# Joint hierarchical Gaussian process model with application to personalized prediction in medical monitoring

**Leo L. Duan**[a], **Xia Wang**[b], **John P. Clancy**[c], and **Rhonda D. Szczesniak**[d,iD]

[a]Department of Statistical Science, Duke University, P.O. Box 90251, Durham, NC 27708, USA

[b]Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH 45221, USA

[c]Division of Pulmonary Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA

[d]Division of Biostatistics & Epidemiology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave (MLC 5041), Cincinnati, OH 45229, USA

## Abstract

A two-level Gaussian process (GP) joint model is proposed to improve personalized prediction of medical monitoring data. The proposed model is applied to jointly analyze multiple longitudinal biomedical outcomes, including continuous measurements and binary outcomes, to achieve better prediction in disease progression. At the population level of the hierarchy, two independent GPs are used to capture the nonlinear trends in both the continuous biomedical marker and the binary outcome, respectively; at the individual level, a third GP, which is shared by the longitudinal measurement model and the longitudinal binary model, induces the correlation between these two model components and strengthens information borrowing across individuals. The proposed model is particularly advantageous in personalized prediction. It is applied to the motivating clinical data on cystic fibrosis disease progression, for which lung function measurements and onset of acute respiratory events are monitored jointly throughout each patient's clinical course. The results from both the simulation studies and the cystic fibrosis data application suggest that the inclusion of the shared individual-level GPs under the joint model framework leads to important improvements in personalized disease progression prediction.

### Keywords

bayesian methods; biostatistics; forecasting; longitudinal data; prediction; regression

## 1 Introduction

In clinical care and research of chronic diseases, an important task is to monitor and predict disease progression based on data from multiple biomedical outcomes. Cystic fibrosis (CF),

as a motivating example, is a lethal autosomal disease, marked by progressive loss of lung function over the lifespan. A variety of biomedical measures and outcomes are collected at each patient's clinical visit, and these data are used to inform CF clinical studies. CF clinicians have great interest in prognostic models of monitoring and predicting disease progression. Particularly, there is an increasing emphasis on detection of severe CF disease at its early stage and prediction of its clinical course. Although the median age of survival has markedly improved, patients still continue to experience declining lung function and acute respiratory events during adolescence and early adulthood. Accurate prediction models could be used as a prognostic tool for more timely treatment of respiratory decline.

Despite the potential clinical utility of prediction models in biomedicine, it is widely known that patient-level extrapolation of disease progression is difficult. On one hand, the sparsity and irregularity of the data collection times pose great challenges in extracting sensible and flexible estimates with a reasonable level of confidence. On the other hand, the statistical models used in clinical practice are often based on unrealistic assumptions (such as a simple linear slope) and overlook or even violate many intrinsic properties inherent in the underlying progression of the disease (Szczesniak et al., 2016). Among these properties, nonlinearity and individual heterogeneity, which are pervasive challenges in modeling CF lung function decline (Szczesniak et al., 2017), are of particular importance for modeling the patient-level disease progression. Inspired by the developments in the field of joint modeling, we propose a two-level Gaussian process (GP) joint model for multiple longitudinal outcomes monitored by clinicians, both continuous and discrete. The goal is to address the nonlinear prediction needs in important biomedical outcomes on the personalized level.

Our key contributions focus on two aspects: one is for the data structure we study in the joint model and the other is for the model structure we propose in the joint model. The joint model conveniently accommodates the need to study multiple outcomes simultaneously. Particularly, for the CF data of interest, there are two important severity markers whose association has been evidenced in clinical practice. These two biomarkers are the forced expiratory volume in one second (hereafter, $FEV_1$) and the occurrence of acute respiratory events, known as pulmonary exacerbations (PEs). Details on both biomarkers and previous studies on them are discussed in Section 2. Both $FEV_1$ and PE are longitudinally collected (Figure 1). $FEV_1$ is a continuous measurement, and PE measurements are binary data on recurrent events. Though many longitudinal studies employ joint models to examine a single binary outcome and a number of longitudinal measurements (Horrocks and van Den Heuvel, 2009; Zhang et al., 2014; Li et al., 2015), they are not commonly used in studying longitudinal observations on both continuous and potentially recurring binary outcomes. Furthermore, we follow the shared parameter model approach to form the joint model (Taylor et al., 2013). The key innovation in our model structure is that we introduce the hierarchical GPs in the joint model, where not only the mean processes but also the shared parameters are modeled through GPs.

GP provides a useful paradigm for nonlinear interpolation and allows for borrowing strengths from data collected at all time points (Rasmussen & Williams, 2006). Compared to spline based methods, GP modeling does not require prior knowledge of important time

points (knots). Its estimation is straightforward, and the predictive distribution is simple: the multivariate normal distribution associated with a GP model allows direct quantification of the uncertainty in the conditional variance. GPs have become particularly popular with the availability of fast estimation procedures such as the work by Banerjee et al. (2013). It has observed wide applications for many years in machine learning (Rasmussen & Williams, 2006) and spatial statistics (Cressie, 1988). However, its application has not been adequate in addressing the urgent demand for personalized prediction in biomedical settings.

In the proposed joint hierarchical Gaussian process model (JHGP), we adapt the GP to the inherently multilevel structure of the clinical data. On the population level, each of the two outcomes has its own nonlinear trend via GP modeling; on the individual level, each subject has a unique GP, which serves as the shared parameter to permit information borrowing across outcomes. Assuming the individual GPs are independently and identically distributed overcomes the challenges posed by the small sample size for a single subject and allow more stable estimates of the correlation parameter that characterizes the individual patterns. The model estimation is easy to be carried out through Hamiltonian Monte Carlo using Rstan (Stan Development Team, 2016). The direct goal of our proposed JHGP model is to address the nonlinear prediction needs in both $FEV_1$ and PE on the personalized level. The proposed model framework has wide applications in many other medical monitoring situations beyond the CF disease progression monitoring discussed here.

The remainder of the article is organized as follows. The motivating data set, including details of the clinical markers and previous research results, is described in Section 2. Section 3 presents the proposed JHGP model, including the personalized predictive distribution under the proposed model, prior distributions and posterior sampling. Section 4 shows the results from the simulation studies and assessments of the model prediction performance. Source code to reproduce these findings is included as Supporting Information. Section 5 shows results from applying the JHGP model to the CF clinical data set. Concluding remarks and discussion are presented in Section 6.

## 2 Motivating clinical data

### 2.1 Monitoring cystic fibrosis disease progression

Our JHGP model was motivated by medical monitoring data from the United States Cystic Fibrosis Foundation Patient Registry, which has provided epidemiologic surveillance on CF outcomes for more than 40 years. We obtained data on a sample of 38 patients with a total of 818 entries of observations. Data were analyzed from patients who had PE and $FEV_1$ data observed anytime from 6 to 18 years of age. This age range was selected for the following reasons. First, it was of interest to focus on data collected from childhood to adulthood, covering the most common interval in which there is a prolonged drop in lung function. Second, patients younger than 6 years of age were excluded due to potentially unreliable pulmonary function testing. We used the records from the calendar time January 1, 2003 to December 31, 2013. Encounter-level data available from 2003 onward were considered because PE events were consistently documented in the registry beginning in 2003. Data acquired after lung transplant were excluded. Quarterly ages were used as the time indices for the modeling.

### 2.2 Lung function measurement

As maintaining lung function is essential for survival in individuals with CF, lung function is routinely assessed in CF research and clinical practice. Lung function is monitored as forced expiratory volume in 1 second of percent predicted (denoted as $FEV_1$). $FEV_1$ is obtained as follows. Once a patient performs pulmonary function testing at a clinical encounter, a percent predicted value for his lung function is obtained based on his age, height, gender and race using standard reference equations (Wang et al., 1993; Hankinson et al., 1999). These computed values for each patient encounter are made available from the Cystic Fibrosis Foundation Patient Registry. The $FEV_1$ data in this sample range from 5% to 125% predicted.

As the most routine clinical measure of lung function in CF patients, $FEV_1$ data have motivated various longitudinal models, including recent development on individualized prediction (Taylor-Robinson et al., 2012) and nonlinear fitting of age-related $FEV_1$ progression (Szczesniak et al., 2013; Moss et al., 2016). Studies using CF registries have demonstrated that rapid changes in $FEV_1$ often occur during adolescence and adulthood (Vandenbranden et al., 2012) and that decline is nonlinear with age (Szczesniak et al., 2013). Previous statistical models being used in CF clinical practice for individualized prediction of such markers tend to feature simple functions (e.g., linear slope) or arbitrary rules (e.g., a decrease from maximum $FEV_1$ during the prior year of at least 10% predicted) (Szczesniak et al., 2016).

### 2.3 Pulmonary exacerbation event

Besides $FEV_1$, data on acute respiratory events, known as PEs, are also collected. Clinical diagnosis of a PE is often based on increased pulmonary symptoms, a loss of lung function, energy or weight and physical changes. It can occur multiple times in an individual CF patient. In our analysis, an event is defined as a PE event if it has a recorded treatment of PE with intravenous antibiotics in the registry.

Occurrence of PE has been associated with, among other adverse health outcomes, more rapid decline in $FEV_1$ (Konstan et al., 2007; Sanders et al., 2011). A converse association has been found, with PE occurrence being linked to a failure to recover to pre-PE or "baseline" $FEV_1$ (Sanders et al., 2010). Although PE onset is typically incorporated as a covariate in longitudinal modeling of $FEV_1$, it is a separate event process often clinically diagnosed by changes in lung function. Current joint modeling approaches used in monitoring CF disease progression have not yet incorporated data on PE. Therefore, it is desirable to construct a more informative, efficient approach to a simultaneous prediction of both occurrence of PE events and changes in $FEV_1$.

## 3 Model framework

### 3.1 Joint hierarchical Gaussian process model

Let $\{ Y_{it}, R_{it} \}$ be the observed outcomes for subject $i$ at time $t$, where $Y_{it}$ is the continuous longitudinal outcome and $R_{it}$ is the binary longitudinal outcome. A shared parameter model is prescribed for these data:

$$Y_{\text{it}} \sim \text{N}(\mu_t^y + \psi_{\text{it}}, \sigma^2),$$

$$R_{\text{it}} \sim \text{Bern}\{g(\mu_t^R + c\psi_{\text{it}})\},$$

where N($a$, $b$) represents a normal distribution with a mean $a$ and a variance $b$, Bern($p$) represents a Bernoulli distribution with the probability of success $p$, $g(\cdot)$ is the logistic link function $g(\lambda_{it}) = \{\exp(\lambda_{it})\}/\{1 + \exp(\lambda_{it})\}$ with $\lambda_{\text{it}} = \mu_t^R + c\psi_{\text{it}}$, and $c$ is a scalar in $[-1, 1]$. Covariate effects can be easily added in both submodels, which are temporarily omitted here for focus on exploring the hierarchical GPs.

The uniqueness of this model is that $\mu_t^y$ and $\mu_t^R$ are parameters on the population level and, hence, are the same for all subjects; $\psi_{it}$ is not only the individual deviation from the common trend but also acts as a shared parameter that induces correlation across the two outcomes.

Let $\boldsymbol{\mu}^y$ and $\boldsymbol{\mu}^R$ represent the two population mean processes $\{\mu_t^y\}$ and $\{\mu_t^R\}$, respectively, and $\boldsymbol{\psi}_i$ represents the individual heterogeneity, $i = 1,\ldots, n$. Further denote $\boldsymbol{\lambda}_i = \boldsymbol{\mu}^R + c \cdot \boldsymbol{\psi}_i$. To induce nonlinearity, GPs are used for both the population mean processes and the individual heterogeneities as

$$\boldsymbol{\mu}^y \sim \text{GP}(\mathbf{0}, \textstyle\sum^y),$$

$$\boldsymbol{\mu}^R \sim \text{GP}(\mathbf{0}, \textstyle\sum^R),$$

$$\psi_i \overset{\text{iid}}{\sim} \text{GP}(\mathbf{0}, \textstyle\sum), \quad \text{for } i = 1 \ldots n,$$

where $\mathbf{X} \sim \text{GP}(\mathbf{a}, \mathbf{b})$ means that on any finite time point set $\{t_1, \ldots t_k\}$, the joint distribution of $\mathbf{X}_k = \{X(t_1), \ldots, X(T_k)\}'$ is a k-variate normal distribution $\mathbf{N}(\mathbf{a}_k, \mathbf{b}_k)$ with the mean vector $\mathbf{a}_k = \{a(t_1), \ldots, a(t_k)\}'$ and covariance matrix $\mathbf{b}_k$ with the $(\ell j)^{th}$ entry as $b(t_\ell\, t_j)$, $\ell j = 1, \ldots, k$.

This is a hierarchical GP since $\mathbf{Y}_i | \boldsymbol{\mu}^y \sim \text{GP}(\boldsymbol{\mu}^y, \Sigma + \mathbf{I} \cdot \sigma^2)$ with $\mathbf{I}$ denoting the identify matrix and $\boldsymbol{\lambda}_i | \boldsymbol{\mu}^R \sim \text{GP}(\boldsymbol{\mu}^R, c^2\Sigma)$ where the population mean processes $\boldsymbol{\mu}^y$ and $\boldsymbol{\mu}^R$ are each another GP. Jointly, the shared GP process induces the inter-outcome covariance: $\text{Cov}(\mathbf{Y}_i, \boldsymbol{\lambda}_i) = c\Sigma$.

The population mean processes is $\boldsymbol{\mu}^y$ and $\boldsymbol{\mu}^R$ are common to all subjects and, thus, can be estimated with high precision. A reliable estimation of the individual GP $\boldsymbol{\psi}_i$, though, is challenging due to the limited amount of individual data. To maximize information borrowing across individuals and across outcomes, we allow the common $\Sigma$ among the GP $\boldsymbol{\psi}_i$s' and also include $\boldsymbol{\psi}_i$'s as a shared component in the joint model. Both the simulation

study and the real data application show that this model setup improves the flexibility and the precision in the personalized prediction.

### 3.2 Personalized predictive distribution

Let $\mathbf{Z}_{it}$ denote the joint vector of $\{ Y_{it}, \boldsymbol{\lambda}_{it} \}'$ and $\mathbf{Z}_{i,\mathbf{T}} = \{ \{ Y_{it} \}_{t \in \mathbf{T}}, \{ \boldsymbol{\lambda}_{it} \}_{t \in \mathbf{T}} \}'$, where $\mathbf{T}$ is a collection of time points. The key advantage of the hierarchical Gaussian process is the personalized predictive distribution: given the data $t_i \leq T$, we predict $\mathbf{Z}_{iti}$ at $T < t_i \leq t_{\max}$, where $t_{\max}$ is the maximally observed time in the population. Let $\mathbf{T}_1$ be the recorded set of time points in $t_i \leq T$ for subject $i$, and let $\mathbf{T}_2$ be the time points of interest in $T < t_i \leq t_{\max}$. Using matrix notation,

$$\mathbf{Z}_{i,\mathbf{T}_2} | \mathbf{Z}_{i,\mathbf{T}_1} \sim \mathrm{N} \{ \boldsymbol{\mu}_{\mathbf{T}_2} + \mathbf{K}_{\mathbf{T}_2,\mathbf{T}_1} \mathbf{W}_{\mathbf{T}_1,\mathbf{T}_1}^{-1} (\mathbf{Z}_{i,\mathbf{T}_1} - \boldsymbol{\mu}_{\mathbf{T}_1}), \mathbf{W}_{\mathbf{T}_2,\mathbf{T}_2} - \mathbf{K}_{\mathbf{T}_2,\mathbf{T}_1} \mathbf{W}_{\mathbf{T}_1,\mathbf{T}_1}^{-1} \mathbf{K}_{\mathbf{T}_1,\mathbf{T}_2} \},$$

where $\boldsymbol{\mu}_{\mathbf{T}_j}$ is the joint vector of $(\boldsymbol{\mu}_{\mathbf{T}_j}^y, \boldsymbol{\mu}_{\mathbf{T}_j}^R)', j = 1, 2, \mathbf{W} = \begin{bmatrix} \Sigma + \mathbf{V} & c\Sigma \\ c\Sigma & \Sigma \end{bmatrix}, \mathbf{K}_{\mathbf{T}_2,\mathbf{T}_1}$ is the correlation matrix between $\mathbf{m}\psi_{/\mathbf{T}_2}$ and $\mathbf{m}\psi_{/\mathbf{T}_1}$ and $\mathbf{V}$ is a diagonal matrix $\sigma^2 \cdot \mathbf{I}$. Note that the prediction mean is the sum of two parts: (i) the nonlinear estimate $\boldsymbol{\mu}_{\mathbf{T}_2}$ obtained at the population level and (ii) the individual extrapolation based on its past deviation from the common trend $(\mathbf{Z}_{i;\mathbf{T}_1} - \boldsymbol{\mu}_{\mathbf{T}_1})$.

To illustrate the improvement in predictive accuracy with a two-level GP model, compared to a one-level GP model, extrapolation results taken from a simulated data set are presented in Figure 2. The three models shown are the model with only individual GPs, the model with the single population GP and the hierarchical GP model. The test samples include simulated data from $n = 50$ subjects, with the curve for each individual generated as the sum of a common random curve and an individual AR(1) process with $\rho = -0.8$. One subject was randomly selected, and the second half of the observed data points was removed. In the first test, an individual GP with the AR(1) (Figure 2, leftmost panel) is fitted to each subject. Although the fitted line shows that the model has adequate flexibility, it performs poorly in terms of extrapolation: the fitted line in the forecast range ($t = 15$ and forward) rapidly reverts to the constant mean due the underestimation of the autocorrelation. In the second test, all subjects are fitted with only the population GP with the squared exponential covariance; the prediction benefits from the similarity of trajectories and, hence, obtains some improvement (Figure 2, middle panel). In the last test, the hierarchical GP is used and leads to more gains, especially on the points near $t = 15$ (Figure 2, rightmost panel). Because the prediction is based on the individual level, the credible band is narrower than the population based GP when used alone.

### 3.3 Prior distributions and posterior sampling

A Bayesian method was employed to estimate the model parameters. Here, the squared exponential covariance $\phi \exp(-\rho|t_1 - t_2|^2)$ was used in $\Sigma^y$ and $\Sigma^R$ to induce smoothness on the population level. For $\Sigma$, Brownian motion $\phi \min(t_1, t_2)$ or an autoregressive AR(1) structure with the $(i, j)^{th}$ component as $\phi \rho^{|t_j - t_i|}$ was used to capture the fluctuation. Different covariance structures can be included based on the data structure. It is worth noting that $\phi$ and $\rho$ are allowed to vary across the three processes defined in Section 3.1, but this notation has been suppressed for ease of exposition. Weakly informative proper priors were used for

all the scale parameters, $\phi \sim C^+(5)$, $\sigma^2 \sim C^+(5)$ or an inverse Gamma prior IG(2, 1) with a mean 1 and an infinite variance, $\rho \sim C^+(5)$ in $\Sigma^y$ and $\Sigma^R$ and $\rho \sim$ Uniform(−1, 1) in $\Sigma$ with AR(1), where $C^+(5)$ denotes a half Cauchy distribution with a scale of 5. The posterior sampling was carried out in Hamiltonian Monte Carlo via the software Rstan (Stan Development Team, 2016). The simulation file and source code are available as Supporting Information.

## 4 Simulation studies

### 4.1 Model estimation

To demonstrate estimation of the JHGP model, the following simulations were conducted. We simulated data sets with 50 subjects ($i = 1, \ldots, 50$), each with observed $\mathbf{Y}_i$ and $\mathbf{R}_i$ at 25 time points ($t = 1, \ldots, 25$) under three parameter settings, Sim 1 ($\rho = -0.8$, $c = 0.9$), Sim 2 ($\rho = -0.5$, $c = -0.3$) and Sim 3 ($\rho = -0.1$, $c = 0.01$), corresponding to three different levels of autocorrelation and inter-outcome association. Under each of the simulation settings, the data were simulated using the following framework:

> **Unobserved population and individual GPs:**
> $\mu_t^y = 50 \sin\{(t - 20)/100\} \cos\{-(t - 10)/15\}, \quad t = 1, \ldots, T,$
> $\mu_t^R = 4 \sin\{(t - 10)/5\} \cos\{t/10\}, \quad t = 1, \ldots, T,$
> $\psi_i \overset{\text{indep}}{\sim} \mathbf{N}(\mathbf{0}, \Sigma)$, where $\Sigma$ is based on AR(1) with parameter $\rho$;
> **Observed contionuous and binary longitudinal data:**
> $\mathbf{Y}_i \sim \mathbf{N}(\boldsymbol{\mu}^y + \psi_i, 0.01\mathbf{I}),$
> $\mathbf{R}_i \sim \text{Bern}\{g(\boldsymbol{\lambda}_i)\}, \quad$ where $\boldsymbol{\lambda}_i = \boldsymbol{\mu}^R + c\psi_i.$

The proposed JHGP model was fit to these three sets of data. The estimation of parameters is shown in Table 1; plots of estimated population processes ($\boldsymbol{\mu}^y, \boldsymbol{\mu}^R$) and a randomly selected individual processes ($\boldsymbol{\lambda}_i$) for subject $i$ are shown in Figure 3. The model correctly identified the values of autocorrelation $\rho$ and the association parameter $c$. Moreover, the unobserved population processes ($\boldsymbol{\mu}^y, \boldsymbol{\mu}^R$) were accurately estimated. The estimates for $\boldsymbol{\lambda}_i$, which are hidden behind the binary outcomes $\mathbf{R}_i$, show high correlation with the true values.

### 4.2 Different covariance functions for individual deviation

The choice of covariance function affects the behavior of extrapolation curves. A simple comparison between stationary (AR(1) process) and nonstationary (Brownian motion process) covariance functions was conducted. As shown in Figure 4, the stationary AR(1) process with a negative $\rho$ tends to oscillate around the mean. This property is useful if the subject-specific trajectories tend to have similar progression while still accounting for the "saw-tooth" shape found in $FEV_1$ profiles in the CF data. On the other hand, Brownian motion as a martingale process always shows a constant difference from the mean process. This can reflect the notion that a loss or gain at a certain time is permanent for an individual.

### 4.3 Robustness in association detection

In the joint model, the association between two responses is established by the shared individual GP, and its strength is measured by the association parameter $c$. We now assess its robustness under various levels of interference.

The same framework as in Section 4.1 was used to generate test samples with $c = 0.5$, except that an additional noise vector $\tau_i \sim \mathrm{N}(0, \sigma_\tau^2 \cdot \mathbf{I})$ was added into $\lambda_i$. We then gradually increase $\sigma_\tau^2$ in order to disturb the estimation of $c$. The noise-signal ratio is controlled by $\sigma_\tau^2/(c^2 \phi_\psi)$, where $\phi_\psi$ is the scale parameter in the GP for $\psi_i$. The results are shown in Table 2. The numerical estimates only start to degrade around noise-signal ratio of 8.0, yet the association remains significant until the magnitude reaches 32.0. These findings suggest that the model is very robust in detecting the association between two responses in the presence of disturbance.

### 4.4 Forecasting performance

We censored each subject in Sim1 data set (See Section 4.1) using random time $C_i = \min\{t_{\max}, t_c\}$ where $t_c \sim \mathrm{Uniform}[0, 2t_{\max}]$ and $t_{\max}$ is the largest recorded time. This mechanism results in censoring in about 50% of the subjects at random time points. Figure 5 plots the predicted values against the true values of $\mathbf{Y}$ and $g(\lambda)$. The predicted values are highly correlated to the true values, with a Pearson correlation 0.86 for $\mathbf{Y}$ and 0.73 for $g(\lambda)$.

## 5 Application in cystic fibrosis medical monitoring

### 5.1 Analysis cohort selection

We applied the model to characterize the association between the longitudinal marker of lung function, $FEV_1$, and the occurrence of acute respiratory events, PEs, for patients with CF disease. The primary goal was to provide longitudinal predictions for each clinical marker. The analysis cohort and study variables are described in Section 2. Among the 38 patients included in the analysis cohort, 19 patients were randomly selected to assess prediction performance. For each patient in this sub-cohort, the more recent 50% of his or her observations (both $FEV_1$ and PE) were masked. Having this sub-cohort results in a training and testing split of about 75% and 25%, respectively. We use a random intercept for each patient (indexed by $i$) to offset the starting difference.

### 5.2 Parameter estimation

The JHGP model detects a strong autocorrelation ($\rho = -0.82 \pm 0.05$) in the shared Gaussian process $\psi_i$, $i = 1, \ldots, 38$; the variations of $FEV_1$ and the PE probability have a weak negative inter-outcome correlation ($c = -0.29 \pm 0.13$). The fitted $FEV_1$ and PE probability are shown, along with corresponding 95% credible intervals, in Figure 6. The population mean processes and individual AR(1) processes appear to be reasonably estimated. Among which, the population-level estimate of age-related $FEV_1$ progression is consistent with what was found in previous CF studies using penalized splines (Szczesniak et al., 2013; Szczesniak et al., 2017). The mean evolution of lung function exhibits a gradual decline in earlier ages and subsequently becomes more progressive as patients enter adolescence and early adulthood.

The risk of PE appears heightened around ages 10–11, which corresponds to the time preceding more progressive lung function decline. The overall PE risk has smooth estimates, yet it does not resemble any common parametric distribution. The stochasticity of individual variation in the PE risk is also captured by the model.

### 5.3 Prediction performance

The predicted $FEV_1$ values are plotted against the corresponding observed yet masked values in Figure 7, which shows all points lie around the 45° reference line. We further evaluate the results and compare the effects of the hierarchical model with the nonhierarchical model using the median absolute deviation, root mean squared error and Pearson correlation. The three evaluation metrics median absolute deviation, root mean squared error and Pearson correlation are 6.43, 8.47 and 0.93, respectively, by the JHGP model, compared to 8.66, 10.63 and 0.89, respectively, by the single population GP model. Although the results from the population GP seem adequate for predicting the future trend, including the individual GPs apparently further improves the estimation accuracy.

### 5.4 Assessment of sensitivity and specificity

Sensitivity and specificity of the PE predictions using three different model settings were examined using Receiver Operating Characteristic curves (Figure 8). The models included the JHGP of $FEV_1$ and PE, a single logistic model that included a population-based GP to model PE events and a traditional linear-logistic model of PE events. In the later two single models, $FEV_1$ was included as a covariate. Both the JHGP and the single logistic GP models had higher AUC than the traditional linear-logistic model, indicating that the nonlinear attributes of the GP models improved predictive accuracy. The joint model with hierarchical GPs had better sensitivity, compared to the single GP model.

## 6 Discussion

We have developed a joint model with hierarchical GPs in each submodel to accommodate the analysis and prediction of monitored clinical data. We consider a two-level design: the first (population) level describes the overall progression of the markers; the second stochastic (individual) level captures the finer, personalized variation in both continuous and binary outcomes. The hierarchical design is conceptually simple and allows information to be shared across individuals as well as different outcome types. This leads to an attractive form of the predictive distribution that joins population change and personal history.

Several extensions to this work can be made in the future. If the longitudinal data exhibit several distinct types of progression, then the population hierarchy can be replaced with a mixture of Gaussian processes as suggested in Shi et al. (2005). Additional levels could be incorporated in the model to accommodate a more refined hierarchical structure in the data; for example, different medical centers tend to have distinct practice patterns. Our JHGP application to the CF data suggests the presence of more heterogeneity between longitudinal $FEV_1$ trajectories, compared to PE hazard curves in Figure 6. This has not been previously characterized in the CF epidemiological literature, but additional work could be performed to standardize the processes and assess heterogeneity in this context. Findings on prediction

performance for $FEV_1$ suggest the possibility of bimodality (Figure 7), but further assessments with additional covariate information, or perhaps an extension of the JHGP model to accommodate latent classes, would be needed to determine if results are indicative of subgroups. Also, the joint hierarchical GP modeling framework is general and can be extended to more than two outcomes. Although a pediatric cohort was examined in the current work, such an extension could accommodate censoring arising from death if studying the broader adult CF cohort or other data in which there is a potential for survivorship bias.

Often clinicians are interested in survival times of recurrent events. This would require, as in the motivating CF example, joint modeling of the longitudinal measurements on $FEV_1$ as well as the survival times of the recurrent PE events. The implementation of the proposed hierarchical GP model framework to this situation is currently under study. Furthermore, separate assessments of each component in the joint model help assess the goodness of fit for each component and is of clinical importance (Zhang et al., 2014; Zhang et al., 2017). It is of practical and theoretical interests to investigate how to achieve separate assessments in this JHGP model based on the work of Zhang et al. (2017) as a direct application of their criteria faces computational challenges due to the inclusion of the shared GPs component.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

## References

Banerjee A, Dunson DB, Tokdar ST. Efficient Gaussian process regression for large data sets. Biometrika. 2013; 100(1):75–89. [PubMed: 23869109]

Cressie N. Spatial prediction and ordinary kriging. Mathematical Geology. 1988; 20(4):405–421.

Hankinson J, Odencrantz J, Fedan K. Spirometric reference values from a sample of the general U.S. population. American Journal of Respiratory & Critical Care Medicine. 1999; 159:179–187. [PubMed: 9872837]

Horrocks J, van Den Heuvel M. Prediction of pregnancy: a joint model for longitudinal and binary data. Bayesian Analysis. 2009; 4:523–538.

Konstan M, Morgan W, Butler S, Pasta D, Craib M, Silva S, Stokes D, Wohl M, Wagener J, Regelmann W. Risk factors for rate of decline in forced expiratory volume in one second in children and adolescents with cystic fibrosis. Journal of Pediatrics. 2007; 151:134–139. [PubMed: 17643762]

Li D, Wang X, Song S, Zhang N, Dey DK. Flexible link functions in a joint model of binary and longitudinal data. Stat. 2015; 4:320–330.

Moss A, Juarez-Colunga E, Nathoo F, Wagner B, Sagel S. A comparison of change point models with application to longitudinal lung function measurements in children with cystic fibrosis. Statistics in Medicine. 2016; 35(12):2058–2073. [PubMed: 27118629]

Rasmussen, CE., Williams, CKI. Gaussian Processes for Machine Learning. MIT Press; Cambridge, MA: 2006.

Sanders D, Bittner R, Rosenfeld M, Hoffman L, Redding G, Goss C. Failure to recover to baseline pulmonary function after cystic fibrosis pulmonary exacerbation. American Journal of Respiratory and Critical Care Medicine. 2010; 182:627–632. [PubMed: 20463179]

Sanders DB, Bittner RC, Rosenfeld M, Redding GJ, Goss CH. Pulmonary exacerbations are associated with subsequent FEV1 decline in both adults and children with cystic fibrosis. Pediatric Pulmonology. 2011; 46(4):393–400. [PubMed: 20967845]

Shi JQ, Murray-Smith R, Titterington D. Hierarchical Gaussian process mixtures for regressionx. Statistics and Computing. 2005; 15(1):31–41.

Stan Development Team. [Last accessed: 2 Nov, 2017] Rstan: the R interface to Stan. 2016. R package version 2.14.1http://mc-stan.org

Szczesniak R, Heltshe S, Stanojevic S, Mayer-Hamblett N. Use of FEV1 in cystic fibrosis epidemiologic studies and clinical trials: a statistical perspective for the clinical researcher. Journal of Cystic Fibrosis. 2017; 16:318–326. [PubMed: 28117136]

Szczesniak R, Li D, Su W, Brokamp C, Pestian J, Seid M, Clancy J. Phenotypes of rapid cystic fibrosis lung disease progression during adolescence and young adulthood. American Journal of Respiratory & Critical Care Medicine. 2017; 196:471–478. [PubMed: 28410569]

Szczesniak R, McPhail G, Li D, Amin R, Clancy J. Predicting future lung function decline in cystic fibrosis patients: statistical methods and clinical connections. Pediatric Pulmonology. 2016; 51:217–218. [PubMed: 26677806]

Szczesniak RD, McPhail GL, Duan LL, Macaluso M, Amin RS, Clancy JP. A semiparametric approach to estimate rapid lung function decline in cystic fibrosis. Annals of Epidemiology. 2013; 23(12):771–777. [PubMed: 24103586]

Taylor JM, Park Y, Ankerst DP, Proust-Lima C, Williams S, Kestin L, Bae K, Pickles T, Sandler H. Real-time individual predictions of prostate cancer recurrence using joint models. Biometrics. 2013; 69:206–213. [PubMed: 23379600]

Taylor-Robinson D, Whitehead M, Diderichsen F, Olesen H, Pressler T, Smyth R, Diggle P. Understanding the natural progression in % FEV1 decline in patients with cystic fibrosis: a longitudinal study. Thorax. 2012; 67(10):860–866. [PubMed: 22555277]

Vandenbranden SL, McMullen A, Schechter MS, Pasta DJ, Michaelis RL, Konstan MW, Wagener JS, Morgan WJ, McColley SA. Lung function decline from adolescence to young adulthood in cystic fibrosis. Pediatric Pulmonology. 2012; 47(2):135–143. [PubMed: 22241571]

Wang X, Dockery D, Wypij D, Fay M, Ferris BJ. Pulmonary function between 6 and 18 years of age. Pediatric Pulmonology. 1993; 15:75–88. [PubMed: 8474788]

Zhang D, Chen MH, Ibrahim JG, Boye ME, Shen W. Bayesian model assessment in joint modeling of longitudinal and survival data with applications to cancer clinical trials. Journal of Computational and Graphical Statistics. 2017; 26:121–133. [PubMed: 28239247]

Zhang D, Chen MH, Ibrahim JG, Boye ME, Wang P, Shen W. Assessing model fit in joint models of longitudinal and survival data with applications to cancer clinical trials. Statistics in Medicine. 2014; 33:4715–4733. [PubMed: 25044061]

Zhang N, Chen H, Zou Y. A joint model of binary and longitudinal data with non-ignorable missingness, with application to marital stress and late-life major depression in women. Journal of Applied Statistics. 2014; 41:1028–1039.
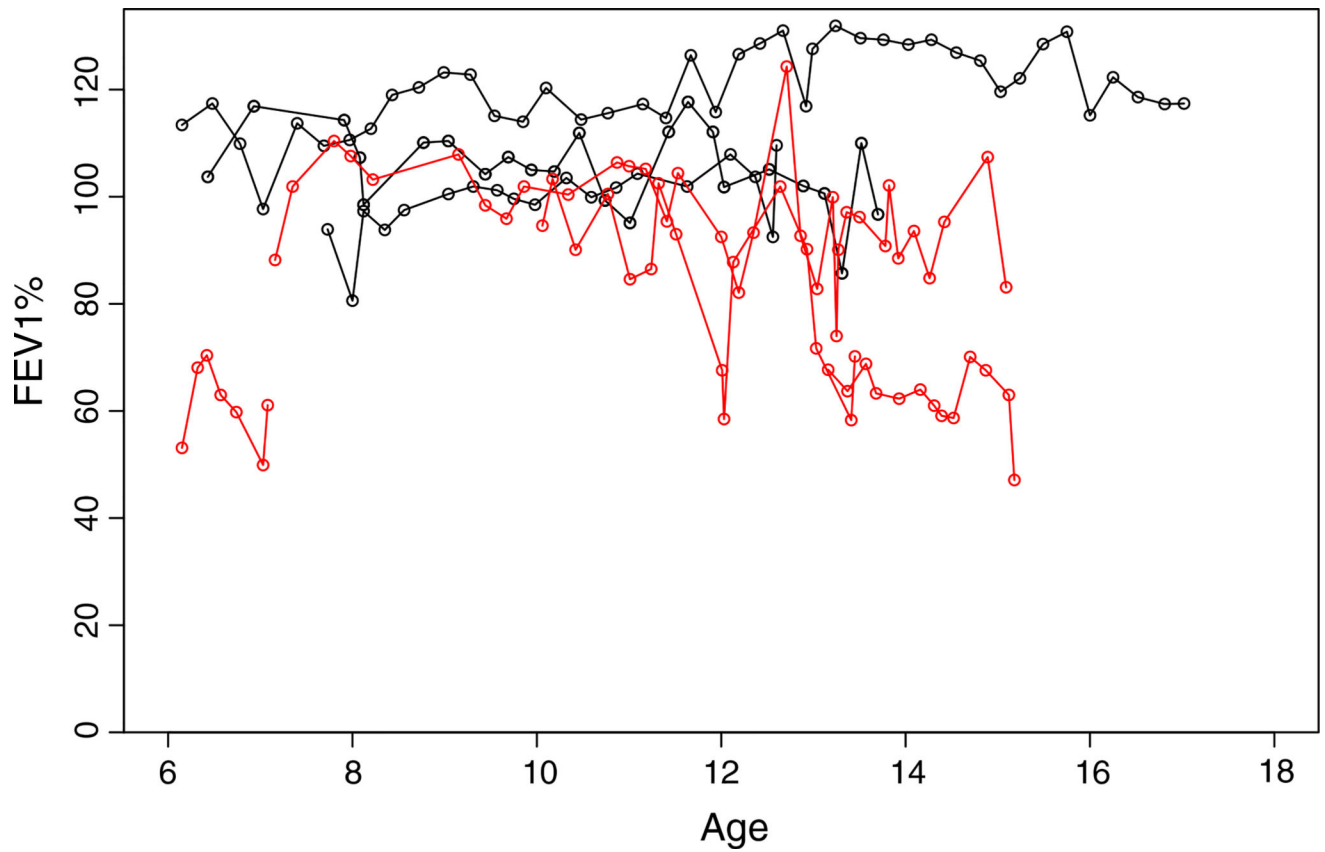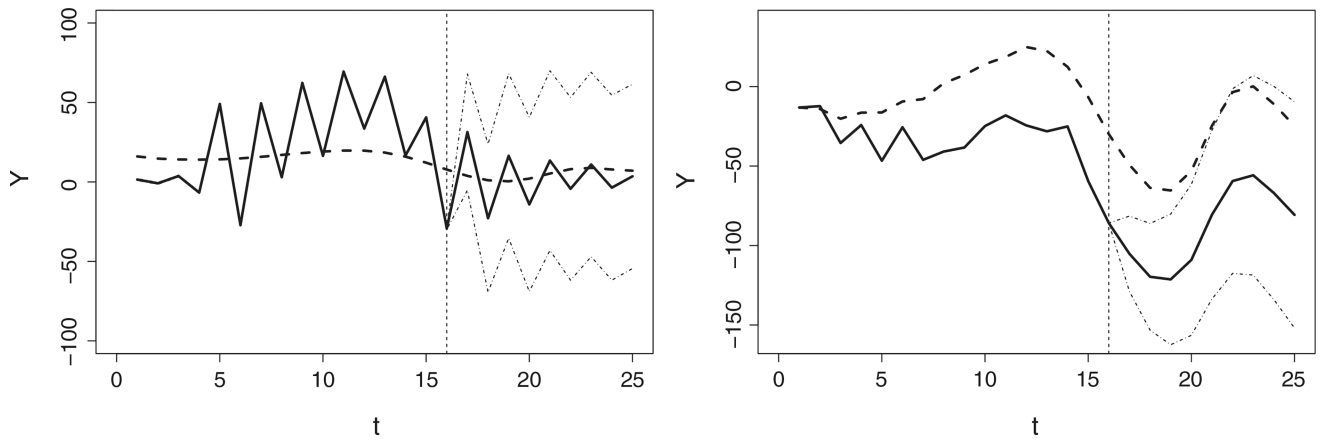
**Figure 1.**
Forced expiratory volume (FEV)$_1$ versus age (in years) from seven representative patients with data taken from the Cystic Fibrosis Foundation Patient Registry. Each set of points connected by a line represents observed FEV$_1$ over age for an individual patient. Trajectories in red represent individuals who had a pulmonary exacerbation, while trajectories in black represent individuals who were pulmonary exacerbation free during the observation period.
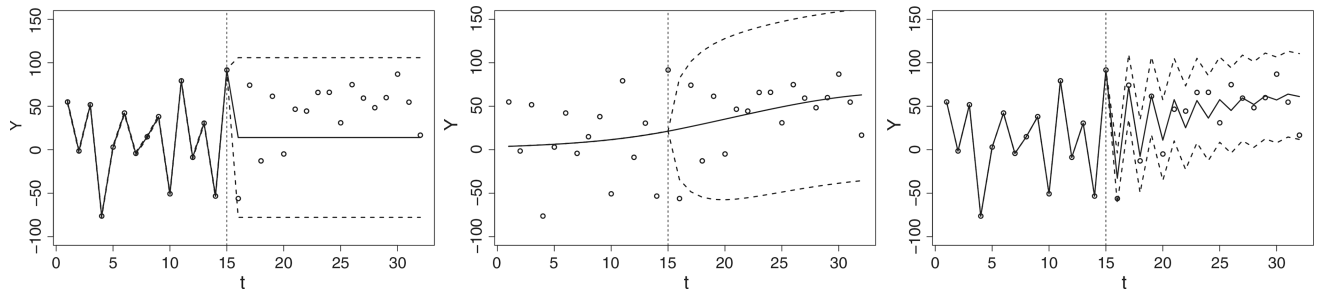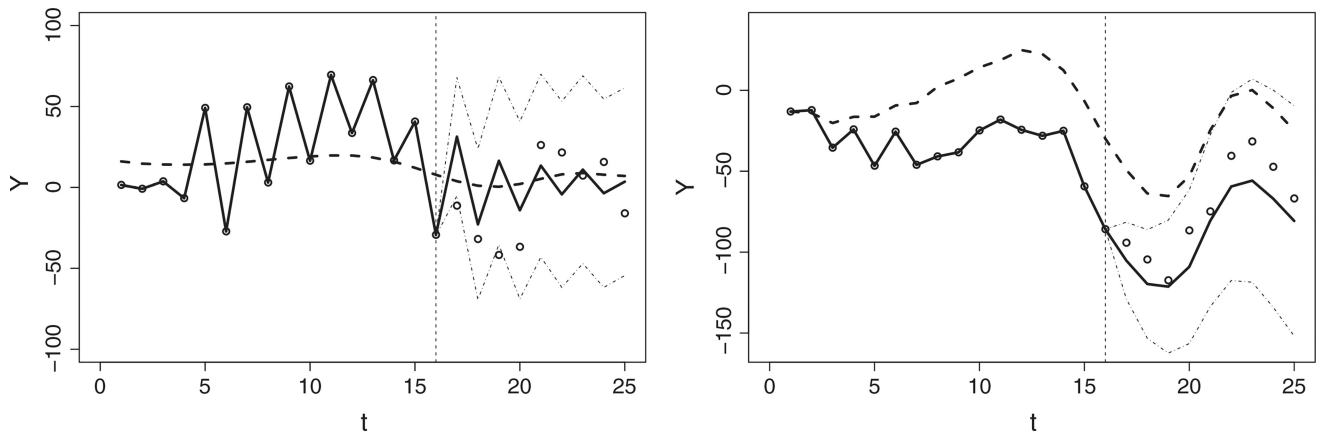
**Figure 2.**
Illustration of the hierarchical Gaussian process (GP) in prediction based on simulated data. For a single subject, beginning with leftmost panel using autocorrelation of an individual GP; middle panel using the estimated $\mu$ from the population GP (with this subject excluded); ending with rightmost panel showing forecasting based on the estimated $\mu$ and the autocorrelation from the joint hierarchical Gaussian process model. In each panel, the vertical bar at $t = 15$ marks the beginning of the extrapolation.

**Figure 3.**
Estimation of the unobserved nonlinear trends in simulation studies. The true (unknown) processes are shown in blue. The estimated values and corresponding 95% credible band are shown in red. The first two columns depict each of the longitudinal submodel results, while the last column shows the observed event occurrences against time for a randomly selected subject *i*.
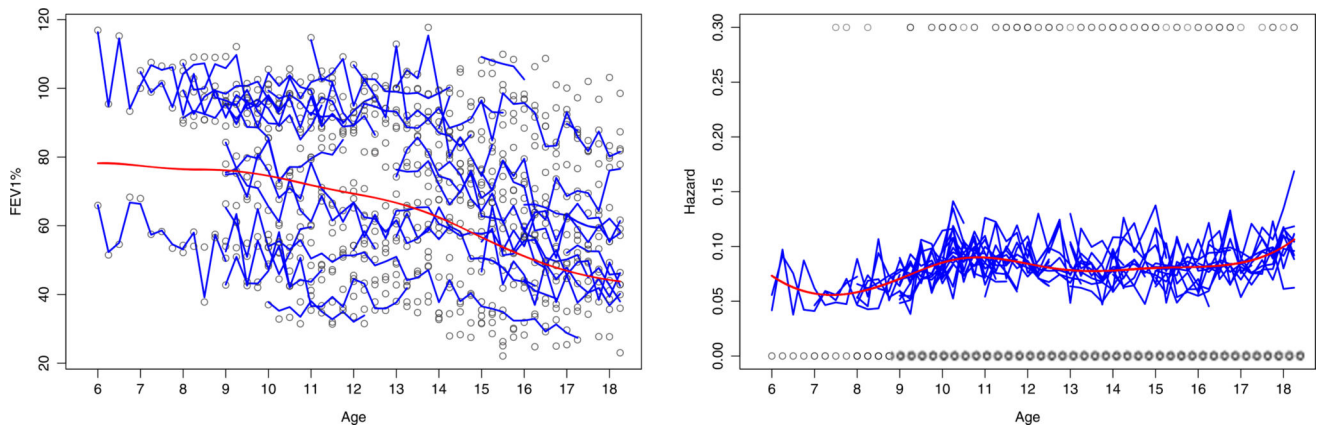
**Figure 4.**
Simulated data illustrating different covariance functions in individual deviations (solid line) from the population mean (dashed line). Predictions with AR(1) (left) with negative $\rho$ fluctuate around the common trend in the center; predictions with Brownian motion (right) tend to remain separate from the mean.
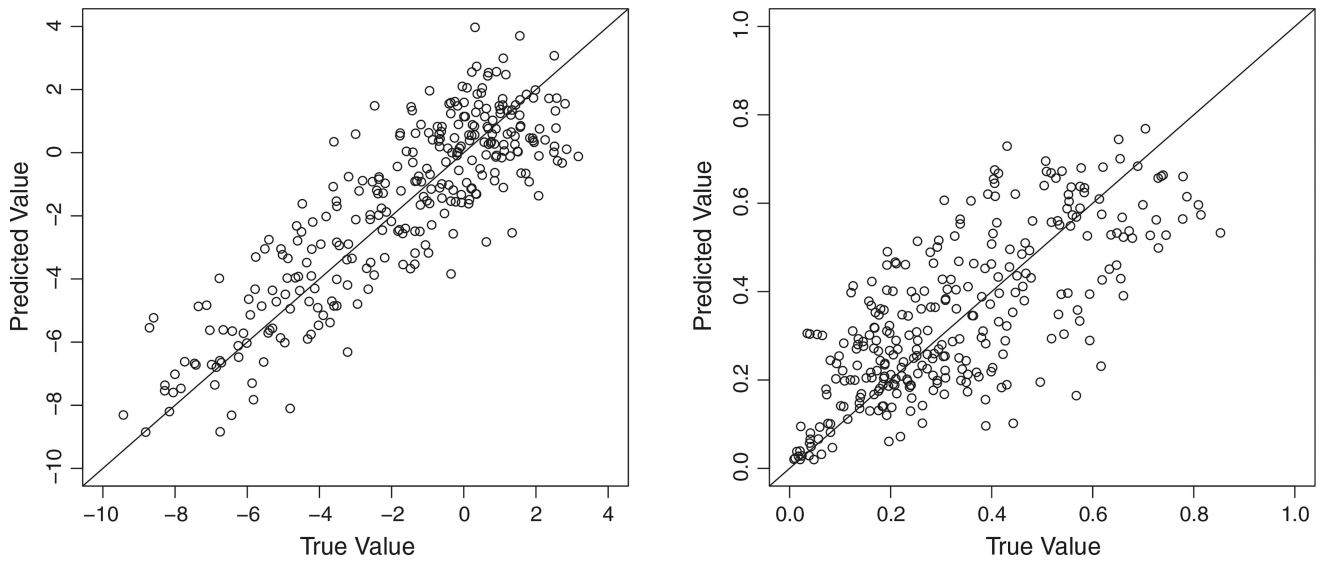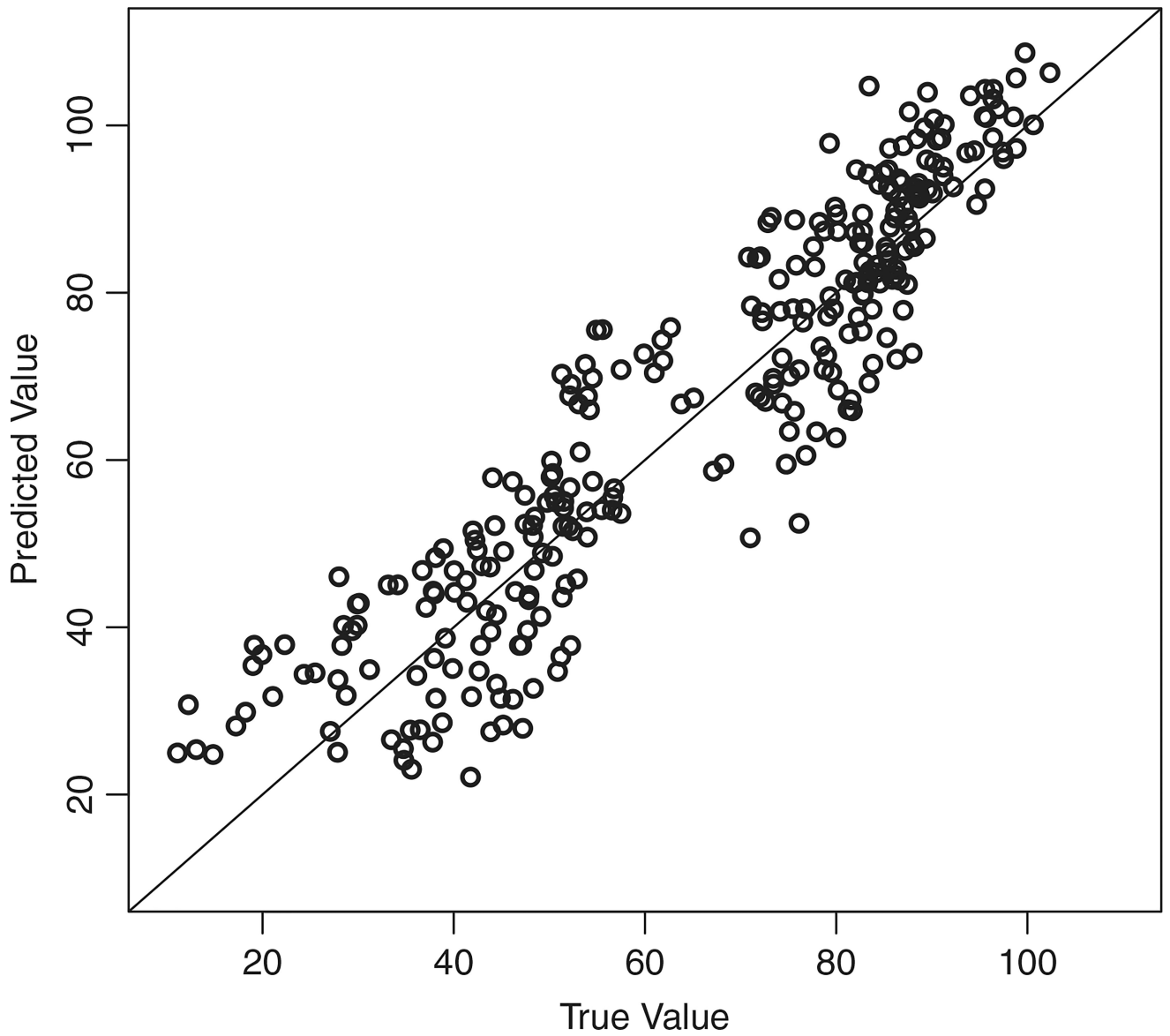
**Figure 5.**
Prediction performance: predicted values versus true values in Sim1 data set for modeling the longitudinal process for $Y$ (left) and the event process for $g(\lambda)$ (right).
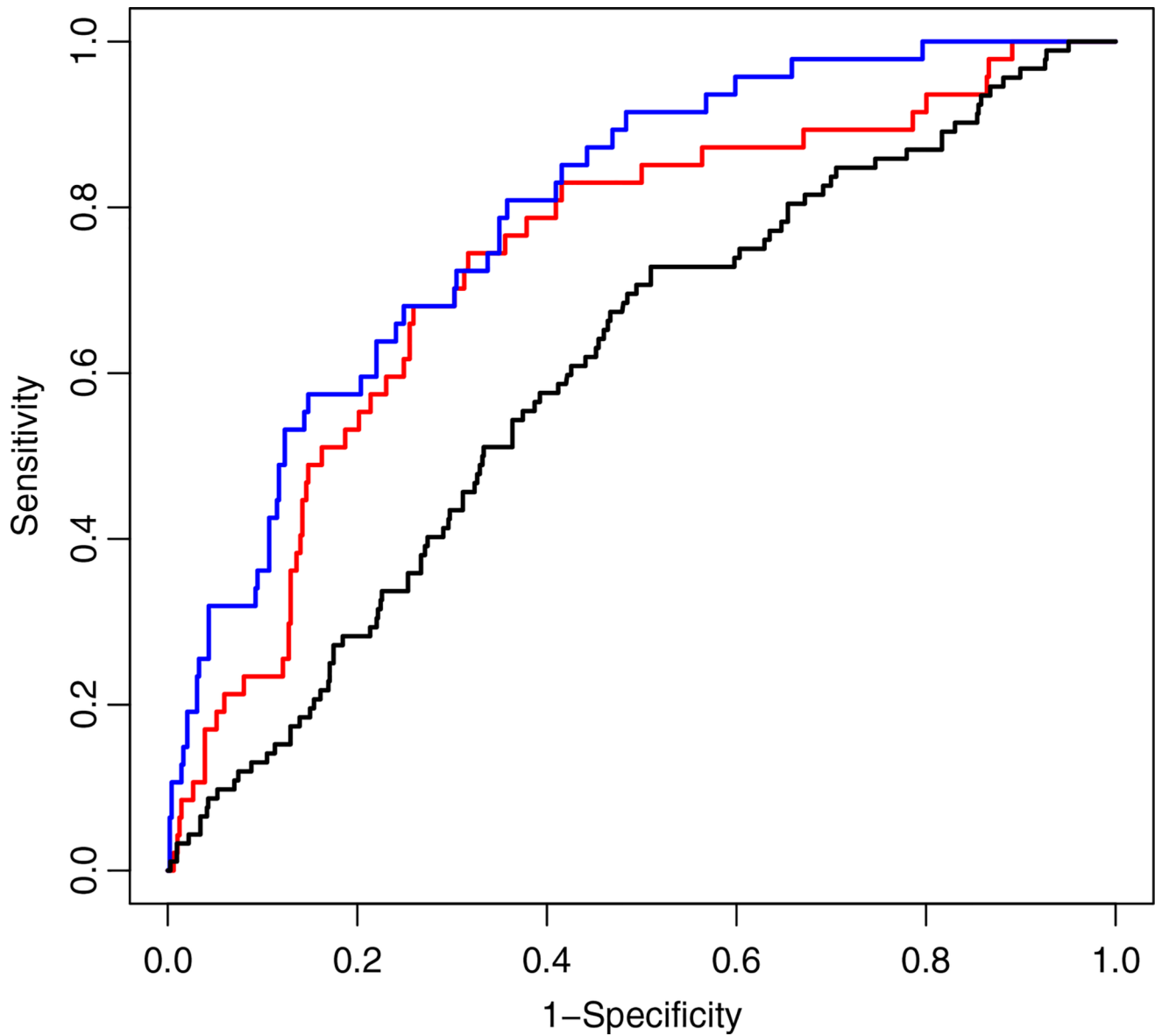
**Figure 6.**
Fitted longitudinal forced expiratory volume (FEV$_1$) (left) and pulmonary exacerbation event (right) data over age for representative cystic fibrosis patients using the joint hierarchical Gaussian process model. Each dot in the left panel represents observed FEV$_1$ for a patient at a given age, while each dot in the right panel indicates whether or not a given patient has experienced a pulmonary exacerbation. The smooth red line in each graph is the estimate of age-related progression at the population level. The individual blue lines are the fitted profiles representing personalized predictions based on the joint hierarchical Gaussian process.

**Figure 7.**
Prediction performance of the joint hierarchical Gaussian process longitudinal submodel for forced expiratory volume from the cystic fibrosis clinical data. The graph shows the predicted values from joint hierarchical Gaussian process against the observed values of forced expiratory volume with the 45° reference line.

**Figure 8.**
Receiver Operating Characteristic curves of each model of the pulmonary exacerbation (PE) event data from the cystic fibrosis clinical data. From left to right, the joint hierarchical Gaussian process model (blue curve) shows highest AUC at 0.735, followed by the single logistic Gaussian process model of PE (red curve: 0.683) and lastly the traditional linear-logistic model of PE (black curve: 0.605).

**Table 1**

Simulation with different values of the AR(1) parameter $\rho$ and inter-outcome scalar $c$. The posterior means (with 95% credible intervals) show that these parameters are correctly recovered.

| Sim No. (true values) | $\rho$ | $c$ |
|---|---|---|
| Sim 1 ($\rho = -0.8$, $c = 0.9$) | $-0.77(-0.81,-0.74)$ | $0.86\ (0.69,\ 1.00)$ |
| Sim 2 ($\rho = 0.5$, $c = -0.3$) | $-0.53(-0.48,-0.57)$ | $-0.28(-0.44,-0.12)$ |
| Sim 3 ($\rho = -0.1$, $c = 0.01$) | $-0.09(-0.14,-0.02)$ | $0.03(-0.10,\ 0.18)$ |

**Table 2**

Association measures under different noise levels to assess model robustness.

| Noise/Signal $(\sigma_\tau^2 / c^2 \phi_\psi)$ | *c* (true value: 0.50) |
|---|---|
| 0.1 | 0.50 (0.34, 0.65) |
| 0.5 | 0.46 (0.29, 0.58) |
| 1.0 | 0.51 (0.37, 0.66) |
| 2.0 | 0.44 (0.30, 0.58) |
| 4.0 | 0.45 (0.30, 0.60) |
| 8.0 | 0.25 (0.13, 0.37) |
| 16.0 | 0.23 (0.11, 0.36) |
| 32.0 | 0.09 (−0.03, 0.22) |