Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Biometrics. Author manuscript; available in PMC 2026 May 08.

Published in final edited form as:

Biometrics. 2023 June ; 79(2): 722–733. doi:10.1111/biom.13646.

Functional data analysis for longitudinal data with informative observation times

Caleb Weaver,

Luo Xiao,

Wenbin Lu

Department of Statistics, North Carolina State University, Raleigh, North Carolina, USA

Abstract

In functional data analysis for longitudinal data, the observation process is typically assumed to be noninformative, which is often violated in real applications. Thus, methods that fail to account for the dependence between observation times and longitudinal outcomes may result in biased estimation. For longitudinal data with informative observation times, we find that under a general class of shared random effect models, a commonly used functional data method may lead to inconsistent model estimation while another functional data method results in consistent and even rate-optimal estimation. Indeed, we show that the mean function can be estimated appropriately via penalized splines and that the covariance function can be estimated appropriately via penalized tensor-product splines, both with specific choices of parameters. For the proposed method, theoretical results are provided, and simulation studies and a real data analysis are conducted to demonstrate its performance.

Keywords

functional data analysis; informative observation times; longitudinal data; penalized splines

1 | INTRODUCTION

In functional data analysis for longitudinal data, the observation process is typically assumed to be noninformative, that is, itis assumed tobe independent of longitudinal outcomes. However, this assumption is often violated in real applications. For example, in a longitudinal study, the frequency and timing of a subject's observations may be correlated with the values of longitudinal outcomes. Data collected in such a setting are referred to as longitudinal data with informative observation times.

A motivating example is a longitudinal data set of trajectories of patient-reported symptom severity of Parkinson's disease (Mischley et al., 2017). The data consist of self-reported severity of each of 33 symptoms on a scale of 0–100 for 371 patients, with the higher

This is an open access article under the terms of the creative commons Attribution license which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Correspondence Caleb Weaver, Department of Statistics, North Carolina State University, 2311 Stinson Dr, Raleigh, NC 27606, USA. cjweave2@ncsu.edu.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

number reflecting greater symptom severity. The sum of these symptoms generates the total patient-reported outcome in Parkinson's disease (PRO-PD) score. The number of observations for each patient ranges from 2 to 38, and observations are made on a voluntary basis at varying times since a patient's initial diagnosis. Patients entered the study at arbitrary times since diagnosis. There are no instances of dropout, and we only consider data observed less than 20 years following any patient's initial diagnosis. Figure 1 illustrates the 371 trajectories along with smoothed group-mean profiles of subjects with 10 or fewer observations and of subjects with more than 10 observations in relation to patient-reported quality of life (Cella et al., 2010). The average (standard error) PRO-PD score of subjects with 10 or fewer observations is 828.3 (16.5), and the average PRO-PD score of subjects with more than 10 observations is 752.4 (13.2). It is apparent that subjects reporting lower symptom severity tend to report more frequent observations, which provides evidence against the assumption of noninformative observation times.

It is well understood that standard methods for longitudinal data that ignore the informative observation processes may result in bias in parameter estimation. Hence, a number of methods have been proposed for the situation in which longitudinal outcomes are correlated with observation times (Lin et al., 2004; Sun et al., 2005, 2007, 2012; Liang et al., 2009; Zhu et al., 2011; Cai et al., 2012; Song et al., 2012; Zhao et al., 2012; Fang et al., 2016). Among them, a commonly used modeling strategy is to consider shared random effects models (eg, Sun et al., 2005; Liang et al., 2009). Such models can allow flexible dependence between longitudinal outcomes and observation times.

However, existing literature for longitudinal data with informative observation times mostly focuses on the estimation of an unspecified smooth mean function or integrated mean function. In the analysis of longitudinal data, besides the mean function, it might also be interesting to consider a nonparametric covariance function, which captures the patterns of subject-specific trajectories. A popular nonparametric modeling approach for estimating both mean and covariance functions of longitudinal data is via methods of functional data analysis (Ramsay & Silverman, 2006), in which one views the values of a longitudinal trajectory as observations from a random smooth function; see Guo (2002) and Yao et al. (2003, 2005). The advantage of functional data methods compared to traditional parametric longitudinal methods is that they allow for unspecified mean/covariance functions, which may better capture complex variation patterns of subject-specific trajectories.

Theoretical properties of functional data analysis have been well studied in several papers (Li & Hsing, 2010; Cai & Yuan, 2011; Zhang & Wang, 2016; Xiao, 2020). In particular, Zhang and Wang (2016) and Xiao (2020) considered general weighting schemes for aggregating multiple observations of each subject in the estimation of mean and covariance functions of functional data. The theories therein assume that the observation times are either fixed or randomly sampled and independent from the longitudinal outcomes. Hence, existing theories are not directly applicable to settings in which the observation process is informative. Nevertheless, it might be helpful to clarify that, as shall be seen by our model specification, we consider longitudinal data where each subject only has only a finite number of observations, and hence the sparse to dense phase transition of functional data in

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

terms of the number of observations per subject as those in Cai and Yuan (2011) and other works is not relevant here.

In this paper, we consider functional data methods for longitudinal data with informative observation times. We use a heuristic argument to show that a commonly used functional data method may be inconsistent for such data. Using the same argument, we identify one particular functional data method that could give consistent estimation under the considered shared random effect model. The difference in these methods lies in how the multiple observations of each subject are aggregated to formulate the weighted least square estimation. This contrasts dramatically with existing functional data theories, which found that both methods lead to consistent estimation. For the suggested method with penalized splines, we derive the corresponding rates of convergence, which are either optimal or the best in the literature. It is worthy of pointing out that the theoretic derivation is far from a straightforward application of existing theoretic results which consider noninformative observation times while the theory in the paper has to handle informative time points. Simulation studies demonstrate the desired performance of the suggested method and the inconsistent behavior of the aforementioned commonly used method in functional data analysis. Finally, the suggested method is applied to the motivating data and we conclude the paper with some discussion.

2 | METHODS

2.1 | Model specification

Let $Y_i(t)$ be subject i's underlying longitudinal outcome ($i = 1, \ldots, n$) at time $t \in \mathcal{T}$ where $\mathcal{T}$ is a compact time interval. Without loss of generality, let $\mathcal{T} = [0, 1]$. The observations $Y_{ij} = Y_i(T_{ij})$ of subject i are taken at time points $T_{i1} < T_{i2} < \ldots < T_{im_i}$. Let $C_i \in \mathcal{T}$ denote the censoring time of the ith subject. The cumulative number of observations for the ith subject by time t is $N_i(t) = N_i^*(t \wedge C_i)$, where $N_i^*(t) = \sum_{j \geq 1} I(T_{ij} \leq t)$. Here, I is the indicator function and $s \wedge t$ denotes the minimum of s and t. The observed values of $Y_i(t)$ are obtained at the jump points of $N_i(t)$. Consider the functional data model

$$Y_i(t) = \mu^*(t, Z_i) + X_i(t) + \epsilon_i(t), \tag{1}$$

where $\mu^*(\cdot, Z_i)$ is a conditional mean function parameterized by the latent frailty variable Z_i, $X_i(\cdot)$ is a zero-mean random function that models subject i's smooth deviation from the mean, and $\epsilon_i(\cdot)$ is zero-mean white noise with finite variance σ_ϵ^2. Define the marginal mean function $\mu(t) = \mathbb{E}\{\mu^*(t, Z_i)\}$, the covariance function of the conditional mean functions $\sigma_{\mu^*}(s, t) = \mathrm{cov}\{\mu^*(s, Z_i), \mu^*(t, Z_i)\}$, and the covariance function of the random subject-specific functions $\sigma_X(s, t) = \mathrm{cov}\{X_i(s), X_i(t)\}$. Then, the responses can be considered as observations from a random function with mean function $\mu(t)$ and covariance function $\sigma(s, t) = \sigma_{\mu^*}(s, t) + \sigma_X(s, t)$, contaminated with measurement error.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

As commonly considered in shared random effects models (eg, Liang et al., 2009), we make the following assumptions: (1) censoring time C_i is independent of the observation times $\{T_{ij}, j \geq 1\}$ and longitudinal outcomes $\{Y_i(T_{ij}), j \geq 1\}$; (2) given the frailty variable Z_i, the observation process $N_i^*(\cdot)$ is independent of the longitudinal outcomes $\{Y_i(T_{ij}), j \geq 1\}$; and (3) given the frailty variable Z_i, the observation process $N_i^*(t)$ follows a Poisson process with the intensity function $\lambda(t, Z_i) = Z_i\lambda_0(t)$, where $\lambda_0(t)$ is an unspecified intensity function. Define $\Lambda_0(t) = \int_0^t \lambda_0(s)\,\mathrm{d}s$, the baseline cumulative intensity function.

The dependence between longitudinal outcomes and observation times is induced by the latent frailty variable Z_i, but the form and magnitude of the dependence is left completely unspecified for flexibility. To simplify theoretical analysis, we allow m_i to be 0. In formulas and equations involving m_i^{-1}, the term m_i^{-1} should always be interpreted as $m_i^{-1}I(m_i \neq 0)$, which equals 0 if $m_i = 0$.

2.2 | Estimation of mean function

The mean function $\mu(t)$ is approximated by a spline function $B^{\mathsf{T}}(t)\boldsymbol{\theta}$ where $B(t) = \{B_1(t), \ldots, B_K(t)\}^{\mathsf{T}} \in \mathbb{R}^K$ is a vector of rth order B-spline basis functions constructed from equally spaced knots in $\mathcal{T}$, and $\boldsymbol{\theta}$ is a vector of coefficients. The coefficient vector $\boldsymbol{\theta}$ is estimated via the minimization

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left[\sum_{i=1}^{n} \sum_{j=1}^{m_i} w_i \left\{ Y_{ij} - B^{\mathsf{T}}(T_{ij})\boldsymbol{\theta} \right\}^2 + \lambda \boldsymbol{\theta}^{\mathsf{T}} P \boldsymbol{\theta} \right], \tag{2}$$

where w_is are nonnegative weights that may depend on m_i and C_i, λ is a smoothing parameter that balances fit and smoothness, and P is a penalty matrix such that $\boldsymbol{\theta}^{\mathsf{T}} P \boldsymbol{\theta}$ equals the squared sum of the qth order consecutive differences of $\boldsymbol{\theta}$ as in Eilers and Marx (1996). The mean function is then estimated by $\hat{\mu}(t) = B^{\mathsf{T}}(t)\hat{\boldsymbol{\theta}}$.

A common choice of the weights is $w_i = 1$, which means that each observation has equal weight in forming the least squares in (2). We will denote this choice of weights by "OBS," following Zhang and Wang (2016) and Xiao (2020). To our knowledge, this is the most often used choice for formulating the least squares in functional data analysis. Another choice of weights is to $w_i = m_i^{-1}$, which implies that the set of observations from each subject has equal weight (eg, Li & Hsing, 2010). We will denote this choice of weights by "SUBJ." Both choices of weights could lead to consistent and even rate-optimal estimation in functional data analysis as shown in Zhang and Wang (2016) and Xiao (2020) for functional data with noninformative observation times. However, in what follows, we shall heuristically derive different weighting schemes will affect estimation for functional data with informative observation times.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Let $\mathscr{A}_i(t) = w_i\{Y_i(t) - \mu_n(t)\}\,\mathrm{d}N_i(t)$, where $\mu_n(t)$ is the solution to the estimating equation $\sum_{i=1}^{n}\mathscr{A}_i(t) = 0$. Then, we have

$$\mu_n(t) = \frac{\sum_{i=1}^{n} w_i Y_i(t)\mathrm{d}N_i(t)}{\sum_{i=1}^{n} w_i \mathrm{d}N_i(t)}. \tag{3}$$

As shown in Liang et al. (2009),

$$\mathbb{E}\{\mathrm{d}N_i(t) \mid m_i, C_i\} = I(C_i \geq t) m_i \frac{\mathrm{d}\Lambda_0(t)}{\Lambda_0(C_i)}.$$

Thus,

$$\begin{aligned}
&\mathbb{E}\{w_i Y_i(t)\mathrm{d}N_i(t) \mid C_i\} \\
&= \mathbb{E}[w_i \mathbb{E}\{Y_i(t)\,\mathrm{d}N_i(t) \mid m_i, C_i\} \mid C_i] \\
&= \mathbb{E}\left[w_i I(C_i \geq t) m_i \frac{\mathrm{d}\Lambda_0(t)}{\Lambda_0(C_i)} \mathbb{E}\{\mu^*(t, Z_i) \mid m_i, C_i\} \mid C_i\right] \\
&= I(C_i \geq t)\frac{\mathrm{d}\Lambda_0(t)}{\Lambda_0(C_i} \mathbb{E}\{\mu^*(t, Z_i) w_i m_i \mid C_i\}.
\end{aligned}$$

Similarly, we derive that

$$\mathbb{E}\{w_i\,\mathrm{d}N_i(t) \mid C_i\} = I(C_i \geq t)\frac{\mathrm{d}\Lambda_0(t)}{\Lambda_0(C_i)} \mathbb{E}(w_i m_i \mid C_i).$$

It follows by Equation (3) that as n increases to infinity, $\mu_n(t)$ converges to

$$\mu^o(t) \equiv \frac{\mathbb{E}\left[\frac{I(C_i \geq t)}{\Lambda_0(C_i)} \mathbb{E}\{\mu^*(t, Z_i) w_i m_i \mid C_i\}\right]}{\mathbb{E}\left\{\frac{I(C_i \geq t)}{\Lambda_0(C_i)} \mathbb{E}(w_i m_i \mid C_i)\right\}}. \tag{4}$$

We now evaluate expression (4) with either the OBS weights or the SUBJ weights. First, consider the SUBJ weights for which $w_i = m_i^{-1}$. Then $w_i m_i = 1$. Because Z_i is independent of C_i, $\mu^o(t) = \mathbb{E}\{\mu^*(t, Z_i) \mid C_i\} = \mathbb{E}\{\mu^*(t, Z_i)\} = \mu(t)$. Hence, for the SUBJ weights, $\mu_n(t)$ converges to $\mu(t)$ as desired. Second, consider the OBS weights for which $w_i = 1$. In general, in this case $\mu^o(t)$ in (4) is different from $\mu(t)$.

To give an example, consider the simulation setting in Section 4, where we let $\mu^*(t, Z) = Z(t^2 + 2/3)$, $Z = (1 - \rho) + \rho Z^*$ for some $0 < \rho < 1$, and Z^* is from a distribution with mean 1 and variance $\sigma^2_{Z^*}$. In addition, m follows a Poisson distribution with mean $10Z\Lambda_0(C)$. Here, (Z, m, C) has the same distribution as (Z_i, m_i, C_i). Then we derive that $\mathbb{E}(m \mid C) = 10\mathbb{E}\{Z\Lambda_0(C) \mid C\} = 10\Lambda_0(C)$, $\mu^*(t, Z)m = Zm(t^2 + 2/3)$, and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$\mathbb{E}\{\mu^*(t,Z)m \mid c\} = (t^2+2/3)\mathbb{E}(Zm \mid C) = 10(t^2+2/3)\mathbb{E}\{Z^2\Lambda_0(C) \mid C\} = 10\Lambda_0(C)(t^2+2/3).(1+\rho^2\sigma_{Z^*}^2)$

It follows that $\mu^o(t) = (t^2+2/3)(1+\rho^2\sigma_{Z^*}^2)$, which differs from $\mu(t) = t^2+2/3$. Indeed, our simulation study in Section 4 will provide empirical confirmation of the above derivations. Another general example is when $C_i = 1$ for all i, that is, no censoring. Then it is easy to see that (4) reduces to $\mu^o(t) = \mathbb{E}\{\mu^*(t,Z)m\}/\mathbb{E}(m)$, which differs from $\mu(t) = \mathbb{E}\{\mu^*(t,Z)\}$ unless $\mu^*(t,Z)$ and m are uncorrelated.

The above derivations show that the OBS weights lead to inconsistent estimation of the mean function while in contrary the SUBJ weights might give consistent estimation. This differs from traditional functional data analysis, where both OBS and SUBJ weights will lead to consistent estimation (Zhang & Wang, 2016; Xiao, 2020). Thus, for functional data with informative observation times under the shared random effects model, we propose to use the SUBJ weights. Accordingly, we shall rigorously derive the rate of convergence of the mean function estimation in Section 3.

2.3 | Estimation of covariance function

Let $\hat{\mu}(t)$ be the estimate of the mean function using the SUBJ weights. Let $\tilde{e}_{ij} = Y_{ij} - \hat{\mu}(T_{ij})$ and define the auxiliary variables $\tilde{\sigma}_{ij_1j_2} = \tilde{e}_{ij_1}\tilde{e}_{ij_2}$, which is an empirical estimate of $\sigma(T_{ij_1}, T_{ij_2})$. Then $\{(T_{ij_1}, T_{ij_2}, \tilde{\sigma}_{ij_1j_2}): 1 \le j_1 \ne j_2 \le m_i, i = 1,\ldots,n\}$ forms a data set for estimating the covariance function.

The covariance function $\sigma(s,t)$ is approximated by a tensor-product spline, $H(s,t) = \sum_{1\le k,l\le K}\theta_{kl}B_k(s)B_l(t)$, where $\Theta = (\theta_{kl})_{1\le k\le K, 1\le l\le K}$ is a coefficient matrix. To simplify notation, we use the same univariate B-splines as in the mean function estimation. Since the covariance function is symmetric, it is natural to impose the constraint that $\Theta = \Theta^{\mathsf{T}}$, which ensures that $H(s,t) = H(t,s)$. Let $\mathrm{vec}(\cdot)$ be the vectorization operator, which is invertible, and define $\boldsymbol{\theta}_\sigma = \mathrm{vec}(\Theta)$. Let $B(s,t) = B(t)\otimes B(s)$, where $\otimes$ is the Kronecker product. Then, $H(s,t) = B^{\mathsf{T}}(s,t)\boldsymbol{\theta}_\sigma$. The coefficient matrix Θ is estimated via the minimization

$$\widehat{\Theta} = \arg\min_{\Theta:\Theta=\Theta^{\mathsf{T}}} \sum_{i=1}^{n}\left[v_i \sum_{1\le j_1\ne j_2\le m_i}\{\tilde{\sigma}_{ij_1j_2} - H(T_{ij_1},T_{ij_2})\}^2\right] + \lambda K^{-1}\boldsymbol{\theta}_\sigma^{\mathsf{T}}(I_K\otimes P)\boldsymbol{\theta}_\sigma, \tag{5}$$

where v_is are nonnegative weights, λK^{-1} is a smoothing parameter, and P is the same penalty matrix for mean function estimation. We use the same notation λ as in the mean function estimation to simplify the notation in theoretical analysis. The above method was proposed in Xiao et al. (2018) and can be shown to be a special case of the bivariate tensor-product P-splines in Eilers and Marx (2003).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Let $\widehat{\boldsymbol{\theta}}_\sigma = \mathrm{vec}(\widehat{\Theta})$. Then, the covariance function $\sigma(s,t)$ is estimated by $\hat{\sigma}(s,t) = B^{\mathsf{T}}(s,t)\widehat{\boldsymbol{\theta}}_\sigma$. As in the mean function estimation, we consider two choices of weights. We will denote the choice of $v_i = 1$ by "OBS" and the choice of $v_i = \{m_i(m_i-1)\}^{-1}$ by "SUBJ." Similar to the mean function estimation, the former weight implies each auxiliary variable carries the same weight in forming the weighted least squares while the latter implies the set of auxiliary variables from each subject carries the same weight. Then, by similar arguments (omitted) as in the mean function estimation, it can be shown that the SUBJ weights should be adopted. And we shall also derive rates of convergence of the above proposed estimator with the SUBJ weights.

3 | THEORETICAL PROPERTIES

3.1 | Notation

For a univariate function g over $\mathcal{T}$, $\|g\|$ denotes its supreme norm. The same notation is also used for a bivariate function. We also use $g^{(1)}$ to denote the derivative of a function g. For two scalars, let $a \wedge b = \min(a,b)$ and $a \vee b = \max(a,b)$. We let $\underline{m} = \{m_i, 1 \le i \le n\}$ and $\underline{C} = \{C_i, 1 \le i \le n\}$.

3.2 | Asymptotic properties of mean function estimator

Assumption 1. The nonnegative latent variables $Z_i (1 \le i \le n)$ are independent and identically distributed and satisfy, for any positive constant $c > 0$, $0 < \mathbb{E}(Z_1 e^{-cZ_1}) < \infty$.

Assumption 2. The censoring times $C_i (1 \le i \le n)$ are independent and identically distributed.

Assumption 3. The baseline hazard function $\Lambda_0(t)$ has a positive and continuous density function on $\mathcal{T}$.

Assumption 4. Let $\rho_1(t) = \mathbb{E}\{\Lambda_0^{(1)}(t \wedge C_1) \,/\, \Lambda_0(C_1)\}$ and $\rho_2(s,t) = \mathbb{E}\Lambda_0^{(1)}(t \wedge C_1)\Lambda_0^{(1)}(s \wedge C_1) \,/\, \Lambda_0^2(C_1)\}$. Then, ρ_1 and ρ_2 are continuous and bounded away from 0 and infinity, over $\mathcal{T}$ and $\mathcal{T}^2$, respectively.

Assumption 5. (a) $\|\sigma\| < \infty$; (b)$\sigma_\epsilon^2 < \infty$.

Assumption 6. (a) The number of basis functions K satisfies $K \ge n^{\delta_1}$ for some constant $\delta_1 > 0$ and $K = o(n)$; (b) The smoothing parameter λ satisfies $\lambda = o(n^{-\delta_2})$ for some constant $\delta_2 > 0$; (c) $\log n \,/\, n = o(K^{-4})$.

We now discuss Assumption 4. First, $\rho_1(t)$ is the density function of the theoretical counterpart of the empirical distribution of the observed time points. In nonparametric regression, it is usually required this function is bounded away from 0 and infinity. When there are no censoring, that is, $C_i = 1$, then $\rho_1(t) = \Lambda_0^{(1)}(t)$ and $\rho_2(s,t) = \Lambda_0^{(1)}(s)\Lambda_0^{(1)}(t)$. Because of Assumption 3, Assumption 4 will always hold. When there are censoring, we could show

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

that one generating distribution for C_i in the simulation study also ensures that Assumption 4 holds.

For theoretical results, we shall focus on the SUBJ weights and scale the weights w_i with $w_i = (nm_i)^{-1}$ so that $\sum_{i=1}^{n} w_i m_i = 1$. The notation r denotes the order of splines and q denotes the order of penalty. Let $h = K^{-1}$ and $h_e = h \vee \lambda^{1/(2q)}$. By Xiao (2019), h_e plays the role of bandwidth parameter for penalized splines.

Theorem 1. Suppose that Assumptions 1-6 hold. If $\mu \in \mathscr{C}^p(\mathscr{T})$ with $q \leq p \wedge r$, then

$$\mathbb{E}\left[\int_{\mathscr{T}} \{\hat{\mu}(t) - \mu(t)\}^2 \, \mathrm{d}t \mid \underline{m}, \underline{C}\right] = O(h^{2r}) + o(h^{2p}) + O\left(\lambda^2 h_e^{-2q}\right) + O(n^{-1} h_e^{-1}), \ a.s.$$

of Theorem 1. The proof of Theorem 3.3 in Xiao (2020) can be adapted if Lemmas A.7, A.8, and A.9 in Xiao (2020) hold almost surely, which are given in Lemma 1 and proved in Section A of the Supporting Information. These lemmas deal with empirical distributions of the observed time points, which are informative, and hence different proofs are required.

The derived L_2 convergence rate in Theorem 1 is the same as the those in Xiao (2020) for the scenario with a finite number observations per subject and randomly observed time points. Moreover, as shown in Claeskens et al. (2009), Xiao (2019), and Huang and Su (2021), there are two types of asymptotics for penalized splines: (1) when $\lambda = o(h^{p+q})$, then $h_e = h$ and the rates become $O(h^{2r}) + o(h^{2p}) + O(h^{n^{-1}h^{-1}})$, which are the asymptotic rates for regression splines; (2) when $\lambda^{-1} = o(h^{-2q})$, then $h_e = \lambda^{-1/(2q)}$ and the rates become $O(\lambda) + O(n^{-1}\lambda^{-1/(2q)})$, the asymptotic rates for smoothing splines. Finally, both types of asymptotics could lead to the best rate, $n^{-2p/(2p+1)}$, which is the optimal rate for estimating univariate smooth functions (Stone, 1982).

3.3 | Asymptotic properties of covariance function estimator

To simplify the theoretic analysis, we shall assume that the mean function $\mu(t)$ is known and the auxiliary variables in Section 2.3 are accordingly constructed.

Assumption 7. (a) $\sup_{t \in \mathscr{T}} \mathbb{E}\{\mu^*(t, Z_i) - \mu(t)\}^4 < \infty$; (b) $\sup_{t \in \mathscr{T}} \mathbb{E}\{X_i(t)\}^4 < \infty$; (c) $\sup_{t \in \mathscr{T}} \mathbb{E}\{\epsilon_i(t)\}^4 < \infty$.

Assumption 8. Let

$$\rho_3(s, t_1, t_2) = \mathbb{E}\{\Lambda_0^{(1)}(s \wedge C_1)\Lambda_0^{(1)}(t_1 \wedge C_1)\Lambda_0^{(1)}(t_2 \wedge C_1) / \Lambda_0^3(C_1)\}$$

and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\begin{aligned}&\rho_4(s_1, s_2, t_1, t_2)\\&\quad = \mathbb{E}\{\Lambda_0^{(1)}(s_1 \wedge C_1)\Lambda_0^{(1)}(s_2 \wedge C_1)\Lambda_0^{(1)}\\&\quad \times (t_1 \wedge C_1)\Lambda_0^{(1)}(t_2 \wedge C_1) / \Lambda_0^4(C_1)\}.\end{aligned}$$

Both ρ_3 and ρ_4 are continuous and bounded away from 0 and infinity, over $\mathscr{T}^3$ and $\mathscr{T}^4$, respectively.

Theorem 2. Suppose that Assumptions 1-8 hold. If $\sigma \in \mathscr{C}^p(\mathscr{T})$ with $q \leq p \wedge r$, then

$$\begin{aligned}&\mathbb{E}\left[\int\int \{\hat{\sigma}(s,t) - \sigma(s,t)\}^2 \,\mathrm{d}s\,\mathrm{d}t \mid \underline{m}, \underline{C}\right]\\&\quad = O_{\mathbb{P}}(h^{2r}) + o_{\mathbb{P}}(h^{2p}) + O_{\mathbb{P}}\left(\lambda^2 h_e^{-2q}\right) + O_{\mathbb{P}}(n^{-1}h_e^{-2}).\end{aligned}$$

The proof of Theorem 2 is given in Section B of the Supporting Information. The derived L_2 rate in Theorem 2 can be shown to achieve the best rate in the literature (Zhang and Wang, 2016; Xiao, 2020), $n^{-2p/(2p+2)}$, which is the optimal rate for estimating bivariate smooth functions (Stone, 1982).

4 | SIMULATION STUDY

4.1 | Simulation settings

In this section, we conduct simulation studies to evaluate the performance of the proposed method, that is, penalized splines with SUBJ weights, and compare it with that of penalized splines with OBS weights. In the following simulations, the maximum follow-up time is $\tau = 1$ and we consider three censoring time distributions: (1) $C_i = 1$, which corresponds to no censoring; (2) C_i is sampled from a mixture distribution of Uniform(0.2, 1) and a point mass at 1, with mixing factors of 0.8 and 0.2, respectively; and (3) $C_i \sim \text{Uniform}(0, 1)$. For the second generating distribution for the censoring times, which we call mixed censoring, Assumptions 4 and 8 still hold and hence our theoretical results will still hold; see Section C of the Supporting Information for a proof. The third generating distribution, which we call uniform censoring, serves for the purpose of a sensitivity analysis. The frailty variable is chosen as $Z_i = (1-\rho) + \rho Z_i^*$, where Z_i^* is sampled from a gamma distribution or lognormal distribution, both with mean 1 and variance $\sigma^2_{Z^*} = 0.5$ or $\sigma^2_{Z^*} = 1$, and $\rho = 0.5$ or 0.85, corresponding to the informative observation process settings, and $\rho = 0$, corresponding to the noninformative observation process settings. The observation process follows a Poisson process with the intensity function $10Z_i$. Given Z_i and C_i, the number of observations m_i of subject i then follows a Poisson distribution with the mean $10Z_iC_i$.

The conditional mean function is $\mu^*(t, Z_i) = Z_i(t^2 + 2/3)$, implying a marginal mean function of $\mu(t) = t^2 + 2/3$. The covariance function of the random functions X_i is chosen as $\sigma_X(s,t) = 0.05\exp(s^2) \times 0.05\exp(t^2) \times (1 - |s-t|)$. The observed values $Y_i(T_{ij})$ are then generated from (1), where $\epsilon_i(T_{ij}) \sim \mathcal{N}(0, \sigma_\epsilon^2)$ and σ_ϵ^2 is set to be 0.5 in the informative

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

observation process settings, and 1 in the noninformative observation process settings. The number of subjects is set to $n = 150$ or $n = 300$.

We replicate each simulation setting 200 times. For univariate penalized splines, we use 10 cubic B-spline bases with equally spaced knots in the unit interval and for the tensor-product penalized splines, we use 10 marginal cubic B-splines with equally spaced knots in both dimensions. The smoothing parameters are selected using generalized cross-validation via a modified version of the function *fpca.sparse* in the R package *face*, which performs the fast covariance estimation (FACE) method for sparse functional data (Xiao et al., 2018).

4.2 | Simulation results

Define the integrated squared error as either $\int\{\hat{\mu}(t) - \mu(t)\}^2 \mathrm{d}t$ or $\int\int\{\hat{\sigma}(s,t) - \sigma(s,t)\}^2 \mathrm{d}s\,\mathrm{d}t$ where $\hat{\mu}(t)$ and $\hat{\sigma}(s,t)$ are estimates of $\mu(t)$ and $\sigma(s,t)$. For each simulation setting, we calculate the median and interquartile range of the integrated squared error of the estimates using the OBS and SUBJ choices of weights.

Figure 2 plots the averages of the estimated mean functions from penalized splines with either OBS or SUBJ weights for a few simulation settings with gamma frailty variables, for no censoring and mixed censoring times. In these plots, $\sigma^2_{Z^*} = 0.5$. A nonzero value of ρ in these plots indicate informative observation times. We first observe from the plots that penalized splines with OBS weights converge to $\mu^o(t) = (t^2 + 2/3)(1 + \rho^2\sigma^2_{Z^*})$, the derived expression in (4) for OBS weights, which differs from the true $\mu(t) = t^2 + 2/3$ and the difference becomes more pronounced as ρ increases. Next we see that penalized splines with SUBJ weights converge to the true mean function as desired, irrespective of the values of ρ. Thus, the plots demonstrate the inconsistent estimation of the mean function by penalized splines with OBS weights and the desired model estimation by penalized splines with SUBJ weights. The plots for uniform censoring and lognormal frailty variables are similar and are presented in Web Appendix D.

Table 1 gives the median and interquantile range of integrated squared error for estimating the mean function using both methods in all simulation settings with informative observation times. The numerical results show that penalized splines with SUBJ weights have much smaller median integrated squared error for estimating the mean function, in agreement with the above plots. In addition, the estimators based on the SUBJ weights have much smaller variation in terms of interquartile range in the presence of information observation times. While there is some additional variability in the estimates at the larger time points in the mixed censoring settings compared to the no censoring settings, censoring seems to have a less significant effect on the estimates. Similar simulation results for the uniform censoring setting are found and given in Web Appendix D.

The numerical results for the covariance function estimation under the informative observation times settings are summarized in Table 2, and the corresponding numerical results under the noninformative observation times settings are included in Section D of the Supporting Information. The estimate using the SUBJ weights clearly outperforms the estimate using the OBS weights in both cases. Similar to the results for the mean

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

function estimation, the estimators based on the SUBJ weights have much smaller integrated squared error and variation than those based on the OBS weights under all the settings with information observation times. For illustration, Figure 3 presents the estimated covariance function from both methods along with the true covariance function from two simulated data sets.

In simulation settings with noninformative observations, that is, $\rho = 0$, we find no substantial difference in the integrated squared errors for estimating the mean and covariance functions using either the SUBJ or OBS weights. Indeed, penalized splines with both types of weights perform similarly and both estimate accurately the mean function. The numerical results are given in Web Appendix D.

Based on the above empirical results, we can recommend the use of the SUBJ weights for any longitudinal data with or without informative observations, without concern for suboptimal estimation in the case that the observation process is noninformative.

5 | DATA APPLICATION

In this section, we apply the discussed methods to the Parkinson's disease symptom severity data set described in Section 1. Recall that in this data set, there was evidence of an informative observation process in which the average severity score of subjects with more than 10 observations is less than that of subjects with less than or equal to 10 observations. The intuition behind the selection of weights is clearly illustrated by this example. When weighting each observation equally, the estimates are more heavily influenced toward subjects with more observations. Thus, the mean function is likely to be underestimated. Similarly, the estimate of the covariance function is likely to be biased. Also recall that the time at which a patient entered the study is unrelated to the time since their initial diagnosis, there is no instance of dropout, and we only consider data observed less than 20 years following a patient's diagnosis. There is therefore no informative truncation or censoring present in the data. Therefore, it seems appropriate to use the proposed functional method with SUBJ weights.

The trajectories are plotted in Figure 4 along with the mean function estimates of the penalized spline estimator with the OBS and SUBJ weights. As expected, the mean estimate when using SUBJ weights is greater than the estimate when using OBS weights over the whole time domain. By the estimate from penalized splines with SUBJ weights, we see a monotonic increase in severity over the first 20 years after diagnosis, which agrees with existing research (Group, 2004; Venuto et al., 2016; Holden et al., 2018).

The estimated correlation and variance functions of the penalized spline estimator with SUBJ weights are displayed in Figure 4. It is interesting to see that the variance function takes its smallest value around 8 years after diagnosis and is greatest at the maximal follow-up time. As a comparison, the estimated variance function with OBS weights is also plotted and has uniformly higher values. We also consider the spectral decomposition of the estimated covariance function. The first three eigenfunctions are shown in Figure 4. These first three eigenfunctions account for 60%, 30%, and 6% of the total variation, respectively.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The first eigenfunction corresponds to a contrast between early and late time after diagnosis. The second accounts for deviation from the mean trajectory around 8 years after diagnosis. The estimated correlation function and associated eigenfunctions using OBS weightings are displayed in Section E of the Supporting Information and very similar to the above results using SUBJ weights.

6 | DISCUSSION

In this paper, we studied some commonly used functional data methods for analysis of longitudinal data with informative observation times and identified a method with a proper subject-specific weighting scheme that can achieve the consistent estimation of the mean and covariance functions under a class of shared random effect models. The suggested method can be easily implemented using existing software. The theoretical properties of the estimator were studied under the setting with the informative observation process, and special care must be taken to handle the distribution of observation times for establishing the convergence rate results.

In our current work, no covariates are included in the model. But the proposed functional data method can be easily extended to accommodate covariates. In addition, censoring times are assumed to be independent of longitudinal outcomes and observation times, which we think is reasonable for our data application. It is certainly of interest to extend the proposed method to the setting with the informative observation process and informative dropout. In such a setting, the shared random effects model can be extended to account for dependence between censoring times and longitudinal outcomes and observation times. Then, the nonparametric approach in Wang et al. (2001) can be used to estimate the baseline cumulative intensity function. Note that the derivations that lead to Equation (4) remain valid for informative censoring. Therefore, Equation (4) may be used for selecting an appropriate set of weights in aggregating the observations. Such a research direction merits further investigation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors thank the editor and the associate editor for their constructive and helpful comments, which improved the paper. This work was partially supported by the NIH grants R56 AG064803 (LX), R01 AG064803 (LX), and R01 NS112303 (LX and CW). This work represents the opinions of the researchers and not necessarily that of the granting organizations.

Funding information

National Institutes of Health, Grant/Award Numbers: R01 AG064803, R01 NS112303, R56 AG064803

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

DATA AVAILABILITY STATEMENT

The data that support the findings in this paper are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

REFERENCES

Cai N, Lu W and Zhang HH (2012) Time-varying latent effect model for longitudinal data with informative observation times. Biometrics, 68, 1093–1102. [PubMed: 23025338]

Cai TT and Yuan M (2011) Optimal estimation of the mean function based on discretely sampled functional data: phase transition. Annals of Statistics, 39, 2330–2355.

Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. (2010) The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. Journal of Clinical Epidemiology, 63, 1179–1194. [PubMed: 20685078]

Claeskens G, Krivobokova T and Opsomer JD (2009) Asymptotic properties of penalized spline estimators. Biometrika, 96, 529–544.

Eilers PH and Marx BD (1996) Flexible smoothing with B-splines and penalties. Statistical Science, 11, 89–102.

Eilers PH and Marx BD (2003) Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. Chemometrics and Intelligent Laboratory Systems, 66, 159–174.

Fang S, Zhang H and Sun L (2016) Joint analysis of longitudinal data with additive mixed effect model for informative observation times. Journal of Statistical Planning and Inference, 169, 43–55.

Group, P.S. (2004) Levodopa and the progression of Parkinson's disease. New England Journal of Medicine, 351, 2498–2508. [PubMed: 15590952]

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

Holden SK, Finseth T, Sillau SH and Berman BD (2018) Progression of MDS-UPDRS scores over five years in De Novo Parkinson disease from the Parkinson's progression markers initiative cohort. Movement Disorders Clinical Practice, 5, 47–53. [PubMed: 29662921]

Huang JZ and Su Y (2021) Asymptotic properties of penalized spline estimators in concave extended linear models: rates of convergence. Annals of Statistics, 49, 3383–3407.

Li Y and Hsing T (2010) Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. Annals of Statistics, 38, 3321–3351.

Liang Y, Lu W and Ying Z (2009) Joint modeling and analysis of longitudinal data with informative observation times. Biometrics, 65, 377–384. [PubMed: 18759841]

Lin H, Scharfstein DO and Rosenheck RA (2004) Analysis of longitudinal data with irregular, outcome-dependent follow-up. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 66, 791–813.

Mischley LK, Lau RC and Weiss NS (2017) Use of a self-rating scale of the nature and severity of symptoms in Parkinson's Disease (PRO-PD): correlation with quality of life and existing scales of disease severity. npj Parkinson's Disease, 3, 1–7.

Ramsay JO and Silverman BW (2006) Functional Data Analysis. New York: Springer.

Song X, Mu X and Sun L (2012) Regression analysis of longitudinal data with time-dependent covariates and informative observation times. Scandinavian Journal of Statistics, 39, 248–258.

Stone CJ (1982) Optimal global rates of convergence for nonparametric regression. Annals of Statistics, 10, 1040–1053.

Sun J, Park D-H, Sun L and Zhao X (2005) Semiparametric regression analysis of longitudinal data with informative observation times. Journal of the American Statistical Association, 100, 882–889.

Sun J, Sun L and Liu D (2007) Regression analysis of longitudinal data in the presence of informative observation and censoring times. Journal of the American Statistical Association, 102, 1397–1406.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Sun L, Song X, Zhou J and Liu L (2012) Joint analysis of longitudinal data with informative observation times and a dependent terminal event. Journal of the American Statistical Association, 107, 688–700.

Venuto CS, Potter NB, Ray Dorsey E and Kieburtz K (2016) A review of disease progression models of Parkinson's disease and applications in clinical trials. Movement Disorders, 31, 947–956. [PubMed: 27226141]

Wang M-C, Qin J and Chiang C-T (2001) Analyzing recurrent event data with informative censoring. Journal of the American Statistical Association, 96, 1057–1065.

Xiao L. (2019) Asymptotic theory of penalized splines. Electronic Journal of Statistics, 13, 747–794.

Xiao L. (2020) Asymptotic properties of penalized splines for functional data. Bernoulli, 26, 2847–2875.

Xiao L, Li C, Checkley W and Crainiceanu C (2018) Fast covariance estimation for sparse functional data. Statistics and Computing, 28, 511–522. [PubMed: 29449762]

Yao F, Müller H-G, Clifford AJ, Dueker SR, Follett J, Lin Y, et al. (2003) Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate. Biometrics, 59, 676–685. [PubMed: 14601769]

Yao F, Müller H-G and Wang J-L (2005) Functional data analysis for sparse longitudinal data. Journal of the American Statistical Association, 100, 577–590.

Zhang X and Wang J-L (2016) From sparse to dense functional data and beyond. Annals of Statistics, 44, 2281–2321.

Zhao X, Tong X and Sun L (2012) Joint analysis of longitudinal data with dependent observation times. Statistica Sinica, 22, 317–336.

Zhu L, Sun J, Tong X, and Pounds S (2011) Regression analysis of longitudinal data with informative observation times and application to medical cost data. Statistics in Medicine, 30, 1429–1440. [PubMed: 21341298]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

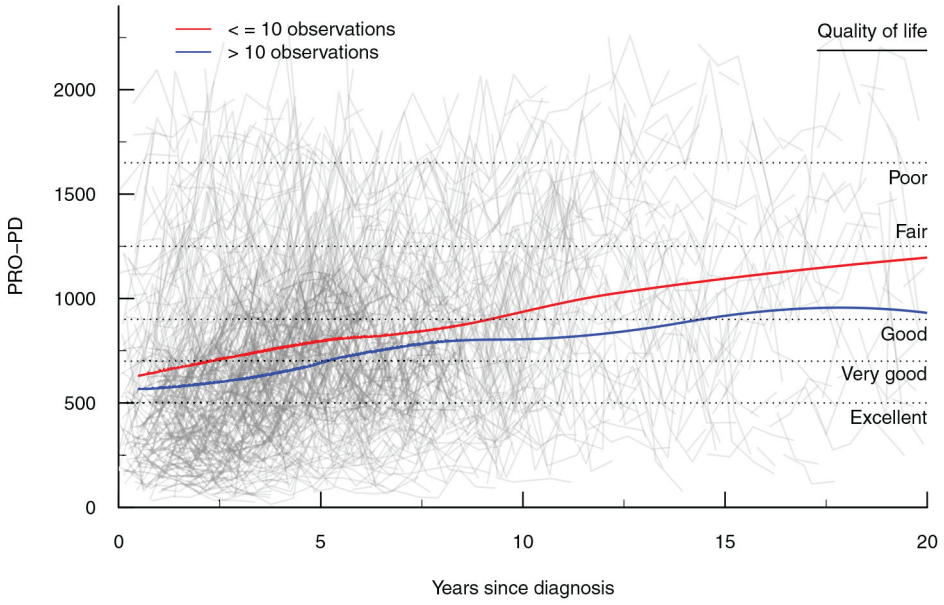


FIGURE 1.

The 371 trajectories of symptom severity along with smoothed mean profiles of subjects with less than or equal to 10 observations and of subjects with more than 10 observations. Reference lines indicate the average PRO-PD of subjects answering the prompt “In general, would you say your quality of life is:” with either “poor,” “fair,” “good,” “very good,” or “excellent.”

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

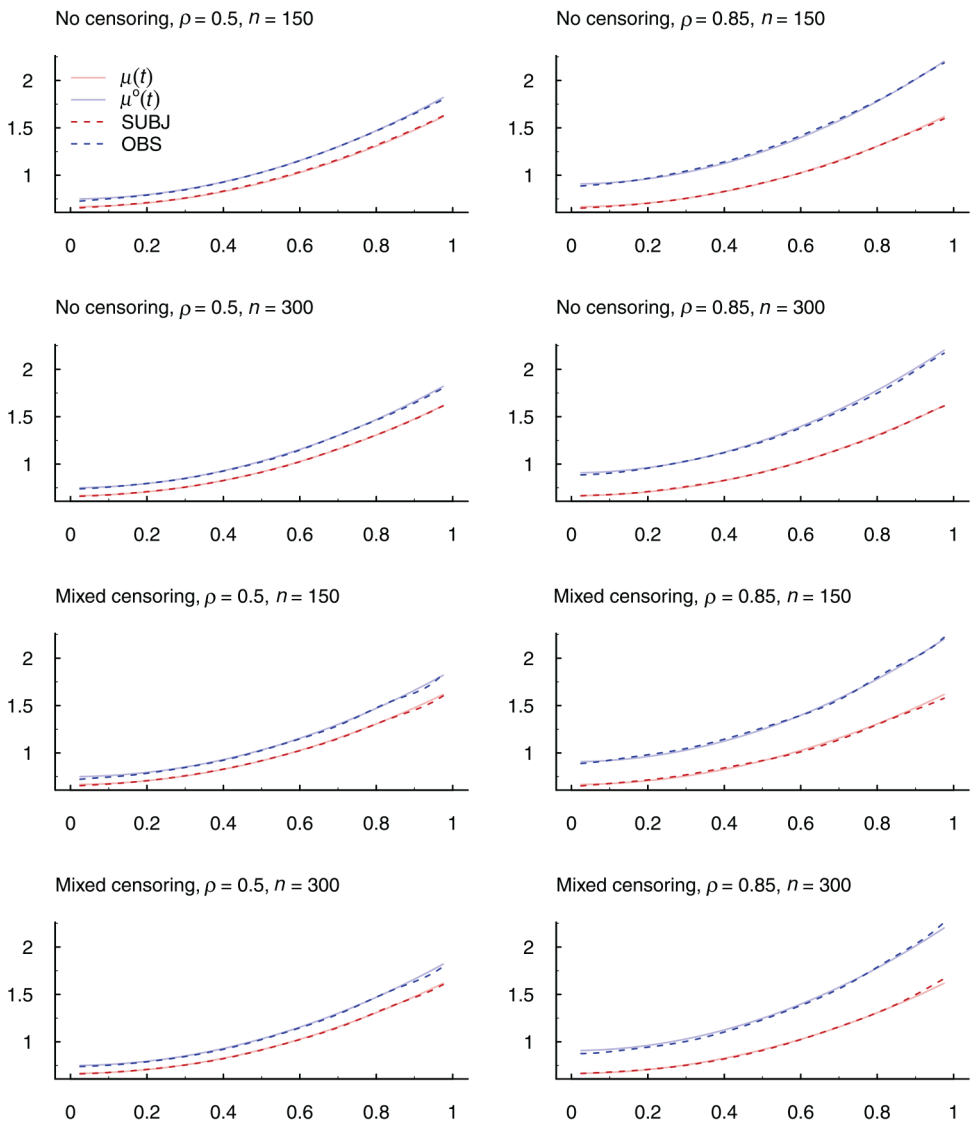


FIGURE 2.

Average of 200 estimates of the mean function using penalized splines with both OBS and SUBJ weights for simulation settings in which the frailty variable Z_i^* follows a gamma distribution with variance $\sigma^2_{Z^*} = 0.5$. The true mean function is $\mu(t) = t^2 + 2/3$ and $\mu^o(t)$ is the formula in (4) associated with OBS weights, specifically $\mu^o(t) = (t^2 + 2/3)(1 + \rho^2\sigma^2_{Z^*})$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

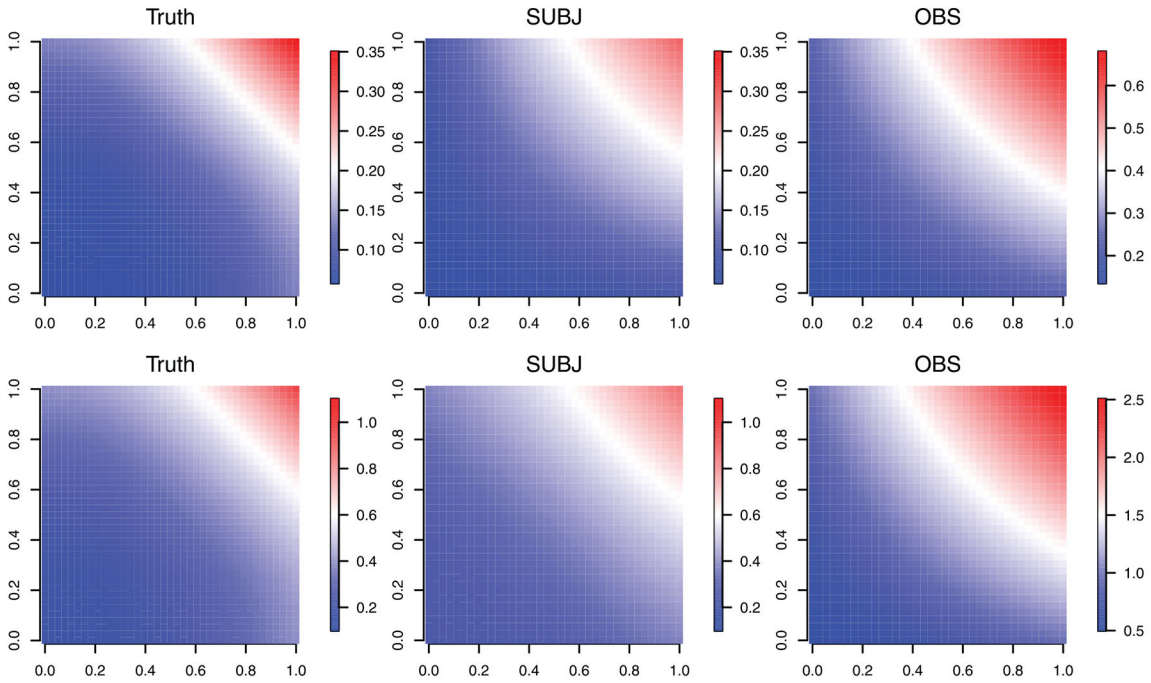


FIGURE 3.

True covariance function and corresponding estimates using penalized splines with both OBS and SUBJ weights for two simulated data sets in which the frailty variable Z_i^* follows a gamma distribution with variance $\sigma^2_{Z^*} = 0.5$, with $\rho = 0.5$ (top row) and $\rho = 0.85$ (bottom row).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

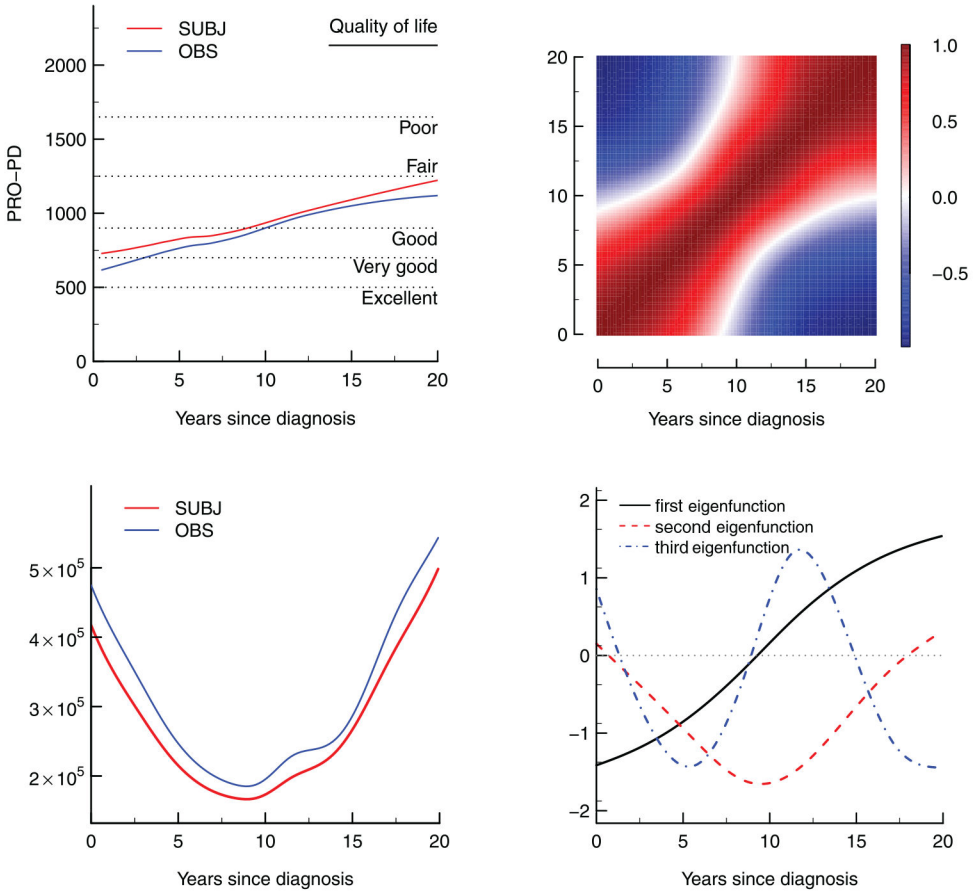


FIGURE 4.

(Top left) Mean function estimates using both the SUBJ and OBS weights compared to average PRO-PD by reported quality of life. (Top right) Estimated correlation function using the SUBJ weights. (Bottom left) Estimated variance function using both the SUBJ and OBS weights. (Bottom right) First three estimated eigenfunctions using the SUBJ weights.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

TABLE 1

Median (interquartile range) of integrated squared error for estimating the mean function using penalized splines with both OBS and SUBJ weights

			Gamma			
			No censoring		Mixed censoring	
ρ	n	$\sigma^2_{Z^*}$	OBS	SUBJ	OBS	SUBJ
0.50	150	0.5	0.011 (0.017)	0.001 (0.002)	0.022 (0.021)	0.002 (0.002)
		1.0	0.059 (0.060)	0.002 (0.002)	0.068 (0.039)	0.004 (0.004)
	300	0.5	0.016 (0.008)	0.000 (0.000)	0.011 (0.007)	0.001 (0.002)
		1.0	0.069 (0.037)	0.001 (0.001)	0.073 (0.042)	0.003 (0.003)
0.85	150	0.5	0.171 (0.109)	0.003 (0.003)	0.102 (0.116)	0.007 (0.010)
		1.0	0.557 (0.467)	0.007 (0.009)	0.409 (0.343)	0.016 (0.038)
	300	0.5	0.141 (0.056)	0.001 (0.002)	0.102 (0.057)	0.005 (0.008)
		1.0	0.596 (0.184)	0.003 (0.003)	0.593 (0.297)	0.011 (0.014)
			Lognormal			
			No censoring		Mixed censoring	
ρ	n	$\sigma^2_{Z^*}$	OBS	SUBJ	OBS	SUBJ
0.50	150	0.5	0.016 (0.024)	0.001 (0.002)	0.012 (0.027)	0.002 (0.003)
		1.0	0.073 (0.120)	0.001 (0.002)	0.075 (0.081)	0.005 (0.005)
	300	0.5	0.017 (0.014)	0.000 (0.000)	0.017 (0.017)	0.001 (0.001)
		1.0	0.048 (0.061)	0.001 (0.001)	0.054 (0.078)	0.002 (0.003)
0.85	150	0.5	0.141 (0.048)	0.003 (0.004)	0.098 (0.137)	0.009 (0.008)
		1.0	0.477 (0.344)	0.006 (0.009)	0.297 (0.684)	0.014 (0.012)
	300	0.5	0.102 (0.052)	0.001 (0.002)	0.107 (0.091)	0.005 (0.005)
		1.0	0.579 (0.348)	0.003 (0.003)	0.375 (0.283)	0.005 (0.009)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

TABLE 2

Median (interquartile range) of integrated squared error for estimating the covariance function using penalized splines with both OBS and SUBJ weights

			Gamma			
			No censoring		Mixed censoring	
ρ	n	$\sigma^2_{Z^*}$	OBS	SUBJ	OBS	SUBJ
0.50	150	0.5	0.018 (0.015)	0.000 (0.001)	0.011 (0.066)	0.001 (0.002)
		1.0	0.224 (0.653)	0.004 (0.009)	0.242 (1.141)	0.008 (0.014)
	300	0.5	0.009 (0.010)	0.000 (0.000)	0.006 (0.041)	0.001 (0.001)
		1.0	0.220 (0.472)	0.003 (0.004)	0.211 (0.789)	0.005 (0.005)
0.85	150	0.5	0.324 (0.581)	0.008 (0.014)	0.033 (0.226)	0.011 (0.018)
		1.0	5.359 (21.07)	0.043 (0.157)	5.568 (22.08)	0.077 (0.049)
	300	0.5	0.235 (0.399)	0.003 (0.003)	0.068 (0.577)	0.006 (0.013)
		1.0	3.022 (15.30)	0.045 (0.060)	2.556 (5.849)	0.054 (0.059)
			Lognormal			
			No censoring		Mixed censoring	
ρ	n	$\sigma^2_{Z^*}$	OBS	SUBJ	OBS	SUBJ
0.50	150	0.5	0.032 (0.109)	0.001 (0.002)	0.011 (0.160)	0.002 (0.004)
		1.0	2.885 (7.116)	0.011 (0.043)	0.341 (1.786)	0.012 (0.028)
	300	0.5	0.093 (0.113)	0.001 (0.002)	0.043 (0.088)	0.001 (0.002)
		1.0	0.614 (2.761)	0.005 (0.006)	0.747 (12.21)	0.010 (0.021)
0.85	150	0.5	0.691 (1.745)	0.007 (0.011)	0.652 (5.084)	0.031 (0.042)
		1.0	12.37 (74.19)	0.034 (0.106)	9.179 (96.96)	0.176 (0.238)
	300	0.5	0.550 (1.728)	0.004 (0.005)	0.562 (2.247)	0.009 (0.021)
		1.0	32.76 (93.44)	0.076 (0.133)	10.48 (129.7)	0.060 (0.078)