Biostatistics (2015), **16**, 4, *pp.* 754–771
doi:10.1093/biostatistics/kxv015
Advance Access publication on May 18, 2015

Methods to assess an exercise intervention trial based on 3-level functional data

HAOCHENG LI*

Departments of Oncology and Community Health Sciences, University of Calgary, Calgary, AB, T2N 4N2, Canada

haocheng.li@ucalgary.ca

SARAH KOZEY KEADLE

Nutritional Epidemiology Branch, National Cancer Institute, 9609 Medical Center Drive, Bethesda, MD 20892, USA

JOHN STAUDENMAYER

Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003-9305, USA

HOUSSEIN ASSAAD AND JIANHUA Z. HUANG

Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, USA

RAYMOND J. CARROLL

Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, USA and Department of Mathematics and Statistics, University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Australia

Summary

Motivated by data recording the effects of an exercise intervention on subjects' physical activity over time, we develop a model to assess the effects of a treatment when the data are functional with 3 levels (subjects, weeks and days in our application) and possibly incomplete. We develop a model with 3-level mean structure effects, all stratified by treatment and subject random effects, including a general subject effect and nested effects for the 3 levels. The mean and random structures are specified as smooth curves measured at various time points. The association structure of the 3-level data is induced through the random curves, which are summarized using a few important principal components. We use penalized splines to model the mean curves and the principal component curves, and cast the proposed model into a mixed effects model framework for model fitting, prediction and inference. We develop an algorithm to fit the model iteratively with the Expectation/Conditional Maximization Either (ECME) version of the EM algorithm and eigenvalue decompositions. Selection of the number of principal components and handling incomplete data issues are incorporated into the algorithm. The performance of the Wald-type hypothesis test is also discussed. The method is applied to the physical activity data and evaluated empirically by a simulation study.

*To whom correspondence should be addressed.

Keywords: Longitudinal data; Mixed-effects model; Penalized splines; Physical activity measurement; Principal components.

1. Introduction

Motivated by data from an exercise intervention trial, we consider the problem of evaluating the effects of the intervention in the presence of 3-level functional data, with possibly missing, or unbalanced, observations. The data are from Kozey-Keadle *and others* (2014) and consist of estimates of relative energy expenditure (metabolic units, or METs) on 63 inactive individuals every 5 min. We consider data from 5 days a week (Monday through Friday) for 5 separate weeks (study weeks 0, 3, 6, 9 and 12). Figure 1 displays data from one subject as an example. The treatment assignment was made after the baseline week and consisted of assignment to either a control arm or an exercise intervention where subjects completed a standardized aerobic exercise program. The data come from electronic monitors worn by each subject, and incomplete data occur; a subject's record may be missing for some time points, days or weeks.

Our goal is to evaluate the effects of the intervention by comparing the change in relative energy expenditure in weeks 3, 6, 9 and 12 to that of the baseline across the treatment groups. Additionally, we are interested in the patterns of physical activity for individuals at different time scales (within days, across days of week and across weeks) and how these patterns vary across the treatment groups. As a result, our goal is to model the responses influenced by 2 factors, days within weeks, while also addressing the fact that these are nested within a third factor, subject.

The data are functional, and we approach this problem using a combination of linear mixed models, penalized B-spline smoothing, and principal component-based dimension reduction. The paper's primary contribution is in its application. Combined with functional data analysis methods, emerging physical activity monitor technology provides a surprising amount of insight into personal behaviors, and these detailed data give a new way to test for the effects of a physical activity intervention. In addition, the 3-level model in our motivating example, presents computational challenges. Current studies have mainly focused on 2-level models. For example, Di *and others* (2009), Zipunnikov *and others* (2011) and Serban and Jiang (2012) discuss 2-level multilevel functional analysis approaches with subunit data nested within each unit. Methodology for the analysis of 3-level functional data is limited, and developing a general methodology for 3-level functional data is necessary.

Although the extension from a 2-level to a 3-level model may seem straightforward, our saturated 3-level model involves much more complicated structures, ones that present non-trivial modeling and computational challenges. To be specific, the fixed effects in the 3-level model involve functional curves on different weeks and days: importantly, we allow the possibility of interactions between them and stratify by treatment. In addition, besides the random effects for the subject level, our model includes random curves for the week, the day of the week and interactions between days and weeks. We thus propose a general model with mean structures for week-specific, day-specific and week $\times$ day interaction effects stratified by treatment, and random effects structures for subject-specific, week-specific, day-specific and week $\times$ day interaction variation.

Since we employ a 3-level structure of random curves, dimensionality problems arise that can affect computation and model explanation. The problem can be handled via the dimension reduction approach discussed by Zhou *and others* (2008); see also Zhou *and others* (2010), which uses a few important principal components to summarize random curves. A typical issue for using dimension reduction techniques is to determine the number of principal components and estimate the principal component vectors. Zhou *and others* (2008) use a model-based approach, which estimates the parameters with an EM algorithm. For improved computational efficiency, we propose a new algorithm that includes features of a likelihood approach and eigen-decomposition.

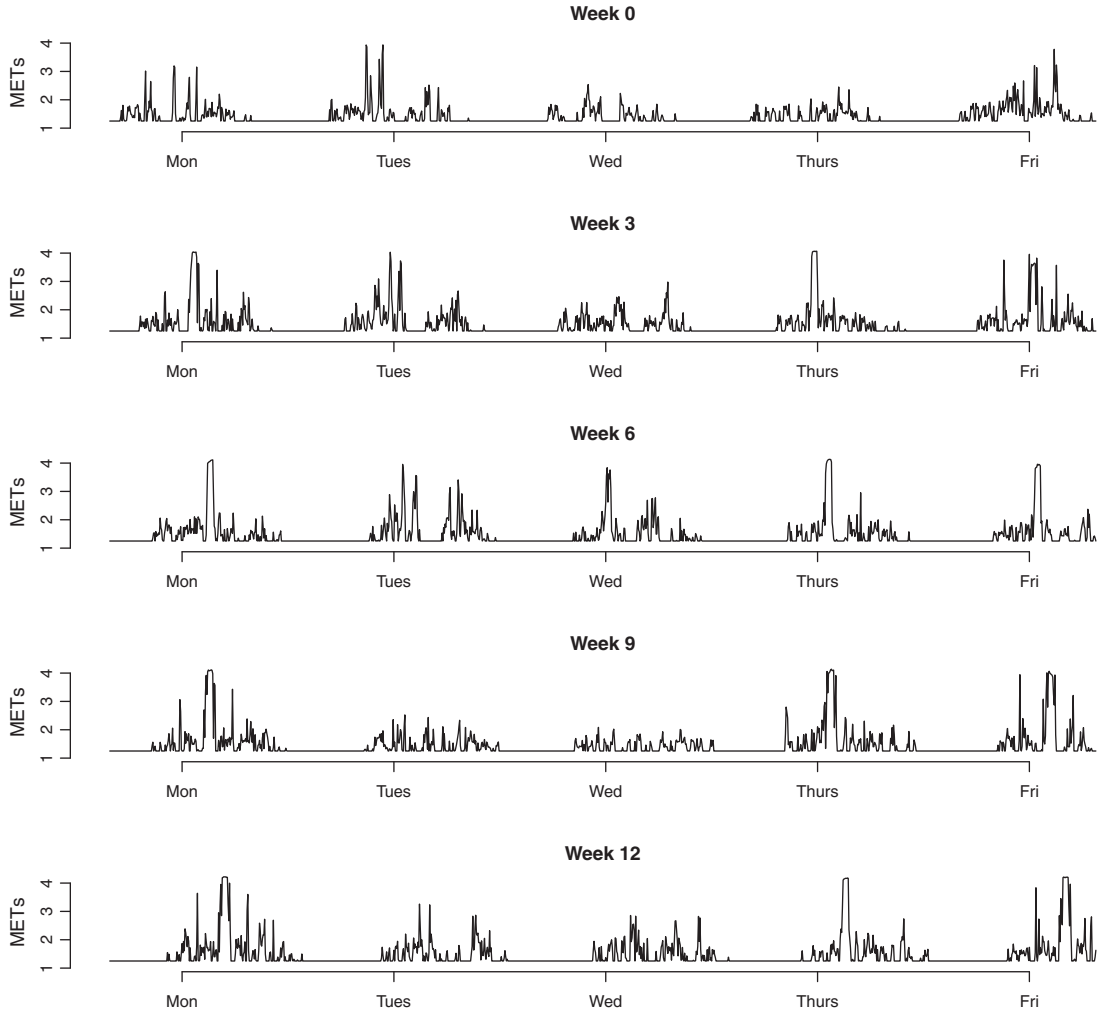

Fig. 1. An example of the estimates of METs for a subject in the exercise group.

Current functional data frameworks focus on complete data (Zhou *and others*, 2010). Missing or incomplete data are common in applications though, and estimation methods developed for complete data scenarios may fail in that case. Assuming our model specification is correct, we only require that the data be missing at random (Little and Rubin, 2002), and we develop a modified estimation algorithm to accommodate the incomplete data. In particular, for each person, complete data have identical dimensions of design matrices in both fixed and random effects. All individuals also have the same covariance structures. However, missing data may not have such features. Thus, our algorithm is updated to adjust the differences in design matrices and variance structures between subjects.

Since our model involves a treatment and week-specific, day-specific and week $\times$ day interaction effects for the fixed effects mean structure, techniques to test whether these fixed effect mean components are necessary. We study the performance of the Wald test under our model settings. We show that care must be taken when the parameter estimates are obtained from the penalized likelihood. The statistics from the

penalized likelihood method lead to systematically anti-conservative p-values, but we use simulation to demonstrate that a Wald statistic based on an unpenalized likelihood can have an acceptable test level.

The paper is organized as follows. Section 2 describes the model, and Section 3 describes our algorithm for model fitting. Section 4 gives an updated algorithm for incomplete data, Section 5 discusses hypothesis testing issues, and simulation studies are in Section 6. We analyze data from the physical activity intervention trial in Section 7. Concluding remarks are given in Section 8. R programs are available from the first author.

2. Model setup

2.1 *The mixed effects model*

In this section, we establish notation for our data, specify the functional data model's fixed and random effects, and describe the covariance structure that the model induces. Let $Y_{ijk}^{(g)}(t)$ be the outcome observation at time t for subject i $(i = 1, \ldots, n)$ in week j $(j = 1, \ldots, J)$ on day k $(k = 1, \ldots, K)$ in treatment group g $(g = 1, \ldots, G)$. Our model is

$$Y_{ijk}^{(g)}(t) = \mu_{\cdot\cdot}^{(g)}(t) + \mu_{j\cdot}^{(g)}(t) + \mu_{\cdot k}^{(g)}(t) + \mu_{jk}^{(g)}(t) + \boldsymbol{\xi}_i(t) + \boldsymbol{\eta}_{ij}(t) + \boldsymbol{\zeta}_{ik}(t) + \boldsymbol{\gamma}_{ijk}(t) + \epsilon_{ijk}(t), \qquad (2.1)$$

where the superscript $^{(g)}$ denotes treatment group, $\mu_{\cdot\cdot}^{(g)}(t)$ are the population mean curves, $\mu_{j\cdot}^{(g)}(t)$, $\mu_{\cdot k}^{(g)}(t)$ and $\mu_{jk}^{(g)}(t)$ are week-specific, day-specific and week × day interaction mean curves, $\boldsymbol{\xi}_i(t)$, $\boldsymbol{\eta}_{ij}(t)$, $\boldsymbol{\zeta}_{ik}(t)$ and $\boldsymbol{\gamma}_{ijk}(t)$ are mutually independent subject-specific, week-within-subject, day-within-subject and week × day interaction-within-subject random effects curves, and $\epsilon_{ijk}(t)$ denotes random noise with mean zero and variance σ^2. Although there is a single curve for each subject-day-week combination, similarly to random effects linear regression with random slopes, the smooth reduced rank modeling strategy allows the model to include both this curve and a residual error. We do not vary the random effect distributions by group, but we implement a model to investigate that possibility in our application in Section 7. For ease of notation, we present the model for the case when each subject contributes data from J weeks and K days. Necessary modifications to the model and algorithm to accommodate the situation when this is not the case are in Section 4. Motivated by our application, we consider the case where each subject is only in one treatment group. For identifiability, we constrain $\mu_{1\cdot}^{(g)}(t) = \mu_{\cdot 1}^{(g)}(t) = \mu_{1k}^{(g)}(t) = \mu_{j1}^{(g)}(t) = 0$ for all g, k, j, t.

Each subject is only in one group, and observations from different subjects are modeled to be independent. Within subject, the covariance structure depends on whether observations are from the same week and/or day. The 4 covariance structures are:

1. same subject, week and day:

$$\begin{aligned}\operatorname{cov}\{Y_{ijk}^{(g)}(t), Y_{ijk}^{(g)}(s)\} = {} & \operatorname{cov}\{\boldsymbol{\xi}_i(t), \boldsymbol{\xi}_i(s)\} + \operatorname{cov}\{\boldsymbol{\eta}_{ij}(t), \boldsymbol{\eta}_{ij}(s)\} \\ & + \operatorname{cov}\{\boldsymbol{\zeta}_{ik}(t), \boldsymbol{\zeta}_{ik}(s)\} + \operatorname{cov}\{\boldsymbol{\gamma}_{ijk}(t), \boldsymbol{\gamma}_{ijk}(s)\},\end{aligned}$$

2. same subject, week, different day $(k \neq k')$:

$$\operatorname{cov}\{Y_{ijk}^{(g)}(t), Y_{ijk'}^{(g)}(s)\} = \operatorname{cov}\{\boldsymbol{\xi}_i(t), \boldsymbol{\xi}_i(s)\} + \operatorname{cov}\{\boldsymbol{\eta}_{ij}(t), \boldsymbol{\eta}_{ij}(s)\},$$

3. same subject, day, different week $(j \neq j')$:

$$\operatorname{cov}\{Y_{ijk}^{(g)}(t), Y_{ij'k}^{(g)}(s)\} = \operatorname{cov}\{\boldsymbol{\xi}_i(t), \boldsymbol{\xi}_i(s)\} + \operatorname{cov}\{\boldsymbol{\zeta}_{ik}(t), \boldsymbol{\zeta}_{ik}(s)\}$$

4. and, same subject only ($j \neq j', k \neq k'$):

$$\operatorname{cov}\{Y_{ijk}^{(g)}(t), Y_{ij'k'}^{(g)}(s)\} = \operatorname{cov}\{\boldsymbol{\xi}_i(t), \boldsymbol{\xi}_i(s)\}.$$

2.2 *Principal component form of the random effects*

We write the random effects in principal component form in this section. This allows us to consider the possibility that subject-to-subject variability is a linear combination of a limited number of principal component functions:

$$\boldsymbol{\xi}_i(t) = \sum_{\ell=1}^{P_\xi} f_{\xi,\ell}(t)\alpha_{\xi,i,\ell}; \quad \boldsymbol{\eta}_{ij}(t) = \sum_{\ell=1}^{P_\eta} f_{\eta,\ell}(t)\alpha_{\eta,ij,\ell};$$
$$\boldsymbol{\zeta}_{ik}(t) = \sum_{\ell=1}^{P_\zeta} f_{\zeta,\ell}(t)\alpha_{\zeta,ik,\ell}; \quad \boldsymbol{\gamma}_{ijk}(t) = \sum_{\ell=1}^{P_\gamma} f_{\gamma,\ell}(t)\alpha_{\gamma,ijk,\ell}.$$

The ℓth principal component functions for subject-specific, week-specific, day-specific and week $\times$ day interaction random effects curves are $f_{\xi,\ell}(t)$, $f_{\eta,\ell}(t)$, $f_{\zeta,\ell}(t)$, $f_{\gamma,\ell}(t)$, respectively, $\alpha_{\xi,i,\ell}$, $\alpha_{\eta,ij,\ell}$, $\alpha_{\zeta,ik,\ell}$, $\alpha_{\gamma,ijk,\ell}$ are the corresponding components' random loadings, and P_ξ, P_η, P_ζ, P_γ are the numbers of principal components for the 4 random effect variables. The principal components functions are orthogonal if $\ell = \ell^*$,

$$\int f_{\xi,\ell}(t) f_{\xi,\ell^*}(t)\,\mathrm{d}t = \int f_{\eta,\ell}(t) f_{\eta,\ell^*}(t)\,\mathrm{d}t = \int f_{\zeta,\ell}(t) f_{\zeta,\ell^*}(t)\,\mathrm{d}t = \int f_{\gamma,\ell}(t) f_{\gamma,\ell^*}(t)\,\mathrm{d}t = 1,$$

and the integrals are 0 otherwise. As is standard in an eigenvalue decomposition, we order indices so that $\operatorname{var}(\alpha_{\xi,i,1}) > \cdots > \operatorname{var}(\alpha_{\xi,i,P_\xi})$, $\operatorname{var}(\alpha_{\eta,ij,1}) > \cdots > \operatorname{var}(\alpha_{\eta,ij,P_\eta})$, $\operatorname{var}(\alpha_{\zeta,ik,1}) > \cdots > \operatorname{var}(\alpha_{\zeta,ik,P_\zeta})$ and $\operatorname{var}(\alpha_{\gamma,ijk,1}) > \cdots > \operatorname{var}(\alpha_{\gamma,ijk,P_\gamma})$ which makes the model identifiable.

In order to model the correlation of each random effect using only the principal component functions, we assume that the $\alpha_{\xi,i,\ell}$, $\alpha_{\eta,ij,\ell}$, $\alpha_{\zeta,ik,\ell}$ and $\alpha_{\gamma,ijk,\ell}$ are mutually independent random effects variables, and different instances of each from the 4 levels are assumed to be independent and identically distributed. The model is specified with notation in the next subsection where we also describe how we represent the functions.

2.3 *Modeling with B-splines*

We model the fixed effect and principal component functions using B-splines which we chose because our application requires a smooth function and their compact support can prevent computational instability. We define them as follows. Let $b(t) = \{b_1(t), \ldots, b_q(t)\}^{\mathrm{T}}$ be the $q \times 1$ vector of B-splines basis functions evaluated at t. Define $\beta^{(g)}$, $(\beta_{1\cdot}^{(g)\mathrm{T}}, \ldots, \beta_{J\cdot}^{(g)\mathrm{T}})^{\mathrm{T}}$, $(\beta_{\cdot 1}^{(g)\mathrm{T}}, \ldots, \beta_{\cdot K}^{(g)\mathrm{T}})^{\mathrm{T}}$ and $(\beta_{11}^{(g)\mathrm{T}}, \ldots, \beta_{JK}^{(g)\mathrm{T}})^{\mathrm{T}}$ to be group-specific fixed effects coefficients for population, week-specific, day-specific and week $\times$ day interaction mean structures, respectively. For identifiability, we set $\beta_{1\cdot}^{(g)} = \beta_{\cdot 1}^{(g)} = \beta_{j1}^{(g)} = \beta_{1k}^{(g)} = \mathbf{0}$ for all g, j, k. The non-zero fixed effects coefficients are grouped together in $\boldsymbol{\beta}$. We let $\mathbf{g}_{\xi,\ell}$, $\mathbf{g}_{\eta,\ell}$, $\mathbf{g}_{\zeta,\ell}$ and $\mathbf{g}_{\gamma,\ell}$ denote $q \times 1$ principal components vectors. Then we have

$$\mu_{\cdot\cdot}^{(g)}(t) = b^{\mathrm{T}}(t)\beta^{(g)}; \quad \mu_{j\cdot}^{(g)}(t) = b^{\mathrm{T}}(t)\beta_{j\cdot}^{(g)}; \quad \mu_{\cdot k}^{(g)}(t) = b^{\mathrm{T}}(t)\beta_{\cdot k}^{(g)}; \quad \mu_{jk}^{(g)}(t) = b^{\mathrm{T}}(t)\beta_{jk}^{(g)};$$
$$f_{\xi,\ell}(t) = b^{\mathrm{T}}(t)\mathbf{g}_{\xi,\ell}; \quad f_{\eta,\ell}(t) = b^{\mathrm{T}}(t)\mathbf{g}_{\eta,\ell}; \quad f_{\zeta,\ell}(t) = b^{\mathrm{T}}(t)\mathbf{g}_{\zeta,\ell}; \quad f_{\gamma,\ell}(t) = b^{\mathrm{T}}(t)\mathbf{g}_{\gamma,\ell}.$$

To maintain the orthogonality restrictions in Section 2, $b(t)$, $\mathbf{g}_{\xi,\ell}$, $\mathbf{g}_{\eta,\ell}$, $\mathbf{g}_{\zeta,\ell}$ and $\mathbf{g}_{\gamma,\ell}$ are restricted to be orthogonal. We use orthogonal B-spline basis functions from the R package "orthogonalsplinebasis".

Model (2.1) then becomes

$$\begin{aligned} Y_{ijk}^{(g)}(t) &= b^{\mathrm{T}}(t)\beta^{(g)} + b^{\mathrm{T}}(t)\beta_{j\cdot}^{(g)} + b^{\mathrm{T}}(t)\beta_{\cdot k}^{(g)} + b^{\mathrm{T}}(t)\beta_{jk}^{(g)} \\ &\quad + b^{\mathrm{T}}(t)\mathbf{G}_{\xi}\boldsymbol{\alpha}_{\xi,i} + b^{\mathrm{T}}(t)\mathbf{G}_{\eta}\boldsymbol{\alpha}_{\eta,ij} + b^{\mathrm{T}}(t)\mathbf{G}_{\zeta}\boldsymbol{\alpha}_{\zeta,ik} + b^{\mathrm{T}}(t)\mathbf{G}_{\gamma}\boldsymbol{\alpha}_{\gamma,ijk} + \epsilon_{ijk}(t), \end{aligned} \tag{2.2}$$

where $\mathbf{G}_{\xi} = (\mathbf{g}_{\xi,1}, \ldots, \mathbf{g}_{\xi,P_{\xi}})$, $\mathbf{G}_{\eta} = (\mathbf{g}_{\eta,1}, \ldots, \mathbf{g}_{\eta,P_{\eta}})$, $\mathbf{G}_{\zeta} = (\mathbf{g}_{\zeta,1}, \ldots, \mathbf{g}_{\zeta,P_{\zeta}})$ and $\mathbf{G}_{\gamma} = (\mathbf{g}_{\gamma,1}, \ldots, \mathbf{g}_{\gamma,P_{\gamma}})$ are $q \times P_{\xi}$, $q \times P_{\eta}$, $q \times P_{\zeta}$ and $q \times P_{\gamma}$ matrices, respectively, and $\boldsymbol{\alpha}_{\xi,i} = (\alpha_{\xi,i,1}, \ldots, \alpha_{\xi,i,P_{\xi}})^{\mathrm{T}}$, $\boldsymbol{\alpha}_{\eta,ij} = (\alpha_{\eta,ij,1}, \ldots, \alpha_{\eta,ij,P_{\eta}})^{\mathrm{T}}$, $\boldsymbol{\alpha}_{\zeta,ik} = (\alpha_{\zeta,ik,1}, \ldots, \alpha_{\zeta,ik,P_{\zeta}})^{\mathrm{T}}$ and $\boldsymbol{\alpha}_{\gamma,ijk} = (\alpha_{\gamma,ijk,1}, \ldots, \alpha_{\gamma,ijk,P_{\gamma}})^{\mathrm{T}}$ are $P_{\xi} \times 1$, $P_{\eta} \times 1$, $P_{\zeta} \times 1$ and $P_{\gamma} \times 1$ vectors. The orthogonality restrictions mean that $\mathbf{G}_{\xi}^{\mathrm{T}}\mathbf{G}_{\xi}$, $\mathbf{G}_{\eta}^{\mathrm{T}}\mathbf{G}_{\eta}$, $\mathbf{G}_{\zeta}^{\mathrm{T}}\mathbf{G}_{\zeta}$ and $\mathbf{G}_{\gamma}^{\mathrm{T}}\mathbf{G}_{\gamma}$ are identity matrices.

Next, we develop notation for the variance components of the component loadings. Let $\Delta_{\xi} = \mathrm{cov}(\boldsymbol{\alpha}_{\xi,i})$, $\Delta_{\eta} = \mathrm{cov}(\boldsymbol{\alpha}_{\eta,ij})$, $\Delta_{\zeta} = \mathrm{cov}(\boldsymbol{\alpha}_{\zeta,ik})$ and $\Delta_{\gamma} = \mathrm{cov}(\boldsymbol{\alpha}_{\gamma,ijk})$. With the restrictions and independence assumptions described in Section 2, Δ_{ξ}, Δ_{η}, Δ_{ζ} and Δ_{γ} are $P_{\xi} \times P_{\xi}$, $P_{\eta} \times P_{\eta}$, $P_{\zeta} \times P_{\zeta}$ and $P_{\gamma} \times P_{\gamma}$ diagonal matrices with decreasing positive diagonal elements.

To obtain the fixed and random effects curves, we need estimates of $\boldsymbol{\beta}$, $\mathbf{G}_{\xi}$, $\mathbf{G}_{\eta}$, $\mathbf{G}_{\zeta}$ and $\mathbf{G}_{\gamma}$. Estimates of σ^2, Δ_{ξ}, Δ_{η}, Δ_{ζ} and Δ_{γ} are also needed in order to estimate the random effects covariance structure and to make inferences.

2.4 *Link to the linear mixed model*

In this section, we demonstrate that our model is in fact related to the familiar linear mixed model. If we build 4 sets of q-dimensional random variables,

$$\mathbf{u}_{\xi,i} = \mathbf{G}_{\xi}\boldsymbol{\alpha}_{\xi,i}; \quad \mathbf{u}_{\eta,ij} = \mathbf{G}_{\eta}\boldsymbol{\alpha}_{\eta,ij}; \quad \mathbf{u}_{\zeta,ik} = \mathbf{G}_{\zeta}\boldsymbol{\alpha}_{\zeta,ik}; \quad \mathbf{u}_{\gamma,ijk} = \mathbf{G}_{\gamma}\boldsymbol{\alpha}_{\gamma,ijk},$$

then model (2.2) can be rewritten as

$$\begin{aligned} Y_{ijk}^{(g)}(t) &= b^{\mathrm{T}}(t)\beta^{(g)} + b^{\mathrm{T}}(t)\beta_{j\cdot}^{(g)} + b^{\mathrm{T}}(t)\beta_{\cdot k}^{(g)} + b^{\mathrm{T}}(t)\beta_{jk}^{(g)} \\ &\quad + b^{\mathrm{T}}(t)\mathbf{u}_{\xi,i} + b^{\mathrm{T}}(t)\mathbf{u}_{\eta,ij} + b^{\mathrm{T}}(t)\mathbf{u}_{\zeta,ik} + b^{\mathrm{T}}(t)\mathbf{u}_{\gamma,ijk} + \epsilon_{ijk}(t). \end{aligned} \tag{2.3}$$

The variance components are $\mathrm{cov}(\mathbf{u}_{\xi,i}) = \psi_{\xi} = \mathbf{G}_{\xi}\Delta_{\xi}\mathbf{G}_{\xi}^{\mathrm{T}}$, $\mathrm{cov}(\mathbf{u}_{\eta,ij}) = \psi_{\eta} = \mathbf{G}_{\eta}\Delta_{\eta}\mathbf{G}_{\eta}^{\mathrm{T}}$, $\mathrm{cov}(\mathbf{u}_{\zeta,ik}) = \psi_{\zeta} = \mathbf{G}_{\zeta}\Delta_{\zeta}\mathbf{G}_{\zeta}^{\mathrm{T}}$ and $\mathrm{cov}(\mathbf{u}_{\gamma,ijk}) = \psi_{\gamma} = \mathbf{G}_{\gamma}\Delta_{\gamma}\mathbf{G}_{\gamma}^{\mathrm{T}}$. If $P_{\xi} = P_{\eta} = P_{\zeta} = P_{\gamma} = q$, then ψ_{ξ}, ψ_{η}, ψ_{ζ} and ψ_{γ} would be full rank, and (2.3) is equivalent to a 3-level linear mixed model with unstructured random effect covariance matrices (Laird and Ware, 1982). Exploiting that connection, our estimation algorithm for (2.2) iteratively combines 2 steps which we describe in the next section. The algorithm takes q as an initial value for P_{ξ}, P_{η}, P_{ζ} and P_{γ}. The latter are then updated during the iterations.

3. Algorithm

The previous section demonstrates that our model is a linear mixed model with reduced rank variance component matrices for the random effects. As a result, our estimation method iterates 2 steps. The first step uses the ECME algorithm (Schafer, 1998) to obtain updates of $(\boldsymbol{\beta}, \sigma^2)$ and full rank versions of ψ_{ξ}, ψ_{η}, ψ_{ζ} and ψ_{γ}. The second step reduces the rank of ψ_{ξ}, ψ_{η}, ψ_{ζ} and ψ_{γ} to obtain $\mathbf{G}_{\xi}$, $\mathbf{G}_{\eta}$, $\mathbf{G}_{\zeta}$, $\mathbf{G}_{\gamma}$, Δ_{ξ}, Δ_{η}, Δ_{ζ} and Δ_{γ}. These steps are described in more detail in the next 2 subsections, and the steps

are repeated until convergence. Liu and Rubin (1994) discuss convergence properties of the algorithm in general.

We note that in this section, we present results rather than derive them since the steps are based on existing methods. That said, care must be taken to avoid bottlenecks caused by large matrix inversions, and methods to address these non-trivial problems are in Section S.2 of Supplementary Material available at *Biostatistics* online.

3.1 *ECME update step*

Using notation described in Section S.1 of Supplementary Material available at *Biostatistics* online, model (2.3) can be expressed as

$$\mathbf{Y}_i^{(g)} = \mathbf{B}_i^{(g)\mu}\boldsymbol{\beta} + \mathbf{B}_i^{U}\mathbf{U}_i + \boldsymbol{\epsilon}_i. \tag{3.1}$$

We further define $\Psi = \text{cov}(\mathbf{U}_i)$, $\mathbf{V}_i(\sigma^2, \Psi) = \mathbf{V}_i = \text{cov}(\mathbf{Y}_i^{(g)})/\sigma^2 = \mathbf{I}_{N_i \times N_i} + \mathbf{B}_i^{U}\Psi\mathbf{B}_i^{U^{\mathrm{T}}}/\sigma^2$ and $\mathbf{S}_i(\sigma^2, \Psi) = \mathbf{S}_i = \text{cov}(\mathbf{U}_i|\mathbf{Y}_i^{(g)})/\sigma^2 = \Psi/\sigma^2 - \Psi\mathbf{B}_i^{U^{\mathrm{T}}}\mathbf{V}_i^{-1}\mathbf{B}_i^{U}\Psi/\sigma^4$. Let the current parameters be $(\boldsymbol{\beta}_{\text{curr}}, \sigma^2_{\text{curr}}, \Psi^{\text{RR}}_{\text{curr}})$, leading to $(\mathbf{S}_{i,\text{curr}}, \mathbf{V}_{i,\text{curr}})$.

Using that notation, parameter updates are

$$\boldsymbol{\beta}_{\text{new}} = \left(\sum_{i=1}^{n} \mathbf{B}_i^{(g)\mu^{\mathrm{T}}}\mathbf{V}_{i,\text{curr}}^{-1}\mathbf{B}_i^{(g)\mu}\right)^{-1}\left(\sum_{i=1}^{n} \mathbf{B}_i^{(g)\mu^{\mathrm{T}}}\mathbf{V}_{i,\text{curr}}^{-1}\mathbf{Y}_i^{(g)}\right),$$

$$\sigma^2_{\text{new}} = N^{-1}\sum_{i=1}^{n}(\mathbf{Y}_i^{(g)} - \mathbf{B}_i^{(g)\mu}\boldsymbol{\beta}_{\text{new}})^{\mathrm{T}}\mathbf{V}_{i,\text{curr}}^{-1}(\mathbf{Y}_i^{(g)} - \mathbf{B}_i^{(g)\mu}\boldsymbol{\beta}_{\text{new}}),$$

and

$$\Psi^{\text{UN}}_{\text{new}} = n^{-1}\sum_{i=1}^{n}(\hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^{\mathrm{T}} + \sigma^2_{\text{new}}\mathbf{S}_{i,\text{curr}}), \quad \text{where } \hat{\mathbf{U}}_i = \mathbf{S}_{i,\text{curr}}\mathbf{B}_i^{U^{\mathrm{T}}}(\mathbf{Y}_i^{(g)} - \mathbf{B}_i^{(g)\mu}\boldsymbol{\beta}_{\text{new}}).$$

The first two updates come from maximizing the likelihood with other parameters held fixed, and the third update is an EM step.

The variance component matrices $\psi_{\xi,\text{curr}}$, $\psi_{\eta,\text{curr}}$, $\psi_{\zeta,\text{curr}}$ and $\psi_{\gamma,\text{curr}}$ can be extracted from the block diagonal elements of $\Psi^{\text{UN}}_{\text{new}}$ with indices that correspond to the subvectors in $\mathbf{U}_i$ as described in Section S.1 of Supplementary Material available at *Biostatistics* online. We describe the rank reduction step which leads to $\Psi^{\text{RR}}_{\text{new}}$ in the next subsection. After that step, we update $\boldsymbol{\beta}_{\text{curr}} = \boldsymbol{\beta}_{\text{new}}$, $\sigma^2_{\text{curr}} = \sigma^2_{\text{new}}$ and $\Psi^{\text{RR}}_{\text{curr}} = \Psi^{\text{RR}}_{\text{new}}$.

3.2 *Reduced rank model implementation*

We obtain $\Psi^{\text{RR}}_{\text{new}}$, eigenvalue decompositions of $(\psi_{\xi,\text{curr}}, \psi_{\eta,\text{curr}}, \psi_{\zeta,\text{curr}}, \psi_{\gamma,\text{curr}})$. These are

$$\psi_{\xi,\text{curr}} = \tilde{\mathbf{G}}_{\xi,\text{new}}\tilde{\Delta}_{\xi,\text{new}}\tilde{\mathbf{G}}^{\mathrm{T}}_{\xi,\text{new}}, \quad \psi_{\eta,\text{curr}} = \tilde{\mathbf{G}}_{\eta,\text{new}}\tilde{\Delta}_{\eta,\text{new}}\tilde{\mathbf{G}}^{\mathrm{T}}_{\eta,\text{new}},$$

$$\psi_{\zeta,\text{curr}} = \tilde{\mathbf{G}}_{\zeta,\text{new}}\tilde{\Delta}_{\zeta,\text{new}}\tilde{\mathbf{G}}^{\mathrm{T}}_{\zeta,\text{new}}, \quad \psi_{\gamma,\text{curr}} = \tilde{\mathbf{G}}_{\gamma,\text{new}}\tilde{\Delta}_{\gamma,\text{new}}\tilde{\mathbf{G}}^{\mathrm{T}}_{\gamma,\text{new}}.$$

The tildes denote that these are full rank decompositions, and we describe the properties of the terms in these decompositions in Section 2.1.

Next, to select the number of principal components (P_ξ, P_η, P_ζ, P_γ), we set a threshold P and select by defining

$$P_{\xi,\text{new}} = \min\left\{q^* : \frac{\text{var}(\alpha_{\xi,i,1,\text{new}}) + \cdots + \text{var}(\alpha_{\xi,i,q^*,\text{new}})}{\text{var}(\alpha_{\xi,i,1,\text{new}}) + \cdots + \text{var}(\alpha_{\xi,i,q,\text{new}})} \geqslant P\right\},$$
$$P_{\eta,\text{new}} = \min\left\{q^* : \frac{\text{var}(\alpha_{\eta,ij,1,\text{new}}) + \cdots + \text{var}(\alpha_{\eta,ij,q^*,\text{new}})}{\text{var}(\alpha_{\eta,ij,1,\text{new}}) + \cdots + \text{var}(\alpha_{\eta,ij,q,\text{new}})} \geqslant P\right\},$$
$$P_{\zeta,\text{new}} = \min\left\{q^* : \frac{\text{var}(\alpha_{\zeta,ik,1,\text{new}}) + \cdots + \text{var}(\alpha_{\zeta,ik,q^*,\text{new}})}{\text{var}(\alpha_{\zeta,ik,1,\text{new}}) + \cdots + \text{var}(\alpha_{\zeta,ik,q,\text{new}})} \geqslant P\right\},$$

and

$$P_{\gamma,\text{new}} = \min\left\{q^* : \frac{\text{var}(\alpha_{\gamma,ijk,1,\text{new}}) + \cdots + \text{var}(\alpha_{\gamma,ijk,q^*,\text{new}})}{\text{var}(\alpha_{\gamma,ijk,1,\text{new}}) + \cdots + \text{var}(\alpha_{\gamma,ijk,q,\text{new}})} \geqslant P\right\}. \tag{3.2}$$

For each of the 4 random effect groups (ξ, η, ζ and γ) we then retain that many of each of the columns for the corresponding eigenvectors (**G**s) and the rows and columns for random loadings (Δs) to define $\mathbf{G}_{\xi,\text{new}}$, $\mathbf{G}_{\eta,\text{new}}$, $\mathbf{G}_{\zeta,\text{new}}$ and $\mathbf{G}_{\gamma,\text{new}}$ and $\Delta_{\xi,\text{new}}$, $\Delta_{\eta,\text{new}}$, $\Delta_{\zeta,\text{new}}$ and $\Delta_{\gamma,\text{new}}$.

Finally, we obtain the updated reduced rank matrices

$$\psi^{\text{RR}}_{\xi,\text{new}} = \mathbf{G}_{\xi,\text{new}}\Delta_{\xi,\text{new}}\mathbf{G}^{\text{T}}_{\xi,\text{new}}, \quad \psi^{\text{RR}}_{\eta,\text{new}} = \mathbf{G}_{\eta,\text{new}}\Delta_{\eta,\text{new}}\mathbf{G}^{\text{T}}_{\eta,\text{new}},$$
$$\psi^{\text{RR}}_{\zeta,\text{new}} = \mathbf{G}_{\zeta,\text{new}}\Delta_{\zeta,\text{new}}\mathbf{G}^{\text{T}}_{\zeta,\text{new}}, \quad \psi^{\text{RR}}_{\gamma,\text{new}} = \mathbf{G}_{\gamma,\text{new}}\Delta_{\gamma,\text{new}}\mathbf{G}^{\text{T}}_{\gamma,\text{new}},$$

and these are combined into $\Psi^{\text{RR}}_{\text{new}}$ as described in Section S.1 of Supplementary Material available at *Biostatistics* online.

To choose P, a subjective choice is usually satisfactory and is often used. We use $P = 0.85$ in our simulation studies which works very well in our experience. Alternatively, cross-validation or model selection methods such as BIC could be used.

3.3 *Maximum penalized likelihood*

The previous discussion focuses on the modeling of the response variables using basis functions, but it is important to introduce roughness penalties to regularize the function fits.

We use penalized maximum likelihood for parameter estimation with maximization of

$$\mathcal{L}_{pen} = \mathcal{L} - \tau_\mu \boldsymbol{\beta}^{\text{T}} D_\mu \boldsymbol{\beta} - \tau_{\mathbf{g}} \left(\sum_{\ell=1}^{P_\xi} \mathbf{g}^{\text{T}}_{\xi,\ell} D \mathbf{g}_{\xi,\ell} + \sum_{\ell=1}^{P_\eta} \mathbf{g}^{\text{T}}_{\eta,\ell} D \mathbf{g}_{\eta,\ell} + \sum_{\ell=1}^{P_\zeta} \mathbf{g}^{\text{T}}_{\zeta,\ell} D \mathbf{g}_{\zeta,\ell} + \sum_{\ell=1}^{P_\gamma} \mathbf{g}^{\text{T}}_{\gamma,\ell} D \mathbf{g}_{\gamma,\ell} \right), \tag{3.3}$$

where $\mathcal{L} = (-N/2)\log(\sigma^2) - (1/2)\sum_{i=1}^n \log(|\mathbf{V}_i|) - (2\sigma^2)^{-1}\sum_{i=1}^n (\mathbf{Y}_i^{(g)} - \mathbf{B}_i^{(g)\mu}\boldsymbol{\beta})^{\text{T}}\mathbf{V}_i^{-1}(\mathbf{Y}_i^{(g)} - \mathbf{B}_i^{(g)\mu}\boldsymbol{\beta})$, the penalty parameters are τ_μ and $\tau_{\mathbf{g}}$, $D = \int b''(t)b''(t)^{\text{T}}\,\text{d}t$, and $D_\mu = \mathbf{I}_{GJKq\times GJKq} \otimes D$.

Using maximum penalized likelihood has minor effect on the equations that are used to update $\boldsymbol{\beta}$ and Ψ. The technical details are in Section S.3 of Supplementary Material available at *Biostatistics* online. A search over a candidate set of parameters is a feasible way to choose the penalty parameters τ_μ and $\tau_{\mathbf{g}}$. We used that approach and 5-fold cross-validation in our numerical work. It would be possible to model the penalty parameters as variance components and estimate them by REML as well.

4. ALGORITHM FOR INCOMPLETE DATA

The model and algorithm discussed in Sections 2 and 3 assume that all subjects contribute the same numbers of weeks and days within each week. Although the observed data likelihood is proper for the missing data, the estimation algorithm and notation need to be modified to accommodate the data incompleteness. Define J_i^{obs}, K_i^{obs}, M_i^{obs} and N_i^{obs} to be number of observed weeks, days, week-days and observed records for subject i, respectively. We write $ij \in O$, $ik \in O$ and $ijk \in O$ to represent that week j, day k and week-day (j, k) are observed for subject i, respectively. Let $\mathbf{B}_i^{(g)\mu,\text{obs}}$ and $\mathbf{B}_i^{U,\text{obs}}$ to be the $\mathbf{B}_i^{(g)\mu}$ and $\mathbf{B}_i^{U}$ matrices with unobserved block rows removed, respectively. Define $\mathbf{U}_i^{\text{obs}} = (\mathbf{u}_{\xi,i}^{\text{obs}^{\mathrm{T}}}, \mathbf{U}_{\eta,i}^{\text{obs}^{\mathrm{T}}}, \mathbf{U}_{\zeta,i}^{\text{obs}^{\mathrm{T}}}, \mathbf{U}_{\gamma,i}^{\text{obs}^{\mathrm{T}}})^{\mathrm{T}}$ to be the $\{(1 + J_i^{\text{obs}} + K_i^{\text{obs}} + M_i^{\text{obs}}) \times q\} \times 1$ vector corresponding to subject-specific, week-specific, day-specific and week $\times$ day interaction random effects for the observed data. Then define $\Psi_i^{\text{obs}} = \text{diag}\,(\psi_\xi, \Psi_{\eta,i}^{\text{obs}}, \Psi_{\zeta,i}^{\text{obs}}, \Psi_{\gamma,i}^{\text{obs}})$, where $\Psi_{\eta,i}^{\text{obs}} = \mathbf{I}_{J_i^{\text{obs}} \times J_i^{\text{obs}}} \otimes \psi_\eta$, $\Psi_\zeta^{\text{obs}} = \mathbf{I}_{K_i^{\text{obs}} \times K_i^{\text{obs}}} \otimes \psi_\zeta$, and $\Psi_\gamma^{\text{obs}} = \mathbf{I}_{M_i^{\text{obs}} \times M_i^{\text{obs}}} \otimes \psi_\gamma$. Because of different missing patterns for each subject, $\mathbf{U}_i^{\text{obs}}$ and Ψ_i^{obs} may not have identical dimensions across i.

We define $\mathbf{V}_i^{-1,\text{obs}}(\sigma^2, \Psi_i^{\text{obs}}) = \mathbf{V}_i^{-1,\text{obs}} = (\mathbf{I}_{N_i^{\text{obs}} \times N_i^{\text{obs}}} + \mathbf{B}_i^{U,\text{obs}} \Psi_i^{\text{obs}} \mathbf{B}_i^{U,\text{obs}^{\mathrm{T}}} / \sigma^2)^{-1}$, and $\mathbf{S}_i^{\text{obs}}(\sigma^2, \Psi_i^{\text{obs}}) = \mathbf{S}_i^{\text{obs}} = \Psi_i^{\text{obs}}/\sigma^2 - \Psi_i^{\text{obs}} \mathbf{B}_i^{U,\text{obs}^{\mathrm{T}}} \mathbf{V}_i^{-1,\text{obs}} \mathbf{B}_i^{U} \Psi_i^{\text{obs}} / \sigma^4$. The joint log-likelihood function for the available data becomes

$$\mathcal{L}^{\text{obs}} = -(N^{\text{obs}}/2)\log(\sigma^2) - (1/2)\sum_{i=1}^{n} \log(|\mathbf{V}_i^{\text{obs}}|)$$
$$- (1/2\sigma^2)\sum_{i=1}^{n} (\mathbf{Y}_i^{(g)\text{obs}} - \mathbf{B}_i^{(g)\mu,\text{obs}}\boldsymbol{\beta})^{\mathrm{T}} \mathbf{V}_i^{-1,\text{obs}} (\mathbf{Y}_i^{(g)\text{obs}} - \mathbf{B}_i^{(g)\mu,\text{obs}}\boldsymbol{\beta}),$$

where $N^{\text{obs}} = \sum_{i=1}^{n} N_i^{\text{obs}}$.

Updates for $\boldsymbol{\beta}$ and σ^2 follow the complete data strategy. On the other hand, since subject-specific $\mathbf{U}_i^{\text{obs}}$ and Ψ_i^{obs} may have different dimensions, additional modification of the algorithm for updating Ψ is necessary. For each subject i, we can obtain unstructured $\Psi_i^{\text{UN,obs}} = \hat{\mathbf{U}}_i^{\text{obs}} \hat{\mathbf{U}}_i^{\text{obs}^{\mathrm{T}}} + \sigma^2 \mathbf{S}_i^{\text{obs}}$, where $\hat{\mathbf{U}}_i^{\text{obs}} = \mathbf{S}_i^{\text{obs}} \mathbf{B}_i^{U\text{obs}^{\mathrm{T}}} (\mathbf{Y}_i^{(g)\text{obs}} - \mathbf{B}_i^{(g)\mu,\text{obs}}\boldsymbol{\beta})$.

Similar to the algorithm used for complete data, we extract the diagonal blocks from $\Psi_i^{\text{UN,obs}}$. For the complete data case, $\widehat{\text{cov}}(\mathbf{u}_{\xi,i})$, $\widehat{\text{cov}}(\mathbf{u}_{\eta,ij})$, $\widehat{\text{cov}}(\mathbf{u}_{\zeta,ik})$ and $\widehat{\text{cov}}(\mathbf{u}_{\gamma,ijk})$ can be extracted directly from Ψ^{UN} because each i contributes 1, J, K and JK components, respectively. However, the incomplete data $\Psi_i^{\text{UN,obs}}$ contributes 1, J_i, K_i and M_i components for $\widehat{\text{cov}}^*(\mathbf{u}_{\xi,i})$, $\widehat{\text{cov}}^*(\mathbf{u}_{\eta,ij})$, $\widehat{\text{cov}}^*(\mathbf{u}_{\zeta,ik})$ and $\widehat{\text{cov}}^*(\mathbf{u}_{\gamma,ijk})$, respectively, which may be different among subjects. Thus, we calculate ψ_ξ, ψ_η, ψ_ζ and ψ_γ by considering the different number of complete cases for each subject:

$$\psi_\xi = n^{-1}\sum_{i=1}^{n} \widehat{\text{cov}}^*(\mathbf{u}_{\xi,i}), \quad \psi_\eta = n^{-1}\sum_{i=1}^{n}\sum_{j,ij\in O} \widehat{\text{cov}}^*(\mathbf{u}_{\eta,ij})/J_i,$$
$$\psi_\zeta = n^{-1}\sum_{i=1}^{n}\sum_{k,ik\in O} \widehat{\text{cov}}^*(\mathbf{u}_{\zeta,ik})/K_i, \quad \psi_\gamma = n^{-1}\sum_{i=1}^{n}\sum_{jk,ijk\in O} \widehat{\text{cov}}^*(\mathbf{u}_{\gamma,ijk})/M_i,$$

which leads to the updated values of Ψ. The estimators of $\mathbf{G}_\xi$, $\mathbf{G}_\eta$, $\mathbf{G}_\zeta$, $\mathbf{G}_\gamma$, Δ_ξ, Δ_η, Δ_ζ and Δ_γ are identical to the method in Section 3.

5. Hypothesis tests for mean curves

In this section, we discuss testing for the fixed effects mean structures, while keeping the random effects structures as previously specified. Model (2.1) involves mean curves for treatments and week-specific, day-specific and week × day interaction effects for each treatment. We next discuss testing whether these effects equal zero. Using the parameterization in model (2.3), inference is equivalent to testing whether a subvector of $\boldsymbol{\beta}$ equals zero, and we discuss the Wald test approach with null hypothesis $\mathbf{L}\boldsymbol{\beta} = \mathbf{0}$, where $\mathbf{L}$ is a known matrix constructed for a specific hypotheses test. Thus, the Wald statistics can be written as $(\mathbf{L}\hat{\boldsymbol{\beta}})^{\mathrm{T}}\{\mathbf{L}\,\widehat{\mathrm{cov}}(\hat{\boldsymbol{\beta}})\,\mathbf{L}^{\mathrm{T}}\}^{-1}\mathbf{L}\hat{\boldsymbol{\beta}}$, where $\widehat{\mathrm{cov}}(\hat{\boldsymbol{\beta}})$ is an estimate of the covariance matrix of $\hat{\boldsymbol{\beta}}$.

The hypothesis test is based on the model discussed in Section 2 and the estimation results obtained from the algorithm discussed in Section 3. When $\hat{\boldsymbol{\beta}}$ is obtained from the unpenalized likelihood function, $\widehat{\mathrm{cov}}(\hat{\boldsymbol{\beta}})$ can be calculated as $\hat{\sigma}^2(\sum_{i=1}^n \mathbf{B}_i^{(g)\mu^{\mathrm{T}}}\hat{\mathbf{V}}_i^{-1}\mathbf{B}_i^{(g)\mu})^{-1}$ with $\hat{\sigma}^2$ and $\hat{\mathbf{V}}_i$ obtained from the algorithm. The Wald statistic then follows asymptotically a χ^2 distribution with degrees of freedom equal to the rank of $\mathbf{L}$. In the simulation reported below in Section 6, this test has acceptable level for a small to moderate number of basis functions, and for both complete and missing data. When the model involves a large number of knots, the test levels deteriorate, especially with incomplete data and a small sample size, but this problem does not occur if the sample size is large.

An alternative is to use the penalized likelihood. For $\hat{\boldsymbol{\beta}}$ obtained from penalized likelihood (3.3), $\widehat{\mathrm{cov}}(\hat{\boldsymbol{\beta}})$ can be estimated by

$$\hat{\sigma}^2\left(\sum_{i=1}^n \mathbf{B}_i^{(g)\mu^{\mathrm{T}}}\hat{\mathbf{V}}_i^{-1}\mathbf{B}_i^{(g)\mu} + 2\tau_\mu D_\mu\right)^{-1}\left(\sum_{i=1}^n \mathbf{B}_i^{(g)\mu^{\mathrm{T}}}\hat{\mathbf{V}}_i^{-1}\mathbf{B}_i^{(g)\mu}\right)\left(\sum_{i=1}^n \mathbf{B}_i^{(g)\mu^{\mathrm{T}}}\hat{\mathbf{V}}_i^{-1}\mathbf{B}_i^{(g)\mu} + 2\tau_\mu D_\mu\right)^{-1}.$$

However, the resulting Wald test performs poorly in terms of test level, leading to too many false positives. An alternative is to use the so-called Bayesian variance estimator $\hat{\sigma}^2(\sum_{i=1}^n \mathbf{B}_i^{(g)\mu^{\mathrm{T}}}\hat{\mathbf{V}}_i^{-1}\mathbf{B}_i^{(g)\mu} + 2\tau_\mu D_\mu)^{-1}$ as the asymptotic covariance matrix, see Ruppert *and others* (2003), Equation (6.13). However, we have found that this leads to much too conservative a test and with little power. Ruppert *and others* (2003) suggest using 10 000 to 100 000 Monte Carlo simulations to calculate the p-values. However, the Monte Carlo approach is very computationally expensive for our 3-level model.

6. Simulation studies

In this section, we illustrate the performance of our methodology. In each simulation run, we have $n = 60$ subjects for $J = 5$ weeks with $K = 5$ days, and each day has 36 measurement times. The probability that a day's records is observed is 50%. Here, we report the setting with $G = 1$, but we also implement the setting with $G = 2$ and the results are very similar. A measurement at time t on day k in week j for subject i results in observation $Y_{ijk}^{(1)}(t)$ which is generated according to

$$\begin{aligned} Y_{ijk}^{(1)}(t) = \mu_{\cdot\cdot}^{(1)}(t) + \mu_{j\cdot}^{(1)}(t) + \mu_{\cdot k}^{(1)}(t) + \sum_{\ell=1}^{2} f_{\xi,\ell}(t)\alpha_{\xi,i,\ell} + \sum_{\ell=1}^{2} f_{\eta,\ell}(t)\alpha_{\eta,ij,\ell} \\ + \sum_{\ell=1}^{2} f_{\zeta,\ell}(t)\alpha_{ik,\ell} + \sum_{\ell=1}^{2} f_{\gamma,\ell}(t)\alpha_{\gamma,ijk,\ell} + \epsilon_{ijk}(t), \end{aligned}$$

with detailed settings of curves and parameters listed in Table 1.

Table 1. *Settings and parameter estimation results for the simulation study in Section 6. The first part of the table displays the true curves used in the model. The second part of the table displays the true value, the average estimates and the mean squared errors (MSE) of the parameters in the joint model.* σ^2, $\Delta_{\xi,1}$, $\Delta_{\xi,2}$, $\Delta_{\eta,1}$, $\Delta_{\eta,2}$, $\Delta_{\zeta,1}$, $\Delta_{\zeta,2}$, $\Delta_{\gamma,1}$, $\Delta_{\gamma,2}$ *are the variances of* $\epsilon_{ijk}(t)$, $\alpha_{\xi,i,1}$, $\alpha_{\xi,i,2}$, $\alpha_{\eta,ij,1}$, $\alpha_{\eta,ij,2}$, $\alpha_{\zeta,ik,1}$, $\alpha_{\zeta,ik,2}$, $\alpha_{\gamma,ijk,1}$ *and* $\alpha_{\gamma,ijk,2}$, *respectively. The number marked with an asterisk is the actual number multiplied by* 1000. *The lower part of the table displays the rate of true null hypothesis for 4 methods with* $\alpha = 0.05$. *PEN*10 *represents the penalized likelihood with* 10 *knots. UN*10, *UN*24 *and UN*30 *are the unpenalized likelihood methods with* 10, 24 *and* 30 *knots, respectively*

$\mu_{..}^{(1)}(t)$	$1 + t/500 + \exp\{-(t-18)^2/500\}$
$\mu_{j.}^{(1)}(t)$	$\exp\{t/(j-1) - 5.5\}/[1 + \exp\{t/(j-1) - 5.5\}] - 0.5$ (when $j > 1$)
$\mu_{.k}^{(1)}(t)$	$0.5 - \exp\{t/(k-1) - 5.5\}/[1 + \exp\{t/(k-1) - 5.5\}]$ (when $k > 1$)
$f_{\xi,1}(t)$	$\sin\{2\pi(t-1)/35\}/\sqrt{17.5}$
$f_{\xi,2}(t)$	$\cos\{2\pi(t-1)/35\}/\sqrt{17.5}$
$f_{\eta,1}(t)$	$\sin\{4\pi(t-1)/35\}/\sqrt{17.5}$
$f_{\eta,2}(t)$	$\cos\{4\pi(t-1)/35\}/\sqrt{17.5}$
$f_{\zeta,1}(t)$	$1/\sqrt{35}$
$f_{\zeta,2}(t)$	$\sqrt{3}\{2(t-1)/35 - 1\}/\sqrt{35}$
$f_{\gamma,1}(t)$	$[6\{(t-1)/35\}^2 - 6\{(t-1)/35\} + 1]/\sqrt{7}$
$f_{\gamma,2}(t)$	$1/\sqrt{35}$

Parameter	σ^2	$\Delta_{\xi,1}$	$\Delta_{\xi,2}$	$\Delta_{\eta,1}$	$\Delta_{\eta,2}$	$\Delta_{\zeta,1}$	$\Delta_{\zeta,2}$	$\Delta_{\gamma,1}$	$\Delta_{\gamma,2}$
True	1.00	8.00	4.00	6.00	3.00	4.00	2.00	2.00	1.00
Mean	1.00	8.33	3.86	6.00	2.86	3.89	2.00	1.95	0.93
MSE	0.08*	2.11	0.55	0.26	0.11	0.18	0.04	0.03	0.02

	Complete data $n = 60$	Missing data $n = 60$	Complete data $n = 250$	Missing data $n = 250$
PEN10	0.37	0.32	0.62	0.46
UN10	0.06	0.08	0.07	0.06
UN24	0.06	0.15	0.05	0.07
UN30	0.10	0.23	0.04	0.07

We use B-spline basis functions with 10 equispaced knots to fit the data. Figures 2 and 3 give results for the fixed effects and principal component curves, respectively. Parameter estimation results are shown in Table 1. The results suggest that our method has excellent performance.

We next study the performance of the Wald statistics discussed in Section 5 to test the null hypothesis that $\mu_{jk}^{(1)}(t) = \mathbf{0}$ for all j, k. The nominal rejection rate is set to 0.05. Four methods are studied: penalized likelihood with 10 knots (PEN10), unpenalized likelihood with 10 knots (UN10), unpenalized likelihood with 24 knots (UN24) and unpenalized likelihood with 30 knots (UN30). For each method, Table 1 displays the rejection rate of 500 replicates under complete and incomplete data scenarios with sample size $n = 60$, and the rejection rate of 200 replicates for sample size $n = 250$. With complete data, unpenalized likelihood works well in terms of test level with 10 and 24 knots, but has an inflated rejection rate if 30 knots are used. Under the missing data scenario, unpenalized likelihood with small sample size performs poorly in terms of rejection rate when more knots are added, but this problem does not occur with large sample size. On the other hand, testing based on penalized likelihood leads to unacceptable Type I errors in all scenarios.

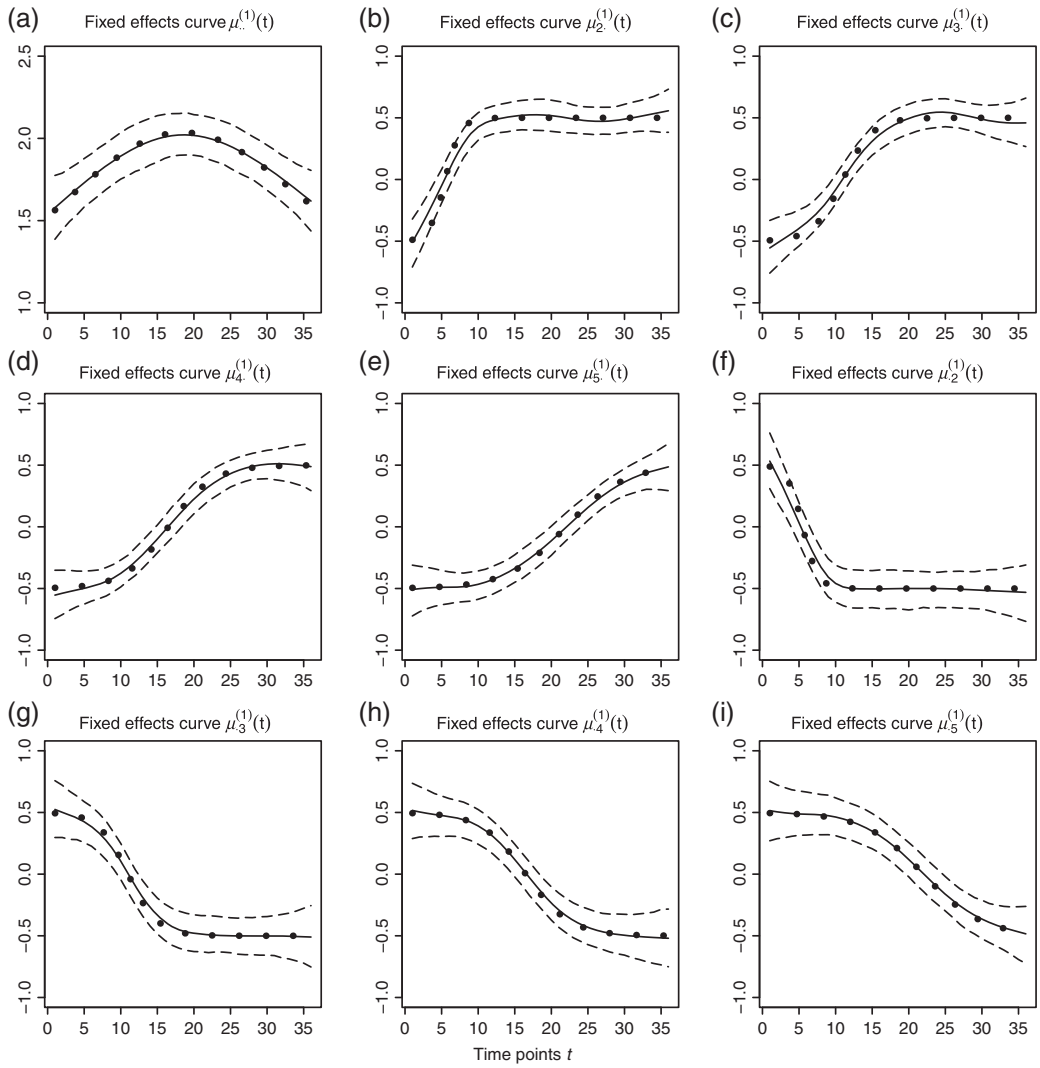

Fig. 2. Fitted fixed effects curves for 500 simulated data sets: (a) fixed effects curve $\mu_{..}^{(1)}(t)$, (b)–(e) fixed effects curves $\mu_{2.}^{(1)}(t)$, $\mu_{3.}^{(1)}(t)$, $\mu_{4.}^{(1)}(t)$, $\mu_{5.}^{(1)}(t)$, (f)–(i) fixed effects curves $\mu_{.2}^{(1)}(t)$, $\mu_{.3}^{(1)}(t)$, $\mu_{.4}^{(1)}(t)$, $\mu_{.5}^{(1)}(t)$. Dotted lines denote true curves. Solid lines represent the averaged values of fitted curves. The upper and lower dashed lines are the 10% and 90% quantiles of the fitted values in 500 simulation studies.

7. Application

7.1 *Background*

In this section, we apply our model to data from the physical activity intervention trial described in Section 1, which uses an activity monitor to estimate the physical activity levels of subjects over time. For this paper, the outcome of interest is energy expenditure, which is expressed in units of metabolic equivalents (METs). A person's MET value for an activity is defined as the ratio of energy expenditure during an activity to resting energy expenditure. The unit for both numerator and denominator of a MET

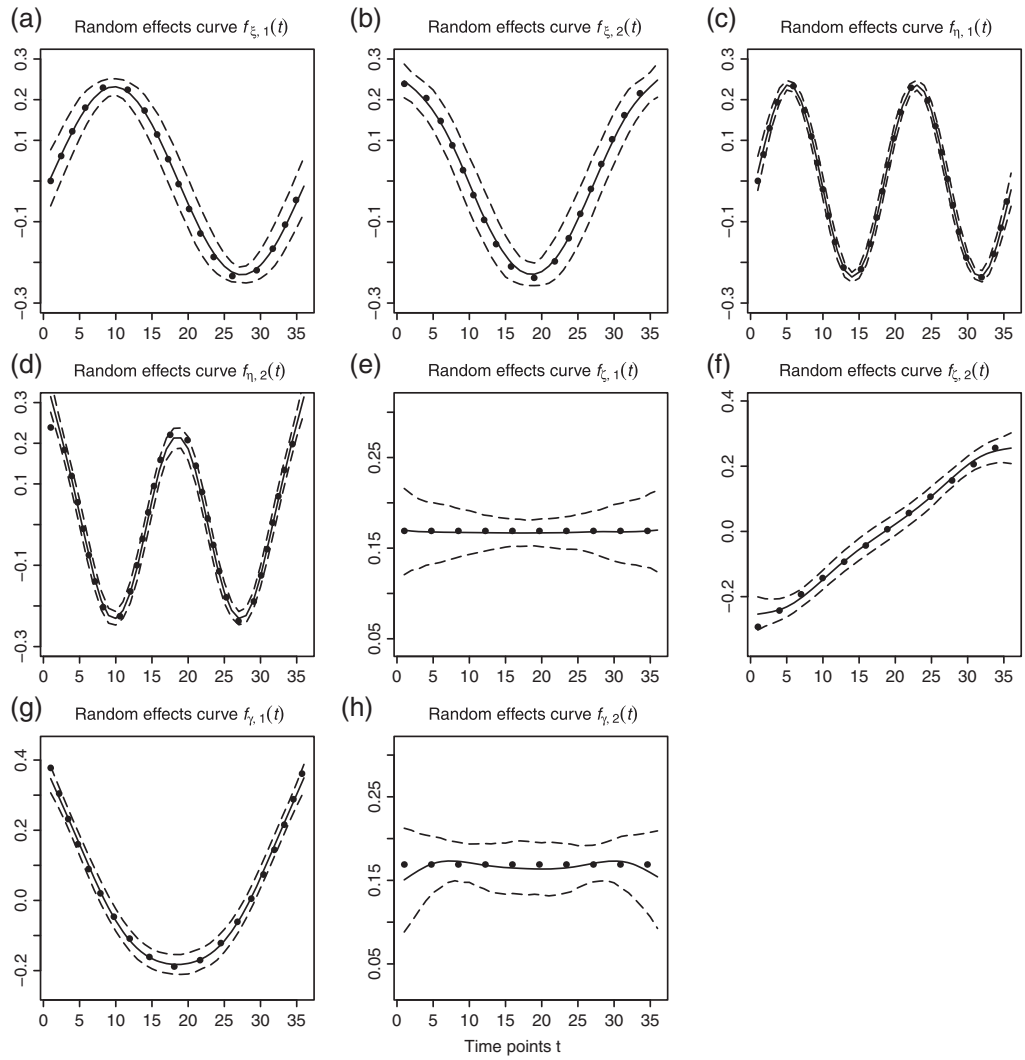

Fig. 3. Fitted principal components curves for 500 simulated data sets: (a) and (b) principal components curves $f_{\xi,1}(t)$ and $f_{\xi,2}(t)$, (c) and (d) principal components curves $f_{\eta,1}(t)$ and $f_{\eta,2}(t)$, (e) and (f) principal components curves $f_{\zeta,1}(t)$ and $f_{\zeta,2}(t)$, (g) and (h) principal components curves $f_{\gamma,1}(t)$ and $f_{\gamma,2}(t)$. Dotted lines denote true curves. Solid lines represent the averaged values of fitted curves. The upper and lower dashed lines are the 10% and 90% quantiles of the fitted values in 500 simulation studies.

is oxygen consumption per kilogram of body weight per minute, and the ratio makes a MET value independent of body weight or time. A MET value of at least 3 defines moderate to vigorous physical activity (MVPA). The gold-standard for measuring energy expenditure is doubly labeled water, but it is expensive and estimates only total daily energy expenditure; thus, the patterns and temporal distribution of intensity (of interest in our application) cannot be estimated. Another approach is to use indirect calorimetry where a person wears a mask and an electronic device that analyzes inhaled and exhaled gases to measure energy expenditure over time. This approach provides temporal detail, but it is impractical for large

scale monitoring outside of controlled settings since the mouth is covered and the equipment is bulky and expensive. As a result, accelerometer-based activity monitoring is widely used in the physical activity and health literature. Accelerometers are small, unobtrusive and provide time-stamped estimates of acceleration. Estimates of energy expenditure from those measurements are based on the principle that acceleration is proportional to net external forces, therefore reflective of the energy cost of movement. Freedson *and others* (2012) provide a recent review of this work, and introduce a special issue of a journal that is devoted to this topic.

The activPAL^{TM} (www.paltech.plus.com) accelerometer activity monitor was used in this study. This device were taped to the front of the thigh and used accelerometers to assess the angle of the thigh and movement. The angle differentiates sitting from standing, and the movement differentiates standing still from stepping. Those measurements are then used by an algorithm in the device to estimate energy expenditure. The algorithm is proprietary, but it appears to yield constant values for sitting and standing, and to be a linear function of a summary of total acceleration when stepping. The calibration is based primarily on locomotive activities, thus for some activities that involve significant upper body movement, energy expenditure may be underestimated. Additionally, it should be noted that acceleration tends not to be a smooth function of time, and the activPAL^{TM}'s estimates of energy expenditure (oxygen use) also tend not to be smooth. However, a person's actual use of oxygen and the resulting MET value are smooth functions of time for physiological reasons (Powers and Howley, 2001, Chapter 4).

7.2 *Analysis of daily METs*

In the study, 63 individuals wore the activPAL^{TM} for 5 weeks (denoted weeks 0, 3, 6, 9, 12), 5 days in a week (Monday to Friday) and measurements were recorded every 5 min during each day. After week zero each individual was assigned randomly to either the treatment group or the control group. Each member of the treatment group received a personal trainer who developed an exercise program for them and supervised 40 min exercise sessions 5 days a week. Members of the control group were instructed to continue their lives as before.

We analyze these data with model (2.1) with $g = 1, 2$ denoting the control and treatment groups, respectively. We also fit models that allow the distributions of the random effects to vary by group, but the results do not change appreciably and are not shown. The model's response is METs over time, and, as discussed in Section 7.1, the activPal's estimate of METs is less smooth in time than actual oxygen consumption. Figure S.1 in Supplementary Material available at *Biostatistics* online shows the activPal METs over time for a particular day and subject overlaid with the model fit, where the fit includes the maximum likelihood estimates of the μ functions and the empirical best linear predictions of the random effects.

We next describe figures for the analysis of MET levels over the course of the entire day, 8:00 (8AM) to 20:00 (8PM). The top panel in Figure 4 illustrates that the mean MET levels in the treatment and control groups are similar before treatment assignment ($p = 0.27$), and the next panel shows the different means ($p < 0.0001$) in the 2 groups after treatment assignment. Both are averaged over days and the second panel is averaged over weeks. It is likely that the modes in the treatment group correspond to times when the treatment group subjects tend to schedule their exercise sessions.

The individual level daily random effect principal components are in the bottom 4 panels of Figure 4. These components are readily interpretable as the times of day when there is likely to be subject level variability in METs irrespective of day and week. The principal components for the other levels of variation are similar to the ones in this figure. One exception is that the other levels of variation show more variation in the evening (not shown).

The 4 levels of the covariance matrix described in Section 2.1 are shown in Figure 5. The top left panel shows that, independent of day and week, some subjects tend to be more or less active early in the

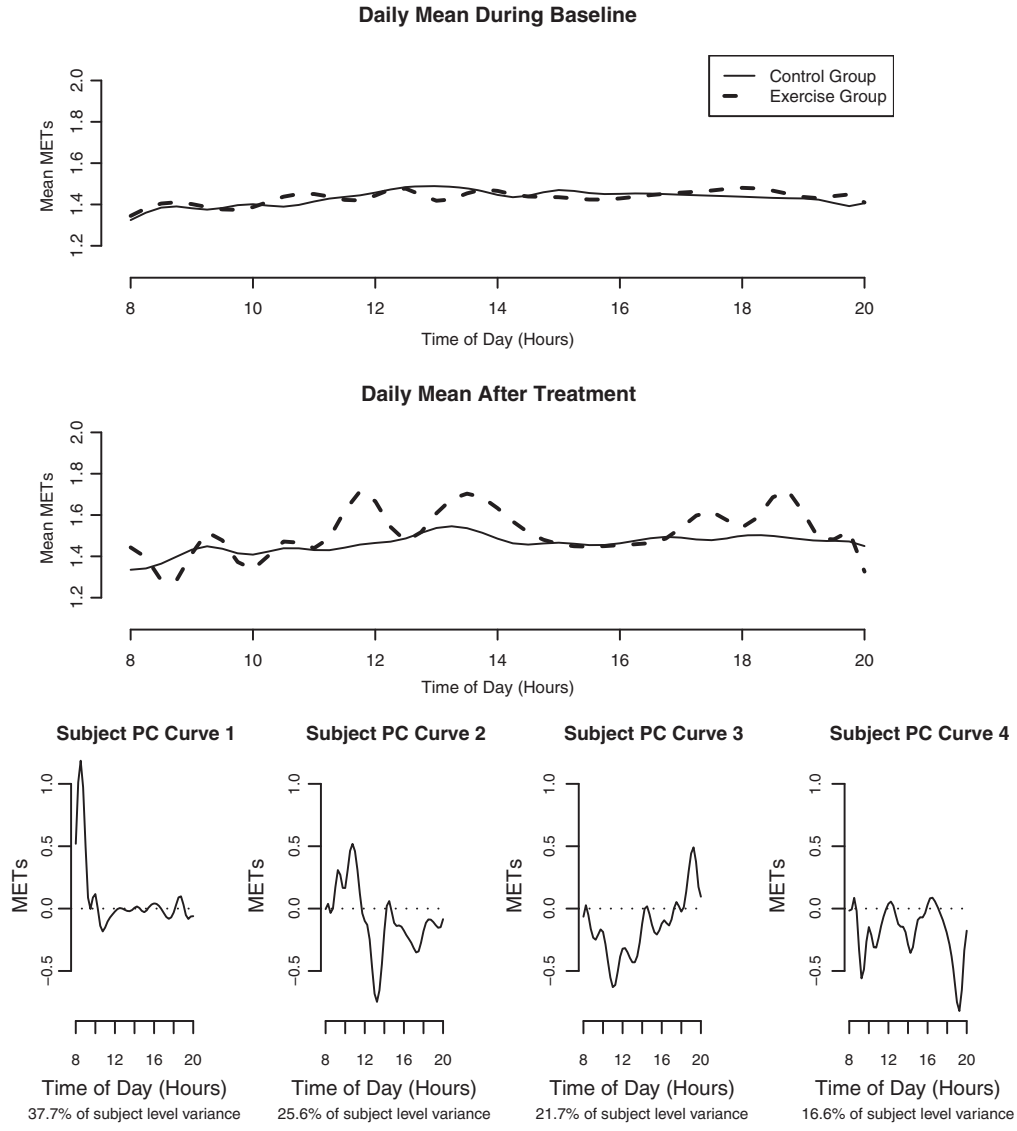

Fig. 4. The estimated mean METs over the course of the day for each treatment group before and after the treatment assignment.

morning with periods of relatively smaller variation just before and just after noon. The other panels show that week, day, and week-by-day levels of variation increase these peaks and also add variation in the evening. In contrast to the morning variation that is purely subject-specific, this variation suggests that the subject level variation in the evening tends to be week- or day-specific. Additionally, we note that most of correlation is positive and off-diagonal regions of the matrix are close to zero.

We also did an analysis using data before and after the first event of MVPA. See Figures S.2 and S.3 in the Supplementary Material available at *Biostatistics* online.

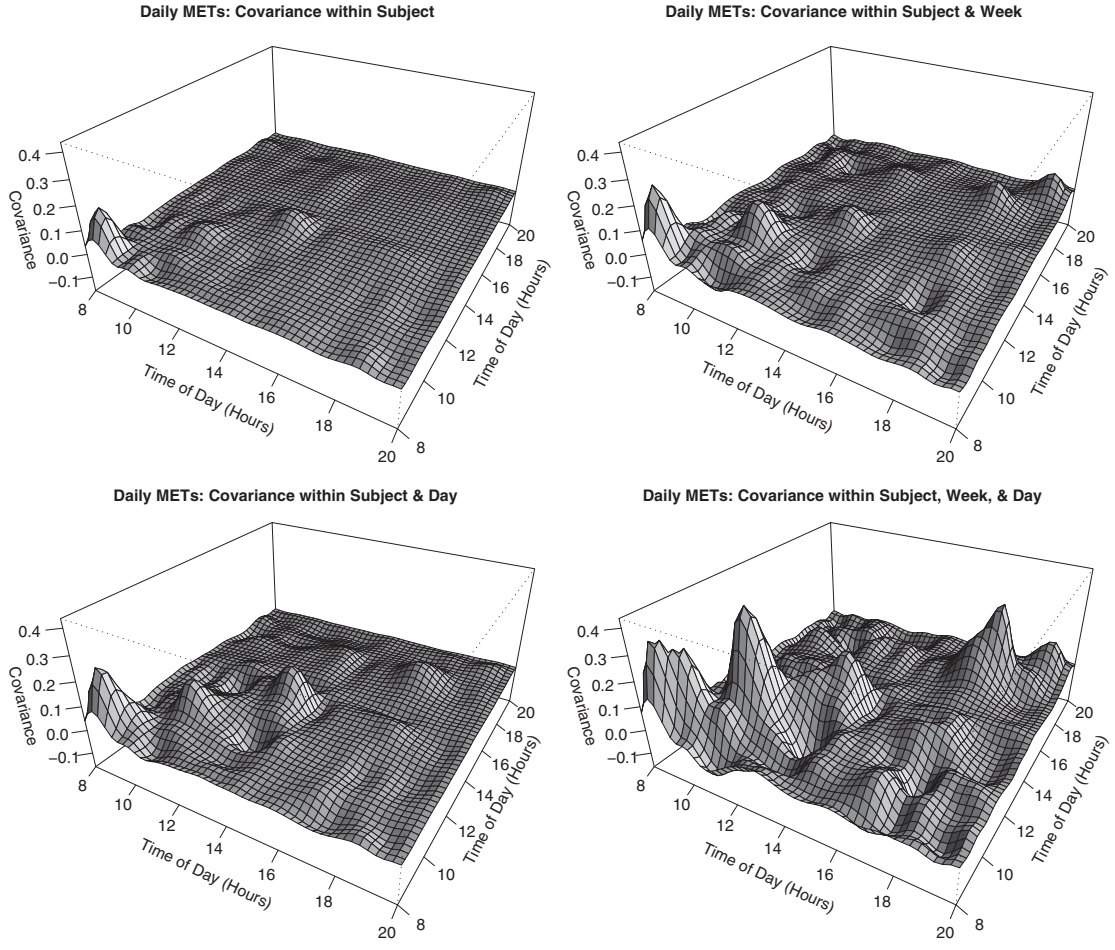

Fig. 5. Plots of the estimated covariance matrix of the subject random effects, including the contributions from the weeks, days and their interactions.

8. Conclusion

We have proposed a general modeling and estimation strategy for a treatment effect in the presence of 3-level correlated functional data. Our algorithm can handle incomplete data and employs a data-based method to select the number of principal components for the random curves. Simulation results for the model fitting strategy are encouraging and indicate little bias. We also discuss the performance of the Wald test to assess the evidence that linear combinations of mean structure parameters are zero. Our simulation results suggest that the Wald test works well for the unpenalized likelihood, but the number of knots should be restricted to be moderate, especially when the data are incomplete and the sample size is small.

The Wald test leads to biased results when using penalized likelihood. Wood (2006, 2013) discusses a modified Wald statistic based on the Bayesian variance estimator described in Section 5. However, his modifications assume that the data responses are independent, and in our case they are correlated with a complicated structure. It is an interesting problem for future research to extend the modified Wald statistic to accommodate the data associations inherent in our model.

We applied our model to analyze data from a physical activity intervention trial. The response data from this trial consisted of measurements of relative energy expenditure (MET level) every five minutes during the day for 5 days during 5 weeks from 63 individuals. The individuals in the trial were randomized into a treatment group that engaged in structured exercise and a control group. The detailed functional data analysis revealed how the patterns of activity in these groups differ in terms of timing and intensity.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at http://biostatistics.oxfordjournals.org.

ACKNOWLEDGMENTS

Conflict of Interest: None declared.

FUNDING

H.L. and R.J.C. were supported by a grant from the National Cancer Institute (R37-CA057030). This research has been partially supported by the Spanish Ministry of Science and Innovation (project MTM 2011-22664 which is co-funded by FEDER). H.A. was supported by a postdoctoral training grant from the National Cancer Institute (R25T-CA090301). S.K.K. and J.S. were supported by a National Cancer Institute grant (R01-CA121005). J.Z.H. was supported by a NSF grant (DMS-1208952).

REFERENCES

DI, C. Z., CRAINICEANU, C. M., CAFFO, B. S. AND PUNJABI, N. M. (2009). Multilevel functional principal component analysis. *The Annals of Applied Statistics* **3**, 458–488.

FREEDSON, P., BOWLES, H. R., TROIANO, R. AND HASKELL, W. (2012). Assessment of physical activity using wearable monitors: recommendations for monitor calibration and use in the field. *Medicine and Science in Sports and Exercise* **44**, S1–S4.

KOZEY-KEADLE, S., LIBERTINE, A., LYDEN, K., STAUDENMAYER, J. AND FREEDSON, P. S. (2014). Changes in sedentary time and spontaneous physical activity in response to an exercise training and/or lifestyle intervention. *Journal of Physical Activity and Health* **11**, 1324–1333.

LAIRD, N. M. AND WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.

LITTLE, R. J. A. AND RUBIN, D. B. (2002) *Statistical Analysis with Missing Data*, 2nd edition. Wiley-Interscience, United States.

LIU, C. AND RUBIN, D. B. (1994). The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. *Biometrika* **81**, 633–648.

POWERS, S. K. AND HOWLEY, E. T. (2001) *Exercise Physiology*. Boston: McGraw-Hill.

RUPPERT, D., WAND, M. P. AND CARROLL, R. J. (2003) *Semiparametric Regression*. New York: Cambridge University Press.

SCHAFER, J. L. (1998). Some improved procedures for linear mixed models. *Technical Report*. The Methodological Center, The Pennsylvania State University.

SERBAN, N. AND JIANG, H. (2012). Multilevel functional clustering analysis. *Biometrics* **68**, 805-–814.

WOOD, S. N. (2006). On confidence intervals for generalized additive models based on penalized regression splines. *Australian and New Zealand Journal of Statistics* **48**, 445–464.

WOOD, S. N. (2013). On p-values for smooth components of an extended generalized additive model. *Biometrika* **100**, 221–228.

ZHOU, L., HUANG, J. Z. AND CARROLL, R. J. (2008). Joint modelling of paired sparse functional data using principal components. *Biometrika* **95**, 601–619.

ZHOU, L., HUANG, J. Z., MARTINEZ, J. G., MAITY, A., BALADANDAYUTHAPANI, V. AND CARROLL, R. J. (2010). Reduced rank mixed effects models for spatially correlated hierarchical functional data. *Journal of the American Statistical Association* **105**, 390–400.

ZIPUNNIKOV, V., CAFFO, B., YOUSEM, D. M., DAVATZIKOS, C., SCHWARTZ, B. S. AND CRAINICEANU, C. (2011). Multilevel functional principal component analysis for high-dimensional data. *Journal of Computational and Graphical Statistics* **20**, 852–873.

[*Received January 1, 2014; revised April 8, 2015; accepted for publication April 9, 2015*]