Estimation and identification issues in the promotion time cure model when the same covariates influence long- and short-term survival

Philippe Lambert*[1,2] and Vincent Bremhorst[2]

[1]*Faculté des Sciences Sociales, Méthodes quantitatives en sciences sociales, Université de Liège, Liège, Belgium.*
[2]*Institut de statistique, biostatistique et sciences actuarielles (ISBA), Université catholique de Louvain, Louvain-la-Neuve, Belgium.*

4 July 2018
Accepted for publication in *Biometrical Journal*
DOI: 10.1002/bimj.201700250

Abstract

The promotion time cure model is a survival model acknowledging that an unidentified proportion of subjects will never experience the event of interest whatever the duration of the follow-up. We focus our interest on the challenges raised by the strong posterior correlation between some of the regression parameters when the same covariates influence long- and short-term survival. Then, the regression parameters of shared covariates are strongly correlated with, in addition, identification issues when the maximum follow-up duration is insufficiently long to identify the cured fraction. We investigate how, despite this, plausible values for these parameters can be obtained in a computationally efficient way. The theoretical properties of our strategy will be investigated by simulation and illustrated on clinical data. Practical recommendations will also be made for the analysis of survival data known to include an unidentified cured fraction.

1 Introduction

When the follow-up of subjects in time-to-event studies is insufficiently long, one cannot reasonably claim that the surviving units are 'cured', i.e. that they will never experience the event of interest whatever the duration of their follow-up. Given the uncertainty on the status of these units, estimating the cure probability can be challenging. However, it does not mean that nothing can be said about it as the number of failures and their timing dynamically inform us not only on the risk evolution, but also on upper (and to some extend lower) bounds for the cure probability. More importantly, non negligible information can be extracted on relative values of conditional risks and of cure probabilities for different covariate profiles. We propose to discuss these issues in the framework of the *bounded cumulative hazard* (BCH) model, also named the *promotion time cure model*.

The promotion time (cure) model was initially motivated by a biological model to analyze the time-to-relapse in cancer studies (Yakovlev and Tsodikov, 1996; Tsodikov, 1998;

*Corresponding author: Philippe Lambert, p.lambert@ulg.ac.be, Phone: +32-4-3665990

Chen *et al.*, 1999; Tsodikov *et al.*, 2003). Assume for example that, after a tumor removal, a subject is still potentially exposed to $N \sim \text{Pois}(\theta)$ undetected latent cancer cells that require independent and identically distributed times $W_1, \ldots, W_N \sim F$ to become detectable tumours. Then, $Y = \min\{W_i : i = 1, \ldots, N\}$ is the time required to diagnose a relapse and it has an improper survival distribution

$$\begin{aligned} S_p(y) &= \Pr(Y > y) = \Pr(N = 0) + \Pr(W_1 > y, \ldots, W_N > y, N \geq 1) \\ &= \mathrm{e}^{-\theta} + \sum_{N=1}^{+\infty} \big(1 - F(y)\big)^N \, \mathrm{e}^{-\theta} \frac{\theta^N}{N!} \\ &= \exp\{-\theta F(y)\}. \end{aligned}$$

The proportion of 'cured' subjects is given by $\Pr(N = 0) = S_p(+\infty) = \exp(-\theta)$. But the preceding biological motivation is not essential to come up with the proposed expression for the population survival function $S_p(t)$, opening its use in non-medical areas such as demography (Bremhorst *et al.*, 2016, 2018). Indeed, if a fraction of the population is really 'cured' or 'non-susceptible' (to experience the event of interest), then the underlying cumulative hazard $\Lambda(t)$ has a finite positive limiting value (say θ), yielding the former expression for $S_p(t)$ with the normalized cumulative hazard $F(t) = \Lambda(t)/\theta$. For that reason, the model is sometimes more explicitly named the *bounded cumulative hazard* (BCH) model (Tsodikov, 1998). We refer to the latter paper for additional insightful information on the genesis of that family of cure survival models.

To describe the effect of covariates $\mathbf{x}$ on cure probability (or long-term survival), one usually takes a log-linear model for θ,

$$\log \theta(\mathbf{x}) = \eta_\theta(\mathbf{x}) = \beta_0 + \mathbf{x}'\boldsymbol{\beta}. \tag{1}$$

In the frequentist framework, Tsodikov (1998) suggests a profile likelihood approach to estimate the promotion time model for a completely unspecified distribution function $F(\cdot)$. This approach is extended by Liu and Shen (2009) when the reported survival times are interval censored. Within the Bayesian paradigm, Chen *et al.* (1999) suggest a parametric (Gamma or Weibull) distribution for $F(\cdot)$, while a piecewise constant function is proposed by Ibrahim *et al.* (2001).

Covariates $\mathbf{z}$ can also be assumed to affect the dynamics in the normalized cumulative hazard function $F(\cdot)$ (or, in other words, short-term survival). This can be done e.g. using a Cox proportional hazards model (Cox, 1972),

$$F(y|\mathbf{z}) = 1 - S_0(y)^{\exp(\eta_F(\mathbf{z}))} \quad \text{with } \eta_F(\mathbf{z}) = \mathbf{z}'\boldsymbol{\gamma}, \tag{2}$$

where $S_0(t)$ denotes the baseline survival function (when $z = 0$). Tsodikov (2002) proposed, in a frequentist framework, a nonparametric estimation of the baseline survival function $S_0(t)$. Using a Bayesian approach, Yin and Ibrahim (2005) suggested a flexible specification of $S_0(t)$ using a piecewise exponential distribution where the number of intervals is selected via the conditional predictive ordinate (CPO) criterion, while Bremhorst and Lambert (2016) opt for a flexible and smooth specification of the baseline distribution using Bayesian P-splines (Eilers and Marx, 1996, 2010; Lang and Brezger, 2004; Jullion and Lambert, 2007) with covariates appearing simultaneously in (1) and (2). They also prove the identifiability of that promotion time model when the follow-up study is sufficiently long and, in particular when the covariate vectors $\mathbf{x}$ and $\mathbf{z}$ potentially share some components.

Starting from this last result, one of our goals is to point the identifiable quantities in the promotion time cure model when the follow-up is insufficiently long and when the same covariates potentially affect the cure probability and the dynamics in the population hazard function, see Sections 2.1 and 2.2. Based on this information, a reparametrization of the conditional promotion time cure model is suggested in Section 2.3 with some practical recommendations for the modelling of survival data in the presence of an unidentified cured

fraction. That reparametrization turns to be useful whatever the duration of the follow-up as it substantially reduces the posterior correlation of regression parameters and, hence, improves the mixing of Markov chain Monte Carlo in a Bayesian framework or facilitates the convergence of nonlinear optimizers in likelihood based estimation procedures. The choice of priors for the regression parameters is discussed in Section 2.4 as slightly informative and meaningful specifications for these distributions can help to avoid the exploration of unrealistic parameter combinations during the inferential process, see Section 2.5. An extensive simulation study in Section 3 will not only confirm the theoretical expectations regarding the identification issues, but also lead to additional practical recommendations for the analysis of survival data with a cured fraction, in particular when the follow-up is insufficiently long, see also Section 5. Times to recurrence in patients with colon cancer are analyzed in Section 4 to illustrate the proposed modelling strategy. We conclude the paper with a discussion and practical recommendations in Section 5.

2 The conditional promotion time cure model

2.1 Definition

Denote by Y_i and C_i the event and right censoring times for the ith of n units ($i = 1, \ldots, n$). Further assume that Y_i and C_i are independent given covariate values $(\mathbf{x}_i', \mathbf{z}_i')$ in $\mathbb{R}^{p_1+p_2}$ and that only (T_i, δ_i) is observed where $T_i = \min\{Y_i, C_i\}$ and $\delta_i = \mathrm{I}(Y_i < C_i)$. Given the covariates, the conditional promotion time cure model assumes that the population survival function is given by

$$\begin{aligned} S_p(y_i|\mathbf{x}_i, \mathbf{z}_i) &= \Pr(Y_i > y_i|\mathbf{x}_i, \mathbf{z}_i) \\ &= \exp\{-\theta(\mathbf{x}_i)F(y_i|\mathbf{z}_i)\}, \end{aligned} \tag{3}$$

where $F(\cdot|\mathbf{z}_i)$ is a conditional distribution function, see Section 1 for a motivation of this expression. Hence, $\exp\{-\theta(\mathbf{x}_i)\} = S_p(+\infty|\mathbf{x}_i, \mathbf{z}_i)$ is the conditional probability to be cured for some positive function $\theta(\cdot)$ of the covariates.

For the inclusion of baseline covariates, we follow Bremhorst and Lambert (2016) and Bremhorst *et al.* (2016) with the log-linear model (1) for the cure probability and a (flexible) proportional hazards (PH) model (2) to complete the specification of the event time distribution. The baseline survival function,

$$S_0(y_i) = \exp\left(-\int_0^{y_i} h_0(t)\mathrm{d}t\right) = \exp\left\{-\int_0^{y_i} \exp\left(\sum_{k=1}^{K} b_k(t)\phi_k\right)\mathrm{d}t\right\}, \tag{4}$$

is specified through the log of the baseline hazard $h_0(t)$ (Rosenberg, 1995) using a cubic B-spline basis $\{b_k(\cdot) : k = 1, \ldots, K\}$ on $(0, t_{\max})$ where $t_{\max}$ denotes the considered minimum follow-up duration required to ensure that an event-free subject by that time will not experience the event of interest. Mathematically, we assume that $t_{\max}$ is the smallest value of y ensuring that $S_0(y) = 0$. In the context of a cancer study, this might be, for example, 9 years after surgery if a subject is (arbitrarily) defined to be 'cured' when no event is reported by that point in time. Note that the proposed strategy slightly differs from Bremhorst and Lambert (2016) where the largest knot was chosen to be the *observed* maximum follow-up time for the data at hand. However, the two definitions will roughly coincide if the study is designed to ensure that any susceptible unit would be observed to experience the event of interest if monitored up to the maximum follow-up time. Following Eilers and Marx (1996, 2010), we take a large number K of equidistant splines on $(0, t_{\max})$ and penalize rth-order differences of successive B-spline coefficients $\phi = (\phi_1, \ldots, \phi_K)'$ when it comes to estimation, yielding penalized B-splines (shortly named ***P-splines***), see Section 2.4.2 for more details.

2.2 Identification issues

The population survival function in (3) and the associated cumulative hazard, $H_p(y|\mathbf{x},\mathbf{z}) = -\log S_p(y|\mathbf{x}_i,\mathbf{z})$ are key expressions to understand identification issues in the promotion time cure model as they have quite straightforward 'observable' counterparts that can be obtained from the data $\mathcal{D} = \left\{(t_i,\delta_i,\mathbf{x}_i',\mathbf{z}_i') : i = 1,\ldots,n\right\}$ using e.g. the nonparametric Kaplan-Meier estimator (Kaplan and Meier, 1958). It implies that $H_p(y|\mathbf{x},\mathbf{z})$ can be estimated from the data, but not necessarily the two factors $\theta(\mathbf{x}_i)$ and $F(y_i|\mathbf{z}_i)$ entering its definition in the promotion time model, see (3). To fix the ideas, assume that the two sets of regressors built from the same 3 covariates share a single component, more specifically that

$$\begin{array}{rcccccccc} \eta_\theta(\mathbf{x}) & = & \beta_0 + \mathbf{x}'\boldsymbol{\beta} & = & \beta_0 & + & \beta_1 x_1 & + & \beta_2 x_2 \\ \eta_F(\mathbf{x}) & = & \mathbf{x}'\boldsymbol{\gamma} & = & 0 & + & \gamma_2 x_2 & + & \gamma_3 x_3 \,. \end{array}$$

When the follow-up duration t of a susceptible unit is not large enough to ensure that the event was observed, translating in a small value for the baseline c.d.f. $F_0(t) = 1 - S_0(t)$ in the preceding Cox regression model, one can show, using a MacLaurin series, that

$$\begin{aligned} F(t|\mathbf{x}) = 1 - S(t|\mathbf{x}) &= 1 - (1 - F_0(t))^{\exp(\eta_F(\mathbf{x}))} \\ &= \exp\big(\eta_F(\mathbf{x})\big) F_0(t) + \mathcal{O}\left(F_0(t)^2\right). \end{aligned}$$

Therefore, for such values of t, the log of the population cumulative hazard is

$$\begin{aligned} \log H_p(t|\mathbf{x}) &= \log\big(-\log S_p(t|\mathbf{x})\big) = \log\theta(\mathbf{x}) + \log F(t|\mathbf{x}) \\ &\approx \eta_\theta(\mathbf{x}) + \eta_F(\mathbf{x}) + \log F_0(t) \\ &= (\beta_0 + \log F_0(t)) + \beta_1 x_1 + (\beta_2 + \gamma_2) x_2 + \gamma_3 x_3 \,. \end{aligned} \tag{5}$$

This suggests that, when the follow-up is insufficiently long (to ensure that the last monitored units will never experience the event of interest), there are identification issues for the intercept β_0 as there there are no such values for the reported event or right censoring times at which the summed $\log F_0(t)$ is theoretically known. The same is true for the regression coefficients of shared covariates (here, x_2) as they only appear summed in (5) for a small value of t. On the other hand, no identification problems are expected for β_1 and γ_3 as they correspond to non-shared regressors.

Note also that, for small values of t, (5) indicates that everything works as if a proportional hazards model was assumed at the population level with baseline (population) hazard $\mathrm{e}^{\beta_0} f_0(t)$ such that

$$h_p(t|\mathbf{x}) \approx \mathrm{e}^{-\beta_0 + \eta_\theta(\mathbf{x}) + \eta_F(\mathbf{x})}\big(\mathrm{e}^{\beta_0} f_0(t)\big) = \mathrm{e}^{\mathbf{x}'(\boldsymbol{\beta}+\boldsymbol{\gamma})}\big(\mathrm{e}^{\beta_0} f_0(t)\big). \tag{6}$$

For an arbitrary value of t, one has

$$h_p(t|\mathbf{x}) = \mathrm{e}^{\mathbf{x}'(\boldsymbol{\beta}+\boldsymbol{\gamma})}\big(\mathrm{e}^{\beta_0} f_0(t)\big) S_0(t)^{\exp(\boldsymbol{\gamma}'\mathbf{x})-1} \,, \tag{7}$$

where the last factor in the expression induces time-varying (conditional) hazard ratios for different values of a covariate x_k associated to a non-zero value for γ_k.

2.3 Model reparametrization

The preceding identification issues suggest to reparametrize the model, as disentangling covariate effects on the cure probability (i.e. on long-term survival) or on the normalized cumulative hazard dynamics (NCHD) (i.e. on short-term survival) will remain computationally challenging even with longer follow-ups. It will improve the mixing of Monte Carlo Markov chains (MCMC) when exploring the joint posterior under the Bayesian paradigm,

or facilitate the convergence of algorithms used to compute maximum likelihood estimators in a frequentist framework. Given that a pre-defined zero-value for a component in $\boldsymbol{\beta}$ (resp. $\boldsymbol{\gamma}$) indicates that the corresponding regressor is not part of the covariates $\mathbf{x}$ in $\mathbb{R}^p$ entering the definition of $\theta(\mathbf{x})$ (resp. $F(t|\mathbf{z})$), consider $(\beta_0, \boldsymbol{\beta}, \boldsymbol{\gamma}) \longrightarrow (\beta_0, \boldsymbol{\psi}, d\boldsymbol{\psi})$ where, for $k = 1, \ldots, p$,

$$\psi_k = \begin{cases} \beta_k & \text{if} \quad \gamma_k = 0 \quad (k\text{th covariate only in (1)}) \\ \gamma_k & \text{if} \quad \beta_k = 0 \quad (k\text{th covariate only in (2)}) \\ (\beta_k + \gamma_k)/2 & \text{if} \quad \beta_k \gamma_k \neq 0 \quad (k\text{th covariate shared by (1) and (2)}). \end{cases}$$

Then, when $\beta_k \gamma_k \neq 0$, one has $\beta_k = \psi_k + d\psi_k$ and $\gamma_k = \psi_k - d\psi_k$ where $d\psi_k = (\beta_k - \gamma_k)/2$. In the illustrative example of Section 2.2 with the new parametrization and under an insufficiently long follow-up, we expect that limited information can be extracted from the data on components β_0 and $d\psi_2$ (with, thus, bias or relatively large variances for any relevant estimators of these quantities). However, it does not imply that sets of plausible values for these parameters cannot be obtained, as will be illustrated by the simulation study in Section 3.

Concerning the baseline distribution function, $F_0(t) = 1 - S_0(t)$, in (4), we force it to be one whenever $t \geq t_{\max}$. Given the assumed smoothness of the associated baseline hazard $h_0(\cdot)$, the uncertainty on β_0 (that can be identified when summed to $\log F_0(t)$ in (5)) will be reduced when the effective follow-up duration approaches values of t for which $\log F_0(t)$ is close to zero.

2.4 Prior elicitation

2.4.1 Regression parameters

The uncertainty on β_0 and shared regression parameters can be very large. The time spent on a careful elicitation of priors for these parameters is usually worthwhile as it can avoid putting posterior probability mass on meaningless values for parameters or combination of them. Assume for example that $\mathbf{x} = 0$ corresponds to a meaningful combination of the regressors. If helpful, this could be achieved by subtracting from a continuous component a reference value such as its sample mean. Then, β_0 has a clear interpretation in terms of the baseline probability to be cured,

$$\beta_0 = \log \theta(\mathbf{x} = 0) = \log\left(-\log \Pr(\text{cured}|\mathbf{x} = 0)\right) = g\left(\Pr(\text{cured}|\mathbf{x} = 0)\right),$$

that can be used to build an informative prior for β_0. Assume for example that one can claim that the baseline cure probability is with a large probability (say, about 95%) in $(p_0^{\min}, p_0^{\max}) = (.01, .30)$. This could tentatively be translated by the following normal prior for β_0,

$$\beta_0 \sim \mathcal{N}\left(\frac{b_0^{\min} + b_0^{\max}}{2}, \left(\frac{b_0^{\max} - b_0^{\min}}{2 \times 1.96}\right)^2\right) = \mathcal{N}(.86, .335^2), \tag{8}$$

where $b_0^{\min} = g\big(p_0^{\max}\big)$ and $b_0^{\max} = g\big(p_0^{\min}\big)$ correspond to the limiting cure probabilities on the β_0-scale. That would imply that, a priori, β_0 and the baseline cure probability have a 95% probability to be in $(b_0^{\min}, b_0^{\max})$ and in $(p_0^{\min}, p_0^{\max})$, respectively. A similar exercise could be made separately for the shared regression parameters, β_2 and γ_2 in the example of Section 2.2, yielding independent normal priors for these quantities. It will avoid putting a non negligible posterior mass on sub-regions of (β_2, γ_2) where unrealistically large (resp. small) values of β_2 are associated to (partially compensating) unrealistically small (resp. large) values of γ_2. In particular, when the follow-up is insufficiently long, these parameters (approximately) only enter the model likelihood through their sum $(\beta_2 + \gamma_2)$, see Eq. (5). Alternatively, this could be made in the $(\psi_2, d\psi_2)$ parametrization

with the help of the proportional hazard approximation to the conditional population hazard, see (6). Indeed, when $F_0(t)$ is small, twice ψ_2 can be (approximately) interpreted as the log of a conditional hazard ratio,

$$\exp(2\psi_2) = \exp(\beta_2 + \gamma_2) \approx \frac{h_p(t|x_1, x_2 = s+1, x_3)}{h_p(t|x_1, x_2 = s, x_3)},$$

which is helpful to realize what might be (un)realistic values for such a quantity.

2.4.2 Spline parameters

As indicated in (4), the log of the baseline hazard is specified using a linear combination of B-splines associated to a large number of equidistant knots on $(0, t_{\max})$,

$$\log h_0(t) = \sum_{k=1}^{K} b_k(t)\phi_k.$$

The flexibility induced by that large number of B-splines can be counterbalanced by penalizing changes in rth (typically 2nd or 3rd) order differences of B-splines coefficients,

$$\Delta^r \phi_k = \Delta^{r-1}\phi_k - \Delta^{r-1}\phi_{k-1} \text{ with } \Delta^1\phi_k = \phi_k - \phi_{k-1},$$

yielding penalized B-splines or *P-splines*. In a frequentist context, this can done by adding a penalty,

$$\text{pen}(\boldsymbol{\phi}|\lambda) = -\frac{\lambda}{2}\sum_k (\Delta^r \phi_k)^2 = -\frac{\lambda}{2}\boldsymbol{\phi}' D_r' D_r \boldsymbol{\phi}, \tag{9}$$

to the log-likelihood (Eilers and Marx, 1996, 2010). In a Bayesian framework, it translates into a smoothness prior for $\boldsymbol{\phi}$ (Lang and Brezger, 2004),

$$p(\boldsymbol{\phi}|\lambda) \propto \exp\left(-\frac{\lambda}{2}\boldsymbol{\phi}' D_r' D_r \boldsymbol{\phi}\right).$$

A robust prior for the penalty parameter λ is recommended by Jullion and Lambert (2007),

$$(\lambda|\delta) \sim \mathcal{G}(.5\nu, 5\nu\delta) \;\; ; \;\; \delta \sim \mathcal{G}(a_\delta, b_\delta),$$

as results were shown to be potentially sensitive to the choice of hyperparameters with the simple gamma priors suggested in Lang and Brezger (2004). One can show that, when $\nu = 1$ and $a_\delta = b_\delta = .5$, the marginal prior density for λ is

$$p(\lambda) = \int_0^{+\infty} p(\lambda|\delta)p(\delta)\mathrm{d}\delta \propto \frac{1}{\sqrt{\lambda}(1+\lambda)}, \tag{10}$$

i.e. $\lambda \sim \text{BetaPrime}(.5, .5)$, or equivalently, $\lambda/(1+\lambda) \sim \text{Beta}(.5, .5)$. Given the symmetric U-shape of the corresponding Beta density, this indicates that some more prior weight is set symmetrically on values of λ close to 0 or tending to $+\infty$, corresponding to no penalty ($\lambda = 0$) or an extremely large one ($\lambda \to +\infty$), yielding a polynomial of order $(r-1)$ for $\log h_0(t)$ in the latter case. Note that (10) is equivalent to taking a half-Cauchy prior (Polson and Scott, 2012) for $\sqrt{\lambda}$, thereby expressing the limited prior information assumed on λ using this particular specification of the Jullion and Lambert (2007) prior.

2.5 Inference

We focus on likelihood-based inferential methods. Given the data setting and model specification in Section 2.1, the log-likelihood is

$$\log L(\beta_0, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi}; \mathcal{D}) = \sum_{i=1}^{n} \left\{ \delta_i \log h_p(t_i|\mathbf{x_i}, \mathbf{z_i}) - H_p(t_i|\mathbf{x_i}, \mathbf{z_i}) \right\},$$

with the following expressions for the (population) cumulative hazard and hazard functions,

$$\begin{aligned} H_p(t_i|\mathbf{x_i}, \mathbf{z_i}) &= \theta(\mathbf{x_i}; \beta_0, \boldsymbol{\beta}) F(t_i|\mathbf{z_i}; \boldsymbol{\phi}, \boldsymbol{\gamma}) \\ &= \exp(\beta_0 + \boldsymbol{\beta}'\mathbf{x_i}) \left\{ 1 - \exp\left(- \exp(\boldsymbol{\gamma}'\mathbf{z}_i) H_0(t_i; \boldsymbol{\phi}) \right) \right\}; \\ h_p(t_i|\mathbf{x_i}, \mathbf{z_i}) &= \exp(\beta_0 + \boldsymbol{\beta}'\mathbf{x_i} + \boldsymbol{\gamma}'\mathbf{z_i})\, h_0(t_i; \boldsymbol{\phi}) \exp\left(- \exp(\boldsymbol{\gamma}'\mathbf{z}_i) H_0(t_i; \boldsymbol{\phi}) \right). \end{aligned}$$

In a frequentist setting, the penalty (9) is added to the log-likelihood. The penalized log-likelihood can be maximized using e.g. a Newton-Raphson algorithm to obtain parameter estimates for a given value of the penalty parameter λ.

In a Bayesian framework, given the potentially large posterior correlation between regression parameters, we strongly suggest to reparametrize the model along the suggestions made in Section 2.3. The joint posterior is given by

$$\begin{aligned} &p(\beta_0, \boldsymbol{\psi}, d\boldsymbol{\psi}, \boldsymbol{\phi}, \lambda, \delta|\mathcal{D}) \\ &\quad \propto L(\beta_0, \boldsymbol{\psi}, d\boldsymbol{\psi}, \boldsymbol{\phi}; \mathcal{D})\, p(\beta_0)\, p(\boldsymbol{\psi})\, p(d\boldsymbol{\psi})\, p(\boldsymbol{\phi}|\lambda)\, p(\lambda|\delta)\, p(\delta). \end{aligned}$$

Except for λ and δ for which

$$\begin{aligned} (\lambda|\ldots, \mathcal{D}) &\sim \mathcal{G}\left(\frac{\nu + K}{2}, \frac{\nu\delta + \boldsymbol{\phi}' D_r' D_r \boldsymbol{\phi}}{2} \right), \\ (\delta|\ldots, \mathcal{D}) &\sim \mathcal{G}\left(a_\delta + \frac{\nu}{2}, b_\delta + \frac{\nu\delta}{2} \right), \end{aligned}$$

the conditional posteriors of the other model parameters do not belong to familiar distributions. Therefore, we suggest to explore the joint posterior using a Metropolis-within-Gibbs algorithm with Gibbs steps for λ and δ and Metropolis steps for the other parameters. A good starting value can be obtained by maximizing the joint posterior for given rough guesses of λ and δ. The inverse Hessian at the so-obtained (conditional) posterior mode can be used to improve the mixing of the chains by making Metropolis steps along the eigenvectors of that matrix. The variances of the proposal distributions in the Metropolis steps are tuned automatically using the adaptive procedure proposed by Haario *et al.* (2001) during the burn-in to achieve the targeted optimal acceptance rates (Gelman *et al.*, 1996; Roberts and Rosenthal, 2001). Strategies relying on Laplace approximations could also be used (Gressani and Lambert, 2018).

3 Simulation study

A large simulation study was performed to confirm the identification issues suggested by the theory of Section 2.2 and to evaluate the merits of the strategy proposed in Sections 2.3 to 2.5 to obtain plausible values for the model parameters.

The data were generated using the conditional promotion time cure model defined in Section 2.1 using three independent covariates $X_1, X_2, X_3 \sim \mathcal{N}(0, 1)$. The first two regressors were used to define the cure probability by taking $\eta_\theta(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ with $(\beta_0, \beta_1, \beta_2) = (.75, -.50, .80)$, while only the last two regressors were assumed to

influence the detection of a given latent factor for a susceptible subject with $\eta_F(x_2, x_3) = \gamma_2 x_2 + \gamma_3 x_3$, where $(\gamma_2, \gamma_3) = (.40, -.40)$. Denote by Weib$(a, b)$ a Weibull distribution with shape and scale parameters a and b. The baseline $F_0(\cdot)$ distribution for the time-to-detection of a given latent factor was taken to be Weib$(2, 9)$ with mean 8.0 and standard deviation 4.2. The independent right censoring process was taken to be a Weib$(3, 25)$ truncated on $(0, t_R)$ where t_R denotes the maximum follow-up time.

Five-hundreds datasets were generated for two different sample sizes ($n = 300$ or 500) and three possible maximum follow-up durations ($t_R = 25$, 15 or 10 with $F_0(t_R)$ equal to .9995, .937 and .707, respectively), yielding increasing average proportions of right censored units (24%, 31% or 41%, respectively) including an average common 20% proportion of (unidentified) cured units.

In what follows, the support of the response was assumed to be $(0, t_{\max}) = (0, 30)$ with a basis of ten B-splines defined on that interval for the specification of the baseline hazard (cf. Section 2.4.2). Such a value for $t_{\max}$ is obviously large enough given that $F_0(30) = .999985$. We also use the recommendations following from Eq. (8) and Section 2.4 to specify the priors for the regression coefficients in the ψ–$d\psi$ parametrization:

$$\psi_0 = \beta_0 \sim \mathcal{N}(.86, .335^2) \; ; \; \psi_k, d\psi_k \sim \mathcal{N}(0, 1.5^2).$$

The first one implicitly assumes that the baseline cure probability is in $(.01, .30)$, while the other priors just ensure that completely unrealistic values are not generated during the exploration of the joint posterior using MCMC.

3.1 Correctly specified model

Assume that the correct model family is chosen (here: the promotion time cure model) and that, in addition, the set of regressors in the cured probability and Cox proportional hazards regression models are correctly specified, namely $\mathbf{x} = (x_1, x_2)$ and $\mathbf{z} = (x_2, x_3)$. As can be seen from the simulation results in Table 1 for two different sample sizes, when the maximum follow-up time t_R is sufficiently large (such as in Setting 1 where one can reasonably assume that all right-censored units at $t_R = 25$ correspond to cured ones given that $F_0(25) = .9995$), the posterior means taken as estimators of the regression parameters in the two parametrizations have excellent frequentist properties: biases are virtually zero and the coverage of 95% credible intervals are close to their nominal values. Note also that the empirical standard errors for β_2 and γ_2 (sharing the same covariate x_2) tend to be relatively large.

However, when the duration of the follow-up is really short such as in Setting 3 (with $F_0(t_R) = .707$), one can see (as expected from the theory in Section 2.2) that the estimators for the regression parameters β_2 and γ_2 sharing the same covariate x_2 have relatively large standard errors and display biases that tend to compensate. However, their credible intervals have coverages close to their nominal values, suggesting that the exploration of their posterior distributions was realized properly by pointing in a satisfactory way plausible values for β_2 and γ_2 (given the observed data). One can also see that their (half) sum, ψ_2, is estimated without bias and that the associated 95% credible interval has a compatible frequentist coverage. Their (half) difference $d\psi_2$ is, not surprisingly, estimated with bias, but again with properly explored plausible values. As also announced by the theory, the intercept $\beta_0 = \psi_0$ cannot be estimated, but its sum to $\log F_0(t)$ can when $t \leq t_R$ as illustrated with $t = t_R$. Standard errors of regression coefficient estimators for shared covariates also tend to be larger, as a consequence of the identification issue.

3.2 Misspecified model

Assume now that the correct model family is chosen, but that unnecessary regressors are included in the cured probability part and in the Cox proportional hazards model. More

specifically, let us wrongly assume that $\mathbf{x} = \mathbf{z} = (x_1, x_2, x_3)$ like when one does not know a priori whether a given covariate acts on long- or short-term survival.

Simulation results in Tables 2 & 3 clearly suggest that, whatever the considered follow-up duration, the regression parameters for the covariate x_1 acting only on cure probability are estimated without significant biases and with a frequentist coverage of the HPD credible intervals close to the 95% nominal value. In particular, the absence of an effect of x_1 on NCHD ($\gamma_1 = 0$) is clearly revealed, suggesting to drop x_1 from the regression model in (2).

Except when the follow-up is sufficiently long (such as in Setting 1 where all regression parameters are properly estimated), the estimation of a regression parameter associated to a covariate having (at least) a true effect on NCHD (i.e. on short-term survival) (such as x_2 and x_3) is biased with compensating biases for β_k and γ_k ($k = 2, 3$). While the credible intervals for the regression parameters β_2 and γ_2 associated to the truly shared covariate x_2 have the expected frequentist coverages, it is not the case for the intervals for β_3 and γ_3 associated to the covariate x_3 (truly) acting on NCHD only. The latter intervals tend to have an effective frequentist coverage slightly smaller than their nominal values. However, the (half) sum of these parameters, $\psi_3 = (\beta_3 + \gamma_3)/2$, has a credible interval with the expected coverage.

3.3 Conclusions

The lessons of this simulation study are particularly useful (and not only a reassuring indication that our code is working properly and with results in agreement with our prior theoretical expectations). Given that, in practice, one rarely knows in advance how covariates should be divided between the two regression model parts, one will most likely start with a misspecified model, cf. Section 3.2.

If the follow-up is sufficiently long, regression parameter estimation will proceed properly (with unbiased estimators and a good quantification of uncertainty) and provide reliable information on covariate effects.

On the other hand, if the follow-up is insufficiently long, one should be very careful when analyzing estimation results. Unless substantive knowledge indicates that a covariate only has a potential impact on cure probability or if strong informative priors are available for the regression parameters, one should only rely on point estimation for $\psi_k = (\beta_k + \gamma_k)/2$. In addition, slightly larger credibility levels should be considered to examine plausible values for the regression parameters to compensate for the potential undercoverage of intervals if the covariate only happens to (truly) impact short-term survival (and not the long-term cure probability).

Therefore, when the maximum follow-up is insufficiently long and unless the division of covariates between the two regression model parts is clear, the safest option is to consider the simplified promotion time cure model corresponding to a proportional hazards model with (population) survival function

$$S_p(t|\mathbf{x}, \boldsymbol{\varphi}) = \exp\left(-\exp(\mathbf{x}'\boldsymbol{\varphi})H_0(t)\right) = S_0(t)^{\exp(\mathbf{x}'\boldsymbol{\varphi})}, \tag{11}$$

where the baseline cumulative hazard $H_0(t)$ remains constant for $t \geq t_{\max}$, the maximum follow-up time required to observe the monitored event on a susceptible subject. The latter function $H_0(t)$ corresponds to $\mathrm{e}^{\beta_0} F_0(t)$ in the first order approximation discussed in Section 2.2, see Eq. (5). However, in the presence of a cured fraction, indications on the specific impact of the selected covariates can be obtained from the extended conditional promotion time model presented in Section 2.1 and credible regions with large nominal levels for the pairs (β_k, γ_k). Potential identification issues in that context and the preceding simulation results suggest that one can reasonably construct interpretations for covariate effects using credible intervals, but that one should be very cautious with point estimates.

4 Application

Moertel *et al.* (1990) and Moertel *et al.* (1991) report the results of a successful treatment of colon cancer based on the combination of Levamisole (used to treat parasitic worm infections) and Fluorouracil (5-FU), a now widely used medication to fight cancer. We focus here on the time to cancer recurrence for the $n = 888$ patients involved in the trial with complete information on sex, age, indicators of obstruction of the colon by the tumour and of adherence to nearby organs, differentiation of cancer cells, the extent of the cancer and the number of positive lymph nodes. The median follow-up time was 4.26 years with a maximum value slightly exceeding 9 years ($t_R = 9.12$). About half of the patients were observed to experience a cancer recurrence with the largest recurrence time observed after 7.38 years. The other 442 patients were right censored after a median censoring time of 6.27 years.

A plot of the Kaplan-Meier Meier estimates of the survival curves for the recurrence times under each treatment level, see Fig. 1, suggests that a cured fraction exists and that the addition of the chemotherapeutic agent 5-FU decreases the chance of a relapse. A first selection of the covariates was made using a stepwise procedure with the Cox proportional hazards model. The flexible promotion time cure model described in Section 2.1 was fitted to the preceding data with the pre-selected covariates simultaneously included to describe the cure probability and their effects on the population hazard dynamics. Priors were elicited for the regression parameters using the suggestions of Section 2.4. In particular, a Normal prior with mean -0.512 and standard deviation 0.494 was taken for β_0, translating a 95% prior belief that the cure probability is within (.20,.80) in the patient subgroup with baseline values for the covariates. Normal priors with mean 0 and standard deviation 1.5 were taken for the other regression parameters: they correspond to flat priors in the study context, but will prevent the generation of unrealistic $(\psi_k, d\psi_k)$ parameter values when generating Monte Carlo Markov chains (MCMC).

A Metropolis-within-Gibbs algorithm was used to sample the joint posterior. The model was reparametrized following the recommendations in Section 2.3, enabling the generation of well mixing chains. A chain of length 10,000 followed a burn-in of 2,000 iterations during which the standard deviations of the univariate normal proposal distributions were tuned (Haario *et al.*, 2001) to reach the desired acceptance probabilities (Roberts and Rosenthal, 2001). A summary of the MCMC results can be found in Table 4 giving the estimated posterior mean, standard deviation and several quantiles for each of the regression parameters. The significance of a contrast between a covariate category and its reference value was measured using the minimum of the posterior probabilities of the parameter to be larger and smaller than 0, respectively, with bolded values when it is smaller than 0.05/2.

Given that the Cox PH model is a special case of the promotion time cure model, it is not surprising to discover that each covariate is associated to a least one regression parameter with only non-zero plausible values (given by a 95% credible interval). Starting from there, the model was progressively simplified by regrouping covariate categories (such as the *Control* and *Levamisole* treatment levels, the first two categories for the number of positive lymph nodes, and the *Submucosa* and *Muscle* categories in the description of the extent of the cancer) and by determining whether they were significantly influencing the probability to be cured or/and NCHD. It led to the final flexible cure promotion time model in Table 5. We can conclude from it that treatment, the number of positive lymph nodes (`Nodes`) and the extent of the cancer significantly influence the probability of 'cure', i.e. of not experiencing a cancer recurrence within (about) 9 years. The combination of Levamisole and 5-FU significantly improves prognosis (as compared to Levamisole alone or the absence of treatment), while the risk of relapse significantly increases with the spread of the cancer (qualitatively described by the extent of the tumour into the wall of the colon, and also quantified by the number of contaminated lymph nodes).

From the bottom of Table 5, we conclude that the time to recurrence also tends to be

significantly smaller when the number of positive lymph nodes exceeds two. In particular, it suggests that hazard ratios contrasting different categories of `Nodes` are not proportional (conditionally on the other covariates), but change over time. The other selected covariate suggests that a recurrence tends to occur faster with patients with poorly differentiated cancer cells (conditionally on other covariates), but its absence in the upper panel of the table reveals that it does not significantly influence the long term risk of recurrence.

Our results are more insightful than with traditional analyses based e.g. on the Cox proportional hazards model where one does not try to disentangle the effect of covariates on long term risk from their short term influence on hazard dynamics.

To illustrate the influence of the maximum follow-up duration on the estimation of the regression parameters, the preceding selected cure survival model was refitted on the same data with the response artificially right censored at increasing values of t_R (between 1 and 9 years by steps of six months) when an individual follow-up originally exceeds t_R. The plots of the so-obtained posterior means and 95% credible intervals are available in Fig. 2. The first two rows, related to the quantification of the `Nodes` effects, clearly illustrate the large uncertainty in the estimation of β_k and γ_k when the corresponding covariate simultaneously appears in the two regression model parts and when the maximum follow-up t_R is insufficiently long. In addition, a large (resp. small) value for β_k tends to be compensated by a small (resp. large) one for γ_k as a consequence of the identification issues anticipated in Section 2.2. On the other hand, the estimate for the sum of these coefficients, $\beta_k + \gamma_k (= 2\psi_k)$, does not change much with t_R and has good theoretical properties (in terms of bias and effective coverage of credible intervals, cf. Section 3). The same is true for regression parameters associated to covariates involved only in one of the two regression model parts (see the bottom row in Fig. 2). The evolution with t_R of the estimates of the regression parameter φ_k in the Cox PH model with the same covariates can also be visualized as dashed curves on the same figure. In the first two rows of Fig. 2, one can see (as expected from the theory in Section 2.2) that the estimates for $(\beta_2 + \gamma_2)$ and $(\beta_3 + \gamma_3)$ are very close to these for φ_2 and φ_3, respectively, when the maximum follow-up duration t_R is small, but tend to diverge as t_R increases. On the other hand, when a covariate only enters the sub-model defining the cure probability, the estimation of β_k in the promotion time model and of φ_k in the Cox PH model are equivalent whatever t_R (see the first three plots in the third row of Fig. 2 where the solid and dashed lines are are hardly distinguishable). From the credible intervals for the β_k's and the sign of their plausible values, one can see that clear indications on the effect of these covariates on long-term survival are already available from the data with a rather short maximum follow-up time t_R. Effects on short-term survival require larger t_R values to be established for shared covariates.

5 Discussion

Cure rate models are attractive alternatives to classical survival model when it is known a priori that a unidentified proportion of subjects will never experience the event of interest whatever the duration of the follow-up. When the study duration is long enough to declare that survivors with the maximum follow-up time are 'cured', estimating a promotion time cure model with covariates simultaneously influencing the cure probability (or long-term survival) and the population hazard dynamics (or short-term survival) works perfectly: estimators of the regression parameters are unbiased and credible intervals have the expected frequentist coverage.

On the other hand, for shorter follow-up, although the inclusion of informative priors for the regression parameters could be helpful, it is probably not sufficient to obtain unbiased estimations when the concerned covariate truly affects the hazard dynamics. Credible intervals with the expected frequentist coverage can be obtained for the regression parameters provided that the corresponding covariate (truly) affects the probability to be cured.

Thus, in practice, when the follow-up is insufficiently long and unless it is known a priori that a covariate can only affect either the probability to be cured or the population hazard dynamics, one should be very cautious and probably opt for a simpler model in a first step. A standard proportional hazards model is a possibility as it can be seen as a first order approximation of the extended promotion time cure model. Then, for a short maximum follow-up duration, a regression coefficient φ_k in the PH model can approximately be seen as the sum of the corresponding (non disentangled) regression coefficients β_k and γ_k in the associated promotion time cure model with shared covariates. Thus, compared to a classical procedure, variable selection and parameter estimation would not differ, but the interpretation of a regression coefficient as a conditional log hazard ratio (for a subject truly at risk) would not hold anymore. Indications on whether the selected covariates affect the long- or short-term survival can be obtained by fitting, in a second step, the promotion time model described in Section 2.1. If the credible interval for β_k (resp. γ_k) does not include zero, then one has good hints on the conditional qualitative effect of the corresponding covariate on long-term (resp. short-term) survival. But when a covariate simultaneously affects long- and short-term survival, one should not rely on point estimates for its regression coefficients. An illustration of the whole modelling strategy can be found in Section 4, in particular when the follow-up is insufficiently long.

It would be interesting to explore the same research questions for extensions of the promotion time model. For example, Zeng *et al.* (2006) motivated the inclusion of a subject-specific frailty term in the population survival function as the independence assumption for the W_i's in the biological derivation of the promotion time model might be unrealistic. Alternatives to the Poisson distribution for the number of latent factors N can also be considered (Cooner *et al.*, 2007). In the special case of a Bernoulli distribution for N, one ends up with the mixture cure model (Berkson and Gage, 1952) where the population survival function is modelled as

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

where $\pi(\mathbf{x})$ denotes the conditional cure probability and $S_u(t|\mathbf{z})$ the survival function for the susceptible subjects. We plan to report on these issues in an additional paper.

Acknowledgements

The authors would like to thank the two anonymous Referees for their constructive and insightful comments. The authors also acknowledge financial support from IAP research network P7/06 of the Belgian Government (Belgian Science Policy), and from the contract 'Projet d'Actions de Recherche Concertées' (ARC) 11/16-039 of the 'Communauté française de Belgique', granted by the 'Académie universitaire Louvain'.

Conflict of Interest
The authors have declared no conflict of interest.

References

Berkson, J. and Gage, R. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, **47**(259), 501–515.

Bremhorst, V. and Lambert, P. (2016). Flexible estimation in cure survival models using Bayesian P-splines. *Computational Statistics and Data Analysis*, **93**, 270–284.

Bremhorst, V., Kreyenfeld, M., and Lambert, P. (2016). Fertility progression in Germany: An analysis using flexible nonparametric cure survival models. *Demographic Research*, **35**, 505–534.

Bremhorst, V., Kreyenfeld, M., and Lambert, P. (2018). Nonparametric double additive cure survival models : an application to the estimation of the nonlinear effect of age at first parenthood on fertility progression. *Statistical Modelling*. (in press).

Chen, M.-H., Ibrahim, J., and Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, **94**(447), 909–919.

Cooner, F., Banerjee, S., Carlin, B., and Sinha, D. (2007). Flexible cure rate modeling under latent activation schemes. *Journal of the American Statistical Association*, **102**(478), 560–572.

Cox, D. (1972). Regression models and life tables. *Journal of the Royal Statistic Society: Series B (Methodological)*, **34**, 187–202.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**, 89–102.

Eilers, P. H. C. and Marx, B. D. (2010). Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics*, **2**, 637–653.

Gelman, A., Roberts, G. O., and Gilks, W. R. (1996). Efficient Metropolis jumping rules. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 5*, pages 599–607. Oxford: Oxford University Press.

Gressani, O. and Lambert, P. (2018). Fast Bayesian inference using Laplace approximations in a flexible promotion time cure model based on P-splines. *Computational Statistics & Data Analysis*, **124**, 151–167.

Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, **7**(2), 223–242.

Ibrahim, J., Chen, M., and Sinha, D. (2001). Bayesian semiparametric models for survival data with a cure fraction. *Biometrics*, **57**(2), 383–388.

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**(5), 2542–2558.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**(282), 457–481.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, **13**(1), 183–212.

Liu, H. and Shen, Y. (2009). Semiparametric parametric regression cure model for interval-censored data. *Journal of the American Statistical Association*, **104**, 1168–1178.

Moertel, C. G., Fleming, T. R., MacDonald, J. S., Haller, D. G., Laurie, J. A., Goodman, P. J., Ungerleider, J. S., Emerson, W. A., Tormey, D. C., Glick, J. H., Veeder, M. H., and Maillard, J. A. (1990). Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *New England J of Medicine*, **332**, 352–358.

Moertel, C. G., Fleming, T. R., MacDonald, J. S., Haller, D. G., Laurie, J. A., Tangen, C. M., Ungerleider, J. S., Emerson, W. A., Tormey, D. C., Glick, J. H., Veeder, M. H., and Maillard, J. A. (1991). Fluorouracil plus Levamisole as an effective adjuvant therapy after resection of stage II colon carcinoma: a final report. *Annals of Internal Medicine*, **122**: 321–326.

Polson, N. G. and Scott, J. G. (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis*, **7**(4), 887–902.

Roberts, G. O. and Rosenthal, J. S. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, **16**(4), 351–367.

Rosenberg, P. S. (1995). Hazard function estimation using B-splines. *Biometrics*, **51**, 874–887.

Tsodikov, A. (1998). A proportional hazard model taking account of long-term survivors.

Biometrics, **54**, 1508–1516.

Tsodikov, A. (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 Medicine*, **21**, 895–920.

Tsodikov, A., Ibrahim, J., and Yakovlev, A. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models. *Journal of the American Statistical Association*, **98**(464), 1063–1078.

Yakovlev, A. and Tsodikov, A. (1996). *Stochastic Models for Tumor of Latency and Their Biostatistical Applications*. World Scientific Publishing Singapore.

Yin, G. and Ibrahim J.G. (2005). Cure rate models : a unified approach. *The Canadian journal of statistics*, **33**, 559–570.

Zeng, D., Yin, G., and Ibrahim, J. (2006). Semiparametric transformation models for survival data with a cure fraction. *Journal of the American Statistical Association*, **101**(474), 670–684.

Table 1: Simulation results for a **correctly specified** model fitted on S=500 datasets of size n with a maximum follow-up time t_R yielding an average percentage RC of right censored units involving an average of $0.20n$ cured units. Point estimates refer to the posterior mean, 'Cov' to the estimated coverage of 95% HPD credible intervals and $\beta_0^* = \beta_0 + \log F_0(t_R)$.

n = 500

		Setting 1: t_R=25, $\beta_0^* = 0.750$, RC: 24%					Setting 2: t_R=15, $\beta_0^* = 0.685$, RC: 31%					Setting 3: t_R=10, $\beta_0^* = 0.403$, RC: 41%				
	True	Mean	Bias	ESE	RMSE	Cov	Mean	Bias	ESE	RMSE	Cov	Mean	Bias	ESE	RMSE	Cov
ψ_0	0.75	0.759	0.009	0.077	0.078	0.944	0.718	-0.032	0.124	0.128	0.880	0.591	-0.159	0.154	0.222	0.726
ψ_1	-0.50	-0.500	0.000	0.056	0.056	0.948	-0.501	-0.001	0.058	0.058	0.938	-0.501	-0.001	0.062	0.062	0.954
ψ_2	0.60	0.608	0.008	0.041	0.042	0.936	0.606	0.006	0.042	0.043	0.942	0.606	0.006	0.045	0.046	0.952
ψ_3	-0.40	-0.404	-0.004	0.068	0.068	0.930	-0.399	0.001	0.070	0.070	0.944	-0.414	-0.014	0.076	0.077	0.946
$d\psi_2$	0.20	0.186	-0.014	0.110	0.111	0.946	0.225	0.025	0.143	0.145	0.946	0.293	0.093	0.197	0.218	0.950
β_0^*	...	0.756	0.006	0.076	0.076	0.942	0.673	-0.012	0.085	0.086	0.934	0.405	0.002	0.087	0.087	0.952
β_0	0.75	0.759	0.009	0.077	0.078	0.944	0.718	-0.032	0.124	0.128	**0.880**	0.591	**-0.159**	0.154	0.222	**0.726**
β_1	-0.50	-0.500	0.000	0.056	0.056	0.948	-0.501	-0.001	0.058	0.058	0.938	-0.501	-0.001	0.062	0.062	0.954
β_2	0.80	0.793	-0.007	**0.103**	0.103	0.948	0.831	0.031	**0.134**	0.137	0.948	0.898	**0.098**	**0.184**	0.209	0.934
γ_2	0.40	0.422	0.022	**0.130**	0.132	0.932	0.381	-0.019	**0.163**	0.164	0.956	0.313	**-0.087**	**0.219**	0.236	0.954
γ_3	-0.40	-0.404	-0.004	0.068	0.068	0.930	-0.399	0.001	0.070	0.070	0.944	-0.414	-0.014	0.076	0.077	0.946

n = 300

ψ_0	0.75	0.762	0.012	0.092	0.093	0.954	0.708	-0.042	0.118	0.125	0.928	0.616	-0.134	0.153	0.203	0.866
ψ_1	-0.50	-0.506	-0.006	0.074	0.074	0.948	-0.508	-0.008	0.078	0.078	0.948	-0.509	-0.009	0.084	0.084	0.940
ψ_2	0.60	0.607	0.007	0.055	0.056	0.932	0.605	0.005	0.058	0.058	0.936	0.604	0.004	0.061	0.061	0.938
ψ_3	-0.40	-0.404	-0.004	0.092	0.092	0.932	-0.400	0.000	0.096	0.096	0.928	-0.411	-0.011	0.105	0.105	0.932
$d\psi_2$	0.20	0.193	-0.007	0.141	0.142	0.950	0.242	0.042	0.185	0.189	0.946	0.316	0.116	0.259	0.284	0.924
β_0^*	...	0.760	0.010	0.090	0.091	0.950	0.684	-0.001	0.107	0.107	0.946	0.429	0.026	0.100	0.103	0.966
β_0	0.75	0.762	0.012	0.092	0.093	0.954	0.708	-0.042	0.118	0.125	0.928	0.616	**-0.134**	0.153	0.203	**0.866**
β_1	-0.50	-0.506	-0.006	0.074	0.074	0.948	-0.508	-0.008	0.078	0.078	0.948	-0.509	-0.009	0.084	0.084	0.942
β_2	0.80	0.800	0.000	**0.131**	0.131	0.954	0.846	0.046	**0.171**	0.177	0.946	0.920	**0.120**	**0.242**	0.270	0.930
γ_2	0.40	0.414	0.014	**0.170**	0.170	0.948	0.363	-0.037	**0.214**	0.217	0.928	0.287	**-0.113**	**0.288**	0.310	0.928
γ_3	-0.40	-0.404	-0.004	0.092	0.092	0.932	-0.400	0.000	0.096	0.096	0.928	-0.411	-0.011	0.105	0.105	0.932

Table 2: Simulation results for a **misspecified model** fitted on S=500 datasets of size $\boldsymbol{n = 500}$ with a maximum follow-up time t_R yielding an average percentage RC of right censored units involving an average of $0.20n$ cured units. Point estimates refer to the posterior mean, 'Cov' to the estimated coverage of 95% HPD credible intervals and $\beta_0^* = \beta_0 + \log F_0(t_R)$.

		Setting 1: t_R=25, $\beta_0^* = 0.750$, RC: 24%					Setting 2: t_R=15, $\beta_0^* = 0.685$, RC: 31%					Setting 3: t_R=10, $\beta_0^* = 0.403$, RC: 41%				
	True	Mean	Bias	ESE	RMSE	Cov	Mean	Bias	ESE	RMSE	Cov	Mean	Bias	ESE	RMSE	Cov
ψ_0	0.75	0.771	0.021	0.088	0.090	0.956	0.740	-0.010	0.147	0.147	0.908	0.614	-0.136	0.160	0.210	0.826
ψ_1	-0.25	-0.254	-0.004	0.035	0.035	0.950	-0.254	-0.004	0.035	0.036	0.956	-0.254	-0.004	0.039	0.039	0.954
ψ_2	0.60	0.609	0.009	0.042	0.042	0.940	0.606	0.006	0.042	0.042	0.954	0.603	0.003	0.046	0.046	0.948
ψ_3	-0.20	-0.203	-0.003	0.034	0.034	0.932	-0.199	0.001	0.036	0.036	0.950	-0.199	0.001	0.040	0.040	0.952
$d\psi_1$	-0.25	-0.244	0.006	0.095	0.095	0.934	-0.248	0.002	0.121	0.121	0.966	-0.253	-0.003	0.164	0.164	0.966
$d\psi_2$	0.20	0.187	-0.013	0.112	0.113	0.940	0.237	0.037	0.154	0.159	0.944	0.328	0.128	0.206	0.242	0.946
$d\psi_3$	0.20	0.198	-0.002	0.092	0.092	0.934	0.149	-0.051	0.127	0.137	0.892	0.087	-0.113	0.155	0.192	0.904
β_0^*	...	0.767	0.017	0.084	0.086	0.958	0.683	-0.002	0.092	0.092	0.954	0.407	0.004	0.091	0.091	0.960
β_0	0.75	0.771	0.021	0.088	0.090	0.956	0.740	-0.010	0.147	0.147	**0.908**	0.614	**-0.136**	0.160	0.210	**0.826**
β_1	-0.50	-0.498	0.002	0.091	0.091	0.938	-0.502	-0.002	0.114	0.114	0.964	-0.507	-0.007	0.152	0.152	0.966
γ_1	0.00	-0.010	-0.010	0.110	0.111	0.948	-0.006	-0.006	0.137	0.137	0.962	-0.000	-0.000	0.183	0.183	0.962
β_2	0.80	0.796	-0.004	0.105	0.106	0.956	0.843	**0.043**	0.147	0.153	0.940	0.931	**0.131**	0.192	0.233	0.944
γ_2	0.40	0.422	0.022	0.133	0.135	0.950	0.369	**-0.031**	0.172	0.175	0.950	0.276	**-0.124**	0.228	0.259	0.956
β_3	0.00	-0.006	-0.006	0.084	0.084	0.944	-0.050	**-0.050**	0.118	0.128	**0.906**	-0.112	**-0.112**	0.140	0.179	**0.904**
γ_3	-0.40	-0.401	-0.001	0.109	0.109	0.940	-0.347	**0.053**	0.144	0.154	**0.910**	-0.285	**0.115**	0.178	0.212	**0.914**

Table 3: Simulation results for a **misspecified model** fitted on S=500 datasets of size $\boldsymbol{n = 300}$ with a maximum follow-up time t_R yielding an average percentage RC of right censored units involving an average of $0.20n$ cured units. Point estimates refer to the posterior mean, 'Cov' to the estimated coverage of 95% HPD credible intervals and $\beta_0^* = \beta_0 + \log F_0(t_R)$.

		Setting 1: t_R=25, $\beta_0^* = 0.750$, RC: 24%					Setting 2: t_R=15, $\beta_0^* = 0.685$, RC: 31%					Setting 3: t_R=10, $\beta_0^* = 0.403$, RC: 41%				
	True	Mean	Bias	ESE	RMSE	Cov	Mean	Bias	ESE	RMSE	Cov	Mean	Bias	ESE	RMSE	Cov
ψ_0	0.75	0.776	0.026	0.102	0.105	0.956	0.732	-0.018	0.131	0.132	0.942	0.658	-0.092	0.158	0.182	0.948
ψ_1	-0.25	-0.259	-0.009	0.048	0.048	0.930	-0.259	-0.009	0.051	0.051	0.930	-0.258	-0.008	0.054	0.054	0.926
ψ_2	0.60	0.609	0.009	0.055	0.056	0.934	0.605	0.005	0.057	0.057	0.948	0.601	0.001	0.061	0.061	0.948
ψ_3	-0.20	-0.204	-0.004	0.046	0.046	0.922	-0.199	0.001	0.049	0.049	0.922	-0.196	0.004	0.054	0.054	0.930
$d\psi_1$	-0.25	-0.245	0.005	0.128	0.128	0.930	-0.253	-0.003	0.164	0.164	0.938	-0.268	-0.018	0.227	0.228	0.946
$d\psi_2$	0.20	0.195	-0.005	0.153	0.154	0.950	0.260	0.060	0.194	0.203	0.936	0.367	0.167	0.269	0.317	0.922
$d\psi_3$	0.20	0.200	0.000	0.117	0.117	0.938	0.145	-0.055	0.155	0.164	0.908	0.070	-0.130	0.211	0.248	0.908
β_0^*	0.75	0.774	0.024	0.100	0.103	0.954	0.703	0.018	0.115	0.116	0.962	0.452	0.049	0.103	0.114	0.950
β_0	0.75	0.776	0.026	0.102	0.105	0.956	0.732	-0.018	0.131	0.132	0.942	0.658	-0.092	0.158	0.182	0.948
β_1	-0.50	-0.504	-0.004	0.120	0.120	0.942	-0.512	-0.012	0.150	0.151	0.944	-0.526	-0.026	0.208	0.210	0.946
γ_1	0.00	-0.014	-0.014	0.152	0.152	0.936	-0.005	-0.005	0.191	0.191	0.924	0.009	0.009	0.256	0.256	0.946
β_2	0.80	0.804	0.004	0.143	0.143	0.942	0.865	**0.065**	0.182	0.193	0.932	0.968	**0.168**	0.254	0.304	0.922
γ_2	0.40	0.415	0.015	0.181	0.182	0.938	0.345	**-0.055**	0.221	0.228	0.934	0.234	**-0.166**	0.296	0.339	0.926
β_3	0.00	-0.003	-0.003	0.110	0.110	0.942	-0.055	**-0.055**	0.144	0.154	**0.912**	-0.126	**-0.126**	0.192	0.229	**0.902**
γ_3	-0.40	-0.404	-0.004	0.140	0.140	0.926	-0.344	**0.056**	0.179	0.187	**0.898**	-0.265	**0.135**	0.240	0.276	**0.908**

Table 4: Colon cancer data: regression parameter estimation in the flexible promotion time cure model with all the pre-selected covariates included. n_{eff} indicates the effective size of the MCMC sample out of a chain of length $10,000$; P(Coef) $= \min\{\Pr(\text{Coef} > 0|\mathcal{D}), \Pr(\text{Coef} < 0|\mathcal{D})\}$.

					Quantiles					
Variable	Category	Coef	Mean	sd	2.5%	5%	95%	97.5%	P(Coef)	n_{eff}
Intercept		β_0	-0.429	0.125	-0.673	-0.632	-0.222	-0.184	0.000	939
Treatment	Control	–	0.000							
	Lev(amisole)	β_1	-0.064	0.118	-0.299	-0.260	0.128	0.165	0.299	1480
		γ_1	0.075	0.156	-0.234	-0.181	0.328	0.378	0.318	785
	Lev + 5-FU	β_2	-0.521	0.131	-0.780	-0.738	-0.303	-0.264	**0.000**	1668
		γ_2	-0.034	0.173	-0.379	-0.321	0.243	0.293	0.423	1318
#Positive lymph nodes	≤ 1	–	0.000							
	2	β_3	0.114	0.163	-0.209	-0.147	0.381	0.432	0.242	1184
		γ_3	0.188	0.212	-0.230	-0.156	0.539	0.598	0.189	454
	$[3, 5]$	β_4	0.484	0.142	0.205	0.248	0.721	0.762	**0.000**	993
		γ_4	0.396	0.181	0.034	0.094	0.699	0.755	**0.016**	380
	≥ 6	β_5	0.875	0.142	0.597	0.643	1.111	1.155	**0.000**	1289
		γ_5	0.437	0.192	0.065	0.131	0.756	0.824	**0.011**	575
Cell differentiation	Well/Moderate	–	0.000							
	Poor	β_6	0.099	0.131	-0.165	-0.117	0.316	0.359	0.221	1597
		γ_6	0.703	0.158	0.388	0.443	0.956	1.005	**0.000**	1331
Extent of cancer	Submucosa or Muscle	β_7	-0.563	0.185	-0.929	-0.870	-0.266	-0.201	**0.002**	1617
		γ_7	-0.030	0.242	-0.572	-0.455	0.334	0.398	0.475	1379
	Serosa	–	0.000							
	Contiguous structures	β_8	0.275	0.232	-0.184	-0.117	0.651	0.719	0.121	1611
		γ_8	0.525	0.275	-0.044	0.056	0.960	1.033	0.034	1619

– DIC=2319.63 ; p_D=21.49 ; BIC=2407.74 –

Table 5: Colon cancer data: regression parameter estimation in the flexible promotion time cure model after the final covariate selection. n_{eff} indicates the effective size of the MCMC sample out of a chain of length $10,000$; P(Coef) $= \min\{\Pr(\text{Coef} > 0|\mathcal{D}), \Pr(\text{Coef} < 0|\mathcal{D})\}$.

					Quantiles					
Variable	Category	Coef	Mean	sd	2.5%	5%	95%	97.5%	P(Coef)	n_{eff}
Long-term survival: $\Pr(\text{cured}\|\mathbf{x}) = \exp\left(-\exp(\beta_0 + \mathbf{x}'\boldsymbol{\beta})\right)$										
Intercept		β_0	-0.403	0.085	-0.574	-0.548	-0.266	-0.245	0.000	1192
Treatment	Control / Lev	–	0.000							
	Lev + 5-FU	β_1	-0.506	0.108	-0.719	-0.687	-0.335	-0.301	**0.000**	1937
#Positive lymph nodes	≤ 2	–	0.000							
	$[3, 5]$	β_2	0.446	0.117	0.214	0.252	0.635	0.669	**0.000**	1514
	≥ 6	β_3	0.846	0.124	0.604	0.641	1.052	1.089	**0.000**	1541
Extent of cancer	Submucosa/Muscle	β_5	-0.572	0.171	-0.912	-0.858	-0.289	-0.241	**0.000**	1699
	Serosa	–	0.000							
	Contig. Structures	β_6	0.462	0.205	0.038	0.112	0.785	0.837	**0.017**	1601
Short-term survival (NCHD): $F(y\|\mathbf{z}) = 1 - S_0(y)^{\exp(\mathbf{z}'\boldsymbol{\gamma})}$										
#Positive lymph nodes	≤ 2	–	0.000							
	$[3, 5]$	γ_2	0.304	0.149	0.000	0.051	0.544	0.591	**0.025**	681
	≥ 6	γ_3	0.347	0.156	0.040	0.091	0.605	0.649	**0.012**	808
Cell differentiation	Well/Moderate	–	0.000							
	Poor	γ_4	0.733	0.147	0.438	0.491	0.971	1.020	**0.000**	1260

– DIC=2308.98 ; p_D=12.97 ; BIC=2362.17 –

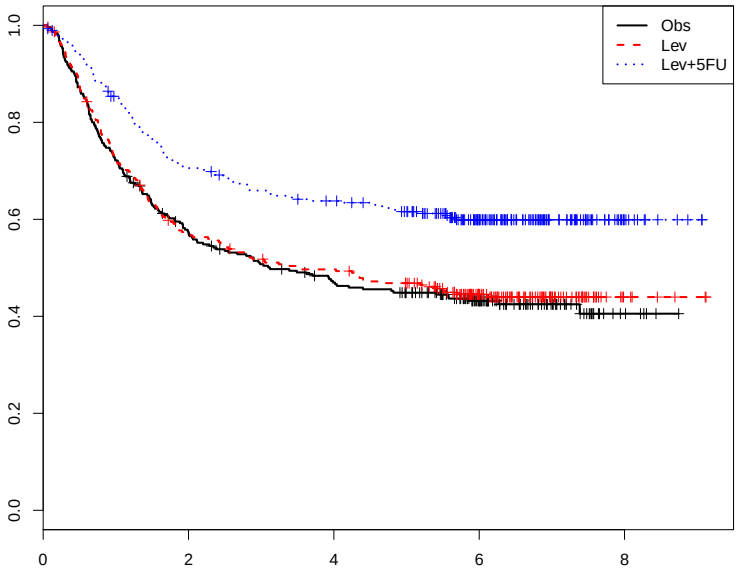

Figure 1: Colon cancer dataset: Kaplan-Meier estimates of the survival functions for the time (in years) to recurrence for each of the treatment groups.

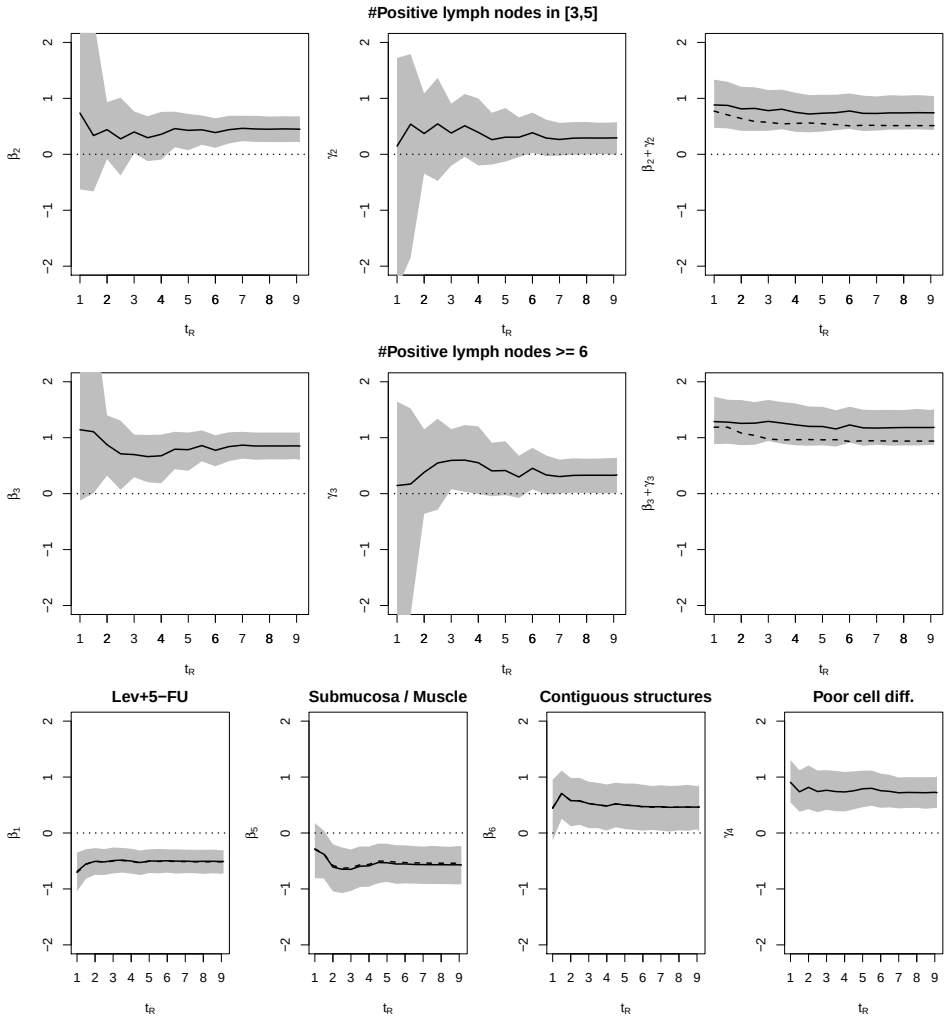

Figure 2: Colon cancer dataset: regression parameter estimation for growing maximum follow-up time t_R (in years) in the final model: posterior mean (solid line) and 95% quantile-based credible interval (grey region) for β_k, γ_k and $(\beta_k + \gamma_k)$ in the flexible promotion time model ; dashed curves, when present, provides the corresponding parameter estimates for φ_k in the Cox PH model.