A marginal approach to reduced-rank penalized spline smoothing with application to multilevel functional data

Huaihou Chen[1], **Yuanjia Wang**[2], **Myunghee Cho Paik**[3], and **H. Alex Choi**[4]

[1]Department of Child and Adolescent Psychiatry, New York University School of Medicine New York, NY 10016, U.S.A

[2]Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY 10032, U.S.A

[3]Department of Statistics, Seoul National University, 1 Gwanakro, Gwanakgu, Seoul, Korea 151-742

[4]Department of Neurosurgery and Neurology, The University of Texas Health Science Center at Houston Medical School, Houston, Texas 77030, U.S.A

Abstract

Multilevel functional data is collected in many biomedical studies. For example, in a study of the effect of Nimodipine on patients with subarachnoid hemorrhage (SAH), patients underwent multiple 4-hour treatment cycles. Within each treatment cycle, subjects' vital signs were reported every 10 minutes. This data has a natural multilevel structure with treatment cycles nested within subjects and measurements nested within cycles. Most literature on nonparametric analysis of such multilevel functional data focus on conditional approaches using functional mixed effects models. However, parameters obtained from the conditional models do not have direct interpretations as population average effects. When population effects are of interest, we may employ marginal regression models. In this work, we propose marginal approaches to fit multilevel functional data through penalized spline generalized estimating equation (penalized spline GEE). The procedure is effective for modeling multilevel correlated generalized outcomes as well as continuous outcomes without suffering from numerical difficulties. We provide a variance estimator robust to misspecification of correlation structure. We investigate the large sample properties of the penalized spline GEE estimator with multilevel continuous data and show that the asymptotics falls into two categories. In the small knots scenario, the estimated mean function is asymptotically efficient when the true correlation function is used and the asymptotic bias does not depend on the working correlation matrix. In the large knots scenario, both the asymptotic bias and variance depend on the working correlation. We propose a new method to select the smoothing parameter for penalized spline GEE based on an estimate of the asymptotic mean squared error (MSE). We conduct extensive simulation studies to examine property of the proposed estimator under different correlation structures and sensitivity of the variance estimation to the choice of smoothing parameter. Finally, we apply the methods to the SAH study to evaluate a recent debate on discontinuing the use of Nimodipine in the clinical community.

Keywords

Penalized spline; GEE; Semiparametric models; Longitudinal data; Functional data

1 Introduction

Multilevel functional data is often collected in many biomedical studies. For example, in a study of the effect of Nimodipine on patients diagnosed with subarachnoid hemorrhage (SAH) introduced in section 1.1 below, each patient is administered with one of the two doses of Nimodipine during multiple 4-hour treatment cycles, and their clinical outcomes were recorded every 5 seconds and the averages in each 10-minute interval were reported (Choi et al. 2012). The data has a multilevel structure with treatment cycles nested within subjects and repeated outcome measurements nested within cycles.

Modeling multilevel functional data has recently received extensive attention. Brumback and Rice (1998) used smoothing splines to analyze nested samples of functional data. Guo (2002) proposed a functional mixed effects model with functional random effects fitted by Kalman filtering. Zhou et al. (2008) proposed jointly modeling of paired sparse functional data with reduced rank principal components. Baladandayuthapani et al. (2008) and Staicu et al. (2010) developed a functional mixed effects model based Bayesian approach for correlated multilevel spatial data. Crainiceanu et al. (2009) proposed methods for functional regression with multilevel functional predictors under a mixed effects model framework. Apanasovich et al. (2008) proposed a composite likelihood based approach for correlated binary data. Di et al. (2009) developed a functional multivariate analysis of variance which used a few functional principal components to reduce dimensionality.

The above methods on multilevel functional data in the literature focus on conditional approaches through a functional mixed effects model or functional principal components analysis. In some clinical trials, such as the SAH study described in section 1.1 (Choi et al. 2012), the goal is to estimate the population average effect, or the group difference. To achieve this goal, marginal approaches are more suitable than conditional approaches. There is a wealth of literature on nonparametric marginal regression models through local polynomial or kernel based methods (see for example, Lin and Carroll 2000, Welsh et al. 2002, Lin et al. 2004). In particular, Welsh et al. (2002) compared the efficiency of the local kernel based methods with spline based methods for marginal models with single-level functional data. However, it is not straightforward to apply kernel smoothing to accommodate the multilevel data structure. A few other works that propose marginal models fitted by smoothing splines include Ibrahim and Suliadi (2010a, 2010b). In a variable selection setting, Fu (2003) proposed penalized generalized estimating equation to handle collinearity among variables.

The pros and cons of marginal versus conditional model for longitudinal data has been debated extensively in literature (see for example, Diggles et al. 2002). Marginal models provide a direct estimation of the population average effect. For continuous outcomes, parameters in the marginal and conditional model have the same interpretation as the population average effect. In contrast, for generalized outcomes, conditional models do not directly give estimators of population averaged marginal effects due to a non-identity link function. Therefore, when marginal effects are of interest, subject-specific random effects need to be integrated out, usually through numerical integration. In addition, a potential computational advantage of the marginal regression is that since the procedure only requires the specification of the first two moments of the marginal distribution, it is particularly effective for modeling correlated generalized outcomes. Numerical algorithms for

conditional approaches for multilevel functional data with generalized outcomes may not always converge. In the SAH study, the functional mixed effects model with a two-level random effects did not converge for the primary binary outcome. Furthermore, a widely known advantage of using a robust sandwich variance estimator in marginal models is that it remains consistent under a misspecified working correlation structure. For a parametric model, the estimated mean parameters are asymptotically efficient when the true correlation is used. However, for nonparametric models fitted by local polynomials, such property does not hold (Lin and Carroll 2000). To take into account the within-cluster correlation to improve efficiency, seemingly unrelated kernel estimator should be used (Wang 2003, Lin et al. 2004). It may not be straightforward to adapt local kernel based approaches to effectively account for more complicated multilevel functional data.

There is few literature on marginal approaches for multilevel functional data through reduced-rank penalized spline smoothing (P-spline; Eilers and Marx 1996; Ruppert et al. 2003). In this work, we study semiparametric marginal regression models with multilevel continuous or generalized functional data. The developed penalized spline GEE and robust variance estimator provide tools to evaluate the population average effect without requiring integrating over the distribution of the random effects. The rest of the manuscript is organized as follows. In section 1.1, we provide an overview of the clinical study that motivated this research. In section 2, we present the penalized spline GEE for marginal models along with a robust variance estimator. In section 3, we investigate large sample properties of the proposed estimator and show that similar to independent data, the asymptotics fall into two scenarios. For the small knots scenario, the estimated population mean function is asymptotically efficient when the true correlation function is used and the asymptotic bias does not depend on the working correlation matrix. For the large knots scenario, both the asymptotic bias and variance depend on the working correlation. In section 4, we use the asymptotic results to develop a new method to select the smoothing parameter for marginal regressions based on an estimated asymptotic average mean squared error (MSE). In section 5, we carry out extensive simulation studies to examine the performance of the approaches under various models. In section 6, we apply the proposed methods to the SAH study to evaluate a debatable recommendation in the clinical community to discontinue the use of Nimodipine among SAH patients. Lastly, in section 7 we conclude with some remarks.

1.1 Motivating example: Nimodipine and the SAH study

Subarachnoid hemorrhage (SAH) is an acute cerebrovascular event caused by rupture of a cerebral aneurysm. It can have devastating consequences, causing serious morbidity and mortality. Nimodipine is the only medication shown in phase III trials to improve clinical outcomes after SAH (Dorhout et al. 2007). Although initial clinical studies did not document low blood pressure as a side effect, a decrease in the blood pressure and even a decrease in brain oxygen delivery has been observed during routine clinical usage (Stiefel et al. 2004). In light of these clinical findings, the effectiveness of Nimodipine has been challenged. In fact, recent clinical guidelines have suggested discontinuing the use of Nimodipine when administration is associated with significant decreases in blood pressure. Although this is a strong recommendation, the committee admits to little clinical data supporting their recommendation (Dringer et al. 2011). In this work, we aim to quantify the effect of Nimodipine on various physiologic outcomes from an observational study of SAH patients admitted to a neurological intensive care unit (Choi et al. 2012).

Nimodipine is administered to patients with SAH at one of the two doses every 4 hours, creating multiple 4-hour treatment cycles. Within each treatment cycle, subjects' vital signs such as mean arterial blood pressure (MAP) and brain tissue oxygenation are recorded continuously and averaged over 10-minute intervals to reduce noise. Every 4 hours a patient

receives a high dose or a low dose of Nimodipine depending on his or her clinical profile. This scenario creates a multi-level data structure with treatment cycles nested within subjects and repeated outcome measurements nested within cycles. Our primary research interest is to estimate mean physiologic outcomes averaged across treatment cycles and across subjects to evaluate the acute effects of Nimodipine on systemic and brain physiology. Specific research questions include whether Nimodipine increases or reduces the MAP and its effect on the risk of cerebral autoregulation loss.

2 Marginal nonparametric or semiparametric models and reduced rank smoothing

2.1 Single-level continuous functional data

Let $i = 1, \cdots, n$ index subject and let $j = 1, \cdots, n_i$ index observations within a subject. Let $Y_i = (Y_{i1}, \cdots, Y_{in_i})^T$ denote a vector of outcomes on the ith subject, let X_{ij} denote a vector of covariates and let $X_i = (X_{i1}, \cdots, X_{in_i})^T$. For simplicity in illustration, we present methods for a nonparametric model. It is straightforward to extend it to semiparametric models such as a partially linear model. Consider the marginal regression,

$$E(Y_{ij}|X_{ij})=f(X_{ij}), \quad \text{cov}(Y_i|X_i)=\sum\nolimits_i,$$

where $f(\cdot)$ is an unspecified smooth function. Let $B(x)$ denote an l-dimensional vector of spline basis functions such as B-splines or truncated polynomials. For the pth order truncated polynomial with K knots, $B(x)=[1, x, \cdots, x^p, (x-\tau_1)_+^p, \cdots, (x-\tau_K)_+^p]^T$, where $\tau_1, \cdots, \tau_K$ is a sequence of knots. Let $B_i = [B(X_{i1}), \cdots, B(X_{in_i})]^T$ denote the $n_i \times l$ matrix of basis functions. Given the covariance matrix Σ_i, the usual penalized spline estimator with the qth order penalty minimizes a weighted least-square,

$$\sum_{i=1}^n (Y_i - B_i\theta)^T \sum\nolimits_i^{-1} (Y_i - B_i\theta) + \lambda \int_a^b \left\{ [B^T(x)\theta]^{(q)} \right\}^2 dx,$$

where θ is a vector of basis coefficients and λ is a smoothing parameter. Using a difference-based penalty matrix, the above can be expressed as:

$$\sum_{i=1}^n (Y_i - B_i\theta)^T \sum\nolimits_i^{-1} (Y_i - B_i\theta) + \lambda\theta^T D_q\theta,$$

where D_q is an appropriate penalty matrix depending on the chosen basis. For example, for the pth order truncated polynomial basis, we have $q = p + 1$ and $D_q = \text{diag}(\mathbf{0}_{p+1}, \mathbf{1}_K)$. The fitted value at a fixed point is $\hat{f}(x) = B^T(x)\hat{\theta}$ and its standard error is estimated from

$$B^T(x)\left(\sum_{i=1}^n B_i^T \sum\nolimits_i^{-1} B_i + \lambda D_q\right)^{-1} \sum_{i=1}^n B_i^T \sum\nolimits_i^{-1} B_i \left(\sum_{i=1}^n B_i^T \sum\nolimits_i^{-1} B_i + \lambda D_q\right)^{-1} B(x). \quad (1)$$

In practice, Σ_i is often unknown and will be estimated under a parametric model. A mis-specified parametric model would lead to an inconsistent estimate of the standard error of $\hat{f}(x)$.

Next, consider the GEE for a parametric mean model with a design matrix Z_i which solves the estimating equation

$$\sum_{i=1}^{n} Z_i^T V_i^{-1}(Y_i - Z_i \eta) = 0,$$

where V_i is a working covariance matrix of Y_i not necessary equal to the true covariance Σ_i. Although no likelihood is assumed for the GEE-based approaches, the estimating equation can be treated as the score equation for mean parameters from a partly exponential model (Zhao, Prentice and Self 1992). For a model with a nonparametric mean function, adding a roughness penalty to a partly exponential model and taking the partial derivative with respect to the basis coefficients for the mean function motivates the penalized spline GEE,

$$\sum_{i=1}^{n} B_i^T V_i^{-1}(Y_i - B_i \theta) - \lambda D_q \theta = 0,$$

where again V_i is a working covariance matrix. When ignoring the penalty term, the penalized spline GEE reduces to a regular parametric GEE. The solution is

$$\hat{\theta}_\lambda = \left(\sum_{i=1}^{n} B_i^T V_i^{-1} B_i + \lambda D_q \right)^{-1} \sum_{i=1}^{n} B_i^T V_i^{-1} Y_i, \quad (2)$$

and the sandwich covariance formula for $\hat{\theta}_\lambda$ is

$$\mathrm{cov}(\hat{\theta}_\lambda) = H_{n,\lambda}^{-1} M_n H_{n,\lambda}^{-1}, \quad (3)$$

where $H_{n,\lambda} = \sum_{i=1}^{n} B_i^T V_i^{-1} B_i + \lambda D_q$, and $M_n = \sum_{i=1}^{n} B_i^T V_i^{-1}(Y_i - B_i \theta)(Y_i - B_i \theta)^T V_i^{-1} B_i$.

The sandwich variance for $\hat{f}(x)$ is

$$\mathrm{var}[\hat{f}(x)] = B^T(x)\mathrm{cov}(\hat{\theta}_\lambda)B(x).$$

Let τ index a finite dimensional parameter vector for V_i and let $\hat{V}_i = V_i(\hat{\tau})$. The variance is then estimated by $\hat{H}_{n,\lambda} = \sum_{i=1}^{n} B_i^T \hat{V}_i^{-1} B_i + \lambda D_q$ and $\widehat{M}_n = \sum_{i=1}^{n} B_i^T \hat{V}_i^{-1}(Y_i - B_i \hat{\theta}_0)(Y_i - B_i \hat{\theta}_0)^T \hat{V}_i^{-1} B_i$ in (3), where $\hat{\theta}_0$ is an initial regression spline estimator.

Note that this new variance estimator (3) differs from the usual model-based estimator in (1). It shares the robustness property as the sandwich variance estimator for the parametric marginal regressions: it remains consistent even if the correlation structure is misspecified.

2.2 Single-level generalized functional data

Again, first consider a nonparametric model

$$E(Y_{ij}|X_{ij})=\mu_{ij}, \quad g(\mu_{ij})=f(X_{ij}),$$

where $g(\cdot)$ is a known link function and $f(\cdot)$ is an unspecified smooth function. Let $\mu(\cdot) = g^{-1}(\cdot)$ denote the inverse of the link function, and with a little abuse of notation, let $\mu(B_i\theta)=[\mu(B_{i1}^T\theta),\cdots,\mu(B_{in_i}^T\theta)]^T$. The penalized spline GEE for generalized outcomes is then

$$\sum_{i=1}^{n}D_i^T(\theta)[V_i(\theta)]^{-1}[Y_i-\mu(B_i\theta)]-\lambda D_q\theta=0, \quad (4)$$

where $D_i(\theta)=\dfrac{\partial\mu(B_i\theta)}{\partial\theta}$, $V_i(\theta)=A_i^{1/2}(\theta)R_i(\tau)A_i^{1/2}(\theta)$, $A_i(\theta) = \text{diag}[\text{var}(Y_{i1}), \cdots, \text{var}(Y_{in_i})]$, and $R_i(\tau)$ is a working correlation matrix. Similar to the continuous outcome model, the sandwich covariance estimator for $\hat{\theta_\lambda}$ takes the same form as (3) with

$$H_{n,\lambda}(\theta)=\sum_{i=1}^{n}B_i^TA_i(\theta)[V_i(\theta)]^{-1}A_i(\theta)B_i+\lambda D_q \quad \text{and}$$

$$M_n(\theta)=\sum_{i=1}^{n}B_i^TA_i(\theta)[V_i(\theta)]^{-1}[Y_i-\mu(B_i\theta)][Y_i-\mu(B_i\theta)]^T[V_i(\theta)]^{-1}A_i(\theta)B_i,$$

which can be estimated by replacing θ with $\hat{\theta_\lambda}$ in the above expressions.

The estimating equation in (4) and the variance estimator are different from the likelihood based conditional approaches. The resulting fitted function and parameters also have different interpretations (population average effects) than the ones obtained from a conditional models (subject-specific effects).

2.3 Multilevel functional data

For multilevel functional data, let $Y_{ij}(t_{ijk})$ denote the measurement on the ith subject during the jth cycle at the kth time point, where $i = 1, \cdots, n, j = 1, \cdots, n_i$ and $k = 1, \cdots, n_{ij}$. The marginal methods presented in previous sections can be applied under the working assumption that all measurements on the ith subject are independent. However, a good choice of working covariance matrix may improve estimation efficiency. To obtain a reasonable working covariance, we present a two-way functional analysis of variance (ANOVA) working model as,

$$Y_{ij}(t_{ijk})=\mu(t_{ijk})+\eta_j(t_{ijk})+\xi_i(t_{ijk})+\gamma_{ij}(t_{ijk})+\varepsilon_{ijk}, \quad (5)$$

where $\mu(t)$ is the grand mean function, $\eta_j(t)$ is the deviation of the jth cycle from the grand mean, or the cycle effect, $\xi_i(t)$ is the subject-specific deviation from the cycle-specific mean

function (or the subject effect), $\gamma_{ij}(t)$ is the interaction effect, and $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$ are the residual measurement errors.

Using the spline basis expansion, we have

$$\mu(t) \approx B^T(t)\mu, \ \eta_j(t) \approx B^T(t)\eta_j, \ \xi_i(t) \approx B^T(t)\alpha_i, \ \gamma_{ij}(t) \approx B^T(t)\gamma_{ij},$$

where μ, η_j, α_i and γ_{ij} are basis coefficients. Let $Y_{ij} = [Y_{ij}(t_{ij1}), \cdots, Y_{ij}(t_{ijn_{ij}})]^T$ and $B_{ij} = [B_{ij}(t_{ij1}), \cdots, B_{ij}(t_{ijn_{ij}})]^T$. Then a working model using regression splines can be expressed as

$$Y_{ij} = B_{ij}\mu + B_{ij}\eta_j + B_{ij}\alpha_i + B_{ij}\gamma_{ij} + \varepsilon_{ij},$$
$$\eta_j \sim N(0, \Theta), \quad \alpha_i \sim N(0, \Lambda), \quad \gamma_{ij} \sim N(0, \Gamma), \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2 I_{n_{ij}}). \quad (6)$$

Under the model (6), a working covariance matrix is computed to improve estimation efficiency. Other working covariance can also be used. The parameters are obtained by restricted maximum likelihood estimation under the working regression spline model (6) (see for example, Rice and Wu 2001; Wu and Zhang 2006, Chapter 5.4). We do not assume the covariance structure to be correctly specified and will use the robust sandwich formula to compute the standard error of the mean function. For generalized outcomes, a similar functional ANOVA model can be defined.

3 Asymptotic properties

In the online appendix, we examine the asymptotics of the penalized spline estimator in a marginal model for correlated continuous data. We show that similar to independent data case (e.g., Claeskens et al. 2008), the aymptotics for correlated data falls under a small knots and a large knots scenario depending on the rate of increase of the number of knots. The small knots scenario is close to regression spline, i.e., the optimal rate of mean squared error (MSE) attained by the penalized spline estimator is similar to a regression spline estimator shown in Zhu et al. (2008). In this case, the shrinkage bias becomes negligible when smoothing parameter $\lambda = O(n^\gamma)$ is small, that is when $\gamma \quad (p + 2 - q) = (2p + 3)$. Therefore the asymptotic MSE is dominated by the squared approximation bias and asymptotic variance. The large knots is close to smoothing spline, i.e., the optimal rate of MSE attained by the penalized spline estimator is similar to a smoothing spline estimator shown in Lin et al. (2004). In this case, the approximation bias becomes negligible when the number of knots $K = O(n^\nu)$ is large, that is when $\nu \geq \frac{q}{(2q+1)(p+1)}$. Therefore, the asymptotic MSE is dominated by the squared shrinkage bias and asymptotic variance. This property is useful for developing methods to choose smoothing parameter.

We show the rate of convergence of the asymptotic bias, variance, and mean squared error in the appendix. In the small knots scenario, since the shrinkage bias is negligible, the asymptotic bias does not depend on the choice of working covariance matrix or the design density. The asymptotic variance is minimized when the true covariance is used, and therefore the asymptotic MSE is minimized when the working covariance is chosen as the true covariance. In the large knots scenario, the shrinkage bias is not negligible and the asymptotic bias depends on the working covariance matrix, the true covariance matrix, and the design density. Therefore the penalized spline estimator is not "design-adaptive" in the sense of Fan (1992). When the smoothing parameter λ converges to infinity (or when λ/n converges to zero) at a certain rate, we show in the appendix that the asymptotic variance is

minimized when the true covariance is used, which is similar to that reported in Welsh et al. (2002). Lastly, we prove a corollary on the asymptotic normality of the fitted mean function.

4 Selection of the smoothing parameter

For penalized spline smoothing, there are two tuning parameters to be determined: the number of knots of the spline basis and the smoothing parameter. Both empirical and theoretical work have suggested that when the number of knots is sufficiently large, increasing it further does not guarantee improvement in the quality of fit (Ruppert 2002; Li and Ruppert 2008). With a sufficiently large number of knots, the choice of smoothing parameter is critical for satisfactory performance. Popular methods to choose smoothing parameter include information criterion based approaches such as AIC and BIC, cross-validation (CV), generalized cross-validation (GCV, Craven and Wahba 1979), generalized maximum likelihood (GML, Wahba 1985), and restricted maximum likelihood (REML, Wand 2003) where the smoothing parameter is estimated as a ratio of two variance components. Opsomer et al. (2001) compared various methods for choosing smoothing parameter with correlated data and found that GCV may tend to under-smooth data. For marginal models, no likelihood is specified; thus an AIC, BIC or REML-based smoothing parameter is not available.

Here we assume a sufficient number of knots is used and propose a new method to select the smoothing parameter by minimizing an estimate of the asymptotic average mean squared error. The asymptotic analysis in section 3 reveals that the bias is decomposed as the sum of the approximation bias and the shrinkage bias. Since the approximation bias does not depend on λ, we propose to select the smoothing parameter by minimizing an estimate of the asymptotic MSE as the sum of the squared shrinkage bias and the asymptotic variance. To be specific, we choose the smoothing parameter by

$$\min_{\lambda} \left\{ \widehat{MSE}(\lambda) \right\},$$

where

$$\widehat{MSE}(\lambda) = \left(\frac{1}{M} \sum_{j=1}^{M} \left\{ \widehat{b}_{\lambda}^2(x_j, \widehat{V}) + \widehat{\mathrm{var}}[\widehat{f}(x_j)] \right\} \right), \quad (7)$$

and $x_j, j = 1, \cdots, M$, belong to a grid set covering the range of X_i. Note that the shrinkage bias is the difference between the bias of the penalized spline estimator and the approximation bias, or the bias due to the shrinkage effect. It can be estimated by the difference between a regression spline estimator and a penalized spline estimator through nonparametric bootstrap. Specifically, with a given λ and a given x, for each bootstrap copy of data we obtain a penalized spline estimator, $\widehat{f}_{\lambda}^{(b)}(x)$, and a regression spline estimator $\widehat{f}_{\mathrm{reg}}^{(b)}(x)$. We repeat this procedure B times, where B is large, and estimate the squared shrinkage bias by

$$\widehat{b}_{\lambda}^2(x, \widehat{V}) = \frac{1}{B} \sum_{b=1}^{B} [\widehat{f}_{\lambda}^{(b)}(x) - \widehat{f}_{\mathrm{reg}}^{(b)}(x)]^2.$$

We compare the proposed MSE-based choice of smoothing parameter with other existing alternatives, such as CV or GCV, in simulation studies.

5 Simulation studies

To study the performance of the proposed approaches, we conduct five simulation studies. The first two studies investigate the proposed methods for single-level functional data and the next two studies assess methods for multilevel data. The last study investigates the sensitivity of the sandwich variance to the choice of tuning parameters. In each case, we carried out 500 simulation runs. For penalized spline estimators, we used a truncated quadratic polynomial base with 20 knots.

5.1 Scenario I: Single-level functional data

Study I: Continuous outcome—The continuous outcomes are generated from the model

$$Y_{ij} = f(X_{ij}) + \varepsilon_{ij}, \quad i = 1, \cdots, n, \quad j = 1, \cdots, m, \quad (8)$$

with $n = 200$ and $m = 3$. The covariates X_{ij} are independently generated from a uniform distribution, $U(0, 1)$. The random errors are generated from a multivariate normal, uniform or Laplace distribution with compound symmetry correlation and $\rho = 0.2$. The true underlying function $f(x)$ is $\log(x)$, $2\exp(x)$, or $2\sin(2\pi x)$.

We compare the proposed P-spline approach with a regression spline approach (R-spline) where the number of knots is chosen by leave-ten-subjects-out cross validation. For the P-spline estimator, we compare two methods for choosing the smoothing parameter: the proposed MSE-based and the GCV. The GCV for correlated continuous data minimizes

$$\mathrm{GCV}(\lambda) = \frac{\sum_{ij}(\tilde{Y}_{ij} - \tilde{B}_{ij}^T \hat{\beta}_\lambda)^2}{\{1 - \frac{1}{N}\mathrm{trace}[H_n^{-1}(\hat{\beta}_\lambda)G_n]\}^2},$$

where $\tilde{Y}_i = \widehat{\sum}_0^{-1/2} Y_i$, $\tilde{B}_i = \widehat{\sum}_0^{-1/2} B_i$, $G_n = \sum_i \tilde{B}_i^T \tilde{B}_i$, and $\hat{\Sigma}_0$ is estimated based on an initial regression spline estimator.

Table 1 summarizes the mean of average MSE, i.e., $\frac{1}{N}\sum_{ij}[\hat{f}(X_{ij}) - f(X_{ij})]^2$, over 500 simulation repetitions for all estimators. We see that in all scenarios, the P-spline with MSE-based smoothing parameter is more efficient than the other two approaches. The efficiency gain can be up to 18%. In several scenarios with non-normal random errors, the MSE-based P-spline estimator has 50% lower mean average MSE than the R-spline estimator. When the true underlying function is $2\sin(2\pi x)$, the P-spline with GCV to choose smoothing parameter is the least efficient, where its mean average MSE is about five times as large as the other approaches. A close inspection of our simulations suggest that in some cases, GCV tends to under-smooth correlated data, which is consistent with results reported in literature (Opsomer et al. 2001; Welsh et al. 2002).

In Table 2, we show the mean estimated pointwise standard error using the sandwich estimator under a compound symmetry or a working independent covariance structure. We compare the sandwich estimator with the empirical standard deviation and the model-based

standard error estimators. When the underlying covariance structure is correctly specified as compound symmetry, both the sandwich estimator and the model-based estimator are close to the empirical standard deviation of $\hat{f}(x)$. However, when assuming an incorrectly specified working independent covariance structure, the model-based standard error underestimates variability of $\hat{f}(x)$, while the sandwich estimator is still close to the empirical standard deviation. The $\hat{f}(x)$ fitted with a correctly specified compound symmetry covariance has a lower empirical variance than $\hat{f}(x)$ fitted with an incorrectly specified working independent covariance, indicating some efficiency gain in choosing an appropriate correlation structure. Similar results are obtained for other functions of $\hat{f}(x)$, which are not shown here.

Study II: Binary outcome—The binary outcomes are generated from the marginal model,

$$\text{logit}[\text{Pr}(Y_{ij}=1)]=f(X_{ij}), \quad i=1,\cdots,n, \quad j=1,\cdots,m, \quad (9)$$

where $n = 100$ or 30, $m = 5$, 20, or 80 and the within subject correlation is compound symmetry with $\rho = 0.2$. For $m = 5$, the data for each subject is sparse, while for $m = 20$ or 80, the data for each subject is dense. The covariates X_{ij} are independently generated from $U(0, 1)$. We examined three different functions $f(x) = \sin(2\pi x)$, $\exp(x) - 2$, and $2 - 16x + 30x^2 - 15x^3$. Since the standard GCV does not apply to correlated binary data, we compare the MSE-based smoothing parameter selection with leave-ten-subjects-out cross validation. Table 3 and 4 summarize the mean average MSE of $\hat{f}(x)$ and pointwise standard deviation. In all the cases, the P-spline with MSE-based smoothing parameter selection is more efficient than the other two approaches. The efficiency gain of P-spline (MSE) over P-spline (CV) or R-spline is up to 20%.

We assess performance of the standard error estimation with $f(x) = \sin(2\pi x)$ under a compound symmetry and a working independent covariance structure. The pointwise sandwich standard error estimator is close to the empirical standard deviation of $\hat{f}(x)$ under both correlation structures. The results for the other two functions are similar and thus are not shown here. Again, when working independence is assumed, the model-based standard error is much smaller than the empirical standard deviation of $\hat{f}(x)$. Similar to Study I, using a correctly specified covariance structure improves estimation efficiency of $\hat{f}(x)$.

5.2 Scenario II: Multilevel functional data

Study I′: Continuous outcome—We generated the outcomes from a three-level partially linear model,

$$Y_{ijk}=f(X_{ijk})+Z_i\beta+\alpha_i+\eta_{ij}+\varepsilon_{ijk}, \\ i=1,\cdots,n, \quad j=1,\cdots,J, \quad k=1,\cdots,m, \quad (10)$$

where $n = 30$, $J = 5$, $m = 10$, 20, or 80, and $\alpha_i \sim N(0, 1)$ are subject-level random effects, and $\eta_{ij} \sim N(0, 1)$ are subject-specific and cycle-level random effects. Here for $m = 10$, the data in each cycle is sparse, while for $m = 20$ or 80, the data in each cycle is dense. Note that for continuous data, $f(\cdot)$ and β in model (10) are marginal means and the random effects are merely used to simulate correlation among outcomes. The covariates X_{ijk} are independently generated from $U(0, 1)$, and the measurement errors ε_{ijk} are independently generated from $N(0, 1)$. The subject-level covariates Z_i are i.i.d. and follow $N(0, 1)$ and the coefficient $\beta = 0.4$. We examined two different functions, $f(x) = 2\sin(2\pi x)$ and $f(x) = 2 - 16x + 30x^2 - 15x^3$, and three working correlation structures: assuming all observations are independent, assuming observations from different cycles are independent (between-cycle independence),

and the true correlation structure (accounting for both between- and within-cycle correlation of the observations on the same subject). For the P-spline estimator, the proposed MSE method was used to select the smoothing parameter.

Tables 5 and 6 summarize the simulation results. In Table 5, we show the mean average MSE of the fitted nonparametric function and the standard error of the parametric estimate. In terms of the mean average MSE, using a correctly specified correlation structure yields the most efficient estimator, while accounting for the within-cycle correlation but ignoring the between-cycle correlation ranks the second. Using working independent covariance for all observations on a subject provides the least efficient estimator. Compared to the R-spline, the P-spline estimator has a smaller mean average MSE. For the estimation of the parametric part, all the approaches lead to estimators with small biases and similar variances. Table 6 shows the pointwise mean estimated standard error of $\hat{f}(x)$ under the sparse data cases (the results for the dense cases are similar, which are not shown here). For all the three correlation structures, the sandwich estimates are close to the corresponding empirical variances. However, properly accounting for correlation increases the efficiency of the estimator. When the correct correlation structure is used, the model-based pointwise standard error estimate is close to the empirical estimate as well. We see that the pointwise empirical standard deviation is slightly higher when using working independence covariance than using independent cycle or correctly specified correlation.

Study II′: Binary outcome—Here we generate binary outcomes from the following three-level model:

$$\text{logit}[\Pr(Y_{ijk}=1)]=f(X_{ijk})+Z_i\beta,$$
$$i=1,\cdots,n, \quad j=1,\cdots,J, \quad k=1,\cdots,m, \quad (11)$$

where the between-cycle correlation is 0.07 and within-cycle correlation is 0.3, $n = 50$, $J = 5$, and $m = 5$. The correlation at both levels assume exchangeable structure. The covariates X_{ijk} are independently generated from $U(0, 1)$. The subject level covariates Z_i are generated from $U(0, 1)$ with the coefficient $\beta = 0.2$. We examined two functions, $f(x) = \sin(2\pi x)$ and $\exp(x) - 2$. We compare the estimator obtained assuming working independence of all observations on a subject to the one assuming between cycle independence. For the P-spline estimators, the proposed MSE-based method is used to select the smoothing parameter.

The simulation results are shown in Tables 7 and 8. Table 7 summarizes the mean average MSE of the nonparametric component in (11) and the standard error of the parametric component. The results are analogous to those in study I′ for the continuous outcome. In general, properly accounting for the correlation leads to a smaller mean average MSE estimate. For the parametric coefficient, all the approaches result in estimators with small biases and similar variances. For the nonparametric function, the P-spline estimators are more efficient than the R-spline estimator. Table 8 summarizes the mean pointwise standard error estimate of the fitted nonparametric function. We observe the sandwich variance estimates to be close to the corresponding empirical variance estimates using either correlation structure.

5.3 Sensitivity analysis of the sandwich variance estimator

In this section, we conducted simulation studies to investigate whether the sandwich variance estimator is sensitive to the choice of the tuning parameter λ. The simulation settings are similar to those in the Scenario I. We used the mean function $f(x) = 2\exp(x)$ with $n = 200$, $m = 10$ for continuous outcome and $n = 100$, $m = 10$ for binary outcome. We computed the sandwich variance estimator under ten equally spaced λ on the log-scale, that is, $\log_{10}\lambda = -5, -4, \cdots, 3, 4$.

The results based on 500 simulations are summarized in Figure A1 (online appendix). The top two panels show $\widehat{\mathrm{MSE}}(\lambda)$ as defined in (7) using different smoothing parameters for the continuous (left) and binary (right) outcomes, respectively. From the MSE, we can define a proper range of smoothing parameter as λ's such that no significant improvement in the $\widehat{\mathrm{MSE}}(\lambda)$ is observed using the 1-standard-deviation rule. In this case, the proper range of λ is $[1, 10^4]$ for both types of outcomes, which is also reflected in the top two subfigures. The bottom two panels show the estimated sandwich variance under these values of the smoothing parameter in the proper and improper range. We observe that the sandwich variance is non-increasing in λ at any given time point, which is expected: the smaller the smoothing parameter, the less penalty is placed on the roughness and the fitted curve is more wiggly, leading to a smaller bias but larger variance. If the tuning parameter λ is in a proper range defined by the MSE, the variation of the sandwich variance is ignorable or minor. In conclusion, the estimated sandwich variance is not sensitive to the choice of λ, given that the λ is in a proper range defined by the estimated MSE.

6 Data analysis

Per clinical protocol in the SAH study, Nimodipine was administered orally every 4 hours (Choi et al. 2012). Each patient received a dose of 30 mg (low dose) or 60 mg (high dose). Patients underwent multiple treatment cycles and their physiologic outcomes, such as MAP and brain oxygenation, during each treatment cycle were recorded. The dose level does not change within a treatment cycle of the same patient, but can change from cycle to cycle depending on the patient's clinical profile. The oxygen reactivity index (ORX) was calculated post-hoc as the running Pearson correlation coefficient between the brain tissue oxygenation and cerebral perfusion pressure, which takes a value between –1 and 1. The ORX is an index of cerebral autoregulation, a reflection of the cerebral vasculatures ability to control blood flow to the brain, independent of the systemic blood pressure. Higher ORX values indicate a higher risk of poor outcome after acute brain injury (Jaeger et al. 2006).

Physiologic variables were measured continuously every 5 seconds from General Electric (GE) Solar 8000i monitors and averages over each 10-minute interval were reported (Choi et al. 2012). Patients were monitored for 90 minutes before each dose, making for 9 measurements, and 120 minutes after the dose, making for 12 measurements. Including the time of administration, each cycle had a total of 22 equally spaced measurements. We observed 562 treatment cycles, among which 30 mg Nimodipine was given in 279 cycles and 60 mg Nimodipine was given in 283 cycles in a total of sixteen patients. On average, each patient has around 35 treatment cycles. The total number of observations is 11482. Among the patients 62.5% are female and their mean age is 50.

In Figure 1, we show scatter plots of all observed MAP for three subjects with 76, 29, and 45 treatment cycles, and we superimpose a local polynomial smoothing line using data from each subject. We observe considerable between- and within-subject variability. While the data on each subject is dense, some aggregation and modeling is needed to reveal the general trend in a treatment cycle. In the lower right panel of Figure 1, we show the sample mean MAP obtained by averaging measurements at the same time point across subjects and treatment cycles. We observe a larger effect on mean MAP from high dose group. In Figure 2, we show snap shots of a subject's MAP during several treatment cycles and superimpose a scatterplot smoothing line for each individual figure.

The primary goal of the study is to estimate the effect of Nimodipine on various physiologic outcomes in patients with SAH averaged across treatment cycles and across patients. The research interest is the outcome trajectory during each treatment cycle where the cycle-level data contains 22 equally spaced measurements in a natural time order nested within each

subject. Here the three levels of analysis units are: subjects, cycles, and average physiological outcomes (e.g., MAP) in each 10-minute interval. The correlation between measurements taken at different cycles on a subject and repeated measurements within a cycle may be difficult to model. Such correlation is not of scientific interest but needs to be accounted for. Hence, the marginal approach focusing on average effect with a robust standard error estimate is the preferred analysis. For the continuous outcome of interest, MAP, we fitted the marginal model under two working covariance structures: (1) assuming independence between cycles and exchangeable correlation within cycles; and (2) the two-way ANOVA in (5) accounting for both levels of correlation. The marginal mean is specified with a varying coefficient model,

$$E(Y_{ijk}|X_{ijk}, W_{ij}, Z_{ij})=f(X_{ijk})+\beta(X_{ijk})W_{ij}+Z_i^T\gamma, \quad (12)$$

where X_{ijk} is the time in a treatment cycle centered at the point of administering Nimodipine, W_{ij} is indicator of being on the higher dose, Z_i is a vector of baseline covariates including age and gender, $f(\cdot)$ is the MAP for the lower dose cycle, and $\beta(\cdot)$ is the difference in MAP between the two dose cycles.

In the upper left panel of Figure 3, we show the estimated mean MAP for each dose cycle obtained from model (12) assuming independence between cycles and exchangeable correlation within cycles along with its pointwise 95% confidence interval (CI). In the upper right panel, we present the estimated MAP assuming the working two-way ANOVA model in (5) along with its pointwise 95% CI. The range of the between-cycle correlation is from 0.21 to 0.46, while the within-cycle correlation ranges from 0.21 to 0.57. The smoothed estimates obtained from model (12) reflect a similar trend to the sample average (lower right panel of Figure 1). Using different working correlation structures gives similar point estimates. However, accounting for between-cycle correlation provides an estimator with a narrower CI than when the between-cycle correlation is ignored. As expected, Nimodipine has a larger effect on decreasing the MAP in high dose cycles than on the low dose cycles: the mean MAP in a high dose cycle decreases from 120.7 (95% CI: [119.1, 122.3]) to 116.4 (95% CI: [114.7, 118.0]), while in the low does cycles it decreases from 106.5 (95% CI: [105.4, 107.6]) to 105.1 (95% CI: [104.0, 106.2]). The analysis also suggests that the effect of Nimodipine in the high dose cycles lasts longer than the low dose cycles. We also fitted a conditional mixed effects model with both subject- and cycle-specific random intercepts and obtained similar estimated mean curves (results omitted).

In the bottom panels of Figure 3, we plot the estimated difference between MAP at a given time point (time t) post-medication and right before taking the medication (time zero) in both dose cycles together with the corresponding pointwise 95% CIs (dark gray) and simultaneous confidence bands (light gray). Constructing simultaneous confidence band for functional data is an important problem. Two key issues to maintain the nominal confidence level are to correct for bias in the nonparametric estimation and account for extra variability introduced by selecting the smoothing parameter. For single-level normal functional data, Krivobokova et al. (2010) proposed simultaneous confidence band based on Bayesian framework, frequentist framework, and an intermediate mixed model representation framework. For correlated functional data, Craineceanu et al. (2012) proposed bootstrap-based approaches. Here our simultaneous 95% confidence band was also computed using a bootstrap based approach similar to Crainiceanu et al. (2012), where we conditioned on subjects and resampled cycles within a subject. For each bootstrap sample, we applied the proposed marginal GEE model with an independent cycle working correlation. We compute the maximal standardized difference across the entire range of t in each bootstrap sample as

$d_b=\max_t\{|\hat{f}_b(t)-\hat{f}_b(0)-[\overline{f}(t)-\overline{f}(0)]|/\overline{\mathrm{sd}}(t)\}$ for the bth bootstrap sample. The

simultaneous CI is obtained as $\overline{f}(t) - \overline{f}(0) \pm q_{1-\alpha}\overline{\mathrm{sd}}(t)$, where $q_{1-\alpha}$ is the $(1-\alpha)$th empirical quantile of the maximal statistic d_b (Crainiceanu et al. 2012).

From Figure 3, we see that in the low dose cycles there is a slight dip in the mean MAP and it bounces back 50 minutes post-medication (lower left panel of Figure 3). In the high dose cycles, the mean MAP continues to decrease until about 90 minutes after administering the medication (lower right panel of Figure 3). From the simultaneous confidence bands in Figure 3, we see that Nimodipine significantly decreases the MAP in the high dose cycles over the entire course of a treatment cycle, while it only slightly decreases the MAP in the low dose cycles early on: in the low dose cycles, the pointwise CI shows a significant decrease during the first 50 minutes of receiving the medication, while the simultaneous confidence band barely shows a difference in this period, suggesting no effect after adjusting for multiple comparisons.

The other goal of the study is to estimate the effect of Nimodipine on cerebral autoregulation. Loss of autoregulation is defined as the oxygen reactivity index (ORX) greater than 0.2. Patients with a prolonged loss of cerebral autoregulation are at risk for worse outcomes. Let R_{ijk} be the at risk indicator for subject i in cycle j at time point k. Algorithm for fitting multilevel functional mixed effects model with a logit link and normal random effects failed to converge for this outcome. We fit the following marginal model to assess the risk of loss of autoregulation,

$$\mathrm{logit}[\Pr(R_{ijk}=1)] = f(X_{ijk}) + \beta(X_{ijk})W_{ij} + Z_i^T\gamma, \quad (13)$$

with a working covariance assuming exchangeable within-cycle correlation and independent between-cycle correlation on the same subject. The top two panels of Figure 4 show the estimated risk of cerebral autoregulation loss in the low and high dose cycles. For the low dose cycles, the probability of autoregulation loss increases slightly before the medication and continues to until 33 minutes post-medication, at which the maximal probability of 28% (95% CI [0.25, 0.31]) is attained. Then afterwards, the probability decreases to the minimal risk of 20% (95% [0.16, 0.25]) at the end of the treatment cycle. For the high dose cycles, the risk of cerebral autoregulation loss is always more than 30% and varies between 31% and 35.5%. Analogous to the MAP outcome, we plot the odds ratio of cerebral autoregulation loss at a time point t post-medication to right before taking the medication (time zero) in both dose cycles together with the corresponding pointwise 95% CIs (dark grey) and simultaneous confidence bands (light grey). The bottom two panels Figure 4 show that the estimated odds ratio is greater than one until about 65 minutes after administering Nimodipine in the low dose cycles. For the high does cycles, the estimated odds ratio stays above one until about 95 minutes after the administration. However, there is no significant difference in the odds ratio of post-medication risk comparing to right before medication between the two dosage groups for the entire treatment period. The simultaneous confidence bands show no significant effect in either group.

In summary, we found some evidence of Nimodipine reducing the mean MAP when administered at the 60 mg dose, but not at the 30 mg dose. Nimodipine does not appear to have a significant effect on cerebral autoregulation. These findings can be used to evaluate the safety concerns of Nimodipine and the recommendation of discontinuing the use of Nimodipine in SAH patients that is proposed in Diringer et al. (2011).

7 Discussion

The proposed marginal approach provides an effective alternative to analyze multilevel functional data when the population average effects are of interest. The robust sandwich variance estimator can be used for both conditional models and marginal models to protect against misspecification of correlation matrix, especially when the data has a complicated multilevel structure. Our investigation of the asymptotic properties reveals that for the small knots scenario, the asymptotic bias does not depend on the working correlation matrix and the estimated mean function is asymptotically efficient when the working correlation is correctly specified. For the large knots scenario, both the asymptotic bias and variance depend on the working correlation. A practical use of the asymptotic properties is to develop a new method to select the smoothing parameter in marginal approaches based on minimizing the asymptotic mean squared error. Without a likelihood framework, information criteria such as AIC or BIC are not applicable to choose the smoothing parameter. However, for logistic regression with random intercepts, under a bridge distribution (Wang and Louise 2003) the marginal model takes a logistic form; therefore, the regression parameters in a conditional model also has a marginal interpretation. Likelihood based inference can then be obtained under a conditional model and it may be possible to estimate the smoothing parameter from the likelihood using a bridge distribution for single level data.

Our methods can be applied to other marginal models such as an additive model,

$$g[E(Y_{ijk}|X_{ijk}, W_{ijk})] = f_1(X_{ijk}) + f_2(W_{ijk}),$$

where $f_1(\cdot)$ and $f_2(\cdot)$ are smooth functions. For the multilevel MAP data in our example, we used a two-way ANOVA to obtain a working covariance function. Other techniques, such as functional principal components can also be used to obtain an efficient working covariance function and the standard error will be calculated by the robust sandwich formula. Although consistency is guaranteed by the sandwich variance estimator, effective choice of covariance structure for multilevel binary data deserves further research.

Finally, some remarks on computation. The proposed methods can handle large scale data under a working independent covariance structure. For more complicated correlation structure, when the dimension of the cycle-level data increases, the methods face computational challenges such as matrix inversion. Computational manipulations such as singular value decomposition would be helpful in these cases. For the proposed MSE-based tuning parameter selection method, since it is based on bootstrap, it may not be applicable for large scale data. In addition, when there are multiple covariates fitted as smooth nonparameteric functions (for example, in an additive model), choosing smoothing parameters using an iterative cross validation procedure is computationally intensive. Choosing two smoothing parameters by cross validation is still feasible (Wang and Chen 2012), however, with more than two components the computation may be more challenging. Wood (2011) has proposed methods to choose multiple smoothing parameters for conditional models. Lastly, to appropriately account for the uncertainty introduced by the data-driven smoothing parameter choice especially for small to moderate sample size, some alternative approaches such as Bayesian methods can be used. It has been shown that Bayesain methods have improved coverage properties and are insensitive to the exact choice of smoothing parameter (e.g. see Marra and Wood 2012; Goldsmith, Greven, and Crainiceanu 2012). In practice, how to adjust for the variability and uncertainty introduced

by the data-driven smoothing parameters is an interesting and important topic that worth further investigating in the future..

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Apanasovich T, Ruppert D, Lupton J, Popovic N, Turner N, Chapkin R, Carrol R. Aberrant Crypt Foci and Semiparametric Modeling of Correlated Binary Data. Biometrics. 2008; 64:490–500. [PubMed: 17725810]

Baladandayuthapani V, Mallick B, Turner N, Hong M, Chapkin R, Lupton J, Carroll RJ. Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis. Biometrics. 2008; 64:64–73. [PubMed: 17608780]

Barrow DL, Smith PW. Asymptotic properties of best L_2[0,1] approximation by splines with variable knots. Quarterly of Applied Mathematics. 1978; 36:293–304.

Brumback B, Rice J. Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion). Journal of the American Statistical Association. 1998; 93:961–994.

Chen H, Wang Y. A penalized spline approach to functional mixed effects model analysis. Biometrics. 2011; 67:861–870. [PubMed: 21155747]

Choi HA, Ko SB, Chen H, Gilmore E, Carpenter AM, Lee D, Claassen J, Mayer SA, Schmidt JM, Lee K, Connelly ES, Paik M, Badjatia N. Acute Effects of Nimodipine on Cerebral Vasculature and Brain Metabolism in High Grade Subarachnoid Hemorrhage Patients. Neurocritical Care. 2012 Jun; 16(3):363–367. [PubMed: 22262041]

Claeskens G, Kivobokova T, Opsomer JD. Asymptotic properties of penalized spline estimators. Biometrika. 2009; 96(3):529–544.

Crainiceanu CM, Staicu AM, Di C. Generalized Multilevel Functional Regression. Journal of the American Statistical Association. 2009; 104(488):1550–1561. [PubMed: 20625442]

Crainiceanu CM, Staicu AM, Ray S, Punjabi N. Bootstrap-based inference on the difference in the means of two correlated functional processes. Statistics and Medicine. 2012; 31(26):3223–3240.

Craven, Wahba G. Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the methods of generalized cross-validation. Numerische Mathematik. 1979; 31:377–403.

Di C, Crainiceanu CM, Caffo BS, Punjabi NM. Multilevel functional principal component analysis. Annals of Applied Statistics. 2009; 3:458–488. [PubMed: 20221415]

Diggle, PJ.; Liang, KY.; Zeger, SL. Analysis of Longitudinal Data. 2. Oxford: Oxford University Press; 2002.

Diringer MN, Bleck TP, Claude HJ, Menon D, Shutter L, Vespa P, Bruder N, Connolly ES, Citerio G, Gress D, Hanggi D, Hoh BL, Lanzino G, Le RP, Rabinstein A, Schmutzhard E, Stocchetti N, Suarez JI, Treggiari M, Tseng MY, Vergouwen MD, Wolf S, Zipfel G. Neurocritical Care Society. Critical care management of patients following aneurysmal subarachnoid hemorrhage: recommendations from the Neurocritical Care Society's Multidisciplinary Consensus Conference. Neurocritical Care, 2011. 2011 Sep; 15(2):211–240.

Dorhout SM, Rinkel GJ, Feigin VL, Algra A, van den Bergh WM, Vermeulen M, van Gijn J. Calcium antagonists for aneurysmal subarachnoid haemorrhage. Cochrane Database System Review, 2007. 2007 Jul 18.3:CD000277.

Eilers P, Marx B. Flexible smoothing with B-splines. Statistical Science. 1996; 11:89–121.

Fan J. Design-adaptive nonparametric regression. Journal of the American Statistical Association. 1992; 87:998–1004.

Fu W. Penalized Estimating Equations. Biometrics. 2003; 59:126–132. [PubMed: 12762449]

Goldsmith J, Greven S, Crainiceanu C. Corrected confidence Bands for Functional Data Using Principal Components. Biometrics. 2012

Guo W. Functional mixed effects models. Biometrics. 2002; 58(1):121–128. [PubMed: 11890306]

Huang JZ, Zhang L, Zhou L. Efficient estimation in marginal partially linear models for longitudinal/clustered data using splines. Scandinavian Journal of Statistics. 2007; 34:451–477.

Jaeger M, Schuhmann MU, Soehle M, Meixensberger J. Continuous assessment of cerebrovascular autoregulation after traumatic brain injury using brain tissue oxygen pressure reactivity. Critical Care Medicine. 2006; 34(6):1783–1788. [PubMed: 16625135]

Krivobokova T, Kneib T, Claeskens G. Simultaneous confidence bands for penalized spline estimators. Journal of the American Statistical Association. 2010; 105:852–863.

Ibrahim A, Suliadi. GEE-smoothing spline in semiparametric model with correlated nominal Data: Estimation and simulation study. Latest trends on applied mathematics, simulation, modeling. 2010a

Ibrahim, A.; Suliadi. WSEAS Transactions on Systems and Control. 2010b. Analyzing Longitudinal Data Using Gee-Smoothing Spline.

Li Y, Ruppert D. On the asymptotics of penalized splines. Biometrika. 2008; 95:415–36.

Lin X, Carroll R. Nonparametric function estimation for clustered data when the predictor is measured without/with error. Journal of the American Statistical Association. 2000; 95:520–534.

Lin X, Wang N, Welsh A, Carroll R. Equivalent kernels of smoothing splines in nonparametric regression for clustered data. Biometrika. 2004; 92:177–193.

Marra G, Wood SN. Coverage properties of confidence intervals for generalized additive model components. Scandinavian Journal of Statistics. 2012; 39(1):53–74.

Opsomer JD, Wang Y, Yang Y. Nonparametric regression with correlated errors. Statistical Science. 2001; 16:134–153.

Rice J, Wu C. Nonparametric mixed effects models for unequally sampled noisy curves. Biometrics. 2001; 57:253–259. [PubMed: 11252607]

Ruppert D. Selecting the number of knots for penalized splines. JCGS. 2002; 11:735–757.

Ruppert, D.; Wand, MP.; Carroll, RJ. Semiparametric Regression. New York: Cambridge University Press; 2003.

Staicu A, Crainiceanu C, Carroll C. Fast methods for spatially correlated multilevel functional data. Biostatistics. 2010; 11 (2):177–194. [PubMed: 20089508]

Stiefel MF, Heuer GG, Abrahams JM, Bloom S, Smith MJ, Maloney-Wilensky E, Grady MS, LeRoux PD. The effect of nimodipine on cerebral oxygenation in patients with poor-grade subarachnoid hemorrhage. Journal of Neurosurgery. 2004; 101:594–599. [PubMed: 15481712]

Wahba G. A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. The Annals of Statistics. 1985; 4:1378–1402.

Wand MP. Smoothing and Mixed Models. Computational Statistics. 2003; 18:223–249.

Wang N. Marginal Nonparametric kernel regression accounting for within-subject correlation. Biometrika. 2003; 90:43–52.

Wang Y, Chen H. On testing a variance component in a linear mixed effects model with multiple variance components. Biometrics. 2012 In press.

Wang Z, Louise T. Matching Conditional and Marginal Shapes in Binary Random Intercept Models Using a Bridge Distribution Function. Biometrika. 2003; 90:765–775.

Welsh A, Lin X, Carroll R. Marginal longitudinal nonparametric regression: locality and efficiency of spline and kernel methods. Journal of the American Statistical Association. 2002; 97:482–493.

Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society Series B. 2011; 73:3–36.

Wu, H.; Zhang, J. Nonparametric Regression Methods for Longitudinal Data Analysis Mixed-Effects Modeling Approaches. Wiley; 2006.

Zhao LP, Prentice R, Self S. Multivariate Mean Parameter Estimation by using a Partly Exponential Model. Journal of the Royal Statistical Society, Series B. 1992; 54(3):805–811.

Zhou L, Huang J, Carroll R. Joint modeling of paired sparse functional data using principal components. Biometrika. 2008; 95:601–619. [PubMed: 19396364]

Zhu Z, Fung WK, He X. On the asymptotics of marginal regression splines with longitudinal data. Biometrika. 2008; 95:907–917.
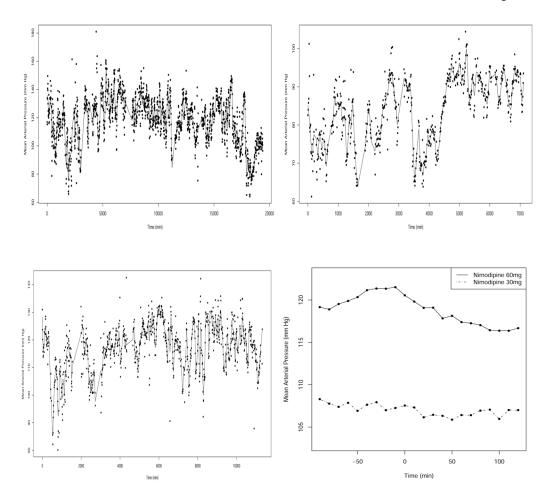
Figure 1.
Scatter plots of MAP versus continuous time for three subjects (upper and lower left). Dots: observed MAP; Solid line: local polynomial smoothing using observed MAP. Sample average at each time point in a cycle (lower right) averaged across subjects and cycles.
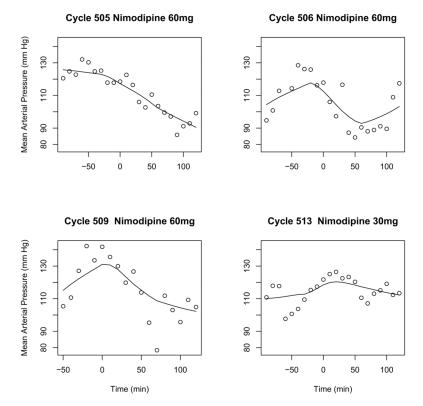
Figure 2.
Scatter plot of the MAP versus time measured on a subject during four treatment cycles.
Dots: observed MAP; Solid line: local polynomial smoothing using observed MAP in a
cycle. Time is centered at administration of Nimodipine, so that negative values are before
using the medication and positive values are after using the medication.
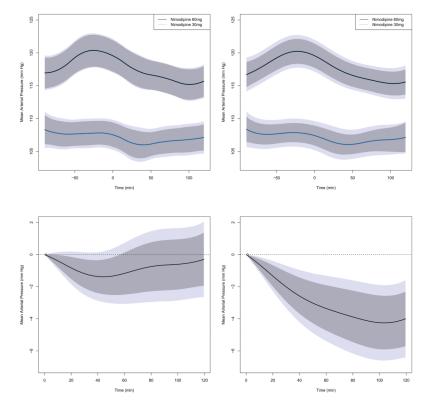
Figure 3.
Top two panels: Estimated effect of Nimodipine on MAP and its 95% pointwise confidence band. Left: assuming independence between cycles. Right: using a working two-way ANOVA model based correlation structure. Bottom two panels: Estimated differences $\hat{\mu}(t) - \hat{\mu}(0)$ and $\hat{f}(t) - \hat{f}(0)$ for MAP, with $\hat{\mu}(t) = \hat{f}(t) + \hat{\beta}(t)$, in low dose group (left) and high dose group (right). Dark grey shade represents the 95% pointwise confidence bands and light grey shade represents 95% simultaneous band based on bootstrap.
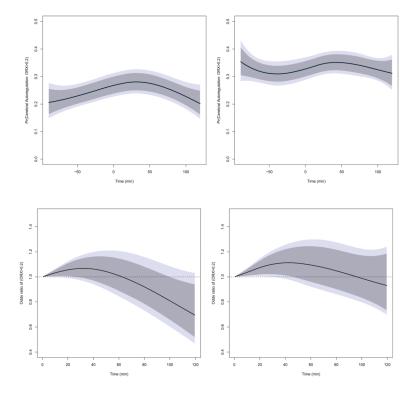
Figure 4.
Top two panels: Estimated effect of Nimodipine on at risk of cerebral autoregulation loss in the low dose group (left) and the high dose group (right). Bottom two panels: Estimated odds ratio $\exp[\hat{\mu}(t)]/\exp[\hat{\mu}(0)]$ and $\exp[\hat{f}(t)]/\exp[\hat{f}(0)]$ of at risk for autoregulation loss, with $\hat{\mu}(t) = \hat{f}(t) + \hat{\beta}(t)$, in low dose group (left) and high dose group (right). Dark grey shade represents the 95% pointwise confidence bands and light grey shade represents 95% simultaneous band based on bootstrap.

Table 1

Mean average MSE of $\hat{f(x)}$ using various smoothing techniques and smoothing parameter selectors, continuous outcome, $n = 200$, $m = 3$, 500 simulations.

$f(x)$	Error dist.	P-spline (MSE)	P-spline (GCV)	R-spline
log(x)	N(0,1)	0.015	0.023	0.018
log(x)	U(−3,3)	0.044	0.055	0.052
2 exp(x)	N(0,1)	0.007	0.007	0.014
2 exp(x)	Laplace(0,1)	0.021	0.021	0.041
2 sin($2\pi x$)	N(0,1)	0.011	0.065	0.013
2 sin($2\pi x$)	Laplace(0,1)	0.021	0.106	0.027

Table 2

Pointwise standard deviation, continuous outcome, $f(x) = 2\sin(2\pi x)$, compound symmetry correlation ($\rho = 0.2$), normal random error, $n = 200$, $m = 3$, 500 simulations.

x		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
CS	Empirical	0.11	0.10	0.10	0.10	0.096	0.099	0.10	0.10	0.11
	Model-based*	0.11	0.10	0.10	0.10	0.097	0.097	0.099	0.11	0.10
	Sandwich	0.11	0.10	0.10	0.10	0.097	0.097	0.099	0.11	0.11
WI	Empirical	0.12	0.12	0.12	0.12	0.10	0.12	0.12	0.11	0.12
	Model-based**	0.099	0.096	0.093	0.093	0.092	0.091	0.093	0.094	0.098
	Sandwich	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.12

* : Under correctly specified compound symmetry correlation

** : Under incorrectly specified working independence correlation

Table 3

Mean average MSE of $\hat{f(x)}$ using various smoothing techniques and smoothing parameter selection, binary outcome, 500 simulations.

	f(x)	**P-spline (MSE)**	**P-spline (CV)**	**R-spline**
$n = 100, m = 5$	$\sin(2\pi x)$	0.059	0.064	0.063
$n = 100, m = 5$	$2 - 16x + 30x^2 - 15x^3$	0.060	0.066	0.065
$n = 100, m = 5$	$\exp(x) - 2$	0.047	0.057	0.058
$n = 100, m = 20$	$\exp(x) - 2$	0.026	0.028	0.028
$n = 30, m = 80$	$\exp(x) - 2$	0.074	0.075	0.076

Table 4

Pointwise standard deviation with binary outcome, exchangeable correlation ($\rho = 0.2$), $f(x) = \sin(2\pi x)$, $n = 100$ or 200, $m = 5$, 500 simulations.

$n = 100$	x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
CS	Empirical	0.27	0.25	0.24	0.24	0.23	0.22	0.22	0.24	0.24
	Model-based*	0.25	0.24	0.23	0.22	0.21	0.21	0.22	0.23	0.23
	Sandwich	0.25	0.23	0.23	0.22	0.21	0.21	0.22	0.23	0.24
WI	Empirical	0.28	0.26	0.25	0.25	0.23	0.23	0.23	0.24	0.24
	Model-based**	0.23	0.21	0.20	0.19	0.18	0.18	0.20	0.21	0.21
	Sandwich	0.26	0.24	0.23	0.22	0.21	0.21	0.23	0.24	0.24
$n = 200$	x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
CS	Empirical	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17
	Model-based*	0.18	0.17	0.17	0.16	0.16	0.16	0.16	0.17	0.17
	Sandwich	0.18	0.17	0.17	0.16	0.16	0.16	0.16	0.17	0.17
WI	Empirical	0.17	0.16	0.17	0.17	0.16	0.16	0.17	0.17	0.17
	Model-based**	0.16	0.15	0.15	0.14	0.14	0.14	0.14	0.15	0.16
	Sandwich	0.18	0.17	0.17	0.16	0.16	0.16	0.16	0.17	0.18

*: Under correctly specified compound symmetry correlation

**: Under incorrectly specified working independence correlation

Table 5

Mean average MSE of $f(\hat{x})$ and SE of $\hat{\beta}$ using different correlation structures, continuous outcome, multilevel model, $n = 30$, $J = 5$, $m = 10$, 20 or 80, 500 simulations.

	R-spline	P-spline (WI)	P-spline (Ind cycles)	P-spline (True)
$n = 30$, $J = 5$, $m = 10$ $f(x) = 2 - 16x + 30x^2 - 15x^3$				
AMSE[$\hat{f(\cdot)}$]	0.044	0.042	0.041	0.040
mean $\hat{\beta}$	0.401	0.401	0.401	0.401
mean $\widehat{SE}(\hat{\beta})$	0.243	0.243	0.243	0.242
$n = 30$, $J = 5$, $m = 10$ $f(x) = 2\sin(2\pi x)$				
AMSE[$\hat{f(\cdot)}$]	0.045	0.047	0.044	0.043
mean $\hat{\beta}$	0.395	0.394	0.394	0.395
mean $\widehat{SE}(\hat{\beta})$	0.221	0.221	0.221	0.221
$n = 30$, $J = 5$, $m = 20$ $f(x) = 2\sin(2\pi x)$				
AMSE[$\hat{f(\cdot)}$]	0.044	0.046	0.044	0.043
mean $\hat{\beta}$	0.400	0.400	0.400	0.400
mean $\widehat{SE}(\hat{\beta})$	0.216	0.216	0.216	0.216
$n = 30$, $J = 5$, $m = 80$ $f(x) = 2\sin(2\pi x)$				
AMSE[$\hat{f(\cdot)}$]	0.041	0.041	0.040	0.040
mean $\hat{\beta}$	0.390	0.390	0.390	0.390
mean $\widehat{SE}(\hat{\beta})$	0.220	0.220	0.220	0.220

Table 6

Pointwise standard deviation, continuous outcome, multilevel model, normal random error, $n = 30$, $J = 5$, $m = 10$, 500 simulations.

$f(x) = 2 \sin(2\pi x)$									
x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Empirical (WI)	0.22	0.22	0.21	0.21	0.20	0.22	0.23	0.22	0.22
Sandwich (WI)	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21
Empirical (Ind cycles)	0.21	0.21	0.21	0.20	0.20	0.21	0.21	0.21	0.21
Sandwich (Ind cycles)	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21
Empirical (True)	0.21	0.21	0.21	0.20	0.20	0.21	0.21	0.21	0.21
Model-based (True)	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21
Sandwich (True)	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20

$f(x) = 2 - 16x + 30x^2 - 15x^3$									
x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Empirical (WI)	0.20	0.20	0.20	0.20	0.20	0.20	0.21	0.21	0.21
Sandwich (WI)	0.21	0.21	0.20	0.20	0.20	0.20	0.21	0.21	0.21
Empirical (Ind cycles)	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.21	0.20
Sandwich (Ind cycles)	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
Empirical (True)	0.20	0.20	0.19	0.19	0.20	0.20	0.20	0.20	0.20
Model-based (True)	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21
Sandwich (True)	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20

Table 7

Mean average MSE of $f(\hat{x})$ and SE of $\hat{\beta}$ using different correlation structures, binary outcome, multilevel model, $n = 50$, $J = 5$, $m = 5$, 500 simulations

	$f(x) = \sin(2\pi x)$		
	R-spline	P-spline (WI)	P-spline (Ind cycles)
AMSE[$f(\hat{\cdot})$]	0.077	0.075	0.074
mean $\hat{\beta}$	0.205	0.203	0.203
mean $\widehat{SE}(\hat{\beta})$	0.428	0.426	0.425
	$f(x) = \exp(x) - 2$		
	R-spline	P-spline (WI)	P-spline (Ind cycles)
AMSE[$f(\hat{\cdot})$]	0.075	0.073	0.071
mean $\hat{\beta}$	0.191	0.191	0.191
mean $\widehat{SE}(\hat{\beta})$	0.434	0.436	0.433

Table 8

Pointwise standard deviation, binary outcome, multilevel model, $n = 50$ or 100, $J = 5$, $m = 5$, 500 simulations.

$n = 50$				$f(x) = \sin(2\pi x)$				
x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Empirical (WI)	0.27	0.27	0.27	0.26	0.26	0.27	0.27	0.27	0.27
Sandwich (WI)	0.26	0.26	0.25	0.25	0.25	0.25	0.25	0.25	0.26
Empirical (Ind cycles)	0.27	0.27	0.27	0.26	0.26	0.26	0.26	0.27	0.27
Sandwich (Ind cycles)	0.25	0.25	0.25	0.24	0.24	0.24	0.25	0.25	0.25

$n = 50$				$f(x) = \exp(x) - 2$				
x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Empirical (WI)	0.26	0.26	0.25	0.25	0.26	0.26	0.26	0.26	0.26
Sandwich (WI)	0.25	0.25	0.24	0.24	0.24	0.24	0.24	0.24	0.25
Empirical (Ind cycles)	0.26	0.26	0.24	0.25	0.25	0.25	0.25	0.25	0.25
Sandwich (Ind cycles)	0.25	0.24	0.24	0.24	0.23	0.23	0.23	0.24	0.25

$n = 100$				$f(x) = \exp(x) - 2$				
x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Empirical (Ind cycles)	0.18	0.17	0.16	0.16	0.16	0.16	0.16	0.17	0.17
Sandwich (Ind cycles)	0.18	0.17	0.16	0.16	0.16	0.16	0.16	0.17	0.17
Empirical (WI)	0.18	0.17	0.16	0.16	0.16	0.16	0.17	0.17	0.17
Sandwich (WI)	0.18	0.17	0.17	0.16	0.16	0.16	0.16	0.17	0.17