



Published in final edited form as:

*J R Stat Soc Ser C Appl Stat.* 2015 November 1; 64(5): 731–753. doi:10.1111/rssc.12102.

## Modeling Short- and Long-Term Characteristics of Follicle Stimulating Hormone as Predictors of Severe Hot Flashes in Penn Ovarian Aging Study

Bei Jiang<sup>1</sup>, Naisyin Wang<sup>2,1</sup>, Mary D. Sammel<sup>3</sup>, and Michael R. Elliott<sup>1,4</sup>

<sup>1</sup>Department of Biostatistics, University of Michigan, Ann Arbor, USA

<sup>2</sup>Department of Statistics, University of Michigan, Ann Arbor, USA

<sup>3</sup>Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, USA

<sup>4</sup>Survey Methodology Program, Institute for Social Research, University of Michigan, Ann Arbor, USA

### Abstract

The Penn Ovarian Aging Study tracked a population-based sample of 436 women aged 35-47 years to determine associations between reproductive hormone levels and menopausal symptoms. We develop a joint modeling method that uses the individual-level longitudinal measurements of follicle stimulating hormone (FSH) to predict the risk of severe hot flashes in a manner that distinguishes long-term trends of the mean trajectory, cumulative changes captured by the derivative of mean trajectory, and short-term residual variability. Our method allows the potential effects of longitudinal trajectories on the health risks to vary and accumulate over time. We further utilize the proposed methods to narrow the critical time windows of increased health risks. We find that high residual variation of FSH is a strong predictor of hot flash risk, and that the high cumulative changes of the FSH mean trajectories in the 52.5-55 year age range also provides evidence of increased risk above and beyond that of short-term FSH residual variation by itself.

### Keywords

Joint modeling; Bayesian penalized B splines; functional regression; short- and long-term characteristics; increased risk window; robust inference

### 1 Introduction

The Penn Ovarian Aging Study (Freeman et al., 2011) is a longitudinal study of a population-based sample of 436 women aged 35-47 years selected via random digit dialing in Philadelphia County, PA during 1996-97, and followed biannually through 2010. The study goal is to explore the associations between reproductive hormone levels and symptoms in the transition to menopause. Changes in hormone levels alter menstrual bleeding patterns, culminating in the cessation of menstruation, which marks the end of a

---

Address for correspondence: Bei Jiang, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109-2029, USA. beijiang@umich.edu.

woman's reproductive years. This course of events, termed perimenopause, can last for 5 or more years, and coincides for a majority of women with the development of hot flashes, night sweats, and other symptoms. The extent to which these symptoms are associated with reproductive hormone levels, trends over time, or fluctuations is not well understood. This lack of understanding is due in part to limited prospectively collected data, and is also due to limitations in our ability to model various aspects of this dynamic process.

In this paper, we focus on the relationship between follicle stimulating hormone (FSH) and presence and severity of hot flashes. FSH stimulates folliculogenesis, an important factor in ovarian aging; thus there has been interest in using longitudinal FSH information to define menopause transition stages as discussed by Sowers et al. (2008). While elevated FSH is an indicator of ovarian aging, Sowers et al. (2008) found both acceleration and deceleration periods in FSH levels were predictive of time to final menstrual period, suggesting that features other than just the level of FSH may give rise to menopausal symptoms. Exploratory analysis of the FSH data in the Penn Ovarian Aging Study shows both acute and gradual increase periods of FSH levels in the population level, and have given rise to clinical questions about whether it is the rate of increase in FSH that signal risks of severe menopausal symptoms. Moreover, identifying critical ages when women are at increased risk for symptoms would be helpful for making treatment decisions. To better understand the association between trajectories of FSH and risk of severe menopausal symptoms in perimenopausal women, we develop a joint modeling method that 1) makes efficient use of the available information in the longitudinal FSH trajectories, by including long-term trends captured by the mean trajectories or the time varying change rates in the long-term trends captured by the derivatives of the mean trajectories as potential predictors in the primary outcome submodel while also adjusting for the previously identified effect of the short term variation captured by the variance of the residuals (Jiang et al., 2014); and 2) allows selection of the longitudinal FSH features within certain clinically relevant time windows to predict the risk of hot flash severities in the primary outcome submodel, where the effects outside this particular time window are assumed to be negligible.

Joint models of longitudinal and health outcome data have been extensively developed in the literature. The early developments of such joint models were mainly motivated by HIV/AIDS clinical trials and cancer research and often focused on summarizing mean longitudinal trends as time-varying predictors in survival outcome models (Tsiatis et al., 1995; Muthén and Shedden, 1999; Wang and Taylor, 2001; Law et al., 2002; Song et al. 2002; Brown and Ibrahim, 2003a, 2003b; Ibrahim et al., 2004; Yu et al., 2008; among many others). In our work, we extend the existing joint modeling approaches and shift the focus to relating scalar response and functional predictors in a functional data analysis (FDA) paradigm. Our modeling strategies are motivated by the need to properly account for three key features of the FSH trajectories in the longitudinal submodel: nonlinear trajectories observed at unequally spaced time points, short-term elevated variation, which is shown by the residual variance, and the heterogeneity nature among individuals, which is shown by the mixture components in both the mean trajectory and the residual variance. Briefly, our work brings together advanced statistical ideas including FDA, robust and semi-parametric inference, and joint longitudinal and outcome modeling in novel ways.

Unlike the typical FDA practice to smooth each individual trajectory independently of one and another, we formulate a robust semi-parametric mixed effect model for all trajectories, where we simultaneously model both the underlying mean and residual variance of the longitudinal FSH trajectories. We consider the Bayesian penalized spline approach by Lang and Brezger (2004), a Bayesian version of the penalized splines proposed by Eilers and Marx (1996), to estimate the underlying mean FSH trajectories. In contrast to fully parametric splines, penalized splines are not as sensitive to the exact number and location of the knots as long as enough knots are being used, since “unnecessary” knots will be smoothed away by shrinking random effects toward 0. This feature enhances the flexibility to accommodate individual curve fitting of FSH values when these subject-level fitted curves may differ from each other. Examples of applications of penalized B splines for longitudinal data include Durban et al. (2005), who modeled the individual heights of children suffering from acute lymphoblastic leukemia from a clinical trial conducted at Dana Farber Cancer Institute, and Chen and Wang (2011), who considered modeling longitudinal systolic blood pressure data from Framingham Heart Study. For the residual variance, instead of treating it as a nuisance parameter as many others did, we follow Elliott et al. (2012) and Jiang et al. (2014) to model the within-subject residual variance in the FSH trajectories and study its prediction ability in the primary outcome submodel. Finally, considering the bimodal nature in the FSH trajectories as suggested in Jiang et al. (2014), also shown in Figures 3 and 7 in Sections 3 and 4, respectively, we allow for mixtures for both mean trajectories and residual variances to reflect early or late rising patterns in the FSH mean trajectories, crossed with high or low level of short-term variation patterns. The assumed structure nicely reflects the heterogeneity features in the FSH observations. Besides modeling individual trajectory via spline fitting, we extend the normal-error assumptions of Jiang et al. (2014) by allowing for heavier tailed t-distributions for residual errors to avoid the potential influence of outlying observations.

In the primary outcome submodel, while also adjusting for the effect of the residual variance, we treat the smooth mean trajectories estimated from the longitudinal submodel, or the corresponding derivatives as functional predictors linked to the risk of hot flash severities through a FDA regression model in the sense of Ramsay and Dalzell (1991) and James (2002) among many others. This modeling strategy implicitly allows the effects of FSH histories (i.e., FSH values up to a particular time point) or the time-varying change rates of FSH histories that are represented by functional coefficient curves to be time varying and accumulative over time. To estimate the functional coefficient curves, we also propose to use the Bayesian penalized spline approach by Lang and Brezger (2004). In addition to the desirable semi-parametric features mentioned above, the Bayesian penalized spline approach also allows for simultaneous evaluation of the uncertainty of the estimated functional coefficient curves by providing point-wise Bayesian credible intervals, which leads to identification of critical time windows of increased risk of health outcome of interest, while such intervals are typically obtained by bootstrap methods in frequentist FDA regression. To the best of our knowledge, such a modeling strategy has not been considered in the joint modeling literature. Instead, most of the joint modeling developments have focused on using 1) a summary of important features in the longitudinal trajectories, such as the random effects (RE) and the latent classes (LC); or 2) the last available “true” value as a

time dependent covariate, with the earlier values being considered irrelevant to the outcome of interest. In the context of joint modeling of continuous longitudinal data and a binary outcome, Jiang et al. (2014) contrasted the use of RE and LC approaches and discussed how to utilize the information they jointly provide to fully take advantage of each approach. Thorough reviews of the joint modeling of continuous longitudinal data and time-to-event outcomes are given by Tsiatis and Davidian (2004), Ibrahim et al. (2010) and Rizopoulos (2012).

The rest of this paper is organized as follows. In Section 2, we provide the statistical modeling, inference and model-checking procedures that are needed to conduct the proposed analysis of Penn Ovarian Aging data. In Section 3, we present the key features in Penn Ovarian Aging data, which has motivated the modeling and methodology strategies given in Section 2, as well as how we use these strategies to reach new scientific findings and discoveries in linking severe hot flashes risk to FSH longitudinal features for the Penn Ovarian Aging Study. We conclude with a discussion in Section 4. Algorithms to implement the Gibbs sampler for our proposed models are available in the Web-based supporting materials.

## 2 The proposed model

In this section, we present our joint FDA regression models for the longitudinal FSH levels to predict severity of hot flashes modeled using ordinal multinomial probit models.

- Specifically, the longitudinal submodel for the FSH data is given by:

$$\begin{aligned}
 Y_{ij} | \mathbf{b}_i &= \mu(\mathbf{b}_i; t_{ij}) + \varepsilon_{ij} \\
 \varepsilon_{ij} &\sim t_v(0, \sigma_i^2), \text{ which is equivalent to} \\
 \varepsilon_{ij} &\sim N(0, \sigma_i^2/m_{ij}) \text{ with } m_{ij} \sim \text{Gamma}(v/2, v/2) \quad (1) \\
 \mu(\mathbf{b}_i; t_{ij}) &= \sum_{l=1}^L b_{il} \phi_l(t_{ij})
 \end{aligned}$$

where  $Y_{ij}$  denotes the observed longitudinal FSH values for subject  $i$ ,  $i = 1, \dots, n$  at time  $t_{ij}$ ,  $j = 1, \dots, n_i$ ,  $\mu_i(t) = \mu(\mathbf{b}_i; t)$  denotes the mean of  $Y_{ij}$  at time  $t$ , and the vector  $\boldsymbol{\mu}_i = (\mu(\mathbf{b}_i; t_{i1}), \dots, \mu(\mathbf{b}_i; t_{in_i}))^T$  defines the mean trajectory or trajectory for subject  $i$ , where  $\mathbf{b}_i = (b_{i1}, \dots, b_{iL})$  is the vector of the random effects that reflects the subject-level trajectory patterns, and  $\phi_l(t_{ij})$ ,  $l = 1, \dots, L$  are the B spline basis functions.

To flexibly model the mean trajectory  $\boldsymbol{\mu}_i$ , we use truncated power splines consisting of piecewise polynomials of certain order connected at pre-specified knot locations (Ruppert et al., 2003). Given the same order and knot locations, truncated power splines and B splines are equivalent in the sense that there exist unique one-to-one linear transformations between these two sets of spline basis functions (Ruppert et al., 2003), leading to the same fitted values from these two splines in the regression setup. However, the B spline is more numerically stable than the truncated power spline because the B spline basis functions are almost orthogonal while the truncated power spline basis functions are not. Therefore, we use B spline basis

functions  $\phi_l(t_{ij}) \equiv \phi_{l,d}(t_{ij})$ ,  $l = 1, \dots, L$  of degree  $d = 3$ , where  $\phi_{l,3}(t_{ij})$  is obtained by the recursion relation:

$$\phi_{l,d}(t_{ij}) = \frac{t_{ij} - k_l}{k_{l+d} - k_l} \phi_{l,d-1}(t_{ij}) + \frac{k_{l+1+d} - t_{ij}}{k_{l+d+1} - k_{l+1}} \phi_{l+1,d-1}(t_{ij})$$

for knots at points  $k_1, \dots, k_{L-d-1}$ , where  $\phi_{l,0}(t_{ij}) = I(k_l \leq t_{ij} < k_{l+1})$ . The number of interior knots is denoted by  $J_{\mu(t)}$ , such that  $\sum_{l=1}^L \phi_l(t) = 1$  with  $L = J_{\mu(t)} + d + 1$ . We defer the discussion of the selection of knot points to Section 2.5.

To allow for “heterogeneity” in the mean trajectory in the sense of Growth Mixture Models (Verbeke and Lesaffre, 1996; Muthén and Shedden, 1999; Jiang et al., 2014), we consider a finite mixture of normal distributions for the random effect  $\mathbf{b}_i$ ,

$$\mathbf{b}_i = (b_{i1}, \dots, b_{iL})^T \mid D_i = d \sim \mathcal{N}(\boldsymbol{\beta}_d, \boldsymbol{\Sigma}_d), d = 1, \dots, K_D. \quad (2)$$

where,  $D_i$  defines the corresponding latent class membership for the mean trajectory class and  $\boldsymbol{\beta}_d = (\beta_{d1}, \dots, \beta_{dL})^T$ . Thus, the fixed effect coefficients  $\beta_{dl}$ ,  $l = 1, \dots, L$  determine the shape and also the smoothness of the mean trajectory for the  $d^{th}$  latent class, defined as  $f_d(t_{ij}) = \sum_{l=1}^L \beta_{dl} \phi_l(t_{ij})$ ,  $d = 1, \dots, K_D$ . Following Lang and Brezger (2004), we use Gaussian random walk priors on  $\boldsymbol{\beta}_d$  to penalize large differences among coefficients of the adjacent spline basis and therefore control the smoothness of the mean trajectory curve to avoid potential overfitting. The specific prior distributions are given in Section 2.3. The random coefficients  $b_{il}$ ,  $l = 1, \dots, L$  then capture the individual deviations from the class specific mean trajectory.

The residual  $\varepsilon_{ij}$  denotes the deviation of  $Y_{ij}$  from the subject specific mean at  $t_{ij}$  and is assumed to follow a Student's t-distribution with  $\nu$  degrees of freedom, assuming mean 0 and scale  $\sigma_i^2$ . The value of  $\nu$  is assumed to be known. Thus the variance of  $Y_{ij}$  is equal to  $\frac{\nu}{\nu - 2} \sigma_i^2$ , which can be interpreted as a measurement of the short term variability around the mean trajectory  $\boldsymbol{\mu}_i$ . In the case of  $\nu = \infty$ ,  $\varepsilon_{ij}$  is normally distributed with mean 0, variance  $\sigma_i^2$  and  $m_{ij} \equiv 1$ . To allow for over-dispersion and “heterogeneity” in the within-subject scale parameter  $\sigma_i^2$ , we assume a mixture of log normal distributions,

$$\sigma_i^2 \mid C_i = c \sim \log - \mathcal{N}(\mu_c, \tau^2), c = 1, \dots, K_C. \quad (3)$$

where  $C_i$  defines the corresponding latent class membership for the variance class and we assume  $C_i \perp\!\!\!\perp D_i$  so that the common assumption that for subject  $i$ , the mean trajectory  $\boldsymbol{\mu}_i(t)$  and the residual  $\varepsilon_{ij}$  are independent still holds.

- The outcome submodel for hot flash severities is defined through an ordinal probit model that assumes there exists a latent continuous variable underlying the observed ordinal outcomes. Specifically, let  $W_i$  denote this underlying latent variable. We observe the ordinal outcome  $o_i = s$ ,  $s = 0, \dots, S$ , if this latent variable  $W_i$  falls between the cutoff  $\gamma_s$  and  $\gamma_{s+1}$ , that is,

$$o_i = s \iff \gamma_s < W_i \leq \gamma_{s+1}$$

where these cutoffs between categories is subject to the common constraint that  $-\infty = \gamma_0 < \gamma_1 < \dots < \gamma_{S+1} = \infty$  with one reference cutoff, usually  $\gamma_1$ , fixed at value 0. Then the distribution of this latent variable  $W_i$  is specified conditional on individual longitudinal mean trajectories and variances as follows:

$$W_i \sim N(\eta_i^W, 1) \\ \eta_i^W = \alpha_0 + \mathbf{x}_i^T \boldsymbol{\lambda}_0 + \int_T \mu_i(t) \theta_0(t) dt, \quad (4)$$

where  $\mathbf{x}_i$  is a vector of baseline covariates with associated (constant) parameter  $\boldsymbol{\lambda}_0$ , and the functional coefficient function  $\theta_0(t)$  represents the effect of subject specific mean trend  $\mu_i(t)$  at time  $t$  while adjusting for the mean trends at other time points within the time window  $T$ . The purpose of considering the integral over the chosen time domain  $T$ , i.e.,  $\int_T \mu_i(t) \theta_0(t) dt$  is to identify critical time windows of elevated outcome risks, which have several advantages over simply summing up over the observed time points  $t_{ij}$ ,  $j = 1, \dots, n$ . First, longitudinal observations often have missing values and can be measured at different time points (known as unbalanced data) and hence summation over the observed time points becomes problematic. Second,  $\mu_i(t)$  is a smoothed functional representation of the underlying mean function with the individual level variability “captured” by  $\sigma_i^2$ . Third, since we have considered a mixed effect model to smooth all individual-level curves and hence borrow strength across individuals, we obtain more stable estimates of  $\mu_i(t)$  in comparison to smoothing  $\mu_i(t)$  individually. Fourth, an integral over a chosen time domain implicitly uses the information at infinite time points within time window  $T$  while summation only uses the information at finitely observed time points. As in the mean trajectories, we express  $\theta_0(t) = \sum_{k=1}^{K_0} \tilde{\theta}_{0k} \psi_k^0(t)$  using cubic B spline basis  $\boldsymbol{\psi}^0(t) = (\psi_1^0(t), \dots, \psi_{K_0}^0(t))^T$  and the associated coefficient vector  $\tilde{\boldsymbol{\theta}}_0 = (\tilde{\theta}_{01}, \dots, \tilde{\theta}_{0K_0})^T$ , with  $\tilde{\theta}_{0K}$  following a random walk prior, given in Section 2.3, to avoid overfitting. Given that we express  $\mu_i(t)$  by  $\mathbf{b}_i^T \boldsymbol{\phi}(t)$  and  $\theta_0(t)$  by  $\boldsymbol{\psi}^0(t)^T \tilde{\boldsymbol{\theta}}_0$ , thus  $\int_T \mu_i(t) \theta_0(t) dt = \int_T \mathbf{b}_i^T \boldsymbol{\phi}(t) \boldsymbol{\psi}^0(t)^T \tilde{\boldsymbol{\theta}}_0 dt = \mathbf{b}_i^T G_T^0 \tilde{\boldsymbol{\theta}}_0$ , where  $\boldsymbol{\phi}(t)$  is a vector of  $L$  basis functions chosen to express  $\mu_i(t)$  in the longitudinal submodel and  $\boldsymbol{\psi}^0(t)$  is a vector of  $K_0$  basis functions;  $G_T^0 = \int_T \boldsymbol{\phi}(t) \boldsymbol{\psi}^0(t)^T dt$ . We can calculate or evaluate numerically  $G_T^0$  for any given spline basis functions and the estimation of unknown parameters in the outcome primary model becomes fully parametric.

Alternatively, one may postulate that the cumulative changes of the individual trajectories are potentially predictive of the outcome of interest. To accommodate such a possibility, we can consider the first derivative of  $\mu_i(t)$  i.e.,  $\mu'_i(t) = \partial\mu_i(t)/\partial t$  as a functional predictor by taking advantage of the nice properties of B spline of continuity and replace the specification (4) for the outcome model by the following alternative form,

$$W_i \sim N(\eta_i^W, 1) \\ \eta_i^W = \alpha_1 + \mathbf{x}_i^T \boldsymbol{\lambda}_1 + \int_T \mu'_i(t) \theta_1(t) dt, \quad (5)$$

where, as for  $\theta_0(t)$ , the functional coefficient function  $\theta_1(t)$  can be interpreted as the effect of the derivative of mean trend  $\mu'_i(t)$  or the rate of change in  $\mu'_i(t)$  at time  $t$  while adjusting for the values of  $\mu'_i(t)$  at other time points within the time window  $T$ . To emphasize the fact that we can use different spline basis functions to express  $\theta_1(t)$ , we express

$\theta_1(t) = \sum_{k=1}^{K_1} \tilde{\theta}_{1k} \psi_k^1(t)$  using a different set of B spline basis  $\psi^1(t) = (\psi_1^1(t), \dots, \psi_{K_1}^1(t))^T$  and the associated coefficient vector  $\tilde{\boldsymbol{\theta}}_1 = (\tilde{\theta}_{11}, \dots, \tilde{\theta}_{1K_0})^T$ . A penalized approach was used by requiring a random walk prior on  $\tilde{\boldsymbol{\theta}}_1$ , i.e.,  $\tilde{\theta}_{1k} \sim N(\tilde{\theta}_{1k-1}, \tau_{\theta_1}^2)$ ,  $k=2, \dots, K_1$ . Similarly, we have  $\int_T \mu'_i(t) \theta_1(t) dt = \int_T \mathbf{b}_i^T \boldsymbol{\phi}'(t) \boldsymbol{\psi}^1(t)^T \tilde{\boldsymbol{\theta}}_1 dt = \mathbf{b}_i^T G_T^1 \tilde{\boldsymbol{\theta}}_1$ , where  $\boldsymbol{\phi}(t) = \boldsymbol{\phi}(t)/t$  given  $\boldsymbol{\phi}(t)$  is a vector of  $L$  basis functions chosen to express  $\mu_i(t)$  in the longitudinal submodel and  $\boldsymbol{\psi}^1(t)$  is a vector of  $K_1$  basis functions;  $G_T^1 = \int_T \boldsymbol{\phi}'(t) \boldsymbol{\psi}^1(t)^T dt$ .

### 2.1 Likelihood specification

Let  $\boldsymbol{\phi} = (\pi_d^D, \boldsymbol{\beta}_d, \sum_d, d=1, \dots, K_D; \pi_c^C, \mu_c, c=1, \dots, K_C; \tau^2, \alpha_0, \boldsymbol{\lambda}_0, \tilde{\boldsymbol{\theta}}_0, \boldsymbol{\gamma})$ , where we assume each parameter in  $\boldsymbol{\phi}$  has an independent prior distribution, with the joint prior distribution denoted by  $\pi(\boldsymbol{\phi})$ , and  $\mathbf{z}$  includes all unobserved latent variables, i.e.,  $\mathbf{z} = (\mathbf{b}, \boldsymbol{\sigma}, \mathbf{C}, \mathbf{D})'$ . The observed data  $\mathbf{x}$  consists of the longitudinal trajectories  $\mathbf{y}_1, \dots, \mathbf{y}_n$  and the observed outcomes  $o_1, \dots, o_n$ . Then the complete data likelihood of  $\boldsymbol{\phi}$  based on  $(\mathbf{x}, \mathbf{z})$  is given by

$$f(\mathbf{x}, \mathbf{z} | \boldsymbol{\phi}) \propto \left\{ \prod_{i=1}^n \left\{ \prod_{d=1}^{K_D} \left[ \pi_d^D (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}_d|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{b}_i - \boldsymbol{\beta}_d)' \boldsymbol{\Sigma}_d^{-1} (\mathbf{b}_i - \boldsymbol{\beta}_d) \right\} \right]^{I_{(D_i=d)}} \prod_{c=1}^{K_C} \left[ \pi_c^C (2\pi\tau^2)^{-\frac{1}{2}} \sigma_i^{-2} \exp \left\{ -\frac{(\log \sigma_i^2 - \mu_c)^2}{2\tau^2} \right\} \right] \right\}$$

where  $\Phi(\cdot)$  denotes the cumulative distribution function for standard normal distribution and

$$p(y_{ij}; v, \mathbf{b}_i, \sigma_i^2) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left\{ -\frac{(y_{ij} - \sum_{l=1}^L b_{il} \phi_l(t_{ij}))^2}{2\sigma_i^2} \right\} & \text{if } v = \infty; \\ \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2}) \sqrt{2\pi\sigma_i^2}} \left\{ 1 + \frac{(y_{ij} - \sum_{l=1}^L b_{il} \phi_l(t_{ij}))^2}{2\sigma_i^2} \right\}^{-\frac{v+1}{2}} & \text{if } v < \infty. \end{cases}$$



## 2.2 Data augmentation step to impute missing data

Given the minimum number of available repeatedly measured FSH levels in our final sample (ranging between 6 and 26 per woman), we are limited as to the number of knots when choosing cubic B spline basis functions to express  $\mu_i(t)$ . To maximize the number of knots we can consider, we fill in those with fewer than 26 observations based on data augmentation within each iteration of Gibbs sampling (Chapter 10 in Little and Rubin, 2002). When assuming missing at random (MAR) missing data mechanism, this data augmentation procedure proceeds as follows,

- draw  $\mathbf{Y}_{miss}^{(t+1)}$  from  $p(\mathbf{Y}_{mis} | \boldsymbol{\varphi}, \mathbf{X}_{obs})$
- draw  $\boldsymbol{\varphi}^{(t+1)}$  from  $p(\boldsymbol{\varphi} | \mathbf{X}_{obs}, \mathbf{Y}_{mis})$

where  $\boldsymbol{\varphi}$  denotes model parameters,  $\mathbf{Y}_{mis}$  denotes the missing longitudinal observations of FSH levels, and  $\mathbf{X}_{obs}$  denotes all observed data including observed longitudinal observations and primary outcome of interest. The above simulation leads to draws from the joint distribution of  $(\boldsymbol{\varphi}, \mathbf{y}_{mis})$  given observed data  $\mathbf{X}_{obs}$ . Therefore, this procedure leads to the same inference about  $\boldsymbol{\varphi}$  as when we only focus on the marginal distribution of  $\boldsymbol{\varphi}$  given observed data  $\mathbf{X}_{obs}$ . This trick allows us to put in more knots to fully take advantage of the penalized spline approach that is free from knot location selection given a sufficient number of knots.

## 2.3 Prior specification

We propose a fully Bayesian approach to estimate model parameters. For the mixture normal distribution of the random effects, we assume a first-order Gaussian random walk prior as proposed by as Lang and Brezger (2004):  $\beta_{dl} \sim N(\beta_{d,l-1}, \tau_{\beta_{dl}}^2)$ ,  $l=2, \dots, L$  with diffuse prior  $\beta_{d1} \sim N(0, 100)$  for the initial coefficient, and  $\tau_{\beta_{dl}}^2 \sim IG(1, 0.005)$  to control the smoothness of the fitted curves. We do not impose restrictions on the structure of the variance-covariance matrix for the random effects  $\Sigma_d$ . To avoid problems with unbounded likelihoods in normal mixture models with unstructured variance-covariance matrices (Day 1969), we use an empirical Bayes prior proposed by Kass and Natarajan (2006):  $\Sigma_d \sim$  Inverse-Wishart(df =  $r$ ,  $\Lambda$ ), where  $\Lambda = r \left( \sum_{i=1}^n \widehat{\text{Cov}}(\tilde{\mathbf{b}}_i)^{-1} / n \right)^{-1}$ , where  $\tilde{\mathbf{b}}_i$  is given by OLS estimator of  $\mathbf{b}_i$  for subject  $i$ , and  $r$  is the dimension of  $\mathbf{b}_i$ .

For the mixture log normal distribution for the residual variances, we used diffuse priors:  $\mu_c \sim N(0, \nu)$ ,  $\tau^2 \sim IG(a, b)$  with  $\nu = 1000$  and  $a = b = .001$ . For the class membership probabilities, we assume conjugate Dirichlet(4, ..., 4) on both  $\boldsymbol{\pi}^C = (\pi_1^C, \dots, \pi_{K_C}^C)^T$  and  $\boldsymbol{\pi}^D = (\pi_1^D, \dots, \pi_{K_D}^D)^T$  (Frühwirth-Schnatter 2006); this is equivalent to assuming *a priori* 4 observations in each class, avoiding the existence of empty classes.

Lastly, in the probit submodel we assign independent priors  $N(0, 9/4)$  for the  $\alpha_0$  and every element of  $\boldsymbol{\lambda}_0$ ; for the coefficients associated with functional coefficient function  $\theta_0(t)$ ,  $\hat{\boldsymbol{\theta}}_0 = (\hat{\theta}_0, \dots, \hat{\theta}_{0K_0})^T$ , similarly we use a first-order Gaussian random walk prior, i.e.,



$\tilde{\theta}_{0k} \sim N(\tilde{\theta}_{0k-1}, \tau_{\tilde{\theta}_0}^2), k=2, \dots, K_0$  with  $\hat{\theta}_{01} \sim N(0, 9/4)$  and  $\tau_{\tilde{\theta}_0}^2 \sim \text{IG}(1, 0.005)$ , where the prior variance 9/4 is chosen to bound the probabilities of  $o_i = s, s = 0, \dots, S$  to be away from 0 and 1 (Garrett and Zeger, 2000; Elliott et al., 2007 and Neelon et al., 2011). We put flat uniform priors on  $\gamma_s$  for  $s \notin [0, 1, S + 1]$ , that is,  $\gamma_s \sim \text{Uniform}(-\infty, \infty)$ .

## 2.4 Posterior computation

Gibbs sampling is used to obtain draws from the corresponding posterior distributions. For  $(\alpha_0, \lambda_0, \boldsymbol{\theta} \mid \mathbf{b}, \boldsymbol{\sigma}, \mathbf{o})$  we use the Albert and Chib (1993) data augmentation method for probit regression models. The draws of  $(\sigma_i^2 \mid C_i, \mu_c, \boldsymbol{\gamma}, \mathbf{b}_i, o_i, W_i, \{y_{ij}\}_j)$  for  $i = 1, \dots, n$  are obtained by the inverse cumulative distribution method. The exact specification of all priors and MCMC sampling procedures are provided in the Web-based supporting materials.

For each model, we ran three chains of 100,000 iterations from diverse starting points, discarding the first 50,000 as burn-in and retaining every 10th draw to reduce autocorrelation. Gelman-Rubin statistic  $\hat{R}$  (Gelman et al., 2003) (square root of total variance to within-chain variance ratio) were used to assess the convergence of the MCMC chains. For the population level parameters, the maximum  $\hat{R} = 1.030$  for models assuming less than 3 classes; and when assuming 3 classes for either mean trajectory or the variance class, the maximum  $\hat{R} = 1.184$ . For the well-documented issue of “label switching” in finite mixture modeling (Redner and Walker 1984), various solutions have been proposed, including the relabeling algorithms by Stephens (2000), Jasra et al. (2005) and Rodríguez and Walker (2012). We applied the post-processing relabeling algorithm by Stephens (2000), which considers all possible permutations of class assignments at each iteration of the Gibbs sampler and chooses the one which minimizes Kullback-Leibler (KL) divergence of the estimated vs. true probabilities of class membership, thus maximizing the posterior probability so that the labeling of classes was consistent with the previous assignments. We post-process the MCMC chains using Stephen’s algorithm to “untangle” the draws for model parameters.

All the calculations were performed by calling stand alone C++ codes in R, developed using an open source C++ library for statistical computation, the Scythe statistical library (Pemstein et al., 2007), which is available for free download at <http://scythe.wustl.edu>.

## 2.5 The choice of the number of classes and number of knots in penalized splines

We consider the deviance information criterion (DIC), proposed by Spiegelhalter et al. (2002), to select both the number of components for the latent classes and to choose the number of knots in the penalized splines. DIC uses the discrepancy between the posterior mean of the deviance  $\overline{D(\phi)} = E_{\phi} \{-2 \log f(\mathbf{x} \mid \phi) \mid \mathbf{x}\}$  and the deviance evaluated at the posterior mean  $D(\bar{\phi}) = -2 \log f\{\mathbf{x} \mid E(\bar{\phi}, \mathbf{x})\}$  to estimate the effective number of degrees of freedom in the model  $p_D$ . DIC is then given by the analog of the Akaike Information Criterion (AIC):

$$\text{DIC}(\mathbf{x}) = \overline{D(\phi)} + p_D = 2\overline{D(\phi)} - D(\bar{\phi}) = -4E_{\phi} \{\log f(\mathbf{x} \mid \phi) \mid \mathbf{x}\} + 2\log f\{\mathbf{x} \mid E(\bar{\phi}, \mathbf{x})\}.$$

In our setting,  $f(\mathbf{x} | \boldsymbol{\phi})$  where  $\mathbf{x} = (\mathbf{y}_{obs}, \boldsymbol{o})'$  consisting of the fully-observed data is not available in closed form; instead we use the approach outlined in Celeux et al. (2006) to obtain

$$DIC(\mathbf{x}) = E_{\mathbf{z}}\{DIC(\mathbf{x}, \mathbf{z})\} = -4E_{\mathbf{z}, \boldsymbol{\phi}}\{\log f(\mathbf{x}, \mathbf{z} | \boldsymbol{\phi}) | \mathbf{x}\} + 2E_{\mathbf{z}}[\log f\{\mathbf{x}, \mathbf{z} | E_{\boldsymbol{\phi}}(\boldsymbol{\phi} | \mathbf{x}, \mathbf{z})\} | \mathbf{x}]$$

where integration over the latent variables  $\mathbf{z} = (\mathbf{b}, \boldsymbol{\sigma}, \mathbf{C}, \mathbf{D}, \mathbf{y}_{mis})'$  is obtained via numerical methods.

## 2.6 Goodness of fit evaluation

We assessed the model goodness of fit to the data in two ways: pivotal discrepancy measures (PDMs) (Johnson, 2007; Yuan and Johnson, 2012), which yields an overall goodness-of-fit measure for the longitudinal predictor component, and area under the receiver-operator characteristic (ROC) curve (AUC), a goodness-of-fit measure focusing on prediction of the ordinal outcome of interest.

In contrast to more general posterior predictive distribution measure of fit (Gelman et al., 1996), PDMs are defined to depend only on the data and the model parameters with a known distribution. If the model is correctly specified, the PDMs evaluated at the true parameter value and the draws from the posterior distribution should have the same sampling distribution. Therefore, model adequacy can be tested by treating the PDMs as a test statistic to obtain a uniformly distributed  $p$  value. However, the posterior samples of PDMs are not independent as they are all derived from the observed data (Johnson, 2004), thus  $p$ -value calculation is difficult. Instead, Johnson (2007) and Yuan and Johnson (2012) focus on the upper bound of  $p$  values and hence the upper bound of a  $p$  value being less than 0.05 definitely provided strong evidence of model inadequacy.

To examine the fit of the longitudinal trajectories, we consider subject level PDMs, where for subject  $i$ , we let  $D_i = \sum_{j=1}^{n_i} m_{ij} [y_{ij} - \mu_i(t_{ij})]^2 / \sigma_i^2$ . When the assumed longitudinal submodel defined in (1) is correct, the PDM  $D_i$  is  $\chi_{n_i-1}^2$  distributed. We use repeated posterior draws to obtain the sampling distribution of PDMs and compute the upper bounds of the  $p$  values based on the ordered statistics of PDMs using the approach by Yuan and Johnson (2012).

Second, we assessed the prediction of the outcome using receiver-operator characteristic (ROC) curves, in particular the area under the ROC curve (AUC). ROC curves plot true positive rate (TP) versus false positive rate (FP) for all possible cutoffs based on predicted  $P(o_i = s) = \Phi(\mathbf{Z}'_i \boldsymbol{\eta})$  obtained from (4) for  $s = 0, \dots, S$ . The ROC curve and AUC were computed at each MCMC iteration using the ROCR package in R (Sing et al. 2005). The ROC is computed by ordering the observations  $(i) = 1, \dots, n$  so that  $P(\hat{o}_{(i)} = 1) - P(\hat{o}_{(i+1)} = 1)$ , computing changepoints  $c = 2, \dots, n_c, n_c - n$  where the observations change from positive to negative (i.e.,  $o_{(c-1)} = 1, o_{(c)} = 0$ ), and plotting

$$\sum_{(i)=1}^c (1 - o_{(i)}) / \sum_{(i)=1}^n (1 - o_{(i)}) \text{ on the horizontal axis versus } \sum_{(i)=1}^c o_{(i)} / \sum_{(i)=1}^n o_{(i)} \text{ on}$$

the vertical axis. Area under the ROC is then computed using a trapezoidal approximation. The posterior mean AUC is calculated as the average AUC's across MCMC iterations. To obtain the posterior mean and the pointwise 95% credible interval of ROC curve, we choose 250 points equally spaced along the FP axis and take the vertical average or 95% quantiles of TP's at the 250 chosen points. This approach is referred to as vertical averaging of ROC curves at fixed FP rates by Fawcett (2006).

### 3 Predicting risks of hot flash severities from longitudinal follicle stimulating hormone data

In the Penn Ovarian Aging Study, participating women had their hormone measures taken annually during the early follicular phase of a menstrual cycle for 2 sequential menstrual cycles, with up to 13 years of follow-up available at the time of our analysis. We focus our analysis on the 234 women who 1) had not experienced hot flash symptoms at baseline, 2) had baseline measurements of BMI and smoking status (0 or 1) that are to be included as baseline covariates in the outcome submodel, and 3) had at least 6 measurements of FSH levels. Among this restricted sample, 144 (62%) women had fully participated in the study. Among the remaining 90 (38%) women, 42 of them dropped out after at least 6 assessment periods, while 48 of them had either sporadically skipped the assessments or dropped out of the study in the very beginning but came back to the study later on when increased incentives were offered. Nelson et al. (2004) examined the factors that may predict the participation after six assessment periods and concluded that dropping out was likely random; for those who came back to the study because of increased incentives, their initial dropout was likely due to personal reasons that were not symptom related. FSH values could be missing due to lab errors or missing blood samples (7.1%), which is likely to be missing at random. Further, FSH values were censored if a woman was 1) pregnant and/or breast feeding (0.3%) 2) hysterectomy with or without oophorectomy (3.0%) 3) taking exogenous hormone replacement therapy (1.4%) 4) taking oral contraceptives (2.5%) 5) taking cancer treatment medications (0.6%) 6) taking other estrogen (0.2%) during the follow-up. The average number of available FSH levels per woman is 18.7 (range: 6-26) in our final sample.

We let  $y_{ij}$  denote the natural log transformed FSH levels i.e.,  $\log(\text{FSH})$  and  $o_i$  denote the ordinal outcome of interest, severity of hot flashes (0, 1, and 2), defined as  $o_i = 0$  if never had severe hot flashes (that is, severity score  $< 2$  throughout the follow up period);  $o_i = 1$  if had severe but not more severe hot flashes (that is, severity score at least once  $=2$  or once  $=3$  that occurred before 40 yrs. old) and  $o_i = 2$  if had more severe hot flashes (that is, severity score at least once  $=3$  after 40 yrs old). In our final sample, 117 (50%) never experienced any severe hot flashes during follow-up (severity score  $=0$ ), 80 (34%) had a severity score of 1, and 37 (16%) had a severity score of 2. Since most women start to experience menopausal related symptoms between the age of 45 and 50 and reach menopause by the age 55, we consider  $T = [45, 55]$  as a potential risk time window in our analysis for the impact of changes in FSH levels on risk of severe hot flashes.

We use the longitudinal submodel defined in (1) to describe longitudinal measured FSH and the outcome model defined in (4) to relate long- and short-term FSH characteristics to the risk of severe hot flashes. Preliminary analysis suggested using cubic B spline basis

functions with 1 to 3 inner knots to express  $\mu_i(t_{ij})$  and cubic B spline basis functions with 1 to 5 inner knots to express the functional coefficient function  $\theta_0(t)$ . Thus we consider models with 1, 3 or 5 knots, putting these knots at the equally spaced quantiles of the distinctly observed ages of these women (Ruppert et al. 2003). This is equivalent to assuming piecewise cubic orthogonal polynomials connected at those chosen knot locations. Next, we consider the number of components for both mean trajectory and variance classes. Previous analysis of fitting mixture distributions for both the random effects and variances (Jiang et al. 2014) successfully identified 1 mean trajectory class and 2 variance classes under normality assumption for  $\varepsilon_{ij}$ . However, our current approach assumes a  $t$ -distribution for  $\varepsilon_{ij}$  that will potentially impact the effect of any outliers on estimation of the mean trajectories, which may alter the optimal numbers of components for the mean trajectory and variance classes. With all these considerations, we consider  $K_D = 1, 2$  and  $3$  and  $K_C = 1, 2$  and  $3$  in our analysis. We attempted to estimate the degrees of freedom  $\nu$  of the  $t_\nu$  distribution by treating it as a true parameter in our model, but found its estimation unstable without use of a strongly informative prior. Hence we perform a sensitivity analysis, comparing results from a normal model with a submodel with  $t_4$  and  $t_7$  assumptions, respectively, based on Jeffreys' (1973, p.65) suggestion to replace the normality assumption with a  $t$ -distribution with degrees of freedom in the range of 4 to 15. We chose these three scenarios as representative settings to reflect the assumptions of presence of extreme outliers, mild outliers or absence of outliers relative to a normal distribution in the FSH data.

Table 1 presents the DIC statistics for all models considered: 1,2, or 3 latent classes for the mean trajectories and variances; normal,  $t_7$  and  $t_4$  assumptions for the errors in longitudinal submodel; and 1,3 or 5 knots for the longitudinal trajectories or functional varying coefficient function respectively. In general, DIC suggests that joint models with  $t_4$  assumption for the longitudinal submodel fits the data better than  $t_7$  and much better than the normal model.  $K_D = K_C = 2$  is selected for both  $t_4$  and  $t_7$ . Given these selected number of components for both the mean trajectory and variance classes for each model, DIC further suggests that 1 knot (i.e.,  $J_{\mu(t)} = 1$ ) at 46.6 years of age for the longitudinal trajectories and 3 knots (i.e.,  $J_{\theta_0(t)} = 3$ ) at 41.6, 46.6 and 51.5 years of age for the functional varying coefficient function offers the best balance between goodness of fit and smoothness under all these three longitudinal submodel assumptions. Thus we will focus on these three best fitting models:

- best fitting normal model:  $K_D = 1, K_C = 2$  with  $J_{\mu(t)} = 1$  at 46.6 years of age and  $J_{\theta_0(t)} = 3$  at 41.6, 46.6 and 51.5 years of age
- best fitting  $t_7$  and  $t_4$  models:  $K_D = K_C = 2$  with  $J_{\mu(t)} = 1$  at 46.6 years of age and  $J_{\theta_0(t)} = 3$  at 41.6, 46.6 and 51.5 years of age

For these best fitting models, PDMs also confirmed our previous finding based on model selection criterion DIC that the  $t_4$  model fits the longitudinal FSH trajectories better than the  $t_7$  and normal distribution. Figure 1 shows the upper bounds of the p values based on PDMs for longitudinal trajectories fitted by all three final models. If the upper bound of a p value is less than 0.05, there is strong evidence of inadequate fit. We see that the normal model fits the large majority of subjects well, with 7 individual trajectories being considered to have inadequate fit by PDMs. Out of these 7 individual trajectories, assuming a  $t$ -distribution

with 7 degrees of freedom improved the fits of 4 individual trajectories, leaving 3 individual trajectories with inadequate fit; among the 3 individual trajectories, assuming a  $t$ -distribution with 4 degrees of freedom resulted in only 2 individual trajectories with inadequate fit. Figure 2(a) shows the 2 trajectories that are considered to have inadequate fits by all three best fitting models based on PDMs. Figure 2(b) shows the 4 trajectories that have upper bounds of  $p$  values less than 0.05 by our best fitting normal model but upper bounds of  $p$  values greater than 0.05 by both our best fitting  $t_7$  and  $t_4$  models. Clearly, these plots suggest that  $t$  models with 4 and 7 degrees of freedom show considerably less influence by outlying observations than the normal model and they both have almost identical fits visually. Finally, Figure 2(c) shows random selected 4 trajectories that have upper bounds of  $p$  values greater than 0.05 by all three of our best fitting models: the normal and  $t_7$  and  $t_4$  show very similar fits. Therefore, the inadequate fit of longitudinal FSH trajectories identified by PDMs is likely due to these varying degrees of extreme outliers. Although we could consider even smaller degrees of freedom of  $t$ -distribution or more heavily tailed distribution for the longitudinal submodel to accommodate these extreme outlying observations, the  $t$  model with either 4 or 7 degrees of freedom already shows almost identical robustness to them and seems to provide reasonably good fit to more than 99% of the FSH data.

Next, we contrast the estimation results from these models to demonstrate the influence of not appropriately accommodating outlying observations. Figure 3 presents the mean trajectory components and two variance components identified by the three best fitting models. Consistent with the finding reported in Jiang et al. (2014), under the normal model assumption, a single-component mean trajectory is favored by DIC. In contrast, under both the  $t_7$  and  $t_4$  model assumptions, a two-component mean trajectory is favored by DIC: the major mean class (86% of women) whose FSH levels begin increasing in their late 40s and the minor mean class (14% of women) with increasing FSH levels starting around age 40 capturing a proportion of women who might transition into menopause at an earlier age. The variance class has different meanings under  $t$  and normal assumptions but in both scenarios measure the short term variations in FSH levels: according to their magnitudes, both  $t$  and normal models would classify them to either “low” or “high” variance classes. Based on the posterior estimates of these component-specific parameters given in Table 1 in the Web-based supporting material, we can see more subtle differences in these estimated mixture components under varying assumptions.

Table 2 shows that all three models reach the same broad conclusions: high short term variability (its effect is represented by  $\theta_3$ ) in the FSH levels is strongly associated with increased risks of more severe hot flashes; smoking (its effect is represented by  $\theta_2$ ) is marginally associated with more severe hot flashes, and there was no association with BMI (its effect is represented by  $\theta_1$ ) or the individual mean trajectories between age 45 and 55 (its cumulative time varying effect is represented by  $\theta_0(t)$ ). The most dramatic difference between the different df models occurs for the estimated functional coefficient  $\theta_0(t)$  that captures the cumulative time varying effect of the mean trajectory  $\mu_t(t)$ . Figure 4 (a), (b) and (c) show the estimated functional coefficient  $\theta_0(t)$  by our best fitting normal,  $t_7$  and  $t_4$  models, respectively. The estimated  $\theta_0(t)$  under our best fitting normal model tends to have

larger effect size (larger magnitude in  $\theta_0(t)$ ) before age 53 and an overall wider pointwise 95% credible interval than the estimated  $\theta_0(t)$ 's under our best fitting  $t_4$  and  $t_7$  models. All three coefficient curves suggest that, when adjusting for the whole history of mean FSH levels over the age range of age 45 to age 55, higher mean FSH levels before age 53 reduce risk of severe hot flashes, while higher mean FSH levels between age 53 and age 55 increase this risk, but there is no conclusive evidence of a true association between the FSH trajectory histories and the risk of more severe hot flashes.

Finally, to consider the effect of the derivative of the mean trajectory  $\mu'_i(t)$ , or the rate of change in the mean trajectory  $\mu_i(t)$ , we focus on the best-fitting  $t_4$  model. Figure 5(a) considers the effect of cumulative changes in the mean trajectories across the age range  $T = [45, 55]$ , while Figure 5(b) considers the equivalent effect across the age range  $T = [50, 55]$ , potentially a more clinically relevant age range since the median age of menopause is 51 and therefore the hormone dynamics in this time window are more likely to play a role in the menopause-related symptoms. When fit over the wider age range, higher values of  $\mu'_i(t)$  decrease risk slightly before age 50 and increase it over age 50, although the 95% credible intervals include 0 by a wide margin. In contrast, a more narrowly-focused age range of  $T = [50, 55]$  suggested significantly increased risk of severe hot flash associated with higher values of  $\mu'_i(t)$  in the age range of 52.5-55, with  $\theta_1(\hat{52.5}) = 0.408$  (95% CI=0.019, 0.843) and  $\theta_1(\hat{55}) = 0.514$  (95% CI=0.003, 1.290).

Figure 6 shows the receiver-operator characteristic (ROC) curves for the best-fitting  $t_4$  model, comparing the use of the  $\mu_i(t)$  and  $\mu'_i(t)$  between age 45 and 55 to discriminate each of the hot flash severities (0, 1 and 2), along with the other predictors (residual variance, BMI, and smoking status). These ROCs and their associated area under the curve (AUCs) suggest that using either functional predictors led to moderately accurate classifications of different hot flash severities. Visually, there is not much differences in these ROC curves; a further comparison of AUCs also suggests that the predictive performances by using both  $\mu_i(t)$  and  $\mu'_i(t)$  have negligible differences ( AUCs for severity 0, 1, and 2 are -0.012 (-0.097, 0.070), -0.002 (-0.073, 0.071) and -0.020 (-0.131, 0.091) respectively).

## 4 Conclusions and Discussion

In this paper we develop a novel joint modeling approach to answer the scientifically important research question of how long-term history of FSH values or their rate of change affects the risk of hot flash severity, a symptom almost every woman experiences during the menopausal transition. While many joint models have been developed in the context of cancer research and HIV/AIDS clinical trials in the past decade, most methods focus on the features in the “true” underlying longitudinal process (i.e., mean trajectory) that take the forms of random effects or latent classes; or alternatively the last available “true” underlying value as a time-dependent covariate. Following Elliott et al. (2012) and Jiang et al. (2014), we seek the useful longitudinal features in both the mean trajectories and the short-term variability. Further we allow the mean of the longitudinal process and the corresponding derivatives to be time varying, and their effects on the response to be accumulative over time. To summarize, we propose a broadly applicable joint modeling approach that



1. extends conventional functional data analysis to the framework of joint modeling of both the longitudinal (functional predictor) and outcome data, which allows us to study different aspects of the features in the dynamics of longitudinal process as functional predictors. In particular, we focus on the values and derivatives of the mean trajectories at certain time windows as potential functional predictors. This will allow us to identify ages of vulnerability and test hypotheses about the association between the functional predictors (FSH level, rate of change) and our outcome, severe hot flashes, while also adjusting for the previously identified effect of short term variability captured by the variance of the residuals (Jiang et al., 2014).
2. uses flexible mixed effects models with Bayesian penalized B spline basis and latent classes in the longitudinal submodel, which relaxes assumptions about the specific form of the trajectories and allows uneven spacing and unequal length that are densely or sparsely measured to be used as functional predictors.
3. allows the effects of FSH histories (the mean value or derivative) to be time varying and to accumulate over time. Statistical tests of these functional coefficient functions in the primary outcome submodel for hot flashes can then be used to identify critical time windows where the association exists. Using a Bayesian approach allows easy calculation of pointwise credible intervals for the functional coefficient functions in comparison to frequentist approaches.
4. uses a robust model to decrease the influence of outlying observations in the FSH data.

To realize these modeling goals, we use a penalized spline approach to allow the flexible modeling of longitudinal features and the functional coefficient curve representing the time varying effect of the longitudinal features. Since the ultimate goal is to simultaneously model both the mean trajectories and the residual variability but distinguish between their effects in the outcome submodel, we choose a  $t$ -distribution to properly model residual variability to avoid the impact of outlying FSH values. In particular, we demonstrate the importance of assuming this robust distribution assumption instead of the typical normal assumption used in most of the joint modeling literature. However, due to the limited number of longitudinal observations for some women (i.e., ranging from 6 to 26), there is insufficient information in the data to assume individually varying degrees of freedom in the  $t$ -distribution; thus we are limited to assume a global degrees of freedom common to all trajectories. In addition, our attempts to use the data to estimate even the global degrees of freedom parameter using the informative exponential distribution proposed by Geweke (1993), the truncated uniform prior on the inverse of the degrees of freedom suggested in Lange et al. (1989) and Gelman and Hill (2007) and the Jefferys prior derived by Fonseca et al. (2008) all failed: the estimated global degrees of freedom were always close to a prior cutoff value, implying the existence of extreme outliers in the FSH data that tend to drive the degrees of freedom in  $t$ -distribution to low values. Given that the fitted values are only modestly affected by different values of degrees of freedom in  $t$ -distribution (Lange et al., 1989), we chose to fix the degrees of freedom parameter at a small number of fixed values and conduct a sensitivity analysis using DIC to choose among the models.



The proposed model also allows latent heterogeneities in both the individual level mean trajectories and the residual variability as in Jiang et al. (2014). Under our best fitting  $t_4$  model, as shown in Figure 3 (e), the mean FSH trajectories can be separated into two classes, one minor class with 14% of trajectories and the other major one with 86% of trajectories. Both classes are reflective of three typical FSH change patterns for women in the transition to menopause (Burger et al., 1999) in that FSH is relatively flat prior to the menopause transition, has an increasing period during the menopause transition, and eventually plateauing once women are 2 years post menopause; but women in the minor class tend to have an earlier increases in FSH along with higher FSH values than the women in the major class. Figure 7 plots the fitted mean FSH curves for the 28 women assigned to the minor class and a random sample of 20 women assigned to the major class based on the posterior mode. This once again shows the heterogeneous nature in the mean FSH trajectories that is supported by our model selection criterion DIC and implies that the women in the minor class tend to reach menopause at a much earlier age. Also, as shown in Figure 3, even with the use of the  $t$ -distribution to account for extreme outlying observations, it seemed that there still exists a true mixture in residual variability, with a low-variance class consisting of one in three to one in five women, with the remainder in a high variance class.

In summary, the proposed model gives added insights about hormone changes in the menopausal transition and their associations with severe hot flashes. First, whether the robust or normal models were used, we identified a strong association between residual variability in FSH and hot flashes as in Jiang et al. (2014), and similarly to what has been reported for depressive symptoms (Freeman et al., 2006). In addition, we identified latent heterogeneities in both the individual level mean trajectories. Under our best fitting  $t_4$  model, as shown in Figure 3 (e), the mean FSH trajectories can be separated into two classes, one minor class with 14% of trajectories and the other major one with 86% of trajectories. Both classes are reflective of three typical FSH change patterns for women in the transition to menopause (Burger et al., 1999) in that FSH is relatively flat prior to the menopause transition, has an increasing period during the menopause transition, and will eventually plateau once women get about 2 years post menopause; but women in the minor class tend to have an earlier increase in their FSH trajectories along with higher FSH values than the women in the major class. As shown in Figure 7, the fitted mean FSH curves for the total 28 women assigned to the minor class and a random sample of 20 women assigned to the major class based on the posterior mode were plotted. This once again shows the heterogeneous nature in the mean FSH trajectories that is supported by our model selection criterion DIC and implies that the women in the minor class tend to reach menopause at a much earlier age. Also, as shown in Figure 3, even with the use of the  $t$ -distribution to account for extreme outlying observations, it seemed that there still exists a true mixture in residual variability. Another interesting finding is illustrated in Figure 5(b) depicting the association between increases in hot flashes and the functional coefficient which describes the rate of change in FSH between the ages 50-5. This age window corresponds precisely to when hot flashes are reported to be most likely (Harlow et al., 2012). These findings have important ramifications for treatment of hot flashes with hormone replacement therapy. These medications impact the levels of FSH and Estradiol, and reduce variability. The

current recommendation is for women to take these medications for no more than 3 to 5 years, however, the optimal timeframe and duration for treatment is unknown.

Generally, the functional coefficient curves  $\theta_0(t)$  and  $\theta_1(t)$  can be fit by any spline basis with or without penalty parameters. In particular, if the shape of  $\theta_0(t)$  or  $\theta_1(t)$  is known – for example,  $\theta_0(t)$  is a linear function – then we can let  $\psi^0(t) = (1, t)$  and assume a regular normal prior on the coefficients associated with basis function 1 and  $t$ . When the true shape of  $\theta_0(t)$  or  $\theta_1(t)$  is unknown, we recommend starting the analysis using a more flexible penalized approach to get some idea of the shape of  $\theta_0(t)$  or  $\theta_1(t)$ , which may be further reduced to simple parametric form to stabilize estimation of model parameters and reduce the length of pointwise credible or confidence intervals for  $\theta_0(t)$  or  $\theta_1(t)$ .

The methods presented for data augmentation of unobserved FSH values assumes MAR. For the FSH values missing due to age at enrollment or reasons such as a subject did not deliver a blood sample at a certain visit, we can reasonably assume MAR. One known non-random source of missingness would be when women went on hormone replacement therapy (HRT) for relief of menopausal symptoms. These hormone values during HRT were censored. In this subset who were symptom free at baseline, 31/234 = 13% reported any hormone therapy use over the 13 years of follow-up, and the majority, 26/31=84%, reported use at only 1 or 2 visits. Among the remaining 5 women, 3 women reported HRT use at 6 visits, 1 woman reported use at 4 visits, and 1 at 3 visits. However, women with skipped visits or dropout during the first 5 years (i.e., 10 visits) were less likely to be due to menopausal symptoms (Nelson et al., 2004). Furthermore, when fitting the individual's FSH trajectory assuming MAR, we did not observe noticeable irregular residual patterns from the FSH values collected before and after skipped visits; therefore the impact from assuming MAR for the sporadic missingness should be minimal. We may under-estimate the short-term variation if the missingness is associated with a high level of FSH fluctuation and this could be a worthy future research topic. For dropout, we may expect an impact if those that dropped out had different profiles after they left than those that stayed. There are a total of 26 women who dropped out after being in the program for more than 5 years. Among them, 10 women contribute 20 or more observations prior to the dropout and 5 women dropped out at age 54 or older. A preliminary study that examined FSH patterns and values in the visits prior to the dropout did not reveal a reason behind the dropout. Nor could we find an explanation behind their dropouts based on factors such as their history of hot flash severity, menopausal stage or HRT use. Future work will develop methods to thoroughly examine the sensitivity to different missing data mechanisms through pattern mixture models or selection models within our modeling framework, although the sensitivity of our results to failures of the MAR assumption as anticipated would be relatively minor given the limited amount of missing data.

Another direction for future work is to make use of the fact that longitudinal studies often measure several variables repeatedly, for example, in the Penn Ovarian Study several other hormone trajectories are available. Developing methods to model these potentially correlated longitudinal trajectories simultaneously while also using this information effectively to predict or relate to the outcome of interest is a key area for future research.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

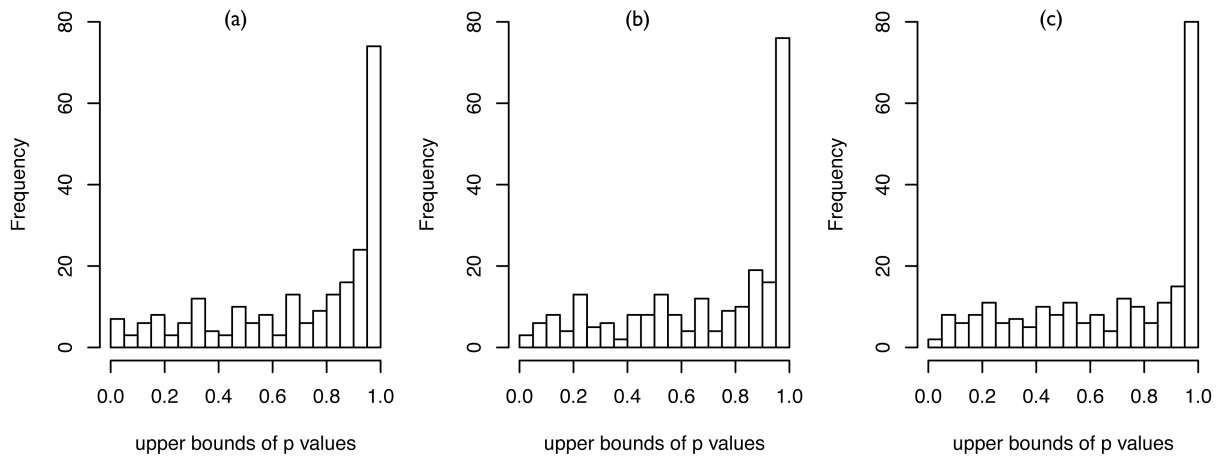
This work was supported in part by Grant Number R03AG031980 from the US National Institute of Aging and by Grant Number R01CA74552 from the US National Cancer Institute. The authors thank Ellen W. Freeman, PhD for sharing her data, as well as two reviewers and the Associate Editor for comments that helped to improve the manuscript.

## References

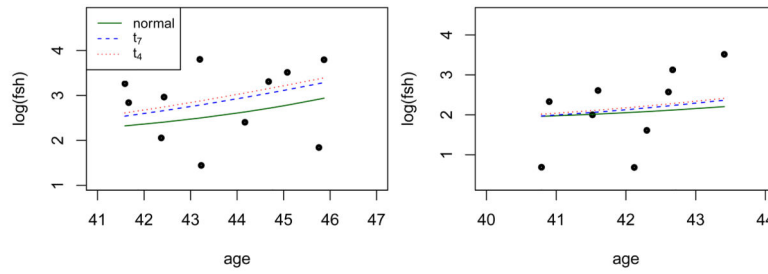
- Albert JH, Chib S. Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*. 1993; 88:669–679.
- Brown ER, Ibrahim JG. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*. 2003a; 59:221–228. [PubMed: 12926706]
- Brown ER, Ibrahim JG. Bayesian approaches to joint cure-rate and longitudinal models with applications to cancer vaccine trials. *Biometrics*. 2003b; 59:686–693. [PubMed: 14601770]
- Burger HG, Dudley EC, Hopper JL, Groome N, Guthrie JR, Green A, Dennerstein L. Prospectively measured levels of serum follicle-stimulating hormone, estradiol, and the dimeric inhibins during the menopausal transition in a population-based cohort of women. *Journal of Clinical Endocrinology & Metabolism*. 1999; 84:4025–4030. [PubMed: 10566644]
- Celeux G, Forbes F, Robert CP, Titterington DM. Deviance information criteria for missing data models. *Bayesian Analysis*. 2006; 1:651–673.
- Chen H, Wang Y. A penalized spline approach to functional mixed effects model analysis. *Biometrics*. 2011; 67:861–870. [PubMed: 21155747]
- Day NE. Estimating the components of a mixture of normal distributions. *Biometrika*. 1969; 56:463–474.
- Durbán M, Harezlak J, Wand MP, Carroll RJ. Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine*. 2005; 24:1153–1167. [PubMed: 15568201]
- Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. *Statistical Science*. 1996; 11:89–121.
- Elliott MR. Identifying latent clusters of variability in longitudinal data. *Biostatistics*. 2007; 8:756–771. [PubMed: 17267391]
- Elliott MR, Sammel MD, Faul J. Associations between variability of risk factors and health outcomes in longitudinal studies. *Statistics in Medicine*. 2012; 31:2745–2756. [PubMed: 22815213]
- Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters*. 2006; 27:861–874.
- Fonseca TC, Ferreira MA, Migon HS. Objective Bayesian analysis for the student-t regression model. *Biometrika*. 2008; 95:325–333.
- Freeman EW, Sammel MD, Lin H, Liu Z, Gracia CR. Duration of menopausal hot flushes and associated risk factors. *Obstetrics and Gynecology*. 2011; 117:1095. [PubMed: 21508748]
- Freeman EW, Sammel MD, Lin H, Nelson DB. Associations of hormones and menopausal status with depressed mood in women with no history of depression. *Archives of General Psychiatry*. 2006; 63:375–382. [PubMed: 16585466]
- Frühwirth-Schnatter, S. *Finite Mixture and Markov Switching Models*. New York: Springer; 2008.
- Garrett ES, Zeger SL. Latent class model diagnosis. *Biometrics*. 2000; 56:1055–1067. [PubMed: 11129461]
- Gelman, A.; Carlin, JB.; Stern, HS.; Rubin, DB. *Bayesian Data Analysis*. second. London: CRC press; 2003.
- Gelman A, Goegebeur Y, Tuerlinckx F, Van Mechelen I. Diagnostic checks for discrete data regression models using posterior predictive simulations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 2000; 49:247–268.

- Gelman, A.; Hill, J. *Data Analysis using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press; 2007.
- Gelman A, Meng XL, Stern H. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*. 1996; 6:733–760.
- Geweke J. Bayesian treatment of the independent student-t linear model. *Journal of Applied Econometrics*. 1993; 8:S19–S40.
- Harlow SD, Gass M, Hall JE, Lobo R, Maki P, Rebar RW, Sherman S, Sluss PM, de Villiers TJ. Executive summary of the stages of reproductive aging workshop+ 10: addressing the unfinished agenda of staging reproductive aging. *Climacteric*. 2012; 15:105–114. [PubMed: 22338612]
- Ibrahim JG, Chen MH, Sinha D. Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Statistica Sinica*. 2004; 14:863–884.
- Ibrahim JG, Chu H, Chen LM. Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*. 2010; 28:2796–2801. [PubMed: 20439643]
- James GM. Generalized linear models with functional predictors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2002; 64:411–432.
- Jasra A, Holmes C, Stephens D. Markov chain monte carlo methods and the label switching problem in bayesian mixture modeling. *Statistical Science*. 2005; 20:50–67.
- Jeffreys, H. *Scientific Inference*. third. New York: Cambridge University Press; 1973.
- Jiang B, Elliott MR, Sammel MD, Wang N. Joint modeling of cross-sectional health outcomes and longitudinal predictors via mixtures of means and variances. Submitted Manuscript. 2014
- Johnson VE. A Bayesian  $\chi^2$  test for goodness-of-fit. *The Annals of Statistics*. 2004; 32:2361–2384.
- Johnson VE. Bayesian model assessment using pivotal quantities. *Bayesian Analysis*. 2007; 2:719–734.
- Kass RE, Natarajan R. A default conjugate prior for variance components in generalized linear mixed models (comment on article by Browne and Draper). *Bayesian Analysis*. 2006; 1:535–542.
- Lang S, Brezger A. Bayesian P-splines. *Journal of Computational and Graphical Statistics*. 2004; 13:183–212.
- Lange KL, Little RJ, Taylor JM. Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*. 1989; 84:881–896.
- Law NJ, Taylor JM, Sandler H. The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure. *Biostatistics*. 2002; 3:547–563. [PubMed: 12933597]
- Little, RJ.; Rubin, DB. *Statistical analysis with missing data*. New York: Wiley; 2002.
- Muthén B, Shedden K. Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*. 1999; 55:463–469. [PubMed: 11318201]
- Neelon B, O'Malley AJ, Normand SLT. A Bayesian two-part latent class model for longitudinal medical expenditure data: assessing the impact of mental health and substance abuse parity. *Biometrics*. 2011; 67:280–289. [PubMed: 20528856]
- Nelson DB, Sammel MD, Freeman EW, Liu L, Langan E, Gracia CR. Predicting participation in prospective studies of ovarian aging. *Menopause*. 2004; 11:543–548. [PubMed: 15356407]
- Pemstein D, Quinn KM, Martin AD. The scythe statistical library: An open source C++ library for statistical computation. *Journal of Statistical Software*. 2007; 1:29.
- Proust-Lima C, Séne M, Taylor JM, Jacqmin-Gadda H. Joint latent class models for longitudinal and time-to-event data: a review. *Statistical Methods in Medical Research*. 2012; 23:74–90. [PubMed: 22517270]
- Ramsay JO, Dalzell C. Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 1991; 53:539–572.
- Redner RA, Walker HF. Mixture densities, maximum likelihood and the EM algorithm. *SIAM review*. 1984; 26:195–239.
- Rizopoulos, D. *Joint Models for Longitudinal and Time-to-Event Data: with Applications in R*. CRC Press; 2012.

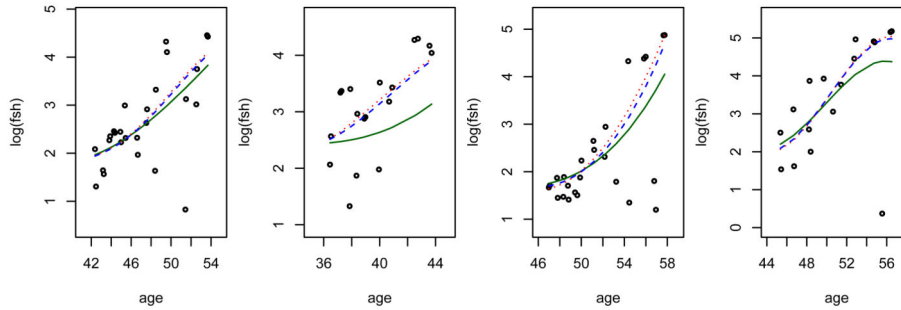
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. proc: an open-source package for r and s+ to analyze and compare ROC curves. *BMC bioinformatics*. 2011; 12:77. [PubMed: 21414208]
- Rodríguez CE, Walker SG. Label switching in Bayesian mixture models: Deterministic relabeling strategies. *Journal of Computational and Graphical Statistics*. 2012; 23:25–45.
- Ruppert, D.; Wand, MP.; Carroll, RJ. *Semiparametric Regression*. New York: Cambridge University Press; 2003.
- Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCr: visualizing classifier performance in R. *Bioinformatics*. 2005; 21:3940–3941. [PubMed: 16096348]
- Song X, Davidian M, Tsiatis AA. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics*. 2002; 58:742–753. [PubMed: 12495128]
- Sowers MR, Zheng H, McConnell D, Nan B, Harlow S, Randolph JF. Follicle stimulating hormone and its rate of change in defining menopause transition stages. *Journal of Clinical Endocrinology & Metabolism*. 2008; 93:3958–3964. [PubMed: 18647816]
- Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2002; 64:583–639.
- Stephens M. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2000; 62:795–809.
- Tsiatis A, Degruittola V, Wulfsohn M. Modeling the relationship of survival to longitudinal data measured with error. applications to survival and cd4 counts in patients with aids. *Journal of the American Statistical Association*. 1995; 90:27–37.
- Verbeke G, Lesaffre E. A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*. 1996; 91:217–221.
- Wang Y, Taylor JMG. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*. 2001; 96:895–905.
- Xu J, Zeger SL. The evaluation of multiple surrogate endpoints. *Biometrics*. 2001; 57:81–87. [PubMed: 11252622]
- Yu M, Taylor JMG, Sandler HM. Individual prediction in prostate cancer studies using a joint longitudinal survival–cure model. *Journal of the American Statistical Association*. 2008; 103:178–187.
- Yuan Y, Johnson VE. Goodness-of-fit diagnostics for Bayesian hierarchical models. *Biometrics*. 2012; 68:156–164. [PubMed: 22050079]



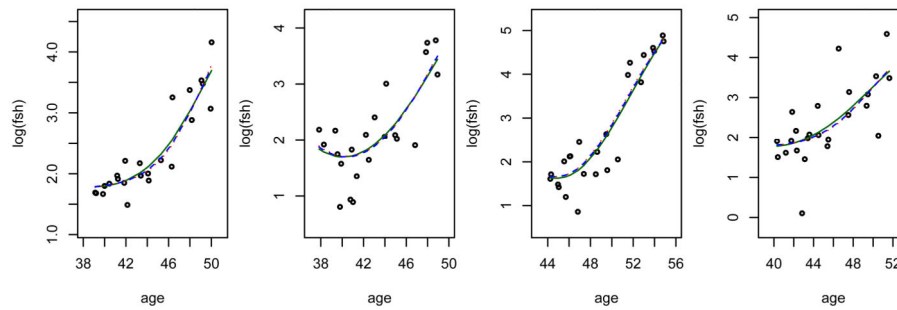
**Figure 1.** Upper bounds of p values based on PDMs for individual trajectories fit by our best fitting models with  $\mu_i(t)$   $i = 1, \dots, n$  within the time window  $T = [45, 55]$  as a functional predictor in primary outcome submodel: (a) best fitting normal model; (b) best fitting  $t_7$  model; (c) best fitting  $t_4$  model.



(a) Fitted trajectories by all three models have upper bounds of p values based on PDMs for individual trajectories of less than 0.05



(b) Fitted trajectories by the normal model have upper bounds of p values based on PDMs for individual trajectories less than 0.05 under the normality assumption but upper bounds of p value greater than 0.05 under  $t_4$  and  $t_7$  model assumptions

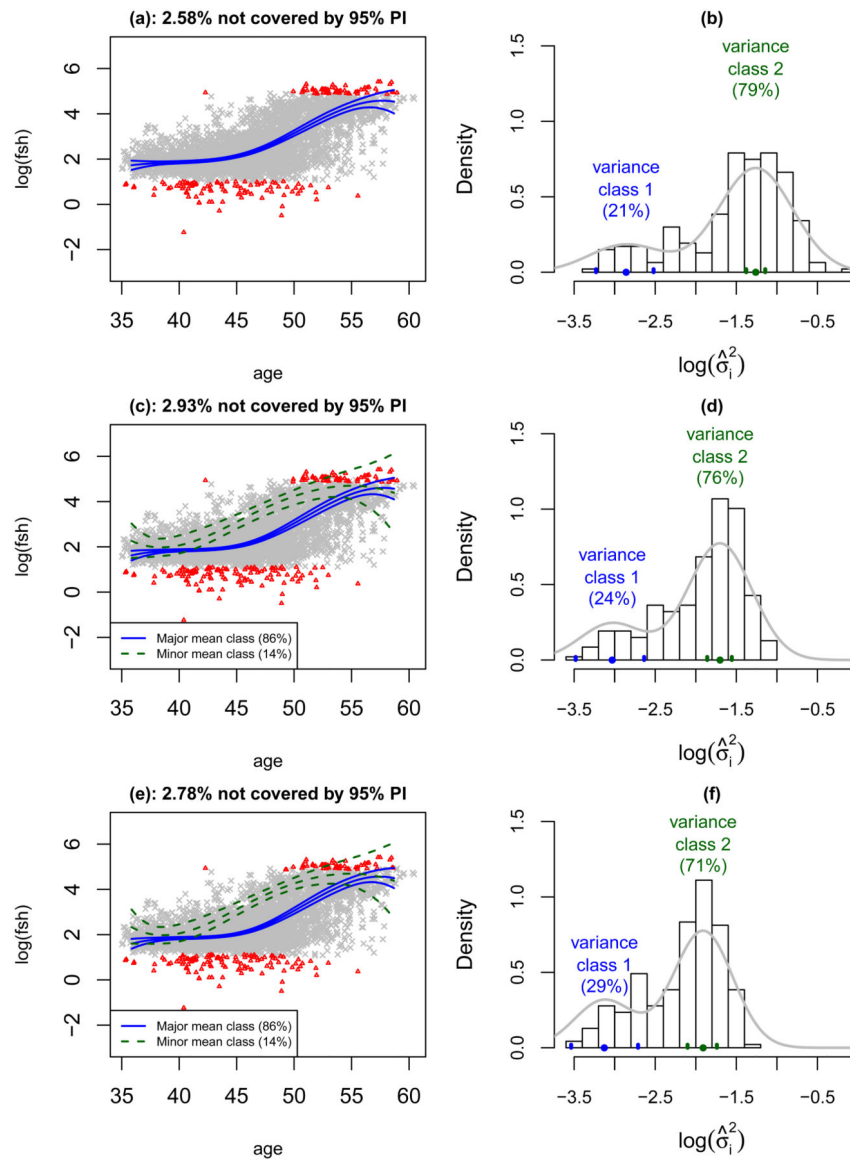


(c) Fitted trajectories under all three models have upper bounds of p values based on PDMs for individual trajectories of greater than 0.05

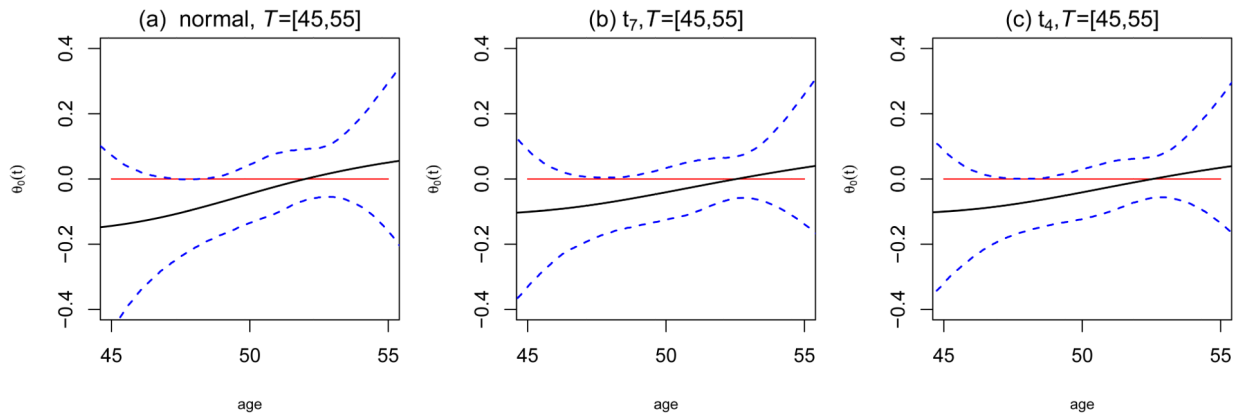
**Figure 2.**

Selected individual FSH trajectories from the Penn Ovarian Aging Study fitted by our best-fitting joint models with  $\mu_i(t)$   $i = 1, \dots, n$  within the time window  $T = [45, 55]$  as a functional predictor in primary outcome submodel.



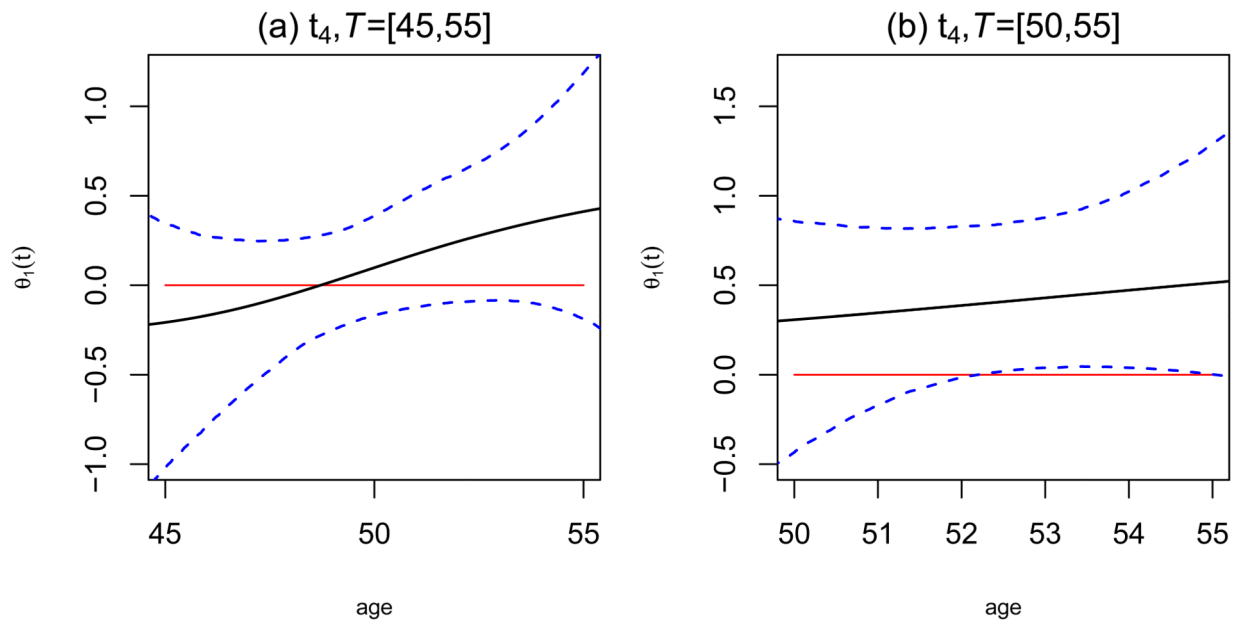


**Figure 3.** Longitudinal mean trajectories for the Penn Ovarian Aging Study from our final models with  $J_\mu = 1$  and  $K_D = K_C = 2$  in longitudinal submodel;  $\mu_i(t)$  as functional predictor with time window  $T = [45, 55]$  and  $J_{\theta_0} = 3$  in primary outcome submodel with different assumptions for longitudinal submodel: a) and b) under normal assumption; c) and d) under  $t_7$  assumption; e) and f) under  $t_4$  assumption.



**Figure 4. Functional coefficient function**

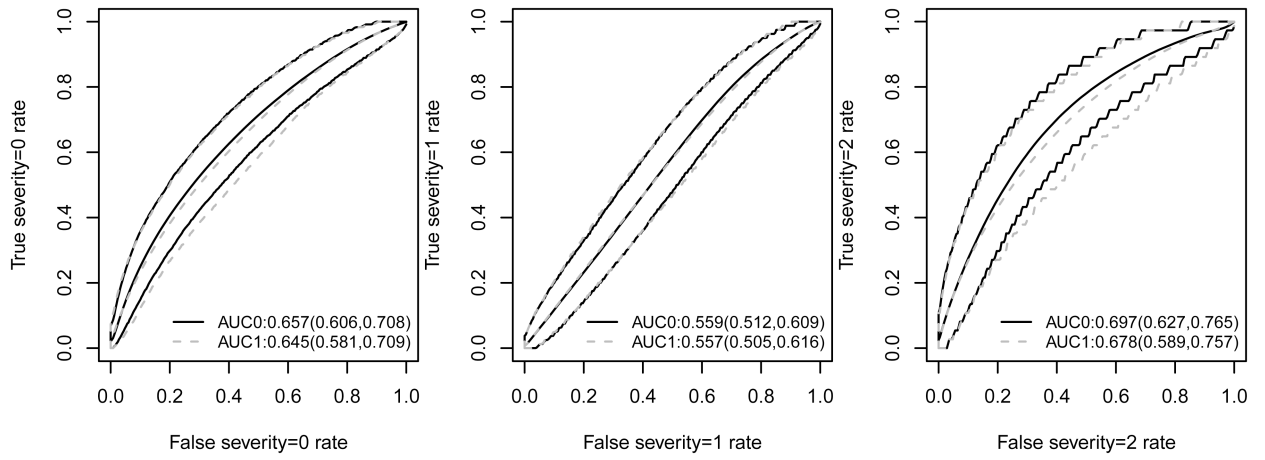
$\theta_0(t)$  for the Penn Ovarian Aging Study from our best fitting  $t_4$ ,  $t_7$  and normal models with  $\mu_i(t)$ ,  $i = 1, \dots, n$  within the time window  $T = [45, 55]$  as a functional predictor in primary outcome submodel.



**Figure 5. Functional coefficient function**

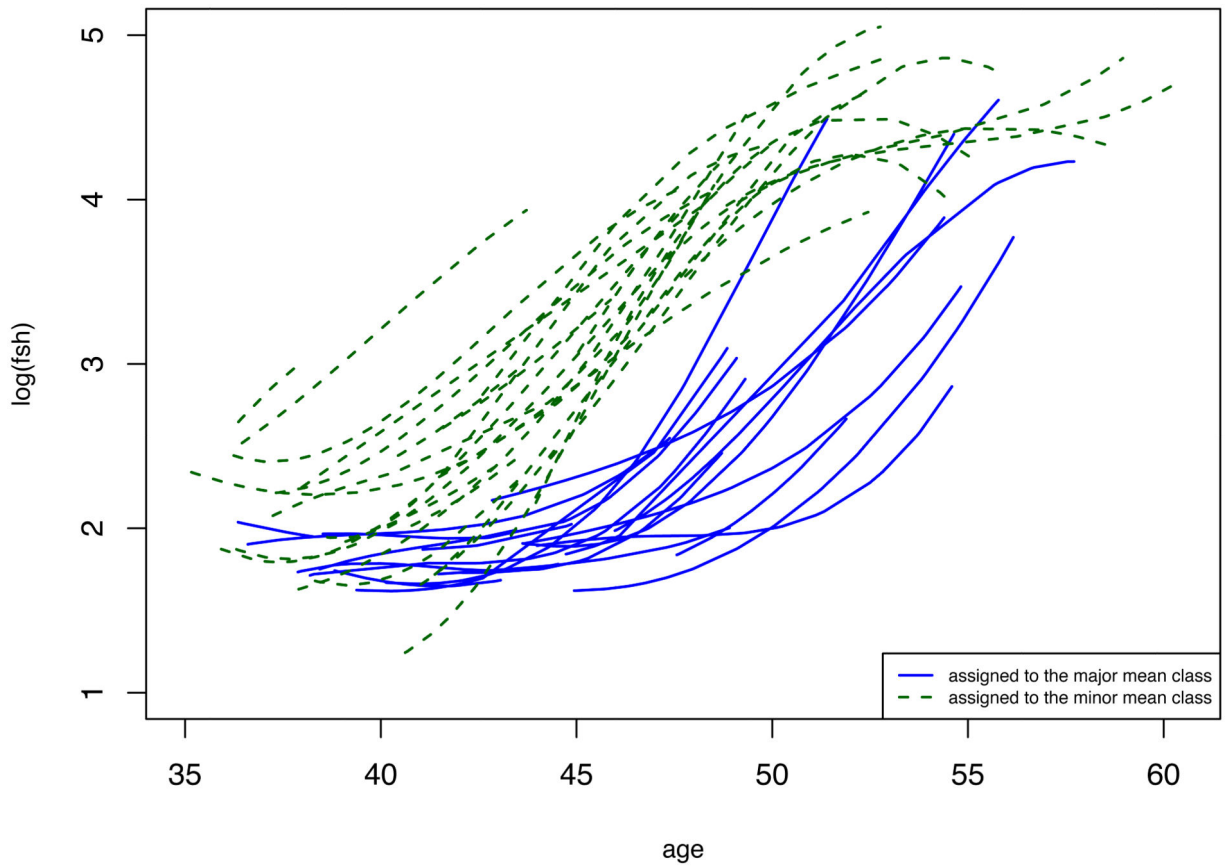
$\theta_1(t)$  for the Penn Ovarian Aging Study from our best-fitting model with  $J_\mu = 1$  and  $K_D = K_C = 2$  in longitudinal submodel with  $t_4$  assumption; and  $J_{\theta_1} = 3$  in primary outcome submodel:

a)  $\mu'_i(t)$  as functional predictor with  $T = [45, 55]$  and b)  $\mu'_i(t)$  as functional predictor with  $T = [50, 55]$ .



**Figure 6. ROC curves**

for the Penn Ovarian Aging Study from our final t model: AUC0 is obtained by using  $\mu_i(t)$  with  $J_{\theta(t)} = 3$  within the time window  $T = [45, 55]$  as a functional predictor in outcome submodel and AUC1 is obtained by using  $\mu_i'(t)$  with  $J_{\theta_1(t)} = 3$  within the time window  $T = [45, 55]$  as a functional predictor with in outcome submodel.



**Figure 7.** Individual FSH trajectories from the Penn Ovarian Aging Study that are assigned to the minor and major mean trajectory classes by our best fitting  $t_4$  model with  $\mu_i(t)$   $i = 1, \dots, n$  within the time window  $T = [45, 55]$  as a functional predictor in primary outcome submodel.

DIC from different joint models for the analysis of the Penn Ovarian Aging data, assuming normal,  $t_7$  and  $t_4$  distribution for the longitudinal submodel and using  $J_{\mu(t)} i = 1, \dots, n$  within the time window  $T = [45, 55]$  as a functional predictor in primary outcome submodel.

**Table 1**

Model	$K_C = 1$			$K_C = 2$			$K_C = 3$		
	$K_D = 1$	$K_D = 2$	$K_D = 3$	$K_D = 1$	$K_D = 2$	$K_D = 3$	$K_D = 1$	$K_D = 2$	$K_D = 3$
<b>normal</b>									
$J_{\mu(t)} = 1, J_{\theta(t)} = 1$	11439.0	11477.2	11492.9	<b>11333.6</b>	11369.1	11399.3	11511.6	11545.1	11560.7
$J_{\mu(t)} = 1, J_{\theta(t)} = 3$	11437.5	11487.9	11501.8	<b>11327.7</b>	11364.9	11386.7	11506.8	11542.9	11561.5
$J_{\mu(t)} = 1, J_{\theta(t)} = 5$	11435.0	11480.5	11493.3	<b>11330.6</b>	11369.1	11385.7	11500.7	11552.2	11574.9
$J_{\mu(t)} = 2, J_{\theta(t)} = 1$	11923.4	11912.4	11924.6	11809.6	11788.7	11798.9	12000.1	11977.5	11984.4
$J_{\mu(t)} = 2, J_{\theta(t)} = 3$	11923.8	11901.3	11915.5	11807.0	11803.5	11799.8	11995.0	11971.6	11997.1
$J_{\mu(t)} = 2, J_{\theta(t)} = 5$	11924.7	11892.4	11919.2	11799.7	11788.2	11801.4	11993.1	11965.6	11991.5
$J_{\mu(t)} = 3, J_{\theta(t)} = 1$	12419.3	12400.5	12418.6	12319.9	12308.2	12316.5	12506.2	12489.0	12499.3
$J_{\mu(t)} = 3, J_{\theta(t)} = 3$	12421.8	12398.8	12412.5	12317.6	12306.7	12320.6	12506.5	12486.7	12489.2
$J_{\mu(t)} = 3, J_{\theta(t)} = 5$	12416.6	12399.3	12409.5	12317.0	12298.1	12307.5	12504.7	12472.7	12485.0
<b><math>t_4</math></b>									
$J_{\mu(t)} = 1, J_{\theta(t)} = 1$	10335.0	10257.5	10271.0	10303.3	<b>10215.4</b>	10246.8	10425.0	10326.3	10347.2
$J_{\mu(t)} = 1, J_{\theta(t)} = 3$	10333.2	10255.7	10272.5	10308.8	<b>10210.8</b>	10235.5	10419.9	10330.3	10374.1
$J_{\mu(t)} = 1, J_{\theta(t)} = 5$	10331.2	10260.0	10273.9	10298.5	<b>10230.4</b>	10228.3	10432.3	10322.7	10371.9
$J_{\mu(t)} = 2, J_{\theta(t)} = 1$	10831.8	10823.6	10826.4	10803.1	10774.6	10778.2	10947.6	10906.7	10889.1
$J_{\mu(t)} = 2, J_{\theta(t)} = 3$	10830.0	10821.0	10833.2	10821.3	10776.0	10812.1	10929.6	10897.9	10934.2
$J_{\mu(t)} = 2, J_{\theta(t)} = 5$	10828.0	10818.8	10822.3	10818.0	10780.1	10791.6	10936.8	10914.8	10922.0
$J_{\mu(t)} = 3, J_{\theta(t)} = 1$	11280.6	11259.2	11256.8	11287.8	11255.8	11257.5	11406.5	11369.9	11397.4
$J_{\mu(t)} = 3, J_{\theta(t)} = 3$	11275.4	11251.5	11256.8	11276.3	11251.4	11271.0	11393.9	11356.3	11382.0
$J_{\mu(t)} = 3, J_{\theta(t)} = 5$	11278.3	11250.5	11265.0	11298.1	11253.6	11264.5	11409.9	11381.4	11384.1
<b><math>t_7</math></b>									
$J_{\mu(t)} = 1, J_{\theta(t)} = 1$	10626.5	10585.0	10606.3	10566.9	<b>10518.2</b>	10533.3	10679.8	10603.3	10652.2

Model	$K_C = 1$			$K_C = 2$			$K_C = 3$		
	$K_D = 1$	$K_D = 2$	$K_D = 3$	$K_D = 1$	$K_D = 2$	$K_D = 3$	$K_D = 1$	$K_D = 2$	$K_D = 3$
$J_{\mu(t)} = 1, J_{\theta(t)} = 3$	10624.0	10584.2	10600.6	10567.8	<b>10511.5</b>	10532.0	10694.9	10633.9	10648.5
$J_{\mu(t)} = 1, J_{\theta(t)} = 5$	10622.5	10579.8	10598.3	10558.1	<b>10512.0</b>	10536.6	10670.4	10615.5	10628.7
$J_{\mu(t)} = 2, J_{\theta(t)} = 1$	11127.3	11114.8	11125.2	11065.8	11051.9	11067.9	11214.9	11205.2	11201.2
$J_{\mu(t)} = 2, J_{\theta(t)} = 3$	11123.7	11116.2	11132.3	11074.7	11062.0	11061.8	11210.6	11195.2	11207.4
$J_{\mu(t)} = 2, J_{\theta(t)} = 5$	11126.5	11115.4	11128.0	11069.1	11055.4	11056.6	11225.2	11185.2	11206.9
$J_{\mu(t)} = 3, J_{\theta(t)} = 1$	11604.1	11582.4	11585.9	11570.0	11550.0	11544.7	11652.8	11651.3	11661.6
$J_{\mu(t)} = 3, J_{\theta(t)} = 3$	11601.5	11577.1	11588.5	11572.0	11541.7	11547.4	11687.8	11644.0	11672.1
$J_{\mu(t)} = 3, J_{\theta(t)} = 5$	11600.6	11586.8	11587.9	11569.2	11540.3	11548.9	11672.2	11671.7	11651.9



Estimates of the regression coefficients in the outcome model for the Penn Ovarian Aging Study by our best fitting models with  $\mu_i(t) i = 1, \dots, n$  within the time window  $T = [45, 55]$  as a functional predictor in primary outcome submodel.

**Table 2**

	normal				$t_7$				$t_4$			
	mean	se	95% CI		mean	se	95% CI		mean	se	95% CI	
$\alpha_0$ (intercept)	0.305	0.995	(-.631, 2.329)		0.279	0.972	(-.637, 2.268)		0.012	0.985	(-.886, 1.979)	
$\lambda_{01}$ (log(BMI))	0.068	0.277	(-.501, 0.607)		0.039	0.264	(-.497, 0.573)		0.101	0.273	(-.449, 0.627)	
$\lambda_{02}$ (smoking)	<b>0.386</b>	0.170	(0.052, 0.717)		<b>0.370</b>	0.170	(0.039, 0.708)		<b>0.371</b>	0.171	(0.036, 0.708)	
$\lambda_{03}$ (variance)	<b>1.576</b>	0.565	(0.498, 2.703)		<b>1.887</b>	0.747	(0.451, 3.394)		<b>1.960</b>	0.723	(0.579, 3.403)	