Far East J. Theo. Stat. **16**(2) (2005), 269-284
(A special volume on Biostatistics and related areas)

A METROPOLIS WITHIN GIBBS SAM PLING IN RELATIVE SURVIVAL

ROCH GIORGI, ARIANE SAHEL, JEAN-PIERRE DAURES[*] and JOANNY GOUVERNET

LERTIM, Faculté de Médecine
Université de la Méditerranée
27 Boulevard Jean Moulin, 13385 Marseille Cedex, France
e-mail: roch.giorgi@ap-hm.fr

[*]Laboratoire de Biostatistique
Institut Universitaire de Recherche Clinique
641 Avenue Gaston Giraud, 34093 Montpellier, France

Abstract

Relative survival analysis is a method which provides an estimate of the effect on survival corrected for the effect of other independent causes of death, using the natural mortality in the underlying general population as the reference. This method is frequently used when the specific causes of deaths are uncertain or unavailable as in some population or hospital-based registries. We proposed a Markov Chain Monte Carlo (MCMC) approach to perform relative survival analysis using a proportional hazards regression model. We used gamma and normal prior distributions, respectively, for the baseline mortality hazard function and the regression parameters and we established the likelihood function. Conditional posterior distributions cannot be reduced analytically to well-known distributions and we used a Metropolis within Gibbs sampling to obtain samples from the conditional posterior distributions. The accuracy of the estimates obtained by this MCMC approach were evaluated in simulations

2000 Mathematics Subject Classification: 62Nxx.

Key words and phrases: relative survival, proportional hazards model, Bayesian methods, Gibbs sampling, Metropolis sampling, survival analysis.

Received May 13, 2005

studies. Data from a population-based of pharyngeal cancer were used to illustrate our approach.

1. Introduction

Survival analysis is often conduced using the popular Cox [8] proportional hazards (PH) model, which constrains the hazard ratio, describing the effect of a prognostic factor on disease-specific survival, to be constant over time. However, when the specific causes of death are uncertain or unavailable, as in some population or hospital-based registries, Cox model does not differentiate whether covariates such as age and sex, are strictly related to the disease specific mortality, the natural mortality in the source population, or to both (Monnet et al. [29]). This is an important point in cancer survival since age and sex are used to modify therapeutic approaches. Indeed, comparisons between population or hospital-based prognostic studies are difficult to interpret because of the differences in the natural mortality of the populations (Esteve et al. [10]; Monnet et al. [29]).

These problems can be solved using the so-called relative survival analysis which provides an estimate of the patients' survival corrected for the effect of other independent causes of death, using the natural mortality in the general population (Berkson and Gage [1]). Previous study suggested the importance of this method to fully account for the relationships between prognostic factors and mortality, and to separate the impact of prognostic factors on cancer-related deaths from their effects on other causes of death (Monnet et al. [29]). Several relative survival regression models have been proposed (Buckley [5]; Hakulinen and Tenkanen [17]; Esteve et al. [10]; Sasieni [31]) for which both the baseline mortality hazard function and the regression parameters of the prognostic factors need to be estimated.

Several Bayesian approaches to analysing the classical Cox model have been developed (see for example Kalbfleisch [24]; Hjort [19]; Clayton [7]; Laud et al. [26]; Chen et al. [6]; Ibrahim et al. [23]), and the accuracy of the results and the potential advantages of using Bayesian methods for analysing survival data have been presented. A potential advantage of jointly modelling baseline hazard function and regression parameters of

the prognostic factors comes from the use of Markov Chain Monte Carlo (MCMC) simulation techniques to compute posterior quantities of interest with accuracy.

To our knowledge, the use of MCMC simulation techniques have not yet been applied in the framework of relative survival regression analysis. We proposed a Metropolis within Gibbs sampling scheme for PH relative survival regression model. The model and the algorithm used to sample from the posterior distributions are described in Section 2. In Section 3, we present results of simulations performed under different assumptions about the censoring process and we illustrate our approach using a real data set from a population-based study of pharyngeal cancer. Section 5 contains a brief discussion.

2. The Method

2.1. The proportional hazards relative survival model

In a regressive relative survival model, the observed-mortality hazard, λ_o at time t after diagnosis of an individual aged x at diagnosis with a vector of covariates $\mathbf{z}$ is expressed as

$$\lambda_o(t, \mathbf{z}, x) = \lambda_e(t + x, z_1) + \lambda_b(t)\exp(\boldsymbol{\beta}\mathbf{z}), \quad (1)$$

where λ_e represents the expected hazard function in a general population and depends only on z_1, a subvector of $\mathbf{z}$. The expected hazard function is obtained from relevant mortality statistics using external sources (Esteve et al. [10]) and therefore is usually assumed to be known (in our study, λ_e is quantified based on published age-and-sex specific mortality rates, hence, x is age at diagnosis and z_1 is sex). Terms on the right side of equation (1) represent the disease-related mortality hazard function, where $\lambda_b(t)$, usually modelled parametrically, is the disease-related mortality hazard function, i.e., the mortality hazard at time t for patients with $\mathbf{z} = 0$, and where $\boldsymbol{\beta}$ is a vector of regression parameters. In this model, according to the PH assumption, the covariate effect on the disease-related hazard ratio is assumed to be constant over time.

Thus for a Bayesian purpose we have to establish the likelihood function and to specify prior distributions for both the regression parameters and the baseline mortality hazard function (the expected hazard function is known and obtained from relevant mortality statistics).

2.2. Prior distributions

Several prior distributions have been used for the baseline mortality hazard function in the Cox [8] PH model (Kalbfleisch [24]; Clayton [7]; Hjort [19]; Damien et al. [9]; Laud et al. [26]; Chen et al. [6]; Ibrahim et al. [23]). Among these prior distributions, gamma prior distribution has shown its efficiency and its ability to model the baseline mortality hazard (Ibrahim and Chen [20]; Ibrahim et al. [22]; Chen et al. [6]). To implement gamma prior distribution for the baseline mortality hazard function $(\lambda_b(t))$ in the regressive relative survival PH model, we have to divide the time axis into M intervals such as $0 \leq s_0 < s_1 < \cdots < s_M$, with $s_M > t_n$ for all the n individuals of the data set $(i = 1, ..., n)$. In our study, the partitioning of the time axis is based on the quantiles of the distribution of the observed failure times in order to ensure an approximate equal number of failures in each of the intervals $]s_{k-1}, s_k]$. Denoted $\delta_k = \lambda_b(s_k) - \lambda_b(s_{k-1})$ the increment in the baseline mortality hazard in the interval $]s_{k-1}, s_k]$, $k = 1, ..., M$. Assuming that the baseline mortality hazard is random and it remains constant within each of the intervals, the δ_k's are random variables, independent *a priori*, and have gamma distributions with shape and scale parameters defined, respectively, by $s_k - s_{k-1}$ and ρ_k. Therefore, the joint prior density of $\boldsymbol{\Delta} = (\delta_1, ..., \delta_M)$ is the product of M independent gamma distributions with form parameters $s_k - s_{k-1}$ and scale parameter ρ_k.

Elicitation of pre-data knowledge about the regression coefficient results in an appropriate choice of an informative or noninformative prior distribution. Normal (informative) prior has proved to be a flexible and useful class of priors for many regression problems even if, in information theories of $\mathbb{R}$, normal prior is the less informative prior (Bernardo and Smith [2]). Furthermore, this choice allows the use of information coming

from previous studies measuring the same covariates as in the current study. Therefore, we choose a multivariate normal prior for $\boldsymbol{\beta}$.

Assuming an *a priori* independence between the baseline mortality hazard function and the regression parameters, the joint prior density of $(\boldsymbol{\beta}, \boldsymbol{\Delta})$ is $\pi_1(\beta)\pi_2(\Delta)$, where $\pi(\cdot)$ is the prior distribution.

2.3. The likelihood function

The cumulative distribution function determined by equation (1) at time s is

$$F(s) = 1 - \exp\left\{-\int_0^s (\lambda_e(t) + \lambda_b(t)\exp(\boldsymbol{\beta}\mathbf{z}))dt\right\}$$

$$\cong 1 - \exp\left\{-\int_0^s \lambda_e(t)dt - \exp(\boldsymbol{\beta}\mathbf{z})\left((s-s_0)^+\lambda_b(s_0) + \sum_{k=1}^{M}\delta_k(s-s_{k-1})^+\right)\right\}, \quad (2)$$

where $\lambda_e(t) = \lambda_e(t + x, z_1)$, $(t)^+ = t$ if $t > 0$, 0 otherwise. Assuming that $\lambda_b(s_0) = 0$ and $F(s) = 1$ for $s > s_M$ allows simplifications in (2). The probability of a failure in the interval $]s_{k-1}, s_k]$, $k = 1, ..., M$, is thus $p_k = F(s_k) - F(s_{k-1})$ and since

$$\sum_{j=1}^{k}\delta_j(s_k - s_{j-1}) = (s_k - s_{k-1})\sum_{j=1}^{k}\delta_j + \sum_{j=1}^{k-1}\delta_j(s_{k-1} - s_{j-1})$$

we obtain

$$p_k = \exp\left\{-\left(\int_0^{s_{k-1}}\lambda_e(t)dt + \exp(\boldsymbol{\beta}\mathbf{z})\sum_{j=1}^{k-1}\delta_j(s_{k-1} - s_{j-1})\right)\right\}$$

$$\times\left[1 - \exp\left\{-\left(\int_{s_{k-1}}^{s_k}\lambda_e(t)dt + \exp(\boldsymbol{\beta}\mathbf{z})(s_k - s_{k-1})\sum_{j=1}^{k}\delta_j\right)\right\}\right].$$

In the kth interval, the contribution to the likelihood function is p_k for a failure and $S(s_k) = 1 - F(s_k)$ for a right censored observation. For notational convenience, observations are ordered so that in each of the

kth intervals the first d_k observations are failures and the remaining c_k are censored. Thus, the likelihood function over all M intervals is

$$L(\boldsymbol{\beta}, \boldsymbol{\Delta} \mid Data) = \prod_{k=1}^{M} \prod_{i=1}^{d_k} p_k^i \prod_{k=1}^{M} \prod_{i=d_{k+1}}^{d_k+c_k} S(s_i^k)$$

and, after some algebra, we obtain

$$L(\boldsymbol{\beta}, \boldsymbol{\Delta} \mid Data) = \left(\prod_{k=1}^{M} \prod_{i=1}^{d_k} \exp\left\{ - \int_0^{s_k} \lambda_{e,i}(t) dt \right\} \right) \left(\prod_{k=1}^{M} \exp\{ - \delta_k (a_k + b_k) \} \right)$$

$$\times \left(\prod_{k=1}^{M} \prod_{i=1}^{d_k} \left[\exp\left\{ \int_{s_{k-1}}^{s_k} \lambda_{e,i}(t) dt \right\} - \exp\{ - u_{ki}(\boldsymbol{\beta}) T_k(\boldsymbol{\Delta}) \} \right] \right), \quad (3)$$

where $\lambda_{e,i}(t)$ is the expected hazard for the ith individual, and where

$$u_{ki}(\boldsymbol{\beta}) = \exp(\boldsymbol{\beta} z_{ki}), \quad a_k = \sum_{j=k+1}^{M} \sum_{i=1}^{d_j} u_{ki}(\boldsymbol{\beta})(s_{j-1} - s_{k-1}),$$

$$b_k = \sum_{j=k}^{M} \sum_{i=d_j+1}^{d_j+c_j} u_{ki}(\boldsymbol{\beta})(s_j - s_{k-1}), \quad T_k(\boldsymbol{\Delta}) = (s_k - s_{k-1}) \sum_{j=1}^{k} \delta_j.$$

2.4. The Metropolis within Gibbs sampling

In the Bayesian framework, posterior inference can be implemented using Markov Chain Monte Carlo methods. Samples from $\pi(\boldsymbol{\beta}, \boldsymbol{\Delta} \mid Data)$ can be obtained using Gibbs sampling (Geman and Geman [12]; Gelfand and Smith [11]), which require to sample iteratively from (a) $\pi(\Delta \mid \boldsymbol{\beta}, Data)$ and then from (b) $\pi(\boldsymbol{\beta} \mid \boldsymbol{\Delta}, Data)$, i.e., new sample of $\boldsymbol{\Delta}$ is used to sample from (b) and the new sample of $\boldsymbol{\beta}$ obtained in this way is used to sample from (a), and so on.

The conditional posterior distribution of $\boldsymbol{\Delta} = (\delta_1, ..., \delta_M)$ is given by

$$\pi(\Delta \mid \beta, Data) \propto L(\beta, \Delta \mid Data) \pi_2(\Delta) \quad (4)$$

and this of $\boldsymbol{\beta}$ is given by

$$\pi(\beta \mid \Delta, \mathit{Data}) \propto L(\beta, \Delta \mid \mathit{Data})\pi_1(\beta). \tag{5}$$

However, in our case, both conditional posterior distributions of $\boldsymbol{\Delta}$ (equation (4)) and $\boldsymbol{\beta}$ (equation (5)) cannot be reduced analytically to well-known distributions and therefore it is not possible to sample directly by standard methods. As suggested by Tierney [32], a common way to solve this problem is to used a Metropolis sampling (Metropolis et al. [28]) within the Gibbs sampling scheme in order to obtain samples from the conditional posterior distributions: to sample from (4) we generated a proposal value from a multivariate normal distribution $N(\boldsymbol{\Delta}^{(c)}, \boldsymbol{\kappa}_{\boldsymbol{\Delta}}\mathbf{V}_{\boldsymbol{\Delta}})$, and to sample from (5) we generated a proposal value from a multivariate normal distribution $N(\boldsymbol{\beta}^{(c)}, \boldsymbol{\kappa}_{\boldsymbol{\beta}}\mathbf{V}_{\boldsymbol{\beta}})$, where $\boldsymbol{\Delta}^{(c)}$ and $\boldsymbol{\beta}^{(c)}$ are, respectively, the current values of $\boldsymbol{\Delta}$ and $\boldsymbol{\beta}$, $\mathbf{V}_{\boldsymbol{\Delta}}$ and $\mathbf{V}_{\boldsymbol{\beta}}$ are variances-covariances matrix, and where $\boldsymbol{\kappa}_{\boldsymbol{\Delta}}$ and $\boldsymbol{\kappa}_{\boldsymbol{\beta}}$ are scaling factors (the detail of the algorithm is given in Appendix). For the Metropolis algorithm to work effectively, $\boldsymbol{\kappa}_{\boldsymbol{\beta}}$, $\boldsymbol{\kappa}_{\boldsymbol{\Delta}}$, $\mathbf{V}_{\boldsymbol{\beta}}$ and $\mathbf{V}_{\boldsymbol{\Delta}}$ need to be chosen carefully: the scaling factors must be chosen so that the acceptance ratio (the proportion of the proposed values from the posterior distribution accepted by the algorithm) was around 0.3 and the variances-covariances matrix may be elicited using a priori information about the variances-covariances matrix (Roberts [30]).

3. Results

3.1. Simulation studies

We conduced simulation studies to assess our sampling method. Survival data was generated from an exponential distribution assuming that the observed-mortality hazard function is the sum of the expected hazard function in a general population and of the disease-related mortality hazard function. The general population all-causes mortality was assumed to depend only on age and gender and the disease-related mortality hazard function was assumed to depend on two binary

prognostic factors (z_1 and z_2, which are independent of age and sex), having effects $\beta_1 = \ln(1.5)$ and $\beta_2 = \ln(2)$ on the disease-related part of the hazard function.

Two scenarios were considered when determining the "true" distribution of the binary covariates. Firstly, we assumed $\text{corr}(z_1, z_2) = 0$ and z_1 and z_2 were generated as Bernoulli variables with parameter 1/2 each. In the second scenario, we assumed $\text{corr}(z_1, z_2) = 0.7$ and the two covariates were generated from the binary distributions with $P[z_1 = 0] = P[z_1 = 1] = 0.5$ and $P[z_2 = 1 | z_1 = 1] = 0.7$ for z_1 and z_2, respectively.

Censoring times were always generated from the exponential distribution, with hazard selected so as to obtain approximately 15% or 50% overall censoring level. Then, individual's observed time was determined as $T_i = \min(S_i, C_i)$, where S_i and C_i denoted the individual's survival and censoring time, respectively. For each case, we generated 1000 random samples of size 100.

For all generated samples, analysed independently of each other, we used starting values and prior parameters close to the true values of regression parameters. These choices are pragmatic insofar as analysis based on real data set frequently used maximum likelihood estimates for priors (Mallick et al. [27]; Kozumi [25]). The prior for the δ_k was $G(s_k - s_{k-1}, 1)$, with $]s_{k-1}, s_k]$ chosen to contain approximately the same number of failures.

Since some variability may occur in simulated data, the length of the chain was chosen so that stabilisation was achieved. Thanks to preliminary tests (data not shown), we ran the sampler for an initial burn-in period (2000 iterations), after which some stabilisation was appeared, and estimations was based on the last of the sample (45000 iterations), where the stabilisation was achieved.

Table 1. Results of simulation studies obtained with MCMC and Esteve et al. [10] Maximum Likelihood Method (MLM) according to the right random censorship and to the correlation of (z_1, z_2)

	True values		15% of censor		50% of censor	
			MCMC Method	MLM	Bayesian Method	MLM
$\text{corr}(z_1, z_2) = 0$	$\beta_1 = \ln(1.5)$	Bias[a]	-0.031	-0.003	-0.116	-0.051
		SD[b]	0.173	0.251	0.141	0.370
		ASD[c]	0.183	0.246	0.204	0.351
		ECR[d]	95.7%	94.5%	97.2%	94.7%
	$\beta_2 = \ln(2)$	Bias[a]	0.035	-0.020	-0.096	-0.111
		SD[b]	0.179	0.273	0.140	0.367
		ASD[c]	0.182	0.252	0.203	0.356
		ECR[d]	95.4%	93.1%	97.5%	92.5%
$\text{corr}(z_1, z_2) = 0.7$	$\beta_1 = \ln(1.5)$	Bias[a]	-0.043	-0.018	-0.161	-0.081
		SD[b]	0.125	0.455	0.128	0.606
		ASD[c]	0.164	0.426	0.183	0.798
		ECR[d]	98.9%	93.7%	93.7%	96.7%
	$\beta_2 = \ln(2)$	Bias[a]	-0.007	0.000	-0.134	-0.166
		SD[b]	0.126	0.450	0.130	0.637
		ASD[c]	0.166	0.424	0.185	0.817
		ECR[d]	99.5%	94.7%	97.2%	92.2%

[a]The bias corresponds to $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$, where $\hat{\boldsymbol{\beta}}$ represents the mean of the estimates of the true value $\boldsymbol{\beta}$.

[b]SD denotes the sample standard deviation of the estimates.

[c]ASD denotes the average of the standard error estimates.

[d]ECR denotes the proportion of samples in which the 95% credible interval, or the 95% confidence interval includes the true value $\boldsymbol{\beta}$.

Table 1 shows that the bias for Bayesian estimates is minor with 15% of right censored data and, generally, Bayesian estimates and maximum likelihood estimates (MLE) obtained with the Esteve et al. [10] PH model are not very different. As expected, more the censorship increases and

more the bias is greater. Furthermore, the accuracy is better with the MCMC method (the sample standard deviation (SD) of the estimates ranging from 0.125 to 0.179), while, considering the previous case, SD obtained from the MLE method ranging from 0.251 to 0.637. The average of the standard error estimates obtained with our MCMC method is smaller than those obtained from the Esteve et al. method. Finally, the proportion of samples in which the 95% credible interval includes the true value of the parameter is greater than 95% in almost all the cases with the MCMC method (the smaller value is 93.7%). The proportion of samples in which the 95% confidence interval includes the true value of the parameter is between 92.2% and 96.7% with the frequentist Esteve et al. method.

3.2. Analyses of pharyngeal cancer data

Here, we illustrate our approach using a subset of previously analysed population-based survival data (Giorgi et al. [16]) for 306 incident cases of pharyngeal male cancer, diagnosed between 1985 and 1987 and identified by the Bas-Rhin cancer registry (France). The analysis focus on the three prognostic factors of the Tumour Nodes Metastasis (TNM) classification (size of the tumour: *T*1-*T*2 *vs* *T*3-*T*4; nodal invasion: absent *vs* present: metastasis extension: absent *vs* present; see Table 2).

Table 2. Description of the pharyngeal male cancer prognostic factors analysed

Prognostic	Factors	Number	(Percent)*	Deaths at 5 years	(%)†
T	*T*1-*T*2	152	(49.7%)	97	(43.7%)
	*T*3-*T*4	154	(50.3%)	125	(56.3%)
N	Absent	122	(39.9%)	74	(33.3%)
	Present	184	(60.1%)	148	(66.7%)
M	Absent	293	(95.7%)	209	(94.1%)
	Present	13	(4.3%)	13	(5.9%)
Overall		306	(100%)	222	(72.5%)

*Percent of all 306 patients.

†Percent of patients, in a given category, who died within first 5 years after diagnosis.

Bayesian estimates of the regression parameter for the TNM prognostic factors were obtained using the MCMC algorithm described in Section 2. The MCMC simulation was run for an initial burn-in period of 2000 iterations and estimations were based on the last 45000 iterations. Convergence of the chain was achieved using visual inspection of its trace plus the Heidelberger and Welch [18] convergence diagnostic, a robust convergence diagnostic test limiting calculus using just a single chain and provided with some others tests in the free software CODA (Best et al. [3]).

Table 3. Estimates of adjusted* regression parameters, with standard deviation, obtained for relative survival, with our MCMC method and the Esteve et al. [10] Maximum Likelihood Method (MLM), and for crude survival, with the Cox PH model [8], in a population-based study of 306 French pharyngeal male cancer

	Relative survival				Crude survival	
	MCMC Method		Esteve et al. MLM		Cox model	
Prognostic factors	Estimate	SD	Estimate	SD	Estimate	SD
*T*3-*T*4†	0.515	0.158	0.474	0.146	0.404	0.140
N present†	0.695	0.165	0.663	0.156	0.558	0.147
M present†	1.269	0.198	1.168	0.295	0.937	0.302

*All estimates are adjusted for age.

†Referent categories are: *T*1-*T*2, *N* absent and *M* absent, respectively.

Table 3 shows the values of the regression coefficients, with standard deviation, obtained from the relative survival analyses. The prognostic effects of these three factors on relative survival are important and significant (all the three 95% credible intervals, and the three confidence intervals, do not include 0). Bayesian estimates and Esteve et al. maximum likelihood estimates are in the same way, but, as in simulation studies, these latter are slightly lower (discrepancy ranging from 0.03 to 0.15). Thus, we can assume that estimates obtained from the maximum likelihood method are under-estimates. Because there is few cases with metastasis (Table 2), the estimate of the effect of metastasis prognostic factor is less accurate than estimates of the other prognostic factors.

Table 3 compares also estimates adjusted for age obtained with the relative survival methods (the MCMC method and the Esteve et al. [10] maximum likelihood method) to those obtained with the crude survival method using the Cox [8] model. It shows that relative survival approach changes the estimated effects of TNM prognostic factors compared with Cox PH estimates. Similar results, in other clinical applications, have been found before (Monnet et al. [29]; Bolard et al. [4]; Giorgi et al. [15]) suggesting that relative survival approach is important to fully account for the relationships between prognostic factors and mortality, and to separate their impact on cancer-related deaths from their effects on other causes of death.

4. Discussion

In this article, we propose a Markov Chain Monte Carlo approach to perform relative survival analysis using a proportional hazards regression model. A Metropolis within Gibbs sampling scheme allows us to obtain estimates of the regression parameters whereas our posterior distributions cannot be reduced to well-known distributions. Simulations indicate that our estimates are reasonably unbiased, even when the covariates are correlated between them and when the censoring level increases. Furthermore, compared to the maximum likelihood method, our MCMC method reduces the variability of the estimates. Our application to pharyngeal cancer data shows the differences between results of crude survival approach and results of the relative survival, which allows to separate the effect of prognostic factors on cancer-related deaths from their effects on other causes of death.

As explained previously, we could not sample directly from the posterior distributions implying the use of MCMC techniques. In order to improve our algorithm, we tried to reduce posterior distributions to well-known distributions using latent variables, as in Chen et al. [6] for the Cox model, but without success. Another way to have algorithm less computationally expensive is to use adaptive rejection sampling (Gilks [13]) which requires the posterior distribution to be log-concave. Nevertheless, in our relative survival model the posterior distribution given in (5) is not log-concave and consequently we could not use such

algorithms. If necessary, a possible improvement of the algorithm used for this work will consist, for example, on the use of the adaptive rejection Metropolis sampling (ARMS) of Gilks et al. [14] which have the advantage to deal with non-log-concavity.

As for all Bayesian methods, results can depend on the choice of the prior parameters. There are several ways for prior elicitation: one can use literature knowledge, MLE of the current data set (Mallick et al. [27]; Kozumi [25]), or estimates obtained from a previous study similar to the current study (Ibrahim et al. [23]). Weak informative prior distribution was used in our simulation studies and results suggested that our method performs well. It is not the purpose of this paper to perform a robustness analysis of the prior distribution of the regression parameters what could be done, for example, using a mixing of unimodal symmetric distributions. The use of a gamma prior distribution for the baseline mortality hazard function implies to divided the time axis into M intervals and several schemes for the choice of the subintervals $]s_{k-1}, s_k]$, $k = 1, ..., M$, could be considered in the way to have, for example: (i) approximately the same number of failures and/or censored, (ii) equal lengths with at least one failure observation in each intervals or (iii) a decreasing number of failures. To access the impact of the choice of the subintervals, we conduced another analysis on the pharyngeal data with $]s_{k-1}, s_k]$ having approximately equal lengths, with at least one failure observation in each intervals (results not shown). As in Ibrahim and Chen [21], which used an extended Gamma process prior for the baseline mortality hazard function in the Cox PH model and which considered different partitioning schemes for the time axis, our estimates seem not to be too sensitive to the choice of the subintervals. This is convenient since it allows the users some flexibility in the choice of the subintervals.

The R/S-PLUS code for implementing our Bayesian method is freely available on our Web site at http://cybertim.timone.univ-mrs.fr/LERTIM /Recherche/RSBayes.txt.

Appendix

The Metropolis within Gibbs sampling algorithm is as follows:

0. Set $c = 0$.

1. Initialise $\boldsymbol{\Delta}^{(c)}$ and $\boldsymbol{\beta}^{(c)}$.

2. Generate $\boldsymbol{\Delta}^*$ from $N(\boldsymbol{\Delta}^{(c)}, \boldsymbol{\kappa}_{\boldsymbol{\Delta}}\mathbf{V}_{\boldsymbol{\Delta}})$.

Generate u from uniform (0, 1) and

if $u \leq \min\left[1, \dfrac{\pi(\boldsymbol{\Delta}^* \mid \boldsymbol{\beta}^{(c)}, Data)}{\pi(\boldsymbol{\Delta}^{(c)} \mid \boldsymbol{\beta}^{(c)}, Data)}\right]$, then

set $\boldsymbol{\Delta}^{(c+1)} = \boldsymbol{\Delta}^*$.

Otherwise

set $\boldsymbol{\Delta}^{(c+1)} = \boldsymbol{\Delta}^{(c)}$.

3. Generate $\boldsymbol{\beta}^*$ from $N(\boldsymbol{\beta}^{(c)}, \boldsymbol{\kappa}_{\beta}\mathbf{V}_{\beta})$.

Generate u from uniform (0, 1) and

if $u \leq \min\left[1, \dfrac{\pi(\boldsymbol{\beta}^* \mid \boldsymbol{\Delta}^{(c+1)}, Data)}{\pi(\boldsymbol{\beta}^{(c)} \mid \boldsymbol{\Delta}^{(c+1)}, Data)}\right]$, then

set $\boldsymbol{\beta}^{(c+1)} = \boldsymbol{\beta}^*$.

Otherwise

set $\boldsymbol{\beta}^{(c+1)} = \boldsymbol{\beta}^{(c)}$.

4. Set $c = c + 1$.

Repeat Step 2 to Step 4 a sufficiently long time, until the process becomes stationary .

Acknowledgement

We are grateful to Professor P. Schaffer of the Bas-Rhin cancer registry (France) for providing data of pharyngeal cancer.

References

[1] J. Berkson and R. P. Gage, Calculation of survival rates for cancer, Proceedings of the Staff Meetings of the Mayo Clinic 25 (1950), 270-286.

[2] J. M. Bernardo and A. F. M. Smith, Bayesian Theory, Wiley, Chichester, 1993.

[3] N. G. Best, M. K. Cowles and S. K. Vines, CODA Convergence Diagnosis and Output Analysis Software for Gibbs Sampler Output, Version 0.3, MRC Biostatistics Unit, Cambridge, UK, 1996.

[4] P. Bolard, C. Quantin, J. Esteve, J. Faivre and M. Abrahamowicz, Modelling time-dependent hazard ratios in relative survival, Application to colon cancer, Journal of Clinical Epidemiology 54 (2000), 986-996.

[5] J. D. Buckley, Additive and multiplicative models for the relative survival rates, Biometrics 39 (1983), 941-948.

[6] M. H. Chen, Q. M. Shao and J. G. Ibrahim, Bayesian computation for proportional hazards models, Monte Carlo Methods in Bayesian Computation, M. H. Chen, Q. M. Shao and J. G. Ibrahim, eds., Springer, New York, 2000.

[7] D. G. Clayton, A Monte Carlo method for Bayesian inference in frailty models, Biometrics 47 (1991), 467-485.

[8] D. R. Cox, Regression models and life tables, J. Roy. Statist. Soc. Ser. B 34 (1972), 187-220.

[9] P. Damien, P. W. Laud and A. F. M. Smith, Implementation of Bayesian nonparametric inference based on beta processes, Scand. J. Statist. 23 (1996), 27-36.

[10] J. Esteve, E. Benhamou, M. Croasdale and L. Raymond, Relative survival and the estimation of net survival: elements for further discussion, Statistics in Medicine 9 (1990), 529-538.

[11] A. E. Gelfand and A. F. M. Smith, Sampling-based approaches to calculating marginal densities, J. Amer. Statist. Assoc. 85 (1990), 398-409.

[12] S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, I.E.E.E. Transactions on Pattern Analysis and Machine Intelligence 6 (1984), 721-741.

[13] W. R. Gilks, Derivative-free adaptive rejection sampling for Gibbs sampling, Bayesian Statistics 4, J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds., Clarendon Press, Oxford, U. K., 1992.

[14] W. R. Gilks, N. G. Best and K. K. C. Tan, Adaptive rejection Metropolis sampling within Gibbs sampling, Applied Statistics 44 (1995), 455-472.

[15] R. Giorgi, M. Abrahamowicz, C. Quantin, P. Bolard, J. Esteve, J. Gouvernet and J. Faivre, A relative survival regression model using *B*-splines functions, Statistics in Medicine 22 (2003), 2767-2784.

[16] R. Giorgi, H. Hedelin and P. Schaffer, Impact of therapeutic changes in relative survival for upper aerodigestive tract cancer in the Bas-Rhin region between 1975-1977 and 1985-1987, Journal Français d' ORL 48 (1999), 401-408.

[17] T. Hakulinen and L. Tenkanen, Regression analysis of survival rates, Applied Statistics 36 (1987), 309-317.

[18] P. Heidelberger and P. Welch, Simulation run length control in the presence of an initial transient, Oper. Res. 31 (1983), 1109-1144.

[19] N. L. Hjort, Nonparametric Bayes estimation based on beta processes in models for life history data, Ann. Statist. 18 (1990), 1259-1294.

[20] J. G. Ibrahim and M. H. Chen, Prior distributions and Bayesian computation for proportional hazards models, Sankhya Ser. B 60 (1998), 48-64.

[21] J. G. Ibrahim and M. H. Chen, Bayesian methods for variable selection in the Cox model, Generalized Linear Models: A Bayesian Perspective, D. Dey, S. K. Ghosh and B. K. Mallick, eds., Marcel Dekker, New York, 2000.

[22] J. G. Ibrahim, M. H. Chen and S. N. MacEachern, Bayesian variable selection for proportional hazards models, Canad. J. Statist. 27 (1999), 701-717.

[23] J. G. Ibrahim, M. H. Chen and D. Sinha, Bayesian Survival Analysis, Springer-Verlag, New York, 2001.

[24] J. D. Kalbfleisch, Nonparametric Bayesian analysis of survival time data, J. Roy. Statist. Soc. Ser. B 40 (1978), 214-221.

[25] H. Kozumi, Bayesian analysis of discrete survival data with a hidden Markov chain, Biometrics 56 (2000), 1002-1006.

[26] P. W. Laud, P. Damien and A. F. M. Smith, Bayesian nonparametric and covariate analysis of failure time data, Practical Nonparametric and Semiparametric Bayesian Statistics, D. Dey, P. Mueller and D. Sinha, eds., Springer-Verlag, New York, 1998.

[27] B. K. Mallick, D. G. Denison and A. F. Smith, Bayesian survival analysis using a MARS model, Biometrics 55 (1999), 1071-1077.

[28] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, Equation of state calculations by fast computing machines, Journal of Chemical Physics 21 (1953), 1087-1092.

[29] E. Monnet, M.C. Boutron, P. Arveux, C. Milan and J. Faivre, Different multiple regression models for estimating survival: use in a population-based series of colorectal cancers, Journal of Clinical Epidemiology 45 (1992), 267-273.

[30] G. O. Roberts, Markov chain concepts related to sampling algorithms, Markov Chain Monte Carlo in Practice, W. R. Gilks, S. Richardson and D. J. Spiegelhalter, eds., Chapman & Hall, London, 1995.

[31] P. D. Sasieni, Proportional excess hazards, Biometrika 83 (1996), 127-141.

[32] L. Tierney, Markov chains for exploring posterior distributions, Ann. Statist. 22 (1994), 1701-1762.

■