Chapter 3
Bayesian Functional Mixed Models for Survival Responses with Application to Prostate Cancer

Veerabhadran Baladandayuthapan, Xiaohui Wang, Bani K. Mallick and Kim-Anh Do

Abstract In this chapter, we propose a flexible approach to model functional measurements for survival outcomes. Often the class of models for functional observations are assumed to be linear, which may be too restrictive in some cases. We propose an alternative model, in which the simple linear mixed model has been modified by a more flexible semiparametric spline-based functional mixed model, wherein the usage of splines simplifies parameterizations and the joint modeling framework allows synergistic benefit between the regression of functional predictors and the modeling of survival data. We explicitly model the number and location of change points such that our formulation allows for an unknown set of basis functions characterizing the population-averaged and patient-specific trajectories. In addition, we propose a novel auxiliary variable scheme for a fully Bayesian estimation of our model, which not only allows dimension reduction of the parameter space but also allows efficient sampling from the conditional distributions. We illustrate our approach with a recent prostate cancer clinical trial study.

3.1 Introduction

Metastatic prostate cancer is the second most common cancer-related cause of death in North American men (Greenlee et al. 2000). Hormonal treatments such as androgen ablation (AA) have been preferred treatments for metastatic prostate cancer for more than 50 years. Such therapies work by altering the natural history of the disease by specifically disrupting the growth-promoting effects mediated by androgen

V. Baladandayuthapani (✉) K. Do
Department of Biostatistics, UT MD Anderson Cancer Center, Houston, TX 77030, USA
e-mail: veera@mdanderson.org

X. Wang
Department of Mathematics, University of Texas-Pan American, Edinburg, TX 78539, USA
e-mail: xhwang@utpa.edu

B. K. Mallick
Department of Statistics, Texas A & M University, College Station, TX 77840, USA

Z. Chen et al. (eds.), *Applied Statistics in Biomedicine and Clinical Trials Design,*
ICSA Book Series in Statistics, DOI 10.1007/978-3-319-12694-4_3

receptor signaling. Regardless of the mode of administration of AA, most patients with clinically detectable metastatic disease will eventually progress to androgen-independent prostate cancer (AIPC) with a median of 12–18 months (Eisenberger et al. 1986). After progression to AIPC, only symptoms are treatable and patients survive with a median of less than a year (Tannock et al. 1996). Despite major efforts, most studies with various cytotoxic drugs have provided little hint of the disease-altering activity for AIPC. However, in a recent phase II study at the University of Texas M.D. Anderson Cancer Center, a regimen based on chemotherapy demonstrated a survival advantage over historical results (Ellerhorst et al. 1997). This regimen of ketoconazole and doxorubicin alternating with vinblastine and estramustine, termed KA/VE, produced obvious palliation in the majority of treated patients.

Based on these results, a phase III trial (Millikan et al. 2008) was conducted at M.D. Anderson Cancer Center to compare conventional hormonal therapy (AA) to chemohormonal (CH) therapy combined with three 8-week cycles of KA/VE (AA + CH) in patients with metastatic androgen-driven prostate cancer. The hypothesis of interest was that early intervention of KA/VE to standard, sustained AA would delay the emergence of AIPC and ultimately prolong survival. The primary end point of interest was the time to progression to AIPC.

In addition to the time to progression, the longitudinal measurements of prostate-specific antigen (PSA) level from each patient over time were recorded. PSA, a glycoprotein produced by the prostate gland, is considered a useful biomarker for prostate cancer since significant positive correlation has been observed between the levels of PSA and the volume of the prostate (Catalona et al. 1991). Monitoring PSA levels has not only been established as a good diagnostic tool but is also considered an important indicator of response to treatment, with low levels indicating good prognosis. PSA measures are easy to collect via a routine laboratory assay of the blood samples. Thus, given the two sets of measurements: PSA profiles and time to progression (to AIPC), and since the measurements are inherently correlated, our main goal of this chapter is to investigate methods for the joint analysis of both end points.

In practice (and as in our case), the latent functional process is often unobservable due to measurement error and is not available at all times, especially when failure occurs. It is well known that conventional partial likelihood approaches for the Cox model cannot avoid biased inference by using imputation of the latent functional process, such as last value carried forward method (Prentice 1982), smoothing techniques (Raboud et al. 1993), and any other generic two-stage approaches (Bycott and Taylor 1998; Tsiatis et al. 1995). This invoked the consideration of using functional and event processes simultaneously via *joint modeling,* a subject that has recently attracted substantial interest (see Ibrahim et al. 2001; Tsiatis and Davidian 2004 for an overview).

Suppose the data are comprised of a vector of observations $\{T_i, \mathbf{L}_i, \mathbf{Y}(\mathbf{t}_i),\ \mathbf{t}_i \geq 0\}$ for the ith subject, where T_i is an event time (possibly censored), $\mathbf{L}_i$ is a vector of baseline covariates, and $\{\mathbf{Y}(\mathbf{t}_i),\ t_{i\ell} \geq 0,\ i = 1, \cdots, n,\ \ \ell = 1, \cdots, p_i\}$ is the functional marker trajectory for all times $t_{i\ell} \geq 0$, where p_i is the number of functional

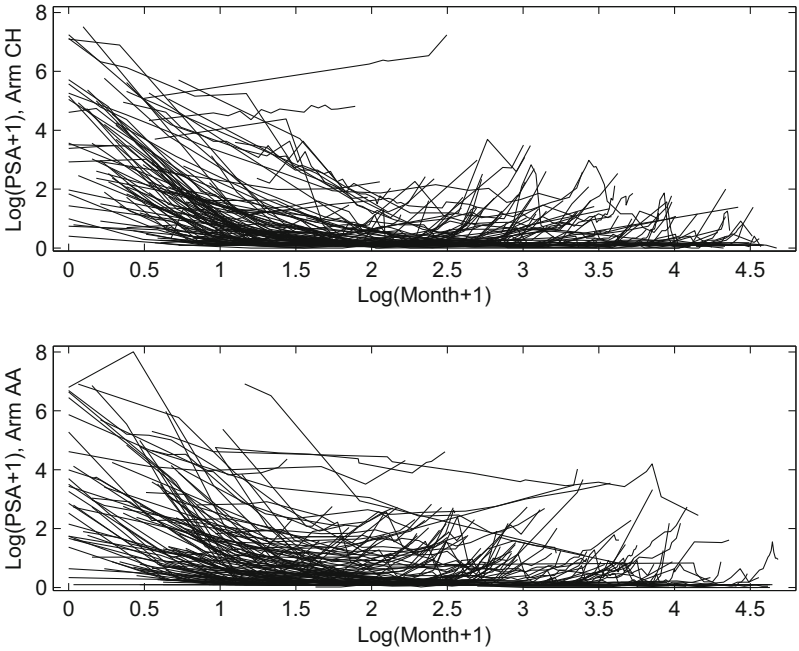

Fig. 3.1 Prostate-specific antigen (PSA) profiles for patients in arm androgen ablation (*AA*; upper panel) and arm chemohormonal (*CH*; lower panel)

measurements for subject i. One simple strategy is to introduce subject-specific random effects and then subsequently couple this model with a model on the survival process such as a proportional hazards model (Wulfsohn and Tsiatis 1997; DeGruttola and Tu 1994; Hogan and Laird 1997). A similar Bayesian method was explored by Faucett and Thomas (1996). Wang and Taylor (2001) introduced an integrated Orstein–Uhlenbeck (IOU) process into the functional modeling. Brown and Ibrahim (2003) started with a model similar to the ones in Wulfsohn and Tsiatis (1997) and Faucett and Thomas (1996) for their Bayesian semiparametric joint model; however, they used a quadratic form for the functional part and introduced a nonparametric specification for the distribution of the random effects, $\boldsymbol{\theta}_i$'s. Recent works include Zhang et al. (2009) proposing a semiparametric model based on Pólya trees and Guo and Carlin (2004) comparing separate and joint modeling of functional and event time data.

In most of these approaches, the form of the functional process or the *trajectory function* is assumed to be a simple parametric form. Although conceptually simple and easily implementable, this is a rather rigid assumption and may not hold in some cases such as the one we describe here. Figure 3.1 shows the overlapping PSA levels for the two treatment arms (AA and CH) posttreatment. The horizontal axes present the time (in logarithm of months) and the vertical axes present the log(PSA+1) measurements.

There are three key aspects of the PSA trajectories which need to be considered for any downstream analysis. First, there seems to be a definite overall pattern in the PSA trajectories for both treatment arms. The PSA levels decrease from time units 0 to 1, then stabilize and finally increase again (around 2.5) after the effect of treatment wears off. The patients have been normalized such that all patients receive their treatment at time 0. Thus, the profiles exhibit a nonlinear characteristic with definite change points at both the subject-specific and population levels—hence the need for flexible models for the functional process. Second, a further complication occurs since the number of PSA measurements for each patient are taken at different times, which causes them to be sparse and irregular. Third, there seems to be considerable heterogeneity among the patients in both treatment groups.

Due to these characteristics, the above mentioned parametric models might not be suitable for modeling such data. Specific to joint modeling of PSA and survival outcomes, Pauler and Finkelstein (2002) used a joint Bayesian model that consisted of piecewise linear functional model and Cox proportional hazard model. Their piecewise linear regression model adopted single unknown change point for each patient and implied independence assumption over functional measurements from the same patient. Ye et al. (2008) gave likelihood-based two-stage regression calibration methods to study the dependence of the risk of prostate cancer recurrence on the PSA level as well as time-independent covariates. Ye et al. (2008) provided a Bayesian-based joint modeling approach with added mixture structure to predict individual disease progression that results in either cure by treatment or susceptible to recurrence. The Ye et al. method models the PSA level with a nonlinear exponential decay and exponential growth model. We propose an alternative model in which the simple linear (or polynomial) model has been modified by a more flexible nonparametric model that cannot only capture nonlinear complex processes but also adopt unknown number of change points at both patient and population levels. We compare two different treatments, explore the effects of PSA level as well as several covariates on the survival outcome, and identify the PSA trajectory change points as patient disease progresses.

There has been an increasing interest in functional data analysis (FDA), analysis of data that are in the form of a (smooth) sample of curves or functions (Ramsay and Silverman 2005; Ngo and Wand 2004; Yao 2007; and Brown et al. 2005), in which the functions form the basic unit of data. Most functional data analyses focus on data which are frequently and regularly sampled across individuals and are not applicable here due to "sparseness and irregularity" of our data. We focus on methods for sparse functional data where not only the number and timing vary across subjects but also some subjects may be sampled at very few time points. Our mixed model uses a flexible spline basis; the usage of this basis simplifies the parameterizations and the joint modeling framework, thus allowing synergistic benefits between the regression of functional and survival data. Further, we explicitly model the number and location of change points such that our formulation allows for an unknown set of basis functions characterizing the population-averaged and patient-specific trajectories. We set up the spline-based model without the assumption of independence over functional measurements from the same patient. Meanwhile,

the novelty of the proposed Bayesian model lies in its ability to draw information from the functional data as well as from the associated event time data by unifying the spline-based functional regression and survival models. In addition, we propose a novel auxiliary variable scheme for a fully Bayesian estimation of our model, which not only allows for dimension reduction of the parameter space but also allows for efficient sampling from the conditional distributions and greatly reduces the computational burden.

The rest of the chapter is organized as follows. Section 3.2 discusses our Bayesian joint hierarchical model, where we set up the functional regression model and the Cox proportional hazards model in an unified framework. Section 3.3 concerns elicitation of prior distributions for the proposed model. Section 3.4 compares our model with other joint models with parametric regression segments based on various model selection criteria. The novel proposed model is illustrated by a motivating example, prostate cancer data set, in Sect. 3.5. The chapter is concluded with a discussion in Sect. 3.6. All technical details are collected into the Appendix.

3.2 Probability Model

In this section, we propose a joint survival and functional model in which the functional curves are modeled nonparametrically via splines. In addition, we explicitly model the change points present in the profiles via a functional variable selection approach, which results in a more flexible and robust model. For ease of exposition, we assume a univariate functional outcome, although our method is easily generalizable to multiple functional outcomes, as we show in Sect. 3.6.

3.2.1 Regression Model for the Functional Covariates

Suppose our data construct for n subjects consists of the following: $\{\mathrm{T_i}, \mathrm{C_i}, \mathbf{L_i}, \mathbf{Y_i(t)}\}$, where for the ith subject we observe a time-independent baseline covariates vector $\mathbf{L}_i$ of dimension m and time-dependent covariates $\mathbf{Y}_i(\mathbf{t})$ measured at time points $\mathbf{t}$. In addition, each individual has a lifetime $\mathrm{T_i}$ and a (right) censored time $\mathrm{C_i}$. Thus, one observes $T_i = \min(\mathrm{T_i}, \mathrm{C_i})$ and the failure indicator δ_i, defined as

$$\delta_i = \begin{cases} 1 & \text{if} \quad \mathrm{T_i} \leq \mathrm{C_i}, \\ 0 & \text{if} \quad \mathrm{T_i} > \mathrm{C_i}. \end{cases}$$

We further assume that the censoring mechanism is independent of all other survival and covariates information. For the functional covariate predictor $\mathbf{Y}_i(\mathbf{t})$, we posit the following functional regression model:

$$\mathbf{Y}_i(\mathbf{t}) = \boldsymbol{\mu}(\mathbf{t}) + \mathbf{b}_i(\mathbf{t}) + \boldsymbol{\epsilon}_i(\mathbf{t}) \quad 0 \leq t \leq T, \;\; 1 \leq i \leq n,$$

where $\boldsymbol{\mu}(\bullet)$ is the overall mean profile and $\mathbf{b}_i(\bullet)$ is the ith subject's deviation from the mean profile, measured intermittently between times $[0, T]$ for an individual i with measurement error $\boldsymbol{\epsilon}_i(\bullet)$. We also assume that the error process $\boldsymbol{\epsilon}_i(\bullet)$ is independent of true functional process and follows a Gaussian process with mean zero and constant variance $\sigma^2_{\boldsymbol{\epsilon}}$. Other forms of correlations such as an autocorrelation process can be used for the errors, but to keep the exposition simple we do not consider that case here.

Our focus is on modeling $\boldsymbol{\mu}(\bullet)$ and $\mathbf{b}_i(\bullet)$ in a flexible manner. We achieve this via a basis function projection:

$$\boldsymbol{\mu}(\mathbf{t}) = \mathbf{X}(\mathbf{t})\boldsymbol{\beta}, \quad \mathbf{b}_i(\mathbf{t}) = \mathbf{X}(\mathbf{t})\boldsymbol{\beta}_i,$$

where $\mathbf{X}(\mathbf{t})$ is any generic basis function and the associated regression coefficients are denoted by $\boldsymbol{\beta}$ for the overall mean and $\boldsymbol{\beta}_i$ for subject i. In practice, we only observe the latent functional process on a finite number of time points $\mathbf{t}_i = (t_{i1}, \ldots, t_{ip_i})$ for the ith subject with p_i as the number of measurements, which varies from subject to subject. The discretized version of the model for the observed PSA measurements Y_{ij} for subject i and time t_{ij} is of the form:

$$Y_{ij} = Y_i(t_j) = \mathbf{X}(t_{ij})\boldsymbol{\beta} + \mathbf{X}(t_{ij})\boldsymbol{\beta}_i + \epsilon_{ij}, \tag{3.1}$$

where $\mathbf{X}(t_{ij})$ is the basis function evaluated at $t_{ij}, i = 1, \ldots, n, j = 1, \ldots, p_i$. There are various basis functions that one could potentially use for modeling the functional predictors, such as smoothing splines, B-splines, and wavelets, among others, depending on the application. For our model exposition, we use a truncated power series basis function (Ruppert et al. 2003) given its nice connections to mixed models (Ngo and Wand 2004).

Let dimension $K = 1 + p + K^*$, where p is the degree of the spline and K^* is the number of interior knots. We rewrite (3.1) in matrix notation as

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \tag{3.2}$$

where $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ip_i})'$, $\mathbf{X}_i$ is the $p_i \times K$ basis matrix for the ith subject, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K)'$ and $\boldsymbol{\beta}_i = (\boldsymbol{\beta}_{i1}, \ldots, \boldsymbol{\beta}_{iK})'$ are the K-dimensional regression coefficient vectors. The jth row of $\mathbf{X}_i$ can be written as

$$\mathbf{X}_{ij} = [1 \;\; t_{ij} \;\; t_{ij}^2 \;\; \cdots \;\; t_{ij}^p \;\; (t_{ij} - \mathrm{t}_1)_+^p \;\; \cdots \;\; (t_{ij} - \mathrm{t}_{K^*})_+^p],$$

where $\{\mathrm{t}_1, \cdots, \mathrm{t}_{K^*}\}$ are the interior knots. We assume that the subject level regression coefficients follow a Gaussian distribution, $\boldsymbol{\beta}_i \sim MVN(\mathbf{0}, \boldsymbol{\Omega})$, which are the random effects corresponding to the systematic deviation from the population mean $\boldsymbol{\beta}$ with a variance–covariance matrix $\boldsymbol{\Omega}$. This distribution implicity makes two key assumptions. First, it induces the same basis function and hence the same amount of smoothing for both the subject-specific and population level functions. This might seem a little restrictive in some sense, since the individual curves could be assumed to be more spatially heterogeneous than the population level curve. But for sparse

functional data (as in our case), the assumption of the same degree of smoothness at both the population and subject level is a reasonable one given the low number of observations per individual. Second, conditional on the choice of basis function and treating the basis matrix as fixed, the model in (3.2) is essentially a semiparameteric random effects model, with the prior on $\boldsymbol{\beta}_i$ admitting the within-subject covariance $V(\mathbf{Y}_i) = \mathbf{X}_i'\boldsymbol{\Omega}\mathbf{X}_i + \sigma_{\boldsymbol{\epsilon}}^2\mathbf{I}_{p_i}$. Hence, the within-subject independence assumption is relaxed to allow within-subject correlation for the observed curve $\mathbf{Y}_i$.

Having posited the above model on the functional (PSA) profiles, conditional on the basis matrix $\mathbf{X}$, we can proceed with estimation using a variety of Bayesian or frequentist techniques. However, two related issues remain. First, the number and position of the knots or breakpoints need to be chosen, and second, conditional on the number of knots, the dimension of $\boldsymbol{\Omega}$, if left unstructured, is of dimension $K \times K$; thus, we need to estimate $K(K+1)/2$ unique parameters. From a practical and methodological point of view, it is useful to reduce dimensionality. This is essentially a model selection problem. Various approaches to solving this problem include using model selection procedures such as conditional predictive ordinate (CPO) or deviance information criteria (DIC), as proposed by Brown et al. (2005), or a fully Bayesian framework using free-knot spline methodology (Denison et al. 1998; Holmes and Mallick 2003). For our application, it is of interest to model the exact location and number of change points in the PSA profiles since drastic changes in PSA might directly impact on the survival of the patient. This is also evident in Fig. 3.1, where one notices a sharp drop in PSA levels initially and then an increase in PSA levels in the later stages of the disease.

We handle the problem of choosing the number of change points in a Bayesian framework via latent indicators (Smith and Kohn 1996; Thompson and Rosen 2008). Essentially, we start with a large pool of potential breakpoints and an associated latent indicator vector, which we denote as $\boldsymbol{\gamma}$. The elements of the latent indicator vector equal 1 if the corresponding change point is included in the model and 0 otherwise—this implies keeping or deleting one basis function in (3.2). Thus, conditional on $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{K^*})$, where K^* is the number of the set of potential change points, our model in (3.2) can be written as

$$\mathbf{Y}_i = \mathbf{X}_{i,\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}} + \mathbf{X}_{i,\boldsymbol{\gamma}}\boldsymbol{\beta}_{i,\boldsymbol{\gamma}} + \boldsymbol{\epsilon}_i, \qquad \boldsymbol{\epsilon}_i \sim MVN(\mathbf{0}, \sigma_{\boldsymbol{\epsilon}}^2\mathbf{I}_{p_i}), \tag{3.3}$$

where each $\mathbf{X}_{i,\boldsymbol{\gamma}}$ is the basis matrix corresponding to change points for the ith individual, and $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ and $\boldsymbol{\beta}_{i,\boldsymbol{\gamma}}$ are the corresponding regressed coefficients of size $1 + p + K_{i,\boldsymbol{\gamma}}^*$, where p is the degree of the spline and $K_{i,\boldsymbol{\gamma}}^*$ is the number of ones in the vector $\boldsymbol{\gamma}$ and within the span of the ith individual curve. Conditional on the latent indicator parameter $\boldsymbol{\gamma}$ (and basis function), model (3.3) is still essentially a Bayesian linear model for which an attractive conjugate prior distribution for parameters exists for efficient Gibbs sampling. Since this model is only a component of our joint functional survival model, we defer our discussion of appropriate priors to Sect. 3.3, after we present our joint modeling framework. Note that we have not included any covariate affecting the functional process in our model above; this is easy to handle in our framework by adding a term corresponding to the covariate in the regression model (3.3).

3.2.2 Joint Survival Model

Having specified our functional submodel above, we now proceed to model the relationship between the functional measures $\mathbf{Y}$ and event time T. We do so by constructing the likelihood in a prospective manner, $P(T, \mathbf{Y}) = P(T|\mathbf{Y})P(\mathbf{Y})$, rather than a retrospective manner using reverse factorization by conditioning on the survival process. The probability model for $P(\mathbf{Y})$ is as specified in (3.3). In this section, we describe how we characterize the distribution $P(T|\mathbf{Y})$.

We model the failure time via a proportional hazards model. Following Cox (1972, 1975), and under the conditions discussed by Kalbfleisch and Prentice (2002), we use the original Cox model formulation, in which the hazard depends on the (true) functional process $\mathbf{Y}_i(t)$ through its current value (and/or other time-dependent covariates) and time-independent covariates $\mathbf{L}_i$. The framework for characterizing associations among the functional and survival processes, as well as other covariates, is then given by

$$\begin{aligned} h(t) &= \lim_{dt \to 0} P\{t < T_i < t + dt | T_i \geq t, \mathbf{Y}_i^H(t), \mathbf{L}_i\} \\ &= h_0(t) \exp\{\boldsymbol{\theta}_1 \mathbf{Y}_i(t) + \boldsymbol{\theta}_2 \mathbf{L}_i\}, \end{aligned}$$

where the coefficients $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ reflect the association of interest and $\mathbf{Y}_i^H(t) = \{\mathbf{Y}_i(u) : 0 < u < t\}$ is the history of the functional process $\mathbf{Y}_i$ up to time t. Note here that this implementation is complicated by two facts. First, the functional covariate is subject to measurement error and is observed only intermittently for each subject at $\mathbf{t}_i = (t_{i1}, \ldots, t_{ip_i})$. Second, and more important, plugging in the entire (smoothed) functional profile leads us to a high-dimensional integral in the likelihood:

$$\begin{aligned} f(T_i, \delta_i | \mathbf{Y}_i) = \{h_0(T_i) \exp[\boldsymbol{\theta}_1 \mathbf{Y}_i(T_i) + \boldsymbol{\theta}_2 \mathbf{L}_i]\}^{\delta_i} \times \\ \exp\left\{ - \int_0^{T_i} h_0(t) \exp[\boldsymbol{\theta}_1 \mathbf{Y}_i(t) + \boldsymbol{\theta}_2 \mathbf{L}_i] \mathrm{d}t \right\}. \end{aligned}$$

The high-dimensional integral in the likelihood does not have an analytical solution for the functional profile specified via a spline representation. Brown et al. (2005) use a trapezoidal rule to approximate the above integral. We present an exact Bayesian analysis via the use of auxiliary variables. To this effect, define a latent auxiliary variable w_i as

$$w_i = \boldsymbol{\beta}'_{\boldsymbol{\gamma}(i)} \boldsymbol{\theta}_{1\boldsymbol{\gamma}} + \mathbf{L}_i' \boldsymbol{\theta}_2 + e_i, \quad e_i \sim N(0, \tau^2), \tag{3.4}$$

where $\boldsymbol{\beta}_{\boldsymbol{\gamma}(i)} = \boldsymbol{\beta}_{\boldsymbol{\gamma}} + \boldsymbol{\beta}_{i,\boldsymbol{\gamma}}$ represents the ith PSA trajectory, $\boldsymbol{\theta}_{1\boldsymbol{\gamma}}, \boldsymbol{\theta}_2$ are the regression coefficient vectors corresponding to the time-dependent and time-independent covariate information, and e_i is an error term. The auxiliary variables summarize the time-dependent covariate effects via a simple projection. The functional submodel is then coupled with the survival model via these latent auxiliary variables by imputing the w_i into the proportional hazards model via the hazard function as:

$$h(t \mid \mathbf{Y}_i, \mathbf{L}_i) = h_0(t) \exp(w_i), \tag{3.5}$$

where $\mathbf{Y}_i$ is the ith individual time-dependent covariates vector, $\mathbf{L}_i$ is the time-independent covariates, and $h_0(t)$ is the baseline hazard function at time point t, free of the covariates.

The introduction of latent auxiliary variables not only eases the high-dimensional integration in the likelihood but also serves three purposes. The first concerns dimension reduction, wherein the information from the potentially high-dimensional regression coefficient $\boldsymbol{\beta}$ is passed along to the survival model via a simple projection into a lower dimensional subspace. Second, in adopting this Gaussian residual effect, many of the conditional distributions for the model parameters are now of a standard form, which greatly aids in the computations. To be specific, conditional on w_i's, model (3.3) is independent of the event time model (3.5) and can be written as a standard Bayesian linear regression on the basis space defined by $\mathbf{X}$, as we show in Sect. 3. The use of the residual component e_i is consistent with the belief that there may be unexplained sources of variation in the data, perhaps due to the lack of a linear relationship. Finally, the latent auxiliary variable formulation allows us to easily generalize our model to handle multiple functional covariates (Sect. 3.6).

We assume that the baseline hazard is a piecewise function as:

$$h_0(t) = \lambda_j \quad (s_{j-1} \leq t < s_j), \quad j = 1, \ldots, J. \tag{3.6}$$

In theory, increasing J approximates semiparametric methods. Other nonparametric priors (such as the gamma process and the beta process) can be easily incorporated within our framework. Based on (3.5) and (3.6), we write the cumulative hazard function for the ith individual as

$$\int_0^{T_i} h_0(t)\exp(w_i)\mathrm{d}t = \sum_{j=1}^{J} I(T_i > s_{j-1}) \int_{s_{j-1}}^{\min(s_j, T_i)} \exp(w_i)\lambda_j \mathrm{d}t,$$

where the indicator function $I(T_i > s_{j-1})$ yields 1 if the survival time is within or later than the jth interval and 0 otherwise.

3.3 Prior Distributions

The parameters and random variables to estimate in our model are

$$\mathcal{M} = \{\boldsymbol{\beta}, \sigma_{\epsilon}^2, \boldsymbol{\mu}_{\boldsymbol{\gamma}}, \boldsymbol{\Omega}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \boldsymbol{\theta}_{\boldsymbol{\gamma}}, \tau^2, \boldsymbol{\lambda}\}.$$

We shall discuss each of the priors and distributions for the regression and survival models, respectively.

3.3.1 Priors for The Regression Model

We assign a Gaussian prior distribution, $MVN(\mathbf{0}, \boldsymbol{\Omega}_{i,\boldsymbol{\gamma}})$, to the subject level regression coefficients $\boldsymbol{\beta}_{i,\boldsymbol{\gamma}}$. Based on the fact that the ith curve may not span the complete

set of selected change points, we use $\boldsymbol{\Omega}_{i,\boldsymbol{\gamma}}$ to denote the subject-specific realizations of parameter $\boldsymbol{\Omega}_{\boldsymbol{\gamma}}$ that respectively represent the population curve covariance corresponding to the latent variable $\boldsymbol{\gamma}$. In the implementation of our methodology for our particular example, the number of basis functions K is relatively small. At least in principle, we can then allow the covariance matrices $\boldsymbol{\Omega}_{\boldsymbol{\gamma}}$ to be general. However, from both a practical and methodological point of view, it is crucial to lower the dimensionality of $\boldsymbol{\Omega}_{\boldsymbol{\gamma}}$. There are a variety of approaches available to this end. For example, Shi et al. (1996) achieve parsimony using a principal component decomposition of the covariance matrix of random effects. In a different context, Daniels and Pourahmadi (2002) provide a Bayesian method based on Cholesky decomposition. Since in our application we work with truncated power series basis functions, dimension reduction has a natural form that exploits the mixed model representation of such basis functions (Ruppert et al. 2003; Baladandayuthapani et al. 2008). The essential idea is to take the coefficients at the knots to be independent while allowing the polynomial part to have an unstructured covariance matrix. Thus, if p is the degree of the regression splines and $K^*_{\boldsymbol{\gamma}}$ is the number of selected knots, then we take $\boldsymbol{\Omega}_{\boldsymbol{\gamma}} = \text{diag}(\boldsymbol{\Sigma}, \sigma^2 \mathbf{I}_{K^*_{\boldsymbol{\gamma}}})$, where $\boldsymbol{\Sigma}$ is an unstructured $p \times p$ matrix. Further, we specify a Gaussian prior distribution on the population level profile or the fixed effects, $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$, as $\boldsymbol{\beta}_{\boldsymbol{\gamma}} \sim MVN(\mathbf{0}, \mathbf{c}\mathbf{I}_{\mathbf{K}_{\boldsymbol{\gamma}}})$, where we set c to be 100, and $K_{\boldsymbol{\gamma}} = 1 + p + K^*_{\boldsymbol{\gamma}}$. We adopt an Inverse–Wishart prior distribution for $\boldsymbol{\Sigma}$ and an inverse-gamma prior distribution for σ^2. For the regression model (3.3), we assume an inverse-gamma prior distribution for the constant variance $\sigma^2_{\boldsymbol{\epsilon}}$.

The selected change points are identified by the vector $\boldsymbol{\gamma}$. We use a Bernoulli prior for each element of this indicator vector, $\gamma_k \sim \text{Bernoulli}(\pi_k)$, and let $\pi_k = \pi$ for all k. The hyperprior for the probability of being a change point is specified as a beta prior, $\pi \sim \text{Beta}(a_\pi, b_\pi)$. Kohn et al. (2001) pointed out a flexible approach specifying beta prior hyperparameters according to a certain expectation or prior knowledge.

3.3.2 *Priors for The Survival Model*

We use conjugate prior distributions for the parameter pair $\boldsymbol{\theta}_{\boldsymbol{\gamma}}$ and τ^2, defined as $\boldsymbol{\theta}_{\boldsymbol{\gamma}} \sim MVN(\mathbf{0}, \tau^2 \mathbf{V}_{\boldsymbol{\gamma}})$ and $\tau^2 \sim IG(a_\tau, b_\tau)$, where $\mathbf{V}_{\boldsymbol{\gamma}} = diag(\mathbf{h})$. The hyperprior for the vector $\mathbf{h} = \{h_\ell\}$ is specified elementwise as inverse-gamma distribution, $h_\ell \sim IG(c_\ell, d_\ell)$. For the survival model, the prior distribution for a piecewise baseline hazard functions, $\boldsymbol{\lambda} = \{\lambda_j\}$ is $\lambda_j \sim IG(a_j, b_j)$, where a_j and b_j can be specified for each interval.

We proceed with the estimation of the above model setup via Markov chain Monte Carlo (MCMC) methods. The full conditional distributions are presented for the regression and survival models in the Appendix. We use the Gibbs sampler (Gelfand and Smith 1990) to obtain samples from the posterior distribution. Two parameters, w_i and γ_k, do not have close forms in their conditionals. Therefore, we use Metropolis–Hastings algorithm (Metropolis et al. 1953; Hastings 1970) to sample those two parameters.

3.4 Model Selection Criteria

For model selection and comparison, we use two comparison statistics: the DIC and the CPO (Gelfand et al. 1992). The DIC is the sum of the deviance estimated using posterior estimates of the parameters and twice the effective number of parameters (Spiegelhalter et al. 2002). A better fit will have a smaller DIC. The DIC for our joint models can be expressed as

$$DIC = 2\frac{1}{Q}\sum_{q=1}^{Q}\sum_{i=1}^{n} log\, f(T_i, \delta_i, \mathbf{Y}_i \mid \Theta^{(q)}) - \sum_{i=1}^{n} log\, f(T_i, \delta_i, \mathbf{Y}_i \mid \bar{\Theta}),$$

where $\Theta^{(q)}$ denotes the parameter samples at the qth iteration of the MCMC method and $\bar{\Theta}$ represents the means of the posterior samples. Chen et al. (2000) showed that a Monte Carlo approximation of the integral in the CPO calculation can be used. For our joint models, we have

$$\widehat{\text{CPO}_i} = \left(\frac{1}{Q}\sum_{q=1}^{Q}\frac{1}{f(T_i, \delta_i, \mathbf{Y}_i \mid \Theta^{(q)})}\right)^{-1}.$$

Models with greater $\sum_{i=1}^{n} log(\widehat{\text{CPO}_i})$'s indicate a better fit.

Computing DIC and $\sum_{i=1}^{n} log(\widehat{CPO_i})$ is straightforward based on the samples from the MCMC method and the joint likelihood function:

$$\begin{aligned} f(T_i, \delta_i, \mathbf{Y}_i) &= f(T_i, \delta_i \mid \mathbf{Y}_i) f(\mathbf{Y}_i) \\ &\propto \left[h_0(T_i)\exp(w_i)\right]^{\delta_i} \exp\left\{-\int_0^{T_i} h_0(u)\exp(w_i)\mathrm{d}u\right\} |2\pi\sigma_\epsilon^2\mathbf{I}|^{-\frac{1}{2}} \\ &\times \exp\left\{-\frac{1}{2\sigma_\epsilon^2}(\mathbf{Y}_i - \mathbf{X}_{i,\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}} - \mathbf{X}_{i,\boldsymbol{\gamma}}\boldsymbol{\beta}_{i,\boldsymbol{\gamma}})'(\mathbf{Y}_i - \mathbf{X}_{i,\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}} - \mathbf{X}_{i,\boldsymbol{\gamma}}\boldsymbol{\beta}_{i,\boldsymbol{\gamma}})\right\}. \end{aligned}$$

3.5 Application to Prostate Cancer Data

We now consider a data set from a phase III trial of prostate cancer patients conducted at M.D. Anderson Cancer Center (Millikan et al. 2008). The clinical trial studied 286 patients with metastatic or locally advanced prostate cancer who were randomized and treated with either AA alone (arm AA) or chemo/hormonal therapy plus AA (arm CH) between August 1996 and March 2003. A complete medical history was obtained from each patient. All patients also underwent a physical examination. For each patient, we have a record of the time (in days) between the trial starting day and progression to AIPC or end of study, an indicator of censoring, which treatment the patient received, day of each visit measured from registration, and PSA level

measured on that day. The functional laboratory results of PSA, the leading diagnostic marker for prostate cancer, is considered a predictor variable in our application. The failure time variable, the time to progression of AIPC, is a right-censored variable. Four time-independent covariates are also considered in the analysis. Their age at diagnosis, prior local treatment, stratification via bone volume, and pretreatment PSA doubling time (PSADT). For prior local treatment, patients either did or did not receive definitive treatment. Patients were also stratified as follows: high-volume bone or visceral disease, low-volume bone disease (one or two spots on bone scan), local/nodal disease with prior definitive local therapy, or local/nodal disease without prior definitive therapy. For simplicity, the three low-volume groups of patients were combined into one category yielding two categories: high-volume disease or low-volume disease. Since we have (intermittent) PSA measurements from the patients before therapy, we include pretreatment PSADT as a time-independent covariate in our survival model. It is a categorical variable stratified, as 0 if data are not available to determine a doubling time, as 1 if doubling time is less than 3 months, and as 2 if doubling time is greater than 3 months.

The number of PSA observations for each patient varied from 1 to 65. We use the data set after a screening procedure removes those patients with fewer than four observations. We transform the PSA levels into a log scale after adding 1. This transformation is usually done so that residuals satisfy the assumption of homoscedasticity and also to reduce the influence of outliers. We also transform the time axis, via a one-to-one function, onto a log scale after dividing by 30 (the change from day to month) and adding 1. Figure 3.1 depicts the overlapping PSA levels for 134 patients in arm AA (upper panel) and 132 patients in arm CH (lower panel). The sparsity of the profiles is suggested by the percentage of patients who have measurements at or spanning the particular time point. More than 50 % of patients do not have measurements before day 18 (equal to 0.47 in the unit of log(month+1)) and after day 1210 (equal to 3.72 in the unit of log(month+1)).

We use a quadratic truncated power series basis function (Ruppert et al. 2003) to model the subject and population PSA profiles. To construct the candidate pools of change points we use 11 equally spaced knots for each arm, since it suffices for this application. For the baseline hazard step function in the proportional hazards model, we include ten-step intervals starting from day 0 to the last day. For the proposed unified Bayesian model, we wish to impose proper but weak prior information. For inverse-gamma priors, we let the shape hyperparameter to be larger than 1, allowing existence of the expectation of the inverse-gamma distribution. For Inverse–Wishart priors, we choose to use the degrees of freedom that are the smallest integers such that the expectation of the distribution exists. The scale matrix is specified as the identity matrix. We employ the following hyperparameter settings: (a_σ, b_σ), (c_σ, d_σ), (a_τ, b_τ), (c_ℓ, d_ℓ), and (a_j, b_j) are specified as (2, 2), and $(\mathbf{A}, b)$ is specified as identity matrices and 4. The hyperparameter c is specified as 100 to produce a non-informative prior for $\boldsymbol{\beta}_\gamma$. We found that the results are insensitive to moderate modifications of these priors. For the hyperparameter pair (a_π, b_π), we use the method by Kohn et al.

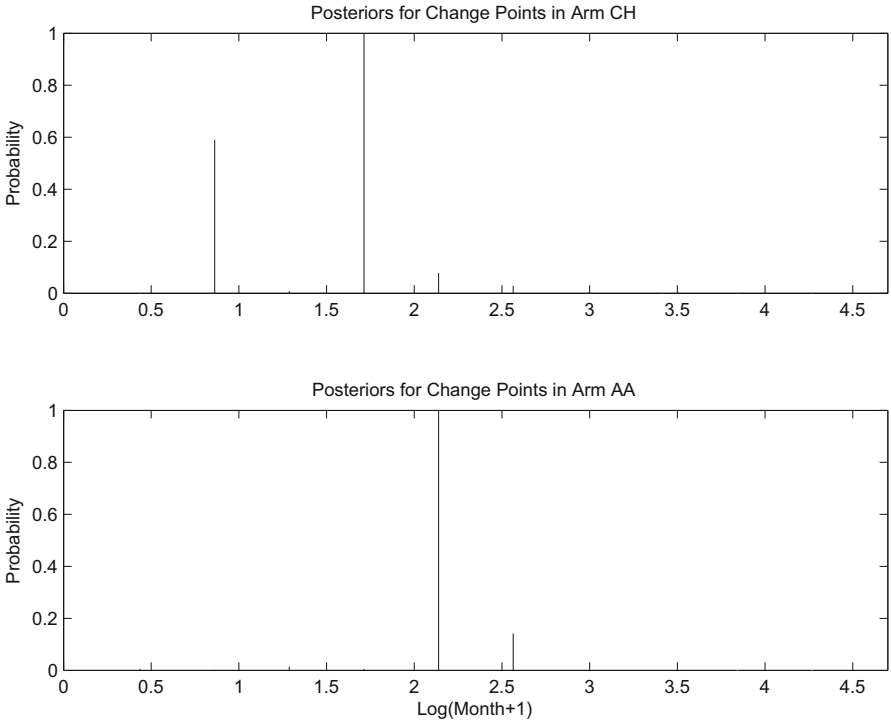

Fig. 3.2 The posterior probabilities of change points for arm chemohormonal (*CH*; *top panel*) and arm androgen ablation (*AA*; *bottom panel*). The vertical axes are the posterior probabilities and the horizontal axes are the location of the change points

(2001) to calculate the priors $a_\pi = 1.077$ and $b_\pi = 4.846$, with $E(K_\gamma^*) = 2$ and $\text{std}(K_\gamma^*) = 2$ so that the number of selected knots, K_γ^*, is likely to range from 0 to 8. We run the MCMC chain for 60,000 iterations with 20,000 burn-in iterations. To verify the stability of the algorithm, we run several different chains with various starting knot vectors; the results show that the change point identification is quite stable. Figure 3.2 shows the posterior probabilities of 11 equally spaced change points for the treatment arms CH and AA.

Our results suggest that arm CH has two change points located at 0.86 (day 41) and 1.71 (day 136), while arm AA has one change point that is located at 2.14 (day 225) with posterior probabilities very close to 1. Thus, our model seems to correctly identify the change points of PSA trends for both arms, as suggested by Fig. 3.1. The PSA levels usually decrease sharply due to the effect of the therapy, since the therapy directly affects the prostate gland; but over time their effect wears off and the PSA levels remain constant before increasing and causes prostate cancer.

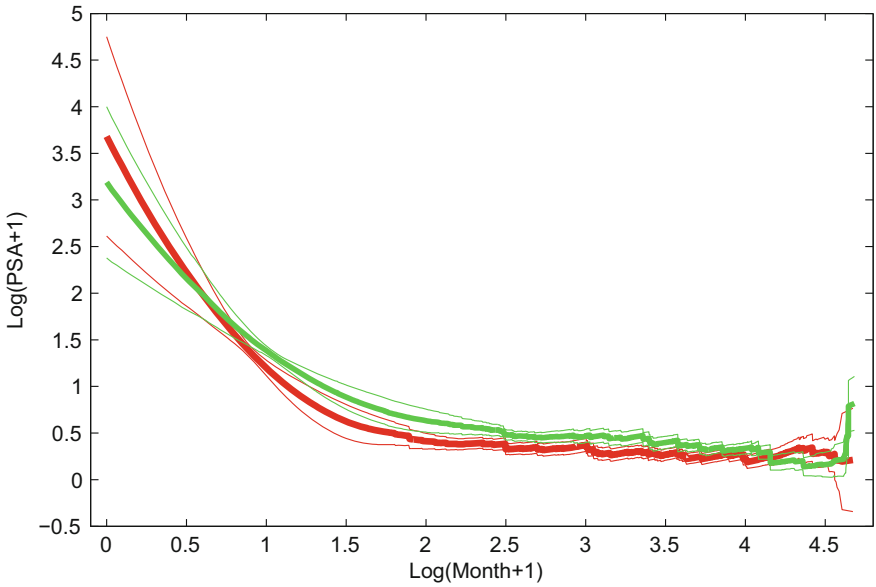

Fig. 3.3 Based on our proposed unified Bayesian model, the estimated trends of prostrate-specific antigen (*PSA*) levels for arm chemohormonal (CH; *red* line) and arm androgen ablation (AA; *green* line) with their 95 % credible intervals

Since it has been established that the volume of the prostate has a significant positive correlation with the level of PSA found by a blood test, we want to estimate the true trends of PSA levels over time for both arms. Figure 3.3 gives the estimated population-level PSA trajectory for both the arms along with the 95 % credible interval obtained using our proposed joint model, showing an L-shaped pattern.

The population profiles intersect for the most part except between time units 1–2. The difference in change points can explain the slight separation of the trends in the two arms, as depicted by Fig. 3.3. Because an increasing PSA level usually indicates prostate malfunction, we see the patients in arm AA deteriorating a bit at the end of the time period as compared to arm CH. We also see evidence that the drop in PSA levels is higher for arm CH than for arm AA. However, near the end of the study (with $\log(\text{month}+1) \geq 4.5$), the PSA difference between the two arms needs careful interpretation. This is because less than 10 % of patients have PSA observations at or spanning this period and some of those patients have extremely high PSA levels that may impose a larger influence on the estimation. Figure 3.4 shows the estimated individual PSA trajectories with 95 % posterior credible intervals for four randomly selected patients treated with CH.

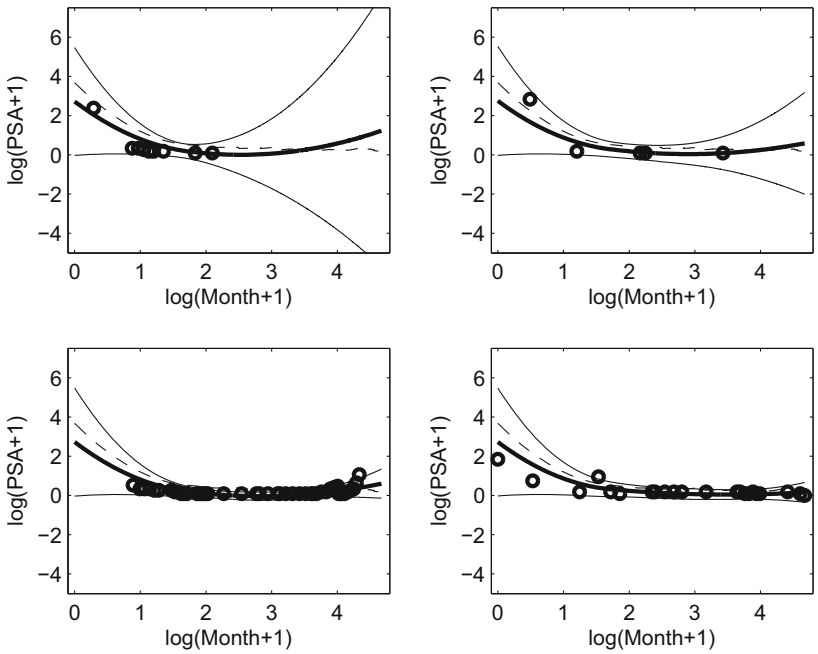

Fig. 3.4 Estimated individual prostrate-specific antigen (*PSA*) trajectories with 95 % posterior credible intervals for four patients treated in arm chemohormonal (CH). *Circles* are actual PSA measurements, *dashed* lines indicate the overall mean trend for arm CH

The figure reveals how we can borrow strength across subjects through our Bayesian model to estimate the PSA trajectories when there is little or no information. It is not surprising that the parts of the trajectories with little or no data have wider pointwise intervals.

For prostate cancer data, the effectiveness of treatment on time to AIPC is of interest. We apply our model to each of the two arms and compare the estimated time-to-event survival curves. The upper panel of Fig. 3.5 shows two superimposed survival curves based on our model and the Kaplan–Meier method with 95 % credible intervals for the two arms.

The lower panel depicts two superimposed cumulative hazard curves for the two arms based on our model, with 95 % credible intervals. The close approximation of estimated survival curves to the Kaplan–Meier curves indicates a fair fit of the model to the observed data. There is no apparent improvement for those in arm CH, but there is some evidence that arm CH may perform marginally better than arm AA because the estimated survival curve for arm AA is a little lower than the one for arm CH. On the other hand, the estimated time-to-AIPC expectancy is 1249 days for arm AA and 1527 days for arm CH. The 95 % credible bands are (881, 1768) and (1157, 2011) for arms AA and CH, respectively. The overlapping of the two credible bands

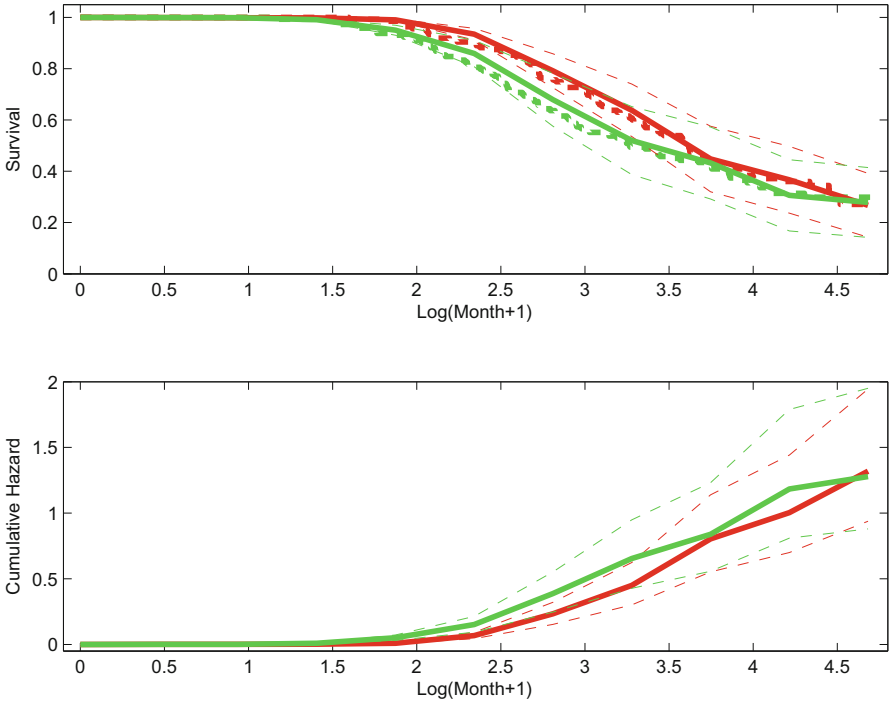

Fig. 3.5 Upper panel: Survival curves for arms androgen ablation (AA; *green* lines) and chemo-hormonal (CH; *red* lines), Kaplan–Meier curve (*thick dotted* line), our estimated survival curves (*solid* lines), and their 95 % credible intervals (*thin dashed* lines). Lower panel: Cumulative hazard curves for two arms and their 95 % credible intervals

means that there is no significant difference in time-to-AIPC expectancy between the two arms. The hazard curves exhibit a similar pattern.

In our model setup, the auxiliary variable w serves as a bridge parameter between the functional regression model and the survival model and captures the relationship between the functional predictor and the survival time. Figure 3.6 shows the box plots of the estimated w's and the observed PSA levels with high or low w's. The top two plots are for arm AA.

The top left plot overlaps ten observed PSA levels (*dotted* lines) that are for patients with the highest estimated w's, and ten other levels (*solid* lines) that are for patients with the lowest estimated w's. The mean observed survival time for the patients with the highest and lowest w's are 1.95 and 4.46, respectively. The separation of PSA levels for two groups of patients shows that long-survived patients have PSA levels that drop to very low levels and remain low after treatment, while short-survived patients have PSA levels that drop slightly yet bounce back quickly. Therefore, we conclude that the mostly nonzero w's reveal the validity of our joint model for these data based on the fact that the functional PSA levels have a (negative) effect on the progress to AIPC. The effect of informative scalar w is further illustrated by the top

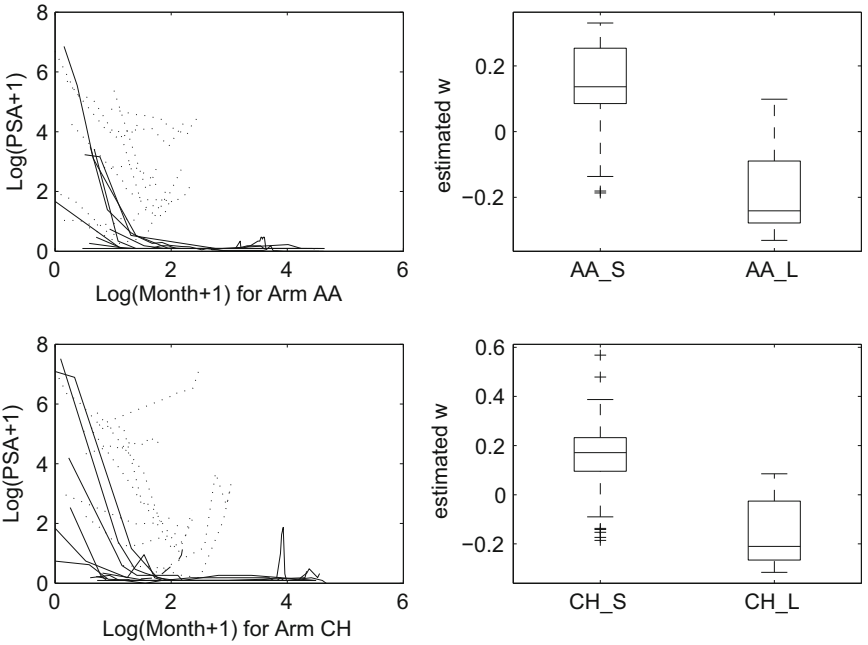

Fig. 3.6 Left: observed prostrate-specific antigen (*PSA*) trajectories for 20 patients. Ten patients with the highest estimated w's are plotted in *dotted* lines, and ten patients with the lowest estimated w's are plotted in *solid* lines. Right: box plots for posterior means of scalar w. For both arms, "_S" means patient's time to progression of androgen-independent prostate cancer (AIPC) is short than or equal to 32 months, and "_L" means patient's time to progression of AIPC is longer than 32 months

Table 3.1 The estimation of coefficients for time-independent covariates. The values in parentheses are the estimated standard deviations

	Age at *R*x	Definitive	Stratification	PSADT
Arm AA	−0.020(0.038)	−0.628(2.151)	0.595(1.984)	−0.569(1.308)
Arm CH	−0.056(0.033)	−1.397(1.733)	0.455(1.740)	−0.636(1.015)

PSADT prostrate-specific antigen doubling time, AA androgen ablation, CH chemohormonal

right plot in Fig. 3.6, where two box plots are stratified by long-term and short-term survivors according to the threshold of 32 months. We see that the w's are negatively associated with survival time. The bottom two plots in Fig. 3.6 are for arm CH, and the findings on w are the same as those for arm AA.

Table 3.1 gives the estimates of coefficients, which are the last four elements of vector $\boldsymbol{\theta}$, corresponding to the four time-independent covariates. The estimation shows that, for both arms, elder age at diagnosis, definitive treatment, low volume of stratification, and pretreatment PSADT longer than 3 months lead to lower hazards. However, there is only one significant covariate for arm CH: age at diagnosis.

Table 3.2 The model comparison measurements for both arms

	Model	DIC	$\sum_{i=1}^{n} \log(\widehat{\text{CPO}_i})$
Arm AA	Change point model	2362.7	−2652.7
	Quadratic model	2402.4	−2706.0
	Linear model	2616.3	−3033.6
Arm CH	Change point model	2129.5	−2405.5
	Quadratic model	2326.3	−2490.3
	Linear model	2859.9	−2951.0

AA androgen ablation, CH chemohormonal, DIC deviance information criteria

For comparison, we also consider two other models without the change points. The model setups are similar to that in Sect. 3.2, except that any term with the indicator vector $\boldsymbol{\gamma}$ is dropped. One model uses the linear basis:

$$\mathbf{X}_i^{Linear}(\mathbf{t}) = \begin{bmatrix} 1 & t_{i1} \\ \vdots & \vdots \\ 1 & t_{ip_i} \end{bmatrix},$$

and the other uses the quadratic basis

$$\mathbf{X}_i^{Quad}(\mathbf{t}) = \begin{bmatrix} 1 & t_{i1} & t_{i1}^2 \\ \vdots & \vdots & \vdots \\ 1 & t_{ip_i} & t_{ip_i}^2 \end{bmatrix}.$$

The model comparison measures DIC and $\sum_{i=1}^{n} \log(\widehat{\text{CPO}_i})$ are reported in Table 3.2 for both arms.

The change point model is a joint model using the change points selected by our proposed model (as shown in Fig. 3.2), two change points included for arm CH, and one change point included for arm AA. For both arms, the joint change point model is the best fit to the data with the largest CPO and smallest DIC. This result is consistent with the outcome in Fig. 3.2, where our flexible change point selection model identifies those significant change points.

3.6 Discussion

Motivated by the analysis of the data from a prostate cancer phase III clinical trial data, we present a joint modeling approach for functional and survival data using a nonparametric regression model and a proportional hazards model. Further, we allow random change points in the functional observations, both in terms of locations and number, to capture the important curvatures of the trajectory. This unified framework

combines the information from both functional predictors and time to progression to generate reliable results for regression and survival analysis. Moreover, a novel auxiliary variable scheme for a fully Bayesian estimation of our model is proposed. This novel scheme reduces the dimension of the parameter space, and greatly eases the computations in Bayesian estimation. Our results indicate that this scheme aids in the understanding or interpretation of the linkage between the functional predictor and time to progression.

Our model can also benefit from several refinements and extensions. We propose to model the survival end point via Cox's proportional hazards model, mainly due to its ease of implementation and interpretability. Other survival models, such as accelerated failure time models and cure rate models, can easily be accommodated in our framework. In some situations one may want to consider the effect of time-independent covariates, such as age at diagnosis, on the progress of the disease. In the joint model, allowing interaction between $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ could address such concerns. Further, one may observe multiple functional predictors and may want to assess their impacts on survival. Suppose that for the ith individual we observe the κth functional covariate $\mathbf{Y}_{i\kappa}$, the basis matrix can be denoted by $\mathbf{X}_{i,\boldsymbol{\gamma}_\kappa}$, and the fixed and random regressed coefficients can be denoted by $\boldsymbol{\beta}_{\boldsymbol{\gamma}_\kappa}$ and $\boldsymbol{\beta}_{i,\boldsymbol{\gamma}_\kappa}$. Then one can express the regression model as

$$\mathbf{Y}_{i\kappa} = \mathbf{X}_{i,\boldsymbol{\gamma}_\kappa}\boldsymbol{\beta}_{\boldsymbol{\gamma}_\kappa} + \mathbf{X}_{i,\boldsymbol{\gamma}_\kappa}\boldsymbol{\beta}_{i,\boldsymbol{\gamma}_\kappa} + \boldsymbol{\epsilon}_{i\kappa}, \qquad \boldsymbol{\epsilon}_{i\kappa} \sim MVN(\mathbf{0}, \sigma_\kappa^2 \mathbf{I}_{p_{i\kappa}}).$$

The information from multiple functional predictors can be easily absorbed into the survival segment via our novel proposed linear model for the auxiliary scalar w_i,

$$w_i = \sum_{\kappa=1}^{K} \mathbf{B}_{i,\boldsymbol{\gamma}_\kappa}\boldsymbol{\theta}_{\boldsymbol{\gamma}_\kappa} + e_i, \qquad e_i \sim N(0, \tau^2),$$

where $\mathbf{B}'_{i,\boldsymbol{\gamma}_\kappa} = [(\boldsymbol{\beta}_{\boldsymbol{\gamma}_\kappa} + \boldsymbol{\beta}_{i,\boldsymbol{\gamma}_\kappa})', \mathbf{L}'_i]$. The rest of the model setup, including prior and posterior distributions, are analogous to the univariate case. Therefore, we conclude that the auxiliary scalar scheme is not only enabling feasible computing in the joint modeling framework but also exhibiting the potential for generalization to a more complex model.

Acknowledgments

We thank Dr. Randall Millikan from the University of Texas M.D. Anderson Cancer Center for supplying the prostate cancer data set as an application example of our proposed methodology. V. Baladandayuthapani was partially supported by NIH grant R01 CA160736. Both K-A Do and V. Baladandayuthapani were partially supported by the Cancer Center Support Grant (CCSG) (P30 CA016672). Dr. K-A Do was also partially supported by the University of Texas SPORE in prostate cancer National Institute of Health grant (P50 CA140388). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

3.7 Appendix

3.7.1 The Model Summary with Specified Prior Distributions

To summarize the hierarchical model setup, we define

$$
\begin{aligned}
\text{Random function} \quad \mathbf{Y}_i &\sim MVN(\mathbf{X}_{i,\boldsymbol{\gamma}}\boldsymbol{\beta}_{i,\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\epsilon}}^2\mathbf{I}_{p_i}),\\
\sigma_{\boldsymbol{\epsilon}}^2 &\sim IG(a_\sigma, b_\sigma),\\
\boldsymbol{\beta}_{i,\boldsymbol{\gamma}} &\sim MVN(\boldsymbol{\beta}_{\boldsymbol{\gamma}_i}, \boldsymbol{\Omega}_{i,\boldsymbol{\gamma}}),\\
\boldsymbol{\beta}_{\boldsymbol{\gamma}_i} &= \mathbf{J}_i\boldsymbol{\beta}_{\boldsymbol{\gamma}},\\
\boldsymbol{\Omega}_{i,\boldsymbol{\gamma}} &= \mathbf{J}_i\boldsymbol{\Omega}_{\boldsymbol{\gamma}}\mathbf{J}_i' \ \text{ where } \boldsymbol{\Omega}_{\boldsymbol{\gamma}} = \text{diag}(\boldsymbol{\Sigma}, \sigma^2\mathbf{I}_{K_{\boldsymbol{\gamma}}^*}),\\
\boldsymbol{\beta}_{\boldsymbol{\gamma}} &\sim MVN(0, c\mathbf{I}_{K_{\boldsymbol{\gamma}}}),\\
\boldsymbol{\Sigma} &\sim IW(\mathbf{A}, b),\\
\sigma^2 &\sim IG(c_\sigma, d_\sigma),\\
\gamma_k &\sim \text{Bernoulli}(\pi_k), \ \text{ where } \pi_k = \pi \text{ for all } k,\\
\pi &\sim \text{Beta}(a_\pi, b_\pi),\\
\text{Linear predictor} \quad w_i &\sim N(\mathbf{B}_{i,\boldsymbol{\gamma}}'\boldsymbol{\theta}_{\boldsymbol{\gamma}}, \tau^2), \ \text{ where } \mathbf{B}_{i,\boldsymbol{\gamma}}' = [\boldsymbol{\beta}_{i,\boldsymbol{\gamma}}', \mathbf{L}_i'],\\
\boldsymbol{\theta}_{\boldsymbol{\gamma}}, \tau^2|\boldsymbol{\gamma}, \mathbf{V}_{\boldsymbol{\gamma}} &\sim NIG(0, \mathbf{V}_{\boldsymbol{\gamma}}, a_\tau, b_\tau), \ \text{ where } \mathbf{V}_{\boldsymbol{\gamma}} = \text{diag}(\mathbf{h}),\\
h_\ell &\sim IG(c_\ell, d_\ell),\\
\text{Hazard function} \quad h(t \mid \mathbf{Y}_i) &= h_0(t)\exp(w_i),\\
h_0(t) &= \lambda_j \ \ (s_{j-1} \le t < s_j),\\
\lambda_j &\sim IG(a_j, b_j),
\end{aligned}
$$

for $i = 1, \ldots, n$, $j = 1, \ldots, J$, $k = 1, \ldots, K$, and $\ell = 1, \ldots, (K_\gamma + m)$.

The fourth and fifth lines in the above model need special attention. Based on the fact that the ith curve may not span the complete set of selected change points, $\boldsymbol{\beta}_{\boldsymbol{\gamma}_i}$ and $\boldsymbol{\Omega}_{i,\boldsymbol{\gamma}}$ are the subject-specific realizations of parameters $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ and $\boldsymbol{\Omega}_{\boldsymbol{\gamma}}$, where they respectively represent the population curve and its covariance corresponding to the latent variable $\boldsymbol{\gamma}$. The relationship can be expressed via a rectangular indicator matrix $\mathbf{J}_i$ as $\boldsymbol{\beta}_{\boldsymbol{\gamma}_i} = \mathbf{J}_i\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ and $\boldsymbol{\Omega}_{i,\boldsymbol{\gamma}} = \mathbf{J}_i\boldsymbol{\Omega}_{\boldsymbol{\gamma}}\mathbf{J}_i'$ with $\boldsymbol{\Omega}_{\boldsymbol{\gamma}} = \text{diag}(\boldsymbol{\Sigma}, \sigma^2\mathbf{I}_{K_{\boldsymbol{\gamma}}^*})$. For example, suppose there are five change points for the population curve, and the ith individual only spans the first two change points (i.e., does not have measurements beyond the third change point). Because the basis has the quadratic polynomial segment and the change points segment, the dimensions of $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ and $\boldsymbol{\Omega}_{\boldsymbol{\gamma}}$ will be 8 and 8 by 8. However, for the ith individual, the dimensions of $\boldsymbol{\beta}_{\boldsymbol{\gamma}_i}$ and $\boldsymbol{\Omega}_{i,\boldsymbol{\gamma}}$ are 5 and 5 by 5. Therefore,

$\boldsymbol{\beta}_{\boldsymbol{\gamma}_i}$ is linked to $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ via a 5 by 8 rectangular index matrix:

$$\mathbf{J}_i = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}. \tag{3.7}$$

The same $\mathbf{J}_i$ is used to link $\boldsymbol{\Omega}_{\boldsymbol{\gamma}}$ and $\boldsymbol{\Omega}_{i,\boldsymbol{\gamma}}$. These expressions with $\mathbf{J}_i$ enable the derivation of the posterior distributions below.

3.7.2 Posterior Distributions

The conditional distribution for the ith regressed covariates vector $\boldsymbol{\beta}_{i,\boldsymbol{\gamma}}$ is updated using regression likelihood

$$\boldsymbol{\beta}_{i,\boldsymbol{\gamma}} \mid \mathbf{X}_{i,\boldsymbol{\gamma}}, \mathbf{Y}_i, \sigma_{\boldsymbol{\epsilon}}^2, \boldsymbol{\Omega}_{\boldsymbol{\gamma}}, w_i, \tau^2, \boldsymbol{\theta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma} \sim MVN(\boldsymbol{\beta}_{i,\boldsymbol{\gamma}}^*, \tau^2 \boldsymbol{\Omega}_{i,\boldsymbol{\gamma}}^*),$$

where $\boldsymbol{\Omega}_{i,\boldsymbol{\gamma}}^* = (\tau^2(\boldsymbol{\Omega}_{i,\boldsymbol{\gamma}}^{-1} + \mathbf{X}_{i,\boldsymbol{\gamma}}'\mathbf{X}_{i,\boldsymbol{\gamma}}/\sigma_{\boldsymbol{\epsilon}}^2) + \boldsymbol{\theta}_{1i,\boldsymbol{\gamma}}\boldsymbol{\theta}_{1i,\boldsymbol{\gamma}}')^{-1}$ and $\boldsymbol{\beta}_{i,\boldsymbol{\gamma}}^* = \boldsymbol{\Omega}_{i,\boldsymbol{\gamma}}^* \times (\tau^2(\mathbf{X}_{i,\boldsymbol{\gamma}}'\mathbf{Y}_i/\sigma_{\boldsymbol{\epsilon}}^2 + \boldsymbol{\Omega}_{i,\boldsymbol{\gamma}}^{-1}\boldsymbol{\mu}_{i,\boldsymbol{\gamma}}) + (w_i - \mathbf{L}_i'\boldsymbol{\theta}_2)\boldsymbol{\theta}_{1i,\boldsymbol{\gamma}})$. The notation $\boldsymbol{\theta}_{1i,\boldsymbol{\gamma}}$ is the part of the coefficients corresponding to time-dependent covariates $\boldsymbol{\beta}_{i,\boldsymbol{\gamma}}$, while $\boldsymbol{\theta}_2$ is the part of the coefficients corresponding to time-independent covariates $\mathbf{L}_i$ in later posteriors. The model variance $\sigma_{\boldsymbol{\epsilon}}^2$ is updated by

$$\sigma_{\boldsymbol{\epsilon}}^2 | \boldsymbol{\beta}_{i,\boldsymbol{\gamma}}, \mathbf{Y}_i, \mathbf{X}_{i,\boldsymbol{\gamma}} \sim IG(a_\sigma^*, b_\sigma^*),$$

where $a_\sigma^* = a_\sigma + (\sum_{i=1}^n p_i)/2$ and $b_\sigma^* = b_\sigma + [\sum_{i=1}^n (\mathbf{Y}_i - \mathbf{X}_{i,\boldsymbol{\gamma}}\boldsymbol{\beta}_{i,\boldsymbol{\gamma}})'(\mathbf{Y}_i - \mathbf{X}_{i,\boldsymbol{\gamma}}\boldsymbol{\beta}_{i,\boldsymbol{\gamma}})]/2$. The indicator vector $\boldsymbol{\gamma}$ can be updated elementwise using the Metropolis–Hastings algorithm with marginal posterior $\gamma_k \mid \boldsymbol{\gamma}_{-k}, \mathbf{Y}_i, \mathbf{X}_{i,\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\epsilon}}^2, \boldsymbol{\Sigma}, \sigma^2, \boldsymbol{\theta}_{\boldsymbol{\gamma}}, \mathbf{V}_{\boldsymbol{\gamma}}$ proportional to

$$\pi(\gamma_k)\pi(\boldsymbol{\theta}_{\boldsymbol{\gamma}})\pi(\mathbf{V}_{\boldsymbol{\gamma}}) \left[\frac{|\boldsymbol{\Phi}_{\boldsymbol{\gamma}}^{-1}|}{|c\mathbf{I}_{\boldsymbol{\gamma}}|} \prod_{i=1}^n \frac{|\tau^2 \mathbf{M}_{i,\boldsymbol{\gamma}}^{-1}|}{|\boldsymbol{\Omega}_{i,\boldsymbol{\gamma}}|} \right]^{1/2} \exp\left\{ \frac{1}{2\tau^2} \sum_{i=1}^n \boldsymbol{\alpha}_{i,\boldsymbol{\gamma}}' \mathbf{M}_{i,\boldsymbol{\gamma}}^{-1} \boldsymbol{\alpha}_{i,\boldsymbol{\gamma}} \right\}$$
$$\times \exp\left\{ \frac{1}{2} \left(\sum_{i=1}^n \boldsymbol{\alpha}_{i,\boldsymbol{\gamma}}' \mathbf{M}_{i,\boldsymbol{\gamma}}^{-1} \boldsymbol{\Omega}_{i,\boldsymbol{\gamma}}^{-1} \mathbf{J}_i \right) \boldsymbol{\Phi}_{\boldsymbol{\gamma}}^{-1} \left(\sum_{i=1}^n \mathbf{J}_i \boldsymbol{\Omega}_{i,\boldsymbol{\gamma}}^{-1} \mathbf{M}_{i,\boldsymbol{\gamma}}^{-1} \boldsymbol{\alpha}_{i,\boldsymbol{\gamma}} \right) \right\},$$

where $\boldsymbol{\alpha}_{i,\boldsymbol{\gamma}} = \tau^2 \mathbf{X}_{i,\boldsymbol{\gamma}}' \mathbf{Y}_i/\sigma_{\boldsymbol{\epsilon}}^2 + (w_i - \mathbf{L}_i'\boldsymbol{\theta}_2)\boldsymbol{\theta}_{1i,\boldsymbol{\gamma}}$, $\mathbf{M}_{i,\boldsymbol{\gamma}} = \tau^2(\mathbf{X}_{i,\boldsymbol{\gamma}}'\mathbf{X}_{i,\boldsymbol{\gamma}}/\sigma_{\boldsymbol{\epsilon}}^2 + \boldsymbol{\Omega}_{i,\boldsymbol{\gamma}}^{-1}) + \boldsymbol{\theta}_{1i,\boldsymbol{\gamma}}\boldsymbol{\theta}_{1i,\boldsymbol{\gamma}}'$ and $\boldsymbol{\Phi}_{\boldsymbol{\gamma}} = (\sum_{i=1}^n \mathbf{J}_i' \boldsymbol{\Omega}_{i,\boldsymbol{\gamma}}^{-1} \mathbf{J}_i) - \tau^2 (\sum_{i=1}^n \mathbf{J}_i' \boldsymbol{\Omega}_{i,\boldsymbol{\gamma}}^{-1} \mathbf{M}_{i,\boldsymbol{\gamma}}^{-1} \boldsymbol{\Omega}_{i,\boldsymbol{\gamma}}^{-1} \mathbf{J}_i) + (1/c)\mathbf{I}_{K_{\boldsymbol{\gamma}}}$. It is worth to point out that the generation of a candidate $\boldsymbol{\gamma}$ is done by changing one element at a time in fixed sequencing order within each iteration. The conditional

distribution for the informative scalar w_i follows the combination of information from both the regression and proportional hazards models. The likelihood of the PH model leads to its nonstandard form

$$w_i \mid T_i, \delta_i, h_0(t), \mathbf{B}_{i,\boldsymbol{\gamma}}, \boldsymbol{\theta}_{\boldsymbol{\gamma}}, \tau^2 \propto \exp\left\{-\frac{(w_i^2 - 2w_i \mathbf{B}'_{i,\boldsymbol{\gamma}}\boldsymbol{\theta}_{i,\boldsymbol{\gamma}})}{2\tau^2}\right\} \times [h_0(T_i)\exp(w_i)]^{\delta_i} \exp\left\{-\int_0^{T_i} \exp(w_i) h_0(t) du\right\},$$

which can be updated by a Metropolis step.

The following layer includes the regression coefficient as population mean $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$, which can be updated as

$$\boldsymbol{\beta}_{\boldsymbol{\gamma}} \mid \boldsymbol{\beta}_{i,\boldsymbol{\gamma}}, \boldsymbol{\Omega}_{i,\boldsymbol{\gamma}} \sim MVN(\boldsymbol{\beta}^*_{\boldsymbol{\gamma}}, c\mathbf{M}),$$

where $\mathbf{M} = (c\sum_{i=1}^n \mathbf{J}'_i \boldsymbol{\Omega}^{-1}_{i,\boldsymbol{\gamma}} \mathbf{J}_i + \mathbf{I}_{K_{\boldsymbol{\gamma}}})^{-1}$ and $\boldsymbol{\beta}^*_{\boldsymbol{\gamma}} = c\mathbf{M}(\sum_{i=1}^n \mathbf{J}'_i \boldsymbol{\Omega}^{-1}_{i,\boldsymbol{\gamma}} \boldsymbol{\beta}_{i,\boldsymbol{\gamma}})$. The unstructured covariance matrix of the polynomial part for quadratic spline coefficients, $\boldsymbol{\Sigma}$, is updated as

$$\boldsymbol{\Sigma} | \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\beta}_{i,\boldsymbol{\gamma}} \sim IW(\mathbf{A}^*, b^*),$$

where $\mathbf{A}^* = [\mathbf{A}^{-1} + \sum_{i=1}^n (\boldsymbol{\alpha}_{i1}\boldsymbol{\alpha}'_{i1})]^{-1}$, $\boldsymbol{\alpha}_i = \boldsymbol{\beta}_{i,\boldsymbol{\gamma}} - \boldsymbol{\beta}_{\boldsymbol{\gamma}_i} = [\boldsymbol{\alpha}'_{i1}, \boldsymbol{\alpha}'_{i2}]'$, and $b^* = b + n$. Here, the dimensions of $\boldsymbol{\alpha}_{i1}$ and $\boldsymbol{\alpha}_{i2}$ are 3×1 and $K^*_{\boldsymbol{\gamma}_i} \times 1$. Linking to the covariance of the change points part for the quadratic spline coefficients, σ^2, is updated as

$$\sigma^2 \mid \boldsymbol{\beta}'_{i,\boldsymbol{\gamma}} s, \boldsymbol{\beta}_{\boldsymbol{\gamma}} \sim IG(c^*_\sigma, d^*_\sigma),$$

where $c^*_\sigma = c_\sigma + (\sum_{i=1}^n K^*_{\boldsymbol{\gamma}_i})/2$ and $d^*_\sigma = d_\sigma + (\sum_{i=1}^n \boldsymbol{\alpha}'_{i2}\boldsymbol{\alpha}_{i2})/2$. The probability of being change point π can be updated as

$$\pi \mid \boldsymbol{\gamma} \sim Beta(a^*_\pi, b^*_\pi),$$

where $a^*_\pi = a_\pi + K_{\boldsymbol{\gamma}}$ and $b^*_\pi = b_\pi + K + K_{\boldsymbol{\gamma}}$. The common coefficient vector $\boldsymbol{\theta}$ in the linear predictor model is updated as

$$\boldsymbol{\theta}_{\boldsymbol{\gamma}} \mid \mathbf{w}, \mathbf{B}, \tau^2, \mathbf{V}_{\boldsymbol{\gamma}} \sim MVN(\boldsymbol{\theta}^*, \tau^2 \mathbf{V}^*),$$

where $\mathbf{V}^* = (\mathbf{V}^{-1}_{\boldsymbol{\gamma}} + \sum_{i=1}^n \mathbf{J}'_i \mathbf{B}_{i,\boldsymbol{\gamma}} \mathbf{B}'_{i,\boldsymbol{\gamma}} \mathbf{J}_i)^{-1}$ and $\boldsymbol{\theta}^* = \mathbf{V}^*(\sum_{i=1}^n w_i \mathbf{J}'_i \mathbf{B}_{i,\boldsymbol{\gamma}})$. Here the definition of $\mathbf{J}_i$ is similar to its definition in (7.2) with a dimension adjustment to match $\mathbf{B}_{i,\boldsymbol{\gamma}}$. The conjugate inverse gamma prior for variance τ^2 leads to its conditional distribution:

$$\tau^2 \mid \boldsymbol{\theta}_{\boldsymbol{\gamma}}, \mathbf{V}_{\boldsymbol{\gamma}}, \mathbf{w}, \mathbf{B} \sim IG(a^*_\tau, b^*_\tau),$$

where $a^*_\tau = a_\tau + (n + K_{\boldsymbol{\gamma}} + m)/2$ and $b^*_\tau = b_\tau + [\boldsymbol{\theta}'_{\boldsymbol{\gamma}} \mathbf{V}^{-1}_{\boldsymbol{\gamma}} \boldsymbol{\theta}_{\boldsymbol{\gamma}} + \sum_{i=1}^n (w_i - \mathbf{B}'_{i,\boldsymbol{\gamma}} \boldsymbol{\theta}_{i,\boldsymbol{\gamma}})^2]/2$.

The next layer includes scale parameters h_k, which is updated by

$$h_\ell \mid \boldsymbol{\theta}_{\boldsymbol{\gamma}}, \tau^2 \sim IG(c_\ell^*, d_\ell^*),$$

where $c_\ell^* = c_\ell + 1/2$ and $d_\ell^* = d_\ell + \theta_\ell^2/2\tau^2$.

The parameters of baseline hazard step function $h_0(t)$, λ_j's, can be updated using the proportional hazards model:

$$\lambda_j \mid \mathbf{T}, \mathbf{w} \sim IG(a_j^*, b_j^*),$$

where $a_j^* = a_j + \sum_{i=1}^n \delta_i I(s_{j-1} \leq T_i < s_j)$ and $b_j^* = b_j + \sum_{i=1}^n \left[I(T_i > s_{j-1}) \times \int_{s_{j-1}}^{\min(T_i, s_j)} \exp(w_i) \mathrm{d}u \right]$.

References

Baladandayuthapani V, Mallick BK, Young Hong M, Lupton JR, Turner ND, Carroll RJ (2008) Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis. Biometrics 64:64–73

Brown ER, Ibrahim JG (2003) A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. Biometrics 59:221–228

Brown ER, Ibrahim JG, DeGruttola V (2005) A flexible B-spline model for multiple longitudinal biomarkers and survival. Biometrics 61:64–73

Bycott P, Taylor J (1998) A comparison of smoothing techniques for CD4 data measured with error in a time-dependent Cox proportional hazards model. Stat Med 17:2061–2077

Catalona WJ, Smith DS, Ratliff TL, Dodds KM, Coplen DE, Yuan JJ, Petros JA, Andriole GL (1991) Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. N Eng J Med 324:1156–1161

Chen M-H, Shao Q-M, Ibrahim JG (2000) Monte Carlo methods in Bayesian computation. Springer-Verlag, New York

Cox CR (1972) Regression models and life tables. J R Stat Soc, Ser B 34:187–202

Cox DR (1975) Partial likelihood. Biometrika 62:269–276

Daniels M, Pourahmadi M (2002) Bayesian analysis of covariance matrices and dynamic models for longitudinal data. Biometrika 89:553–566

DeGruttola V, Tu XM (1994) Modelling progression of CD4-lymphocyte count and its relationship to survival time. Biometrics 50:1003–1014

Denison DGT, Mallick BK, Smith AFM (1998) Automatic Bayesian curve fiting. J R Stat Soc B 60:333–350

Ding J, Wang JL (2008) Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. Biometrics 64:546–556

Eisenberger M, O'Dwyer P, Friedman M (1986) Gonadotropin-hormone releasing hormone analogues: a new therapeutic approach for prostatic carcinoma. J Clin Oncol 4:414–424

Ellerhorst J, Tu S, Amato R, Finn L, Millikan R, Pagliaro L, Jackson A, Logothetis C (1997) Phase II trial of alternating weekly chemohormonal therapy for patients with androgen-independent prostate cancer. Clin Cancer Res 3:2371–2376

Faucett CJ, Thomas DC (1996) Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. Stat Med 15:1663–1685

Greenlee R, Murray T, Bolden S, Wingo P (2000) Cancer statistics, 2000. CA Cancer J Clin 50:7–33

Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. J Am Stat Assoc 85:398–404

Gelfand AE, Dey DK, Chang H (1992) Model determination using predictive distribution with implementation via sampling based methods. In Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds) Bayesian Statistics 4, 147–167. Oxford University Press, Oxford

Guo X, Carlin BP (2004) Separate and joint modelling of longitudinal and event time data using standard computer packages. Am Stat 58(1):16–24

Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109

Hogan JW, Laird NM (1997) Mixture models for the joint distribution of repeated measures and event times. Stat Med 16:239–257

Holmes CC, Mallick BK (2003) Generalized nonlinear modeling with multivariate free-knot regression splines. J Am Stat Assoc 98:352–368

Hsieh F, Tseng YK, Wang JL (2006) Joint modelling of survival and longitudinal data: likelihood approach revisited. Biometrics 62:1037–1043

Ibrahim JG, Chen M-H, Sinha D (2001) Bayesian survival analysis. Springer-Verlag, New York

Kalbfleisch JD, Prentice RL (2002) The statistical analysis of failure time data, 2nd edn. Wiley, Hoboken

Kohn R, Smith M, Chan D (2001) Nonparametric regression using linear combinations of basis functions. Stat Comput 11:313–322

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. J Chem Phys 21:1087–1092

Millikan R, Wen S, Pagliaro L, Brown M, Moomey B, Do K-A, Logothets C (2008) Phase III trial of androgen ablation with or without three cycles of systemic chemotherapy for advanced prostate cancer. J Clin Oncol 26:5936–5942

Ngo L, Wand MP (2004) Smoothing with Mixed Model Software. J Stat Softw 9:1–54

Pauler DK, Finkelstein DM (2002) Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes. Stat Med 21:3897–3911

Prentice R (1982) Covariate measurement errors and parameter estimation in a failure time regression model. Biometrika 69:331–342

Raboud J, Reid N, Coates RA, Farewell VT (1993) Estimating risks of progressing to AIDS when covariates are measured with error. J R Stat Soc A 156:343–406

Ramsay JO, Silverman BW (2005) Functional data analysis, 2nd edn. Springer-Verlag, New York

Ruppert D, Wand MP, Carroll RJ (2003) Semiparametric regression. Cambridge University Press, New York

Shi SR, Cote RJ, Yang C, Chen C, Xu HJ, Benedict WF, Taylor CR (1996) Development of an optimal protocol for antigen retrieval: a 'test battery' approach exemplified with reference to the staining of retinoblastoma protein (pRB) in formalin-fixed paraffin sections. J Pathol 179:347—352

Slasor P, Laird N (2003) Joint models for efficient estimation in proportional hazards regression models. Stat Med 22:2137–2148

Smith M, Kohn R (1996) Nonparametric regression using Bayesian variable selection. J Econ 75:317–343

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit. J R Stat Soc, Ser B 64:583–639

Tannock I, Osoba D, Stockler M, Ernst D, Neville A, Moore M, Armitage G, Wilson J, Venner P, Coppin C, Murphy K (1996) Chemotherapy with mitoxantrone plus prednisone or prednisone alone for symptomatic hormone-resistant prostate cancer: a Canadian randomized trial with palliative end points. J Clin Oncol 14:1756–1764

Thompson W, Rosen O (2008) A Bayesian model for sparse functional data. Biometrics 64:54–63

Tsiatis AA, Degruttola V, Wulfsohn MS (1995) Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and Cd4 counts in patients with AIDS. J Am Stat Assoc 90:27–37

Tsiatis AA, Davidian M (2004) Joint modeling of longitudinal and time-to-event data: an overview. Stat Sin 14:809–834

Wang Y, Taylor JMG (2001) Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. J Am Stat Assoc 96:895–905

Wulfsohn MS, Tsiatis AA (1997) A joint model for survival and longitudinal data measured with error. Biometrics 53:330–339

Yao F (2007) Functional principal component analysis for longitudinal and survival data. Stat Sin 17:965–983

Ye W, Lin X, Taylor J (2008) Semiparametric modeling of longitudinal measurements and time-to-event data–two-stage regression calibration approach. Biometrics 64:1238–1246

Zhang S, Müller P, Do K-A (2009) A Bayesian semi-parameteric survival model with longitudinal markers. Biometrics 66(2):435–443