Biometrics 000, 000–000 DOI: 000
000 0000

Joint Partially Linear Model for Longitudinal Data with Informative Drop-outs

Sehee Kim[1,*], Donglin Zeng[2], Jeremy M. G. Taylor[1]

[1,*]Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, U.S.A.

[2]Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599, U.S.A.

email: seheek@umich.edu

Summary: In biomedical research, a steep rise or decline in longitudinal biomarkers may indicate latent disease progression, which may subsequently cause patients to drop out of the study. Ignoring the informative drop-out can cause bias in estimation of the longitudinal model. In such cases a full parametric specification may be insufficient to capture the complicated pattern of the longitudinal biomarkers. For these types of longitudinal data with the issue of informative drop-outs, we develop a joint partially linear model, with an aim to find the trajectory of the longitudinal biomarker. Specifically, an arbitrary function of time along with linear fixed and random covariate effects is proposed in the model for the biomarker, while a flexible semiparametric transformation model is used to describe the drop-out mechanism. Advantages of this semiparametric joint modeling approach are the following: 1) it provides an easier interpretation, compared to standard nonparametric regression models, and 2) it is a natural way to control for common (observable and unobservable) prognostic factors that may affect both the longitudinal trajectory and the drop-out process. We describe a sieve maximum likelihood estimation procedure using the EM algorithm, where the Akaike information criterion (AIC) and Bayesian information criterion (BIC) are considered to select the number of knots. We show that the proposed estimators achieve desirable asymptotic properties through empirical process theory. The proposed methods are evaluated by simulation studies and applied to prostate cancer data.

Key words: Joint models; Longitudinal data; Nonparametric maximum likelihood; Partially linear model; Random effects; Sieve maximum likelihood; Transformation models.

1. Introduction

In prostate cancer studies, Prostate-specific Antigen (PSA) has been widely used to make clinical decisions. Higher or rising PSA patterns after treatment are related to an increased risk of prostate cancer recurrence. One important scientific goal is therefore to identify the PSA trajectory among patients who are treated with radiation therapy for localized prostate cancer (Proust-Lima et al., 2008). In cases where the treatment has been successful, PSA levels are expected to drop over the first post-radiation year and then remain stable at low levels (Zagars et al., 1995). In the other cases, however, PSA values after treatment declined for a short period of time and then might rise again at later times, for example, as illustrated in Web Figure 1. Hence, it is desirable to develop a flexible model to capture this nonlinear temporal trend of PSA levels. One key challenge here is that the follow-up of PSA stopped when salvage hormone therapy was initiated, which is known to change the PSA level or when prostate cancer recurred, resulting in possibly informative drop-out which can lead to bias for PSA trajectory estimation if not accounted for properly.

Motivated by the prostate cancer data, we propose a method that flexibly models longitudinal trajectories of biomarkers using a partially linear model, while taking informative drop-outs into account. We treat the informative drop-out as an event which makes subsequent PSA measurements missing, and these unobserved measures are related to the drop-out process via subject-specific random effects. Advantages of this semiparametric joint modeling approach are 1) it provides an easier interpretation of the covariate effects, compared to standard nonparametric regression models, and 2) it is a natural way to control for common (observable and unobservable) prognostic factors that may affect both the longitudinal trajectory and the informative drop-out.

Without considering complications due to the informative drop-out, several authors have studied partially linear models for longitudinal/clustered data (Zeger and Diggle, 1994; Zhang

et al., 1998; Lin and Ying, 2001; Lin and Carroll, 2001). Other related work about joint models with time-varying coefficients, but not considering informative drop-outs, includes Cai et al. (2012) and Lu and Huang (2015). In longitudinal studies, the problem of informative drop-out has received enormous attention. Hogan and Laird (1997) provided an excellent review of model-based approaches to handling incomplete longitudinal measurements, where most of the existing methods are based on the full parametric specification of the longitudinal model and dependence structure of the drop-out mechanism. We refer the readers to Hogan and Laird (1997) for a more detailed explanation and discussion of the limitations of these parametric models. The conventional pattern-mixture and selection models reviewed by them were extended by Roy (2003) and Beunckens et al. (2008), who allowed multiple latent subgroups of subjects (i.e., joint latent class models). The assumption on latent subgroups has been relaxed by Muthén et al. (2011) so that a subject's subgroup can differ for drop-out and outcomes in their pattern-mixture and selection models. However, all these methods assume some parametric or even linear trends for the longitudinal trajectories, therefore not appropriate for our application.

Roy and Lin (2005) considered longitudinal data with missing time-varying covariates in addition to informative drop-outs. They used a generalized linear mixed model for continuous or binary longitudinal outcomes, and used a transition model to estimate missing time-varying covariates. For informative drop-out, they used a conventional selection model where a separate logistic regression model was fit at every scheduled visit, including the missing longitudinal outcome as a covariate. Viewing the drop-out time as a continuous random variable, the proportional hazards model is commonly assumed to characterize the drop-out process, however, it may not be appropriate in some cases. For example, patients receiving more aggressive treatment may suffer elevated risk in the beginning but may benefit in the long term if they tolerate the treatment. Considering a general class of transformations for

such cases will lead to a better model fit. The transformation models for a single survival time and recurrent event times have been extensively studied, dating back to Dabrowska and Doksum (1988) and more recently by Kosorok et al. (2004) and Kim et al. (2012).

In this paper, we propose a joint partially linear model for longitudinal data with the issue of informative drop-out, where the informative drop-out is allowed to be dependent on covariates. By applying general transformation models for informative drop-out, we do not restrict to the proportional hazards cases. The underlying trajectory is estimated using a B-spline approach. A key advantage of a B-spline approach is the computational simplicity of using a small number of knots and a parametric regression-type implementation via the sieve approximation, which counterbalances the model complexity in the joint modeling approach. The knot selection procedures based on AIC and BIC are described in Section 3 and numerically evaluated in Sections 4-5. Through our flexible, but readily interpretable, modeling approach, we can detect significant changes in the trajectory, which may not be found using linear mixed effects models due to their parametric model constraints, while correcting biases caused by the informative drop-outs.

2. Joint Partially Linear Model (JPLM)

Let $Y(t)$ be the longitudinal response at time t, and let T be the informative drop-out time. We define $\mathcal{X} = \{X(t);\, t \geqslant 0\}$ and $\mathcal{Z} = \{Z(t);\, t \geqslant 0\}$ as the covariate processes of fixed and random effects, respectively. The vectors of external covariates $X(t)$ and $Z(t)$ are possibly time-varying, and their subsets are denoted as $X_k(t)$ and $Z_k(t)$ $(k = 1, 2)$ in the models (1) and (2). We consider a partially linear model for $Y(t)$

$$Y(t) \;=\; \alpha(t) + \beta^T X_1(t) + b^T Z_1(t) + \epsilon(t), \tag{1}$$

and a transformed Cox model for T with the cumulative hazard function

$$\Lambda(t \,|\, \mathcal{X}, \mathcal{Z}, b) \;=\; H\left(\int_0^t \exp\{\gamma^T X_2(u) + (\phi \circ b)^T Z_2(u)\}\, d\Lambda(u) \right), \tag{2}$$

where $\alpha(t)$ is the unspecified underlying trajectory, β and γ are the vectors of unknown regression coefficients, $\Lambda(\cdot)$ is an unspecified non-decreasing function, and $\epsilon(t)$ is a white noise process with variance σ_e^2. To account for the correlation between $Y(t)$ and T, we introduce a common latent variable b, following a (multivariate) normal distribution with mean zero and covariance matrix Σ_b. We assume $Y(t)$ and T are independent, conditional on $\mathcal{X}$, $\mathcal{Z}$ and b. In the model (2), ϕ is a set of unknown constants with the same number of elements as b, and $\phi \circ b$ denotes the component-wise product of ϕ and b. For instance, if one would expect that the informative drop-out is related to the current level of a biomarker, as in our application, then $(\phi \circ b)^T Z_2(t)$ can take the form $\tilde{\phi}(b_1 + b_2 t)$ by taking $Z_2(t) = (1,\, t)^T$, $b^T = (b_1, b_2)$ and $\phi^T = (\tilde{\phi},\, \tilde{\phi})$. We note that each patient's longitudinal responses and informative drop-out rate are linked through the unobservable latent factors b as well as the observed common covariates, for example, if $X_1(t) = X_2(t)$. The amount of variation in the informative drop-out process due to the latent factors is characterized by ϕ.

In the model (2), the transformation function $H(\cdot)$ is assumed to be continuously differentiable and strictly increasing, and is required to be specified in the analysis. For example, $H(x)$ can take the form of the logarithmic transformation,

$$H(x) = \begin{cases} \log(1 + \eta x)/\eta, & \eta > 0 \\ x, & \eta = 0. \end{cases}$$

The choices of $\eta = 0$ and $\eta = 1$ lead to the proportional hazards model and the proportional odds model, respectively.

Let C be the non-informative drop-out time (e.g. administrative end date) assumed to be independent of $\{Y(\cdot),\, T,\, b\}$ given $\mathcal{X}$ and $\mathcal{Z}$, and let $V = \min(T, C)$ denote the observed drop-out time. The observed data for the ith subject with n_i repeated measurements are denoted by $O_i = \{Y_i(t_{ij}),\, V_i,\, \Delta_i,\, X(u),\, Z(u);\, t_{ij} \leqslant V_i,\, u \leqslant V_i,\, i = 1, \ldots, n,\, j = 1, \ldots, n_i\}$, where $\Delta_i = I(T_i \leqslant C_i)$ with $I(\cdot)$ being the indicator function. The log-likelihood function

for the observed data is given by

$$\sum_{i=1}^{n} \log \int_b \prod_{j=1}^{n_i} \left[(2\pi\sigma_e^2)^{-1/2} \exp\left\{ -(Y_i(t_{ij}) - \alpha(t_{ij}) - \beta^T X_{1i}(t_{ij}) - b^T Z_{1i}(t_{ij}))^2 / (2\sigma_e^2) \right\} \right]$$
$$\times \left[\lambda(V_i)\, e^{\gamma^T X_{2i}(V_i) + (\phi \circ b)^T Z_{2i}(V_i)}\, H' \left(\int_0^{V_i} e^{\gamma^T X_{2i}(u) + (\phi \circ b)^T Z_{2i}(u)}\, d\Lambda(u) \right) \right]^{\Delta_i}$$
$$\times \exp\left\{ -H \left(\int_0^{V_i} e^{\gamma^T X_{2i}(u) + (\phi \circ b)^T Z_{2i}(u)}\, d\Lambda(u) \right) \right\} \times f(b;\, \Sigma_b)\, db, \tag{3}$$

3. Inference Procedure

We propose a maximum likelihood estimation procedure to estimate the finite dimensional parameters $\theta = (\beta,\, \gamma,\, \phi,\, \sigma_e^2,\, \mathrm{Vec}(\Sigma_b))$, the underlying trajectory function $\alpha(t)$, and the baseline cumulative hazard function $\Lambda(t)$, where $\mathrm{Vec}(\Sigma_b)$ denotes the vector consisting of the upper triangular elements of Σ_b. Specifically, in Section 3.1 we use sieve maximum likelihood estimation for $\alpha(t)$ (Geman and Hwang, 1982; Shen and Wong, 1994) in which $\alpha(t)$ is approximated by a combination of known basis functions (e.g., cubic B-splines) and unknown sieve coefficients, while in Section 3.2 we use nonparametric maximum likelihood estimation (NPMLE) for $\Lambda(t)$ by allowing $\Lambda(t)$ to be any increasing right-continuous function. The transformation $H(x)$ is fixed.

3.1 *Sieve Approximation for* $\alpha(t)$

Suppose that subjects are followed up to a fixed time τ. We approximate $\alpha(t)$ in (1) through a finite number of basis functions in a sieve space of t in $\mathbb{T} = [0, \tau]$ as follows:

$$\alpha(t) \simeq \sum_{k=1}^{m+K_n} \zeta_k B_k^m(t),$$

where $\{B_k^m(\cdot)\}$ is a basis function of t with the highest degree less than m, ζ_k is the regression coefficient with a fixed knot sequence, and K_n is the number of interior points in the sieve space. Rigorously speaking, the sieve space for $\alpha(t)$ is defined as

$$S_n(m, K_n, M_n) = \left\{ \alpha(t) :\ \alpha(t) = \sum_{k=1}^{m+K_n} \zeta_k B_k^m(t),\ \sum_{k=1}^{m+K_n} |\zeta_k| \leqslant M_n \right\},$$

on a finite partition of $\mathbb{T}$

$$\{s_1 = \cdots = s_m = 0 < s_{m+1} < \cdots < s_{m+K_n} < \tau = s_{m+K_n+1} = \cdots = s_{2m+K_n}\},$$

where the constants K_n and M_n depend on the data and the sample size (n). The boundedness condition $\sum_{k=1}^{m+K_n} |\zeta_k| \leqslant M_n$ guarantees the sieve space defined by $S_n(m, K_n, M_n)$ is a bounded set in a finite dimensional space. Unlike parametric regression, both the number of knots and the coefficient of the basis function at each knot need to be estimated from the data. In practice, m is usually chosen to be at least 2, which corresponds to a linear function. In particular, we use cubic B-spline functions ($m = 4$),

$$B_k^m(t) = \frac{t - s_k}{s_{k+m-1} - s_k} B_k^{m-1}(t) + \frac{s_{k+m} - t}{s_{k+m} - s_{k+1}} B_{k+1}^{m-1}(t), \qquad \text{for } t \in [0, \tau]$$

and $B_k^1(t) = I(s_k \leqslant t < s_{k+1})$ for $k = 1, \ldots, (m + K_n)$. By the properties of B-splines, for a given t value, only at most m basis functions among $\{B_k^m(t)\}$ are nonzero, therefore, $\alpha(t)$ is approximated by a linear combination of $\{B_k^m(t)\}$ on m nearest knot points at any point t. Therefore, conditional on $\{B_k^m(t)\}$, and hence K_n and $\{s_k\}$, we can use the methodology that has been developed for the parametric longitudinal data analysis in this nonparametric context. It consequently reduces the computational burden of using nonparametric estimation in both longitudinal and survival components.

For the knot locations $\{s_k\}$, equally-spaced knots are commonly used. For longitudinal studies with the issue of informative drop-outs, however, we suggest choosing $\{s_k\}$ based on the observed data. It can prevent an numerical problem in $\{\zeta_k\}$ caused by sparsity in the later study period. To determine the number of interior knots with the best fit to the data, we use the AIC- or BIC-based selection procedures. The performance of AIC as a knot selection criterion has previously been investigated by Shibata (1981) and Ding and Wang (2008) among others, but that of BIC has not been studied in the joint modeling context. The proposed knot selection procedure is as follows. For a given K_n, locations of interior knots are determined as every distinct $q_{100/(K_n+1)}$, the $100/(K_n + 1)$th percentile of the observed

longitudinal measurement times. For example, when $K_n = 3$ and $K_n = 9$ are considered, $\{q_{25}, q_{50}, q_{75}\}$ and $\{q_{10}, q_{20}, \ldots, q_{90}\}$ are used as the locations of interior knots, respectively. We next repeatedly fit the joint model using the candidate interior points $K_n \in \{2, 3, \ldots\}$, and calculate AIC (or BIC). Then, the model with the smallest AIC (or BIC) is considered the best fitting one. In our simulation study, we observed that AIC and BIC selected the same model approximately 42% of the time, and the number of knots selected by AIC was always greater than or equal to that selected by BIC. Moreover, our numerical evaluation was consistent with the results obtained in Huang et al. (2002); the same estimate $\hat{\alpha}$ can be achieved through different sets of basis functions and their corresponding $\{\hat{\zeta}_k\}$. Further detailed comparisons of the two selection procedures are provided in Sections 4 and 5.

3.2 *Nonparametric Maximum Likelihood Estimation for* $\Lambda(t)$

Using the NPMLE approach, we treat Λ as a nondecreasing step function with jumps only at the observed failure times and replace $\lambda(t)$ with the jump size of Λ at t, denoted by $\Lambda\{t\}$, in the log-likelihood function (3). For commonly used transformation functions such as log-transformation, $\exp\{-H(x)\}$ can be expressed as the Laplace transformation of some function $\delta(\xi)$ for $\xi \geqslant 0$, such that $\exp\{-H(x)\} = \int_0^\infty \exp(-x\xi)\,\delta(\xi)\,d\xi$. For example, if we choose a gamma frailty ξ with mean one and variance η, then it holds that $H(x) = \log(1+\eta x)/\eta$. Applying the Laplace transformation, the observed log-likelihood function (3) can be rewritten as $l_n(\theta, \zeta, \Lambda\{\cdot\})$

$$
\begin{aligned}
= \sum_{i=1}^{n} \log \int_b (2\pi\sigma_e^2)^{-\frac{n_i}{2}} \exp\left[-\sum_{j=1}^{n_i} \frac{\left\{ Y_i(t_{ij}) - \zeta^T B^m(t_{ij}) - \beta^T X_{1i}(t_{ij}) - b^T Z_{1i}(t_{ij}) \right\}^2}{2\sigma_e^2} \right] \\
\times \int_\xi \left[\xi\, \Lambda\{V_i\}\, \exp\{\gamma^T X_{2i}(V_i) + (\phi \circ b)^T Z_{2i}(V_i)\} \right]^{\Delta_i} \\
\times \exp\left\{ -\int_0^{V_i} \xi\, e^{\gamma^T X_{2i}(u) + (\phi \circ b)^T Z_{2i}(u)}\, d\Lambda(u) \right\} \delta(\xi)\, d\xi \times f(b;\, \Sigma_b)\, db, \qquad (4)
\end{aligned}
$$

where $\zeta = (\zeta_1, \ldots, \zeta_{K_n})^T$, $B^m(t) = (B_1^m(t), \ldots, B_{K_n}^m(t))^T$, and ξ is assumed to be independent of b.

The most attractive feature about writing the transformation in this way is that the modified log-likelihood (4) can be seen as the proportional hazards frailty model (Kosorok et al., 2004) with the conditional hazard function

$$\lambda(t \,|\, \mathcal{X}, \mathcal{Z}, \xi, b) = \xi \, \lambda(t) \exp\{\gamma^T X_{2i}(t) + (\phi \circ b)^T Z_{2i}(t)\}.$$

This makes the algorithm more stable and computationally efficient. Now, the MLEs can be obtained by maximizing the modified log-likelihood function over $S_n(m, K_n, M_n)$, θ and all jump sizes of Λ at the observed failure times. Since this maximization involves unobservable variables ξ and b, it can be carried out through the following EM algorithm, treating ξ and b as missing data.

3.3 *EM Algorithm*

We describe the EM algorithm (Dempster et al., 1977) to compute the MLEs of $(\theta, \zeta, \Lambda\{\cdot\})$. In the E-step, we calculate conditional expectations of certain functions of (ξ, b) given the observed data O_i, say $\hat{E}[\xi \, g_i(b) \,|\, O_i]$. Hereafter, we omit to write that the expectations are conditional on the observed data and the current parameter estimates, and abbreviate such expectation $\hat{E}[\xi \, g_i(b) \,|\, O_i]$ as $\hat{E}[\xi \, g_i(b)]$. Computation of this expectation can be simplified by first obtaining the *nested* conditional expectation of ξ, given b and O_i. That is, $\hat{E}[\xi \, g_i(b)]$ can be calculated as $\hat{E}_b[\, \hat{E}_\xi[\xi \,|\, b] \; g_i(b)]$. Since the conditional distribution of ξ given b is proportional to

$$h(\xi, b) = \xi^{\Delta_i} \exp\left\{ - \int_0^{V_i} \xi \, e^{\gamma^T X_{2i}(u) + (\phi \circ b)^T Z_{2i}(u)} \, d\Lambda(u) \right\},$$

the conditional expectation of ξ given b has the form of

$$\hat{E}_\xi \left[\xi \,|\, b \right] = \int \xi \, \frac{h(\xi, b) \, \delta(\xi)}{\int h(\xi, b) \, \delta(\xi) \, d\xi} \, d\xi = H'\big(\tilde{x}_i(b)\big) - \left[\frac{H''\big(\tilde{x}_i(b)\big)}{H'\big(\tilde{x}_i(b)\big)} \right]^{\Delta_i},$$

where $\tilde{x}_i(b) = \int_0^{V_i} e^{\gamma^T X_{2i}(u) + (\phi \circ b)^T Z_{2i}(u)} \, d\Lambda(u)$. Once $\hat{E}_\xi \left[\xi \,|\, b \right]$ is calculated, which is a function of b, the conditional expectation $\hat{E}[\xi \, g_i(b)]$ can be computed using numerical approximation methods such as Gaussian quadrature with Hermite orthogonal polynomials. Note that the

conditional distribution of b given O_i is proportional to $\Gamma(O_i|\, b) f(b; \Sigma_b)$, where

$$\Gamma(O_i|\, b) = \exp\left\{ -\sum_{j=1}^{n_i} \left[\{b^T Z_{1i}(t_{ij})\}^2 - 2b^T Z_{1i}(t_{ij}) \{Y_i(t_{ij}) - \zeta^T B^m(t_{ij}) - \beta^T X_{1i}(t_{ij})\} \right] \right\}$$
$$\times \exp\left\{ \Delta_i (\phi \circ b)^T Z_{2i}(V_i) + \Delta_i \log H'\big(\tilde{x}_i(b)\big) - H\big(\tilde{x}_i(b)\big) \right\}.$$

We thus calculate the conditional expectation by

$$\hat{E}[\xi\, g_i(b)] = \int_b \hat{E}_\xi[\xi \mid b]\, g_i(b)\, \frac{\Gamma(O_i|\, b) f(b; \Sigma_b)}{\int_b \Gamma(O_i|\, b) f(b; \Sigma_b)\, db}\, db.$$

In the M-step, we maximize the expectation of the complete-data log-likelihood function:

$$\sum_{i=1}^{n} \sum_{j=1}^{n_i} \left\{ -\log \sigma_e^2/2 - \hat{E}\left[\{Y_i(t_{ij}) - \zeta^T B^m(t_{ij}) - \beta^T X_{1i}(t_{ij}) - b^T Z_{1i}(t_{ij})\}^2/(2\sigma_e^2) \right] \right\}$$
$$+ \sum_{i=1}^{n} \Delta_i \left\{ \log \xi + \log \Lambda\{V_i\} + \gamma^T X_{2i}(V_i) + \hat{E}[\phi \circ b]^T Z_{2i}(V_i) \right\}$$
$$+ \sum_{i=1}^{n} \left\{ -\hat{E}\left[\int_0^{V_i} \xi\, e^{\gamma^T X_{2i}(u) + (\phi \circ b)^T Z_{2i}(u)}\, d\Lambda(u) \right] + \hat{E}\left[\log \delta(\xi) + \log f(b; \Sigma_b) \right] \right\}.$$

Maximizing the above objective function over (ζ, β, σ_e^2, Σ_b) is as simple as in a classic linear regression; whereas the rest of the parameters (γ, ϕ, $\Lambda\{.\}$) do not yield a closed-form of the maximizers. Using a reliable numerical approach, we solve the following equation for γ:

$$\sum_{i=1}^{n} \Delta_i \left\{ X_{2i}(V_i) - \frac{\sum_{j=1}^{n} R_j(V_i) X_{2j}(V_i)\, \hat{E}\left[\xi\, q_{2j}(V_i)\right]}{\sum_{j=1}^{n} R_j(V_i)\, \hat{E}\left[\xi\, q_{2j}(V_i)\right]} \right\} = 0, \tag{5}$$

and the following equation for ϕ:

$$\sum_{i=1}^{n} \Delta_i \left\{ \hat{E}\left[b \circ Z_{2i}(V_i)\right] - \frac{\sum_{j=1}^{n} R_j(V_i)\, \hat{E}\left[\xi\, q_{2j}(V_i)(b \circ Z_{2j}(V_i))\right]}{\sum_{j=1}^{n} R_j(V_i)\, \hat{E}\left[\xi\, q_{2j}(V_i)\right]} \right\} = 0, \tag{6}$$

where $R_j(t) = I(V_j \geqslant t)$ and $q_{2j}(t) = \exp\{\gamma^T X_{2j}(t) + (\phi \circ b)^T Z_{2j}(t)\}$. In addition, Λ is estimated as a step function with the following jump size at V_i:

$$\Lambda\{V_i\} = \frac{\Delta_i}{\sum_{j=1}^{n} R_j(V_i)\, \hat{E}\left[\xi\, q_{2j}(V_i)\right]}. \tag{7}$$

At each M-step, we update γ and ϕ by solving the equations (5) and (6) through a one-step Newton-Raphson algorithm, and update the jump sizes of Λ by equation (7).

To obtain the MLEs, we iterate the E-step and M-step until the parameter estimates converge. The variances of the MLEs can be estimated from the inverse of the observed

information matrix of all parameters (θ, ζ, $\Lambda\{\cdot\}$). The observed information matrix can be computed from the complete data log-likelihood function denoted by ℓ_i^c for the ith subject using the following Louis formula (Louis, 1982) of

$$-\sum_{i=1}^{n} \hat{E}[\nabla^2 \ell_i^c(b) \,|\, O_i] - \sum_{i=1}^{n} \left\{ \hat{E}[\nabla \ell_i^c(b)^{\otimes 2} \,|\, O_i] - \hat{E}[\nabla \ell_i^c(b) \,|\, O_i]^{\otimes 2} \right\}, \tag{8}$$

where $u^{\otimes 2} = uu^T$, ∇ and ∇^2 are the first and second derivatives with respect to parameters, and $\hat{E}$ is the conditional expectation of a function of b given the observed data.

3.4 *Asymptotic Properties*

One of the attractive features of the proposed estimator is that its large sample properties can be shown by using techniques from empirical process theory. Let ($\hat{\theta}$, $\hat{\alpha}$, $\hat{\Lambda}$) denote the estimator maximizing (4), and let (θ_0, α_0, Λ_0) denote the true parameter values. Zeng and Cai (2005) showed the strong consistency and asymptotic normality of $\hat{\theta}$ and $\hat{\Lambda}(\cdot)$ when $\alpha_0(t)$ is constant over time. We relax their full parametric assumptions for $Y(t)$ by adding a nonparametric functional component into their linear mixed effects models. The key difference in the proof is to find an upper bound of $\sup_{t\in[0,\tau]} |\hat{\alpha}(t) - \alpha_0(t)|$. The upper bound is constructed based on a linear span of $\alpha_0(t)$ into the sieve space $S_n(m, K_n, M_n)$ that consists of functions with uniformly bounded rth ($r \geqslant 2$) derivatives. Under the mild regularity conditions (A1)-(A11) stated in the Web Appendix A, we can show that if the smoothing parameters satisfy $M_n = O(\log\log n)$ and $K_n = O(n^{r_0})$ with $1/(4r) < r_0 < 1/3$, then the MLEs are uniformly consistent in the sense that $\|\hat{\theta} - \theta_0\| = o_p(1)$, $\sup_{t\in[0,\tau]} |\hat{\Lambda}(t) - \Lambda_0(t)| = o_p(1)$ and $\sup_{t\in[0,\tau]} |\hat{\alpha}(t) - \alpha_0(t)| = o_p(1)$, where $\|\theta\| = \sqrt{\theta^T\theta}$. From the consistency and the choice of $M_n = O(\log\log n)$, we can further show that $\|\hat{\theta} - \theta_0\|^2 + \|\hat{\Lambda}(t) - \Lambda_0(t)\|^2_{L_2(P)} + \|\hat{\alpha}(t) - \alpha_0(t)\|^2_{L_2(P)} = o_P(n^{-1/2}) + O_P(K_n^{-2r})$. By choosing $K_n = O(n^{1/(4r-1)})$, for example, we can obtain the $\sqrt{n}$-convergence rate of the MLEs. Finally, it is proved that $n^{1/2}(\hat{\theta} - \theta_0)$ is asymptotically normal with mean zero and $\hat{\theta}$ is a semiparametric efficient estimator for θ_0. The asymptotic covariance of $\hat{\theta}$ can be estimated

by (8), and its finite sample properties are shown to be good in Section 4. A detailed proof of the asymptotic properties is given in Web Appendix A.

4. Simulation Studies

To assess the performance of the proposed method, we conduct extensive simulation studies under settings mimicking the prostate cancer data. Specifically, using the analysis results in Figure 3 and Table 3 as a basis, we assume that true $\alpha(t) = (t+0.5)^{-1.5} + (t/2+1)^3 - 4$ and $\phi = 1.7$, choose $n = 500$, and target a similar early drop-out rate as in the real data where 50% and 75% of informative drop-outs occurred within 3 and 5 years, respectively, over the maximum follow-up time of 14 years. For simplicity we include one observed covariate x, baseline PSA level, following a normal distribution with true $\beta = \gamma = 0.5$. Then, the longitudinal outcomes are generated from $Y(t) = \alpha(t) + \beta x + b_1 + b_2 t + \epsilon(t)$, where $\epsilon(t) \sim N(0, \sigma_e^2)$ with $\sigma_e^2 = 0.1$ and (b_1, b_2) are from a normal distribution with zero means, $\sigma_{b_1}^2 = 0.4$, and $\sigma_{b_2}^2 = 0.2$. The correlation between b_1 and b_2 is $\rho = -0.1$. Informative drop-out times are generated from a transformation model taking the form $\Lambda(t \mid x, b_1) = H(\exp(\gamma x + \phi b_1)\, \Lambda(t))$.

Various missing mechanisms are explored by varying $H(\cdot)$ and ϕ. We consider two popular models for survival data: the proportional hazards model $H(x) = x$ (when $\eta = 0$) and the proportional odds model $H(x) = \log(1 + x)$ (when $\eta = 1$). Assuming true $\phi = 0$ represents missing at random (MAR). True $\phi = 0.5$ and $\phi = 1.7$ represent missing not at random (NMAR) with the same and larger effect size than the coefficient of the observed covariate ($\gamma = 0.5$). These settings produce about 6 measurements of Y, on average.

We also investigate the impact of knot selection procedures on $\alpha(t)$ estimation when AIC or BIC is used. We repeatedly fit the joint model using interior points $K_n \in \{2, 3, \ldots, 20\}$, where the locations of knots are determined by every distinct $100/(K_n + 1)$th percentile of the observed measurement times. The best numbers of interior knots by AIC (or BIC), denoted by k_{AIC} (or k_{BIC}), may be different in each dataset. In our simulation settings, we

observe that k_{AIC} equals k_{BIC} for 42% of the data sets, and k_{AIC} is always greater than or equal to k_{BIC}. For comparison purposes, in Table 1 we also report estimation results when the interior knots are always fixed with $K_n = 2$, $K_n = 4$, or $K_n = 8$.

Simulation results based on 1000 replications are presented in Table 1 and Table 2 for various missing mechanisms by different ϕ and $H(\cdot)$, where Bias is the average of the differences between the true parameter and the estimates, SD is the sample standard deviation of the parameter estimator, SEE is the average of the standard error estimates, and CP is the coverage probability of 95% confidence intervals. The confidence intervals of variance components are constructed based on the Satterthwaite approximation.

Table 1 and Web Tables 1-2 report the performance of $\alpha(t)$ estimation in the joint modeling approach, compared with that using the marginal modeling approach (i.e., ignoring the informative drop-out), in terms of Bias, SD, the mean square error (MSE), and the ratio of the MSE for joint estimates to the marginal estimates (MSER). Under MAR (i.e., $\phi = 0$), both approaches are similar in Bias and SD. However, under NMAR (i.e., $\phi = 0.5$ or 1.7), the marginal approach resulted in a larger Bias but with a similar magnitude of SD, which indicates that the joint modeling approach leads to a more accurate and efficient estimator.

In Table 1, when the number of interior knots (K_n) increases, the SD consistently increases, whereas the Bias decreases or increases in earlier or later observations times, respectively. The AIC-based knot selection procedure tends to result in smaller biases than the BIC-based knot selection procedure but at the cost of larger variation. It is worth noting that, when AIC- or BIC-selected knots are used, part of SD is attributed to the uncertainty of model selection, whereas such variability does not exist with the fixed knots. In Table 1, we observe that SDs with the AIC- and BIC-selected knots are similar to SDs with the fixed knots, indicating that the major source of variations in estimates was attributed to the magnitude of K_n itself, rather than the variability in K_n chosen differently by AIC or BIC.

This might be the reason why the coverage probabilities could remain close to the nominal level 95%, even if not accounting for the uncertainty of model selection. Figure 1 shows the estimates of $\alpha(t)$ for all t, using different knot selection procedures and transformations. The average, minimum and maximum of estimates over 1000 simulated datasets indicate that the proposed estimator $\hat{\alpha}(t)$ behaves well for both transformation models when BIC was used. However, the AIC-estimates appeared to have larger variations in both tails. Therefore, we suggest using the BIC-based knot selection procedure, based on negligibly small biases and consistently smaller SDs in all scenarios we studied.

Table 2 shows good performance of the remaining parameters for both $H(x) = x$ and $H(x) = \log(1 + x)$ when BIC-based knot selection procedure was used. That is, the MLEs are unbiased, the standard error estimates calculated using the Louis formula reflect well the true variations in the proposed estimators, and the coverage probabilities are in a reasonable range. When AIC was used as the knot selection criterion, simulation results are very similar to those in Table 2 and hence omitted here (see Web Table 3).

5. Data Application

We illustrate the application of the proposed method using prostate cancer data from the University of Michigan. The data were collected from a total of 503 patients with average age of 69 (range of 34-86) who received planned radiation therapy as the primary treatment method. The objective of this analysis was to identify the trajectory of post-radiation PSA change, while correctly accounting for the informative drop-out either by the start of salvage hormone therapy or by tumor recurrence. A detailed description of the data and possible clinical impact achieved from the study are discussed in Proust-Lima et al. (2008) and Taylor et al. (2013).

In this cohort, 118 patients (23.5%) dropped out, and the PSA level was measured 9 times on average within a median follow-up time of 4.5 years. The model to characterize the PSA

changes in the presence of informative drop-out was

$$\begin{aligned} Y(t) &= \alpha(t) + \beta^T X + b_1 + b_2 t + \epsilon(t), \\ \Lambda(t \mid X, b) &= H\left(\int_0^t \exp\{\gamma^T X + \phi(b_1 + b_2 u)\}\, d\Lambda(u) \right), \end{aligned}$$

where $Y(t)$ is the observed values of log(PSA(t) + 0.1) at time t, $\Lambda(t \mid X, b)$ is the cumulative hazard function of time to informative drop-out from the end of radiation therapy, and the prognostic factors (denoted as X) are log(baseline PSA + 0.1), T-stage, Gleason score, and age at diagnosis. The log-transformation of PSA values and the choice of which covariates to include were based on the findings in Proust-Lima et al. (2008). A subject-specific random intercept and time slope (i.e., $b_1 + b_2 t$) were included in both longitudinal and survival components to account for the dependence of informative drop-out on the PSA trajectory. The inclusion of $b_1 + b_2 t$ in the survival component implied that the time to informative drop-out was linked to the current PSA level.

To adjust for the informative drop-out mechanism better, we assumed different transformation models $H(x) = \log(1 + \eta x)/\eta$ by varying η values in the range [0, 1]. This was because, when $\eta = 0$ or $\eta = 1$ can be assumed, the chosen transformation provides useful interpretations for the drop-out process. That is, $\eta = 0$ implies that the unit change in a covariate has a linear impact on the log-hazard of dropping-out. The choice of $\eta = 1$ implies the data fits better to a model with a linear increment in the log-odds of dropping-out per unit change in a covariate.

To select the best transformation model η and interior knots K_n, model selection approaches such as AIC and BIC are considered. Under each set of (η, K_n) in $\eta \in \{0, 0.1, \ldots, 1\}$ and $K_n \in \{2, 4, \ldots, 12\}$, we computed the MLEs for the regression coefficients using the proposed method, and the resulting AIC and BIC values were compared. The search range of K_n was chosen to cover the upper and lower bounds of K_n satisfying Conditions (A9) - (A11) in Web Section A. In Figure 2 the smallest BIC value corresponding to $(\eta, K_n) = (1, 5)$

indicated that use of 5 interior knots and the proportional odds model produced the best fit to the data. As a sensitivity analysis, we expanded the range of the transformation parameter to $\eta > 1$, leading to the smallest BIC value at $\eta = 1.6$ and $K_n = 6$ (Web Figure 2). Since the decrement in BIC value was very small, for a better interpretation in practice, we still reported the results from the proportional odds model.

Table 3 summarizes the analysis results under the selected best model. The positive $\hat{\phi} = 1.730$ (P<0.001) indicates that a higher current PSA level was significantly associated with a higher rate of informative drop-out. The rate of informative drop-out statistically significantly increased with a higher baseline PSA level, T-stage and Gleason score and for older patients. On the other hand, there was no significant difference in post-radiation PSA level by T-stage and Gleason score, although it was significantly affected by baseline PSA level and patient's age. We noticed that overall results for time-constant covariates were similar between joint and marginal (i.e., ignoring the informative drop-outs) analyses, whereas the results for the temporal trend in PSA level were different. In Figure 3, circles and dots present the full history of all post-radiation PSA values for patients whose follow-up was informatively and non-informatively dropped out, respectively. The mark "×" on some circles indicates the last observation of PSA before the informative drop-out occurred. Based on the descriptive summary indicated by the circles, dots and ×'s, it was observed that patients who dropped out informatively had higher PSA scores and shorter follow-up time in general. When we compare the curve for the 'Joint Estimate' to that for the 'Marginal Estimate' in Figure 3, we can see that the joint modeling approach led to the estimated underlying PSA trajectory curve being lower than the marginal estimate. This might be because the joint modeling is a method to account for the fact that some of the PSA values were observed under some degree of disease progression, which could cause informative drop-out due to the tumor recurrence or sufficient concern that the patient considered hormone

therapy. Moreover, the reduction in the estimate by the joint modeling approach was larger at longer times.

Sensitivity to the inference based on BIC selection was examined by comparing the results by AIC selection. Web Figure 3 shows that the AIC-based procedure selected the model with $(\eta, K_n) = (1, 7)$, however, the resulting estimates were very similar to those based on the BIC-based selection procedure and hence omitted here (see Web Figure 4 and Web Table 4).

6. Discussion

We have discussed a method of fitting a partially linear model for longitudinal data with informative drop-out, which can handle a covariate-dependent drop-out mechanism through the transformed survival model. For estimation of the model parameters, we have maximized the likelihood, and the resulting MLEs have been theoretically justified. The proposed joint modeling approach has clearly shown the capability to correct biases induced by ignoring informative drop-out using simulated data and a real example.

By exploring a broad class of models for the missing data through varying $H(\cdot)$ and ϕ, the proposed method reduces errors due to the misspecified missing data mechanism to some extent. The proposed methodology, however, does require extensive modeling assumptions, including specification of the drop-out mechanism, the mean and covariance structures, such as linearity in the covariate effects, and a normality assumption of latent variables. Therefore, sensitivity analysis to these assumptions as well as future efforts in research and development of model checking tools are needed to promote practical application of the proposed joint partially linear model (JPLM). In the JPLM, the distribution of the outcomes after drop-out is nonidentifiable, and thereby we assume that it remains the same as before the drop-out. One simple approach to exploring sensitivity to this untestable assumption is to 1) introduce a sensitivity parameter as the difference between the mean of observed and unobserved responses, and then 2) examine how sensitive the results are over a clinically plausible range

of the sensitivity parameter (for more details, see Web Section D). For the assessment of JPLM's fit to observed data, a graphical inspection tool has been illustrated in Web Section C. With regard to more rigorous model diagnostic procedures, the main difficulty is that some model assumptions are made about unobserved variables, and hence the standard model diagnostics based on the observed data alone are not sufficient. One possible research direction is to adopt the multiple-imputation-based diagnostic method by Rizopoulos et al. (2010). The key idea of Rizopoulos et al. (2010) is to create multiple sets of complete data by resampling missing longitudinal outcomes from the posterior distribution given the observed data, and apply standard model diagnostics for mixed effects models and survival models with complete data. Since the general framework needed for the multiple imputation procedures in Rizopoulos et al. (2010) has already been established in equation (5) and Sections 3.3 and 3.4, the extension to survival transformation models appears promising. The assumption on the linear effects of covariate can be relaxed by extending to time-varying coefficients models. When a fixed numbers of knots are used or the number of knots can be assumed to the same for all time-varying coefficients, we do not expect any additional technical challenge in the extension. However, further efforts to reduce the computational burden are needed for selecting different numbers of knots for each covariate.

There are a few other ways in which we can extend our proposal. In this paper, we assume that the observation times of the biomarkers are independent of the level of biomarkers. However, in some cases where biomarkers are observed at hospitalizations or whenever clinicians may suspect some progress of the diseases for example, our JPLM can be extended to accommodate the informative observation process by jointly modeling the additional component. The AIC and BIC were considered to determine both the best transformation and the selection of the number of knots, but we can also explore and compare the validity of other resampling-based model selection criteria such as cross-validation in the future.

Lastly, there is no theoretical justification for any of these model selection procedures, and the proposed inference does not account for the uncertainty of model selection. Post-model selection inference methods under the joint modeling setting are worth pursuing.

7. Supplementary Materials

Web Appendices, Tables and Figures referenced in Sections 1, 3.4 and 4-6, and an R package to implement the JPLM method are available with this paper at the *Biometrics* website on Wiley Online Library.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

References

Beunckens, C., Molenberghs, G., Verbeke, G., and Mallinckrodt, C. (2008). A latent-class mixture model for incomplete longitudinal Gaussian data. *Biometrics* **64,** 96–105.

Cai, N., Lu, W., and Zhang, H. H. (2012). Time-varying latent effect model for longitudinal data with informative observation times. *Biometrics* **68,** 1093–1102.

Dabrowska, D. M. and Doksum, K. A. (1988). Partial likelihood in transformation models with censored data. *Scandinavian Journal of Statistics* **15,** 1–23.

Ding, J. and Wang, J.-L. (2008). Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics* **64,** 546–556.

Geman, S. and Hwang, C.-R. (1982). Nonparametric maximum likelihood estimation by the method of sieves. *Annals of Statistics* **10,** 401–414.

Hogan, J. W. and Laird, N. M. (1997). Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine* **16,** 259–272.

Huang, J. Z., Wu, C. O., and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89,** 111–128.

Kim, S., Zeng, D., Chambless, L., and Li, Y. (2012). Joint models of longitudinal data and recurrent events with informative terminal event. *Statistics in Biosciences* **4,** 262–281.

Kosorok, M. R., Lee, B. L., and Fine, J. P. (2004). Robust inference for univariate proportional hazards frailty regression models. *Annals of Statistics* **32,** 1448–1491.

Lin, D. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association* **96,** 103–126.

Lin, X. and Carroll, R. (2001). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association* **96,** 1045–1056.

Lu, T. and Huang, Y. (2015). Bayesian inference on mixed-effects varying-coefficient joint models with skew-t distribution for longitudinal data with multiple features. *Statistical Methods in Medical Research* page Advance online publication. doi:10.1177/0962280215569294.

Muthén, B., Asparouhov, T., Hunter, A. M., and Leuchter, A. F. (2011). Growth modeling with nonignorable dropout: alternative analyses of the star*d antidepressant trial. *Psychological Methods* **16,** 17–33.

Proust-Lima, C., Taylor, J. M. G., Williams, S. G., Ankerst, D. P., Liu, N., Kestin, L. L., Bae, K., and Sandler, H. M. (2008). Determinants of change in prostate-specific antigen over time and its association with recurrence after external beam radiation therapy for prostate cancer in five large cohorts. *International Journal of Radiation Oncology**

Biology Physics* **72,** 782–791.

Rizopoulos, D., Verbeke, G., and Molenberghs, G. (2010). Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. *Biometrics* **66,** 20–29.

Roy, J. (2003). Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics* **59,** 829–836.

Roy, J. and Lin, X. (2005). Missing covariates in longitudinal data with informative dropouts: Bias analysis and inference. *Biometrics* **61,** 837–846.

Shen, X. and Wong, W. H. (1994). Convergence rate of sieve estimates. *Annals of Statistics* **22,** 580–615.

Shibata, R. (1981). An optimal selection of regression variables. *Biometrika* **68,** 45–54.

Taylor, J. M. G., Park, Y., Ankerst, D. P., Proust-Lima, C., Williams, S., Kestin, L., Bae, K., Pickles, T., and Sandler, H. (2013). Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics* **69,** 206–213.

Zagars, G. K., Pollack, A., Kavadi, V. S., and von Eschenbach, A. C. (1995). Prostate-specific antigen and radiation therapy for clinically localized prostate cancer. *International Journal of Radiation Oncology* Biology* Physics* **32,** 293–306.

Zeger, S. and Diggle, P. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* **50,** 689–699.

Zeng, D. and Cai, J. (2005). Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *Annals of Statistics* **33,** 2132–2163.

Zhang, D., Lin, X., Raz, J., and Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association* **93,** 710–719.

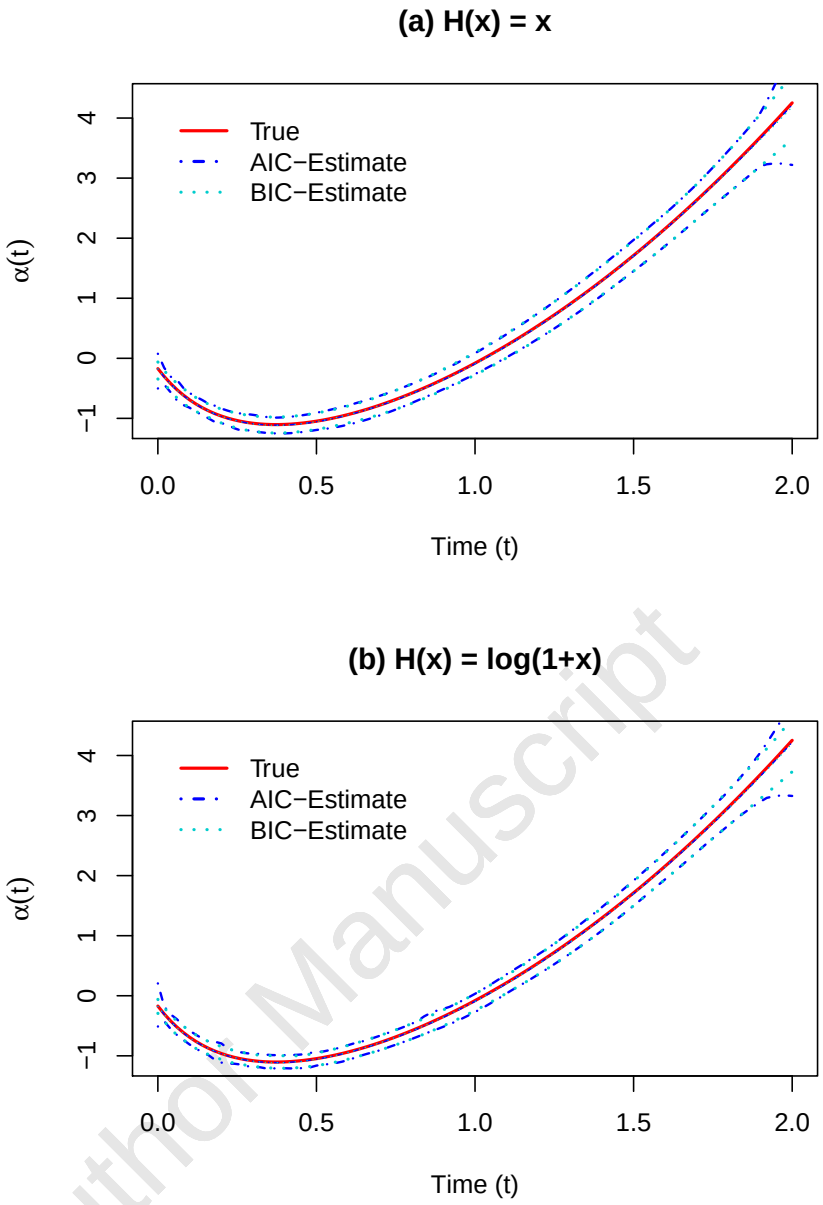

Figure 1. Simulation results for $\alpha(t)$ estimation when $\phi = 1.7$ and (a) $H(x) = x$; and (b) $H(x) = \log(1 + x)$. The solid curve indicates true $\alpha(t) = (t + 0.5)^{-1.5} + (t/2 + 1)^3 - 4$. The dash-dotted (dotted) curves indicate the maximum, average and minimum of the estimated $\alpha(t)$ over 1000 simulated datasets when the AIC-based (BIC-based) knot selection procedure was used. The AIC-estimates appeared to have larger variations in both tails than the BIC-estimates. This figure appears in color in the electronic version of this article.

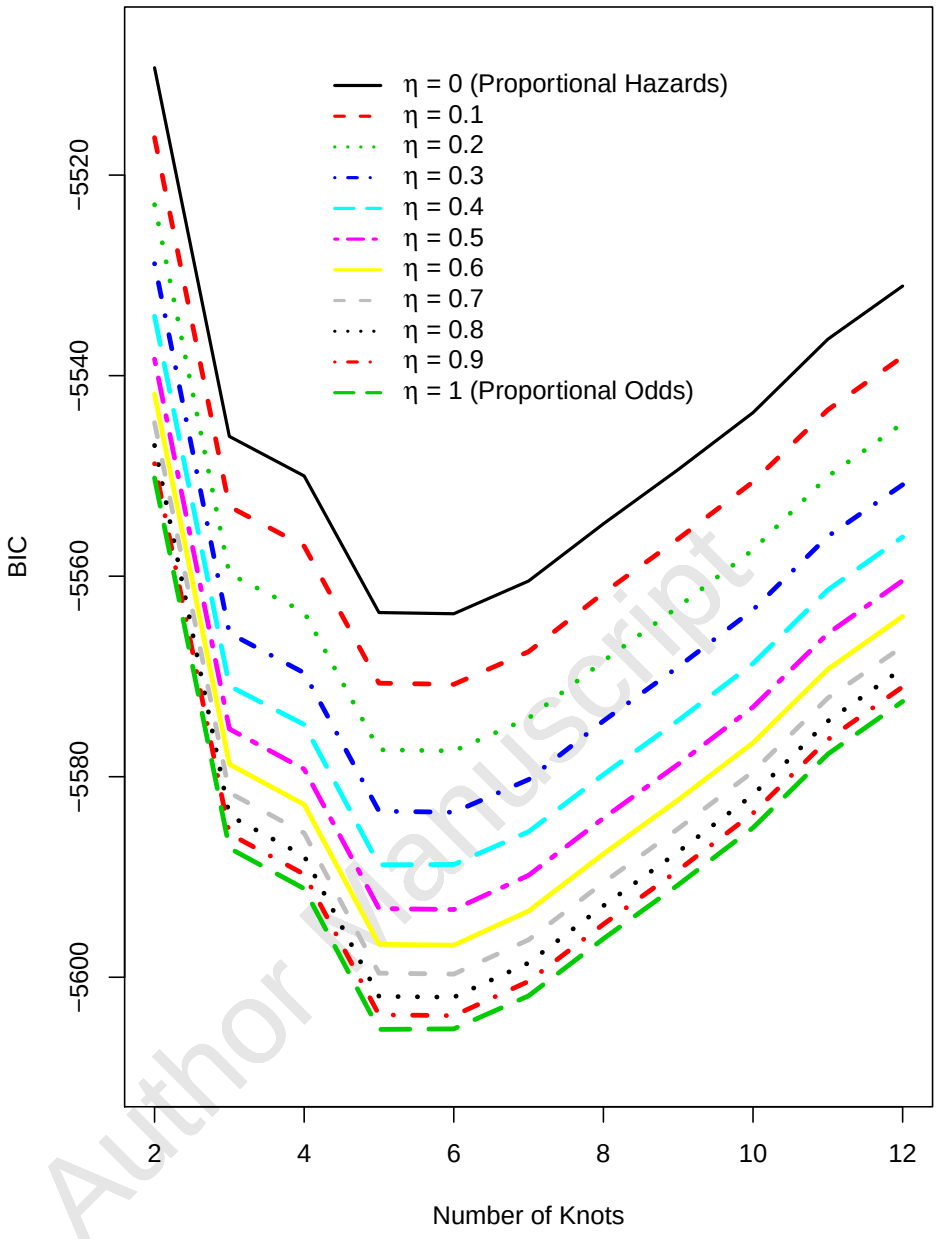

Figure 2. Bayesian information criterion (BIC) plotted for different transformations $H(x) = \log(1 + \eta x)/\eta$ and different numbers of interior knots (K_n).

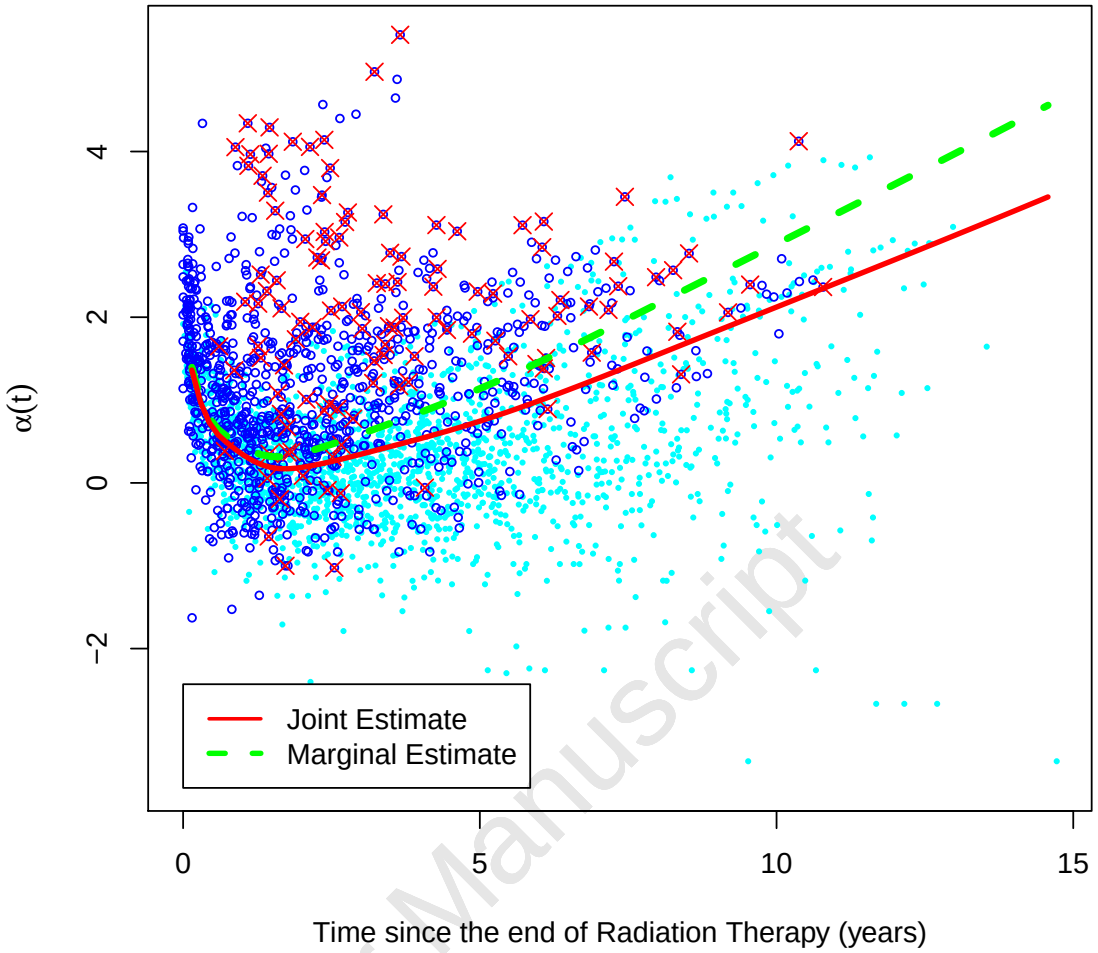

Figure 3. Coefficient function of log PSA score, adjusted by T-stage, gleason score, and age, under the best fit of transformation $H(x) = \log(1 + x)$ and 5 interior points. The solid curve is an estimate from the joint model, and the dashed curve is an estimate from the marginal model. The circles and dots present the full history of all post-radiation PSA values for patients who dropped out informatively and non-informatively, respectively. The mark "×" on some circles indicates the last observation of PSA before the informative drop-out occurred.

Table 1
Simulation results for true $\phi = 1.7$ (i.e., missing not at random) with $K_n \in \{k_{AIC}, k_{BIC}, 2, 4, 8\}$ interior knots of B-spline approximation. True values are $\alpha(\tau_{20}) = -1.10$, $\alpha(\tau_{40}) = -0.58$, and $\alpha(\tau_{80}) = 2.16$, where τ_p represents p% of study duration τ.

		Joint Model					**Marginal Model**			
K_n		*Bias*	*SD*	*SEE*	*CP*	*MSE*	*Bias*	*SD*	*MSE*	*MSER*
						$H(x) = x$				
k_{AIC}	$\alpha(\tau_{20})$	-0.009	0.037	0.035	0.928	0.001	-0.081	0.037	0.008	0.187
	$\alpha(\tau_{40})$	-0.009	0.045	0.044	0.946	0.002	-0.098	0.043	0.011	0.180
	$\alpha(\tau_{80})$	-0.010	0.086	0.082	0.938	0.008	-0.124	0.083	0.022	0.341
k_{BIC}	$\alpha(\tau_{20})$	-0.013	0.035	0.033	0.921	0.001	-0.084	0.035	0.008	0.169
	$\alpha(\tau_{40})$	-0.010	0.043	0.043	0.948	0.002	-0.099	0.041	0.011	0.171
	$\alpha(\tau_{80})$	-0.006	0.084	0.081	0.945	0.007	-0.120	0.081	0.021	0.340
2	$\alpha(\tau_{20})$	-0.013	0.035	0.033	0.922	0.001	-0.085	0.034	0.008	0.166
	$\alpha(\tau_{40})$	-0.010	0.043	0.043	0.946	0.002	-0.099	0.041	0.012	0.169
	$\alpha(\tau_{80})$	-0.006	0.084	0.081	0.945	0.007	-0.120	0.081	0.021	0.340
4	$\alpha(\tau_{20})$	-0.008	0.035	0.034	0.932	0.001	-0.080	0.035	0.008	0.175
	$\alpha(\tau_{40})$	-0.009	0.044	0.043	0.949	0.002	-0.098	0.042	0.011	0.176
	$\alpha(\tau_{80})$	-0.011	0.085	0.082	0.940	0.007	-0.124	0.082	0.022	0.333
8	$\alpha(\tau_{20})$	-0.008	0.039	0.037	0.932	0.002	-0.079	0.038	0.008	0.202
	$\alpha(\tau_{40})$	-0.009	0.046	0.045	0.949	0.002	-0.098	0.044	0.012	0.189
	$\alpha(\tau_{80})$	-0.012	0.088	0.084	0.936	0.008	-0.126	0.084	0.023	0.342
						$H(x) = log(1 + x)$				
k_{AIC}	$\alpha(\tau_{20})$	-0.010	0.035	0.034	0.935	0.001	-0.077	0.034	0.007	0.182
	$\alpha(\tau_{40})$	-0.010	0.041	0.040	0.946	0.002	-0.091	0.040	0.010	0.179
	$\alpha(\tau_{80})$	-0.014	0.070	0.070	0.936	0.005	-0.119	0.068	0.019	0.268
k_{BIC}	$\alpha(\tau_{20})$	-0.014	0.033	0.033	0.932	0.001	-0.081	0.032	0.008	0.166
	$\alpha(\tau_{40})$	-0.008	0.039	0.040	0.953	0.002	-0.089	0.038	0.009	0.167
	$\alpha(\tau_{80})$	-0.012	0.068	0.069	0.943	0.005	-0.117	0.067	0.018	0.264
2	$\alpha(\tau_{20})$	-0.015	0.032	0.033	0.932	0.001	-0.082	0.032	0.008	0.163
	$\alpha(\tau_{40})$	-0.007	0.039	0.040	0.952	0.002	-0.089	0.038	0.009	0.167
	$\alpha(\tau_{80})$	-0.012	0.068	0.069	0.945	0.005	-0.117	0.066	0.018	0.262
4	$\alpha(\tau_{20})$	-0.009	0.033	0.033	0.945	0.001	-0.076	0.033	0.007	0.170
	$\alpha(\tau_{40})$	-0.010	0.039	0.040	0.946	0.002	-0.092	0.038	0.010	0.165
	$\alpha(\tau_{80})$	-0.015	0.068	0.070	0.941	0.005	-0.120	0.067	0.019	0.259
8	$\alpha(\tau_{20})$	-0.009	0.034	0.035	0.946	0.001	-0.076	0.034	0.007	0.178
	$\alpha(\tau_{40})$	-0.011	0.041	0.041	0.946	0.002	-0.092	0.040	0.010	0.176
	$\alpha(\tau_{80})$	-0.018	0.070	0.072	0.950	0.005	-0.122	0.069	0.020	0.266

Table 2

Simulation results by varying ϕ and $H(\cdot)$ when the interior knots of B-spline approximation are selected by BIC (i.e., $K_n = k_{BIC}$). τ_p represents p% of study duration τ.

		$H(x) = x$				$H(x) = \log(1+x)$			
Parameter	*Target*	*Bias*	*SD*	*SEE*	*CP*	*Bias*	*SD*	*SEE*	*CP*
					$\phi = 0.0$				
β	0.5	-0.000	0.031	0.029	0.938	-0.000	0.028	0.029	0.961
σ_e^2	0.1	-0.000	0.003	0.003	0.948	-0.000	0.003	0.003	0.952
σ_{b1}^2	0.4	-0.003	0.028	0.028	0.952	-0.002	0.028	0.028	0.958
σ_{b2}^2	0.2	-0.001	0.030	0.030	0.955	-0.002	0.026	0.027	0.953
ρ	-0.1	0.001	0.087	0.085	0.947	0.003	0.076	0.079	0.962
γ	0.5	0.020	0.065	0.066	0.951	0.020	0.095	0.095	0.947
ϕ	0.0	-0.002	0.106	0.105	0.945	0.001	0.151	0.154	0.958
$\Lambda(\tau_{20})$	0.3	-0.007	0.029	0.029	0.948	-0.011	0.035	0.035	0.943
$\Lambda(\tau_{40})$	0.6	-0.003	0.050	0.050	0.940	-0.006	0.065	0.067	0.954
$\Lambda(\tau_{80})$	1.3	0.023	0.116	0.115	0.950	0.011	0.165	0.165	0.954
					$\phi = 0.5$				
β	0.5	-0.001	0.029	0.030	0.950	-0.001	0.030	0.029	0.939
σ_e^2	0.1	-0.000	0.003	0.003	0.951	-0.000	0.003	0.003	0.939
σ_{b1}^2	0.4	-0.002	0.028	0.028	0.957	-0.002	0.028	0.028	0.947
σ_{b2}^2	0.2	-0.001	0.031	0.030	0.944	-0.001	0.027	0.027	0.950
ρ	-0.1	0.001	0.089	0.086	0.944	0.002	0.082	0.080	0.948
γ	0.5	0.021	0.068	0.068	0.947	0.023	0.099	0.097	0.955
ϕ	0.5	0.019	0.108	0.108	0.945	0.014	0.159	0.158	0.955
$\Lambda(\tau_{20})$	0.3	-0.008	0.031	0.030	0.936	-0.011	0.034	0.036	0.950
$\Lambda(\tau_{40})$	0.6	-0.005	0.052	0.051	0.945	-0.009	0.066	0.068	0.958
$\Lambda(\tau_{80})$	1.3	0.023	0.124	0.119	0.942	0.005	0.161	0.167	0.964
					$\phi = 1.7$				
β	0.5	0.000	0.030	0.030	0.947	-0.000	0.030	0.029	0.945
σ_e^2	0.1	-0.000	0.003	0.003	0.943	-0.000	0.003	0.003	0.948
σ_{b1}^2	0.4	-0.001	0.030	0.029	0.940	-0.001	0.029	0.029	0.947
σ_{b2}^2	0.2	0.000	0.029	0.030	0.958	-0.002	0.027	0.027	0.958
ρ	-0.1	-0.002	0.101	0.101	0.951	0.001	0.088	0.087	0.943
γ	0.5	0.026	0.087	0.086	0.941	0.026	0.112	0.111	0.947
ϕ	1.7	0.091	0.152	0.149	0.920	0.085	0.192	0.195	0.937
$\Lambda(\tau_{20})$	0.3	-0.015	0.036	0.035	0.936	-0.015	0.040	0.041	0.946
$\Lambda(\tau_{40})$	0.6	-0.009	0.067	0.065	0.942	-0.014	0.078	0.080	0.951
$\Lambda(\tau_{80})$	1.3	0.025	0.151	0.153	0.958	-0.000	0.190	0.191	0.952

Table 3

Joint analysis results of the prostate cancer data under the best fit of transformation $H(x) = \log(1 + x)$ and 5 interior knots selected by BIC. The 50:50 mixture of χ^2 distributions is used for testing variances. Reference groups for categorical covariates are T-stage=1, Gleason score between 2 and 6, and age<65. τ_p represents p% of study duration τ.

	Joint Model			**Marginal Model**		
Effect	*Est*	*SE*	*p-value*	*Est*	*SE*	*p-value*
			Longitudinal PSA score			
$\alpha(\tau_{20})$	0.340	0.087	.0001	0.553	0.092	$< .0001$
$\alpha(\tau_{40})$	0.963	0.135	$< .0001$	1.368	0.141	$< .0001$
$\alpha(\tau_{60})$	1.792	0.195	$< .0001$	2.388	0.201	$< .0001$
$\alpha(\tau_{80})$	2.637	0.262	$< .0001$	3.424	0.269	$< .0001$
log(baseline PSA+0.1)	0.510	0.034	$< .0001$	0.509	0.034	$< .0001$
T-stage 2	-0.058	0.064	.3617	-0.045	0.064	.4828
T-stage 3 or 4	0.018	0.117	.8770	0.027	0.117	.8171
Gleason score 7 to 9	0.055	0.060	.3592	0.040	0.060	.4942
Age 65-75 years	-0.190	0.069	.0056	-0.201	0.069	.0037
Age > 75 years	-0.115	0.088	.1882	-0.130	0.088	.1408
σ_e^2	0.117	0.003	$< .0001$	0.118	0.003	$< .0001$
			Random effects			
$\sigma_{b_1}^2$	0.373	0.028	$< .0001$	0.371	0.027	$< .0001$
$\sigma_{b_2}^2$	0.206	0.018	$< .0001$	0.195	0.017	$< .0001$
ρ	-0.121	0.055	.0281	-0.114	0.054	.0359
			Informative drop-out			
log(baseline PSA+0.1)	0.532	0.161	.0010			
T-stage 2	1.193	0.320	.0002			
T-stage 3 or 4	1.815	0.467	.0001			
Gleason score 7 to 9	1.089	0.293	.0002			
Age 65-75 years	-1.180	0.329	.0003			
Age > 75 years	-1.543	0.471	.0011			
ϕ	1.730	0.134	$< .0001$			