Bayesian Estimators for Conditional Hazard Functions

Ian W. McKeague

Department of Statistics, Florida State University, Tallahassee, FL 32306.

email: mckeague@stat.fsu.edu.

and

Mourad Tighiouart

Department of Mathematics and Statistics, Utah State University, Logan, UT 84322.

email: mourad@math.usu.edu.

SUMMARY

This article introduces a new Bayesian approach to the analysis of right-censored survival data. The hazard rate of interest is modeled as a product of conditionally independent stochastic processes corresponding to (1) a baseline hazard function, and (2) a regression function representing the temporal influence of the covariates. These processes jump at times that form a time-homogeneous Poisson process, and have a pairwise dependency structure for adjacent values. The two processes are assumed to be conditionally independent given their jump times. Features of the posterior distribution, such as the mean covariate effects and survival probabilities (conditional on the covariate), are evaluated using the Metropolis–Hastings–Green algorithm. We illustrate our methodology by an application to nasopharynx cancer survival data.

1. Introduction

A central problem in survival analysis is to infer the temporal evolution of covariate effects and baseline hazards from life history data. The proportional hazards model of Cox (1972) is the most popular approach to this problem. This model, however, is not flexible enough for some applications. For instance, in cancer mortality studies, some covariate effects

[1] *Key words:* Life history data; Metropolis–Hastings–Green algorithm; Right censoring; Time-dependent covariate effects.

may diminish with time, and in many practical situations, *a priori* information is available which is difficult to include in a Cox model analysis. In such cases, a Bayesian approach to modeling the conditional hazard function appears to be preferable, especially when limited data make nonparametric frequentist approaches impractical.

In this article, we develop a nonparametric Bayesian approach to the analysis of right-censored survival data. The approach will be illustrated using data on 181 nasopharynx cancer patients. One of the five covariates in this example is a measure of the extent of radiotherapy treatment. Our analysis indicates that the treatment has a significant effect on survival, and this effect is essentially constant over the period of follow-up. The effects of several other covariates, however, are found to vary significantly over time.

Our approach is based on a class of prior models for conditional hazard functions of the form

$$h(t|z) = h_0(t)\exp(\beta(t)'z), \tag{1.1}$$

where z is a p-vector of covariates. The baseline hazard function $h_0(t)$ and covariate effect $\beta(t)$ will be specified as step functions with jump times that form a time-homogeneous Poisson process. Conditional on these jump times, the levels of $\log(h_0(t))$ and $\beta(t)$ are modeled as independent Gaussian Markov random fields. Temporal variation of these processes will be controlled using a pairwise dependency structure for adjacent values.

We briefly mention some related frequentist approaches. A simple extension of the Cox model has been considered by Anderson and Senthilselvan (1982), who proposed piecewise constant $\beta(t)$ with a single jump. They analyzed a data set of cancer recurrence times, and showed that the two-step model fits the data much better than the Cox model. The authors cautioned, however, that in the presence of binary covariates and heavy censoring, allowing more than just a single jump in $\beta(t)$ can result in unstable estimates. Gore et al. (1984, Section 4.4) and Carter et al. (1983) modeled the changes in the covariate effects by assuming that $\log(\beta(t))$ is piecewise linear. A more general approach was developed by Zucker and Karr (1990), and Murphy and Sen (1991). They considered a model of the form (1.1) in which $\beta(t)$ is an arbitrary function of time. No assumptions were placed on the covariate effect $\beta(t)$ and the baseline hazard function $h_0(t)$ apart from smoothness conditions. Zucker and Karr used a penalized partial likelihood approach, resulting in spline estimators, and Murphy and Sen used histogram sieve estimators. Lin and Ying (1994, 1995) and McKeague and Sasieni (1994) have developed additive risk variations of the Cox model. For an extensive review of nonparametric frequentist approaches to inference from survival data, see Andersen et al. (1992).

A review of Bayesian methods for survival data can be found in Sinha and Dey (1997). Most exisiting methods do not adjust for covariate effects, and the cumulative hazard function is specified using a process with independent increments. In practical situations, however, correlated or smooth hazard function priors are more suitable. Gamerman (1991) considered the model (1.1) with the baseline hazard as a step function of the form

$$h_0(t) \;=\; \sum_{i=1}^{k} I(\tau_i < t \leq \tau_{i+1}) h_i, \tag{1.2}$$

where $0 = \tau_1 < \tau_2 < \ldots < \tau_{k+1}$ is a fixed grid of jump times. The log-baseline hazard levels $\lambda_i = \log(h_i)$ form a first-order autoregressive process, $\lambda_{i+1}|\lambda_i \sim N(\lambda_i, \sigma^2)$, and a similar structure is used for the covariate effect $\beta(t)$. West (1992) considered a special case of this type of prior for modeling time-varying hazards and covariate effects. West used a random walk structure for (h_i, β_i), and suggested a method of Bayesian model selection. Several authors have provided heuristic motivation for the choice of the fixed grid of jump times $\{\tau_1, \ldots, \tau_{k+1}\}$. Kalbfleisch and Prentice (1973) recommended that these jump times be selected independently of the data. In an example using cancer mortality data, West (1992) suggested shorter intervals over the first few years of follow-up, and longer intervals in subsequent years, the rationale being that there were more failure times available in the early stages.

A more complete Bayesian approach is obtained by putting a prior distribution on the jump times, achieving a dense support for the prior. In the absence of covariates, Arjas and Gasbarra (1994) modeled $h_0(t)$ as a step function of the form (1.2) with $k = \infty$, where the jump times $\tau_2 < \tau_3 < \ldots$ form a time homogeneous Poisson process, and the levels of the hazard rate $\{h_i, i \geq 1\}$ form a first-order autoregressive process. Estimation of the predictive hazard and survival functions was carried out using a modification of the Gibbs sampler algorithm to allow jump times to be added or deleted during updating. Arjas and Heikkinen (1997) introduced a related Markov random field model for the (prior) intensity of a non-homogeneous Poisson process, but they did not study it in the survival analysis context, and adjustment for covariate effects was not considered.

In this article, the baseline hazard function and the covariate effects are modeled as independent stochastic processes, with sample paths taking the form of step functions. The levels of these step functions form a Markov random field, specified by its local characteristics (componentwise conditional distributions). A pairwise dependency structure for adjacent values of the log-baseline hazard function and the covariate effects is developed. The models proposed by Arjas and Gasbara (1994) and Gamerman (1991) essentially arise as special cases. Increasing, decreasing, or bath-tub shape assumptions on the trend of the hazard rate levels can be easily implemented in our approach. Features of the posterior distribution will be calculated using the Metropolis–Hastings–Green (MHG) algorithm, Metropolis et al. (1953), Hastings (1970), Green (1995).

The article is organized as follows. Section 2 describes the model for simple right censored survival times with covariates. The Metropolis–Hastings–Green algorithm used for sampling from the posterior distribution of the parameter of interest is described in the Appendix. The analysis of the nasopharynx cancer survival data is presented in Section 3. Section 4 contains concluding remarks.

2. The Model

Let $T_1, \ldots, T_n$ be nonnegative independent random variables with associated p-dimensional covariate vectors z_j, $j = 1, \ldots, n$. Assume that the data may be subject to right censoring, i.e., we observe $(X_1, \delta_1, z_1), \ldots, (X_n, \delta_n, z_n)$ where $X_j = \min(T_j, U_j)$, U_j being the censoring time for the j-th individual, and $\delta_j = I\{T_j \leq U_j\}$. Our Bayesian approach consists of putting a prior distribution on the unknown baseline hazard function and p covariate effects.

The structure of the conditional hazard function is given by

$$h(t|z) = \sum_{i=1}^{\infty} I(\tau_i < t \leq \tau_{i+1}) h_i \exp(\beta_i' z), \tag{2.1}$$

where $0 = \tau_1 < \tau_2 < \tau_3 < \ldots$ is an increasing sequence of jump times, h_i is a baseline hazard level, and $\{\beta_i, i \geq 1\} = \{(\beta_{i1}, \ldots, \beta_{ip})', i \geq 1\}$ is a p-dimensional process describing the effect of the covariate z. Let $\tau_{\max} = \max_{1 \leq j \leq n} X_j$ and denote $\lambda_i = \log(h_i)$. We assume

1. The jump times $\tau_2, \tau_3, \ldots$ form a time-homogeneous Poisson process with rate α.

2. Given that there are $m-1$ jumps in the interval $[0, \tau_{\max}]$, $(\lambda_1, \lambda_2, \ldots, \lambda_m)$ is a Gaussian Markov random field, with a nearest neighbor structure, specified by its local characteristics: $\lambda_k | \{\lambda_i, i \neq k\} \sim N(\nu_k, \sigma_k^2)$.

It can be shown that the conditional mean $\nu_k = E(\lambda_k | \{\lambda_i, i \neq k\})$ is given by

$$\nu_k = \mu_k + s_k(\lambda_{k-1} - \mu_{k-1}) + r_k(\lambda_{k+1} - \mu_{k+1}),$$

where the hyperparameters $\mu_k = E(\lambda_k)$ represent the trend in the levels of the log-baseline hazard function, and s_k, r_k are the influences of the left and right neighbors of λ_k, respectively. The models proposed by Arjas and Gasbara (1994) and Gamerman (1991) essentially arise as special cases: use a constant trend and let $r_i \to 0$ and $s_i \to 1$. The joint distribution of $\lambda_1, \ldots, \lambda_m$ (given m fixed) is completely determined by its local characteristics provided they satisfy the following consistency conditions: s_k, r_k are nonnegative with $s_k + r_k < 1$, and $r_k \sigma_{k+1}^2 = s_{k+1}\sigma_k^2$, see, e.g., Besag and Kooperberg (1995). Now we specify s_k, r_k and σ_k^2, the aim being to force the corresponding baseline hazard function to be relatively 'smooth.' Of the two neighbors of λ_k, the one corresponding to the longer interval should have the greatest influence on its mean ν_k. We propose to use

$$r_k = \frac{(\Delta_k + \Delta_{k+1})\, c}{\Delta_{k-1} + 2\Delta_k + \Delta_{k+1}}, \; s_k = \frac{(\Delta_{k-1} + \Delta_k)\, c}{\Delta_{k-1} + 2\Delta_k + \Delta_{k+1}},$$
$$\sigma_k^2 = \frac{2\sigma^2}{\Delta_{k-1} + 2\Delta_k + \Delta_{k+1}},$$

where $\Delta_k = \tau_{k+1} - \tau_k$ is the gap between the k-th and $(k+1)$-st jump times, $2 \le k \le m-1, \sigma^2 > 0$, and $0 \le c < 1$. The influence parameters r_1, s_1, r_m, s_m at the boundaries are defined as above, but identifying the endpoints λ_1 and λ_m as neighbors. It is readily checked that the above consistency conditions are satisfied in this case. Other choices of r_k, s_k, σ_k^2 are also possible.

Arjas and Heikkinen (1997) used the Voronoi tessellation of $\{\tau_i\}$ to specify the jump times in their model for the intensity of a non-homogeneous Poisson process. This is technically appealing because of the one-to-one correspondence it induces between $\tau_1, \ldots, \tau_m$ and the log-intensity levels $\lambda_1, \ldots, \lambda_m$. However, we prefer using $\{\tau_i\}$ as the the jump times to facilitate comparison with the survival analysis models of Arjas and Gasbara (1994) and Gamerman (1991).

Given m, the joint distribution of $(\lambda_1, \lambda_2, \ldots, \lambda_m)$ is Gaussian with mean vector $\boldsymbol{\mu_m}$ and covariance matrix $(I - C)^{-1}M$, where $\boldsymbol{\mu_m} = (\mu_1, \ldots, \mu_m)$, $C = (c_{ij})_{1 \le i,j \le m}$, where $c_{ii+1} = r_i, c_{ii-1} = s_i$, $M = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$, and I is the identity matrix.

The hyperparameter α controls the rate of jump times, (μ_k) is the trend in the log-basekine hazard function, c controls the nearest neighbor interaction, and σ^2 represents the precision of the prior information. Choosing a value of c close to 1 amounts to vague prior knowledge of the trend parameters μ_k. For example, if $\mu_k = \mu$ for all k, then $\nu_k =$

$\mu(1-c) + s_k\lambda_{k-1} + r_k\lambda_{k+1}$ and the influence of μ on ν_k vanishes as $c \to 1$. For simplicity of presentation, we restrict attention to the case $\mu_k = \mu$, which indicates a constant prior level in the mean of the log-baseline hazard function.

To complete the prior specification of $h(t|z)$, we assume that, conditional on the first $m-1$ jump times $\tau_2, \ldots, \tau_m$ in the interval $(0, \tau_{\max})$, and independently of $h_1, \ldots, h_m$, the p covariate effects $(\beta_{1k}, \beta_{2k}, \ldots, \beta_{mk}), k = 1, 2, \ldots, p$, are independent and for each k, $(\beta_{1k}, \beta_{2k}, \ldots, \beta_{mk})$ is a Gaussian Markov random field, specified by its local characteristics. The same expressions for the influences r_i, s_i and conditional variances σ_i^2 used to model the baseline hazard levels are adopted for each covariate effect. The resulting trend, nearest neighbor interaction, and precision of the prior information hyperparameters are denoted by (μ_{ik}), c_k, and σ_k^2, respectively. For $k = 1, \ldots, p$, and $i = 1, \ldots, m$, let

$$W_{ik} = \sum_{\{j: \tau_i < X_j \le \tau_{i+1}, \delta_j = 1\}} z_{jk},$$

with $\tau_{m+1} = \tau_{\max}$, and $W_i = (W_{i1}, W_{i2}, \ldots, W_{ip})$. We assume noninformative censoring, which implies [see, e.g., Andersen et al. (1992), p. 150] that the likelihood is proportional to

$$\begin{aligned} &\prod_{j=1}^{n} (h(X_j|z_j))^{\delta_j} \prod_{j=1}^{n} \exp\left\{ -\int_0^{X_j} h(s|z_j)\, ds \right\} \\ &= \exp\left\{ \sum_{i=1}^{m} (N_i\lambda_i + \beta_i' W_i) - \int_0^{\tau_{\max}} \left[\sum_{j=1}^{n} I(X_j \ge s)\, h(s|z_j) \right] ds \right\}. \end{aligned}$$

Denote by C_k the matrix of spatial dependency parameters for the Markov random field $\boldsymbol{\beta_{mk}} = (\beta_{1k}, \ldots, \beta_{mk})$, $\boldsymbol{\mu_{mk}} = (\mu_{1k}, \ldots, \mu_{mk})$ its mean value, $M_k = \mathrm{diag}(\sigma_{1k}^2, \ldots, \sigma_{mk}^2)$, $B_k = (I - C_k)^{-1} M_k$ the covariance matrix of $\boldsymbol{\beta_{mk}}$ for $k = 1, \ldots, p$, and let $\boldsymbol{\beta_m} = (\boldsymbol{\beta_{m1}}, \boldsymbol{\beta_{m2}}, \ldots, \boldsymbol{\beta_{mp}})$. The posterior density of the $m(p+2)$ dimensional parameter $(\boldsymbol{\tau_m}, \boldsymbol{\lambda_m}, \boldsymbol{\beta_m})$ is proportional the product of the prior and likelihood,

$$\alpha^m (2\pi)^{-\frac{m(1+p)}{2}} |A|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\lambda_m} - \boldsymbol{\mu})' A (\boldsymbol{\lambda_m} - \boldsymbol{\mu})\right\}$$

$$\times \prod_{k=1}^{p} |B_k|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta_{mk}} - \boldsymbol{\mu_{mk}})' B_k (\boldsymbol{\beta_{mk}} - \boldsymbol{\mu_{mk}})\right\}$$

$$\times \exp\left\{\sum_{i=1}^{m}(N_i\lambda_i + \beta_i' W_i) - \int_0^{\tau_{\max}} \left[\sum_{j=1}^{n} I(X_j \geq s)\, h(s|z_j)\right] ds\right\},$$

where $A = M^{-1}(I - C)$.

We have devised a Metropolis–Hastings–Green sampler to extract features of the posterior distribution, such as the mean covariate effects and survival probabilities (conditional on the covariates), see the Appendix.

3. Case Study

The class of priors we proposed in Section 2 is used here to analyze a data set of nasopharynx cancer patients presented in West (1987). In Section 3.1, we describe the data set, and review some previous analyses. The hyperparameters controlling the class of priors are given in Section 3.2. In Section 3.3., we compare our estimates with West's. In Section 3.4, we compare the predictive hazard and survival function estimates.

3.1 *Data and Previous Models*

West (1987) studied data on 181 nasopharynx cancer patients whose cancer careers, culminating in either death (127 cases) or censoring (54 cases) are recorded to the nearest month, ranging from 1 to 177 months. A variety of covariates are available for each subject, none of which are viewed as subject to change during the career of the patient. As mentioned in West (1992), alternative analyses of the data, and some further exploratory investigations, serve to indicate the importance of a subset of the covariates in explaining the observed variation in survival times. The analyses we present here are based on five covariates: (1) Sex of the patient (0 for male, 1 for female); (2) Age of the patient at time $t = 0$, the start of monitoring of the cancer career of that patient (standardized to have zero mean and unit standard deviation across all patients in the study); (3) Dosel, an average measure of the extent of radiotherapy treatment to which the patient has been subjected (also standardized, as with age); (4) Tumor1, a measurement of the extent of the cancer (in terms of an estimate of the number of cancerous cells), taking value 1, 2, 3 or 4; (5) Tumor2, a measure similar to Tumor1, taken from a different X-ray section, again taking values 1, 2, 3 or 4. These tumor variables are measures of tumor growth at the start of monitoring, hence proxies for tumor lifetime. Higher levels of each are expected to be consistent with increased hazards.

West (1992) analyzed the data with the above set of covariates using a random walk structure for the log-baseline hazard function and the covariate effects (h_i, β_i), and obtained the Bayes estimates shown on the left of Figures 1–3. We propose an analysis of this data set based on model (2.1), in which the log-baseline hazard and the five covariate effects are allowed to vary over time.

3.2 *Prior Specification*

The analysis reported in West (1992) required the discretization of the time axis into intervals, with endpoints at every eighth observed death; this gives 16 intervals, with endpoints 3, 5, 7, 9, 10, 12, 13, 16, 19, 21, 25, 29, 35, 45, 60 and ∞, the final interval including just seven observed death times, and 26 censored times. The prior mean number of jump times of the time homogeneous Poisson process is taken as 10.

To allow comparison with the initial priors chosen by West, the trend parameters of the log-baseline hazard function, sex effect, age effect, dosel effect, tumor1 effect, and tumor2 effect are taken to be -3, 0, 0.5, -0.5, 0.5, and 0.5, respectively. The precision of the prior information for the log-baseline hazard function is taken as $\sigma = 1.5$, and $\sigma_i = 0.15, i = 1, \ldots, 5$ for the five covariate effects. Finally, the nearest neighbor interaction hyperparameters for the log-baseline hazard function and the five covariate effects are all taken as 0.98.

3.3 *Comparison of Bayes Estimates*

The solid lines on the left sides of Figures 1, 2, and 3 represent the estimated posterior mean log-baseline hazard function, and the posterior mean effects of the five covariates obtained by West. The dotted lines represent one posterior standard deviation above and below the posterior mean. The plots on the right hand side represent the proposed estimates.

Log-Baseline Hazard. The posterior mean log-baseline hazard function at the top right of Figure 1 appears to be stable for $t \leq 60$ months, and is slightly below the one obtained by West (top left of Figure 1). However, a sharp decrease in our estimate of the log-baseline hazard function occurs at about 80 months, and differs remarkably from to stable trajectory obtained by West. This may not be surprising since the number of observed deaths in the interval $(60, 177]$ is only seven, compared to 120 below sixty months. Note also the departure of the posterior mean from the constant prior trend parameter $\mu = -3$, indicating that the data is 'swamping' the prior information about the mean. Thus, the sharp decrease at 80 months reflects a genuine feature of the data which is not discernible using West's approach. The posterior standard deviation curves we obtained are close to the ones obtained by West for $t \leq 60$ months, but become much wider afterwards. This is due to the nature of our prior process for the jump times, which tries to detect any temporal variation using a random number of jump times, whereas West is estimating a single parameter for $t > 60$ months.

Sex Effect. The posterior mean Sex effect at the bottom right of Figure 1 has a similar pattern as the one obtained by West for $t \leq 70$ months (bottom left of Figure 1). After 70 months, the Sex effect we obtained increases to a maximum value of approximately -1.75, at around 100 months, then decreases later on. This may not be surprising, since there are two observed deaths for females at times 96 and 98 months, and none afterwards. In any case, the posterior mean Sex effects obtained by West and us are negative, indicating consistently lower hazard for female patients relative to males.

Age Effect. The posterior mean Age effect at the top right of Figure 2 has a bath-tub shape for $t \leq 60$ months, similar to the one obtained by West (top left of Figure 2). But then, the age effect we obtained decreases to a minimum value at around 100 months, then increases. Again, this is due to the two observed deaths at times 96 and 98 months. Both of these two patients have ages below average, with 0.853 and 1.759 standard deviations below the mean. The last observed death time is 163 months, with age 0.829 standard deviations above average, causing, perhaps, the Age effect to increase. In any case, the posterior mean Age effect are positive, indicating increased hazards for older patients.

Dosel Effect. The posterior mean Dosel effect shown at the bottom right of Figure 2 appears stable, favoring values around -0.2 across time. The proposed estimate of the Dosel effect is slightly below the one obtained by West (bottom left of Figure 2), overestimating the strength of treatment effect in reducing hazards. Again, in any case, the Dosel effect is negative, indicating the beneficial nature of the treatment.

Tumor1 Effect. The posterior mean Tumor1 effect shown at the top right of Figure 3 is higher than the one obtained by West (top left of Figure 3), for $t \leq 60$ months, but with a similar trajectory. This implies higher hazard rates relative to West's estimate. After 60 months, a significant increase in this effect is observed, as opposed to a decay to around zero for West's model. West noted that this decay could be anticipated in qualitative form by examining those patients with observed death times exceeding six years; there are very few such patients and the death rates are apparently unrelated to the Tumor1 variable. Our estimated Tumor1 effect increases to a maximum value at around 100 months. Again, we examined the two observed death times at 96 and 98 months. The levels of the Tumor1 variable for these two patients are high, 3 and 4 (recall that the possible values of Tumor1 are 1,2,3 and 4). This, perhaps, explains the increased nature of this effect. The slight decrease of this effect after 100 months could be due to the last observed death times at 135 and 163 months. The Tumor1 levels for both patients are 2.0.

Tumor2 Effect. The posterior mean Tumor2 effect at the bottom right of Figure 9 shows a similar pattern with the one obtained by West (bottom left of Figure 9) for $t \leq 60$ months, with much higher values between 40 and 60 months. But then, a similar pattern with Tumor1 effect is observed after 60 months, which again could be explained by the two observed death times at 96 and 98 months. The levels of Tumor2 for these two patients are both 3.0. The last observed death times at 135 and 163 months have Tumor2 levels of 3.0 and 1.0, respectively. The similar pattern of the posterior mean effects of Tumor1 and Tumor2, suggests that a correlated prior process for the two effects is more realistic. This will be investigated in future work.

3.4 *Predictive Hazard and Survival Curves*

Consider a hypothetical male patient, of average age, with Tumor2 level of 0, and treated with the average level of Dosel; thus, the corresponding values of the covariates Sex, Age, and Dosel are all zero. The right hand side of Figure 4 shows our proposed predictive hazard estimates for such a patient, across the levels of Tumor1 variable, 0, 1, 2, 3, 4; lower hazards correspond to lower levels of Tumor1. The corresponding estimates of West are shown on the left side of Figure 4. Our proposed estimates have a similar shape with West's estimates, but with much lower values between 70 and 155 months or so. Below 60 months, both estimates are quite close, and this is reflected by the estimated survival curves shown on Figure 5.

4. Concluding Remarks

We have proposed a nonparametric Bayesian approach to inference from right-censored survival data. Our approach allows the inclusion of prior information on trends in the baseline hazard function and the covariate effects. The class of prior distributions is more flexible than those of earlier approaches; it extends the model proposed by Arjas and Gasbara (1994) by including covariate effects, and is more general than the approach considered by Gamerman (1991) in the sense that (1) the choice of the jump times is not based on heuristic considerations, and (2) Bayes estimates are not step functions, so that no *ad hoc* smoothing procedure is required. Although we did not do so, it would be easy to extend our approach to provide automatic specification of the degree of smoothing by placing a prior on the hyperparameters and have them estimated simultaneously along with the baseline hazard and covariate effects.

Another advantage of our approach is the computational method used to extract features of the posterior distribution. The Metropolis–Hastings–Green algorithm we developed can be extended to handle time-dependent covariates, whereas it is not clear how that could be done for the dynamic version of the Gibbs sampler used by Arjas and Gasbara. In our case, only minor changes in the acceptance probabilities of the Metropolis–Hastings–Green algorithm are needed.

We have illustrated our methodology by giving an analysis of a nasopharynx cancer survival data set. The proposed class of prior processes defining our model was flexible enough to allow the detection of subtle features in the log-baseline hazard function and covariate effects; in particular, we found a sharp decrease in the log-baseline hazard function at a certain time point, which cannot be seen using previous approaches.

In future work it would be worthwhile to extend our approach to allow nonlinear covariate effects. For example, a one-dimensional covariate z could be replaced by a step function $\gamma(z)$ having levels specified by a Gaussian Markov random field.

References

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1992). *Statistical Models Based on Counting Processes* New York: Springer-Verlag.

Anderson, J.A. and Senthilselvan, A. (1982). A Two-Step Regression Model for Hazard Functions. *Applied Statistics* **31,** 44–51.

Arjas, E., and Gasbarra, D. (1994). Nonparametric Bayesian Inference for Right-Censored Survival Data, Using the Gibbs Sampler. *Statistica Sinica* **2,** 505–524.

Arjas, E., and Heikkinen, J. (1997). An Algorithm for Nonparametric Bayesian estimation of a Poisson intensity. *Journal of Computational Statistics* **12,** 385–402.

Besag, J. E. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society, Series B* **36,** 192–225.

Besag, J. E., and Kooperberg, C. (1995). On Conditional and Intrinsic Autoregressions. *Biometrika* **82,** 733–746.

Carter, W. H., Wampler, G. L., and Stablein, D. M. (1983). *Regression Analysis of Survival Data in Cancer Chemotherapy.* New York: Dekker.

Cox, D. R. (1972). Regression Models and Life-Tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34,** 187–220.

Cox, D. R., and Oakes, D. (1984). *Analysis of Survival Data.* London: Chapman and Hall.

Gamerman, D. (1991). Dynamic Bayesian Models for Survival Data. *Applied Statistics* **40,** 63–79.

Gore, S. M., Pocock, S. J., and Kerr, G. R. (1984). Regression Models and Non-Proportional Hazards in the Analysis of Breast Cancer Survival. *Applied Statistics* **33,** 176–195.

Green, P. J. (1995). Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika* **82,** 711–732.

Hastings, W. K. (1970). Monte Carlo Sampling Methods using Markov Chains and their Applications. *Biometrika* **57,** 97–109.

Kalbfleisch, J. D., and Prentice, R. L. (1973). Marginal Likelihoods based on Cox's Regression and Life Model. *Biometrika* **60,** 267–278.

Lin, D. Y., and Ying, Z. (1994). Semiparametric Analysis of the Additive Risk Model. *Biometrika* **81,** 61–71.

Lin, D. Y., and Ying, Z. (1995). Semiparametric Analysis of General Additive-Multiplicative Hazard Models for Counting Processes. *The Annals of Statistics* **23,** 1712–1734.

McKeague, I. W., and Sasieni, P. D. (1994). A Partly Parametric Additive Risk Model. *Biometrika* **81,** 501–514.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics* **21,** 1087–1092.

Meyn, S. P., and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability.* London: Springer-Verlag.

Murphy, S. A., and Sen, P. K. (1991). Time-Dependent Coefficients in a Cox-Type Regression Model. *Stochastic Processes & Application* **39,** 153–180.

Sinha, D., and Dey, D. K. (1997) Semiparametric Bayesian Analysis of Survival Data. *Journal of the American Statistical Association* **92,** 1195–1212.

West, M. (1987). Analysis of Nasopharynx Cancer Data using Dynamic Bayesian Models. Warwick Research Report 109 and Technical Report 7–1987 of the Department of Mathematics, II, University of Rome.

West, M. (1992). Modelling Time-Varying Hazards and Covariate Effects. Survival Analysis: State of the Art, J. P. Klein, P. K. Goel, eds. Kluwer, Boston, 47–62.

Zucker, D. M., and Karr, A. F. (1990). Nonparametric Survival Analysis with Time-Dependent Covariate Effects: A Penalized Partial Likelihood Approach. *The Annals of Statistics* **18,** 329–353.

APPENDIX

Metropolis–Hastings–Green algorithm

To simplify the description of the algorithm, we assume that $p = 2$ and a constant trend in the covariate effects. The extension to more than two covariates is straightforward. The procedure for calculating features of the posterior distribution of $(\tau_m, \lambda_m, \beta_{m1}, \beta_{m2})$ (note that here m is random) consists of running a reversible Markov chain on the state space $\mathcal{S} = \cup_{i\geq 1} S_i$, where $S_i = D_i \times \mathbb{R}^{pi}$, and $D_i = \{(x_1, x_2, \ldots, x_i) : 0 = x_1 < x_2 < \ldots < x_i < \tau_{\max}\}$, using the Metropolis–Hastings–Green algorithm. Suppose that the current state of the chain is $(\tau_m, \lambda_m, \beta_{m1}, \beta_{m2}) \in S_m$, and denote by $(\tau'_{m'}, \lambda'_{m'}, \beta'_{m'1}, \beta'_{m'2}) \in S_{m'}$ the next state of the chain. When $m' = m$, the update can be done using the classical Metropolis–Hastings algorithm, otherwise some adjustments in the transition kernels are needed when transitions are made between subspaces of different dimension. Green (1995) generalized the classical Metropolis–Hastings algorithm by preserving reversibility of the Markov chain when moving between subspaces of different dimension. To simplify the description of the algorithm, let $x = (\tau_m, \lambda_m, \beta_{m1}, \beta_{m2})$ and denote by π the posterior distribution of the parameter. We will consider five transition kernels $P_i(x, A), i = 1, \ldots, 5$, with corresponding state-dependent mixing probabilities $p_i(x)$, satisfying $\sum_{i=1}^{5} p_i(x) = 1$. Denote $Q_i(x, A) = p_i(x)P_i(x, A)$. We also need five symmetric measures $\xi_i(dx, dx')$, such that ξ_i dominates $\pi(dx)Q_i(x, dx')$ for each i. Finally, let

$$f_i(x, x') = \frac{\pi(dx)Q_i(x, dx')}{\xi_i(dx, dx')}.$$

The Metropolis–Hastings–Green algorithm updates the current state x of the Markov chain as follows:

1. Select a proposal kernel P_i with probability $p_i(x)$;
2. Generate x' from $P_i(x, \cdot)$;
3. Accept x' with probability $\min\{1, \frac{f_i(x',x)}{f_i(x,x')}\}$, otherwise stay at x.

It can be shown that the resulting MHG transition kernel is reversible with respect to π, see Green (1995). In the context of our problem, transition from $(\tau_m, \lambda_m, \beta_{m1}, \beta_{m2})$ to a new point $(\tau'_{m'}, \lambda'_{m'}, \beta'_{m'1}, \beta'_{m'2})$ is accomplished by randomly selecting one of five types of moves (H,H1,H2,B,D): H, H1, and H2 correspond to a change of height of a randomly selected level of the baseline hazard function, the first covariate effect, and the second covariate effect, respectively. Moves B and D correspond to birth and death of some jump time. Denote by $p^m_{\mathrm{H}}, p^m_{\mathrm{H1}}, p^m_{\mathrm{H2}}, p^m_{\mathrm{B}}$, and p^m_{D} the probabilities of selecting the five different types of moves H, H1, H2, B, and D when the current state of the Markov chain is (τ_m, λ_m). Note that $N(\tau_{\max}) = \#\{i : \tau_i < \tau_{\max}\}$ has a Poisson distribution with parameter $\alpha\tau_{\max}$. The choice of the state-dependent mixing probabilities will be similar to the ones chosen by Green (1995). We take $p^m_{\mathrm{B}} = \gamma \min\{1, \frac{P(N(\tau_{\max})=m+1)}{P(N(\tau_{\max})=m)}\}$, $p^m_{\mathrm{D}} = \gamma \min\{1, \frac{P(N(\tau_{\max})=m-1)}{P(N(\tau_{\max})=m)}\}$, with $p^m_{\mathrm{B}} + p^m_{\mathrm{D}} = \eta$ where η is a sampler parameter, and γ is completely determined by η. Finally, we set $p^1_{\mathrm{D}} = 0$ and we take $p^m_{\mathrm{H}} = p^m_{\mathrm{H1}} = p^m_{\mathrm{H2}}$. When selecting a moves of type H, H1 or H2, the acceptance probability is the same as in the classical Metropolis–Hastings algorithm:

$$\min\{1, \text{(likelihood ratio)} \times \text{(prior ratio)} \times \text{(proposal ratio)}\},$$

whereas if moves of type B or D are selected, the current state $(\tau_m, \lambda_m, \beta_{m1}, \beta_{m2})$ is mapped onto $(\tau'_{m'}, \lambda'_{m'}, \beta'_{m'1}, \beta'_{m'2})$ by a one-to-one transformation τ. The acceptance probability then takes the form

$$\min\{1, \text{(likelihood ratio)} \times \text{(prior ratio)} \times \text{(proposal ratio)} \times J(\tau)\},$$

where $J(\tau)$ is the Jacobian of the transformation τ. A detailed description of the various transitions and expressions for the acceptance probabilities is given below.

Let

$$\begin{aligned} S(\tau_1, \tau_2, \tau_3, \tau_4) &= \textstyle\sum_{j:\tau_1 < X_j \le \tau_2} (X_j - \tau_1) \exp(\tau_3 z_{j1} + \tau_4 z_{j2}) \\ &\quad + (\tau_2 - \tau_1) \textstyle\sum_{j:X_j > \tau_2} \exp(\tau_3 z_{j1} + \tau_4 z_{j2}). \end{aligned}$$

Move of type H An index k is uniformly selected from $\{1,2,\ldots,m\}$ and V is generated uniformly in the interval $(-\delta,\delta)$, where δ is a sampler parameter. The proposed new level for the log-baseline hazard rate is $\lambda'_k=\lambda_k+V$. The proposed new point is $(\tau'_m,\lambda'_m,\beta'_{m1},\beta'_{m2})$ with $\tau'_m=\tau_m,\beta'_{m1}=\beta_{m1},\beta'_{m2}=\beta_{m2}$, and $\lambda'_i=\lambda_i$ for $i\neq k$.
The likelihood ratio is

$$\exp\left\{N_k(\lambda'_k-\lambda_k)+(e^{\lambda_k}-e^{\lambda'_k})\,S(\tau_k,\tau_{k+1},\beta_{k1},\beta_{k2})\right\}.$$

The prior ratio is

$$\exp\{-\Phi_{\mathrm{H}}(A,\mu,\lambda_m,\lambda'_m)/2\}.$$

where

$$\begin{aligned}\Phi_{\mathrm{H}}(A,\mu,\lambda_m,\lambda'_m) &= a_{kk}(\lambda'_k-\lambda_k)(\lambda_k+\lambda'_k-2\mu)\\ &\quad +2a_{kk-1}(\lambda_{k-1}-\mu)(\lambda'_k-\lambda_k)\\ &\quad +2a_{kk+1}(\lambda_{k+1}-\mu)(\lambda'_k-\lambda_k).\end{aligned}$$

The proposal ratio is 1 by symmetry of the proposal density.

Move of type H1 An index k is uniformly selected from $\{1,2,\ldots,m\}$ and V is generated uniformly in the interval $(-\delta_1,\delta_1)$, where δ_1 is a sampler parameter. The proposed new level for the first covariate effect is $\beta'_{k1}=\beta_{k1}+V$. The proposed new point is $(\tau'_m,\lambda'_m,\beta'_{m1},\beta'_{m2})$ with $\tau'_m=\tau_m,\lambda'_m=\lambda_m,\beta'_{m2}=\beta_{m2}$, and $\beta'_{i1}=\beta_{i1}$ for $i\neq k$.

The likelihood ratio is

$$\exp\left\{W_{k1}(\beta'_{k1}-\beta_{k1})-e^{\lambda_k}\left[S(\tau_k,\tau_{k+1},\beta'_{k1},\beta_{k2})-S(\tau_k,\tau_{k+1},\beta_{k1},\beta_{k2})\right]\right\}.$$

The prior ratio is

$$\exp\{-\Phi_{\mathrm{H}}(B_1,\mu_1,\beta_{m1},\beta'_{m2})/2\}.$$

The proposal ratio is 1 by symmetry of the proposal density. Similar expressions hold for move of type H2.

Move of type B A new jump time $\tau^\star$ is drawn uniformly in the interval $(\tau_1, \tau_{\max})$. Suppose that $\tau^\star \in (\tau_{k-1}, \tau_k)$. This new jump time induces two new levels for the log-baseline hazard rate λ'_{k-1} and λ'_k, two new levels for the first covariate effect $\beta'_{k-1,1}, \beta'_{k1}$, and two new levels for the second covariate effect $\beta_{k-1,2}, \beta_{k2}$ using the transformation $(\tau_{\epsilon,V}, \tau_{\epsilon_1,V_1}, \tau_{\epsilon_2,V_2})$, where V, V_1, V_2 are independent, uniformly distributed over $(-\epsilon,\epsilon), (-\epsilon_1,\epsilon_1), (-\epsilon_2,\epsilon_2)$, respectively, where $\epsilon, \epsilon_1, \epsilon_2$ are sampler parameters, and $\tau_{\epsilon,V}$ is defined as follows:
first, draw V uniformly in the interval $(-\epsilon,\epsilon)$, where ϵ is a sampler parameter, then the new log-hazard levels λ'_{k-1} and λ'_k are taken to be convex combinations of the perturbations $\lambda_{k-1}+V$ and $\lambda_{k+1}-V$ with λ_{k-2} and λ_k, respectively. More specifically, we let

$$\lambda'_{k-1} = \frac{\Delta'_{k-1}}{\Delta_{k-1}}\lambda_{k-2} + \frac{\Delta'_k}{\Delta_{k-1}}(\lambda_{k-1}+V),$$

$$\lambda'_k = \frac{\Delta'_{k-1}}{\Delta_{k-1}}(\lambda_{k-1}-V) + \frac{\Delta'_k}{\Delta_{k-1}}\lambda_k.$$

first, draw V uniformly in the interval $(-\epsilon,\epsilon)$, where ϵ is a sampler parameter, then the new log-hazard levels λ'_{k-1} and λ'_k are taken to be convex combinations of the perturbations $\lambda_{k-1}+V$ and $\lambda_{k+1}-V$ with λ_{k-2} and λ_k, respectively. More specifically, we let

$$\lambda'_{k-1} = \frac{\Delta'_{k-1}}{\Delta_{k-1}}\lambda_{k-2} + \frac{\Delta'_k}{\Delta_{k-1}}(\lambda_{k-1}+V),$$

$$\lambda'_k = \frac{\Delta'_{k-1}}{\Delta_{k-1}}(\lambda_{k-1}-V) + \frac{\Delta'_k}{\Delta_{k-1}}\lambda_k.$$

The likelihood ratio is

$$\exp\Big\{(N'_{k-1}\lambda'_{k-1} + N'_k\lambda'_k - N_{k-1}\lambda_{k-1}) + (\beta'_{1k-1}W'_{1k-1}$$

$$+\beta'_{2k-1}W'_{2k-1} + \beta'_{1k}W'_{1k} + \beta'_{2k}W'_{2k} - \beta_{1k-1}W_{1k-1}$$

$$-\beta_{2k-1}W_{2k-1}) + e^{\lambda_{k-1}}S(\tau_{k-1}, \tau_k, \beta_{1k-1}, \beta_{2k-1})$$

$$-e^{\lambda'_{k-1}} S(\tau'_{k-1}, \tau'_k, \beta'_{1k-1}, \beta'_{2k-1}) - e^{\lambda'_k} S(\tau'_k, \tau'_{k+1}, \beta'_{1k}, \beta'_{2k}) \Big\} .$$

The prior ratio is

$$\alpha\,(2\pi)^{-(p+1)/2} \left(\frac{|A'||B'_1||B'_2|}{|A||B_1||B_2|} \right)^{1/2}$$

$$\times \exp\left\{ \frac{1}{2} (\Phi_{\mathrm{B}}(A, A', \mu, \lambda_m, \lambda'_{m'}) + \Phi_{\mathrm{B}}(B_1, B'_1, \mu_1, \beta_{1m}, \beta'_{1m'}) + \Phi_{\mathrm{B}}(B_2, B'_2, \mu_2, \beta_{2m}, \beta'_{2m'}) \right\},$$

where $\Phi_{\mathrm{B}}(A, A', \mu, \lambda_m, \lambda'_{m'})$ is given by

$$(\lambda_{k-2} - \mu)^2 (a_{k-2k-2} - a'_{k-2k-2}) + (\lambda_k - \mu)^2 (a_{kk} - a'_{kk})$$

$$+a_{k-1k-1}(\lambda_{k-1} - \mu)^2 - a'_{k-1k-1}(\lambda'_{k-1} - \mu)^2 - a'_{k+1k+1}(\lambda_k - \mu)^2$$

$$-2(\lambda_{k-2} - \mu)[a'_{k-2k-1}(\lambda'_{k-1} - \mu) - a_{k-2k-1}(\lambda_{k-1} - \mu)]$$

$$-2(\lambda_k - \mu)[a'_{kk+1}(\lambda'_k - \mu) - a_{k-1k}(\lambda_{k-1} - \mu)]$$

$$-2a'_{k-1k}(\lambda'_{k-1} - \mu)(\lambda'_k - \mu).$$

The proposal ratio for this type of move takes the simple form $2^{p+1} \epsilon\, \epsilon_1\, \epsilon_2 / \alpha$, and the Jacobian of the transformation is

$$2^{p+1}\, \epsilon\, \epsilon_1\, \epsilon_2\, (\Delta'_k \Delta'_{k-1})^{p+1} / \Delta_{k-1}^{2(p+1)}.$$

Move of type D An index k is uniformly selected from $\{2, 3, \ldots, m\}$ corresponding to the removal of the jump time τ_k. The proposed new point is $(\tau'_{m'}, \lambda'_{m'}, \beta'_{m'1}, \beta'_{m'2})$, with $\lambda'_i = \lambda_i, \beta'_{iq} = \beta_{iq}$ for $i \leq k-2$, $\lambda'_j = \lambda_{j+1}, \beta'_{jq} = \beta_{j+1,q}, \tau'_j = \tau_{j+1}$ for $j \geq k$, and $\tau'_i = \tau_i$ for $i \leq k-1, q = 1, 2$.

The likelihood, prior, proposal ratios, and the Jacobian for this type of move are the inverse ratios of the ones for the move of type B with the proper labelling of the jump times and the log-hazard levels.

The above type of moves imply that the corresponding Markov chain is π-irreducible. Denote by $\omega_i(t) = (\lambda_i(t), \beta_{i1}(t), \beta_{i2}(t))$ $i = 1, \ldots, N$ the sample paths generated by a run of length N of the MHG sampler. Then for any real valued function $H : \mathcal{S} \to R$ integrable with respect to π, we have almost surely

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} H(\omega_i(t)) = \int_{\mathcal{S}} H((\tau, \omega)) d\pi((\tau, \omega)|\text{data}),$$

see Meyn and Tweedie (1993, Ch. 17). In particular, the posterior mean baseline hazard function $E_\pi(h_0(t)|\text{data})$ is approximated by

$$\frac{1}{N} \sum_{i=1}^{N} \exp(\lambda_i(t)),$$

and the predictive baseline survival function $P_\pi(T > t|\text{data})$ by

$$\frac{1}{N} \sum_{i=1}^{N} \exp\left\{ - \int_0^t \exp(\lambda_i(u)) du \right\}.$$

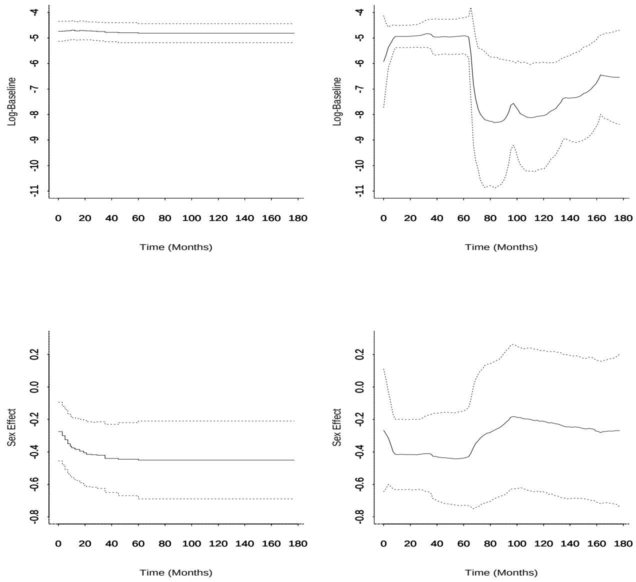

Figure 1: Posterior mean estimates of the log-baseline hazard function and Sex effect.

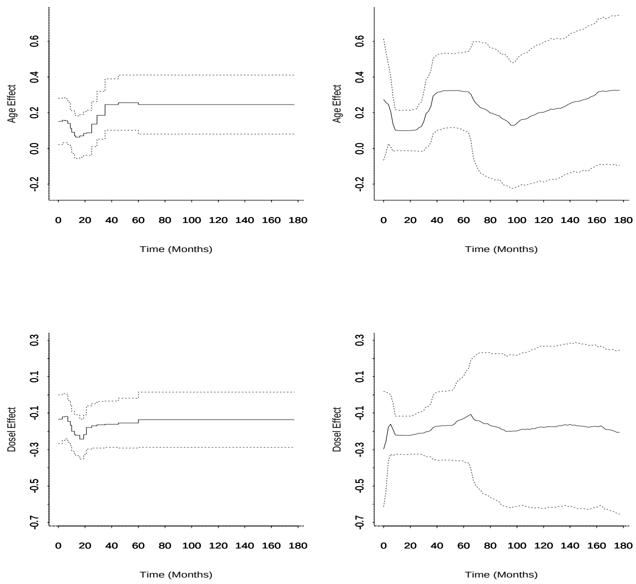

Figure 2: Posterior mean estimates of the Age and DoseI effects.

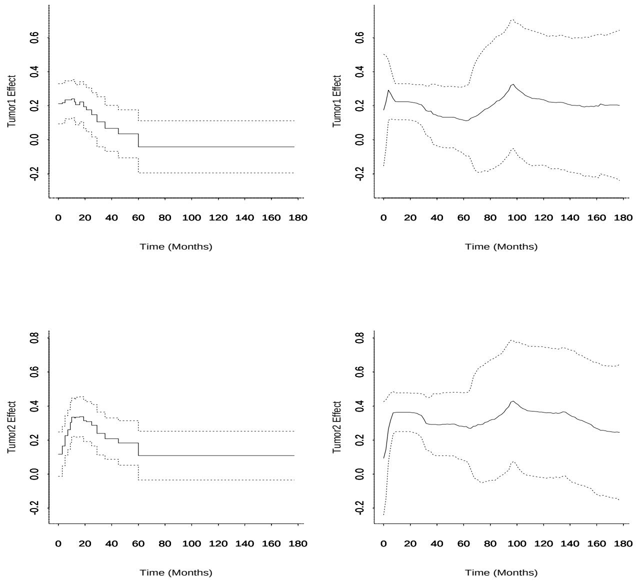

Figure 3: Posterior mean estimates of Tumor1 and Tumor2 effects.

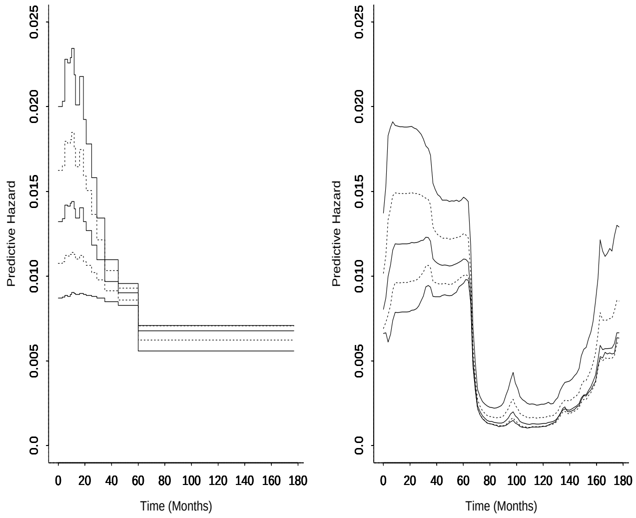

Figure 4: Predictive hazard functions across levels of Tumor1 for a male patient; West's estimate (left), proposed estimate (right).

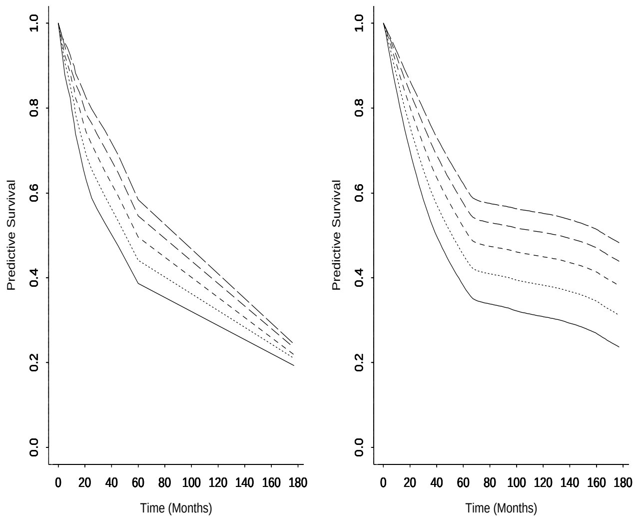

Figure 5: Predictive survival functions across levels of Tumor1 for a male patient; West's estimate (left), proposed estimate (right).