$. However, the result of this is that $\mathbf{M}_{22}$ becomes block diagonal, hence, it can still be inverted in $O(m)$ calculations.

The practical benefits of streamlined variance calculations were explored in a simulation study. Data were generated according to

$$y_{ij} = -\sin(2\pi s_{ij}) + 0.3\,x_{ij} + U_i + \varepsilon_{ij}, \quad 1\leqslant i\leqslant n_i, \;\; 1\leqslant i\leqslant m \tag{5}$$

where the s_{ij} were generated from the uniform distribution on (0, 1), the x_{ij} were generated from the Bernoulli distribution with $P(x_{ij}=0)=P(x_{ij}=1)=\frac{1}{2}$, U_i i.i.d. $\mathrm{N}(0, 0.5^2)$ and ε_{ij} i.i.d. $\mathrm{N}(0, 0.2^2)$. The $0.3\,x_{ij}$ term represents a binary offset from the smooth function in s_{ij}. The within-subject sample sizes n_i were generated uniformly from $\{1, 2, 3, 4\}$ and values of $m \in \{100, 500, 2500, 12\,500\}$ were considered. For each value of m, 25 replicate data sets were generated. All computations were done in the `R` language [13] on a Dell Optiplex SX280 PC with a 2.8 GHz Intel Pentium 4 Processor. The Appendix contains code used for the streamlined variance calculations. The time taken to compute variability bars and the standard error of the offset estimate was recorded. Table II summarizes the results. Note that the naïve method approach failed for $m=12\,500$ due to required storage for $\mathbf{M}$ exceeding memory restrictions.

There is little practical difference between the two methods for $m=100$ and 500. However, for $m=2500$ streamlined variance calculation is much faster—taking about one-tenth of a second on average compared with almost 4 min for the naïve approach. For $m=12\,500$ the streamlined approach is still well under 1 s, while the naïve approach is not viable for typical 2006 computing environments.

Table II. Average times in seconds (standard errors) for computing variability bars and standard error of the binary offset.

m	Naïve	Streamlined	Ratio
100	0.183 (0.294)	0.0114 (0.00810)	15.2
500	1.43 (1.89)	0.0262 (0.00881)	57.9
2500	230 (54.0)	0.110 (0.00872)	2110
12 500	Failed	0.599 (0.0902)	N/A

3. EXTENSION TO SUBJECT-SPECIFIC CURVES

Models (1) and (2) featured *random intercepts*: the difference between the fitted subject response and the estimated population mean curve is constant. This may not in general be realistic; the subject-specific difference may be as complicated as the underlying function f. Durban *et al.* [14] develop a *subject-specific curves* model, based on penalized splines, in which the subject-specific difference is modelled by a random semiparametric function

$$y_{ij}=f(s_{ij})+\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_{\mathbf{x}}+g_i(s_{ij})+\varepsilon_{ij} \tag{6}$$

Earlier work on models of this type includes [2, 3, 5, 15, 16]. In this new model U_i is replaced by

$$g_i(s)=U_{i0}+U_{i1}s+\sum_{i=1}^{K_g}\upsilon_{ik}\,z_{gk}(s),\qquad \upsilon_{gk}\ \text{i.i.d. N}(0,\sigma_\upsilon^2),$$

where z_{gk}, $1\leqslant k\leqslant K_g$, is an appropriate spline basis. It is an advantage of our set-up that g_i need not share the same spline basis as f. Therefore, the splines of f are labelled z_{fk}, $1\leqslant k\leqslant K_f$. Model (6) can be written as a linear mixed model $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u}+\boldsymbol{\varepsilon}$ where

$$\mathbf{X}=\begin{bmatrix}\mathbf{X}_1\\ \vdots\\ \mathbf{X}_m\end{bmatrix},\quad \mathbf{Z}=\begin{bmatrix}\mathbf{Z}_1 & \mathbf{T}_1\mathbf{W}_1 & & \mathbf{0}\\ \vdots & & \ddots & \\ \mathbf{Z}_m & \mathbf{0} & & \mathbf{T}_m\mathbf{W}_m\end{bmatrix}$$

$$\mathbf{X}_i=\begin{bmatrix}1 & s_{i1} & \mathbf{x}_{i1}^{\mathrm{T}}\\ \vdots & \vdots & \vdots\\ 1 & s_{in_i} & \mathbf{x}_{in_i}^{\mathrm{T}}\end{bmatrix},\quad \mathbf{Z}_i=\begin{bmatrix}z_{f1}(s_{i1}) & \cdots & z_{fK_f}(s_{i1})\\ \vdots & & \vdots\\ z_{f1}(s_{in_i}) & \cdots & z_{fK_f}(s_{in_i})\end{bmatrix}$$

$$\mathbf{T}_i=\begin{bmatrix}1 & s_{i1}\\ \vdots & \vdots\\ 1 & s_{in_i}\end{bmatrix},\quad \mathbf{W}_i=\begin{bmatrix}z_{g1}(s_{i1}) & \cdots & z_{gK_g}(s_{i1})\\ \vdots & & \vdots\\ z_{g1}(s_{in_i}) & \cdots & z_{gK_g}(s_{in_i})\end{bmatrix}$$

$$\mathbf{u}=[\mathbf{u}_G^{\mathrm{T}},\mathbf{U}_1^{\mathrm{T}},\boldsymbol{\upsilon}_1^{\mathrm{T}},\ldots,\mathbf{U}_m^{\mathrm{T}},\boldsymbol{\upsilon}_m^{\mathrm{T}}]^{\mathrm{T}},\quad \mathbf{U}_i=[U_{i0},U_{i1}]^{\mathrm{T}},\quad \boldsymbol{\upsilon}_i=[\upsilon_{i1},\ldots,\upsilon_{iK_g}]^{\mathrm{T}}$$

$$\mathbf{G}_G=\operatorname{cov}[\mathbf{U}_1^{\mathrm{T}},\boldsymbol{\upsilon}_1^{\mathrm{T}},\ldots,\mathbf{U}_m^{\mathrm{T}},\boldsymbol{\upsilon}_m^{\mathrm{T}}]^{\mathrm{T}}=\operatorname{blockdiag}(\mathbf{D},\sigma_\upsilon^2\mathbf{I}_{K_g},\ldots,\mathbf{D},\sigma_\upsilon^2\mathbf{I}_{K_g})$$

and $\mathbf{y}$, $\boldsymbol{\beta}$, $\boldsymbol{\varepsilon}$, $\mathbf{u}_G$ and $\mathbf{G}_G$ are the same as in (3). $\mathbf{D}$ is a general, symmetric, 2×2 matrix. Note that the $\mathbf{Z}$ matrix takes a different form from that given in [14] and Section 9.3 of [6]. It has been changed to make the calculations more manageable.

As before, the variance calculations require the estimated covariance matrix $\mathbf{M}^{-1}$. This matrix is now even larger than that of Section 2. Therefore, we propose a streamlined approach, again based around a block decomposition of $\mathbf{M}$:

$$\widehat{\sigma_\varepsilon}^{-2}\begin{bmatrix}\mathbf{M}_{11} & \mathbf{M}_{12}\\ \mathbf{M}_{21} & \mathbf{M}_{22}\end{bmatrix} \equiv \widehat{\sigma_\varepsilon}^{-2}\left[\begin{array}{c|ccc}\mathbf{M}_{11} & \mathbf{H}_1 & \cdots & \mathbf{H}_m\\ \hline \mathbf{H}_1^{\mathrm{T}} & \mathbf{L}_1 & & \mathbf{0}\\ \vdots & & \ddots & \\ \mathbf{H}_m^{\mathrm{T}} & \mathbf{0} & & \mathbf{L}_m\end{array}\right]$$

where

$$\mathbf{M}_{11} = \begin{bmatrix}\sum_{i=1}^m \mathbf{X}_i^{\mathrm{T}}\mathbf{X}_i & \sum_{i=1}^m \mathbf{X}_i^{\mathrm{T}}\mathbf{Z}_i\\ \sum_{i=1}^m \mathbf{Z}_i^{\mathrm{T}}\mathbf{X}_i & \sum_{i=1}^m \mathbf{Z}_i^{\mathrm{T}}\mathbf{Z}_i + \lambda\mathbf{I}_{K_f}\end{bmatrix}, \quad \lambda = \frac{\widehat{\sigma_\varepsilon^2}}{\widehat{\sigma_u^2}}$$

$$\mathbf{H}_i = \begin{bmatrix}\mathbf{X}_i^{\mathrm{T}}\mathbf{T}_i & \mathbf{X}_i^{\mathrm{T}}\mathbf{W}_i\\ \mathbf{Z}_i^{\mathrm{T}}\mathbf{T}_i & \mathbf{Z}_i^{\mathrm{T}}\mathbf{W}_i\end{bmatrix}$$

and

$$\mathbf{L}_i = \begin{bmatrix}\mathbf{T}_i^{\mathrm{T}}\mathbf{T}_i + \widehat{\sigma_\varepsilon^2}\hat{\mathbf{D}}^{-1} & \mathbf{T}_i^{\mathrm{T}}\mathbf{W}_i\\ \mathbf{W}_i^{\mathrm{T}}\mathbf{T}_i & \mathbf{W}_i^{\mathrm{T}}\mathbf{W}_i + \tau\mathbf{I}_{K_g}\end{bmatrix}, \quad \tau = \frac{\widehat{\sigma_\varepsilon^2}}{\widehat{\sigma_v^2}}$$

In this model matrix $\mathbf{M}_{22}$, the contribution from the random subject component, is block diagonal. It can therefore be inverted in $O(m)$ steps, enabling streamlined variance calculations, again using (4):

$$\mathbf{M}^{-1} = \widehat{\sigma_\varepsilon^2}\begin{bmatrix}\mathbf{M}^{11} & \mathbf{M}^{12}\\ \mathbf{M}^{21} & \mathbf{M}^{22}\end{bmatrix}$$

where

$$\mathbf{M}^{11} = \left(\mathbf{M}_{11} - \sum_{i=1}^m \mathbf{H}_i\mathbf{L}_i^{-1}\mathbf{H}_i^{\mathrm{T}}\right)^{-1}$$

$$\mathbf{M}^{12} = [-\mathbf{M}^{11}\mathbf{H}_1\mathbf{L}_1^{-1} \ \cdots \ -\mathbf{M}^{11}\mathbf{H}_m\mathbf{L}_m^{-1}]$$

and

$$\mathbf{M}_{ii}^{22} = \mathbf{L}_i^{-1} + \mathbf{L}_i^{-1}\mathbf{H}_i^{\mathrm{T}}\mathbf{M}^{11}\mathbf{H}_i\mathbf{L}_i^{-1}, \quad 1\leqslant i\leqslant m$$

Since $\mathbf{M}_{11}$, $\mathbf{H}_i$ and $\mathbf{L}_i$ have dimensions much smaller than m, the complexity of the matrix calculations required in these submatrices does not increase as m increases. Therefore, the

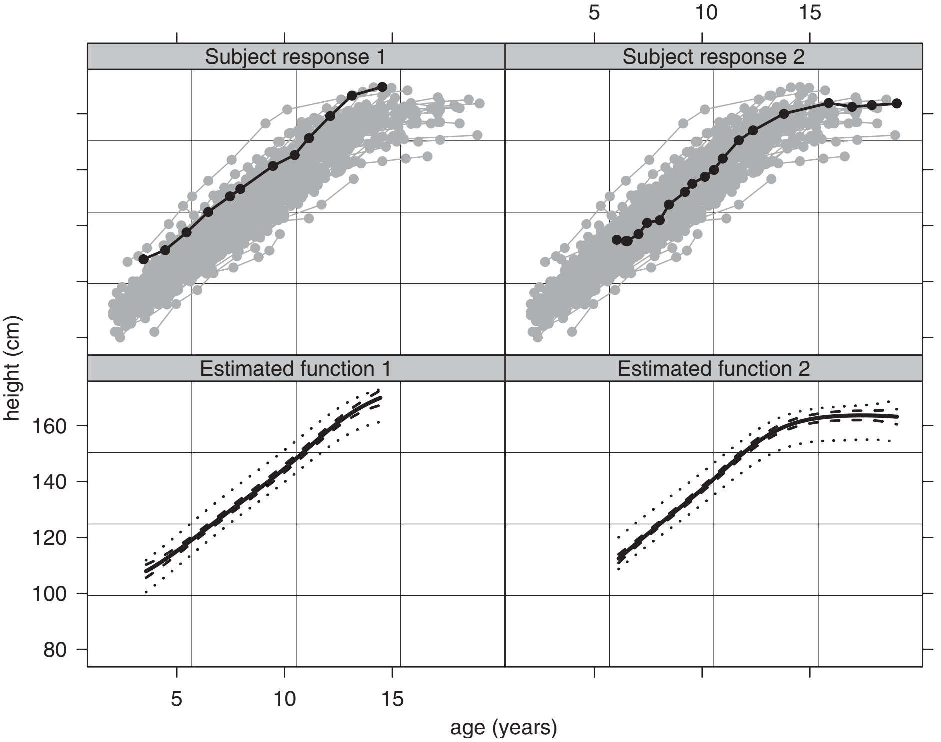

Figure 3. Heights of 190 girls with acute lymphoblastic leukaemia. We have highlighted two subjects (one column for each) with different shaped responses. The top row shows the subjects' individual response profiles, and the bottom row shows the estimated subject-specific fitted responses (solid lines). The bottom panels also show ±2 × estimated standard error bars from both model (7) (dashed lines) and a random-intercepts model (dotted lines).

calculations with the highest order of complexity are the summations in $\mathbf{M}_{11}$ and $\mathbf{M}^{11}$, and the calculation of the m relevant submatrices of $\mathbf{M}^{12}$, $\mathbf{M}^{21}$ and $\mathbf{M}^{22}$. This renders the whole process as $O(m)$, preserving the improvement over naïve inversion that we already witnessed for random intercepts.

Durban *et al.* [14] fit a semiparametric model with subject-specific curves to longitudinal data on the heights of 190 girls with acute lymphoblastic leukaemia. We fit a similar model in which height is modelled as a smooth function of age:

$$\texttt{height}_{ij} = f(\texttt{age}_{ij}) + g_i(\texttt{age}_{ij}) + \varepsilon_{ij} \tag{7}$$

where f and g_i are modelled using radial cubic splines with 15 and 10 knots, respectively.

Figure 3 shows some fitted functions and error bars from this model, superimposed on the data. We have drawn out two subjects' responses that differ from each other in order to show how the subject-specific curves model can produce different estimated smooth curves for each subject. We have plotted error bars according to the subject-specific curves model (7) and also a random-intercepts model such as (2), both obtained *via* the streamlined method. The random-intercepts error bars are much wider than those of the subject-specific curves model.

4. CLOSING REMARKS

The use of semiparametric regression in the analysis of longitudinal data has become commonplace in the last decade. The sample sizes of data sets arising from longitudinal studies are also on the increase, and semiparametric regression methods are susceptible to breakdown if implemented naïvely. The methods developed in this article overcome such problems, even for massive data sets, and therefore are recommended for general practice.

APPENDIX: R IMPLEMENTATION

In this appendix we provide the `R` code that demonstrates the streamlined variance calculations described in Section 2, for data simulated according to (5). In December 2006 this code successfully ran on version 2.4.0 of `R` under the `Linux` operating system on the second author's computer and produced the output

```
Estimated effect of x=1 and standard error is:

  est.  s.e.
0.3080 0.0661
```

and the plot shown in Figure A1. A text file containing this code is available on request from the second author. At the time of writing his e-mail address is `mwand@uow.edu.au`.

Load required packages:

```
library(nlme) ; library(lattice)
```

Generate synthetic data according to (5):

```
set.seed(39402) ; m <- 250 ; nVals <- sample(1:4,m,replace=TRUE)
betaVal <- 0.3 ; sigU <- 0.5 ; sigEps <- 0.2
f <- function(x) return(-sin(2*pi*x))
U <- rnorm(m,0,sigU)
sVals <- NULL ; xVals <- NULL
Uvals <- NULL ; idNum <- NULL
for (i in 1:m)
{
  idNum <- c(idNum,rep(i,nVals[i]))
  stt <- runif(1,0,1-0.05*(nVals[i]-1))
  sVals <- c(sVals,seq(stt,by=0.05,length=nVals[i]))
  xCurr <- sample(c(0,1),1)
  xVals <- c(xVals,rep(xCurr,nVals[i]))
  Uvals <- c(Uvals,rep(U[i],nVals[i]))
}
epsVals <- rnorm(sum(nVals),0,sigEps)
yVals <- f(sVals) + betaVal*xVals + Uvals + epsVals
```

Fit an additive mixed model:

```
numKnots <- 15 ; p <- 1
knots <- quantile(unique(sVals),seq(0,1,length=
           (numKnots+2))[-c(1,(numKnots+2))])
```

DOI: 10.1002/sim

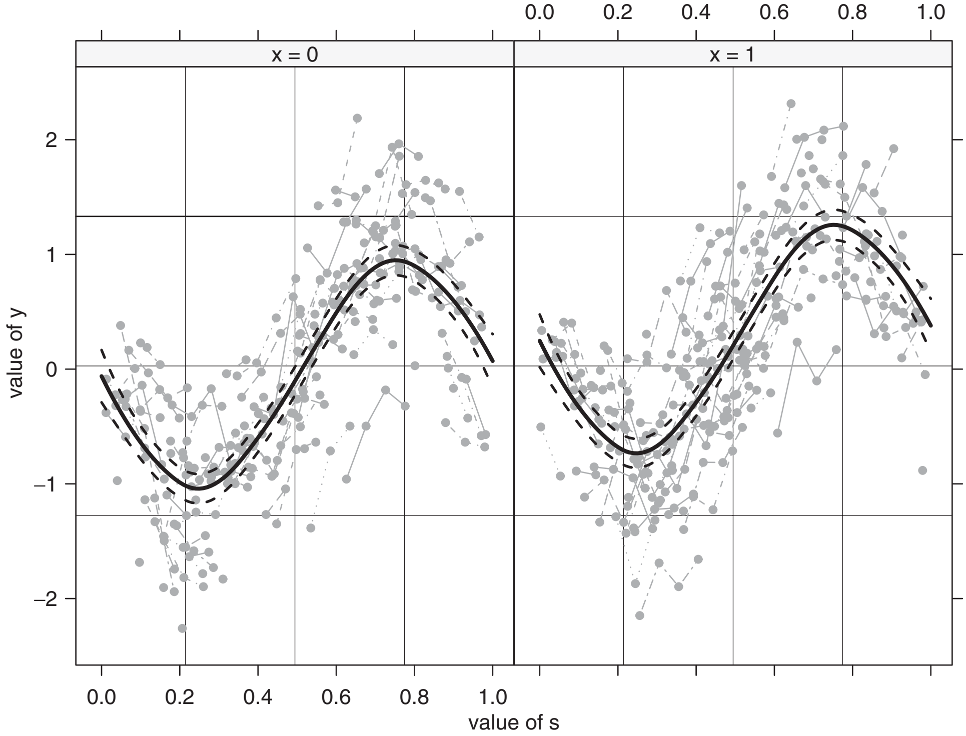

Figure A1. Plot obtained from running the code in this appendix.

```
X <- cbind(rep(1,length(yVals)),sVals,xVals)
svd.Omega <- svd(abs(outer(knots,knots,"-"))^3)
matrix.sqrt.Omega <- t(svd.Omega$v%*%(t(svd.Omega$u)*sqrt(svd.Omega$d)))
ZSpline <- t(solve(matrix.sqrt.Omega,t(abs(outer(sVals,knots,"-")^3))))
groupVec <- factor(rep(1,length(yVals)))
ZBlock <- list(list(groupVec=pdIdent(~ZSpline-1)),list(idNum=pdIdent(~1)))
ZBlock <- unlist(ZBlock,recursive=FALSE)
dataFr <- groupedData(yVals~xVals|groupVec,
              data=data.frame(yVals,X,ZSpline,idNum))
fit <- lme(yVals~-1+X,data=dataFr,random=ZBlock)
betaHat <- fit$coef$fixed
uHat <- unlist(fit$coef$random)
uSplineHat <- uHat[1:ncol(ZSpline)]
```

Perform streamlined variance estimation for variability bars and standard error of binary predictor:

```
sig.eps.hat <- fit$sigma
sig.u.hat <- intervals(fit)$reStruct$groupVec$est
sig.U.hat <- intervals(fit)$reStruct$idNum$est
M11inv <- diag(c(rep(0,p+2),rep(sig.eps.hat^2/sig.u.hat^2,numKnots)))
M11inv <- M11inv + crossprod(cbind(X,ZSpline))
M.21 <- rowsum(cbind(X,ZSpline),idNum)
```

```
for (i in 1:m)
{
  h.i <- M.21[i,] ; n.i <- M.21[i,1]
  M11inv <- (M11inv - sig.U.hat^2*tcrossprod(h.i)/
               (sig.U.hat^2*n.i+sig.eps.hat^2))
}
covMat <- (sig.eps.hat^2)*solve(M11inv)
```

Report results:

```
cat("\n Estimated effect of x=1 and standard error is:\n\n")
outp <- c(betaHat[3],sqrt(covMat[3,3]))
names(outp) <- c("est.","s.e.")
print(signif(outp,3)) ; cat("\n\n")
ng <- 101 ; sValsg <- seq(0,1,length=ng)
ZgSpline <- t(solve(matrix.sqrt.Omega,t(abs(outer(sValsg,knots,"-")^3))))
plotMatrix0 <- cbind(rep(1,ng),sValsg,rep(0,ng),ZgSpline)
fhatgREML <- plotMatrix0 %*% c(betaHat,uSplineHat)
Var0 <- diag(plotMatrix0%*%covMat%*%t(plotMatrix0))
Var1 <- 2*plotMatrix0%*%covMat[,3] + covMat[3,3]
xLabs <- paste("x =",as.character(xVals))
pobj <- xyplot(yVals~sVals|xLabs,groups=idNum,xlab="value of s",
     ylab="value of y",subscripts=TRUE,
     panel=function(x,y,subscripts,groups)
     {
       panel.grid()  ; panel.superpose(x,y,subscripts,groups,
                    type="b",col="grey60",pch=16)
       panelInd <- any(xVals[subscripts]==1)
       panel.xyplot(sValsg,fhatgREML+panelInd*betaHat[3],
                lwd=3,type="l",col="black")
       for (multFac in c(-2,2))
         panel.xyplot(sValsg,fhatgREML+panelInd*betaHat[3]+multFac
           *sqrt(Var0+panelInd*Var1),lty=2,lwd=2,type="l",col="black")
     })
print(pobj)
```

REFERENCES

1. Wang Y. Smoothing spline models with correlated random errors. *Journal of the American Statistical Association* 1998; **93**:341–348.
2. Brumback BA, Rice JA. Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion). *Journal of the American Statistical Association* 1999; **93**:961–994.
3. Verbyla AP, Cullis BR, Kenward MG, Welham SJ. The analysis of designed experiments and longitudinal data by using smoothing splines (with discussion). *Applied Statistics* 1999; **48**:269–312.
4. Lin X, Carroll RJ. Semiparametric regression for clustered data. *Biometrika* 2001; **88**:1179–1865.
5. Rice JA, Wu CO. Nonparametric mixed effect models for unequally sampled noisy curves. *Biometrics* 2001; **57**:253–259.
6. Ruppert D, Wand MP, Carroll RJ. *Semiparametric Regression*. Cambridge University Press: New York, 2003.
7. Wood SN. *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC: Boca Raton, FL, 2006.
8. Wu H, Zhang J-T. *Nonparametric Regression Methods for Longitudinal Data Analysis*. Wiley: New York, 2006.
9. Bachrach LK, Hastie T, Wang M-C, Narasimhan B, Marcus R. Bone mineral acquisition in healthy Asian, Hispanic, Black and Caucasian youth. A longitudinal study. *Journal of Clinical Endocrinology and Metabolism* 1999; **84**:4702–4712.

10. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. Wiley: Hoboken, NJ, 2004.
11. Cormen TH, Leiserson CE, Rivest RL, Stein C. *Introduction to Algorithms*. MIT Press: Cambridge, MA, 2001.
12. Harville DA. *Matrix Algebra from a Statistician's Perspective*. Springer: New York, 2000.
13. R Development Core Team. *R*: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org, 2006.
14. Durban M, Harezlak J, Wand MP, Carroll RJ. Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine* 2005; **24**:1153–1167.
15. Donnelly CA, Laird NM, Ware JH. Prediction and creation of smooth curves for temporally correlated longitudinal data. *Journal of the American Statistical Association* 1995; **90**:984–989.
16. Guo W. Functional mixed effects models. *Biometrics* 2002; **58**:121–128.