A Metropolis within Gibbs sampling in relative survival

Roch GIORGI¹, Ariane SAHEL¹, Jean-Pierre DAURES², Joanny GOUVERNET¹

¹LERTIM, Faculté de Médecine, Université de la Méditerranée, 27 Boulevard Jean Moulin 13385 Marseille Cedex, France.
²Laboratoire de Biostatistique, Institut Universitaire de Recherche Clinique, 641 Avenue Gaston Giraud, 34093 Montpellier, France.

Corresponding Author:
Roch GIORGI
LERTIM
Faculté de Médecine, Université de la Méditerranée
27 Boulevard Jean Moulin
13385 Marseille Cedex, France
Tel number: 33 (0) 491 324 216
Fax number: 33 (0) 491 324 639
Email: roch.giorgi@ap-hm.fr

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

Key words: Relative survival; Proportional hazards model; Bayesian methods; Gibbs sampling; Metropolis sampling; Survival analysis.

1. Introduction
Survival analyses are often conducted using the popular Cox (1972) 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., 1992). 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., 1990, Monnet et al., 1992).

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, 1950). 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., 1992). Several relative survival regression models have been proposed (Buckley, 1983; Hakulinen and Tenkanen, 1987; Esteve et al., 1990; Sasieni, 1996) 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, 1978; Hjort, 1990; Clayton, 1991; Laud et al., 1998; Chen et al., 2000; Ibrahim et al., 2001), 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, $\lambda_o$, at time $t$ after diagnosis of an individual aged $x$ at diagnosis with a vector of covariates $z$ is expressed as

$$\lambda_o(t, z, x) = \lambda_e(t + x, z) + \lambda_b(t) \exp(\beta z)$$  \hspace{1cm} (1)

where $\lambda_e$, represents the expected hazard function in a general population and depends only on $z_1$, a subvector of $z$. The expected hazard function is obtained from relevant mortality statistics using external sources (Esteve et al., 1990) and therefore is usually assumed to be known (in our study, $\lambda_e$ 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 $z = 0$, and where $\beta$ 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 (1972) PH model (Kalbfleisch, 1978; Clayton, 1991; Hjort, 1990; Damien et al., 1996; Laud et al., 1998; Chen et al., 2000; Ibrahim et al., 2001). Among these prior distributions, gamma prior distribution has shown its efficiency and its ability to model the baseline mortality hazard (Ibrahim and Chen, 1998; Ibrahim et al., 1999; Chen et al., 2000).
2000). 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 \(0 \leq s_0 < s_1 < \ldots < s_M\), with \(s_M > t_u\) for all the \(n\) individuals of the data set \((i = 1, \ldots, n)\). 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 \([s_{k-1}, s_k]\), \(k = 1, \ldots, M\). Assuming that the baseline mortality hazard is random and it remains constant within each of the intervals, the \(\delta_k\)'s are random variables, independent \textit{a priori}, and have gamma distributions with shape and scale parameters defined, respectively, by \(s_k - s_{k-1}\) and \(\rho_k\). Therefore, the joint prior density of \(\Delta = (\delta_1, \ldots, \delta_M)\) is the product of \(M\) independent gamma distributions with form parameters \(s_k - s_{k-1}\) and scale parameter \(\rho_k\).

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, 1993). 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 \(\beta\).

Assuming an \textit{a priori} independence between the baseline mortality hazard function and the regression parameters, the joint prior density of \((\beta, \Delta)\) 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

\[
F(s) = 1 - \exp\left\{ - \int_0^s (\lambda_b(t) + \lambda_z(t) \exp(\beta z)) dt \right\} \\
\approx 1 - \exp\left\{ - \int_0^s \lambda_b(t) dt - \exp(\beta z) \left( s - s_0 \right)^* \lambda_b(s_0) + \sum_{k=1}^M \delta_k (s - s_{k-1})^* \right\}
\]

(2)

where \( \lambda_i(t) = \lambda(z_i, t + z_1) \), \( t \) if \( t > 0 \), 0 otherwise. Assuming that \( \lambda(s_i) = 0 \) 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

\[
p_k = \exp \left\{ - \left( \int_{s_{k-1}}^{s_k} \lambda_i(t) dt + \exp(\beta z_i) \sum_{j=1}^{k-1} \delta_j (s_k - s_{j-1}) \right) \right\}
\]

\[
\times \left[ 1 - \exp \left\{ - \left( \int_{s_{k-1}}^{s_k} \lambda_i(t) dt + \exp(\beta z_i) (s_k - s_{k-1}) \sum_{j=1}^{k} \delta_j \right) \right\} \right]
\]

In the \( k \)th 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 \( k \)th intervals the first \( d_k \) observations are failures and the remaining \( c_k \) are censored. Thus, the likelihood function over all \( M \) intervals is

\[
L(\beta, \Delta | Data) = \prod_{k=1}^{M} \prod_{i=1}^{d_k} p_k \prod_{k=1}^{M} \prod_{i=d_k+1}^{d_k+c_k} S(s_k) \]

and, after some algebra, we obtain

\[
L(\beta, \Delta | Data) = \left( \prod_{k=1}^{M} \prod_{i=1}^{d_k} \exp \left\{ - \int_{0}^{s_k} \lambda_i(t) dt \right\} \right) \left( \prod_{k=1}^{M} \exp \left\{ - \delta_i (a_k + b_k) \right\} \right)
\]

\[
\times \left( \prod_{k=1}^{M} \prod_{i=d_k+1}^{d_k+c_k} \left[ \exp \left\{ \int_{s_{k-1}}^{s_k} \lambda_i(t) dt \right\} - \exp \left\{ - u_{kl} (\beta) T_k(\Delta) \right\} \right] \right)
\]

where \( \lambda_i(t) \) is the expected hazard for the \( i \)th individual, and where

\[
u_{kl}(\beta) = \exp(\beta z_{kl}), \quad a_k = \sum_{j=d_k+1}^{d_k+c_k} \sum_{i=1}^{d_j} \nu_{kl}(\beta) (s_{j-1} - s_{k-1}),
\]

\[
b_k = \sum_{j=d_k+1}^{d_k+c_k} \sum_{i=1}^{d_j} \nu_{kl}(\beta) (s_{j-1} - s_{k-1}), \quad T_k(\Delta) = (s_k - s_{k-1}) \sum_{j=1}^{k} \delta_j
\]
2.4. The Metropolis within Gibbs sampling

In the Bayesian framework, posterior inference can be implemented using Markov chain Monte Carlo methods. Samples from \( \pi(\beta, \Delta \mid Data) \) can be obtained using Gibbs sampling (Geman and Geman, 1984; Gelfand and Smith, 1990), which requires to sample iteratively from (a) \( \pi(\Delta \mid \beta, Data) \) and then from (b) \( \pi(\beta \mid \Delta, Data) \), i.e. new sample of \( \Delta \) is used to sample from (b) and the new sample of \( \beta \) obtained in this way is used to sample from (a), and so on.

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

\[
\pi(\Delta \mid \beta, Data) \propto L(\beta, \Delta \mid Data)\pi_2(\Delta) \tag{4}
\]

and this of \( \beta \) is given by

\[
\pi(\beta \mid \Delta, Data) \propto L(\beta, \Delta \mid Data)\pi_1(\beta) \tag{5}
\]

However, in our case, both conditional posterior distributions of \( \Delta \) (equation 4) and \( \beta \) (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 (1994), a common way to solve this problem is to used a Metropolis sampling (Metropolis et al., 1953) 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 \( \mathcal{N}(\Delta^{(c)}, \kappa_\Delta V_\Delta) \), and to sample from (5) we generated a proposal value from a multivariate normal distribution \( \mathcal{N}(\beta^{(c)}, \kappa_\beta V_\beta) \), where \( \Delta^{(c)} \) and \( \beta^{(c)} \) are, respectively, the current values of \( \Delta \) and \( \beta \), \( V_\Delta \) and \( V_\beta \) are variances-covariances matrix, and where \( \kappa_\Delta \) and \( \kappa_\beta \) are scaling factors (the detail of the algorithm is given in Appendix). For the Metropolis algorithm to work effectively, \( \kappa_\beta \), \( \kappa_\Delta \), \( V_\beta \) and \( V_\Delta \) 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, 1995).
3. Results

3.1. Simulation studies

We conducted 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 $\frac{1}{2}$ 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 1 000 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., 1999; Kozumi, 2000). The prior for the $\delta_k$ 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 (2 000 iterations), after which some stabilisation was appeared, and estimations was based on the last of the sample (45 000 iterations) where the stabilisation was achieved.
Table 1. Results of simulation studies obtained with MCMC and Esteve et al. (1990) Maximum Likelihood Method (MLM) according to the right random censorship and to the correlation of $(z_1, z_2)$.

<table>
<thead>
<tr>
<th>True values</th>
<th>15% of censor</th>
<th>50% of censor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MCMC Method</td>
<td>MLM</td>
</tr>
<tr>
<td>corr$(z_1, z_2) = 0$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1=ln(1.5)$</td>
<td>Bias$^a$</td>
<td>-0.031</td>
</tr>
<tr>
<td></td>
<td>SD$^b$</td>
<td>0.173</td>
</tr>
<tr>
<td></td>
<td>ASD$^c$</td>
<td>0.183</td>
</tr>
<tr>
<td></td>
<td>ECR$^d$</td>
<td>95.7%</td>
</tr>
<tr>
<td>$\beta_2=ln(2)$</td>
<td>Bias$^a$</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>SD$^b$</td>
<td>0.179</td>
</tr>
<tr>
<td></td>
<td>ASD$^c$</td>
<td>0.182</td>
</tr>
<tr>
<td></td>
<td>ECR$^d$</td>
<td>95.4%</td>
</tr>
<tr>
<td>corr$(z_1, z_2) = 0.7$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1=ln(1.5)$</td>
<td>Bias$^a$</td>
<td>-0.043</td>
</tr>
<tr>
<td></td>
<td>SD$^b$</td>
<td>0.125</td>
</tr>
<tr>
<td></td>
<td>ASD$^c$</td>
<td>0.164</td>
</tr>
<tr>
<td></td>
<td>ECR$^d$</td>
<td>98.9%</td>
</tr>
<tr>
<td>$\beta_2=ln(2)$</td>
<td>Bias$^a$</td>
<td>-0.007</td>
</tr>
<tr>
<td></td>
<td>SD$^b$</td>
<td>0.126</td>
</tr>
<tr>
<td></td>
<td>ASD$^c$</td>
<td>0.166</td>
</tr>
<tr>
<td></td>
<td>ECR$^d$</td>
<td>99.5%</td>
</tr>
</tbody>
</table>

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

$^b$ SD denote the sample standard deviation of the estimates.

$^c$ ASD denote the average of the standard error estimates.

$^d$ ECR denote the proportion of samples in which the 95% credible interval, or the 95% confidence interval include the true value $\beta$.

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. (1990) 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 are 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 include 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., 1999) 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: T1-T2 vs T3-T4; nodal invasion: absent vs present; metastasis extension: absent vs present; see Table 2).

<table>
<thead>
<tr>
<th>Prognostic Factors</th>
<th>Number (per cent)</th>
<th>Deaths at 5 years (per cent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1-T2</td>
<td>152 (49.7%)</td>
<td>97 (43.7%)</td>
</tr>
<tr>
<td>T3-T4</td>
<td>154 (50.3%)</td>
<td>125 (56.3%)</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>122 (39.9%)</td>
<td>74 (33.3%)</td>
</tr>
<tr>
<td>Present</td>
<td>184 (60.1%)</td>
<td>148 (66.7%)</td>
</tr>
<tr>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>293 (95.7%)</td>
<td>209 (94.1%)</td>
</tr>
<tr>
<td>Present</td>
<td>13 (4.3%)</td>
<td>13 (5.9%)</td>
</tr>
<tr>
<td>Overall</td>
<td>306 (100%)</td>
<td>222 (72.5%)</td>
</tr>
</tbody>
</table>

* Per cent of all 306 patients.
† Per cent 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 2 000 iterations and estimations was based on the last 45 000 iterations. Convergence of the chain was achieved using visual inspection of its trace plus the Heidelberger and Welch (1983) 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., 1996).
Table 3. Estimates of adjusted* regression parameters, with standard deviation, obtained for relative survival, with our MCMC method and the Esteve et al. (1990) Maximum Likelihood Method (MLM), and for crude survival, with the Cox PH model (1972), in a population-based study of 306 French pharyngeal male cancer.

<table>
<thead>
<tr>
<th>Prognostic factors</th>
<th>Relative survival</th>
<th>Crude survival</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MCMC Method</td>
<td>Esteve et al. MLM</td>
</tr>
<tr>
<td>T3-T4 †</td>
<td>0.515 ± 0.158</td>
<td>0.474 ± 0.146</td>
</tr>
<tr>
<td>N present †</td>
<td>0.695 ± 0.165</td>
<td>0.663 ± 0.156</td>
</tr>
<tr>
<td>M present †</td>
<td>1.269 ± 0.198</td>
<td>1.168 ± 0.295</td>
</tr>
</tbody>
</table>

* All estimates are adjusted for age.
† Referent categories are: T1-T2, 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 assumed that estimates obtains 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 others prognostic factors.

Table 3 compares also estimates adjusted for age obtained with the relative survival methods (the MCMC method and the Esteve et al. (1990) maximum likelihood method) to those obtained with the crude survival method using the Cox (1972) 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., 1992; Bolard et al., 2001; Giorgi et al., 2003) 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 allow 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. (2000) for the Cox model, but without success. Another way to have algorithm less computationally expensive is to use adaptive rejection sampling (Gilks, 1992) 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. (1995) 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., 1999; Kozumi, 2000), or estimates obtained from a previous study similar to the current study (Ibrahim et al., 2001). 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 $[s_{k-1}, s_k]$, $k = 1,\ldots,M$, 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 conducted 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 (2000), 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 seems 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 \( \Delta^{(c)} \) and \( \beta^{(c)} \)
2. Generate \( \Delta^* \) from \( N(\Delta^{(c)}, \kappa \Delta V_A) \)
   Generate \( u \) from \( \text{uniform}(0,1) \) and
   If \( u \leq \min \left\{ \frac{\pi(\Delta^* | \beta^{(c)}, \text{Data})}{\pi(\Delta^{(c)} | \beta^{(c)}, \text{Data})}, 1 \right\} \), then
     set \( \Delta^{(c+1)} = \Delta^* \)
   Otherwise
     set \( \Delta^{(c+1)} = \Delta^{(c)} \)
3. Generate \( \beta^* \) from \( N(\beta^{(c)}, \kappa \beta V_B) \)
   Generate \( u \) from \( \text{uniform}(0,1) \) and
   If \( u \leq \min \left\{ \frac{\pi(\beta^* | \Delta^{(c+1)}, \text{Data})}{\pi(\beta^{(c)} | \Delta^{(c+1)}, \text{Data})}, 1 \right\} \), then
     set \( \beta^{(c+1)} = \beta^* \)
   Otherwise
     set \( \beta^{(c+1)} = \beta^{(c)} \)
4. Set \( c = c + 1 \)
Repeat step 2 to step 4 a sufficiently long time, until the process become stationary.
Acknowledgement

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

References


R. Giorgi, H. Hedelin and P. Schaffer, Impact of therapeutic changes in relative survival


