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

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

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

[2]: *Department of Neurology, College of Physicians and Surgeons, Columbia University, New York, NY 10032, 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 recorded 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

*Correspondence author

(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 new 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 marginal penalized spline regression based on an estimate of the asymptotic mean squared error (MSE). Finally, we apply the methods to the SAH study to evaluate a recent debate on discontinuing the use of Nimodipine in the clinical community.

Key words: 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 10 minutes (Choi et al. 2011). 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 based methods to analyze nested samples of function-

al data. Guo (2002) proposed a functional mixed effects model with functional random effects fitted by a Kalman filtering. Zhou et al. (2008) proposed jointly modeling 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 approaches 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 current literature focus on conditional approaches through a functional mixed effects model or functional principal components analysis. In clinical trials, such as the SAH study described in section 1.1 (Choi et al. 2011), the goal is to estimate the population average effect or 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 (penalized spline GEE) 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. 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 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. 2011).

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 minutes. 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 natural multilevel 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,

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

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,

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

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:

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

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 $\widehat{f}(x) = B^T(x)\widehat{\theta}$ and its standard error is estimated from

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

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 $\widehat{f}(x)$.

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

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

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

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

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

$$\text{cov}(\widehat{\theta}_\lambda) = H_{n,\lambda}^{-1} M_n H_{n,\lambda}^{-1}, \tag{3}$$

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 $\widehat{f}(x)$ is

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

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

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

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 $\widehat{\theta}_\lambda$ takes the same form as (3) with

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

which can be estimated by replacing θ with $\widehat{\theta}_\lambda$ 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) from 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,

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

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

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

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

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

Under the model (6), a working covariance matrix is computed to improve estimation efficiency. Other working covariance can also be used. 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

For independent data, Claeskens et al. (2008) showed that the asymptotics of the penalized spline estimator fall into two categories, depending on the number of knots used. Zhu et al. (2008) studied asymptotics for regression spline estimator with correlated data. Here we examine the asymptotics of penalized spline estimator in a marginal model for correlated continuous data.

We assume $\text{cov}(Y_i) = \Sigma$, thus the covariance matrix does not vary across subjects. Let V denote a working covariance matrix of Y_i and let $B = (B_1^T, \cdots, B_n^T)^T$. Let $C^{p+1}([a, b])$ denote the space of all $p+1$ times continuously differentiable functions defined on $[a, b]$. Let $G = (g_{ij})$ and $V^{-1} = (v^{st})$ with

$$g_{ij} = \sum_{s \neq t}^{m} \int_a^b \int_a^b B_i(x) v^{st} B_j(y) \rho_{st}(x, y) dx dy + \sum_{s=1}^{m} \int_a^b B_i(x) v^{ss} B_j(x) \rho_s(x) dx,$$

where $Q_{n,jl}(x,y) = \frac{1}{n}\sum_{i=1}^{n} I(t_{ij} \le x, t_{i,l} \le y), Q_{n,j}(x) = \frac{1}{n}\sum_{i=1}^{n} I(t_{ij} \le x)$. Here, $Q_{jl}(x,y)$ and $Q_j(x)$ are certain distribution functions with positive continuous density functions $\rho_{jl}(x,y)$ and $\rho_j(x)$. Let $W = \{w_{ij}\} = V^{-1}\Sigma V^{-1}$ and $U = \{u_{ij}\}$, where

$$u_{ij} = \sum_{s\neq t}^{m} \int_a^b \int_a^b B_i(x) w_{st} B_j(y) \rho_{st}(x,y) dxdy + \sum_{s=1}^{m} \int_a^b B_i(x) w_{ss} B_j(x) \rho_s(x) dx.$$

Denote the approximation bias as $b_a(t, p+1)$. Zhu et al. (2008) showed that

$$b_a(t, p+1) = -\frac{\mu^{(p+1)}(t)}{(p+1)!} \sum_{i=0}^{K} I(\tau_i \le t < \tau_{i+1}) \delta_i^{p+1} B_{p+1}(\frac{t-\tau_i}{\delta_i}).$$

Denote the shrinkage bias as $b_\lambda(x, V) = -\frac{\lambda}{n} B^T(x)(G + \frac{\lambda}{n} D_q)^{-1} D_q \beta$, where $\beta = (B^T \boldsymbol{V}^{-1} B)^{-1} B^T \boldsymbol{V}^{-1} s_f / n$, $\boldsymbol{V} = \text{diag}(V, \cdots, V)$, and $s_f(\cdot) = B^T(\cdot)\beta$ is the best L_∞ approximation to the function $f(\cdot)$.

Define $K_q = \lambda K^{2q}/n$.

Theorem 1. *Assume that conditions A1 through A4 in the appendix hold.*

1. *If $K_q = o(1)$ and $f(\cdot) \in C^{p+1}[a,b]$, then the following statements hold*

$$E[\widehat{f}(x)] - f(x) = b_a(x, p+1) + b_\lambda(x, V) + o[K^{-(p+1)}] + o\left(\frac{\lambda}{n} K^q\right),$$
$$var[\widehat{f}(x)] = \frac{1}{n} B^T(x)(G + \frac{\lambda}{n} D_q)^{-1} U (G + \frac{\lambda}{n} D_q)^{-1} B(x) + o\left(\frac{K}{n}\right),$$
$$MSE[\widehat{f}(x)] = O\left(\frac{K}{n}\right) + O(K^{-2(p+1)}) + O\left(\frac{\lambda^2}{n^2} K^{2q}\right).$$

For $K = O(n^{\frac{1}{2p+3}})$ and $\lambda = O(n^\gamma)$ for $\gamma \le (p+2-q)/(2p+3)$, the optimal rate for mean squared error (MSE), $n^{-\frac{2p+2}{2p+3}}$, is attained by the penalized spline estimator.

2. *If $K_q = O(1)$ and $f(\cdot) \in C^{p+1}[a,b]$, the following statements hold*

$$E[\widehat{f}(x)] - f(x) = b_a(x, p+1) + b_\lambda(x, V) + o(K^{-q}) + o\left[\left(\frac{\lambda}{n}\right)^{\frac{1}{2}}\right],$$
$$var[\widehat{f}(x)] = \frac{1}{n} B^T(x)(G + \frac{\lambda}{n} D_q)^{-1} U (G + \frac{\lambda}{n} D_q)^{-1} B(x) + o\left[\frac{1}{n}\left(\frac{\lambda}{n}\right)^{-\frac{1}{2q}}\right],$$
$$MSE[\widehat{f}(x)] = O\left[\frac{1}{n}\left(\frac{\lambda}{n}\right)^{-\frac{1}{2q}}\right] + O\left(K^{-2q}\right) + O\left(\frac{\lambda}{n}\right).$$

For $\lambda = O(n^{\frac{1}{2q+1}})$ and $K = O(n^{\frac{1}{2q+1}})$, the optimal rate for MSE, $n^{-\frac{2q}{2q+1}}$, is attained by the penalized spline estimator.

Proof of this Theorem is in the technical appendix. We make the following remarks:

Remark 1. *The asymptotic scenario 1 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* $\gamma \leq (p+2-q)/(2p+3)$*. Therefore the asymptotic MSE is dominated by the squared approximation bias and asymptotic variance.*

Remark 2. *The asymptotic scenario 2 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.*

Remark 3. *In the asymptotic scenario 1, since the shrinkage bias is negligible, the asymptotic bias does not depend on the choice of working covariance matrix or the design density* $Q(x)$*. The asymptotic variance is minimized when the true covariance* $V = \Sigma$ *is used, therefore the asymptotic MSE is minimized when* $V = \Sigma$*.*

Remark 4. *In the asymptotic scenario 2, the shrinkage bias is not negligible and the asymptotic bias depends on the working covariance matrix, the true covariance matrix, and the design density* $Q(x)$*. Therefore the penalized spline estimator is not "design-adaptive" in the sense of Fan (1992). When* λ *converges to infinity at a certain rate, we show in the appendix that the asymptotic variance is minimized when* $V = \Sigma$*, which is similar to that reported in Welsh et al. (2002).*

In many cases, the working covariance matrix is estimated. Let τ index a finite dimensional parameter vector of the V and let $\widehat{V} = V(\widehat{\tau})$. Suppose τ can be estimated at a parametric rate, i.e., $\widehat{\tau} = \tau + o_p(n^{-1/2})$. The next theorem shows that the estimation of $\widehat{\tau}$ does not have any effect on the asymptotic distribution of $\widehat{f}(x)$.

Corollary 1. *Assume* $K_q = o(1)$ *and that there exists* $h > 0$, $C > 0$, *such that* $\sup_{i,j} E|\epsilon_{ij}|^{2+h} \leq C$, *where* $\epsilon_{ij} = Y_{ij} - f(X_{ij})$. *Then*

$$\frac{\widehat{f}(x) - f(x) - b_a(x, p+1) - b_\lambda(x, V)}{\sqrt{var[\widehat{f}(x)]}} \longrightarrow N(0,1)$$

in distribution, as $n \longrightarrow \infty$. *Furthermore, we have* $\widehat{var}[\widehat{f}(x)] = var[\widehat{f}(x)] + o_p(1)$. *Lastly, let* $\tilde{f}(x) = B^T(x)[\sum_{i=1}^n (B_i^T \widehat{V}^{-1} B_i + \lambda D_q)^{-1} \sum_{i=1}^n B_i^T \widehat{V}^{-1} Y_i]$, *then* $\tilde{f}(x) = \widehat{f}(x) + o_p(1)$.

Proof of the corollary 1 is in the appendix. Here, the normality addresses the small knots scenario.

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 λ by

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

where $x_j, j = 1, \cdots, M$, 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, at a given x, and for each bootstrap copy of data, we obtain a penalized spline estimator, $\widehat{f}_\lambda^{(b)}(x)$, and a regression spline estimator $\widehat{f}_{\text{reg}}^{(b)}(x)$. We repeat this procedure B times, where B is large, and estimate the squared shrinkage bias by

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

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 four simulation studies. The first two studies investigate the proposed methods for single-level functional data and the last two studies assess methods for multilevel data. In each case, we carried out 500 simulation runs. For penalized spline estimators, we used a truncated quadratic polynomial base with 20 knots.

Scenario I: Single-level functional data

Study I: Continuous outcome

The continuous outcomes are generated from the model

$$Y_{ij} = f(X_{ij}) + \epsilon_{ij}, \quad i = 1, \cdots, n, \quad j = 1, \cdots, m, \tag{7}$$

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

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

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

Table 1 summarizes the mean of average MSE, i.e., $\frac{1}{N}\sum_{ij}[\widehat{f}(X_{ij}) - f(X_{ij})]^2$, 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 data for 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 $\widehat{f}(x)$. However, when assuming an incorrectly specified working independent covariance structure, the model-based standard error underestimates variability of $\widehat{f}(x)$, while the sandwich estimator is still close to the empirical standard deviation. The $\widehat{f}(x)$ fitted with a correctly specified compound symmetry covariance has a lower empirical variance than $\widehat{f}(x)$ 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 $f(x)$, which are not shown here.

Study II: Binary outcome

The binary outcomes are generated from the marginal model,

$$\text{logit}[\Pr(Y_{ij} = 1)] = f(X_{ij}), \quad i = 1, \cdots, n, \quad j = 1, \cdots, m, \tag{8}$$

where $n = 100$, $m = 5$, and the within subject correlation is compound symmetry with $\rho = 0.2$. The covariates X_{ij} are independently generated from $U(0, 1)$. We use 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 $\widehat{f}(x)$ and pointwise standard deviation. In all three 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 correlation structure. The pointwise sandwich standard error estimator is close to the empirical standard deviation of $\widehat{f}(x)$ 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 $\widehat{f}(x)$. Similar to Study I, using a correctly specified covariance structure improves estimation efficiency of $\widehat{f}(x)$.

Scenario II: Multilevel functional data

Study I′: Continuous outcome

We generated the outcomes from a three-level partially linear model,

$$Y_{ijk} = f(X_{ijk}) + Z_i\beta + \alpha_i + \eta_{ij} + \epsilon_{ijk}, \tag{9}$$
$$i = 1, \cdots, n, \quad j = 1, \cdots, J, \quad k = 1, \cdots, m,$$

where $n = 30$, $J = 5$, $m = 10$, and $\alpha_i \sim N(0, 1)$ are subject-level random effects, and $\eta_{ij} \sim N(0, 1)$ are subject-specific cycle-level random effects. Note that for continuous data, $f(\cdot)$ and β in model (9) 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 used two different functions, $f(x) = 2\sin(2\pi x)$ and $f(x) = 2 - 16x + 30x^2 - 15x^3$. We examine 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 approaches, 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 and 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 $\widehat{f}(x)$. For all the three correlation structures, the sandwich estimates are close to the corresponding empirical variances. How-

ever, properly accounting for correlation increases the efficiency of the estimate. 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 is slightly higher than using independent cycle or correctly specified correlation.

Study II′: Binary outcome

Here we generate binary outcomes from the following three-level model:

$$\text{logit}[\Pr(Y_{ijk} = 1)] = f(X_{ijk}) + Z_i\beta, \tag{10}$$
$$i = 1, \cdots, n, \quad j = 1, \cdots, J, \quad k = 1, \cdots, m,$$

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 $\beta = 0.2$. Two functions, $f(x) = \sin(2\pi x)$ and $\exp(x) - 2$ are used. 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 estimate and the standard error of the parametric estimate. 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.

6 Data analysis

Per clinical protocol in the SAH study, Nimodipine was administered orally every 4 hours (Choi et al. 2011). 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 and averaged over 10 minutes. 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 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. The total number of observations is 11482. In Figure 1, we show the scatter plot of a subject's MAP during several treatment cycles. A scatterplot smoothing line was added to each individual figure.

The primary research 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 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,

$$E(Y_{ijk}|X_{ijk}, W_{ij}, Z_{ij}) = f(X_{ijk}) + \beta(X_{ijk})W_{ij} + Z_i^T\gamma, \tag{11}$$

where X_{ijk} is the time in a treatment cycle centered at the point of administering Nimodipine, W_{ij} is an 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 panel of Figure 2, we show the sample mean MAP obtained by averaging measurements at the same time point across subjects and treatment cycles. In the lower left panel of the same figure, we show the estimated mean MAP for each dose cycle obtained from model (11), assuming working independence of all observations on the same subject. The lower right panel presents the estimated MAP, assuming the working two-way ANOVA model in (5). The smoothed estimates obtained from model (11) reflect a similar trend to the sample average. Using different working correlation structures gives similar point estimates. However, accounting for between-cycle correlation provides an estimator with a narrower confidence band 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 results also suggest that the effect of Nimodipine in the high dose cycles lasts longer than the low dose cycles. In 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. In the low dose cycles, there is a slight dip in the mean MAP and it bounces back 50 minutes post-medication (left panel of Figure 3). In the high dose cycles, the mean MAP continues to decrease until about 90 minutes after administering the medication (right panel of Figure

3). From the pointwise confidence bands in Figure 3, we see that Nimodipine significantly decreases the MAP in the high dose cycles over the course of a treatment cycle, while it only slightly decreases the MAP in the low dose cycles.

The other goal of the study is to estimate the effect of Nimodipine on cerebral autoregulation. We defined loss of autoregulation 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. The multilevel functional mixed effects model with a logit link function and normal random effects failed to converge for this binary outcome. We fit the following marginal model to assess the risk of loss of autoregulation,

$$\text{logit}[\Pr(R_{ijk} = 1)] = f(X_{ijk}) + \beta(X_{ijk})W_{ij} + Z_i^T\gamma. \tag{12}$$

The working covariance assuming exchangeable within cycle correlation and independent between cycle correlation on the same subject is used. Figure 4 shows the estimated risk 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. Figure 5 shows 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.

In summary, we found some evidence of Nimodipine reducing the mean MAP when ad-

ministered 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 that 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.

Acknowledgements

Wang's research is supported by various NIH grants AG031113-01A2 and NS073670-01.

Supplementary Material

The supplementary material section contains the proof of Theorem 1 and Corollary 1.

References

Apanasovich T, Ruppert D, Lupton J, Popovic N, Turner N, Chapkin R, Carrol R. (2008). Aberrant Crypt Foci and Semiparametric Modeling of Correlated Binary Data. *Biometrics*, 64, 490-500.

Baladandayuthapani, V., Mallick, B., Turner, N., Hong, M., Chapkin, R., Lupton, J. and Carroll, R. J. (2008). Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis. *Biometrics*, 64, 64-73.

Barrow, D. L. and Smith, P. W. (1978). Asymptotic properties of best $L_2[0,1]$ approximation by splines with variable knots. *Quarterly of Applied Mathematics* **36**, 293–304.

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

Chen, H. and Wang, Y (2011). A penalized spline approach to functional mixed effects model analysis. *Biometrics*. 67, 861-870.

Choi, H.A., Ko, SB., Chen, H., Gilmore, E., Carpenter, A.M., Lee, D., Claassen, J., Mayer, S.A., Schmidt, J.M., Lee, K., Connelly, E.S., Paik, M., Badjatia, N. (2011). Acute Effects of Nimodipine on Cerebral Vasculature and Brain Metabolism in High Grade Subarachnoid Hemorrhage Patients. Submitted to *Neurocritical Care.*

Claeskens, G., Kivobokova, T. and Opsomer, J.D. (2009). Asymptotic properties of penalized spline estimators. *Biometrika* 96, 3, 529-544.

Crainiceanu, C., Staicu, A., Di, C. (2009), Generalized Multilevel Functional Regression technical report, *Journal of the American Statistical Association*, 104(488), 1550-1561.

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

Di, C., Crainiceanu, C.M., Caffo, B.S., and Punjabi, N.M. (2009). Multilevel functional principal component analysis. *Annals of Applied Statistics*, **3**, 458-488.

Diggle, P. J., Liang, K. Y., and Zeger, S. L. (2002). *Analysis of Longitudinal Data, 2nd ed.* Oxford: Oxford University Press.

Diringer, M.N., Bleck, T.P., Claude, H. J., Menon, D., Shutter, L., Vespa, P., Bruder, N., Connolly, E.S., Citerio, G., Gress, D., Hanggi, D., Hoh, B.L., Lanzino, G., Le, R. P., Rabinstein, A., Schmutzhard, E., Stocchetti, N., Suarez, J.I., Treggiari, M., Tseng, M.Y., Vergouwen, M.D., Wolf, S., Zipfel, G.; Neurocritical Care Society. (2011). Critical care management of patients following aneurysmal subarachnoid hemorrhage: recommendations from the Neurocritical Care Society's Multidisciplinary Consensus Conference. *Neurocritical Care,* 2011 Sep;15(2), 211-240.

Dorhout, S.M., Rinkel, G.J., Feigin, V.L., Algra, A., van den Bergh, W.M., Vermeulen, M.,

van Gijn, J. (2007). Calcium antagonists for aneurysmal subarachnoid haemorrhage. *Cochrane Database System Review,* 2007 Jul 18;(3):CD000277.

Eilers, P., and Marx, B. (1996). Flexible smoothing with B-splines. *Statistical Science*, 11, 89-121.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association,* 87, 998-1004.

Fu W. (2003). Penalized Estimating Equations. *Biometrics*, 59, 126-132.

Guo W. (2002). Functional mixed effects models. *Biometrics*, 58(1):121-128.

Huang, J.Z., Zhang, L. and Zhou, L. (2007). Efficient estimation in marginal partially linear models for longitudinal/clustered data using splines. *Scandinavian Journal of Statistics*, 34, 451-477.

Jaeger, M., Schuhmann, M. U., Soehle, M. and Meixensberger, J. (2006). Continuous assessment of cerebrovascular autoregulation after traumatic brain injury using brain tissue oxygen pressure reactivity. *Critical care medicine*, 34(6): 1783-1788.

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

Ibrahim, A. and Suliadi, (2010b). Analyzing Longitudinal Data Using Gee-Smoothing Spline. *WSEAS Transactions on Systems and Control.* ISSN 1790-5117.

Li, Y. and Ruppert, D. (2008). On the asymptotics of penalized splines. *Biometrika* **95**, 415–36.

Lin, X and Carroll, R. (2000). Nonparametric function estimation for clusterted data when the predictor is measured without/with error. *Journal of the American Statistical Association*, 95, 520-534.

Lin, X, Wang, N., Welsh, A and Carroll, R. (2004). Equivalent kernels of smoothing splines in nonparametric regression for clustered data. *Biometrika*, 92, 177-193.

Opsomer, J.D., Wang, Y. and Yang Y. (2001). Nonparametric regression with correlated errors. *Statistical Science*, 16, 134-153.

Ruppert, D. (2002). Selecting the number of knots for penalized splines, JCGS , 11, 735-757.

Ruppert, D. Wand, M.P. and Carroll, R.J. (2003). *Semiparametric Regression*. New York: Cambridge University Press.

Staicu, A., Crainiceanu, C., and Carroll, C. (2010). Fast methods for spatially correlated multilevel functional data. *Biostatistics*, 11 (2): 177-194.

Stiefel, M.F., Heuer, G.G., Abrahams, J.M., Bloom, S., Smith, M.J., Maloney-Wilensky, E., Grady, M.S., LeRoux, P.D. (2004). The effect of nimodipine on cerebral oxygenation in patients with poor-grade subarachnoid hemorrhage. *Journal of Neurosurgery*, 101, 594-599.

Wahba, G. (1985). A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *The Annals of Statistics*, 4, 1378-1402.

Wand, M.P. (2003). Smoothing and Mixed Models. *Computational Statistics*, 18, 223-249.

Wang, N. (2003) Marginal Nonparametric kernel regression accounting for within-subject correlation. *Biometrika*, 90, 43-52.

Wang, Z, and Louise, T. (2003). Matching Conditional and Marginal Shapes in Binary Random Intercept Models Using a Bridge Distribution Function. *Biometrika*, 90, 765-775.

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

Zhao, L.P., Prentice, R., and Self, S. (1992). Multivariate Mean Parameter Estimation by using a Partly Exponential Model. *Journal of the Royal Statistical Society, Series B*, 54(3), 805-811.

Zhou, L., Huang, J. and Carroll, R. (2008). Joint modeling of paired sparse functional data using principal components. *Biometrika*, **95**, 601-619.

Zhu, Z., Fung, W.K., and He, X. (2008). On the asymptotics of marginal regression splines with longitudinal data. *Biometrika* **95**, 907–917.

Table 1: Mean average MSE of $\widehat{f}(x)$ 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 $\widehat{f}(x)$ using various smoothing techniques and smoothing parameter selection, binary outcome, $n = 100, m = 5$, 500 simulations

$f(x)$	P-spline (MSE)	P-spline (CV)	R-spline
$\sin(2\pi x)$	0.059	0.064	0.063
$\exp(x) - 2$	0.047	0.057	0.058
$2 - 16x + 30x^2 - 15x^3$	0.060	0.066	0.065

Table 4: Pointwise standard deviation with binary outcome, exchangeable correlation ($\rho = 0.2$), $f(x) = \sin(2\pi x)$, $n = 100, m = 5$, 500 simulations

	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

*: Under correctly specified compound symmetry correlation

**: Under incorrectly specified working independence correlation

Table 5: Mean average MSE of $\widehat{f}(x)$ and SE of $\hat{\beta}$ using different correlation structures, continuous outcome, multilevel model, $n = 100, J = 5, m = 10$, 500 simulations

$f(x) = 2\sin(2\pi x)$				
	R-spline	P-spline (WI)	P-spline (Ind cycles)	P-spline (True)
AMSE$[\widehat{f}(\cdot)]$	0.045	0.047	0.044	0.043
mean $\widehat{\beta}$	0.395	0.394	0.394	0.395
mean $\widehat{\text{SE}}(\widehat{\beta})$	0.221	0.221	0.221	0.221
$f(x) = 2 - 16x + 30x^2 - 15x^3$				
	R-spline	P-spline (WI)	P-spline (Ind cycles)	P-spline (True)
AMSE$[\widehat{f}(\cdot)]$	0.044	0.042	0.041	0.040
mean $\widehat{\beta}$	0.401	0.401	0.401	0.401
mean $\widehat{\text{SE}}(\widehat{\beta})$	0.243	0.243	0.243	0.242

Table 6: Pointwise standard deviation, continuous outcome, multilevel model, normal random error, $n = 100, J = 5, m = 10$, 500 simulations

$f(x) = 2\sin(x)$									
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 $\widehat{f}(x)$ and SE of $\widehat{\beta}$ using different correlation structures, binary outcome, multilevel model, $n = 100, J = 5, m = 10$, 500 simulations

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

Table 8: Pointwise standard deviation, binary outcome, multilevel model, $n = 100, J = 5, m = 10$, 500 simulations

$f(x) = \sin(2\pi x)$									
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
$f(x) = \exp(x) - 2$									
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

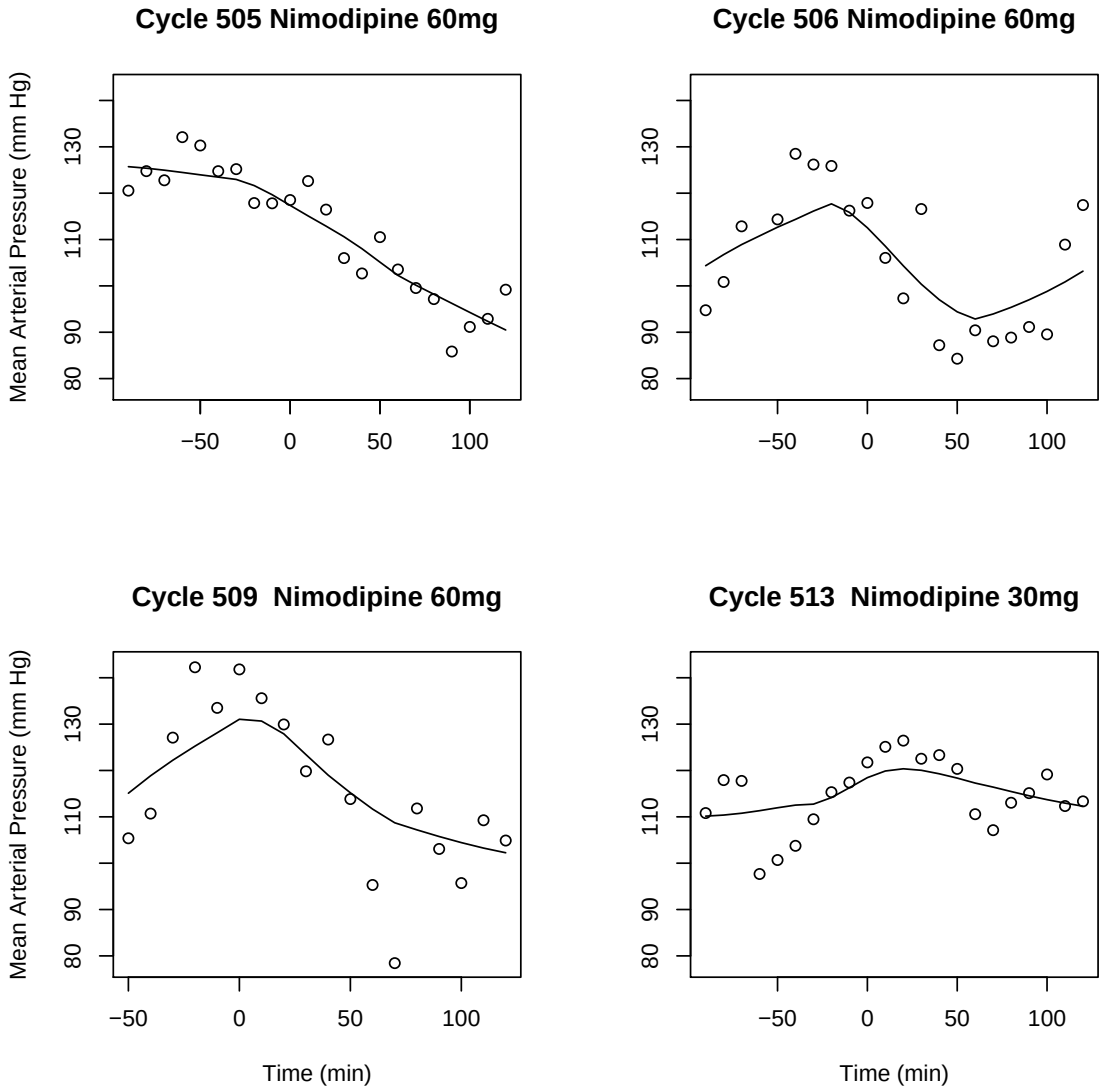

Figure 1: Scatterplot of the MAP versus time measured on a subject during four treatment cycles. Dots: observed MAP; Solid line: local polynomial smoothing using observed MAP in a cycle. Time is centered at administration of Nimodipine, so that negative values are before using the medication and positive values are after using the medication.

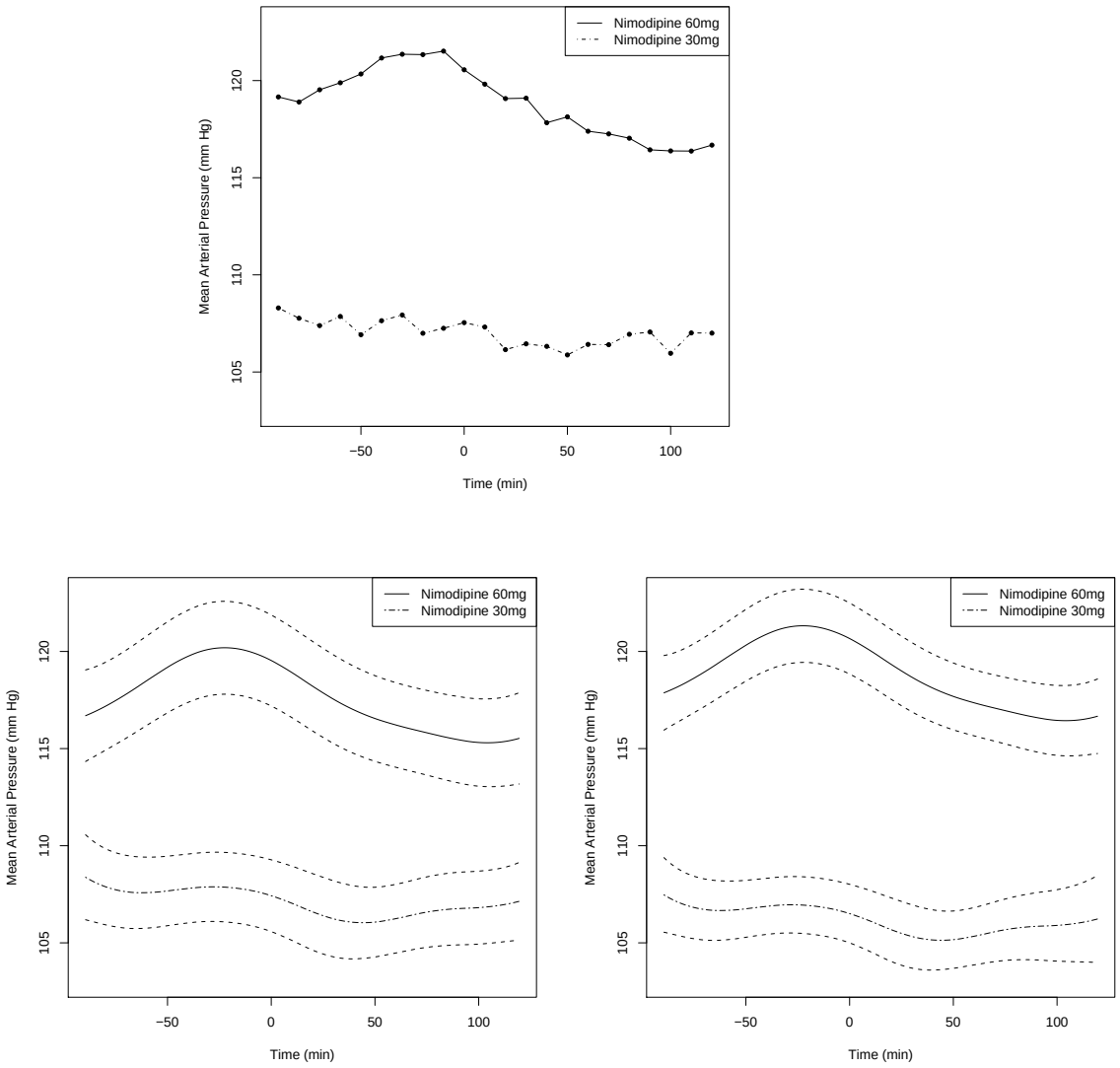

Figure 2: Estimated effect of Nimodipine on MAP and its 95% pointwise confidence band. Upper panel: sample average at each time point. Lower left: assuming independence between cycles. Lower right: using a working two-way ANOVA model based correlation structure.

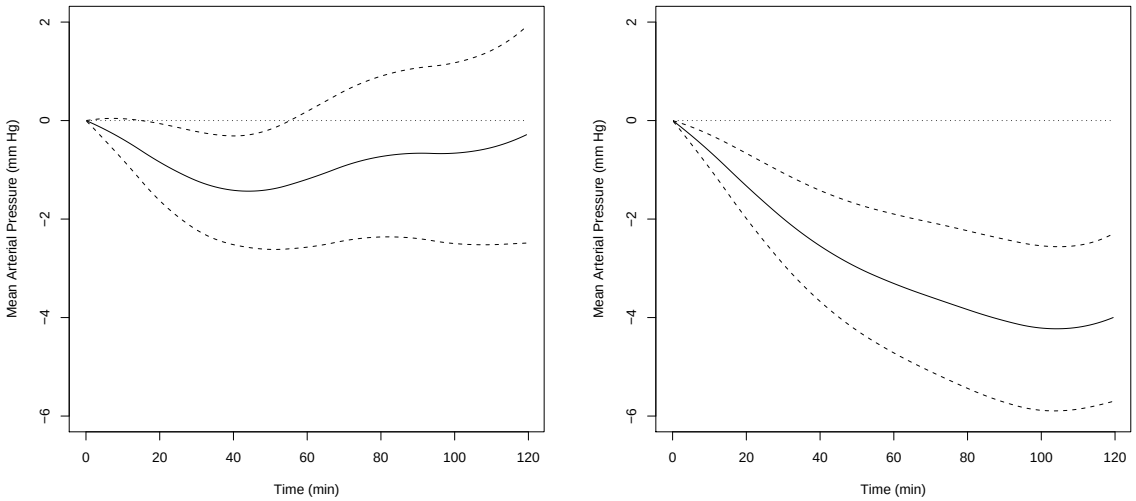

Figure 3: Estimated differences $\widehat{\mu}(t)-\widehat{\mu}(0)$ and $\widehat{f}(t)-\widehat{f}(0)$ for MAP, with $\widehat{\mu}(t)=\widehat{f}(t)+\widehat{\beta}(t)$. Left panel: the low dose group; Right panel: the high dose group. The solid lines are the estimated curves and the dotted lines are associated the 95% pointwise confidence bands.

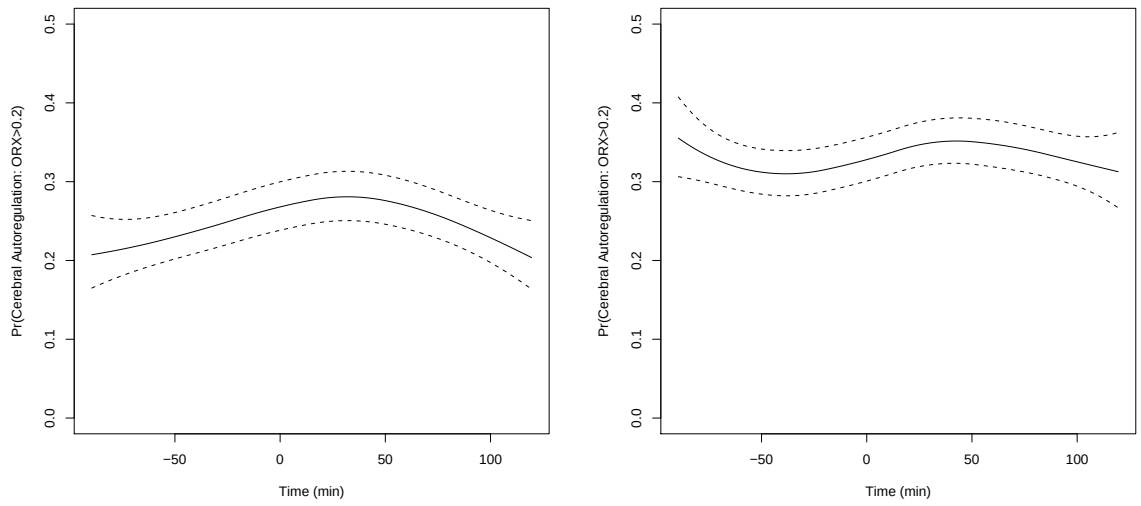

Figure 4: Estimated effect of Nimodipine on being at risk of cerebral autoregulation loss. Left panel: the low dose group. Right panel: the high dose group. The solid lines are the estimated curves and the dotted lines are associated the 95% pointwise confidence bands.

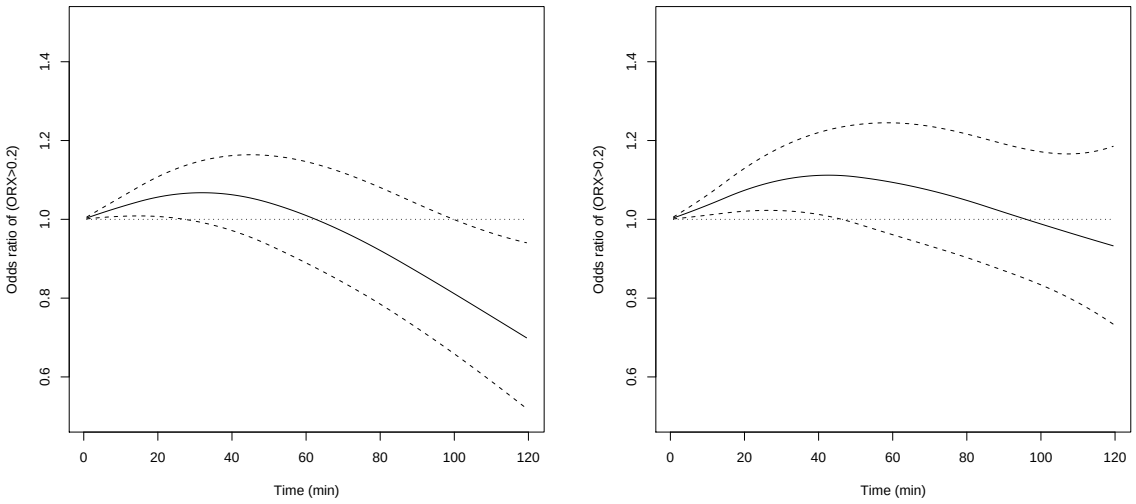

Figure 5: Estimated odds ratio $\exp[\widehat{\mu}(t)]/\exp[\widehat{\mu}(0)]$ and $\exp[\widehat{f}(t)]/\exp[\widehat{f}(0)]$ of at risk for autoregulation loss, with $\widehat{\mu}(t) = \widehat{f}(t) + \widehat{\beta}(t)$. Left panel: the low dose group; Right panel: the high dose group. The solid lines are the estimated curves and the dotted lines are associated the 95% pointwise confidence bands.