A Bayesian Semi-parametric Survival Model with Longitudinal Markers

Song Zhang, Peter Müller and Kim-Anh Do

Song.Zhang@utsouthwestern.edu

Abstract

We consider inference for data from a clinical trial of treatments for metastatic prostate cancer. Patients joined the trial with diverse prior treatment histories. The resulting heterogenuous patient population gives rise to challenging statistical inference problems when trying to predict time to progression on different treatment arms. Inference is further complicated by the need to include a longitudinal marker as a covariate. To address these challenges, we develop a semi-parametric model for the joint inference on longitudinal data and an event time. The proposed approach includes the possibility of cure for some patients. The event time distribution is based on a non-parametric Pólya tree prior. For the longitudinal data we assume a mixed effects model. Incorporating a regression on covariates in a non-parametric event time model in general, and for a Pólya tree model in particular, is a challenging problem. We exploit the fact that the covariate itself is a random variable. We achieve an implementation of the desired regression by factoring the joint model for the event time and the longitudinal outcome into a marginal model for the event time and a regression of the longitudinal outcomes on the event time, i.e., we implicitly model the desired regression by modeling the reverse conditional distribution.

KEYWORDS: Bayesian non-parametric models, Pólya tree, survival, regression

1 Introduction

We discuss inference for data from a phase III clinical trial for treatments of metastatic prostate cancer. The challenging features are patient heterogeneity due to prior treatment history and the need to include a regression on prostate specific antigen (PSA) as an important longitudinal marker. We conduct semi-parametric Bayesian inference to address these challenges.

To achieve the desired data analysis we develop joint inference for event time data and longitudinal observations of a covariate, with the possibility that some patients are cured. Let T be the event time and Y be the longitudinal covariate. Most existing approaches are based on factoring the joint model as $P(T, Y) = P(Y)P(T \mid Y)$. The first factor is the longitudinal submodel $P(Y)$, typically assumed to be a mixed model. The second factor is the survival submodel $P(T \mid Y)$. In the following discussion, we use event time, survival time, time to progression and failure time exchangably.

There is an extensive literature on the joint modeling of longtitudinal and event time data without cured fraction (De Gruttola and Tu, 1994; Tsiatis et al., 1995; Lavalley and De Gruttola, 1996; Wulfsohn and Tsiatis, 1997; Dafni and Tsiatis, 1998; Henderson et al., 2000; Xu and Zeger, 2001; Lin et al., 2002; Ibrahim et al., 2004). A review can be found in Tsiatis and Davidian (2004). Less work has been published on the joint modeling of longitudinal and event time data with cure. Law et al. (2002) proposed a model with the longitudinal process described by an exponential-decay-exponential-growth model and a mixture model to accomodate cure. The imputed values of the longitudinal measurements are covariates in a proportional hazard model. Brown and Ibrahim (2003) and Chen et al. (2004) implemented inference with alternative cure models. The former assumed that the trajectory of longitudinal process affects the hazard function of MCT cells' progression times. The latter assumed that the longitudinal process affects the mean of the Poisson distribution. Yu et al. (2004) provide a recent review of joint longitudinal-survival-cure models.

Specific to modeling PSA, Carter et al. (1992) demonstrated that the use of PSA readings over time leads to more accurate diagnoses. Lin et al. (2002) considered a latent class model to uncover subpopulation structure for both PSA trajectories and a survival outcome. An important feature is that given latent class membership, the longitudinal marker and outcome are assumed independent. A semi-parametric frailty model is assumed, which includes class-specific baseline hazard functions and accommodates possibly time-dependent covariates.

In the joint analysis of longitudinal and event time data, most researchers assume parametric or semi-parametric models for $P(T \mid Y)$. It is difficult to implement non-parametric models for $P(T \mid Y)$ because most non-parametric models do not allow straightforward incorporation of a regression on covariates. We propose to use the alternative factorization, $P(T, Y) = P(T)P(Y \mid T)$. We proceed under the Bayesian paradigm. Choosing a non-parametric model for $P(T)$ is the traditional problem of non-parametric inference for an event time. As a scaler, event time T can be included as a covariate into a parametric $P(Y \mid T)$ with great ease and flexibility. Furthermore, it is straightforward to apply non-parametric Bayesian models for $P(T)$, such as a Dirichlet process prior or a Pólya tree (PT) prior. For $P(Y \mid T)$, we propose a mixed effects model as a default choice. Both factorizations leads to the joint model, $P(T, Y)$, describing the mutual dependence between T and Y. It is this joint model that ultimately allows improved prediction of the event given repeated measurements of the marker. Pawitan and Self (1993) jointly models event time process and longitudinal marker under the framework of $P(T, Y) = P(T)P(Y \mid T)$. Weibull models are assumed for the infection time and disease occurrence time of AIDS. A generalized linear model is specified for the longitudinal measurements of T4 counts and T4/T8 ratio, with the intercept and slope being functions of the event times. Maximum likelihood estimates are obtained for the parameters.

We use a PT prior to model the failure time process, which can be constructed to give probability one to the set of continuous or absolutely continuous probability measures (Lavine, 1992; Mauldin et al., 1992). Muliere and Walker (1997) implemented PT models in

a survival analysis. Walker and Mallick (1997; 1999) demonstrated the application of PT in hierarchical generalized linear models, frailty models, and accelerated failure time models. Hanson and Johnson (2002) developed a general approach to model residual distributions with a mixture of PT. Neath (2003) used PT distributions to model censored data. Paddock et al. (2003) developed randomized PT models, which uses random partitions to smooth out the effect of partitions on posterior inference. See Hanson (2006) for a review of recent development in finite PT models.

Hanson et al. (2007) used mixtures of PTs to construct a joint model for time dependent covariates and survival time. They introduced flexible PT priors for the baseline distributions in the Cox model, the proportional odds model, and an accelerated failure time model accommodating time-dependent covariates. Their approach uses the factorization $P(T, Y) = P(Y)P(T \mid Y)$.

2 A Clinical Study

Androgen ablation (AA) is the preferred treatment for metastatic prostate cancer. AA therapy alters the natural history of the disease by disrupting the growth promoting effects mediated by androgen receptor signaling, which is usually accomplished by medical suppression of testicular endocrine function. Unfortunately, most patients with clinically detectable metastatic disease when the AA therapy started will eventually progress to androgen independent prostate cancer (AIPC). AIPC is a relentlessly progressive disease state, and is the cause of death for the vast majority of men in whom it develops. By this mechanism, prostate cancer leads to an annual death toll of more than 27,000 men in the United States.

To date, no treatment has been found to be curative for AIPC, and it is only fairly recently that some therapies are shown to alter the natural history of the disease. A chemotherapy demonstrated a survival advantage over historical results in a phase II trial conducted at M.D. Anderson Cancer Center (Ellerhorst et al., 1997). This therapy, dubbed KA/VE, treats

patients with ketoconazole and doxorubicin alternating with vinblastine and estramustine.

Then a phase III trial of conventional AA therapy versus AA therapy plus three 8-week cycles of KA/VE was conducted at M.D. Anderson Cancer Center. The aim of this trial was to investigate whether better clinical benefit can be achieved by applying the chemotherapy "early", i.e., before the metastatic prostate cancer develops into the far-advanced AIPC. The two treatment arms are denoted by AA and CH, respectively. The patient population includes metastatic prostate cancer patients whose high risk of developing AIPC justifies long-term, sustained, androgen ablation. The primary endpoint is the time to progression (TTP) to AIPC, which is diagnosed by the following criteria: 1) Symptoms attributed by the treating physician to reflect progressive cancer; 2) Radiographic progression; 3) Rising PSA, with value greater than 1 and doubling time < 9 months; 4) Treatment with chemotherapy. The first 3 also require demonstration of testosterone < 50 and withdrawal of antiandrogens. More details about the clinical trial can be found in Milikan et al. (2007).

Besides the TTP, we also observed the longitudinal measurements of prostate specific antigen (PSA) level from each patient. Carter et al. (2006) demonstrated that PSA velocity is associated with prostate cancer death even 10-15 years before diagnosis. To further improve the understanding of this important marker we propose to build a joint model of the TTP and the PSA measurements.

To statisticians, a challenge posed by this clinical trial is that considerable heterogeneity exists among the patients. Before coming to M.D. Anderson Cancer Center to seek treatment for the metastatic prostate cancer, these patients had been treated by different physicians with different therapies at different institutions. These differences might have a long-term impact on the development of prostate cancer. Second, there is no completely satisfactory way to define "early" in the natural history of metastatic prostate cancer. As a practical solution, the clock start of the trial is defined as the initiation of the AA therapy. Thus at the beginning of the trial, the true stage of cancer might not be exactly the same for each patient.

3 Notation and Model

We use $v = 1, 2$ to denote the two treatment arms (1 for CH and 2 for AA). Let n_v be the number of subjects in each arm. For the i-th subject on arm v, we use $\boldsymbol{y}_{vi} = \{y_{vij}, j = 1, \cdots, m_{vi}\}$ to denote the longitudinal measurements, where m_{vi} is the total number of repeated measurements for patient i under treatment v. We define T_{vi} to be the TTP, which is the time between the start of the CH/AA treatment and the progression to AIPC. We use t_{vi} to denote the censoring time for censored observations, and the actual TTP for non-censored observations. We introduce a failure indicator d_{vi} with $d_{vi} = 1$ if $T_{vi} = t_{vi}$ and $d_{vi} = 0$ if $T_{vi} > t_{vi}$. The number of observed and unobserved TTP in each arm are denoted by n_{v1} and n_{v0} respectively. In summary, the observed data from each subject is $(\boldsymbol{y}_{vi}, t_{vi}, d_{vi})$. We use $[X]$ and $[X \mid Y]$ to generically indicate the probability model for a random variable X and the conditional distribution of X given Y.

3.1 The Likelihood

We define the sampling model for the observed data $(\boldsymbol{y}_{vi}, t_{vi}, d_{vi})$ from each patient. If $d_{vi} = 1$, the progression time T_{vi} is observed. Therefore all subjects with $d_{vi} = 1$ belong to the susceptible group. On the other hand, we only observe $T_{vi} > t_{vi}$ when $d_{vi} = 0$. In this case the subject could be either in the susceptible group or in the cure group. We define a variable $\omega_{vi} = 0/1$ indicating membership in the cure/susceptible group. For $d_{vi} = 1$, we have $\omega_{vi} \equiv 0$ by definition, and $T_{vi} = t_{vi}$. The following discussion simplifies greatly by introducing latent variables T_{vi} and ω_{vi} for subjects with censored TTP, $d_{vi} = 0$.

If $\omega_{vi} = 0$, the subject is at risk of developing the endpoint event. We assume T_{vi} to be a random sample from a distribution G_v, i.e., $[T_{vi} \mid \omega_{vi} = 0, G_v] = g_v(T_{vi})$. Here $g_v(\cdot)$ is the density function of G_v. If $\omega_{vi} = 1$, the subject is a long-term survivor. We assume $T_{vi} = t_c$, where t_c is an extremely long TTP that could not be observed in the clinical trial. Thus the model for TTP is $[T_{vi} \mid \omega_{vi}, G_v] = \{\delta(t_c)\}^{\omega_{vi}}\{g_v(T_{vi})\}^{1-\omega_{vi}}$, where $\delta(t_c)$ denotes a point

mass at t_c. We assume $\omega_{vi} \mid p_v \overset{iid}{\sim} \text{Bernoulli}(p_v)$ with p_v being the probability of cure in treatment arm v. This model assumes that T_{vi} arises from the mixture of a point mass and a continuous distribution. We interpret G_v as the therapy-specific marginal distribution of TTP in the susceptible group.

Given T_{vi}, the longitudinal measurements $\boldsymbol{y}_{vi}$ are assumed to arise from a mixed effects model $[\boldsymbol{y}_{vi} \mid T_{vi}, \boldsymbol{\Psi}]$, indexed by parameters $\boldsymbol{\Psi}$. We include T_{vi} in $[\boldsymbol{y}_{vi} \mid T_{vi}, \boldsymbol{\Psi}]$ via a regression on $u(T_{vi})$, as a subject-specific covariate. Here $u(\cdot)$ is some function of T_{vi}. With different specifications of $u(T_{vi})$ and $[\boldsymbol{y}_{vi} \mid T_{vi}, \boldsymbol{\Psi}]$, we can model various effects of progression time T_{vi} on longitudinal profile $\boldsymbol{y}_{vi}$.

In summary, the likelihood factors corresponding to $(\boldsymbol{y}_{vi}, t_{vi}, d_{vi})$ are

$$\begin{aligned} L_{vi1} &= [\boldsymbol{y}_{vi} \mid T_{vi}, \boldsymbol{\Psi}][T_{vi} = t_{vi} \mid \omega_{vi} = 0, G_v][\omega_{vi} = 0 \mid p_v], \qquad \text{for } d_{vi} = 1, \\ L_{vi0} &= [\boldsymbol{y}_{vi} \mid T_{vi}, \boldsymbol{\Psi}] I(T_{vi} > t_{vi})[T_{vi} \mid \omega_{vi}, G_v][\omega_{vi} \mid p_v], \qquad \text{for } d_{vi} = 0. \end{aligned} \tag{1}$$

Note that L_{vi0} is an augmented likelihood with latent variable T_{vi} and ω_{vi}.

For L_{vi0}, the two values taken by ω_{vi} lead to two models of different dimensions. If $\omega_{vi} = 0$, we have a model with T_{vi} being a random parameter. In contrast, T_{vi} is fixed at $T_{vi} = t_c$ if $\omega_{vi} = 1$. Such a change in dimension complicates posterior simulation (Green, 1995). We use the pseudo prior approach by Carlin and Chib (1995) to avoid this complication. In words, we augment the smaller probability model under $\omega_{vi} = 1$ by defining a prior probability model for a hypothetical T_{vi} (but keep t_c in the regression for $\boldsymbol{y}_{vi}$). The new variable T_{vi} has no meaningful interpretation under $\omega_{vi} = 1$. It is only introduced to match the model dimensions. See Zhang et al. (2008) for details of the pseudo prior choice.

3.2 The Prior Probability Model

We assume prior independence, $[\boldsymbol{\Psi}, p_v, G_v, \ v = 1, 2] = [\boldsymbol{\Psi}] \cdot \prod_{v=1}^{2} [p_v] \cdot [G_v]$. The prior specification $[\boldsymbol{\Psi}]$ and posterior inference for mixed models of repeated measurements have been discussed extensively. See, for example, Ibrahim et al. (2004) and Guo and Carlin

(2004). For priors $[p_v]$, $v = 1, 2$, we assume $p_v \sim Beta(a_p, b_p)$ with a_p and b_p being fixed hyperparameters.

For the unknown survival distribution G_v, we consider two choices. The first choice is a parametric model, which assumes G_v to be indexed by a finite-dimensional parameter vector. In this case the prior specification only involves assigning prior distributions to these parameters. Recall that G_v is the therapy-specific marginal distribution of TTP for the susceptible groups, which contains patients that are different in many important aspects. A parametric model may not suffice to characterize the complexity of G_v. This difficulty motivates the second choice, adopting a non-parametric method. A Bayesian non-parametric prior defines G_v as a random probability measure, i.e., we assume a distribution for the unknown distribution G_v. Specifically, we assume G_v to have a PT prior, denoted by $PT(\Pi_v, \mathcal{A}_v)$. Here Π_v and $\mathcal{A}_v$ are hyper-parameters of the PT prior. See, for example, Hanson (2006) for a recent review of PT priors. We define the parameters $(\mathcal{A}_v, \Pi_v)$ such that $E(G) = \tilde{G}_v$ for a given probability model $\tilde{G}_v$. See Zhang et al. (2008) for a details. Note that G_v is not subject-specific but the posterior TTP distribution given PSA measurements may have very different forms for each subject.

3.3 Posterior Inference and Model Validation

To facilitate discussion, we define the following notation. The set of observed and unobserved TTPs under treatment v are denoted by $\boldsymbol{t}_v^1 = \{t_{vi} : d_{vi} = 1\}$ and $\boldsymbol{T}_v^0 = \{T_{vi} : d_{vi} = 0\}$, respectively. We also define $\boldsymbol{\omega}_v^0 = \{\omega_{vi} : d_{vi} = 0\}$ to be the set of unknown indicators of cure. Without loss of generality, we assume that $d_{vi} = 0$ for $i = 1, \cdots, n_{v0}$, and $d_{vi} = 1$ for $i = n_{v0}+1, \cdots, n_v$. Let $\boldsymbol{\Lambda}$ denote the collection of model parameters, including $\boldsymbol{\Psi}, G_v, \boldsymbol{T}_v^0, \boldsymbol{\omega}_v^0$ and p_v. We have the full posterior distribution:

$$[\boldsymbol{\Lambda} \mid \boldsymbol{Y}, \boldsymbol{t}, \boldsymbol{d}] \propto \prod_{v=1}^{2} \left\{ \left(\prod_{i=1}^{n_{v0}} L_{vi0} \prod_{i=n_{v0}+1}^{n_v} L_{vi1} \right) [p_v][G_v] \right\} [\boldsymbol{\Psi}], \tag{2}$$

where $(\boldsymbol{Y}, \boldsymbol{t}, \boldsymbol{d}) = \{(\boldsymbol{y}_{vi}, t_{vi}, d_{vi}) : v = 1, 2; i = 1, \cdots, n_v\}$. We implement posterior inference by Markov Chain Monte Carlo (MCMC) posterior simulation.

Before proceeding with posterior MCMC, we analytically marginalize (2) with respect to G_v. Recall that each subject with $\omega_{vi} = 0$ is assumed to have a TTP arising from G_v. Define $\boldsymbol{T}_v^s = \boldsymbol{t}_v^1 \bigcup \{T_{vi};\ d_{vi} = \omega_{vi} = 0\}$ to be the set of observed and unobserved TTP in the susceptible group. The size of $\boldsymbol{T}_v^s$ is $n_{vs} = n_v - \sum_{i=1}^{n_v} \omega_{vi}$. Finally, let T_{vj}^s denote the j-th element in $\boldsymbol{T}_v^s$, with the index j assigned arbitrarily. The joint probability model of $\boldsymbol{T}_v^s$ and G_v is $[\boldsymbol{T}_v^s, G_v] = \prod_{j=1}^{n_{vs}} [T_{vj}^s \mid G_v] \cdot [G_v]$. We marginalize G_v by replacing $[\boldsymbol{T}_v^s, G_v]$ with $[\boldsymbol{T}_v^s] = \prod_{j=2}^{n_{vs}} [T_{vj}^s \mid T_{v1}^s, \ldots, T_{v,j-1}^s] \cdot \tilde{G}_v(T_{v1}^s)$, where $[T_{vj}^s \mid T_{v1}^s, \cdots, T_{v,j-1}^s]$ is defined in Expression (5) of Zhang et al. (2008). The marginalization is important. Instead of working with the infinite dimensional random distributions G_v, it allows us to manipulate only the (finite dimensional) set of event times $\boldsymbol{T}_v^s$. Details of the MCMC transition probabilities are in Zhang et al. (2008).

We compare the proposed model with four natural alternatives. Details of the competing models and results are described later, in Section 5. We use the conditional predictive ordinates (CPO) proposed by Gelfand et al. (1992) to compare different models. The CPO for subject i in group v (henceforth subject (v, i)) is defined as the posterior predictive distribution evaluated for the observation from subject (v, i), conditional on all the data minus the response from subject (v, i). Formally, letting $(\boldsymbol{Y}_{(-vi)}, \boldsymbol{t}_{(-vi)}, \boldsymbol{d}_{(-vi)}) = (\boldsymbol{Y}, \boldsymbol{t}, \boldsymbol{d}) \setminus (\boldsymbol{y}_{vi}, t_{vi}, d_{vi})$, we define $\text{CPO}_{vi} = [\boldsymbol{y}_{vi}, t_{vi}, d_{vi} \mid \boldsymbol{Y}_{(-vi)}, \boldsymbol{t}_{(-vi)}, \boldsymbol{d}_{(-vi)}]$. Then we compute a summary statistic called the logarithm of the pseudomarginal likelihood (LPML), $LPML = \sum_{v=1}^{2} \sum_{i=1}^{n_v} \log(CPO_{vi})$. A small value of $LPML$ suggests disagreement between the observations and the model. Gelfand et al. (1992) show how the CPO for each subject, in our case $v = 1, 2$ and $i = 1, \ldots, n_v$, can be evaluated through an importance sampling scheme. We describe the computation of CPO in Zhang et al. (2008). We validate the survival and cure aspect of the model based on subject specific martingale residuals (Barlow and Prentice, 1988; Therneau et al., 1990; Lin et al., 2002).

4 A Phase III Study of Prostate Cancer

We return to the clinical trial from Section 2. The phase III trial for advanced prostate cancer had a total enrollment of 286 patients, with $n_1 = 137$ in the CH arm and $n_2 = 149$ in the AA arm. Starting from the diagnosis of prostate cancer, the PSA level of each patient was monitored for up to 10 years. On average, about 30 PSA measurements were collected from each patient. We use y_{vij} $(j = 1, \cdots, m_{vi})$ to denote the log-transformed longitudinal PSA measurement, $y_{vij} = \log(1 + PSA)$. The age at which y_{vij} was recorded is denoted by s_{vij}. As reference points, the age at diagnosis of prostate cancer is denoted by u_{vi0}, and the age at the initiation of the CH/AA treatment is denoted by u_{vi1}. The number of observed TTP events in the two treatment arms are $n_{11} = 87$ and $n_{21} = 98$, respectively. Figure 1 shows the Kaplan-Meier estimates of the survival function under the two treatments. There are plateaus at the end of the curves. This observation suggests that a significant portion of subjects might have an excessively long event time and a cure model is appropriate.

PSA level normally increases as the prostate enlarges with age. When prostate cancer develops, however, it increases much faster. The typical effect of a treatment on PSA level is a sharp drop in PSA level immediately after the treatment. Then gradually, the body adjusts to offset the treatment effect, and the PSA level bounces back. The speed of rebound depends on the progress of cancer. Figure 2 plots the longitudinal profiles of four randomly selected patients. Note the variability among the profiles. Exploratory analysis indicates a negative correlation between the PSA slope and TTP. Based on these considerations, the longitudinal submodel $[\boldsymbol{y}_{vi} \mid T_{vi}, \boldsymbol{\Psi}]$ is specified as $y_{vij} = f_{vi}(s_{vij}) + e_{vij}$ with

$$f_{vij}(s_{vij}) = \theta_{0vi} + \theta_{1vi} s_{vij} + \gamma_{2vi} \left(e^{-\phi_{0vi}(s_{vij} - u_{vi0})^+} - 1 \right) + \eta_v \left(e^{-\phi_{1v}(s_{vij} - u_{vi1})^+} - 1 \right) + \gamma_{1v}(s_{vij} - u_{vi1})^+ + (\theta_{1vi} + \gamma_{1v})(e^{-\xi_v T_{vi}} - 1)(s_{vij} - u_{vi1})^+, \quad (3)$$

where $(x)^+ = x$ if $x > 0$, and $(x)^+ = 0$ otherwise. We assume independent normal residuals, $e_{vij} \overset{iid}{\sim} N(0, \sigma^2)$. The first two terms define a line with intercept θ_{0vi} and slope θ_{1vi}, describing

the baseline linear trend of PSA over age. The coefficients are subject-specific. Parameters η_v and ϕ_{1v} model the size and the slope of the drop after the intervention with CH or AA. As age s_{vij} moves beyond u_{vi1}, $\eta_v\{\exp[-\phi_{1v}(s_{vij}-u_{vi1})^+]-1\}$ drops from 0 and eventually levels off at $-\eta_v$, i.e., η_v controls the depth and ϕ_{1v} controls the slope of the drop. A smaller value of ϕ_{1v} indicates that the treatment effect persists longer. Similarly, parameters γ_{2vi} and ϕ_{0vi} model the size and the slope of the drop due to the initial therapy right after the diagnosis of prostate cancer. Since we have no information about the initial therapy, we assume γ_{2vi} and ϕ_{0vi} to vary individually. We use γ_{1v} to model the average change of slope in the baseline trend, induced by treatment v. Finally, model (3) reflects our belief that subjects with flatter longitudinal profiles take longer to progress. To see this, first we observe that in (3) the slope after treatment, i.e., the coefficient of $(s_{vij}-u_{vi1})^+$, is $\theta_{1vi}+\gamma_{1v}+(\theta_{1vi}+\gamma_{1v})[\exp(-\xi_v T_{vi})-1]$. Here ξ_v is constrained to be positive. With T_{vi} changing between 0 and $+\infty$, the slope changes from $\theta_{1vi}+\gamma_{1v}$ to 0. We substitute a realistic upper bound for the limiting $T_{vi}\to\infty$ using $t_c=18$ years. Matching with the earlier notation $[\boldsymbol{y}_{vi}\mid T_{vi},\boldsymbol{\Psi}]$ used in (2), we have $\boldsymbol{\Psi}=(\boldsymbol{\theta}_0,\boldsymbol{\theta}_1,\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\boldsymbol{\eta},\boldsymbol{\phi}_0,\boldsymbol{\phi}_1,\boldsymbol{\xi},\sigma^2)$. Here $\boldsymbol{\theta}_0=\{\theta_{0vi},v=1,2;i=1,\cdots,n_v\}$, $\boldsymbol{\eta}=\{\eta_v,v=1,2\}$, and $\boldsymbol{\theta}_1$,$\boldsymbol{\gamma}_2$, $\boldsymbol{\phi}_0$, $\boldsymbol{\phi}_1$, $\boldsymbol{\xi}$ are defined in the same fashion. In (3) we use $u(T_{vi})=\exp(-\xi_v T_{vi})-1$. In summary, besides TTP, the covariates considered include age, treatment, and time under treatment. The dotted lines in Figure 2 show the fitted values of the PSA profiles.

As for the PT priors, $G_v\sim PT(\Pi_v,\mathcal{A}_v)$, we use $\Pi_1=\Pi_2=\Pi$ and $\mathcal{A}_1=\mathcal{A}_2=\mathcal{A}$. Thus $E(G_v)=\tilde{G}$ for $v=1,2$. That is, the two PTs are centered around the same distribution a priori. The matching hyperprior parameters for the two PT priors ensures that posterior inference about differences between the two treatment groups reflects the evidence from data. For the centering measure $\tilde{G}$, we assume a Weibull distribution, $\tilde{G}(t)=\text{Weibull}(t;\tau,\beta)$. Here β and τ are, respectively, the shape and scale parameter. The partition Π is specified by the dyadic quantile sets of $\tilde{G}$. The elements of $\mathcal{A}$ at the mth level are specified to be $c\cdot m^2$, with c being a constant.

The mixture probability p_v is assumed to be $Unif(0,1)$, i.e., $a_p = b_p = 1$. The prior of $\boldsymbol{\Psi}$ and other hyperprior parameters are specified as follows. We assume $(\theta_{0vi}, \theta_{1vi}, \gamma_{2vi})' \mid \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\gamma_{1v} \sim N(0, 100)$, $\eta_v \sim N(0, 100)$, $\xi_v \sim \text{Ga}(a, b)$, $\phi_{0vi} \sim \text{Ga}(a, b)$, $\phi_{1v} \sim \text{Ga}(a, b)$, and $1/\sigma^2 \sim \text{Ga}(a, b)$, all with $a = b = 0.01$. The parameterization of $\text{Ga}(a, b)$ is such that the mean is a/b. We further assume $\boldsymbol{\mu} \sim N_3(\mathbf{0}, 100\boldsymbol{I})$ and $\boldsymbol{\Sigma} \sim IW(3, 0.01\boldsymbol{I}_3)$. Here IW indicates an inverse Wishart prior. The specification of hyper-parameter (τ, β) for $\tilde{G}$ is based on estimation of the Weibull model M_2, described in Section 5. We set $\tau = 4.52$ and $\beta = 1.23$, which are the posterior means of Weibull parameters from the CH group.

5 Results

Model Selection. To validate the proposed model we consider comparisons with four alternative models. Let M_1 denote the proposed model (2). The second model, M_2, is also based on the factorization $P(T, Y) = P(T)P(Y \mid T)$, with $P(Y \mid T)$ being the same as (3). The survival submodel $P(T)$, however, is fully parametric. We assume a Weibull regression model for $(T_{vi} \mid \omega_{vi} = 0)$ with an indicator of treatment being the covariate. The third model, M_3, assumes no cure group. It is obtained from model (2) by setting $\omega_{vi} = 0$ for all patients. The last two models, M_4 and M_5, are constructed under the factorization $P(T, Y) = P(Y)P(T \mid Y)$, where the longitudinal submodel $P(Y)$ is specified as $y_{vij} = f_{vi}(s_{vij}) + e_{vij}$ with

$$f_{vi}(s_{vij}) = \theta_{0vi} + \theta_{1vi}s_{vij} + \gamma_{2vi}\{e^{-\phi_{0vi}(s_{vij}-u_{vi0})^+-1}\}+ \\ \eta_v\{e^{-\phi_{1v}(s_{vij}-u_{vi1})^+-1}\} + \gamma_{1vi}(s_{vij} - u_{vi1})^+$$

and $e_{vij} \sim N(0, \sigma^2)$. The survival submodel $P(T \mid Y)$ is assumed to be a proportional hazard model with a cure fraction p_v. The mean longitudinal process, $f_{vi}(s_{vij})$, together with the PSA slope, $f'_{vi}(s_{vi}) = \partial f_{vi}(s_{vi})/\partial s_{vi}$, are included as time-dependent covariates (Yu et al.,

2004). We assume the following hazard function,

$$h_{vi}(t) = h_{v0}(t) \exp[\zeta_{1v} f_{vi}(u_{vi1} + t) + \zeta_{2v} f'_{vi}(u_{vi1} + t)], \tag{4}$$

where $h_{v0}(t)$ is the baseline hazard and (ζ_{1v}, ζ_{2v}) are scaling parameters. Under M_4, we model $h_{v0}(t)$ by a piecewise constant function. For $0 < q_1 < q_2 < \cdots < q_{J-1} < \infty$, we assume $h_{v0}(t) = \kappa_{v1}$ if $t \leq q_1$, $h_{v0}(t) = \kappa_{v2}$ if $q_1 < t \leq q_2$, $\cdots$, and $h_{v0}(t) = \kappa_{vJ}$ if $t > q_{J-1}$. Gamma priors are assumed for κ_{vj} $(j = 1, \cdots, J)$ to impose positive constraint. More details can be found in Ibrahim et al. (2004). Under M_5, we model $h_{v0}(t)$ by a Weibull hazard, with Gamma priors for the scale and shape parameters. The priors of the other parameters are specified as in M_1.

The estimated LPML under M_1-M_5 are 4833.6, 5115.2, 5007.9, 4889.1, and 5155.0, respectively. Clearly M_1 achieves the best performance. The nonparametric PT model allows the density function to deviate from the form imposed by the Weibull assumption. Assuming a cure group also improves the model fit. Model M_4 has the second best performance, which indicates that the PSA trajectory does play an important role in prostate cancer progression. The inferior performance of M_5 suggests that the Weibull hazard assumption might be too restrictive for our data.

The marginal distribution of TTP. The estimated cure probabilities p_v $(v = 1, 2)$ for the CH and AA treatments are 0.167 and 0.154, respectively. For advanced prostate cancer patients, here "cure" means that those patients take a very long time to progress to AIPC. Figure 3a shows the estimated densities of TTP in the susceptible group, $E(G_v \mid \boldsymbol{Y}, \boldsymbol{t}, \boldsymbol{d})$, under the two treatments. The horizontal axis is in years after the treatments. For comparison, Figure 3b plots the posterior estimate of Weibull densities under M_2. Figure 3 clearly shows deviation from the parametric Weibull distribution. For example, there is a small bump in the CH density curve around 7.5, which is also visible in the Kaplan-Meier estimates in Figure 1. This feature can not be captured by M_2. In Figure 1 we also plot the posterior estimate of the survival function under model M_1, where TTP are assumed to arise from the mixture

of a point mass at t_c and an unknown distribution G_v. Finally, posterior uncertainty on G_v is illustrated in Figure 4 by plotting ten random samples from the posterior distributions. Because Pólya trees with a fixed partition have discontinuities at the partition points, we have conducted kernel smoothing on the PT densities in both Figure 3 and 4.

The dependence of event times on longitudinal profiles. Under model M_1, different PSA profiles lead to different posterior distributions of T_{vi}. In Figure 5 we compare the PSA profiles in the first column, the estimated posterior probability of "cure" $P(\omega_{vi} = 1)$ and the conditional densities of $(T_{vi} \mid \omega_{vi} = 0)$ in the second column, and the estimated hazard curves given $\omega_{vi} = 0$ in the third column, from four patients with censored TTP. Each row corresponds to one patient, with the first two under treatment CH, and the last two under AA. We only plot the PSA profiles after initiation of the therapies. Figure 5 demonstrates the flexible nature of M_1. Each patient has a hazard curve of a different shape.

The longitudinal model parameters. Figure 2 plots the longitudinal PSA profiles of four patients together with their fitted values. Table 1 lists the posterior means and standard deviations of some of the parameters in M_1. The posterior estimates of ξ_v $(v = 1, 2)$ are practically identical, implying that the impact of TTP on the trajectory of PSA profiles are similar across the two treatments. The estimates of γ_{1v} indicate that the PSA profiles of patients in the AA arm on average have a larger slope after treatment. The level and slope of the drop in PSA after the CH/AA treatment are modeled by $l_v(t) = \eta_v[\exp(-\phi_{1v}t) - 1]$, where $t \geq 0$ is the time from the start of treatment v. A larger value of η_v suggests a deeper initial drop in PSA level. On the other hand, the larger the value of ϕ_{1v}, the sooner the treatment effect wears out ($l_v(t)$ becomes flat). We plot $l_v(t)$ in Figure 6. The patients under CH therapy experience a deeper drop in PSA level compared to those under AA therapy. Furthermore, the drop induced by CH therapy lasts longer.

Continuously reassessing the risk of progression. Given a currently observed PSA profile, we can use the proposed method to obtain the predictive distribution of TTP. The predictive distribution provides a good assessment of progression risk. With additional PSA measurements being observed, the predictive distribution can be updated to reflect newly obtained information.

We demonstrate this learning process in Figure 7. The left panel plots the PSA profiles of two hypothetical patients from the AA arm. Each point denotes a PSA measurement. The two patients have their PSA level measured at the same time points after treatment. Within the first two years the two PSA profiles are identical, and then they deviate: the first patient's PSA level stays low, while the second patient's PSA level gradually goes up. The center panel shows the continuously updated posterior estimates of $P(\omega = 1 \mid data)$, based on accumulated PSA measurements up to the points marked by the corresponding grey shades in the left panel. We interpret $P(\omega = 1 \mid data)$ as the individual probability of long term survival. The solid(dotted) line denotes the first(second) patient. Similarly, the right panel shows the continuously updated posterior estimates of $E(T \mid \omega = 0, data)$. The first patients has a flat PSA profile. With more PSA measurements being collected, both the estimates of $E(T \mid \omega = 0, data)$ and $P(\omega = 1 \mid data)$ increase steadily. The second patient presents a different situation. The PSA profile is flat in the first two years, and then it goes up. The proposed model captures this change. The estimated $P(\omega = 1 \mid data)$ first increases, but goes down as PSA level rises. It literally drops to zero when there is strong evidence of cancer progression. The estimates of $E(T \mid \omega = 0, data)$ show a similar change. This learning process helps physicians to continuously reassess the risk of progression, and make appropriate medical decisions.

Sensitivity analysis We conducted a sensitivity analysis to explore the impact of t_c and c on posterior inference, where t_c is the assumed extremely long TTP that could not be observed by the clinical trial and c is a parameter in the PT prior controlling how much G_v can

deviate from $\tilde{G}$. Smaller values of c allow greater flexibility. We tried two values for c, (0.1, 1), and three values for t_c, (15, 18, 20). The posterior means of p_v and $LPML$ are listed in Table 2. Model fit as summarized by the LPML remains essentially uunchanged across the considered models. The estimated cure rate slightly increases with larger c. Larger values for c imply stronger shrinkage of G_v to the parametric centering measure $\tilde{G}$. The parametric Weibull model $\tilde{G}$ can not represent the secondary mode that we see in the data and in the non-parametric inference for G_v. Compensating the missing secondary mode by an increased cure fraction could explain the change in the posterior means of p_v.

6 Discussion

We have proposed an approach for the joint modeling of event times and longitudinal measurements. The proposed model allows researchers to relax parametric assumptions on the survival submodel imposed by existing methods. When we assign non-parametric PT priors for the survival submodel, each subject can have a hazard curve of a different shape.

An important limitation of the model is that $P(T, Y) = P(T)P(Y \mid T)$ describes the effect of survival times T on longitudinal profiles Y, but it does not explicitly state how T is affected by Y. Given a particular longitudinal profile, we need to carry out posterior simulation to learn about the survival distribution. Under the traditional framework $P(T, Y) = P(T \mid Y)P(Y)$, the impact of longitudinal profiles on survival times is described by the longitudinal submodel $P(T \mid Y)$. For example, in a proportional hazard model, there are expressions that explicitly describe the relationship between the hazard function and the individual-specific random effects derived from longitudinal measurements.

The proposed approach can readily be generalized to problems with more than two treatments. Also, the longitudinal data model (3) is appropriate for the discussed application to the prostate cancer trial. In general, any other model with a regression on the event time could be used.

Acknowledgment

We thank Randall Millikan for supplying the data set for analysis. Our work was partially supported by the Cancer Center Support Core Grant CA16672, SPORE in prostate cancer grant CA90270 from the National Cancer Institute, National Institute of Health.

References

Barlow, W. E. and Prentice, R. L. (1988), 'Residuals for relative risk regression', *Biometrika* **75**, 65–74.

Brown, E. R. and Ibrahim, J. G. (2003), 'Bayesian approaches to joint cure-rate and longitudinal models with applications to cancer vaccine trials', *Biometrics* **59**(3), 686–693.

Carlin, B. P. and Chib, S. (1995), 'Bayesian model choice via Markov chain Monte Carlo methods', *Journal of the Royal Statistical Society, Series B: Methodological* **57**, 473–484.

Carter, B., Ferrucci, L., Ketterman, A., Landis, P., Wright, J., Epstein, J. I., Trock, B. and Metter, J. (2006), 'Detection of life-threatening prostate cancer with prostate-specific antigen velocity during a window of curability', *Journal of the National Cancer Institute* **98**, 1521–1527.

Carter, H. B., Morrell, C. H., Pearson, J. D., Brant, L. J., Plato, C. C., Metter, E. J., Chan, D. W., Fozard, J. L. and Walsh, P. C. (1992), 'Estimation of prostatic growth using serial prostate-specific antigen measurements in men with and without prostate disease', *Cancer Research* **52**(12), 3323–3328.

Chen, M.-H., Ibrahim, J. G. and Sinha, D. (2004), 'A new joint model for longitudinal and survival data with a cure fraction', *Journal of Multivariate Analysis* **91**(1), 18–34.

Dafni, U. G. and Tsiatis, A. A. (1998), 'Evaluating surrogate markers of clinical outcome when measured with error', *Biometrics* **54**, 1445–1462.

De Gruttola, V. and Tu, X. (1994), 'Modelling progression of CD4-Lymphocyte count and its relationship to survival time', *Biometrics* **50**, 1003–1014.

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

Gelfand, A., Dey, D. and Chang, H. (1992), Model determination using predictive distribution with implementation via sampling-based methods (with discussion), *in* 'Bayesian Statistics 4 – Proceedings of the Fourth Valencia International Meeting', Oxford University Press.

Green, P. J. (1995), 'Reversible jump Markov chain Monte Carlo computation and Bayesian model determination', *Biometrika* **82**, 711–732.

Guo, X. and Carlin, B. P. (2004), 'Separate and joint modeling of longitudinal and event time data using standard computer packages', *The American Statistician* **58**(1), 16–24.

Hanson, T., Branscum, A. and Johnson, W. (2007), Joint modeling of longitudinal and survival data using mixtures of polya trees, Technical report, University of Minnesota.

Hanson, T. E. (2006), 'Inference for mixtures of finite Polya tree models', *Journal of the American Statistical Association* **101**(476), 1548–1565.

Hanson, T. and Johnson, W. O. (2002), 'Modeling regression error with a mixture of Polya trees', *Journal of the American Statistical Association* **97**(460), 1020–1033.

Henderson, R., Diggle, P. and Dobson, A. (2000), 'Joint modelling of longitudinal measurements and event time data', *Biostatistics (Oxford)* **1**(4), 465–480.

Ibrahim, J. G., Chen, M.-H. and Sinha, D. (2004), 'Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials', *Statistica Sinica* **14**(3), 863–883.

Lavalley, M. P. and De Gruttola, V. (1996), 'Models for empirical Bayes estimators of longitudinal CD4 counts (Disc: P2337-2340)', *Statistics in Medicine* **15**, 2289–2305.

Lavine, M. (1992), 'Some aspects of Polya tree distributions for statistical modelling', *The Annals of Statistics* **20**, 1222–1235.

Law, N. J., Taylor, J. M. G. and Sandler, H. (2002), 'The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure', *Biostatistics (Oxford)* **3**(4), 547–563.

Lin, H., Turnbull, B. W., McCulloch, C. E. and Slate, E. H. (2002), 'Latent class models for joint analysis of longitudinal biomarker and event process data: Application to longitudinal prostate-specific antigen readings and prostate cancer', *Journal of the American Statistical Association* **97**(457), 53–65.

Mauldin, R. D., Sudderth, W. D. and Williams, S. C. (1992), 'Polya trees and random distributions', *The Annals of Statistics* **20**, 1203–1221.

Milikan, R., Wen, S., Lance, C., Brown, M., Moomey, B., Do, K. and Logothetis, C. (2007), Phase iii trial of androgen ablation with or without 3 cycles of systemic chemotherapy for advanced prostate cancer, Technical report, Department of Genitourinary Medical Oncology, MD Anderson Cancer Center, Houston, TX, 77030.

Muliere, P. and Walker, S. (1997), 'A Bayesian non-parametric approach to survival analysis using Polya trees', *Scandinavian Journal of Statistics* **24**(3), 331–340.

Neath, A. A. (2003), 'Polya tree distributions for statistical modeling of censored data', *Journal of Applied Mathematics and Decision Sciences* **7**(3), 175–186.

Paddock, S. M., Ruggeri, F., Lavine, M. and West, M. (2003), 'Randomized Polya tree models for nonparametric Bayesian inference', *Statistica Sinica* **13**(2), 443–460.

Pawitan, Y. and Self, S. (1993), 'Modeling disease marker processes in AIDS', *Journal of the American Statistical Association* **88**, 719–726.

Therneau, T. M., Grambsch, P. M. and Fleming, T. R. (1990), 'Martingale-based residuals for survival models', *Biometrika* **77**, 147–160.

Tsiatis, A. A. and Davidian, M. (2004), 'Joint modeling of longitudinal and time-to-event data: An overview', *Statistica Sinica* **14**(3), 809–834.

Tsiatis, A. A., De Gruttola, V. and Wulfsohn, M. S. (1995), 'Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS', *Journal of the American Statistical Association* **90**, 27–37.

Walker, S. G. and Mallick, B. K. (1997), 'Hierarchical generalized linear models and frailty models with Bayesian nonparametric mixing', *Journal of the Royal Statistical Society, Series B: Methodological* **59**, 845–860.

Walker, S. and Mallick, B. K. (1999), 'A Bayesian semiparametric accelerated failure time model', *Biometrics* **55**, 477–483.

Wulfsohn, M. S. and Tsiatis, A. A. (1997), 'A joint model for survival and longitudinal data measured with error', *Biometrics* **53**, 330–339.

Xu, J. and Zeger, S. L. (2001), 'Joint analysis of longitudinal data comprising repeated measures and times to events', *Journal of the Royal Statistical Society, Series C: Applied Statistics* **50**(3), 375–387.

Yu, M., Law, N. J., Taylor, J. M. G. and Sandler, H. M. (2004), 'Joint longitudinal-survival-cure models and their application to prostate cancer', *Statistica Sinica* **14**(3), 835–862.

Zhang, s., Müller, P. and Do, K.-A. (2008), 'Posterior inference for a novel semi-parametric survival model with longitudinal markers', *Supplemental material* .

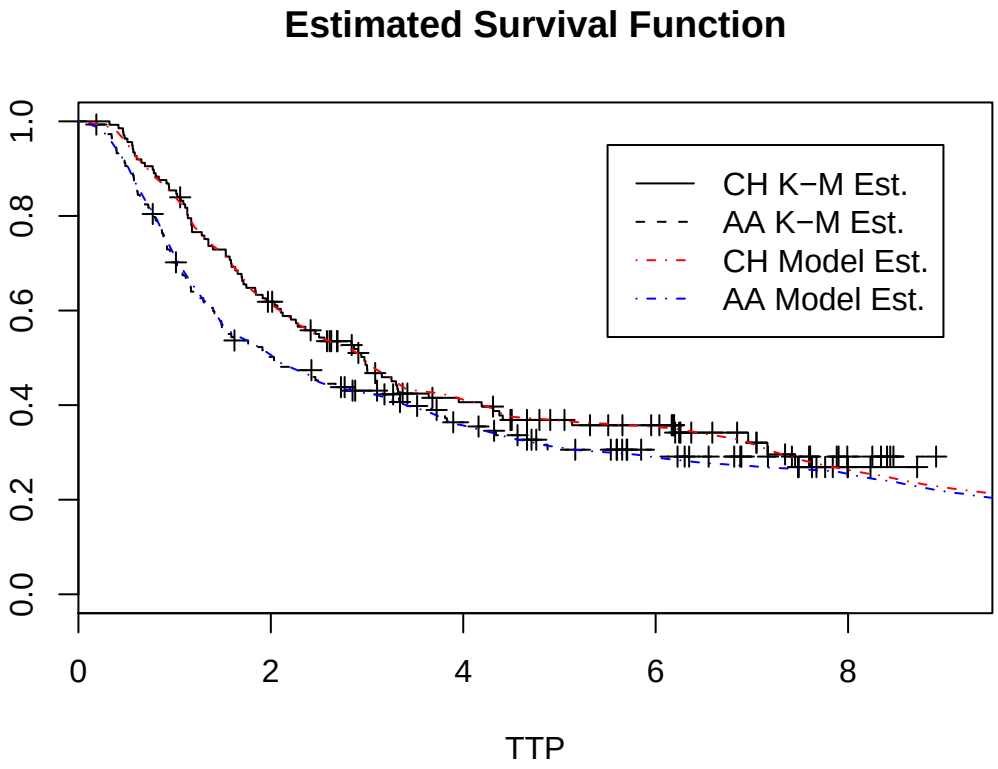

Figure 1: The horizontal axis indicates years after treatment. The censoring times are marked by +. "K-M Est." denotes Kaplan-Meier estimates. "Model Est." denotes estimates based on model M_1, where TTP are assumed to arise from the mixture of a point mass at t_c and an unknown distribution G_v. For $T < 6$ the Kaplan-Meier estimates and the model based estimates of the survival curves are virtually indistinguishable.

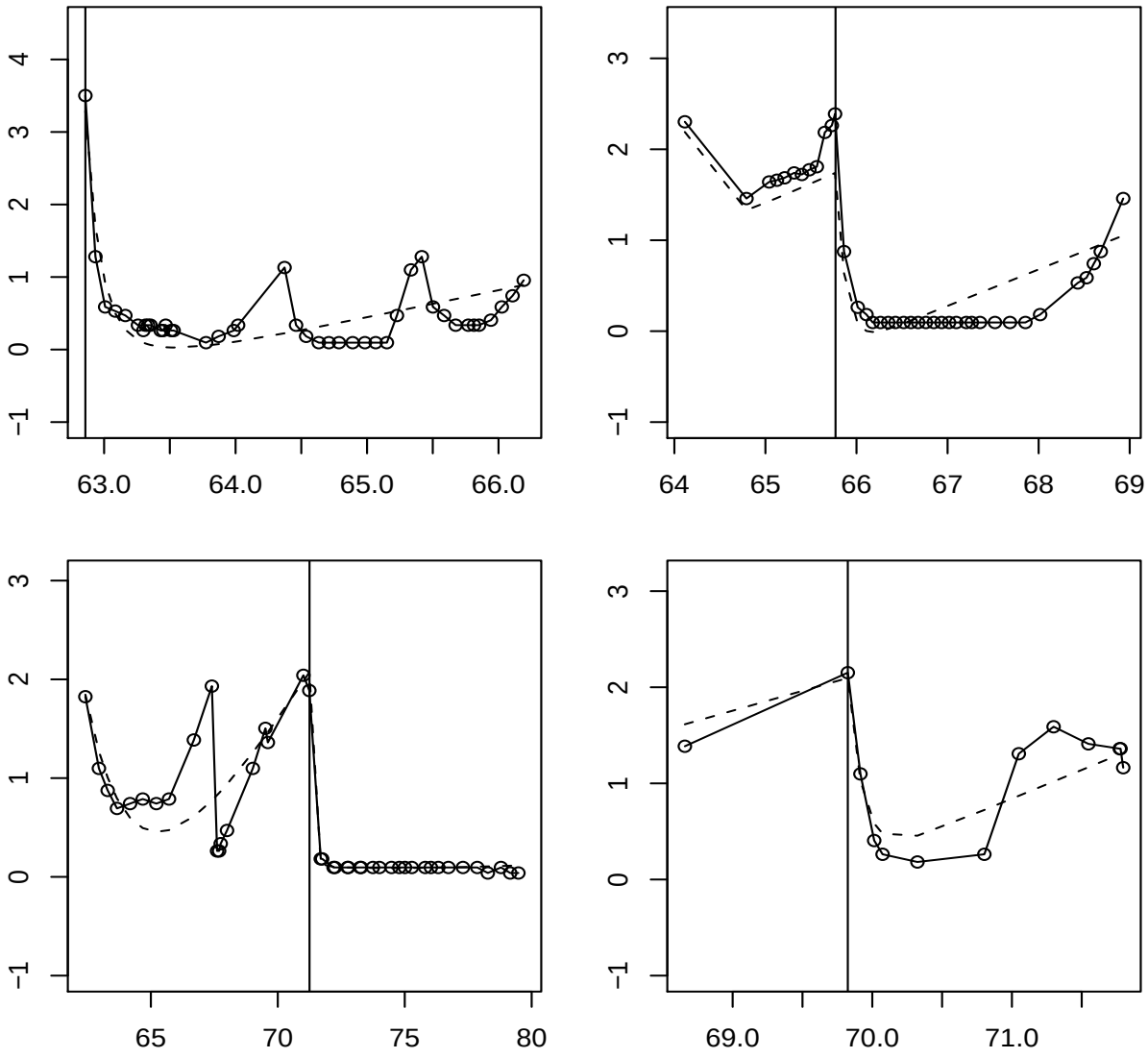

Figure 2: The observed longitudinal profiles and fitted values of 4 randomly selected patients. The vertical axis indicates $\log(PSA+1)$, and the horizontal axis is age in years. Each point denotes a PSA measurement. The dotted lines plot fitted values of the longitudinal profiles. The vertical line marks the initiation of the AA/CH therapy.

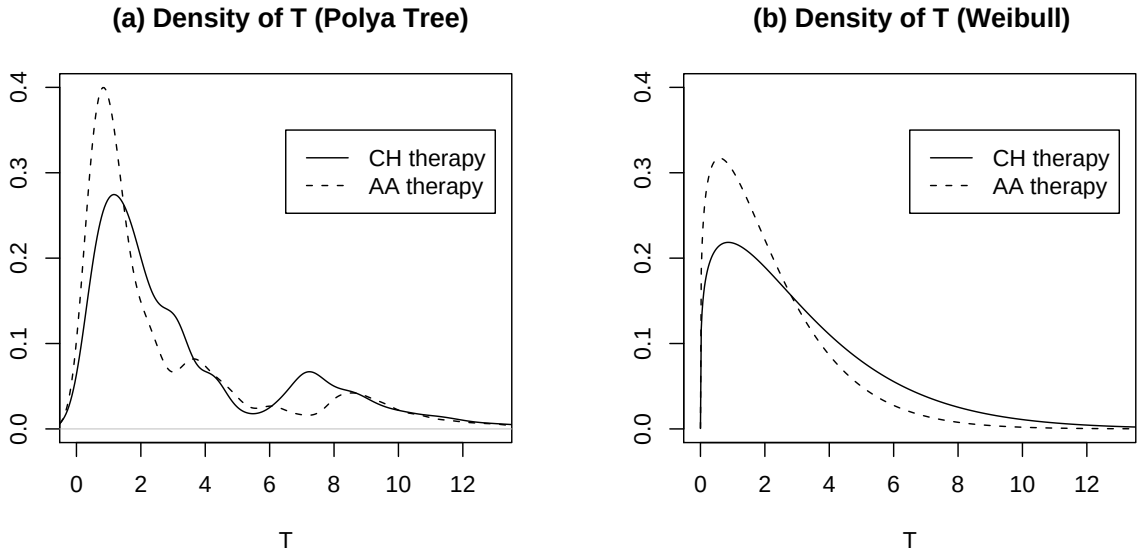

Figure 3: Posterior estimated $E(G_v \mid \boldsymbol{Y}, \boldsymbol{t}, \boldsymbol{d})$ under M_1 and M_2. The horizontal axis shows years after treatment.

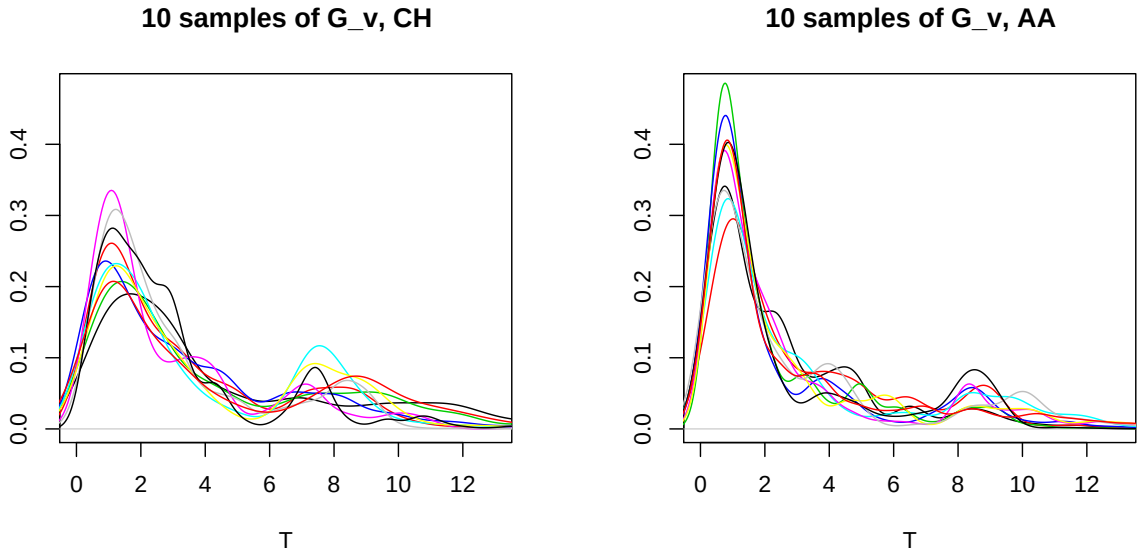

Figure 4: Ten random samples from $[G_v \mid \boldsymbol{Y}, \boldsymbol{t}, \boldsymbol{d}]$. The horizontal axis indicates years after treatment.

Table 1: Parameter Estimates in M_1

	Posterior Mean	Standard deviation
σ^2	0.209	0.005
μ_{θ_0}	-23.310	1.382
$\sigma^2_{\theta_0}$	13.903	2.681
μ_{θ_1}	0.483	0.023
$\sigma^2_{\theta_1}$	0.004	0.001
γ_{11}	0.447	0.122
γ_{12}	0.689	0.139
μ_{γ_2}	3.654	0.211
$\sigma^2_{\gamma_2}$	0.131	0.016
ϕ_{11}	11.077	0.864
ϕ_{12}	9.471	0.585
η_1	1.447	0.046
η_2	1.948	0.047
ξ_1	0.325	0.034
ξ_2	0.326	0.034

Table 2: Sensitivity Analysis

	c=0.1	$c = 1$
$t_c = 15$	(0.169, 0.153), -4930.50	(0.188, 0.159), -4889.98
$t_c = 18$	(0.167, 0.154), -4833.62	(0.182, 0.160), -4936.39
$t_c = 20$	(0.164, 0.154), -4979.03	(0.181, 0.160), -4845.34

In each cell we list the posterior estimations of (p_1, p_2) and $LPML$.

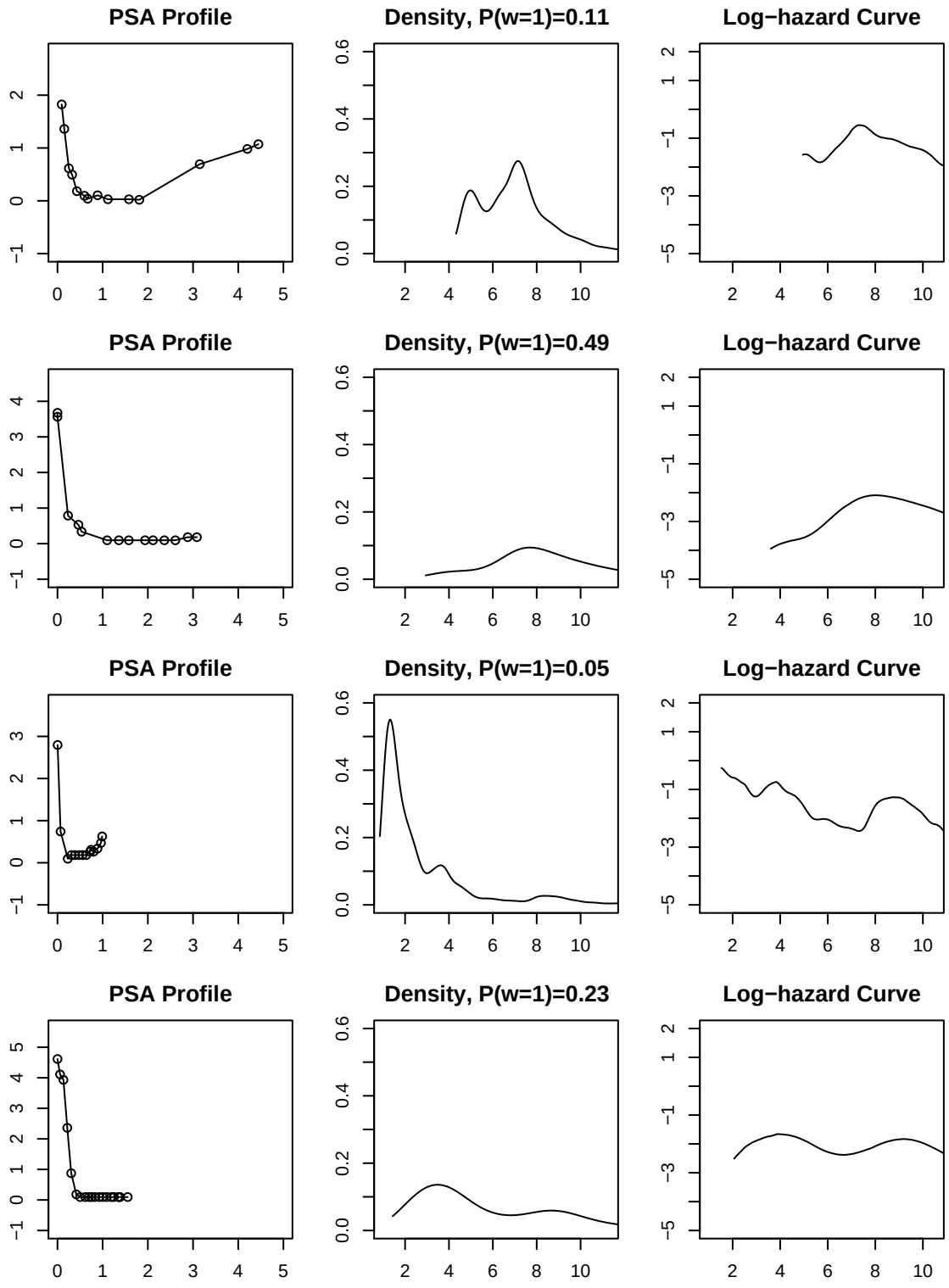

Figure 5: Posterior prediction for four patients with PSA profiles shown in the first column. The horizontal axis is time in years from the start of the AA/CH therapy.

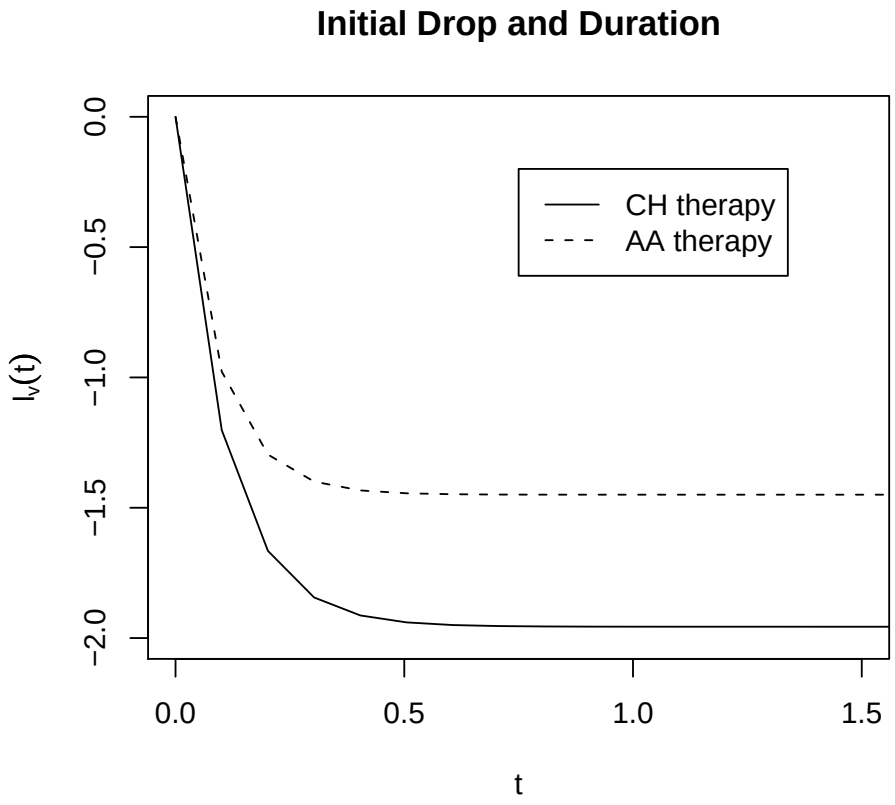

Figure 6: The initial slope and duration of treatment effect modeled by η_v and ϕ_{1v}. Here t is the time in years from the start of treatment v and $l_v(t) = \eta_v\{[\exp(-\phi_{1v}t) - 1]\}$.

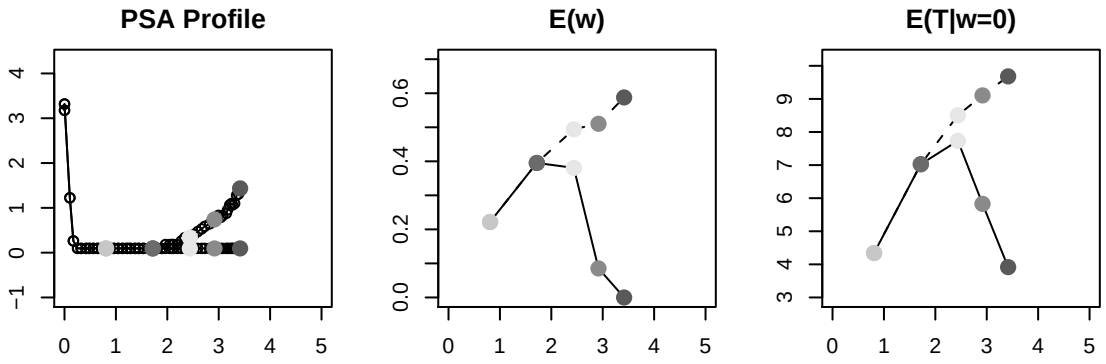

Figure 7: The left panel plots the PSA profiles of two hypothetical patients from the AA arm. The horizontal axis is time in years from the initiation of treatment. Each point denotes a PSA measurement. The first patient's PSA level stays low, while the second patient's PSA level gradually rises. The center panel shows the continuously updated posterior estimates of $P(\omega = 1 \mid data)$, based on accumulated PSA measurements up to the point marked by the corresponding grey shades in the left panel. The solid(dotted) line denotes the first(second) patient. Similarly, the third panel shows the continuously updated posterior estimates of $E(T \mid \omega = 0, data)$.