Flexible estimation in cure survival models using Bayesian P-splines

Vincent Bremhorst[a,*], Philippe Lambert[a,b]

[a]*Université catholique de Louvain, Institut de Statistique, Biostatistique et Sciences Actuarielles, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium*
[b]*Institut des sciences humaines et sociales, Méthodes quantitatives en sciences sociales,Université de Liège, Belgium, boulevard du Rectorat 7, 4000 Liège, Belgium*

Abstract

In the analysis of survival data, it is usually assumed that any unit will experience the event of interest if it is observed for a sufficiently long time. However, it can be explicitly assumed that an unknown proportion of the population under study will never experience the monitored event. The promotion time model, which has a biological motivation, is one of the survival models taking this feature into account. The promotion time model assumes that the failure time of each subject is generated by the minimum of N independent latent event times with a common distribution independent of N. An extension which allows the covariates to influence simultaneously the probability of being cured and the latent distribution is presented. The latent distribution is estimated using a flexible Cox proportional hazard model where the logarithm of the baseline hazard function is specified using Bayesian P-splines. Introducing covariates in the latent distribution implies that the population hazard function might not have a proportional hazard structure. However, the use of P-splines provides a smooth estimation of the population hazard ratio over time. The identification issues of the model are discussed and a restricted use of the model when the follow up of the study is not sufficiently long is proposed. The accuracy of our methodology is evaluated through a simulation study and the model is illustrated on data from a Melanoma clinical trial.

Keywords: Bayesian P-splines, Cox Model, Cure fraction, Promotion time model, Survival analysis

---

*Corresponding author
  Email addresses:* `vincent.bremhorst@uclouvain.be` (Vincent Bremhorst),
`p.lambert@ulg.ac.be` (Philippe Lambert)

1. Introduction

A common hypothesis in the analysis of survival data is that any observed unit will experience the monitored event if it is observed for a sufficient long time. For example, in a cancer clinical trial, one implicitly assumes that all patients will be observed to have a relapse if their follow up is long enough. Hopefully, this is not always a realistic assumption and the consequences of such a wrong hypothesis on the results of the analysis is more and more questioned in the survival literature.

Alternatively, one can explicitly acknowledge that an unknown and unidentified proportion of the population under study is cured and will never experience the event of interest. Such models are refered as cure survival models. There are two well known families of cure survival models. The first one, often refered as the standard mixture cure model, assumes that the population survival function is obtained as a mixture of contributions due to susceptible and cured individuals :

$$S_p(t|\mathbf{x}, \mathbf{z}) = p(\mathbf{x})S_u(t|\mathbf{z}) + (1 - p(\mathbf{x})), \tag{1}$$

where $p(\mathbf{x})$ is the probability of being susceptible and $S_u(t|\mathbf{z})$ is the survival function of the susceptible individuals. This family of cure models was first introduced by Berkson and Gage (1952). They consider $p(\mathbf{x})$ as an unknown constant and $S_u(t|\mathbf{z})$ is related to a parametric model. Farewell (1982 and 1986) extends the model by letting the covariates influence the probability of being susceptible through a logistic regression. Besides that specification for $p(\mathbf{x})$, many authors propose a semiparametric model for the susceptible survival function, see for example Kuk and Chen (1992), Taylor (1995), Peng and Dear (2000), Sy and Taylor (2000), Li and Taylor (2002), Peng (2003), Lu (2010) and Zhang, Peng and Li (2013). Wang, Du and Liang (2012) propose the first completely nonparametric mixture cure model.

The second family, often refered as the promotion time (cure) model or as the non-mixture cure model, was developed and studied by Yakovlev and Tsodikov (1996), Tsodikov (1998) and Chen, Ibrahim and Sinha (1999). The promotion time model is motivated using biological mechanisms in the development of cancer. The model argues that each subject is exposed to a number $N \sim P_{ois}(\theta)$ of carcinogenic cells. For each cell, Y is defined as the time necessary for it to yield a detectable cancer mass. The $Y_i's$ are often refered as the latent event times. We assume that the cancer mass in each cell is detected independently from each other and that only one cell needs to be activated for a subject to fail. The latent event times $\{Y_1, ..., Y_N\}$ are independent with a common proper distribution $F(t)$ independent of N and the observed failure time is defined as $T = \min_i\{Y_i\}$. If the subject is not exposed to carcinogenic cells (if $N = 0$), he or she is considered as cured. Using the biological

derivation (Yakovlev and Tsodikov (1996) and Chen et al. (1999)) or the mathematical properties (Tsodikov (1998)) of the model, one can show that the population survival function is given by :

$$S_p(t) = \exp\left[-\theta F(t)\right] = \exp\left[-\theta(1 - S(t))\right].\qquad(2)$$

Note that, since $F(t)$ is a proper cumulative distribution function, the probability of being cured is given by $P(N = 0) = \lim_{t\to\infty} S_p(t) = \exp(-\theta)$.

When the covariates only influence the probability of being cured, the log-link is usually used on the parameter θ. Many approaches were proposed to specify the latent distribution $F(t)$, see for example Ibrahim, Chen and Sinha (2001), Zeng, Yin and Ibrahim (2006) and Liu and Shen (2009).

Since the covariates might jointly influence the probability to be cured and the time necessary for a cell to yield a detectable tumor, a Cox proportional hazard model can be suggested for the latent distribution : $F(t|\mathbf{z}) = 1 - S_0(t)^{\exp(\mathbf{z}^T\gamma)}$, where $S_0(t)$ is the baseline survival function. Using these two covariates structures, the population survival function defined in (2) becomes :

$$
\begin{aligned}
S_p(t|\mathbf{x}, \mathbf{z}) &= \exp\left[-\theta(\mathbf{x})F(t|\mathbf{z})\right] \\
&= \exp\left[-\exp\left(\beta_0 + \mathbf{x}^T\beta\right)\left(1 - S_0(t)^{\exp(\mathbf{z}^T\gamma)}\right)\right].
\end{aligned}\qquad(3)
$$

Model (3) was already studied, for example by Tsodikov (2002) in a frequentist framework and by Yin and Ibrahim (2005) in a Bayesian framework. Yin and Ibrahim (2005) assume a piecewize exponential distribution for the baseline survival function $S_0(t)$. They use the conditional predictive ordinate criterion to select the appropriate number of intervals.

In this paper, we propose a flexible estimation of the baseline distribution that does not require a reference to a model selection criterion. We suggest to specify the baseline log-hazard function as a linear combination of cubic B-splines associated to a predefined (large) number of equidistant knots. A roughness penalty will be used to counterbalance the flexibility of the B-splines (Eilers and Marx, 1996). The use of P-splines provides a smooth estimation of the hazard ratio (of the whole population or of the susceptible population) over time when the model does not have a proportional hazard structure.

To the best of our knowledge, this is the first time that P-splines are used in a promotion time model. Moreover, since we propose a Bayesian estimation procedure, the confidence bounds of the functional part of the model are directly obtained from the posterior chains.

When working with cure survival models, it is usually stressed that the follow up should be sufficiently long. We investigate the identifiability issues when that assumption is not satisfied and propose a restricted use of the model.

The remainder of the paper is organised as follows. Section 2 investigates the properties of the proposed model : the hazard ratio and identification issues. A flexible specification of the latent distribution is presented in Section 3. Section 4 is devoted to the Bayesian inference. A simulation study and an application on data from the phase III Melanoma e1684 clinical trial are reported in Section 5 and 6, respectively. A discussion concludes the paper.

2. Model properties

Consider the promotion time (cure) model with covariates ($\mathbf{x}$ and $\mathbf{z}$) introduced in Section 1. The population survival function is given by equation (3) where the cure probability is $\exp[-\theta(\mathbf{x})]$. Although a Cox model was used to describe the time necessary for a cancerous cell to develop a detectable tumor, the survival function at the patient level is usually not of a proportional hazards type (see below).

2.1. Hazard ratio issues

The population hazard function $h_p(t|\mathbf{x}, \mathbf{z})$ is defined by

$$
\begin{aligned}
h_p(t|\mathbf{x}, \mathbf{z}) &= \frac{-d\left(\log\left[S_p\left(t|\mathbf{x}, \mathbf{z}\right]\right)\right)}{dt} \\
&= \theta(\mathbf{x})f(t|\mathbf{z}) \\
&= \frac{\exp(\beta_0 + \mathbf{x}^T\beta)\exp(\mathbf{z}^T\gamma)f_0(t)S_0(t)^{\exp(\mathbf{z}^T\gamma)}}{S_0(t)},
\end{aligned}
$$

where $f_0(t)$ is the baseline density function.

The population hazard ratio comparing groups 1 and 2 is given by :

$$
\begin{aligned}
HR_p &= \frac{h_p(t|\mathbf{x}_1, \mathbf{z}_1)}{h_p(t|\mathbf{x}_2, \mathbf{z}_2)} \\
&= \exp\left((\mathbf{x}_1^T - \mathbf{x}_2^T)\beta\right)\exp\left((\mathbf{z}_1^T - \mathbf{z}_2^T)\gamma\right)S_0(t)^{\left(\exp(\mathbf{z}_1^T\gamma) - \exp(\mathbf{z}_2^T\gamma)\right)}.
\end{aligned}
$$

Thus, the population hazard ratio HR_p remains constant over time if $\exp(\mathbf{z}_1^T\gamma) - \exp(\mathbf{z}_2^T\gamma) = 0$, i.e. if $\mathbf{z}_1 = \mathbf{z}_2$. In practice, this means that the population hazard ratio is constant provided that the contrasted groups share the same values for the covariates affecting the development of cancerous cells in the biological model.

4

As mentioned in Chen et al. (1999), the hazard function of the susceptible individuals, given in (4), does not have a proportional hazard structure. Indeed, one can show that the susceptible survival function is given by :

$$
\begin{aligned}
S_u(t|\mathbf{x}, \mathbf{z}) &= P(T > t | N \geq 1, \mathbf{x}, \mathbf{z}) \\
&= \frac{S_p(t|\mathbf{x}, \mathbf{z}) - \exp[-\theta(\mathbf{x})]}{1 - \exp[-\theta(\mathbf{x})]},
\end{aligned}
$$

and the susceptible hazard function is given by :

$$
\begin{aligned}
h_u(t|\mathbf{x}, \mathbf{z}) &= \frac{-d\left(\log\left[S_u\left(t|\mathbf{x}, \mathbf{z}\right)\right]\right)}{dt} \\
&= \frac{S_p(t|\mathbf{x}, \mathbf{z})}{S_p(t|\mathbf{x}, \mathbf{z}) - \exp[-\theta(\mathbf{x})]} h_p(t|\mathbf{x}, \mathbf{z}) \\
&= \frac{h_p(t|\mathbf{x}, \mathbf{z})}{P(T < +\infty | T > t, \mathbf{x}, \mathbf{z})}.
\end{aligned}
\tag{4}
$$

It is straightforward to see that expression (4) does not have a proportional hazard structure.

2.2. Identification issues

Theoretically, the follow up of a study is said to be sufficiently long if the largest censoring time is greater than the largest failure time, i.e. if the follow-up time of any susceptible unit was sufficiently long to observe its failure. In practice, if a plateau is present in the right tail of the estimated population survival function (for example, in the Kaplan Meier estimated curve), the sufficient follow up assumption seems to be reasonable.

<u>Lemma 1.</u>
> *Assumptions :*
> *A1* *The vector $\mathbf{z}$ of covariates does not include an intercept.*
> *A2* *$\mathbf{X}^T\mathbf{X}$ and $\mathbf{Z}^T\mathbf{Z}$ are full rank matrices, where $\mathbf{X}$ and $\mathbf{Z}$ are the design matrices corresponding to covariate vectors $\mathbf{x}$ and $\mathbf{z}$, respectively.*
> *A3* *The baseline cumulative distribution function $F_0(t) = 1 - S_0(t)$ is proper.*
> *Under A1, A2 and A3, we have :*
> *1)* *If the follow up of the study is sufficiently long, then model (3) is identifiable.*

2) *If the follow up of the study is not sufficiently long and if vectors $\mathbf{x}$ and $\mathbf{z}$ do not share some components, then only the estimations of the covariate effects on the cure probability and on failure time for a cancerous cell are identifiable.*

The proof of this lemma is in appendix: it is based on the proof proposed by Liu and Shen (2009) showing the identifiability of the promotion time model when the covariates only influence the probability to be cured.

3. Flexible specification of the baseline distribution

We assume some familiarity with P-splines from the reader. If not, information about P-splines can be found in Eilers and Marx (1996) and in Lang and Brezger (2004). In order to estimate the baseline survival function $S_0(t)$ in (3), we suggest to write the baseline log-hazard as a linear combination of cubic B-splines:

$$h_0(t) \;=\; \exp\left(\sum_{k=1}^{K} b_k(t)\phi_k \right), \tag{5}$$

where $\{b_k(.), k = 1, ..., K\}$ denotes the cubic B-splines basis associated to a predefined number of equidistant knots on $[0, t_{Rcens}]$, where t_{Rcens} is the upper bound of the follow up.

To ensure enough flexibility, Eilers and Marx (1996) suggest to choose a large number of B-splines and to counterbalance the flexibility by adding to the log-likelihood a roughness penalty based on finite differences of adjacent B-spline parameters : $\tau \sum_k (\Delta^r \phi_k)^2 \;=\; \tau\phi^T\mathbf{D}^T\mathbf{D}\phi$, where τ is the penalty parameter and $\mathbf{D}$ is the rth difference penalty matrix. For example, when a third order penalty is specified, the matrix $\mathbf{D}$ is defined as :

$$\mathbf{D} = \begin{bmatrix} 1 & -3 & 3 & -1 & 0 & ... & 0 \\ 0 & 1 & -3 & 3 & -1 & ... & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & ... & 1 & -3 & 3 & -1 \end{bmatrix}.$$

P-splines were already used in many different contexts, see for example Eilers and Marx (1996) and Eilers (2007) in a frequentist framework and Lang and Brezger (2004), Lambert and Eilers (2005), Lambert (2007, 2013) and Cetinyurek and Lambert (2011) in a Bayesian framework. As mentioned in all these references, if K is chosen large enough (between 10 and 20, say), no model selection criterion is needed since all the K's give similar results.

Knowing the relation between the survival function and the hazard function and using (5) as expression for the hazard function, the baseline survival function $S_0(t)$ is specified as :

$$S_0(t) = \exp\left(-\int_0^t \exp\left[\sum_{k=1}^K b_k(u)\phi_k\right] du\right). \tag{6}$$

The integral in (6) has no analytic form and needs to be evaluated numerically. Knowing that our observations are contained in the interval $[0, t_{Rcens}]$, we partition $[0, t_{Rcens}]$ into J (300, say) small bins (of equal width, for simplicity) $J_j = [\tau_{j-1}, \tau_j]$ where $0 = \tau_0 < \tau_1 < ... < \tau_J = t_{Rcens}$. Let u_j and δ_j denote the midpoint and the width of J_j, respectively. Then, using the rectangle method, (6) can be approximated by :

$$S_0(t) \approx \exp\left(-\sum_{j=1}^{j(t)} \exp\left[\sum_{k=1}^K b_k(u_j)\phi_k\right]\delta_j\right), \tag{7}$$

where $j(t)$ indexes the interval containing t.

For identifiability purpose (see Section 2.2), we fix the last spline parameter ϕ_K to a large enough value (10, say). In this way, we force the estimated baseline survival function $\hat{S}_0(.)$ to be 0 at the end of the follow up.

4. Bayesian inference

4.1. Likelihood

For the ith subject under study, we observe the failure or the censoring time t_i, the event indicator ν_i and two sets of covariates $\mathbf{x}_i$ and $\mathbf{z}_i$. We denote these observable variables by $\mathbb{D}_i = (t_i, \nu_i, \mathbf{x}_i, \mathbf{z}_i)$. The set of parameters specific to the chosen model is written as $\mathbf{\Phi}$. Then, the data likelihood is given by :

$$L(\mathbf{\Phi}|\mathbb{D}) = \prod_{i=1}^I h_p(t_i)^{\nu_i} S_p(t_i).$$

4.2. Bayesian Model

In a Bayesian setting, the roughness penalty is translated into a prior distribution for the spline parameters (Lang and Brezger, 2004):

$$\pi(\phi|\tau) \propto \tau^{\frac{K}{2}} \exp\left(-\frac{\tau}{2}\phi^T \mathbf{P}\phi\right),$$

where $\mathbf{P} = \mathbf{D}^T\mathbf{D} + \epsilon\mathbf{I}_K$ is a full rank matrix for some small quantity ϵ (10^{-6}, say). In other words, a normal distribution with mean 0 and variance-covariance matrix $\mathbf{P}^{-1}$ for the spline parameters is considered. As suggested by Jullion and Lambert (2007), we take a robust specification for the roughness penalty prior distribution :

$$
\begin{aligned}
\tau|\delta &\sim G(\frac{\nu}{2}, \frac{\nu\delta}{2}), \\
\delta &\sim G(a_\delta, b_\delta),
\end{aligned}
$$

where $G(a,b)$ denotes a Gamma distribution with mean $\frac{a}{b}$ and variance $\frac{a}{b^2}$.
They showed that if a small value is chosen for a_δ and b_δ (10^{-4}, say), then the choice of ν (here, set equal to 2) does not affect the shape of the estimated curve. If prior knowledge (such as monotonicity) is available about the baseline hazard, it can be expressed during the prior elicitation for the spline parameters. If nothing is known a priori about the covariate effects, a large variance normal prior distribution can be used for all the regression parameters.

Using Bayes' theorem, the joint posterior distribution is given by :

$$
\pi(\mathbf{\Phi}|\mathbb{D}) \propto L(\mathbf{\Phi}|\mathbb{D})\pi(\phi|\tau)\pi(\tau|\delta)\pi(\delta)\pi(\beta_0, \beta)\pi(\gamma). \tag{8}
$$

Given that all the prior distributions are proper, the posterior distribution $\pi(\mathbf{\Phi}|\mathbb{D})$ is proper. Only the conditional posterior of τ and δ belong to known families of distributions :

$$
\begin{aligned}
\tau|\phi, \delta, \mathbb{D} &\sim G\left(\frac{\nu + K}{2}, \frac{\nu\delta + \phi^T\mathbf{P}\phi}{2}\right), \\
\delta|\tau, \mathbb{D} &\sim G(a_\delta + \frac{\nu}{2}, b_\delta + \frac{\nu\tau}{2}).
\end{aligned}
$$

4.3. Posterior sample using MCMC

A Metropolis step will be used to sample the other conditional distributions. As shown by Lambert (2007), the mixing of the chains can be improved by applying the Metropolis algorithm on a reparametrized posterior distribution. An adequate reparametrization can be suggested by a frequentist estimation of the correlation structure of the spline parameters. To reach that goal, one could use a nonlinear optimizer to compute the mode of the joint posterior distribution in (8), for a fixed value of δ. At convergence, the hessian matrix can be used to assess the posterior correlation between the parameters and to suggest a reparametrization yielding less dependent components (see Lambert, 2007, for more details).
Let $\tilde{\beta} = (\beta_0, \beta)$ and $(\phi^{(0)}, \tau^{(0)}, \delta^{(0)}, \tilde{\beta}^{(0)}, \gamma^{(0)})$ be the initial values of the chain selected,

8

for example, using the optimization step described above. The MCMC algorithm consists in five main steps to sample the parameters from the posterior. Iteration m proceeds as follows :

- Draw $\phi^{(m)}$ from $\pi(\phi|\tau^{(m-1)}, \delta^{(m-1)}, \tilde{\beta}^{(m-1)}, \gamma^{(m-1)})$ using univariate Metropolis steps (along directions suggested by the reparametrization) ;
- Draw $\tau^{(m)}$ from $G\left(\frac{\nu+K}{2}, \frac{\nu\delta^{(m-1)}+\phi^{(m)^T}\mathbf{P}\phi^{(m)}}{2}\right)$ in a Gibbs step ;
- Draw $\delta^{(m)}$ from $G(a_\delta + \frac{\nu}{2}, b_\delta + \frac{\nu\tau^{(m)}}{2})$ in a Gibbs step ;
- Draw $\tilde{\beta}^{(m)}$ from $\pi(\tilde{\beta}|\phi^{(m)}, \tau^{(m)}, \delta^{(m)}, \gamma^{(m-1)})$ using univariate Metropolis steps ;
- Draw $\gamma^{(m)}$ from $\pi(\gamma|\phi^{(m)}, \tau^{(m)}, \delta^{(m)}, \tilde{\beta}^{(m)})$ using univariate Metropolis steps ;

The variances of the proposal distributions in the Metropolis steps are tuned automatically using the adaptive procedure proposed by Haario, Saksman and Tamminen (2001) during the burnin to achieve the targeted optimal acceptance rate (Gelman, Roberts and Gilks, 1996 and Roberts and Rosenthal, 2001).

5. Simulation study

5.1. Sufficiently long follow up

The accuracy of the proposed methodology was evaluated using simulations when the follow up is sufficiently long (see Section 2.2). In each setting, the baseline distribution in (3) corresponds to a Weibull with mean 8 and standard deviation 4.18. Two covariates were included in the regression parts : $W_1 \sim N(0,1)$ and $W_2 \sim Bernoulli(0.5)$. Since the sufficient follow up assumption is satisfied, both covariates can be used simultaneously to model the probability of being cured and the time necessary for a cell to yield a detectable tumor without causing an identifiability problem. Thus, we set $\mathbf{X} = \{W_1, W_2\} = \mathbf{Z}$. The regression coefficients associated to W_1 and W_2, in the Cox PH model, are set to 0.4 and -0.4, respectively. The upper bound for the observed failure time was set at 23 as more than 99% of the events occur before that time under the chosen Weibull distribution. Three percentages were considered for the proportion of cured individuals : 15%, 25% and 40%. The value of the regressors (β_0, β) were tuned to get these percentages. Each dataset was generated using the biological motivation of the model as follows : For each subject :

1) Generation of the number of carcinogenic cells using $N \sim Pois(\theta(\mathbf{x}))$ with $\theta(\mathbf{x}) = \exp(\beta_0 + \mathbf{x}^T\beta)$;

9

2) If $N \neq 0$, N latent event times $Y_1, ..., Y_N$ are generated using the Cox proportional hazard model. The observed failure time is defined as $T = \min(Y_1, ..., Y_N)$. This step is repeated until $T < 23$. Note that it had to be repeated more than once in less than 1% of the cases. If $N = 0$, the failure time for the cured individual is set to an arbitrary large value (999, say).

3) The global right censoring rate is controlled by one of the two following censoring distributions :

   a) setting 1 : an uniform distribution on $[20, 25]$. This censoring distribution ensures that almost all the right censored subjects are cured and identifiable (since their censoring time are located in the plateau of the Kaplan Meier estimate of the survival distribution).

   b) setting 2 : a Weibull distribution with mean 22.28 and standard deviation 8.08 truncated at 25. Using this censoring distribution, the censoring time of only 25% of the cured subjects are located in the plateau of the Kaplan Meier estimate and 4% of the suceptible individuals are right censored.

We use the model described in Section 3 with a cubic B-splines basis associated to 12 equidistant knots on $[0, t_{Rcens}]$, where t_{Rcens} is equal to 25 and a third order roughness penalty to counterbalance the flexibility of the B-splines. The simulations were performed on S $= 500$ replicates of sample size n $= 300$ and 600.

Using the procedure described in Section 4.3, we construct a chain of length 23000 (including a burnin of 3000) to explore the joint posterior distribution. The behavior and the convergence of the chains were assessed by an examination of the trace plots and using diagnostics tools such as in Geweke (1992).

For the sake of brevity, we only report the results when the percentage of cured individuals is 25% and 40%. Tables 1 and 2 summarize the simulation results for the regression parameters. One can see that the posterior medians (as estimators) of the regression coefficients show a negligible bias whatever the setting. The empirical standard error and RMSE of the posterior median of the regression parameters decrease slighly when the sample size increases and increase when the proportion of cured individuals with a censoring time greater than the maximum observed failure time decreases and when the percentage of right censoring in the non cured population increases.

In each setting, the coverage probabilities of the 90% and 95% credible intervals are close to their nominal value. The numerical results suggest that the proportion of cured individuals does not affect the accuracy of the estimates.

The estimates of the baseline survival function are plotted in Figures 1 (when $n = 300$) and 2 (when $n = 600$). The variability of the estimated baseline distribution

10

increases slightly when the proportion of cured individuals with a censoring time greater than the maximum observed failure time decreases and when the percentage of right censoring in the non cured population increases and decreases when the sample size increases. A limited bias decreasing with sample size seems to appear in the estimation of the right tail of $S_0(t)$ when the proportion of cured individuals with a censoring time greater than the maximum observed failure time decreases and when the percentage of non cured right censored subjects increases. As for the regression parameters, the percentage of immune individuals does not affect the accuracy of the estimates. The same conclusions as for the regression parameters and for $S_0(t)$ hold for the population log-hazard ratio (see Figures 3, when $n = 300$, and 4, when $n = 600$), while the log-hazard ratio of the susceptible individuals is properly estimated whatever the setting (see Figure 5, when $n = 300$). These hazard ratios are obtained by contrasting the groups induced by the binary covariate (for a median value of the continuous covariate).

Table 1: Simulation results for $S = 500$ replicates and a sample size of $n = 300$ when the follow up is sufficiently long. The percentage of cured individuals, the considered setting and the true value of the regression parameters are defined in the first three columns. The bias, the coverage of the 90% and 95% credible intervals, the empirical standard error (ESE) and the RMSE of the posterior median of the regression parameters are presented for each scenario.

Cure	Setting	Parameters	Bias	$CV_{90\%}$	$CV_{95\%}$	ESE	RMSE
25%	1	$\beta_0 = 0.75$	0.028	88.4	93.8	0.130	0.018
		$\beta_1 = 0.80$	0.016	90.8	94.8	0.112	0.013
		$\beta_2 = $ -0.50	-0.013	91.6	96.6	0.172	0.030
		$\gamma_1 = 0.40$	-0.035	91.0	95.8	0.138	0.142
		$\gamma_2 = $ -0.40	0.008	87.4	94.2	0.221	0.221
	2	$\beta_0 = 0.75$	0.052	85.8	92.6	0.162	0.030
		$\beta_1 = 0.80$	-0.003	93.4	97.8	0.129	0.017
		$\beta_2 = $ -0.50	0.017	88.6	94.0	0.209	0.044
		$\gamma_1 = 0.40$	0.001	92.8	95.8	0.171	0.171
		$\gamma_2 = $ -0.40	-0.029	88.2	93.6	0.283	0.282
40%	1	$\beta_0 = 0.30$	0.006	90.6	94.0	0.128	0.017
		$\beta_1 = 1.00$	0.013	90.6	95.0	0.119	0.014
		$\beta_2 = $ -0.75	-0.008	90.6	94.4	0.182	0.033
		$\gamma_1 = 0.40$	-0.021	90.0	95.6	0.150	0.151
		$\gamma_2 = $ -0.40	0.003	92.0	96.8	0.215	0.215
	2	$\beta_0 = 0.30$	0.043	86.2	91.8	0.155	0.027
		$\beta_1 = 1.00$	-0.014	93.6	97.0	0.137	0.019
		$\beta_2 = $ -0.75	0.024	88.0	93.4	0.222	0.049
		$\gamma_1 = 0.40$	0.005	91.8	96.4	0.183	0.183
		$\gamma_2 = $ -0.40	-0.025	90.4	94.4	0.276	0.276

Table 2: Simulation results for $S = 500$ replicates and a sample size of $n = 600$ when the follow up is sufficiently long. The percentage of cured individuals, the considered setting and the true value of the regression parameters are defined in the first three columns. The bias, the coverage of the 90% and 95% credible intervals, the empirical standard error (ESE) and the RMSE of the posterior median of the regression parameters are presented for each scenario.

| Cure	Setting	Parameters	Bias	$CV_{90\%}$	$CV_{95\%}$	ESE	RMSE
25%	1	$\beta_0 = 0.75$	0.020	90.2	94.0	0.087	0.008
		$\beta_1 = 0.80$	0.017	90.4	95.2	0.077	0.006
		$\beta_2 = -0.50$	-0.020	91.6	95.8	0.115	0.014
		$\gamma_1 = 0.40$	-0.038	87.6	93.6	0.097	0.106
		$\gamma_2 = -0.40$	0.022	89.8	96.4	0.139	0.141
	2	$\beta_0 = 0.75$	0.036	87.2	92.8	0.107	0.013
		$\beta_1 = 0.80$	0.009	93.2	97.2	0.090	0.008
		$\beta_2 = -0.50$	-0.011	88.6	93.6	0.150	0.023
		$\gamma_1 = 0.40$	-0.023	91.0	96.0	0.115	0.118
		$\gamma_2 = -0.40$	0.018	91.6	95.0	0.182	0.183
40%	1	$\beta_0 = 0.30$	0.005	91.2	95.2	0.087	0.008
		$\beta_1 = 1.00$	0.016	91.2	95.6	0.079	0.006
		$\beta_2 = -0.75$	-0.003	91.6	95.6	0.119	0.014
		$\gamma_1 = 0.40$	-0.027	91.2	95.0	0.098	0.101
		$\gamma_2 = -0.40$	0.036	92.4	95.2	0.142	0.146
	2	$\beta_0 = 0.30$	0.021	89.4	94.8	0.094	0.009
		$\beta_1 = 1.00$	-0.004	92.8	95.4	0.093	0.009
		$\beta_2 = -0.75$	0.010	91.2	95.8	0.142	0.020
		$\gamma_1 = 0.40$	-0.002	93.6	97.8	0.115	0.115
		$\gamma_2 = -0.40$	0.025	92.6	96.4	0.184	0.184
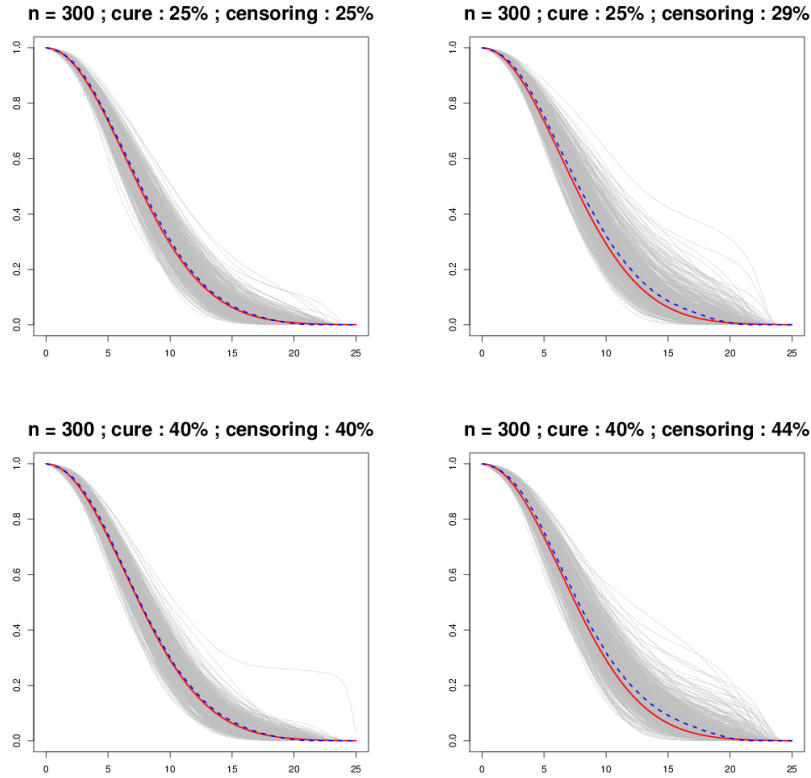
Figure 1: Simulation results when the follow up is sufficiently long : estimation of the baseline distribution $S_0(t)$ for $S = 500$ replications (one grey curve per data set) and sample size $n = 300$. Each row refers to a percentage of cured individuals (row 1 : 25%, row 2 : 40%) with different global right censoring rates (left : setting 1 ; right : setting 2). The solid line corresponds to the true function and the dashed line is the pointwize median of the 500 estimated curves.
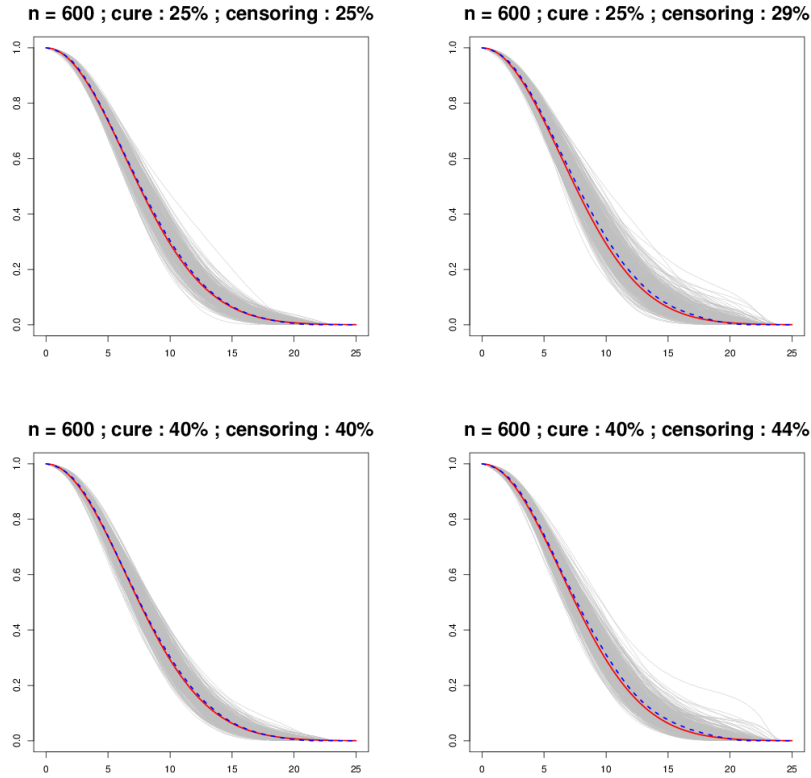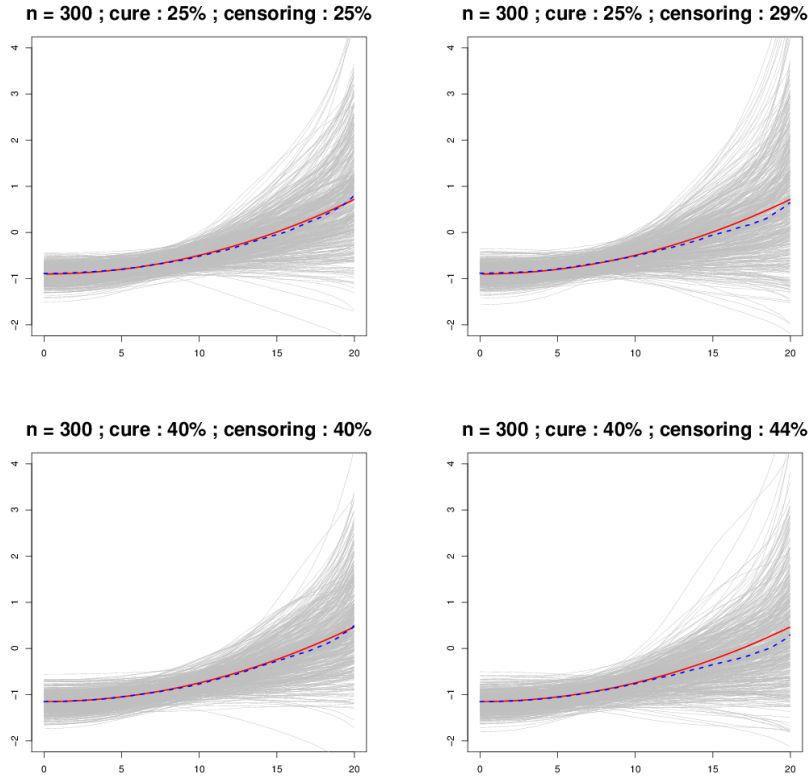
Figure 2: Simulation results when the follow up is sufficiently long : estimation of the baseline distribution $S_0(t)$ for $S = 500$ replications (one grey curve per data set) and sample size $n = 600$. Each row refers to a percentage of cured individuals (row 1 : 25%, row 2 : 40%) with different global right censoring rates (left : setting 1 ; right : setting 2). The solid line corresponds to the true function and the dashed line is the pointwize median of the 500 estimated curves.
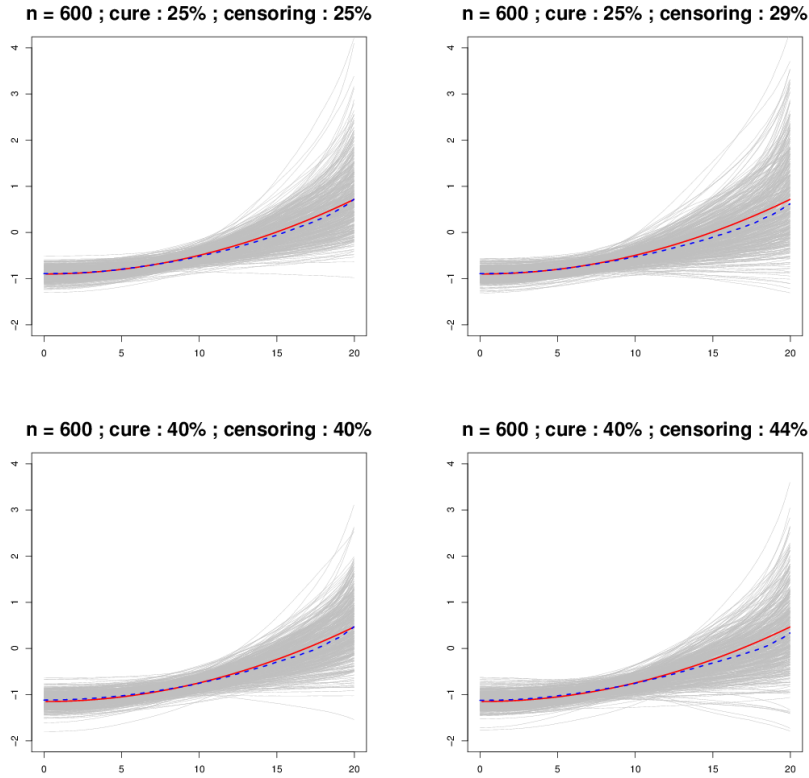
Figure 3: Simulation results when the follow up is sufficiently long : estimation of the population log-hazard ratio $\log(HR_p(t))$ for $S = 500$ replications (one grey curve per data set) and sample size $n = 300$. Each row refers to a percentage of cured individuals (row 1 : 25%, row 2 : 40%) with different global right censoring rates (left : setting 1 ; right : setting 2). The solid line corresponds to the true function and the dashed line is the pointwize median of the 500 estimated curves. The hazard ratio is obtained by contrasting the groups induced by the binary covariate (for a median value of the continuous covariate.)
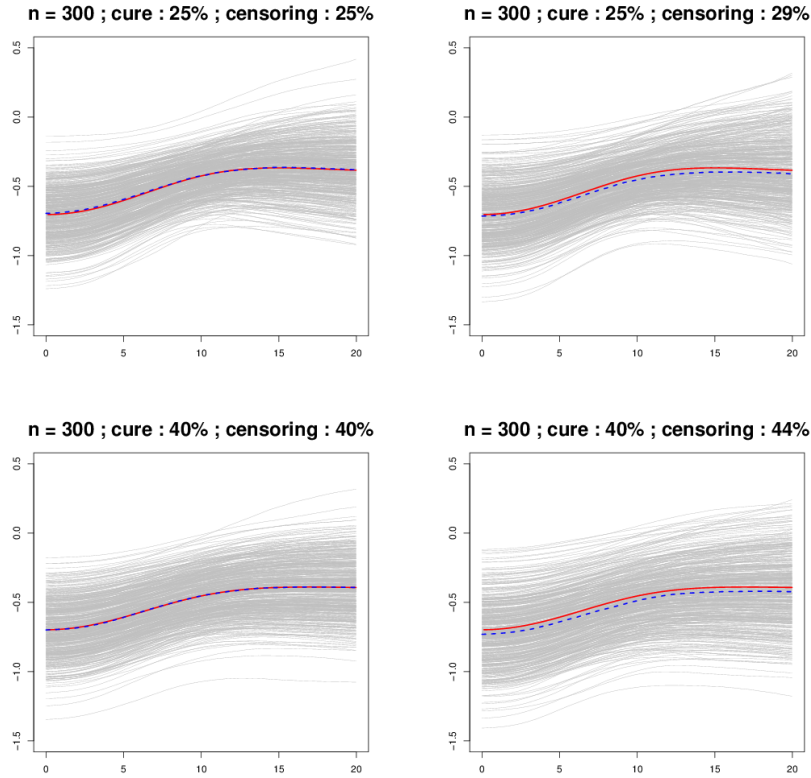
Figure 4: Simulation results when the follow up is sufficiently long : estimation of the population log-hazard ratio $\log(HR_p(t))$ for $S = 500$ replications (one grey curve per data set) and sample size $n = 600$. Each row refers to a percentage of cured individuals (row 1 : 25%, row 2 : 40%) with different global right censoring rates (left : setting 1 ; right : setting 2). The solid line corresponds to the true function and the dashed line is the pointwize median of the 500 estimated curves. The hazard ratio is obtained by contrasting the groups induced by the binary covariate (for a median value of the continuous covariate.)

17

Figure 5: Simulation results when the follow up is sufficiently long : estimation of the log-hazard ratio for the susceptible individuals $\log(HR_u(t))$ for $S = 500$ replications (one grey curve per data set) and sample size $n = 300$. Each row refers to a percentage of cured individuals (row 1 : 25%, row 2 : 40%) with different global right censoring rates (left : setting 1 ; right : setting 2). The solid line corresponds to the true function and the dashed line is the pointwize median of the 500 estimated curves. The hazard ratio is obtained by contrasting the groups induced by the binary covariate (for a median value of the continuous covariate.)

5.2. Not sufficiently long follow up

The restricted use of the promotion time model when the follow up of the study is not sufficiently long is illustrated using simulations. The datasets are generated using the procedure described in Section 5.1. However, since the follow up of the study is not sufficiently long, for identification purpose, W_1 only influences the time for a cell to yield a detectable tumor ($Z = W_1$), and W_2 only influences the probability of being cured ($X = W_2$). Note that if the hazard ratio is obtained by contrasting the groups induced by the binary covariate (for a median value of the continuous covariate), the population hazard ratio has, in this setting, a proportional hazard structure. Three percentages were considered for the proportion of cure individuals : 15%, 25% and 40%. To ensure that the largest censoring time is smaller than the largest observed failure time (which is the feature of non sufficiently long follow up), a Weibull distribution with mean 17.9 and standard deviation 6.5 truncated at 13.7 (setting 3) and at 10.6 (setting 4) is considered for the censoring distribution.

As in Section 5.1, we use the model described in Section 3 with a cubic B-splines basis associated to 12 equidistant knots on $[0, t_{Rcens}]$, where t_{Rcens} is equal to 13.7 or 10.6 depending on the considered censoring distribution and a third order roughness penalty to counterbalance the flexibility of the B-splines. The simulations were performed on $S = 500$ replicates of sample size $n = 300$ and 600. A chain of length 23000 (including a burnin of 3000) is constructed using the procedure described in Section 4.3. As previously, traces and z-scores of all the model parameters are examined to check the convergence of the MCMC algorithm. For the sake of brevity, we only report the results when the percentage of cured individuals is 25% and 40% and when the sample size is 300. Table 3 summarizes the simulation results for the regression parameters. As expected from the theory, in each setting, the posterior medians, as estimators of the intercept, show an underestimation. The biases are close to $\log(F_0(t_{Rcens}))$, where $F_0(.)$ is the cumulative distribution function of the considered baseline Weibull distribution in (3), as can be explained from equation (A.1). The posterior medians of β_1 and γ_1 show a non significant bias whatever the setting. The accuracy of the estimators of the regression parameters increases with the upper bound of the follow up and with sample size. In each setting, the coverage probabilities of the 90% and 95% credible intervals are close to their nominal value except for the intercept due to its underestimation. Figure 6 shows that the baseline distribution $S_0(t)$ is underestimated : this is due to the zero tail constraint. As illustrated on Figure 7, an overestimation appears in the estimation of the log-hazard ratio of the suceptible individuals when the upper bound of the follow up is really to small. It happens when the baseline distribution function in (3) at the maximum possible censoring time is much smaller than 1 : one has $F_0(10.6) = 0.75$

19

Table 3: Simulation results for $S = 500$ replicates and a sample size of $n = 300$ when the follow up is not sufficiently long. The percentage of cured individuals, the considered setting and the true value of the regression parameters are defined in the first three columns. The bias, the coverage of the 90% and 95% credible intervals, the empirical standard error (ESE) and the RMSE of the posterior median of the regression parameters are presented for each scenario.

Cure	Setting	Parameters	Bias	$CV_{90\%}$	$CV_{95\%}$	ESE	RMSE
25%	3	$\beta_0 = 0.70$	-0.093	80.6	87.8	0.129	0.025
		$\beta_1 = -0.70$	0.007	88.4	95.6	0.156	0.024
		$\gamma_1 = 0.40$	-0.029	86.6	92.2	0.111	0.114
	4	$\beta_0 = 0.70$	-0.266	46.6	60.6	0.157	0.096
		$\beta_1 = -0.70$	0.007	91.0	93.8	0.169	0.029
		$\gamma_1 = 0.40$	-0.041	85.6	91.0	0.124	0.130
40%	3	$\beta_0 = 0.30$	-0.112	77.4	85.0	0.142	0.033
		$\beta_1 = -0.80$	0.024	91.0	96.0	0.178	0.032
		$\gamma_1 = 0.40$	-0.039	86.8	93.0	0.127	0.133
	4	$\beta_0 = 0.30$	-0.292	45.4	56.6	0.159	0.104
		$\beta_1 = -0.80$	0.024	92.0	96.2	0.195	0.038
		$\gamma_1 = 0.40$	-0.049	86.2	92.2	0.137	0.145

for the shortest follow up (setting 4) and $F_0(13.7) = 0.9$ in the most favorable setting (setting 3). Similar conclusions can be drawn when the sample size is equal to 600. These simulation results corroborate the theoretical results proved in Lemma 1.
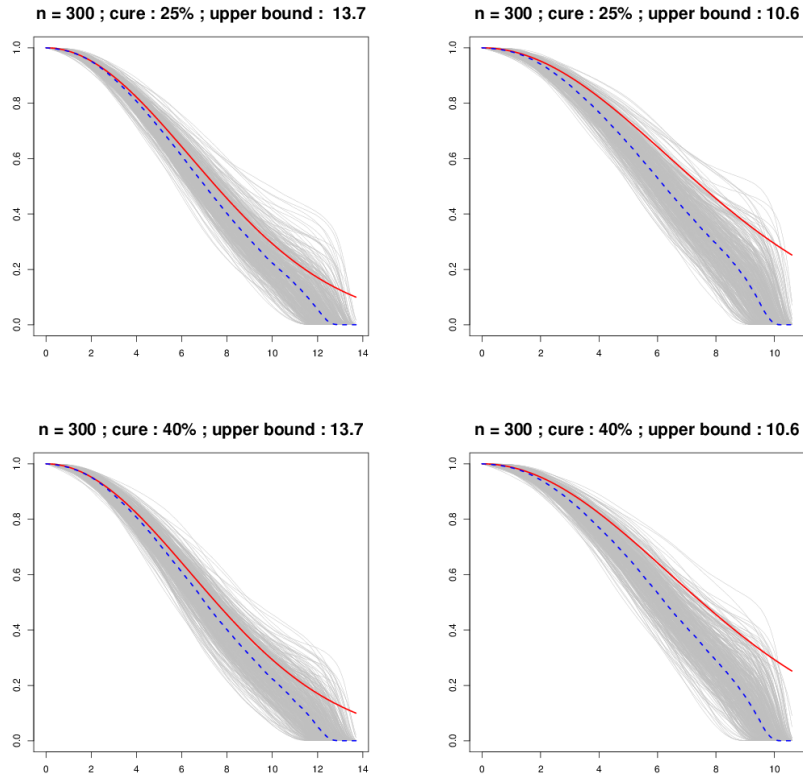
Figure 6: Simulation results when the follow up is not sufficiently long : estimation of the baseline distribution $S_0(t)$ for $S = 500$ replications (one grey curve per data set) and sample size $n = 300$. Each row refers to a percentage of cured individuals (row 1 : 25%, row 2 : 40%) and the columns refer to the upper bound of the follow up (col 1 (setting 3) : 13.7, col 2 (setting 4) : 10.6). The solid line corresponds to the true function and the dashed line is the pointwize median of the 500 estimated curves.
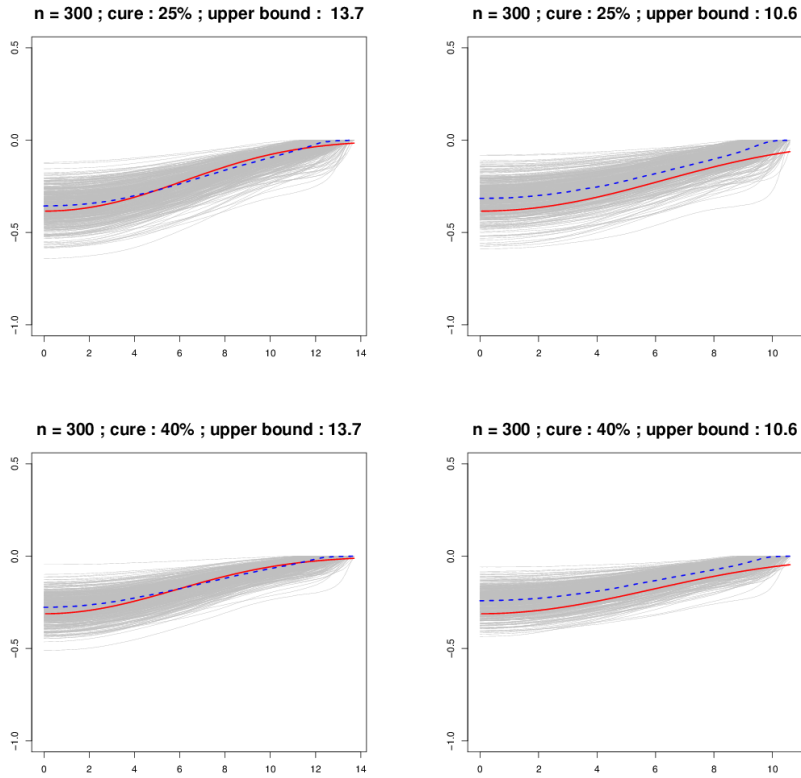
Figure 7: Simulation results when the follow up is not sufficiently long : estimation of the log-hazard ratio for the susceptible individuals $\log(HR_u(t))$ for $S = 500$ replications (one grey curve per data set) and sample size $n = 300$. Each row refers to a percentage of cured individuals (row 1 : 25%, row 2 : 40%) and the columns refer to the upper bound of the follow up (col 1 (setting 3) : 13.7, col 2 (setting 4) : 10.6). The solid line corresponds to the true function and the dashed line is the pointwize median of the 500 estimated curves. The hazard ratio is obtained by contrasting the groups induced by the binary covariate (for a median value of the continuous covariate.)
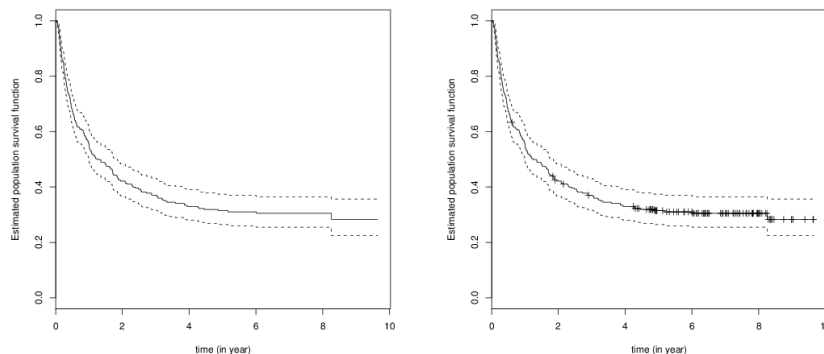
6. Application

Figure 8: Kaplan Meier estimated curve. Right : Each censored individual is marked by a cross. Left : Without cross for the censored individuals.

We illustrate our methodology on data from a phase III Melanoma e1684 clinical trial. The study was conducted by Eastern Cooperative Oncology Group (ECOG) and consist on a two stages randomized clinical trial : Interferon alpha-2b (IFN) versus Control (Kirkwood et al. (1996)). Their study suggests that Interferon alpha-2b has a significant positive effect on the relapse free survival time. Among the 284 patients present in the study, 144 (51%) receive the IFN treatment, 171 (60%) are male, and we observe a relaspe of cancer for 196 (69%) of them. The other subjects were right censored. The age of each subject is known with an average of 47 years and standard deviation of 13 years. The Kaplan Meier estimated curve (Figure 8) shows a plateau. This suggests that the follow up of the study was sufficiently long. Thus, we let sex, age and the randomised treatment influence simultaneously the probability to be cured and the time necessary for a cell to yield a detectable tumor. The survival data of this clinical trial were already studied in Chen et al (1999) and in Cooner et al (2007). However, in their analysis, they assumed that the covariates only infuence the probability to be cured. Here, we also enable the covariates to influence the time necessary for a cell to yield a detectable tumor.

The procedure described in Section 4.3 was use to explore the joint posterior distribution. As for the simulations, 23000 iterations (including a burnin of 3000) were generated. To check the convergence of the MCMC algorithm, the traces of all model parameters were examined critically. The z-scores of the Geweke diagnostics were found to be between -1.96 and 1.96 for all model parameters, suggesting convergence

23

Table 4: Melanoma e1684. Estimation of the posterior median, the 95% HPD interval and the posterior standard deviation of each regression parameter of the model.

	Parameters	Estimation	$HPD_{95\%}$	sd_{post}
	Intercept	0.351	[0.088 ; 0.564]	0.123
$\theta(\mathbf{x})$	AGE	0.100	[-0.046 ; 0.248]	0.074
	IFN	-0.321	[-0.635 ; -0.032]	0.155
	SEX	-0.031	[-0.337 ; 0.276]	0.161
	AGE	-0.136	[-0.303 ; 0.033]	0.085
$F(t|\mathbf{z})$	IFN	-0.060	[-0.417 ; 0.341]	0.190
	SEX	0.053	[-0.333 ; 0.430]	0.195

of the MCMC algorithm.

Table 4 presents the MCMC estimates of the posterior median, the 95% HPD interval and the posterior standard deviation of the regression parameters. One can conclude that treatment only has a significant effect on the probability to be cured. In other words, our model suggests that Interferon alpha-2b significantly reduces the number of carcinogenic cells but does not influence the incubation time of one cell. This conclusion is illustrated on Figures 9 and 10. Figure 9 shows the fitted population survival function (right) and the fitted survival function for the susceptible individuals (left). The only relevant difference is between treatment groups in the population survival function, illustrated for a median value of AGE. Figure 10, for males of median age, shows the estimate of the logarithm of the population hazard ratio $\log(HR_p)$ (left) and of the logarithm of the hazard ratio for the susceptible $\log(HR_u)$ (right) : only $\log(HR_p)$ significantly differs from zero at the beginning of the study. Similar conclusions can be drawn for females.
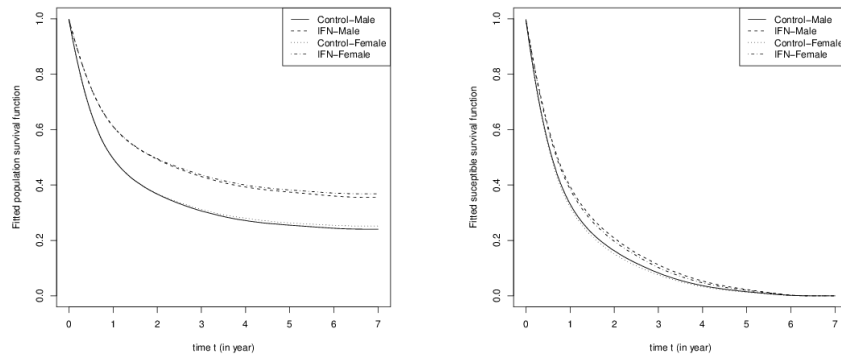
Figure 9: Melanoma e1684. Left : Comparaison of the fitted population survival function between groups. Right : Comparaison of the fitted survival function of the susceptible patients between groups.
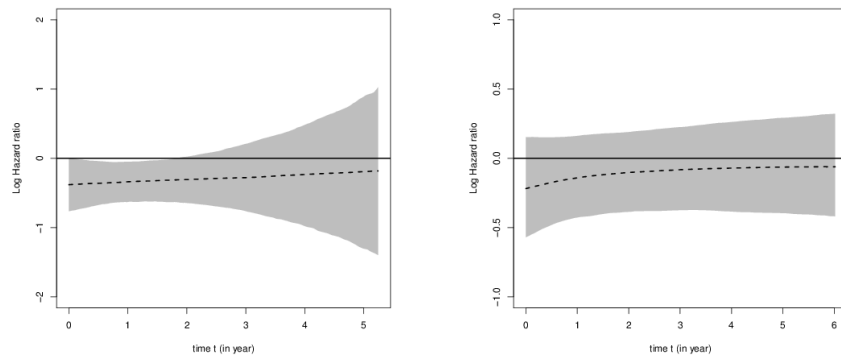

Figure 10: Melanoma e1684. Left : Fitted male population log-hazard ratio with 95% pointwize credible region. Right : Fitted log-hazard ratio for the male susceptible patients with 95% pointwize credible region.

7. Discussion

A flexible version of the promotion time model when the covariates influence simultaneously the probability of being cured and the time necessary for a cell to yield a detectable tumor was proposed. Although the suggested model does not have a proportional hazard structure at the patient level, our specification provides a smooth

25

estimation of the hazard ratios over time.

When the follow up of a study is not sufficiently long, one can use the promotion time model with some restrictions. In this context, it has been proved that of the effects of covariates are identifiable if they are not simultaneously used to model the probability to be cured and the time necessary to detect a tumor growing from a cancerous cell. The use of a logit link (instead of a log one) to model the probability to be cured was investigated to try to solve that identifiability problem but it was not successful.

Lopes and Bolfarine (2012) propose a parametric promotion time model to deal with hierarchical data. We currently work on the extension of the proposed flexible methodology in this context. An extension to interval censored data will also be considered.

Acknowledgements

Appendix A. Proof of lemma 1

A Proof of 1).

    a) Let $(\beta_0, \beta, \gamma, S_0)$ and $(\tilde{\beta}_0, \tilde{\beta}, \tilde{\gamma}, \tilde{S}_0)$ be two sets of parameters that satisfy (3) and let $\mathcal{X}$ be the set of all values of vector $\mathbf{x}$. We need to show that if

$$
\begin{aligned}
S_p(t|\mathbf{x}) &= \exp\left[-\exp(\beta_0 + \mathbf{x}^T\beta)\left(1 - S_0(t)^{\exp(\mathbf{z}^T\gamma)}\right)\right] \\
&= \exp\left[-\exp(\tilde{\beta}_0 + \mathbf{x}^T\tilde{\beta})\left(1 - \tilde{S}_0(t)^{\exp(\mathbf{z}^T\tilde{\gamma})}\right)\right] \\
&= \tilde{S}_p(t|\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{X} \quad ; \quad \forall t \in [0,\infty],
\end{aligned}
$$

then $\beta_0 = \tilde{\beta}_0$, $\beta = \tilde{\beta}$, $\gamma = \tilde{\gamma}$ and $S_0(t) = \tilde{S}_0(t)$ $\forall t \in [0,\infty]$.

    b) Since the exponential function is bijective, we only need to show that if $\forall \boldsymbol{x} \in \mathcal{X}$ and $\forall t \in [0,\infty]$, we have

$$
\exp(\beta_0 + \mathbf{x}^T\beta)\left(1 - S_0(t)^{\exp(\mathbf{z}^T\gamma)}\right) = \exp(\tilde{\beta}_0 + \mathbf{x}^T\tilde{\beta})\left(1 - \tilde{S}_0(t)^{\exp(\mathbf{z}^T\tilde{\gamma})}\right),
$$

then $\beta_0 = \tilde{\beta}_0$, $\beta = \tilde{\beta}$, $\gamma = \tilde{\gamma}$ and $S_0(t) = \tilde{S}_0(t)$ $\quad \forall t \in [0, \infty]$.

c) Since $F_0(t)$ is a proper cumulative distribution function (see A3), it follows immediately that $F(t|\mathbf{z}) = 1 - S_0(t)^{\exp(\mathbf{z}^T \gamma)}$ is also a proper cumulative distribution function. Let us proof that b) is true for $t = \infty$. Knowing that $F(t|\mathbf{z})$ and $\tilde{F}(t|\mathbf{z})$ are proper cumulative distribution functions, we have to show that if

$$\beta_0 + \mathbf{x}^T \beta \ = \ \tilde{\beta}_0 + \mathbf{x}^T \tilde{\beta} \quad \forall \mathbf{x} \in \mathcal{X},$$

then $\beta_0 = \tilde{\beta}_0$ and $\beta = \tilde{\beta}$.

Under A2, this is a straightforward consequence of :

$$P_{\mathcal{X}}(\mathbf{x}^T \beta = \alpha) = 1 \Rightarrow \alpha = 0 \text{ and } \beta = \mathbf{0}.$$

d) Since vector $\mathbf{z}$ does not include an intercept (see A1) and under A2, it is well known that the Cox proportional hazard model is identifiable.

e) Combining the results found in c) and d), we conclude the proof of 1).

B) Proof of 2).

a) The follow-up will be said unsufficiently long if the tumor growing from a cancerous cell cannot be detected by the end of the study at time t. It happens if $F_0(t)$ is too small (i.e. close to 0). Then,

$$
\begin{aligned}
F(t|\mathbf{z}) \ &= \ 1 - S_0(t)^{\exp(\mathbf{z}^T \gamma)} \\
&= \ 1 - \left[(1 - F_0(t))^{\exp(\mathbf{z}^T \gamma)} \right] \\
&= \ 1 - \left[1 - \exp\left(\mathbf{z}^T \gamma\right) F_0(t) \right] + o\left(F_0(t)\right) \\
&\approx \ 1 - \left[1 - \exp\left(\mathbf{z}^T \gamma\right) F_0(t) \right] \\
&= \ \exp\left(\mathbf{z}^T \gamma\right) F_0(t).
\end{aligned}
$$

b) Thus, using this approximation, (3) becomes

$$
\begin{aligned}
S_p(t|\mathbf{x}, \mathbf{z}) \ &\approx \ \exp\left[-\theta(\mathbf{x}) \exp\left(\mathbf{z}^T \gamma\right) F_0(t) \right] \\
&= \ \exp\left[- \exp(\beta_0 + \mathbf{x}^T \beta) \exp\left(\mathbf{z}^T \gamma\right) F_0(t) \right] \\
&= \ \exp\left[- \exp(\beta_0 + \mathbf{x}^T \beta + \mathbf{z}^T \gamma) F_0(t) \right].
\end{aligned}
$$

Then, we conclude that if vectors $\mathbf{x}$ and $\mathbf{z}$ share some components, the estimations of the effects of covariates are not identifiable when the follow-up is not sufficiently long.

c) It remains to prove that the estimations of the effects of covariates are identifiable when the follow up of the study is not sufficienly long and when vectors $\mathbf{x}$ and $\mathbf{z}$ do not share a single component.

To ensure A3, it is custom to force the *zero tail constraint* : One assumes $\hat{S}_0(t)$ to be 0 beyond the last event time t_{max} (Taylor (1995), Zeng et al. (2006), Ma and Yin (2008)). When the sufficient follow up assumption is not satisfied, this constraint is strong and has some consequences :

$$
\begin{aligned}
\theta(\mathbf{x})F(t|\mathbf{z})_{|t_{max}} &\approx \exp(\beta_0 + \mathbf{x}^T\beta + \mathbf{z}^T\gamma)F_0(t)_{|t_{max}} \\
&= \exp(\beta_0 + \mathbf{x}^T\beta + \mathbf{z}^T\gamma + \log(1-\epsilon)) \qquad \text{(A.1)} \\
&= \exp(\tilde{\beta}_0 + \mathbf{x}^T\beta + \mathbf{z}^T\gamma),
\end{aligned}
$$

where $\epsilon = F_0(+\infty) - F_0(t_{max})$.

In conclusion, the estimations of the covariate effects on the cure probability and on failure time for a cancerous cell are not affected by the zero tail constraint and are thus identifiable if vectors $\mathbf{x}$ and $\mathbf{z}$ do not share some components.

This concludes the proof of lemma 1.

References

[1] Berkson, J. and Gage, R.P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association* **47,** 501–515.

[2] Cetinyurek, A. and Lambert, P (2011). Smooth estimation of survival functions and hazard ratios from interval-censored data using Bayesian penalized B-splines *Statistics in Medecine* **30,** 75–90.

[3] Chen, M.-H., Ibrahim, J.G. and Sinha, D. (1999). A New Bayesian Model for Survival Data with a Surviving Fraction. *Journal of the American Statistical Association* **94,** 909–919.

[4] Cooner, F., Banerjee, S., Carlin, B.P. and Sinha, D. (2007). Flexible cure rate modelling under latent activation schemes. *Journal of the American Statistical Association* **102,** 560–572.

[5] Eilers, P.H.C (2007). Ill-posed problems with counts, the composite link model and penalized likelihood. *Statistical Modelling* **7,** 239–254.

[6] Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties (with discussion). *Statistical Science* **11,** 89–121.

[7] Farewell, V.T. (1982). The Use of Mixture Models for the Analysis of Survival Data with Long-Term survivors. *Biometrics* **38,** 1041–1046.

[8] Farewell, V.T. (1986). Mixture models in survival analysis : Are they worth the risk? *The Canadian Journal of Statistics* **14,** 257–262.

[9] Gelman, A., Roberts, G.O. and Gilks, W.R. (1996). Efficient Metropolis jumping rules. *Bayesian Statistics 5* (ed JM Bernardo, JO Berger, AP Dawid, and AFM Smith). Oxford University Press, UK.

[10] Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4* (ed JM Bernardo, JO Berger, AP Dawid, and AFM Smith). Clarendon Press, Oxford, UK.

[11] Haario, H., Saksman, E. and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli* **7,** 223–242.

[12] Ibrahim, J.G., Chen, M.-H. and Sinha, D. (2001). Bayesian Semiparametric Models for Survival Data with a Cure Fraction. *Biometrics* **57,** 383–388.

[13] Jullion, A. and Lambert, P. (2007). Robust specification of the roughness penalty prior distribution in spatially adaptive Bayesian P-splines models. *Computational Statistics and Data Analysis* **51,** 2542–2558.

[14] Kirkwood, J.M., Strawderman, M.H., Ernstoff, M.S., Smith, T.J., Borden, E.C. and Blum, R.H. (1996). Interferon Alpha-2b Adjuvant Therapy of High-Risk Resected Cutaneous Melanoma : The Eastern Cooperative Oncology Group Trial e1684. *Journal of Clinical Oncology* **14,** 7–17.

[15] Kuk, A.Y.C, Chen, M.H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika* **79,** 531–541.

[16] Lambert, P. and Eilers, P. (2005). Bayesian proportional hazards model with time varying regression coefficients: a penalized Poisson regression approach. *Statistics in Medicine* **24,** 3977–3989.

[17] Lambert, P. (2007). Archimedean copula estimation using Bayesian splines smoothing techniques. *Computational Statistics and Data Analysis* **51,** 6307–6320.

[18] Lambert, P. (2013). Nonparametric additive location-scale models for interval censored data. *Statistics and Computing* **23,** 75–90.

[19] Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics* **13,** 183–212.

[20] Li, C.S. and Taylor J.M.G (2002). A semi-parametric accelerated failure time cure model. *Statistics in Medicine* **21,** 3235–3247.

[21] Liu, H. and Shen, Y. (2009). A Semiparametric Parametric Regression Cure Model for Interval-Censored Data. *Jounal of the American Statistical Association* **104,** 1168–1178.

[22] Lopes, C.C.M and Bolfarine, H. (2012). Random effects in the promotion times cure rate models. *Computational Statistics and Data Analysis* **56** 75-87.

[23] Lu, W. (2010). Efficient Estimation for an accelerated failure time model with a cure fraction *Statistica Sinica* **20,** 661–674.

[24] Ma, Y. and Yin, G. (2008). Cure Rate Model with Mismeasured Covariates Under Transformation. *Journal of the American statistical association* **103,** 743–756.

[25] Peng, Y (2003). Estimating baseline distribution in proportional hazards cure models. *Computational Statistics and Data Analysis* **42,** 187–201.

[26] Peng, Y and Dear, K.B.G (2000). A Nonparametric Mixture Model for Cure Rate Estimation. *Biometrics* **56,** 237–243.

[27] Roberts, G.O., Rosenthal, J.S. (2001). Optimal scaling for various Metropolis Hastings algorithms. *Statistical Science* **16,** 351–367.

[28] Sy, J.P. and Taylor, J.M.G. (2000). Estimation in a Cox Proportional hazards Cure Model. *Biometrics* **56,** 227–236.

[29] Taylor J.M.G.(1995). Semi-parametric Estimation in Failure Time Mixture Models *Biometrics* **51,** 899–907.

[30] Tsodikov, A.D. (1998). A Proportional Hazard Model Taking Account of Long-Term Survivors. *Biometrics* **54,** 1508–1516.

[31] Tsodikov, A.D. (2002). Semi-parametric model of long- and short-term survival: an application to the analysis of breast cancer survival in Utah by age and stage. *Statistics in Medecine* **21,** 895–920.

[32] Wang, L., Du, P. and Liang, H. (2012). Two-Component Mixture Cure Rate Model with Spline Estimated Nonparametric Components. *Biometrics* **68,** 726–735.

[33] Yakovlev, A.Y., Tsodikov, A.D. (1996). Stochastic Models for Tumor of Latency and Their Biostatistical Applications. *Singapore: World Scientific Publ.*

[34] Yin, G. and Ibrahim J.G. (2005). Cure Rate Models : A Unified Approach. *The Canadian journal of statistics* **33,** 559–570.

[35] Zeng, D., Yin, G. and Ibrahim J.G. (2006). Semiparametric Transformation Models for Survival Data with a Cure Fraction. *Journal of the American Statistical Association* **101,** 670–684.

[36] Zhang, J. Peng, Y. and Li, H. (2013) A new semiparametric estimation method for accelerated hazards mixture cure model. *Computational Statistics and Data Analysis* **59,** 95–102.