



Published in final edited form as:

J Comput Graph Stat. 2022 ; 31(1): 219–230. doi:10.1080/10618600.2021.1950006.

Fast Univariate Inference for Longitudinal Functional Models

Erjia Cui,

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, USA

Andrew Leroux,

Department of Biostatistics and Informatics, University of Colorado, USA

Ekaterina Smirnova,

Department of Biostatistics, Virginia Commonwealth University, USA

Ciprian M. Crainiceanu

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, USA

Abstract

We propose fast univariate inferential approaches for longitudinal Gaussian and non-Gaussian functional data. The approach consists of three steps: (1) fit massively univariate pointwise mixed effects models; (2) apply any smoother along the functional domain; and (3) obtain joint confidence bands using analytic approaches for Gaussian data or a bootstrap of study participants for non-Gaussian data. Methods are motivated by two applications: (1) Diffusion Tensor Imaging (DTI) measured at multiple visits along the corpus callosum of multiple sclerosis (MS) patients; and (2) physical activity data measured by body-worn accelerometers for multiple days. An extensive simulation study indicates that model fitting and inference are accurate and much faster than existing approaches. Moreover, the proposed approach was the only one that was computationally feasible for the physical activity data application. Methods are accompanied by R software, though the method is “read-and-use”, as it can be implemented by any analyst who is familiar with mixed effects model software.

Keywords

longitudinal functional data; mixed model; DTI; wearable devices

1 Introduction

Longitudinal high dimensional data have become ubiquitous. For example, a Diffusion Tensor Imaging (DTI) study (Greven et al., 2010; Goldsmith et al., 2011, 2012; Scheipl et al., 2015) collected fractional anisotropy (FA) at multiple locations along the corpus

ecui1@jhmi.edu .

SUPPLEMENTARY MATERIALS

FUL_supp.pdf

Comparison of the computing time of two methods to obtain joint confidence bands, and additional simulation results. (.pdf file)

code.zip

The R code for model implementation and simulation. (.zip file)

callosum at multiple visits for each study participant. FA is a measure of water diffusion direction that is thought to be associated with white matter integrity and myelination. Figure 1 displays the FA measures along the corpus callosum for two study participants (left and right panels). Each curve represents data collected during a particular visit (color-coded according to visit number). The visit time, covariates, and health outcomes are collected at each visit but are not displayed; for more details see Goldsmith et al. (2012). Visual inspection of the left panel does not indicate an obvious temporal trend across visits, while each FA profile exhibits substantial within-visit noise. The FA profiles are less noisy for the second study participant (right panel), though the much lower FA values at the second visit could indicate substantial technical or biological variability. Given such data, the scientific goal is to quantify how FA at each location of the corpus callosum changes between visits and how these changes are associated with study-participant characteristics. To conceptualize the data structure, we denote by $Y_{ij}(s)$ the FA value for study participant i at visit j at time t_{ij} and location $s \in \mathcal{S}$ of the corpus callosum.

Another example is the National Health and Nutrition Examination Survey (NHANES) (Leroux et al., 2019; Smirnova et al., 2019; Cui et al., 2020), where minute-level activity counts (AC), a proprietary measure of physical activity (PA), were collected for up to seven consecutive days by hip-worn accelerometers. Data can be further transformed at the minute level to an active/inactive indicator depending on whether the observed AC is above/below a threshold. Here we use the threshold of 100, which was proposed by Matthews et al. (2008) for the NHANES hip accelerometry data and has been used extensively in the literature (Koster et al., 2012). Therefore, the data are binary (active/inactive), functional (minute-level measurements during the day), and multilevel (multiple days). To conceptualize the data structure, $Y_{ij}(s)$ denotes the binary indicator of whether a study participant i was active on day j at minute $s \in \mathcal{S} = \{1, \dots, 1440\}$. For the purpose of this paper, we are interested in studying the association between $Y_{ij}(s)$ and sex, age, day number from the beginning of the study, and day of the week.

Both these examples contain functional data measured at multiple visits, while the sample size in NHANES is over 10 times larger than the DTI study. Such data structures generalize standard longitudinal data, as instead of observing one scalar variable at each visit, one observes a high dimensional function. Functional data methods are sometimes used to model scalar longitudinal data (Yao et al., 2005), but longitudinal functional data has a more complex structure due to the large number of observations with complex correlations *at each visit*. Methods have been proposed for functional data with complex correlation structures (Guo, 2002; Morris and Carroll, 2006; Greven et al., 2010; Zipunnikov et al., 2014; Scheipl et al., 2015; Brockhaus et al., 2015; Scheipl et al., 2016; Shou et al., 2015; Zhu et al., 2019). In particular, Scheipl et al. (2015, 2016) proposed an inferential framework and associated software for correlated functional responses based on additive mixed models. This is an important step forward, though the method cannot currently handle very large data sets. For example, it takes more than 24 hours on a regular laptop (2.7 GHz Dual-Core Intel Core i5, 8GB RAM) to fit a functional random intercept model with $I = 1000$ subjects and an average of 5 visits per subject. The approach of Goldsmith et al. (2015) runs into similar scaling up problems. Indeed, they reported a total computation time of 10 days in their applications

using a dataset with fewer than 600 subjects, 5 visits per subject, and 144 observations per curve. Applying either method to our NHANES example is computationally impractical.

Thus, we conclude that this area of research is still in its initial stages of statistical development. Indeed, there is an increased need for methods that: (1) are scalable both in terms of number of study participants and of the dimension of the functional data; (2) can be applied to Gaussian and non-Gaussian data; and (3) preserve the interpretability of standard mixed effects models. To address this need, we propose fast univariate inferential (FUI) approaches for longitudinal functional data of any type. The approach consists of three steps: (1) fit massively univariate pointwise mixed effects models; (2) apply a smoother along the functional domain; and (3) obtain joint confidence bands using analytic approaches for Gaussian data or a bootstrap of study participants for non-Gaussian data. The first two steps are conceptually similar with Fan and Zhang (2000) and Reiss et al. (2017) for function-on-scalar regressions. However, to the best of our knowledge, these approaches did not model correlated functional responses or provided joint inference that accounts for this correlation. The methods proposed by Park et al. (2018) are the closest to our methods, though there are important differences. First, they estimate the fixed effects under independence across- and within-visits. We use mixed effects models across visits. Second, because Park et al. (2018) rely on fixed effects, their approach cannot be used to predict visit-specific functional effects or missing data. Third, their approach was developed for a narrower set of models.

The remainder of the paper is organized as follows. Section 2 introduces the longitudinal functional model. Section 3 describes two approaches for fixed effects inference, one analytic for Gaussian data and one based on the bootstrap of study participants for any type of data. We further discuss in Section 4 a simple and flexible simulation-based approach to obtain joint confidence bands. Section 5 presents the simulation results and comparisons with existing methods. Section 6 describes the applications of our model on DTI and NHANES study. We close with a discussion in Section 7. All code for model implementation, simulation and application is available on the supplementary material.

2 Massively Univariate Longitudinal Functional Model

Assume that data is of the type $Y_{ij}(s)$ on a grid $\{s_1, s_2, \dots, s_L\}$ of the compact functional domain \mathcal{S} . Data can be Gaussian or non-Gaussian, $i=1, 2, \dots, I$ is the index of the study participant, and $j=1, 2, \dots, J_i$ is the index of the longitudinal visit at time t_{ij} . In addition to the functional outcomes, $\mathbf{X}_{ij} = [X_{ij1}, X_{ij2}, \dots, X_{ijp}]^T \in \mathbb{R}^p$ are the fixed and $\mathbf{Z}_{ij} = [Z_{ij1}, Z_{ij2}, \dots, Z_{ijq}]^T \in \mathbb{R}^q$ are the random effects variables. We posit the following marginal three-step inferential approach:

First step:

At each location $s_l \in \mathcal{S}$, $l=1, 2, \dots, L$, fit a separate pointwise generalized linear mixed model (GLMM) $Y_{ij}(s_l) \sim \text{EF}\{\mu_{ij}(s_l)\}$, where EF denotes the exponential family distribution, $\mu_{ij}(s_l)$ is the conditional mean, and

$$\eta_{ij}(s_l) = g\{\mu_{ij}(s_l)\} = \mathbf{X}_{ij}^T \boldsymbol{\beta}(s_l) + \mathbf{Z}_{ij}^T \mathbf{u}_i(s_l). \quad (1)$$

Here $g(\cdot)$ is a link function and $\mathbf{u}_i(s)$ is a $q \times 1$ dimensional vector of random effects that depends on the location. Denote the estimates of fixed effects as $\boldsymbol{\beta}(s_1), \dots, \boldsymbol{\beta}(s_L)$ and of the linear predictors as $\tilde{\eta}_{ij}(s_1), \dots, \tilde{\eta}_{ij}(s_L)$ obtained from these univariate GLMMs. We refer to this as massively univariate analysis because a GLMM is fit at every location $s_l, l = 1, \dots, L$, where L can be very large (hence, the use of the word “massive”).

Second step:

Smooth the estimated fixed effects $\boldsymbol{\beta}(s_1), \dots, \boldsymbol{\beta}(s_L)$ and/or linear predictors $\tilde{\eta}_{ij}(s_1), \dots, \tilde{\eta}_{ij}(s_L)$ along the functional domain. Denote these smooth estimators by $\{\boldsymbol{\beta}(s), s \in \mathcal{S}\}$ and $\{\hat{\eta}_{ij}(s), s \in \mathcal{S}\}$, respectively. This can use any smoother that is or is not data adaptive, including not smoothing and taking the average over all locations.

Third step:

Obtain joint confidence bands for fixed effects parameters and/or linear predictors using analytic approaches for Gaussian data or a bootstrap of study participants for Gaussian and non-Gaussian data.

The key insight of our method is to decompose the complex correlation structure into longitudinal and functional directions. The first step of the analysis is to use the familiar univariate GLMMs, a procedure that can be easily implemented using parallel computing. This substantially reduces the computational burden of existing methods. The third step allows for joint inferences that take into account within- and between-visit correlations. This approach is not limited to estimating $\boldsymbol{\beta}(s)$ and $\eta_{ij}(s)$ and can be used for any measures of interest, including random effects, quantiles, and group means. We discuss the fixed effects inference of our approach below.

3 Fixed Effects Inference

Developing a principled statistical inferential framework for fixed effects in longitudinal functional models is difficult. Several approaches that account for the complex within- and between-study participant correlations exist and include different Bayesian approaches (Morris and Carroll, 2006; Morris et al., 2006; Zhu et al., 2011; Goldsmith et al., 2015; Zhang et al., 2016). Unfortunately, many of these methods require specialized software, are slow, and do not scale up with the number of study participants and dimension of the functional domain. Our proposed approach is philosophically closer to the methods proposed in Crainiceanu et al. (2012) and Park et al. (2018), which use bootstrap of study participants. For Gaussian functional data we provide an analytic solution, while for all types of data we propose a bootstrap approach.

3.1 Analytic Inference for Gaussian Functional Data

For Gaussian functional data the pointwise linear mixed model has the form

$$Y_{ij}(s_l) = \mathbf{X}_{ij}^T \boldsymbol{\beta}(s_l) + \mathbf{Z}_{ij}^T \mathbf{u}_i(s_l) + \epsilon_{ij}(s_l). \quad (2)$$

For the i th study participant at $s_l \in \mathcal{S}$ data are $\{\mathbf{Y}_i(s_l), \mathbf{X}_i, \mathbf{Z}_i\}$, where $\mathbf{Y}_i(s_l) = [Y_{i1}(s_l), \dots, Y_{iJ_i}(s_l)]$, $\mathbf{X}_i = [X_{i1}, \dots, X_{iJ_i}]$, and $\mathbf{Z}_i = [Z_{i1}, \dots, Z_{iJ_i}]$. The observations for the entire study population are then denoted as $\{\mathbf{Y}(s), \mathbf{X}, \mathbf{Z}\}$, where $\mathbf{Y}(s) = [\mathbf{Y}_1(s), \dots, \mathbf{Y}_I(s)]^T$, $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_I]^T$, and $\mathbf{Z} = \text{diag}(\mathbf{Z}_1^T, \dots, \mathbf{Z}_I^T)$. The matrix notation of equation (2) is $\mathbf{Y}(s) = \mathbf{X}\boldsymbol{\beta}(s) + \mathbf{Z}\mathbf{u}(s) + \boldsymbol{\epsilon}(s)$, where $\mathbf{u}(s) = [\mathbf{u}_1(s)^T, \dots, \mathbf{u}_I(s)^T]^T$ and $\boldsymbol{\epsilon}(s) = [\epsilon_{11}(s), \epsilon_{12}(s), \dots, \epsilon_{IJ_I}(s)]^T$ are mutually independent random coefficients and errors with a joint multivariate Gaussian distribution. The pointwise estimator of the fixed effects is $\hat{\boldsymbol{\beta}}(s) = \{\mathbf{X}^T \mathbf{V}^{-1}(s) \mathbf{X}\}^{-1} \mathbf{X}^T \mathbf{V}^{-1}(s) \mathbf{Y}(s)$, where $\mathbf{V}(s) = \mathbf{Z}\mathbf{H}(s)\mathbf{Z}^T + \mathbf{R}(s)$ and $\mathbf{H}(s)$ and $\mathbf{R}(s)$ are covariance matrices of $\mathbf{u}(s)$ and $\boldsymbol{\epsilon}(s)$, respectively. These fixed effects estimators are correlated across s_l , which needs to be taken into account when conducting joint inference, including when building joint confidence bands or conducting multiple testing. This correlation is modeled intrinsically by assuming $\text{Cov}\{\mathbf{u}(s_{l_1}), \mathbf{u}(s_{l_2})\} = \mathbf{G}(s_{l_1}, s_{l_2})$ and $\text{Cov}\{\boldsymbol{\epsilon}(s_{l_1}), \boldsymbol{\epsilon}(s_{l_2})\} = \mathbf{0}$ for all $s_{l_1} \neq s_{l_2}$. Thus, the covariance of the raw estimates at $s_l \in \mathcal{S}$ is $\text{Var}\{\hat{\boldsymbol{\beta}}(s)\} = \{\mathbf{X}^T \mathbf{V}^{-1}(s) \mathbf{X}\}^{-1}$, and between s_{l_1} and $s_{l_2} \in \mathcal{S}$ is

$$\text{Cov}\{\hat{\boldsymbol{\beta}}(s_{l_1}), \hat{\boldsymbol{\beta}}(s_{l_2})\} = \{\mathbf{X}^T \mathbf{V}^{-1}(s_{l_1}) \mathbf{X}\}^{-1} \mathbf{X}^T \mathbf{V}^{-1}(s_{l_1}) \mathbf{W}(s_{l_1}, s_{l_2}) \mathbf{V}^{-1}(s_{l_2}) \mathbf{X} \{\mathbf{X}^T \mathbf{V}^{-1}(s_{l_2}) \mathbf{X}\}^{-1}. \quad (3)$$

Here $\mathbf{W}(s_{l_1}, s_{l_2}) = \mathbf{Z}\mathbf{G}(s_{l_1}, s_{l_2})\mathbf{Z}^T$.

Estimates of $\mathbf{H}(s)$ and $\mathbf{R}(s)$ can be obtained directly from the mixed effects model software. An additional smoothing approach, for example using penalized splines (Ruppert et al., 2003), can be applied to each entry of these matrices along the functional domain. Estimations of the $\mathbf{G}(s_{l_1}, s_{l_2})$ is not as intuitive. However, the method of moments (MoM) procedure introduced in Greven et al. (2010) provides the blueprint for our procedure. More precisely, for any $s_{l_1} \neq s_{l_2}$

$$\begin{aligned} & E\left[\{Y_{ik}(s_{l_1}) - \mathbf{X}_{ik}^T \boldsymbol{\beta}(s_{l_1})\} \{Y_{ij}(s_{l_2}) - \mathbf{X}_{ij}^T \boldsymbol{\beta}(s_{l_2})\}\right] \\ &= \sum_{v=1}^q \sum_{t=1}^q Z_{ijv} Z_{ikt} \text{Cov}\{u_{it}(s_{l_1}), u_{iv}(s_{l_2})\}, \end{aligned} \quad (4)$$

for any $j, k=1, \dots, J_i$. This suggests a simple approach: regressing linearly the residual products $\{Y_{ik}(s_{l_1}) - \mathbf{X}_{ik}^T \boldsymbol{\beta}(s_{l_1})\} \{Y_{ij}(s_{l_2}) - \mathbf{X}_{ij}^T \boldsymbol{\beta}(s_{l_2})\}$ onto the covariates $\{Z_{ijv} Z_{ikt} : j, k=1, \dots, J_i\}$. A bivariate smoother, for example, the fast bivariate P-splines (Xiao et al., 2013), could be used to further reduce the variability of the covariance estimators. This approach does not guarantee that the resulting $\mathbf{H}(s)$, $\mathbf{R}(s)$, and $\mathbf{G}(s_{l_1}, s_{l_2})$ are positive definite. This is handled

by trimming the negative eigenvalues of these estimators at 0, as suggested in the literature (Yao et al., 2003; Hall et al., 2008; Greven et al., 2010).

While the model allows different smoothers in the second step, we provide the closed form solution for penalized splines. For simplicity, we focus on the r th fixed effect $\beta_r(s)$. Denoted by $\beta_r = [\tilde{\beta}_r(s_1), \dots, \tilde{\beta}_r(s_L)]^T$ the raw estimates of $\beta_r(s)$ obtained from the first step. Let $\mathbf{B}_r = [\mathbf{b}_{r1}, \dots, \mathbf{b}_{rK}]$ be the K -dimensional spline basis matrix, where $\mathbf{b}_{rk} = [b_{rk}(s_1), \dots, b_{rk}(s_L)]^T$, $k = 1, \dots, K$. K is usually chosen to be much smaller than L . Given a smoothing parameter λ_r , penalty matrix \mathbf{P}_r , define $\mathbf{S}_r = \mathbf{B}_r(\mathbf{B}_r^T \mathbf{B}_r + \lambda_r \mathbf{P}_r)^{-1} \mathbf{B}_r^T$. The smoothed estimator of the r th fixed effect is $\hat{\beta}_r = \mathbf{S}_r \beta_r$. The covariance matrix of $\hat{\beta}_r$ is $\text{Cov}(\hat{\beta}_r) = \mathbf{S}_r \text{Cov}(\beta_r) \mathbf{S}_r^T$, a sandwich smoother of $\text{Cov}(\beta_r)$ obtained from the first step.

This inferential approach is explicit when working with linear mixed effects models with Gaussian random effects and errors. While the inferential approach described in this section involves additional notation, the computations are straightforward. This analytic inferential solution offers the potential for substantially reducing the computational burden of performing bootstrap inference for large-scale Gaussian functional data and performs well in our simulation study (close to nominal coverage for 95% confidence bands). For outcomes with general (Gaussian and non-Gaussian) distributional assumptions we now discuss the bootstrap of study participants as a general solution.

3.2 Nonparametric Bootstrap Approach

Bootstrapping functional data is a practical approach for fixed effects inference (Cuevas et al., 2006; Crainiceanu et al., 2012). For complex correlated functional data, Park et al. (2018) proposed both study participant and residual bootstrap. Here we focus only on the study participant bootstrap approach because the residual bootstrap is not defined for generalized outcomes. The approach is described in Algorithm 1.

Algorithm 1:

Nonparametric Bootstrap for fixed effects inference

Data: $\{Y(s_l), l = 1, \dots, L\}, \mathbf{X}, \mathbf{Z}$

Result: $\text{Var}(\hat{\beta}(s_l)), l = 1, \dots, L$.

for $b = 1, \dots, B$ **do**

1. Re-sample / subject indices from $\{1, \dots, I\}$ with replacement. Denote the vector of re-sampled indices as $M_{(b)}$;
2. For the l 'th element of $M_{(b)}$, $l' = 1, \dots, L$, include all observations of the corresponding subject in the bootstrap sample. Denote the b th bootstrap sample as $\left\{ \left\{ Y_{M_{(b)}}(s_l), l = 1, \dots, L \right\}, \mathbf{X}_{M_{(b)}}, \mathbf{Z}_{M_{(b)}} \right\}$;
3. Fit the model in Section 2 using the b th bootstrap sample. Derive the fixed effects estimates $\{\hat{\beta}(s_l)_{(b)}, l = 1, \dots, L\}$;

end

4. For $l = 1, \dots, L$, derive $\text{Var}(\hat{\beta}(s_l))$ from B bootstrap estimates $\{\hat{\beta}(s_l)_{(1)}, \dots, \hat{\beta}(s_l)_{(B)}\}$. In practice we calculate the sample variance and use it as the estimator.

3.3 Extension to Random Effects Inference

The proposed inference procedure has a natural extension to random effects and visit-specific predictions. We provide a brief introduction for this framework. Given equation (2) and notations introduced in Section 3.1, the pointwise BLUP of the random effects is $\mathbf{u}(s) = \mathbf{H}(s)\mathbf{Z}^T\mathbf{V}^{-1}(s)\{\mathbf{Y}(s) - \mathbf{X}\boldsymbol{\beta}(s)\}$. Without loss of generality assume $\mathbf{R}(s) = \sigma^2(s)\mathbf{I}$. The uncertainty of $\mathbf{u}(s)$ can be measured through the conditional variance (Morris, 1983) as $\text{Var}\{\mathbf{u}(s)|\mathbf{Y}(s)\} = \sigma^2(s)\{\mathbf{H}(s) - \mathbf{H}(s)\mathbf{Z}^T\mathbf{V}^{-1}(s)\mathbf{Z}\mathbf{H}(s)\}$. The uncertainty of random effects between locations can be measured similarly using the empirical Bayes estimator.

4 Joint Confidence Bands

Inference for functional data has a natural connection with the problem of multiple testing due to the inherent correlations in the observed data along the functional domain. The pointwise confidence bands described in Sections 3.1 and 3.2 do not provide any information about joint coverage probabilities over the entire domain and are, in fact, valid when averaged across the functional domain. The Bonferroni correction (Bonferroni, 1936) is exact for independent data, but is inappropriate for functional data, which is correlated and can have arbitrary sampling density. For example, a functional domain could be sampled at one hundred or one million equally spaced points. In both situations the point estimators would barely change, whereas the joint Bonferroni corrected confidence intervals would depend on the choice of number of samples within the functional domain. At one extreme, if the number of points is allowed to go to infinity, the length of the joint confidence intervals diverges to infinity. This is unacceptable and cannot be addressed by changing the testing criterion to, say, the Benjamini-Hochberg false discovery rate (Benjamini and Hochberg, 1995). The reason is that both methods are overly conservative when the test statistics are highly correlated. As functional data can be sampled densely and correlations between observations increase with sampling density, the probability of failing to reject the null hypothesis using these corrections rapidly approaches one as the data are more densely sampled. An alternative was proposed by Cox and Lee (2008) for functional data based on the Westfall-Young randomization method (Westfall and Young, 1993).

The construction of joint confidence bands in the context of functional data analysis has been studied using various approaches including local linear estimators (Degras, 2011), piecewise constant splines (Ma et al., 2012), and polynomial splines (Cao et al., 2012). However, most of these methods assume independence between curves, which is not the setting considered here. Crainiceanu et al. (2012) and Park et al. (2018) proposed the bootstrap of study participants as a general inferential procedure for functional data with arbitrarily complex functional correlation structures. Here we follow a similar approach.

Based on a bootstrap of study participants we obtain the following estimators

$\boldsymbol{\beta}_r = [\hat{\beta}_r(s_1), \dots, \hat{\beta}_r(s_L)]^T$ and $\text{Var}(\boldsymbol{\beta}_r)$, where $\{\beta_r(s), s \in \mathcal{S}\}$ is the r th functional fixed effect. Let N_s be the sample size of simulated data. Our approach requires simulations from the multivariate normal distribution $\mathcal{N}\{\boldsymbol{\beta}_r, \text{Var}(\boldsymbol{\beta}_r)\}$. When the dimension L of the functional domain is large, this can be quite slow, but the problem can be addressed using a PCA decomposition of the bootstrap samples $\boldsymbol{\beta}_{r(1)}, \dots, \boldsymbol{\beta}_{r(B)}$; for details see Algorithm 2.

Algorithm 2:Level α joint confidence bands of $\hat{\beta}_r(s)$ **Data:** $\beta_r, \text{Var}(\beta_r), \beta_{r(1)}, \dots, \beta_{r(B)}, N_s$ **Result:** Joint confidence bands of $\{\hat{\beta}_r(s), s \in \mathcal{S}\}$.1. Perform Functional Principal Component Analysis (FPCA) on $[\beta_{r(1)}, \dots, \beta_{r(B)}]^T$. Derive the mean function $\mu = [\mu(s_1), \dots, \mu(s_L)]^T$, eigenvalues $\lambda_1, \dots, \lambda_L$ and eigenfunctions ψ_1, \dots, ψ_L , where $\psi_k = [\psi_k(s_1), \dots, \psi_k(s_L)]^T$, $k = 1, \dots, L$;**for** $n = 1, \dots, N_s$ **do**2. Simulate $\xi_{nk} \sim \mathcal{N}(0, \lambda_k)$ for $k = 1, \dots, K_T$. Calculate $\beta_{r,n} = \mu + \sum_{k=1}^{K_T} \xi_{nk} \psi_k$;3. Calculate $u_n = \max_{s_l \in \mathcal{S}} \{|\beta_{r,n} - \beta_r| / \sqrt{\text{diag}(\text{Var}(\beta_r))}\}$;**end**4. Obtain $q_{1-\alpha}$, the $(1-\alpha)$ empirical quantile of $\{u_1, \dots, u_{N_s}\}$;5. The joint confidence interval at $s_l \in \mathcal{S}$ is calculated as $\hat{\beta}_r(s_l) \pm q_{1-\alpha} \sqrt{\text{Var}(\beta_r)(l, l)}$. The upper and lower bounds of the joint confidence bands can be smoothed.

The number of functional principal component basis K_T in Step 2 is selected based on proportion of variance explained, as suggested in the FPCA literature, and usually does not exceed 100 when using a 95% variance explained threshold. As a result, this modified algorithm reduces the dimension of simulation from potentially large L to a more acceptable K_T ; see supplementary material for a comparison of computing time between this method and direct simulations from multivariate normal distribution.

5 Simulations

An extensive simulation study is used to assess: (1) the performance of the estimators and the pointwise/joint confidence bands; (2) how methods compare to existing approaches. The R code for the simulation study is provided in the supplementary material.

5.1 Simulation Setup

We simulate functional responses on an equally-spaced grid of $\mathcal{S} = [0, 1]$ with length L . For simplicity, we fix $p = 2$, the number of fixed effects for each point s on the functional domain and $q = 1$, the number of random effects for each point on the functional domain. Therefore, $\mathbf{X}_{ij} = [1, X_{ij1}]^T$ and $\mathbf{u}_i(s) = u_i(s)$, though the approach is designed for much larger p and q . For subject i at visit j the data generating model is

$$\eta_{ij}(s) = g\{\mu_{ij}(s)\} = \beta_0(s) + X_{ij1}\beta_1(s) + u_i(s), s \in \mathcal{S}.$$

The fixed effects covariates are simulated as $X_{ij1} \sim \mathcal{N}(0, 4)$, while the random effects are simulated as $u_i(s) = c_{i1}\psi_1(s) + c_{i2}\psi_2(s)$. We use the scaled orthonormal functions $\psi_1(s) \propto 1.5 - \sin(2\pi s) - \cos(2\pi s)$ and $\psi_2(s) \propto \sin(4\pi s)$ to capture the individual-level fluctuations. The random coefficients are generated from $c_{i1} \sim \mathcal{N}(0, 2\sigma_B^2)$ and $c_{i2} \sim \mathcal{N}(0, \sigma_B^2)$, respectively.

Here σ_B^2 is determined by the relative importance of random effects SNR_B , as described below. We consider the following simulation scenarios:

1. Simulation parameters

- Distribution of the functional responses: (a) Gaussian, (b) binary.
- Functional fixed effects $\beta(s)$:
- S1: $\beta_0(s) = -0.15 - 0.1 \sin(2\pi s) - 0.1 \cos(2\pi s)$, $\beta_1(s) = \frac{1}{20} \phi\left(\frac{s-0.6}{0.0225}\right)$;
- S2: $\beta_0(s) = 0.53 + 0.06 \sin(3\pi s) - 0.03 \sin(6.5\pi s)$,
 $\beta_1(s) = \frac{1}{60} \phi\left(\frac{s-0.2}{0.1^2}\right) + \frac{1}{200} \phi\left(\frac{s-0.35}{0.1^2}\right) - \frac{1}{250} \phi\left(\frac{s-0.65}{0.06^2}\right) + \frac{1}{60} \phi\left(\frac{s-1}{0.07^2}\right)$.

2. Sample size parameters

- Number of subjects: $I \in \{50, 100, 200, 400\}$.
- Mean number of visits per subject: $J \in \{5, 10, 20, 40\}$. For subject i the number of visits J_i is drawn from $\text{Poisson}(J)$ with a minimum of 1 visit.
- Dimension of the functional domain: $L \in \{50, 100, 200, 400\}$.

3. Signal-noise parameters

- Relative importance of random effects: $\text{SNR}_B \in \{0.5, 1\}$. Here SNR_B is the standard deviation of the fixed effects functions divided by the standard deviation of the random effects functions; see Scheipl et al. (2015) for detailed descriptions of this parameter.
- Signal-to-noise ratio: $\text{SNR}_\epsilon \in \{0.5, 1\}$ (Gaussian response only). Here SNR_ϵ is the standard deviation of the linear predictors divided by the standard deviation of the noise σ_ϵ .

In terms of fixed effects, the S1 functions are similar to those used by Goldsmith et al. (2015), are smooth and easy-to-estimate. The S2 functions have more complex shapes and are designed to approximate the estimated effects in our DTI application. To be specific, $\beta_0(s)$ mimics a typical fractional anisotropy (FA) trajectory on the corpus callosum in DTI while $\beta_1(s)$ represents a non-monotonic effect of scan date on the FA.

For each scenario we conducted 200 simulations. Considering all combinations of parameters would have been impossible even for our extensive computing resources. Instead, for each combination of simulation parameters (denoted by ‘‘Gaussian S1’’, ‘‘Gaussian S2’’, ‘‘Binary S1’’, ‘‘Binary S2’’ in the simulation results), we: (1) fix the sample size parameters at their baseline ($I = 50$, $J = 5$, $L = 50$), then change the signal-noise parameters; and (2) fix the signal-noise parameters at their baseline ($\text{SNR}_B = 0.5$, $\text{SNR}_\epsilon = 1$), then change the sample size parameter one at a time while fixing the other two sample size parameters at baseline. For example, we fix $J = 5$, $L = 50$ to evaluate model performance for a different number of study participants, I .

5.2 Comparisons to Existing Methods

Many methods have been developed to model correlated functional responses, including Functional Additive Mixed Models (FAMM) (Scheipl et al., 2015, 2016), Generalized Multilevel Function-on-Scalar Regression and Principal Component Analysis (GenMFPCA) (Goldsmith et al., 2015), Functional Linear Array Model (FLAM) (Brockhaus et al., 2015), Wavelet-based Functional Mixed Models (WFMM) (Morris and Carroll, 2006), FMEM (Yuan et al., 2014; Zhu et al., 2019). As described in Scheipl et al. (2015), FAMM outperformed Bayesian WFMM for smooth curves. For FLAM, Brockhaus et al. (2015) reported a similar estimation accuracy with FAMM in their simulations. Although inference for FLAM could potentially be conducted via subject-level bootstrap, this implementation was not available in the FDboost package (Brockhaus et al., 2017). For FMEM we could not identify general purpose software. Therefore, we compare FAMM and GenMFPCA, which have well-documented and easy-to-use implementations for inference.

Our method (FUI) is implemented in the `lfosr3s()` function in the supplementary material, which implements univariate GLMMs and then applies a penalized cubic spline smoother. Results are highly robust to the choice of smoother. For FAMM we use the `pffr()` function from `refund` package (Goldsmith et al., 2020) in R. We used 15 and 20 cubic B-splines bases with first order difference penalty for the population average and global functional intercept respectively; see `bs.yindex` and `bs.int` arguments in the `pffr()` function. We have increased the number of bases from the default to increase the performance of FAMM.

Because Goldsmith et al. (2015) reported considerably larger computing time than FAMM, comparisons with GenMFPCA are restricted to smaller sample sizes. In addition, GenMFPCA software is only applicable to binary response. We have attempted to manually implement GenMFPCA for Gaussian responses, but, probably due to our sub-optimal implementation, our implementation was quite slow. To ensure a fair comparison of computing time, we focus on comparing with GenMFPCA only for binary response. The results are shown in the supplementary material.

5.3 Model Evaluation Criteria

We compare the performance of each method (FUI, FAMM, GenMFPCA) with respect to: (1) accuracy in estimating fixed effects; (2) inference on fixed effects; and (3) computational efficiency.

Accuracy of fixed effects estimation was assessed using the integrated squared error (ISE) of fixed effects, defined as $ISE_k = \int_0^1 (\hat{\beta}_k(s) - \beta_k(s))^2 ds$, $k = 0, 1$. The mean integrated squared error (MISE) is calculated by averaging ISE across simulated datasets. We show ISE of $\beta_1(s)$, as similar results were obtained for $\beta_0(s)$.

Inferential performance was assessed by calculating the empirical coverage probability of 95% pointwise confidence bands at each location, then taking the average along the functional domain. For FUI, we also report the empirical coverage probability of 95% joint confidence bands proposed in Section 4. For FUI we use analytic inference ($\text{mean} \pm 2\text{sd}$) for Gaussian responses and bootstrap inference for other families of distributions. In the

nonparametric bootstrap, we have used the formula $\text{mean} \pm 2.2\text{sd}$ instead of $\text{mean} \pm 2\text{sd}$. This is a simple solution that provides remarkably good results and seems to account for the extra within-subject variability that may be missed by conducting a bootstrap of study participants.

For each scenario, we do 200 simulations on the Joint High Performance Computing Exchange (JHPCE) Cluster with 1 core per simulation. The computing time of each method is obtained under different scenarios.

5.4 Simulation Results: Signal-to-noise parameters

The simulation results for different signal-noise parameters are shown in Figure 2 under the scenario with the smallest number of subjects, mean number of visits per subject, and dimension of the observed functional responses ($I = 50, J = 5, L = 50$). We only display results for the Gaussian response, as similar results were obtained for binary responses. Left two panels: fixed effect is S1. Right two panels: fixed effect is S2. The MISE decreases as signal increases, either by increasing SNR_B or SNR_ϵ . In addition, these two parameters exhibit similar scaling behavior using both estimation methods. Specifically, increasing SNR_B (or SNR_ϵ) from 0.5 to 1 decreases MISE by about 60% in S1 and about 50% in S2 using either FUI or FAMM. For this small sample size, the coverage is close to the nominal level and the computing time for both methods is almost identical (not displayed).

5.5 Simulation Results: Sample Size Parameters

The simulation results for different sample size parameters are shown in Figure 3. As results tend to be quite consistent, we display the results for Gaussian outcomes and fixed effects S1 (denoted by ‘‘Gaussian S1’’ in the title of each panel). Results for other combinations are in the supplementary material. The baseline setting is $I = 50, J = 5, L = 50, \text{SNR}_B = 0.5, \text{SNR}_\epsilon = 1$. All other parameters are fixed at their baseline values when one sample size parameter is changed. Left column: number of subjects (I). Middle column: mean number of visits per subject (J). Right column: dimension of the functional domain (L). The ISE and computing time for 200 simulations are displayed in the top and bottom row, respectively. The inference results are shown in Table 1. For FUI, we report the empirical coverage probability of both joint (denoted as ‘‘Coverage (Joint)’’) and pointwise (denoted as ‘‘Coverage (Pointwise)’’) 95% confidence bands. For FAMM, we report the coverage of the pointwise confidence bands (denoted as ‘‘Coverage’’).

As the number of subjects increases, the MISE for both FUI and FAMM decreases. The estimation accuracy of the two methods is similar, and the coverage of the confidence bands, both joint and pointwise for FUI and pointwise for FAMM, reach their nominal level. However, FAMM computing time increases substantially when the number of subjects increases (see bottom left panel). Indeed, the median computing time exhibits a second-order polynomial shape in number of subjects, with median computing time exceeding 10 hours on the cluster when $I = 400$. In addition, the memory usage for FAMM increases substantially despite the use of the efficient `mgcv::bam` implementation. For example, we were not able to perform simulations for FAMM when $I = 1000$, as both memory (more than 40 GB RAM) and computing time (unknown) exceeded our extensive resources. In contrast,

FUI required only 40 seconds for $I = 500$ study participants and there are no problems with fitting even for $I = 10000$.

As the mean number of visits per subject (second column of panels) and dimension of the functional domain (third column of panels) increase, both methods display similar estimation accuracy. The confidence bands of both methods, including joint and pointwise confidence bands for FUI and pointwise confidence bands for FAMM, also have good coverage close to the nominal level (middle and bottom blocks of Table 1). For $I = 50$, $L = 50$ FAMM requires similar computation time (3 to 7 minutes) for 5 to 40 visits per study participant, while FUI takes on average less than 1 minute in all scenarios. For $I = 50$, $J = 5$ computing time of FUI increases with the dimension of the functional domain (bottom right panel) while FAMM remains unaffected. This increase is expected as we run a GLMM at every location and the time for these GLMMs simply add up. However, our method is easy to parallelize, which would reduce the fitting time to the time it would take to fit a single GLMM. To the best of our knowledge, FAMM does not currently have a parallel implementation.

The computation time advantages of FUI should not be surprising given the way mgcv is used to estimate functional models. Specifically, random effects are incorporated by fully constructing the random effects design matrix and applying ridge penalties. Therefore, for a model with a subject-specific functional random intercept, $u_{0j}(s)$, the design matrix adds k_b columns, where k_b is the number of spline bases used to represent the functional random intercept. This is not a problem when I is in the range of 50 to 100, but it becomes problematic when $I > 200$. Take our physical activity data application for example where $L = 1440$, $J = 7$ and $I = 1680$ and consider a simple functional model $\beta_0(s) + u_{j0}(s)$. Assume that the population mean function $\beta_0(s)$ uses 20 B-spline basis functions (the default for FAMM). The design matrix without random effects is $1440 \times 7 \times 1680 = 16,934,400$ rows and 20 columns. Assume that the model $u_{j0}(s)$ uses 15 B-spline basis functions ($k_b = 15$). Then the full design matrix is as large as $16,934,400 \times 25,220$, since $20 + (15 \times 1680) = 25,220$, amounting to over 400 billion elements. This explains why FAMM runs into substantial computational challenges when the number of study participants increases. The mgcv package does have method for handling large datasets through the `mgcv::bam` function which avoids constructing and performing computations on the full design matrix. However, even with these added efficiencies FAMM runs into substantial computational challenges. This could be addressed in the future, but our current solution provides a practical, “read-and-use”, stable alternative for a moderate to large number of study participants.

5.6 Simulation Summary

Our method achieves similar accuracy with the state-of-the-art FAMM method for fixed effects under different simulation settings, including different signal-to-noise parameters and sample size parameters. FUI is much faster than FAMM when the number of subjects is large. To the best of our knowledge, FUI is the first inferential method that is demonstrated to work with over 1500 study participants. The reason for implementing such approaches is practical, as many datasets, including our NHANES application, contain such sample sizes. Both joint and pointwise confidence bands of FUI exhibit good coverage to the

nominal level. For both FUI and FAMM, estimation accuracy is affected by the change of signal-noise parameters and sample size parameters. These results for FAMM are consistent with the simulation results reported in Scheipl et al. (2015).

6 Applications

In this section, we apply our method to the motivating examples introduced in Section 1.

6.1 DTI Study

Multiple sclerosis (MS) is an autoimmune mediated disease that affects the central nervous system (CNS) and can lead to substantial motor and cognitive disability. While the exact cause of MS remains unknown, modern neuroimaging has played a crucial role in the diagnosis and management of MS. A promising imaging technique is Diffusion Tensor Imaging based on Magnetic Resonance Imaging (DTI-MRI or, shorter, DTI). DTI provides measures of water diffusion in the brain, which are thought to be associated with white matter integrity. Fractional anisotropy (FA) is a measure of diffusion anisotropy derived from DTI. A zero value of FA corresponds to perfectly isotropic diffusion (water diffuses unrestricted in all directions), while a value of one of FA corresponds to perfectly anisotropic diffusion (water diffuses only in one direction). Values of FA fall somewhere within the (0, 1) range with higher values corresponding to more anisotropic (more organized) water diffusion. FA can be calculated at every location in the brain.

Here we focus on the FA calculated along the corpus callosum, a nerve tract connecting the left and right cerebral hemispheres; see Goldsmith et al. (2011) and Greven et al. (2010) for in-depth descriptions of the data. For our purposes, the data set consists of 142 study participants (42 healthy individuals and 100 MS patients). For healthy individuals there is only one visit, whereas for MS patients there are multiple visits with an average of 3.4 and a maximum of 8 visits per MS patient. There were a total of 382 visits across MS patients and healthy individuals. Corpus callosum is a three-dimensional C-shaped nerve fiber bundle that connects the left and right brain hemispheres. For the purpose of this application, several landmarks were manually identified on the brain image and FA was calculated as an average FA at 93 locations along the corpus callosum. Thus, the data consist of a 382×93 dimensional matrix, where each row corresponds to a brain image visit and each column corresponds to a particular location in the corpus callosum, resulting in a total of 35526 observations. The study participant ID, age, sex and date of scan information are also available for each study participant at each visit.

For illustration purposes, we are interested in quantifying the association between age, sex and date of scan with FA measurements along the corpus callosum. Using the notation introduced in Section 2, the longitudinal functional responses are denoted by $Y_{ij}(s)$, and are the observed FA values along the equally-spaced grid of $s \in \mathcal{S} = \{1, \dots, 93\}$. For the i th individual at the j th visit, the fixed effects $\mathbf{X}_{ij} = [1, X_{ij1}, X_{ij2}, X_{ij3}, X_{ij4}]^T$ where X_{ij1} is a binary indicator of case (1 for MS patients and 0 for healthy individuals), X_{ij2} is the date of scan (converted into year unit and treated as numeric), X_{ij3} is a binary indicator of sex (there were no self-identified non-binary participants in this study) and X_{ij4} is the age at baseline scan (in years). For each location we fit a random intercept and slope model

$$Y_{ij}(s) = \beta_0(s) + X_{ij1}\beta_1(s) + X_{ij2}\beta_2(s) + X_{ij3}\beta_3(s) + X_{ij4}\beta_4(s) + u_{i0}(s) + u_{i1}(s)X_{ij2} + \epsilon_{ij}(s),$$

where $(u_{i0}(s), u_{i1}(s))$ have a joint zero-mean bivariate Normal distribution independent of $\epsilon_{ij}(s)$, which are iid $\mathcal{N}(0, \sigma^2(s))$. This standard linear mixed effects model is fit 93 times, once for each location s . After pointwise fitting, a penalized spline smoother is used to smooth each $\beta_l(\cdot)$, $l = 0, 1, 2, 3, 4$ coefficient separately. All five smoothers use a cubic basis with 15 equally spaced knots and REML estimation of the smoothing parameter. The pointwise and joint confidence intervals are obtained as described in Section 3 and Section 4.

Figure 4 displays the point estimators of the fixed effects parameters (dashed blue lines) for the intercept, case, scan date, sex, and age (five panels from left to right), respectively. The dark gray regions correspond to the 95% pointwise confidence bands, while the light gray regions correspond to the 95% joint confidence bands. The intercept estimator is consistent with the geometry of the corpus callosum and previously published literature. Compared with healthy individuals, MS patients of the same sex and age at the same date of scan have significantly lower FA at most locations along the corpus callosum. This result may indicate lower anisotropy corresponds among MS patients, which may be consistent with brain micro-structure damage. The middle panel shows a highly significant *increase* in the FA as a function of scan date at most locations of the corpus callosum. To the best of our knowledge, there is no biological plausible reason for such an increase in anisotropy. Therefore, the result may correspond to the change in technology and software, which led to a large, deterministic, increase in measured FA. The effects of sex and age are not statistically significant at any point along the corpus callosum.

Figure 5 displays the data (left panels), together with the pointwise estimators (middle panels) and smoothed estimators along the functional domain (right panels). The first and second rows correspond to study participants ID 2017 and 2085, respectively. For study participant ID 2017, the point estimators are consistent with substantial reduction in the visit-to-visit variability. Unsurprisingly, after smoothing (right-top panel) visit-specific profiles are slightly smoother along the functional domain, but with a similar reduced visit-to-visit variability. Comparing the middle and right top panels indicates that the pointwise linear mixed effects models did the “heavy lifting”, while the functional smoothing led to mostly cosmetic changes. This need not be the case in general when the noise and correlation structures could be quite different. Results are similar for study participant ID 2085. These results suggest that: (1) there is a statistically significant, but small fixed effect for the date of the visit; and (2) much of the observed variability is due to visit-to-visit fluctuations in FA trajectories (measurement error); and (3) the effect of scan date is largely contained in the fixed effects. The results about the decomposition of the observed residual variability after accounting for fixed effects are consistent with the literature. Indeed, Greven et al. (2010) showed that only 2 to 3% of the observed variability can be attributed to the longitudinal functional slope.

6.2 NHANES Study

The National Health and Nutrition Examination Survey (NHANES) is a large cohort study conducted by the US Centers for Disease Control (CDC) in two-year waves to assess the health and nutritional status of the US population. The objectively measured physical activity (PA) data were collected using hip-worn accelerometers on study participants in the 2003–2006 waves. The accelerometry data are publicly available as minute-level activity counts (AC), a proprietary measure of PA, and can be accessed in an analysis-ready format through the R `rnhanesdata` package (Leroux et al., 2019). Specifically, the accelerometry data were collected on 14631 individuals in the NHANES 2003–2004 and 2005–2006 waves. In this study, we focus on individuals with age between 18 and 30 at the time of accelerometer wear. In addition, we exclude individuals who had less than 3 days of data with at least 10 hours of estimated wear time or were labeled as poor data quality by NHANES. The number of available days vary between individuals with a maximum of 7. The final data include 1680 individuals with 8765 days, each with 1440 observations per day for a total of 12621600 minute-level observations.

We would like to investigate whether being non-sedentary is associated with gender, age, and day of the week (e.g., Monday, Tuesday). For the j th day of the i th study participant, the longitudinal functional response $Y_{ij}(s)$ is now a binary indicator, which equals to 1 if the AC at minute $s \in \{1, \dots, 1440\}$ exceeds 100 and 0 if not. The fixed effect $\mathbf{X}_{ij} = [1, X_{ij1}, X_{ij2}, X_{ij3}, \mathbf{X}_{ij4}^T]^T$ where $X_{ij1} = j$ is the day number, X_{ij2} is a binary indicator of sex (female=1), X_{ij3} is the age, and \mathbf{X}_{ij4} is a 6×1 binary vector indicating the day of the week of day j with order {Mon, Tue, ..., Sat}. For example, for study participant i if day 3 is Tuesday, then the second element of \mathbf{X}_{i34} is 1 while all others are 0; if day 3 is Sunday, all elements in \mathbf{X}_{i34} are 0. Denote $\mathbf{f}(s) = [f_1(s), \dots, f_6(s)]^T$. At every minute s of the day we fit a random intercept and slope model

$$\text{logit}\{\Pr(Y_{ij}(s) = 1 | \mathbf{X}_{ij}, \mathbf{u}_i)\} = \beta_0(s) + X_{ij1}\beta_1(s) + X_{ij2}\beta_2(s) + X_{ij3}\beta_3(s) + \mathbf{X}_{ij4}^T \mathbf{f}(s) + u_{i0}(s) + u_{i1}(s) \cdot j,$$

where $[u_{i0}(s), u_{i1}(s)]^T \sim N\{0, \Sigma_{u_i}(s)\}$. This GLMM is fit 1440 times at every location s . The same penalized spline smoother using cubic basis with 15 equally spaced knots and REML estimated smoothing parameter is applied to smooth estimated coefficients from pointwise fits separately.

Figure 6 displays the estimated coefficients together with the 95% pointwise (dark gray shaded area) and joint (light gray shaded area) confidence bands based on 100 bootstrap replicates. The shape of the functional intercept is consistent with the published literature and indicates less activity during the night, a sharp increase in the morning, sustained activity during the day and a reduction of activity in late evening. The effect of sex in this age group (18 to 30) is statistically significant throughout most of the day with the exception of the late afternoon/early evening period (≈ 4 –10PM). The sex effect is strongest during the late evening/early morning hours, when, on average, females are less active. This result could correspond to more restful sleep, more sleep, or higher compliance to study protocol for women. The effect of age is also highly significant during the night and early to late

morning indicating that older study participants in this age group (18 to 30) tend to have less activity during the night and more activity in the morning and early afternoon. These findings are consistent with those reported by Varma et al. (2017).

The fixed effect of day number ($\hat{\beta}_1(s)$) indicates that individuals are slightly more likely to be active during the nighttime hours and less likely to be active during normal waking hours as a function of day, though the nighttime effect is only pointwise significant during the period roughly corresponding to 4AM-6AM, but not significant when considering the joint confidence bands. This suggests that there may be a small “habituation effect”, particularly during the daytime. Habituation effects was proposed as a potential psychological effect of increasing PA at the beginning of wearing a device merely by its presence. Compared to Sundays (the reference category), weekdays correspond to lower levels of activity during the predawn hours (12AM-4AM), higher levels activity in the morning (6AM-11AM), and about the same levels of activity during the afternoon and evening hours. Fridays and Saturdays correspond to more activity in the evening than Sundays. Saturdays tend to have lower activity in the morning compared to weekdays, but more activity than Sundays. These results are consistent with previous findings that individuals tend to be less active during the night and more active during the day on weekdays. These differences are likely due to social behaviors on the weekends and obligations related to school and/or work during the weekdays.

Computationally, the initial model was fit in 672 minutes, with bootstrapping requiring ~ 67200 additional minutes (1120 hours); results are reported on a standard laptop. However, because of the parallel nature of our method, each location-specific fit can be estimated separately and combined at the end. This would reduce the computation time by 3 orders of magnitude, as most computational time is taken by fitting 1440 univariate GLMMs. A fully parallel implementation the entire procedure would take ~ 30 seconds for one model fit and ~ 1 hour for the inferential procedure. Though this may seem like a long time, we are not aware of any other methods that could fit such a model for this large longitudinal functional data set.

7 Discussion

We have introduced a fast univariate inferential approach for longitudinal functional models, a computationally efficient method for quantifying the association between covariates and a broad family of longitudinal functional outcomes. The model is estimated using a three-step procedure: (1) fit a series of separate standard longitudinal mixed models; (2) smooth estimators along the functional domain; and (3) construct pointwise and joint confidence bands using analytic approaches for Gaussian data or a nonparametric bootstrap of study participants for any type of data. The proposed method is highly computationally efficient because the first step can be parallelized to allow fitting large high-dimensional datasets. The second step is actually optional and one can either smooth or not smooth the resulting coefficients. Building joint confidence bands is a crucial component for conducting joint inference and performing testing multiplicity adjustment. Another major advantage of the proposed approach is its conceptual simplicity and availability in the R software. Most

importantly, methods are “read-and-use”, meaning that data scientists with a working knowledge of GLMMs can easily implement and apply our procedures.

The most important methodological contribution of this paper is to provide practical methods for building pointwise and joint confidence bands for very large longitudinal functional datasets. Simulation results suggest that our method achieves similar estimation accuracy and nominal coverage compared with existing methods, while the computation is much faster when the number of subjects is large (> 100).

Our work is not without limitations. First, the smoothing parameter selection assumes that the residuals of the raw estimated fixed effects around the true coefficient along the functional domain are independent. Second, changing the quantile of the confidence bands from 2 to 2.2 (lengthening the confidence bands by 10%) for bootstrap inference works well in our simulation study, but a more rigorous procedure and associated simulations may be necessary. Third, our method is only applicable to concurrent functional models, and can only take into account functional covariates that are measured on the same grid as the functional responses. Fourth, we focus on the fixed effects inference in this paper. While the inference for visit-specific predictions and other metrics falls into a similar framework, as introduced in Section 3.3, the extension is nontrivial and exceeds the scope of the current paper.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The project described was supported by Award Number R01 NS060910 from the National Institute of Neurological Disorders and Stroke, and training grant T32 AG000247 from the National Institute on Aging. The MRI/DTI data were collected at Johns Hopkins University and the Kennedy-Krieger Institute.

References

- Benjamini Y and Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 57(1):289–300.
- Bonferroni C (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8:3–62.
- Brockhaus S, Rügamer D, and Greven S (2017). Boosting functional regression models with fdboost. arXiv preprint arXiv:1705.10662.
- Brockhaus S, Scheipl F, Hothorn T, and Greven S (2015). The functional linear array model. *Statistical Modelling*, 15(3):279–300.
- Cao G, Yang L, and Todem D (2012). Simultaneous inference for the mean function based on dense functional data. *Journal of Nonparametric Statistics*, 24(2):359–377. [PubMed: 22665964]
- Cox DD and Lee JS (2008). Pointwise testing with functional data using the westfall–young randomization method. *Biometrika*, 95(3):621–634.
- Crainiceanu CM, Staicu A-M, Ray S, and Punjabi N (2012). Bootstrap-based inference on the difference in the means of two correlated functional processes. *Statistics in Medicine*, 31(26):3223–3240. [PubMed: 22855258]

- Cuevas A, Febrero M, and Fraiman R (2006). On the use of the bootstrap for estimating functions with functional data. *Computational Statistics & Data Analysis*, 51(2):1063–1074.
- Cui E, Crainiceanu CM, and Leroux A (2020). Additive functional cox model. *Journal of Computational and Graphical Statistics*, pages 1–14. [PubMed: 33013150]
- Degras DA (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Statistica Sinica*, pages 1735–1765.
- Fan J and Zhang J-T (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(2):303–322.
- Goldsmith J, Bobb J, Crainiceanu CM, Caffo B, and Reich D (2011). Penalized functional regression. *Journal of Computational and Graphical Statistics*, 20(4):830–851. [PubMed: 22368438]
- Goldsmith J, Crainiceanu CM, Caffo B, and Reich D (2012). Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(3):453–469. [PubMed: 22679339]
- Goldsmith J, Scheipl F, Huang L, Wrobel J, Di C, Gellar J, Harezlak J, McLean MW, Swihart B, Xiao L, Crainiceanu C, and Reiss PT (2020). refund: Regression with Functional Data. R package version 0.1–23.
- Goldsmith J, Zipunnikov V, and Schrack J (2015). Generalized multilevel function-on-scalar regression and principal component analysis. *Biometrics*, 71(2):344–353. [PubMed: 25620473]
- Greven S, Crainiceanu C, Caffo BS, and Reich DS (2010). Longitudinal functional principal component analysis. *Electronic Journal of Statistics*, pages 1022–1054. [PubMed: 21743825]
- Guo W (2002). Functional mixed effects models. *Biometrics*, 58(1):121–128. [PubMed: 11890306]
- Hall P, Müller H-G, and Yao F (2008). Modelling sparse generalized longitudinal observations with latent gaussian processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):703–723.
- Koster A, Caserotti P, Patel KV, Matthews CE, Berrigan D, Van Domelen DR, Brychta RJ, Chen KY, and Harris TB (2012). Association of sedentary time with mortality independent of moderate to vigorous physical activity. *PloS One*, 7(6):e37696. [PubMed: 22719846]
- Leroux A, Di J, Smirnova E, McGuffey EJ, Cao Q, Bayatmokhtari E, Tabacu L, Zipunnikov V, Urbanek JK, and Crainiceanu C (2019). Organizing and analyzing the activity data in nhanes. *Statistics in Biosciences*, 11(2):262–287. [PubMed: 32047572]
- Ma S, Yang L, and Carroll RJ (2012). A simultaneous confidence band for sparse longitudinal regression. *Statistica Sinica*, 22:95. [PubMed: 23459083]
- Matthews CE, Chen KY, Freedson PS, Buchowski MS, Beech BM, Pate RR, and Troiano RP (2008). Amount of time spent in sedentary behaviors in the united states, 2003–2004. *American Journal of Epidemiology*, 167(7):875–881. [PubMed: 18303006]
- Morris CN (1983). Parametric empirical bayes inference: theory and applications. *Journal of the American statistical Association*, 78(381):47–55.
- Morris JS, Arroyo C, Coull BA, Ryan LM, Herrick R, and Gortmaker SL (2006). Using wavelet-based functional mixed models to characterize population heterogeneity in accelerometer profiles: a case study. *Journal of the American Statistical Association*, 101(476):1352–1364. [PubMed: 19169424]
- Morris JS and Carroll RJ (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(2):179–199. [PubMed: 19759841]
- Park SY, Staicu A-M, Xiao L, and Crainiceanu CM (2018). Simple fixed-effects inference for complex functional models. *Biostatistics*, 19(2):137–152. [PubMed: 29036541]
- Reiss PT, Huang L, Wu P-S, Chen H, and Colcombe S (2017). Pointwise influence matrices for functional-response regression. *Biometrics*, 73(4):1092–1101. [PubMed: 28405966]
- Ruppert D, Wand MP, and Carroll RJ (2003). *Semiparametric regression*. Number 12. Cambridge university press.
- Scheipl F, Gertheiss J, Greven S, et al. (2016). Generalized functional additive mixed models. *Electronic Journal of Statistics*, 10(1):1455–1492.
- Scheipl F, Staicu A-M, and Greven S (2015). Functional additive mixed models. *Journal of Computational and Graphical Statistics*, 24(2):477–501. [PubMed: 26347592]

- Shou H, Zipunnikov V, Crainiceanu CM, and Greven S (2015). Structured functional principal component analysis. *Biometrics*, 71(1):247–257. [PubMed: 25327216]
- Smirnova E, Leroux A, Cao Q, Tabacu L, Zipunnikov V, Crainiceanu C, and Urbanek JK (2019). The predictive performance of objective measures of physical activity derived from accelerometry data for 5-year all-cause mortality in older adults: National health and nutritional examination survey 2003–2006. *The Journals of Gerontology: Series A*.
- Varma VR, Dey D, Leroux A, Di J, Urbanek J, Xiao L, and Zipunnikov V (2017). Re-evaluating the effect of age on physical activity over the lifespan. *Preventive Medicine*, 101:102–108. [PubMed: 28579498]
- Westfall PH and Young SS (1993). Resampling-based multiple testing: Examples and methods for p-value adjustment, volume 279. John Wiley & Sons.
- Xiao L, Li Y, and Ruppert D (2013). Fast bivariate p-splines: the sandwich smoother. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, pages 577–599.
- Yao F, Müller H-G, Clifford AJ, Dueker SR, Follett J, Lin Y, Buchholz BA, and Vogel JS (2003). Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate. *Biometrics*, 59(3):676–685. [PubMed: 14601769]
- Yao F, Müller H-G, and Wang J-L (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590.
- Yuan Y, Gilmore JH, Geng X, Martin S, Chen K, Wang J. L., and Zhu H (2014). Fmem: Functional mixed effects modeling for the analysis of longitudinal white matter tract data. *NeuroImage*, 84:753–764. [PubMed: 24076225]
- Zhang L, Baladandayuthapani V, Zhu H, Baggerly KA, Majewski T, Czerniak BA, and Morris JS (2016). Functional car models for large spatially correlated functional datasets. *Journal of the American Statistical Association*, 111(514):772–786. [PubMed: 28018013]
- Zhu H, Brown PJ, and Morris JS (2011). Robust, adaptive functional regression in functional mixed model framework. *Journal of the American Statistical Association*, 106(495):1167–1179. [PubMed: 22308015]
- Zhu H, Chen K, Luo X, Yuan Y, and Wang J-L (2019). Fmem: Functional mixed effects models for longitudinal functional responses. *Statistica Sinica*, 29(4):2007. [PubMed: 31745381]
- Zipunnikov V, Greven S, Shou H, Caffo B, Reich DS, and Crainiceanu C (2014). Longitudinal high-dimensional principal components analysis with application to diffusion tensor imaging of multiple sclerosis. *The Annals of Applied Statistics*, 8(4):2175. [PubMed: 25663955]

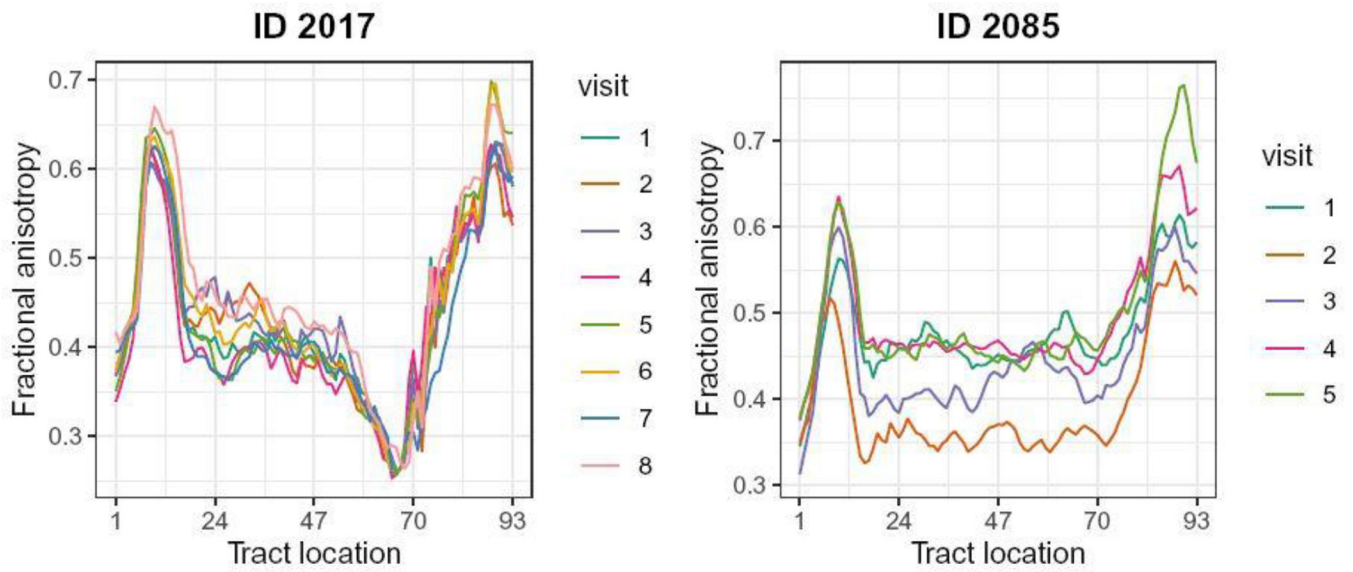


Fig. 1. The fractional anisotropy (FA) tract profiles for the corpus callosum (functional domain) of two study participants in the DTI study. Left panel: ID 2017. Right Panel: ID 2085. For each study participant, each curve represents the tract profiles at one longitudinal visit. The visit number is color coded.

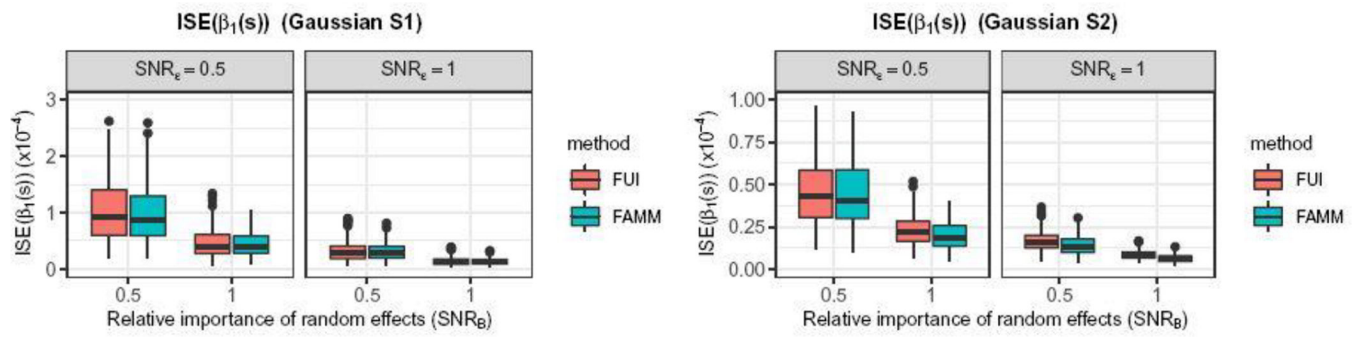


Fig. 2.

Estimation accuracy for FUI (red) and FMM (blue) under different relative importance of random effects (SNR_B , x axis) and signal-to-noise ratios (SNR_e , labels in the gray-shaded area of each panel). Functional response is Gaussian; parameters: $I = 50$, $J = 5$, $L = 50$. Left two panels: S1. Right two panels: S2.

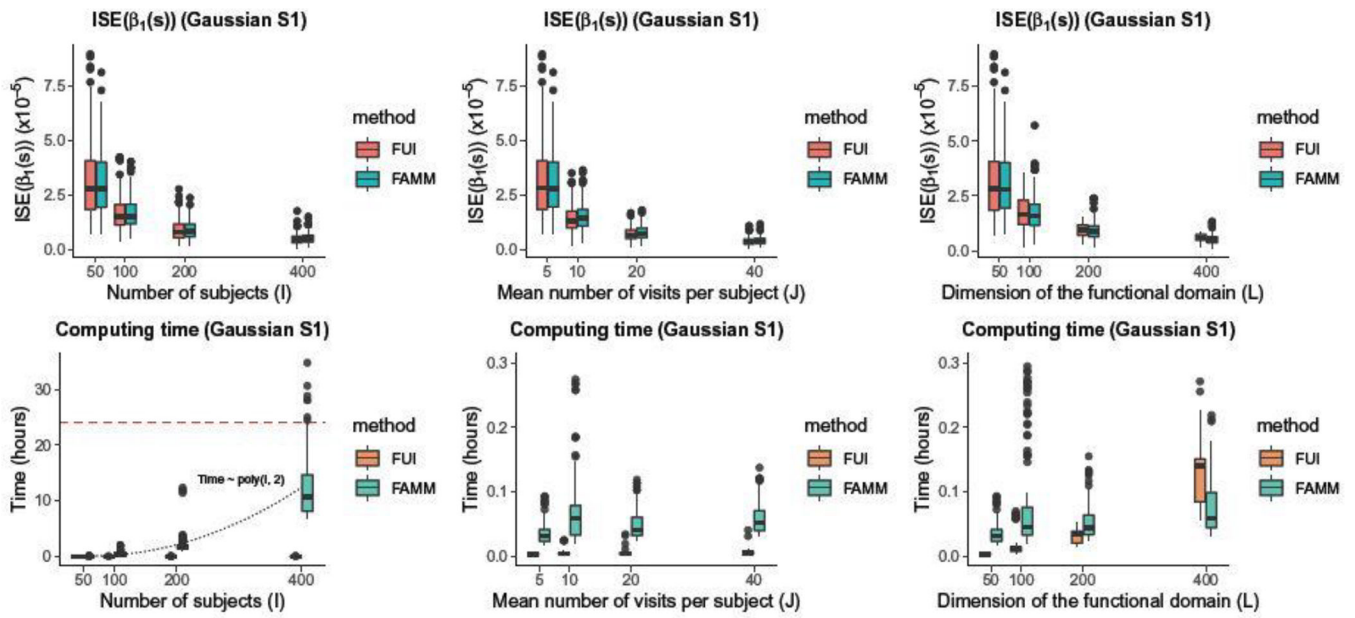


Fig. 3. Estimation accuracy (top row) and computing time (bottom row) for FUI (red) and FAMM (blue) from 200 simulations. Response is Gaussian and the true fixed effects functions are S1. The baseline setting is $I=50, J=5, L=50, \text{SNR}_B=0.5, \text{SNR}_\epsilon=1$. All other parameters are fixed at their baseline values when one sample size parameter is changed. Left column: number of subjects (I). Middle column: mean number of visits per subject (J). Right column: dimension of the functional domain (L).

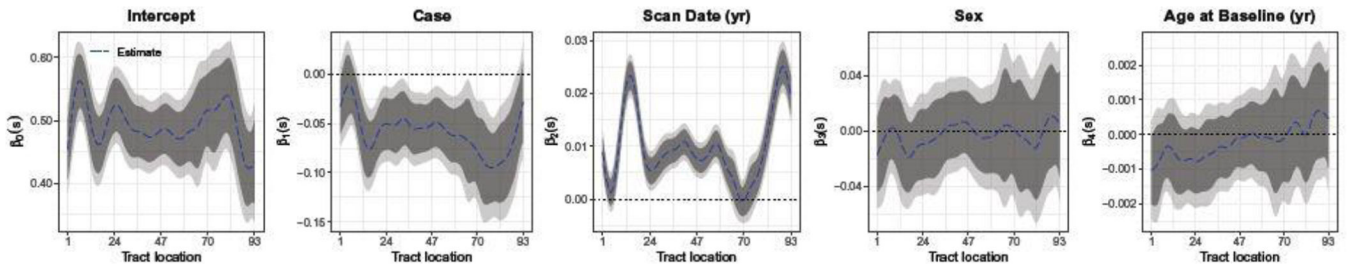


Fig. 4.

Fixed effects estimates (dashed blue line), 95% pointwise confidence intervals (dark gray shaded area), and 95% joint confidence intervals (light gray shaded area) in the DTI study. Panels from left to right: intercept, case, date of scan, sex, age at baseline.

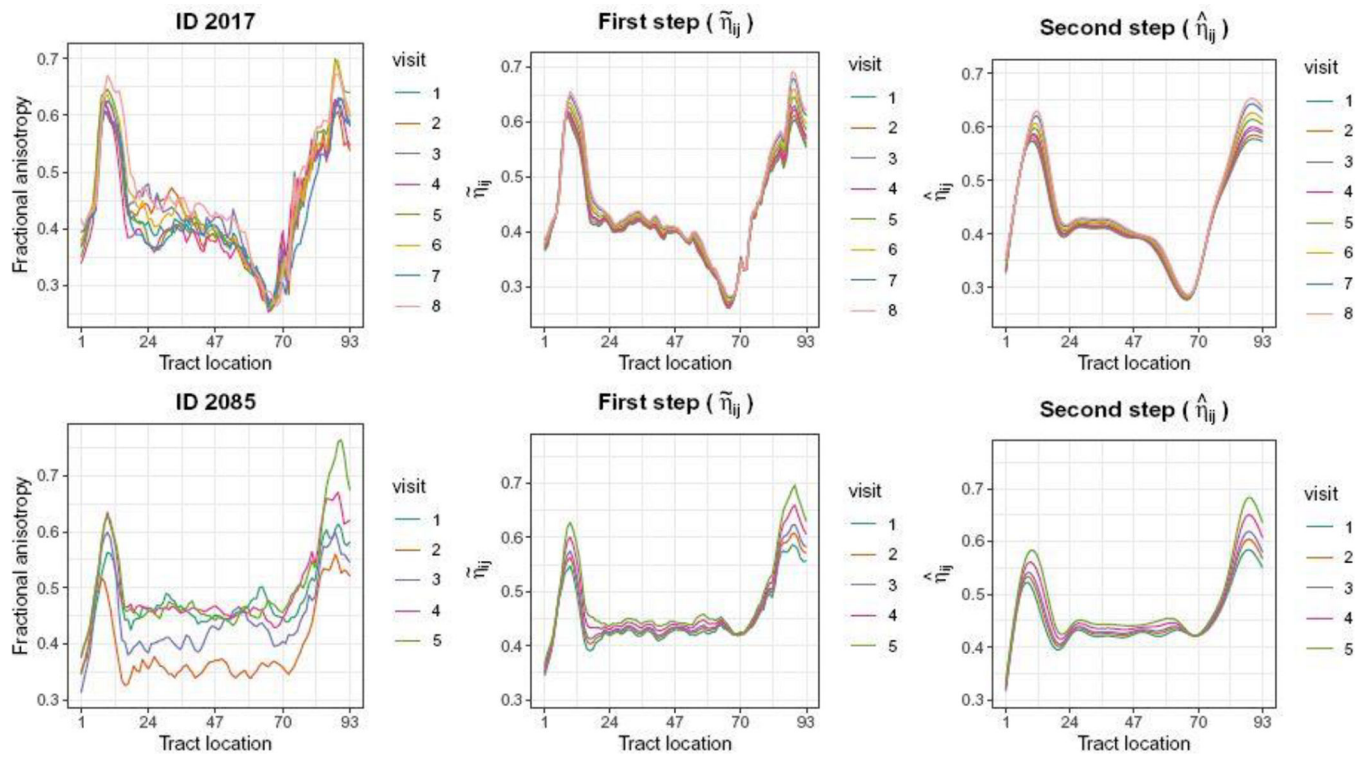


Fig. 5. Fractional anisotropy (FA) tract profiles and estimated predictors for two study participants (first row: ID 2017, second row: ID 2085). First column: FA tract profiles for the corpus callosum over multiple visits. Second column: pointwise estimated predictor $\tilde{\eta}_{ij}$. Third column: smoothed estimated predictor $\hat{\eta}_{ij}$ of the pointwise predictors.

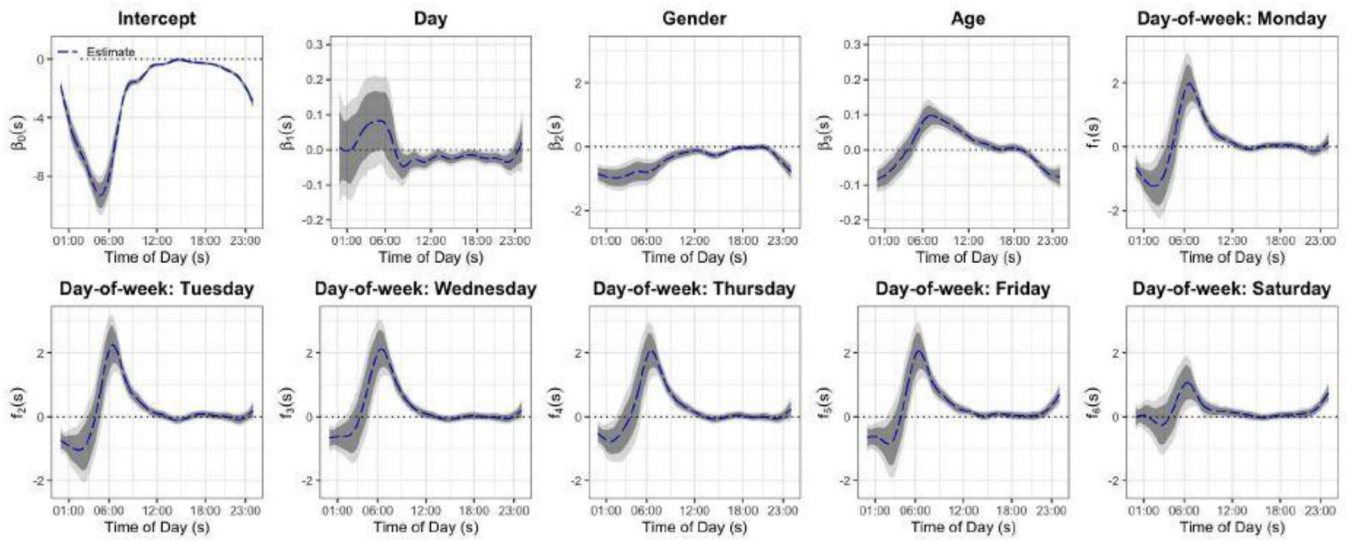


Fig. 6. Estimated coefficients from the NHANES data application. Smoothed coefficient estimates are denoted using blue dashed lines. Pointwise and joint 95% confidence intervals are shown as the dark and light gray shaded area, respectively.

Table 1

Empirical coverage probability of 95% joint and pointwise confidence bands using FUI and 95% pointwise confidence bands using FAMM from 200 simulations. Response is Gaussian and the true fixed effects functions are S1. The pointwise confidence band is constructed as $\text{mean} \pm 2\text{sd}$ and the joint is $\text{mean} \pm q_{0.975} \times \text{sd}$. The baseline setting is $I = 50$, $J = 5$, $L = 50$, $\text{SNR}_B = 0.5$, $\text{SNR}_\epsilon = 1$. All other parameters are fixed at their baseline values when one sample size parameter is changed.

Method	Type	Number of subjects (I)			
		50	100	200	400
FUI	Coverage (Joint)	0.93	0.96	0.94	0.95
	Coverage (Pointwise)	0.94	0.95	0.94	0.95
FAMM	Coverage	0.96	0.96	0.96	0.94
Method	Type	Mean number of visits per subject (J)			
		5	10	20	40
FUI	Coverage (Joint)	0.93	0.95	0.97	0.96
	Coverage (Pointwise)	0.94	0.95	0.95	0.95
FAMM	Coverage	0.96	0.96	0.96	0.96
Method	Type	Dimension of the functional domain (L)			
		50	100	200	400
FUI	Coverage (Joint)	0.93	0.94	0.94	0.94
	Coverage (Pointwise)	0.94	0.94	0.95	0.95
FAMM	Coverage	0.96	0.96	0.96	0.96