

# Laplacian-P-splines for Bayesian inference in the mixture cure model

Oswaldo Gressani<sup>1</sup>  | Christel Faes<sup>1</sup> | Niel Hens<sup>1,2</sup>

<sup>1</sup>Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Data Science Institute, Hasselt University, Hasselt, Belgium

<sup>2</sup>Centre for Health Economics Research and Modelling Infectious Diseases, Vaxinfectio, University of Antwerp, Antwerp, Belgium

## Correspondence

Oswaldo Gressani, Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Data Science Institute, Hasselt University, Hasselt, Belgium.  
Email: oswaldo.gressani@uhasselt.be

## Funding information

European Union's Research and Innovation Action (EpiPose), Grant/Award Number: 101003688

## Abstract

The mixture cure model for analyzing survival data is characterized by the assumption that the population under study is divided into a group of subjects who will experience the event of interest over some finite time horizon and another group of cured subjects who will never experience the event irrespective of the duration of follow-up. When using the Bayesian paradigm for inference in survival models with a cure fraction, it is common practice to rely on Markov chain Monte Carlo (MCMC) methods to sample from posterior distributions. Although computationally feasible, the iterative nature of MCMC often implies long sampling times to explore the target space with chains that may suffer from slow convergence and poor mixing. Furthermore, extra efforts have to be invested in diagnostic checks to monitor the reliability of the generated posterior samples. A sampling-free strategy for fast and flexible Bayesian inference in the mixture cure model is suggested in this article by combining Laplace approximations and penalized B-splines. A logistic regression model is assumed for the cure proportion and a Cox proportional hazards model with a P-spline approximated baseline hazard is used to specify the conditional survival function of susceptible subjects. Laplace approximations to the posterior conditional latent vector are based on analytical formulas for the gradient and Hessian of the log-likelihood, resulting in a substantial speed-up in approximating posterior distributions. The spline specification yields smooth estimates of survival curves and functions of latent variables together with their associated credible interval are estimated in seconds. A fully stochastic algorithm based on a Metropolis-Langevin-within-Gibbs sampler is also suggested as an alternative to the proposed Laplacian-P-splines mixture cure (LPSMC) methodology. The statistical performance and computational efficiency of LPSMC is assessed in a simulation study. Results show that LPSMC is an appealing alternative to MCMC for approximate Bayesian inference in standard mixture cure models. Finally, the novel LPSMC approach is illustrated on three applications involving real survival data.

## KEYWORDS

Approximate Bayesian inference, Laplace approximation, Metropolis-adjusted Langevin algorithm, P-splines, Survival analysis

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Statistics in Medicine* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

Survival analysis is a challenging, yet very attractive area of statistical science that is devoted to the study of time-to-event data. Standard models for survival data typically leave no room for the existence of a cure fraction such that it is implicitly assumed that all subjects of the population under study will experience the event of interest as time unfolds for a sufficiently long period. Technological breakthroughs in medicine during the last decades, especially in cancer research, have led to the development of promising new treatments and therapies so that many diseases previously considered fatal can now be cured. This phenomenon has triggered the necessity to develop models that allow for long-term survivors and gave birth to cure models. A recent complete textbook treatment on cure models is proposed by Peng and Yu<sup>1</sup> and an enriching literature review on cure regression models has been written by Amico and Van Keilegom.<sup>2</sup> Among the large family of cure models that have emerged, the mixture cure model driven by the seminal work of (Boag; Berkson and Gage; Haybittle)<sup>3-5</sup> and later refined by Farewell<sup>6,7</sup> is probably the most prominent as its mathematical formulation allows for a clear and interpretable separation of the population in two categories, namely cured subjects and susceptible subjects who are at risk of experiencing the event of interest.

Let  $T \in [0, +\infty)$  be a continuous random variable representing the survival time. Existence of a cure proportion in the population under study is made possible by allowing the event  $\{T = +\infty\}$  to arise with positive probability. To include covariate information, denote by  $\mathbf{X}$  and  $\mathbf{Z}$  random covariate vectors (with continuous and/or discrete entries) that belong to covariate spaces  $\mathcal{X}$  and  $\mathcal{Z}$ , respectively. In a mixture cure model, the population survival function expresses the separation between the cured and uncured subpopulations as follows:

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

with covariate vectors  $\mathbf{x} = (x_1, \dots, x_p)^\top \in \mathcal{X}$  and  $\mathbf{z} = (z_1, \dots, z_q)^\top \in \mathcal{Z}$  that can share (partially) the same components or can be entirely different. The term  $p(\mathbf{x})$  is frequently called the “incidence” of the model and corresponds to the conditional probability of being uncured, that is,  $p(\mathbf{x}) = P(B = 1|\mathbf{X} = \mathbf{x})$  with binary variable  $B = \mathbb{I}(T < +\infty)$  referring to the (unknown) susceptible status and  $\mathbb{I}(\cdot)$  the indicator function, that is,  $\mathbb{I}(E) = 1$  if condition  $E$  is true. The term  $S_u(t|\mathbf{z})$  is known as the “latency” and represents the conditional survival function of the uncured subjects  $S_u(t|\mathbf{z}) = P(T > t|B = 1, \mathbf{Z} = \mathbf{z})$ . The logistic link is commonly employed to establish a functional relationship between the probability to be uncured and the vector  $\mathbf{x}$ ,<sup>8-10</sup> so that  $p(\mathbf{x}) = \exp(\beta_0 + \mathbf{x}^\top \boldsymbol{\beta}) / (1 + \exp(\beta_0 + \mathbf{x}^\top \boldsymbol{\beta}))$ , with regression coefficients  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$  and  $\beta_0$  an intercept term. The latency part is often specified in a semiparametric fashion by using the Cox<sup>11</sup> proportional hazards (PH) model (see eg, 12,13) and implies the following form for the survival function of the susceptibles  $S_u(t|\mathbf{z}) = S_0(t)^{\exp(\boldsymbol{\gamma}^\top \mathbf{z})}$ , where  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)^\top$  are the regression coefficients pertaining to the latency part and  $S_0(\cdot)$  is the baseline survival function.

The philosophy underlying Bayesian approaches considers that the model parameters are random and that their underlying uncertainty is characterized by probability distributions. After obtaining data, Bayes’ theorem acts as a mechanistic process describing how to update our knowledge and is essentially the key ingredient permitting the transition from prior to posterior beliefs. Unfortunately, the complexity of mixture cure models is such that the posterior distribution of latent variables of interest are not obtainable in closed form. An elegant stochastic method that is widely used in practice is Markov chain Monte Carlo (MCMC) as it allows to draw random samples from desired target posterior distributions and hence compute informative summary statistics. According to Greenhouse et al<sup>14</sup>, the first steps of Bayesian methods applied to mixture models with a cure fraction date back as far as Chen et al<sup>15</sup> to analyze cancer data. Later, Stangl<sup>16</sup> and Stangl and Greenhouse<sup>17</sup> used a Bayesian mixture survival model to analyze clinical trial data related to mental health. The end of 1990s saw the emergence of Bayesian approaches in the promotion time cure model,<sup>18</sup> another family of cure models motivated by biological mechanisms that does not impose a mixture structure on the survival. Some references for Bayesian analysis in the latter model class are Yin and Ibrahim<sup>19</sup>, who proposed a Box-Cox based transformation on the population survival function to reach a unified family of cure rate models embedding the promotion time cure model as a special case; Bremhorst and Lambert<sup>20</sup> used Bayesian P-splines with MCMC for flexible estimation in the promotion time cure model and Gressani and Lambert<sup>21</sup> suggested a faster alternative based on Laplace approximations. More recent uses of Bayesian methods in mixture cure models are Yu and Tiwari<sup>22</sup> in the context of grouped population-based cancer survival data or Martinez et al<sup>23</sup> who consider a parametric specification for the baseline survival of uncured subjects governed by a generalized modified Weibull distribution. In the literature of cure survival models, only scarce attempts have been initiated to propose an alternative to the deep-rooted MCMC instruments. This is especially true for

mixture cure models, where to our knowledge Lázaro et al<sup>24</sup> is the only reference proposing an approximate Bayesian method based on a combination of Integrated Nested Laplace Approximations (INLA)<sup>25</sup> and modal Gibbs sampling.<sup>26</sup> In this article, we propose a new approach for fast approximate Bayesian inference in the mixture cure model based on the idea of Laplacian-P-splines.<sup>21</sup> The proposed Laplacian-P-splines mixture cure (LPSMC) model has various practical and numerical advantages that are worth mentioning. First, as opposed to Lázaro et al<sup>24</sup>, our approach is completely sampling-free in the sense that estimation can be fully reached without the need of drawing samples from posterior distributions. This of course implies a huge gain from the computational side, without even mentioning the additional speed-up effect implied by the analytically available gradient and Hessian of the log-likelihood function in our Laplace approximation scheme. Second, the LPSMC approach delivers approximations to the joint posterior latent vector, while the INLA scheme concentrates on obtaining approximated versions of the marginal posterior of latent variables. A direct positive consequence is that with LPSMC, the “delta” method can be used to compute (approximate) credible intervals for functions of latent variables, such as the cure proportion or the survival function of the uncured, in virtually no time. A third beneficial argument is that the use of P-splines is particularly well adapted in a Bayesian framework and provides smooth estimates of the survival function. Our methodology and its associated algorithms natively account for the underlying mixture specification, making LPSMC a tailored toolkit to fit mixture cure models, contrary to INLA that is originally not adapted to cope with mixture structures.<sup>24</sup> A fully stochastic alternative to LPSMC is also proposed based on a Metropolis-Langevin-within-Gibbs (MLWG) algorithm to sample from the joint posterior of the model parameters. The conditional posterior distribution of the latent vector is explored via a Metropolis algorithm augmented by Langevin diffusions that account for the gradient of the (log) target distribution to generate new proposals. Using the variance-covariance matrix of the Laplace approximation as a basis for the correlation structure of the proposal distribution proves to be beneficial as it improves chain mixing and convergence. The **mixcurelps** package<sup>27</sup> gathers all the routines to implement the LPSMC methodology (and MLWG sampler) and can be used to reproduce the main results of this article as shown in the supplementary materials. For higher efficiency, algorithms related to Laplace approximations and B-splines evaluations are coded in C++ and integrated into the R language via the **Rcpp** package.<sup>28</sup>

The article is organized as follows. In Section 2, the spline specification of the log-baseline hazard is presented and the Bayesian model is formulated along with the prior assumptions. Laplace approximations to the conditional latent vector are derived and an approximate version of the posterior penalty parameter is proposed. The end of Section 2 is dedicated to the construction of approximate credible intervals for (functions of) latent variables. In Section 3, we propose a Markov chain Monte Carlo algorithm with a Metropolis-adjusted Langevin setting to explore the conditional posterior distribution of the latent vector. The steps of the so-called Metropolis-Langevin-within-Gibbs (MLWG) sampler are presented in detail and summarized in a pseudocode. Section 4 aims at assessing the proposed LPSMC and MLWG methodologies in a numerical study with simulated data under different cure and censoring scenarios. Section 5 is dedicated to three real data applications and Section 6 concludes the article.

## 2 | THE LAPLACIAN-P-SPLINES MIXTURE CURE MODEL

We consider that the survival time  $T$  is accompanied by the frequently encountered feature of random right censoring. Rather than observing  $T$  directly, one observes the pair  $(T_{\text{obs}}, \tau)$ , where  $T_{\text{obs}} = \min(T, C)$  is the follow-up time and  $\tau = I(T \leq C)$  is the event indicator ( $\tau = 1$  if the event occurred and  $\tau = 0$  otherwise) and  $C$  is a non-negative random censoring time that is assumed conditionally independent of  $T$  given the covariates, that is,  $C \perp T | \mathbf{X}, \mathbf{Z}$ . At the sample level,  $\mathcal{D}_i = (t_i, \tau_i, \mathbf{x}_i, \mathbf{z}_i)$  denotes the observables for the  $i$  unit, with  $t_i$  the realization of  $T_{\text{obs}}$  and  $\tau_i$  its associated event indicator. Vectors  $\mathbf{x}_i$  and  $\mathbf{z}_i$  represent the observed covariate values of subject  $i$  and the entire information set available from data with sample size  $n$  is denoted by  $\mathcal{D} = \bigcup_{i=1}^n \mathcal{D}_i$ .

### 2.1 | Flexible modeling of the baseline risk function with B-splines

A flexible spline specification of the (log) baseline hazard function  $h_0(\cdot)$  is proposed<sup>29,30</sup> using a linear combination of cubic B-splines, that is,  $\log h_0(t) = \theta^\top b(t)$ , where  $\theta = (\theta_1, \dots, \theta_K)^\top$  is a  $K$ -dimensional vector of B-spline amplitudes and  $b(\cdot) = (b_1(\cdot), \dots, b_K(\cdot))^\top$  is a cubic B-spline basis constructed from a grid of equally spaced knots in the closed interval  $\mathcal{I} = [0, t_u]$ , with  $t_u$  the largest observed follow-up time. Partitioning  $\mathcal{I}$  into  $J$  (say 300) sections of equal length  $\Delta$  with midpoint  $s_j$ , the Riemann midpoint rule is used to approximate the analytically unsolvable baseline survival function:

$$S_0(t) = \exp\left(-\int_0^t \exp(\theta^\top b(s)) ds\right) \approx \exp\left(-\sum_{j=1}^{j(t)} \exp(\theta^\top b(s_j)) \Delta\right),$$

where  $j(t) \in \{1, \dots, J\}$  is an integer enumerating the interval that includes time point  $t$ .

### 2.2 | Latent vector and priors

The latent vector of the model is  $\xi = (\theta^\top, \check{\beta}^\top, \gamma^\top)^\top$  and contains the spline vector  $\theta$ , the vector of regression coefficients belonging to the incidence part (including the intercept)  $\check{\beta} = (\beta_0, \beta^\top)^\top$  and the vector of remaining regression coefficients belonging to the latency part  $\gamma$ , with dimension  $\dim(\xi) = K + (p + 1) + q$ . Based on the idea of Eilers and Marx<sup>31</sup>, we fix  $K$  large enough to ensure flexible modeling of the baseline hazard curve and counterbalance the latter flexibility by imposing a discrete penalty on neighboring B-spline coefficients based on finite differences. In a Bayesian translation,<sup>32</sup> the prior distributional assumption on the B-spline vector is taken to be Gaussian  $\theta | \lambda \sim \mathcal{N}_{\dim(\theta)}(0, (\lambda P)^{-1})$ , with a covariance matrix formed by the product of a roughness penalty parameter  $\lambda > 0$  and a penalty matrix  $P = D_r^\top D_r + \varepsilon I_K$  obtained from  $r$ th order difference matrices  $D_r$  of dimension  $(K - r) \times K$ . An  $\varepsilon$ -multiple of the  $K$ -dimensional identity matrix  $I_K$  is added to ensure full rankedness<sup>33</sup> with typical values for the scalar perturbation being  $\varepsilon = 10^{-6}$ <sup>34</sup> or  $\varepsilon = 10^{-5}$ .<sup>35</sup> Furthermore, a Gaussian prior is imposed on the remaining regression coefficients, with zero mean and small (common) precision  $\zeta = 10^{-6}$ , resulting in the following proper (conditional) prior for the latent vector  $\xi | \lambda \sim \mathcal{N}_{\dim(\xi)}(0, \Sigma_\xi(\lambda))$  with covariance matrix:

$$\Sigma_\xi(\lambda) = \begin{pmatrix} (\lambda P)^{-1} & 0 \\ 0 & \zeta^{-1} I_{(p+1)+q} \end{pmatrix}.$$

The prior precision matrix of the latent vector is denoted by  $Q_\xi(\lambda) = \Sigma_\xi^{-1}(\lambda)$ . For full Bayesian treatment, we impose a robust Gamma prior on the roughness penalty parameter as suggested by Jullion and Lambert.<sup>36</sup> In particular, the prior on the roughness penalty parameter conditional on a dispersion hyperparameter  $\delta > 0$  is taken to be  $\lambda | \delta \sim \mathcal{G}(\varphi/2, (\varphi\delta)/2)$ , where  $\mathcal{G}(a, b)$  denotes a Gamma distribution with mean  $a/b$  and variance  $a/b^2$ . A proper Gamma hyperprior is imposed on  $\delta$ , that is,  $\delta \sim \mathcal{G}(a_\delta, b_\delta)$  with small values for  $a_\delta$  and  $b_\delta$  to make it uninformative (here  $a_\delta = b_\delta = 10^{-4}$ ). For a fixed value of  $\varphi$  (here  $\varphi = 3$ ),<sup>36</sup> have shown that the model fit exhibits almost no sensitivity to the choice of  $a_\delta$  and  $b_\delta$  when the latter are fixed to a small value. This is also confirmed in our setting (see supplementary materials), where a sensitivity analysis is implemented to measure the impact on the mixture cure model fit for different configurations of  $a_\delta$  and  $b_\delta$ .

### 2.3 | Laplace approximations

In a mixture cure model, the full likelihood is given by (see eg, Sy and Taylor<sup>37</sup>):

$$\mathcal{L}(\xi; D) = \prod_{i=1}^n (p(\mathbf{x}_i) f_u(t_i | \mathbf{z}_i))^{\tau_i} (1 - p(\mathbf{x}_i) + p(\mathbf{x}_i) S_u(t_i | \mathbf{z}_i))^{(1-\tau_i)},$$

where  $f_u(t_i | \mathbf{z}_i) = -(d/dt)S_u(t_i | \mathbf{z}_i)$ . Using the Cox PH model specification for the survival function of the susceptibles, one recovers:

$$\begin{aligned} f_u(t_i | \mathbf{z}_i) &= -(d/dt)S_u(t_i | \mathbf{z}_i) \\ &= -(d/dt)S_0(t_i)^{\exp(\mathbf{z}_i^\top \gamma)} \\ &= -\exp(\mathbf{z}_i^\top \gamma) S_0(t_i)^{\exp(\mathbf{z}_i^\top \gamma) - 1} (d/dt)S_0(t_i) \\ &= \exp(\mathbf{z}_i^\top \gamma) S_0(t_i)^{\exp(\mathbf{z}_i^\top \gamma) - 1} f_0(t_i) \end{aligned}$$

$$\begin{aligned} &= \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) S_0(t_i)^{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) - 1} S_0(t_i) h_0(t_i) \\ &= \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) h_0(t_i) S_0(t_i)^{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})}. \end{aligned}$$

It follows that the log-likelihood is:

$$\begin{aligned} \ell(\boldsymbol{\xi}; \mathcal{D}) &:= \log \mathcal{L}(\boldsymbol{\xi}; \mathcal{D}) \\ &= \sum_{i=1}^n \tau_i (\log p(\mathbf{x}_i) + \mathbf{z}_i^\top \boldsymbol{\gamma} + \log h_0(t_i) + \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \log S_0(t_i)) \\ &\quad + (1 - \tau_i) \log \left( 1 - p(\mathbf{x}_i) + p(\mathbf{x}_i) S_0(t_i)^{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \right). \end{aligned}$$

Using the B-spline approximations for the baseline quantities, we get:

$$\begin{aligned} \ell(\boldsymbol{\xi}; \mathcal{D}) &\approx \sum_{i=1}^n \tau_i \left( \log p(\mathbf{x}_i) + \mathbf{z}_i^\top \boldsymbol{\gamma} + \boldsymbol{\theta}^\top b(t_i) - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \sum_{j=1}^{j(t_i)} \exp(\boldsymbol{\theta}^\top b(s_j)) \Delta \right) \\ &\quad + (1 - \tau_i) \log \left( 1 - p(\mathbf{x}_i) + p(\mathbf{x}_i) \exp \left( - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \sum_{j=1}^{j(t_i)} \exp(\boldsymbol{\theta}^\top b(s_j)) \Delta \right) \right). \end{aligned}$$

Let us denote by  $g_i(\boldsymbol{\xi})$  the contribution of the  $i$ th unit to the log-likelihood:

$$\begin{aligned} g_i(\boldsymbol{\xi}) &= \tau_i \left( \log p(\mathbf{x}_i) + \mathbf{z}_i^\top \boldsymbol{\gamma} + \boldsymbol{\theta}^\top b(t_i) - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \sum_{j=1}^{j(t_i)} \exp(\boldsymbol{\theta}^\top b(s_j)) \Delta \right) \\ &\quad + (1 - \tau_i) \log \left( 1 - p(\mathbf{x}_i) + p(\mathbf{x}_i) \exp \left( - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \sum_{j=1}^{j(t_i)} \exp(\boldsymbol{\theta}^\top b(s_j)) \Delta \right) \right), \end{aligned}$$

so that the log-likelihood can be compactly written as  $\ell(\boldsymbol{\xi}; \mathcal{D}) \approx \sum_{i=1}^n g_i(\boldsymbol{\xi})$ . Using Bayes' theorem, the conditional posterior of the latent vector is (up to a proportionality constant):

$$\begin{aligned} p(\boldsymbol{\xi} | \lambda, \mathcal{D}) &\propto \mathcal{L}(\boldsymbol{\xi}; \mathcal{D}) p(\boldsymbol{\xi} | \lambda) \\ &\propto \exp(\ell(\boldsymbol{\xi}; \mathcal{D}) - 0.5 \boldsymbol{\xi}^\top Q_{\boldsymbol{\xi}}(\lambda) \boldsymbol{\xi}) \\ &\propto \exp \left( \sum_{i=1}^n g_i(\boldsymbol{\xi}) - 0.5 \boldsymbol{\xi}^\top Q_{\boldsymbol{\xi}}(\lambda) \boldsymbol{\xi} \right). \end{aligned} \tag{2}$$

A second-order Taylor expansion of the log-likelihood yields a quadratic form in the latent vector  $\boldsymbol{\xi}$  and hence can be used to obtain a Laplace approximation to (2) as shown in appendix. In what follows, we denote by  $\tilde{p}_G(\boldsymbol{\xi} | \lambda, \mathcal{D}) = \mathcal{N}_{\dim(\boldsymbol{\xi})}(\boldsymbol{\xi}^*(\lambda), \Sigma_{\boldsymbol{\xi}}^*(\lambda))$  the Laplace approximation to  $p(\boldsymbol{\xi} | \lambda, \mathcal{D})$  for a given value of  $\lambda$ .

## 2.4 | Approximate posterior of the penalty parameter

The conditional posterior in (2) is a function of the penalty parameter  $\lambda$ . In a frequentist setting, an “optimal” value for  $\lambda$  is generally obtained by means of the Akaike information criterion (AIC) or (generalized) cross-validation. From a Bayesian perspective,  $\lambda$  is random and its associated posterior distribution is of crucial importance for optimal smoothing. Let  $\boldsymbol{\eta} = (\lambda, \delta)^\top$  denote the vector of hyperparameters. Mathematically, the posterior of  $\boldsymbol{\eta}$  is given by:

$$p(\boldsymbol{\eta} | \mathcal{D}) \propto \frac{\mathcal{L}(\boldsymbol{\xi}; \mathcal{D}) p(\boldsymbol{\xi} | \lambda) p(\lambda | \delta) p(\delta)}{p(\boldsymbol{\xi} | \lambda, \mathcal{D})}. \tag{3}$$



Using the Laplace approximation to the denominator  $p(\xi|\lambda, D)$  and replacing the latent vector by its modal value  $\xi^*(\lambda)$  from the Laplace approximation, the marginal posterior in (3) is approximated in the spirit of Tierney and Kadane<sup>38</sup>:

$$\begin{aligned} \tilde{p}(\eta|D) &\propto \frac{\exp(\ell(\xi; D))p(\xi|\lambda)p(\lambda|\delta)p(\delta)}{\tilde{p}_G(\xi|\lambda, D)} \Big|_{\xi=\xi^*(\lambda)} \\ &\propto \sqrt{\det(\Sigma_\xi^*(\lambda))} \exp\left(\sum_{i=1}^n g_i(\xi^*(\lambda)) - 0.5\xi^{*T}(\lambda)Q_\xi(\lambda)\xi^*(\lambda)\right) \lambda^{0.5(K+\varphi)-1} \delta^{0.5\varphi+a_\delta-1} \exp(-\delta(0.5\varphi\lambda + b_\delta)). \end{aligned}$$

Note that in the above formula, the determinant of the block diagonal matrix  $Q_\xi(\lambda)$  (coming from the Gaussian prior on  $\xi$ ) is equal to the product of the determinants of the composing blocks and hence  $\sqrt{\det(Q_\xi(\lambda))} \propto \lambda^{0.5K}$ , which explains part of the power to which the roughness penalty parameter is raised. Note also that  $\delta^{0.5\varphi+a_\delta-1} \exp(-\delta(0.5\varphi\lambda + b_\delta))$  is the kernel of a Gamma distribution with shape parameter  $0.5\varphi + a_\delta$  and rate parameter  $0.5\varphi\lambda + b_\delta$ , so that integrating the latter in the  $(0, +\infty)$  domain yields  $\Gamma(0.5\varphi + a_\delta)(0.5\varphi\lambda + b_\delta)^{-(0.5\varphi+a_\delta)}$ , where  $\Gamma(\cdot)$  is the gamma function. It follows that  $\delta$  can be analytically integrated out from  $\tilde{p}(\eta|D)$ :

$$\begin{aligned} \tilde{p}(\lambda|D) &= \int_0^{+\infty} \tilde{p}(\eta|D) d\delta \\ &\propto \sqrt{\det(\Sigma_\xi^*(\lambda))} \exp\left(\sum_{i=1}^n g_i(\xi^*(\lambda)) - 0.5\xi^{*T}(\lambda)Q_\xi(\lambda)\xi^*(\lambda)\right) \lambda^{0.5(K+\varphi)-1} (0.5\varphi\lambda + b_\delta)^{-(0.5\varphi+a_\delta)}. \end{aligned} \tag{4}$$

For numerical reasons it is more appropriate to work with the log transformed penalty parameter  $v = \log(\lambda)$  as the latter is unbounded. Using the transformation method for random variables, one obtains the following approximated (log) posterior for  $v$ :

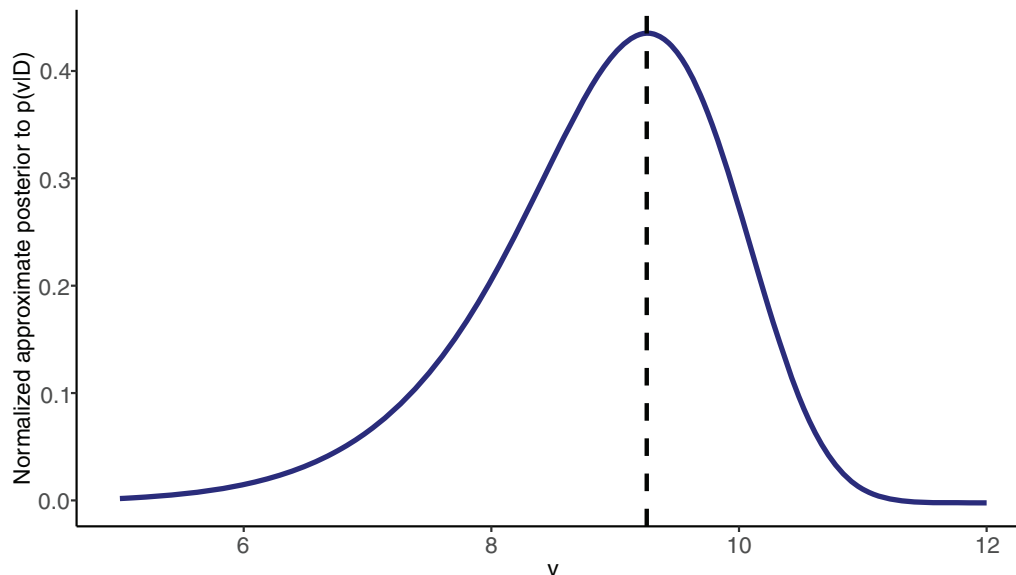
$$\begin{aligned} \log \tilde{p}(v|D) &\doteq 0.5 \log\left(\det(\Sigma_\xi^*(v))\right) + \sum_{i=1}^n g_i(\xi^*(v)) - 0.5\xi^{*T}(v)Q_\xi(v)\xi^*(v) \\ &\quad + 0.5(K + \varphi)v - (0.5\varphi + a_\lambda) \log(0.5\varphi \exp(v) + b_\delta), \end{aligned} \tag{5}$$

where  $\doteq$  denotes equality up to an additive constant. Approximation (5) provides a good starting point for various strategies to explore the posterior penalty space. A possibility is to use grid-based approaches<sup>21,25</sup> or MCMC algorithms<sup>39,40</sup> as often encountered in models with a multidimensional penalty space. In latent Gaussian models, the posterior penalty typically satisfies suitable regularity conditions such as unimodality<sup>41</sup> and not a “too-large” deviation from Gaussianity. This suggests to use a simple and yet efficient type of bracketing algorithm to compute the (approximate) posterior mode  $v^*$  of  $\log \tilde{p}(v|D)$ . Starting with an arbitrarily “large” value, say  $v_0 = 15$ , the algorithm moves in the left direction with a fixed step size  $d$ , that is, at the  $m$ th iteration  $v_m = v_{m-1} - d$ . Movement in the left direction continues until reaching  $v_{\tilde{m}}$ , the point at which the target function starts to point downhill  $\tilde{p}(v_{\tilde{m}}|D) < \tilde{p}(v_{\tilde{m}} + d|D)$ . The approximated modal value is then  $v^* = v_{\tilde{m}} + d/2$ . Figure 1 illustrates the normalized approximate posterior to  $p(v|D)$  along with the modal value (dashed line) obtained with a step size  $d = 0.01$ .

### 2.5 | Approximate credible intervals

The Laplace approximation to the conditional posterior of the latent vector evaluated at the (approximated) modal posterior value  $v^*$  (cf. Section 2.4) is denoted by  $\tilde{p}_G(\xi|v^*, D) = \mathcal{N}_{\dim(\xi)}\left(\xi^*(v^*), \Sigma_\xi^*(v^*)\right)$  and a point estimate for  $\xi$  is taken to be the mean/mode  $\xi^*(v^*)$  with associated variance-covariance matrix  $\Sigma_\xi^*(v^*)$ .

To ensure that the estimated baseline survival function  $\hat{S}_0(\cdot)$  “lands” smoothly on the horizontal asymptote at 0 near the end of the follow-up, we constrain the last B-spline coefficient by fixing  $\theta_K = 1$ . A major advantage of LPSMC is that credible intervals for (differentiable functions of) latent variables can be straightforwardly obtained starting from  $\tilde{p}_G(\xi|v^*, D)$ .



**FIGURE 1** Approximated (normalized) posterior of the log penalty  $v$ . The dashed line is the modal value  $v^*$  obtained with the bracketing algorithm using a step size  $\delta = 0.01$

### 2.5.1 | Credible interval for latent variables

A  $(1 - \alpha) \times 100\%$  (approximate) credible interval for a latent variable  $\xi_h \in \xi$  follows easily from the fact that the Laplace approximated posterior to  $\xi_h$  is  $\tilde{p}_G(\xi_h | v^*, \mathcal{D}) = \mathcal{N}_1(\xi_h^*, \sigma_{\xi_h}^{2*})$ , where  $\xi_h^*$  is the  $h$ th entry of the vector  $\xi^*(v^*)$  and  $\sigma_{\xi_h}^{2*}$  is the  $h$ th entry on the main diagonal of  $\Sigma_{\xi}^*(v^*)$ . It follows that a quantile-based  $(1 - \alpha) \times 100\%$  credible interval for  $\xi_h$  is:

$$CI_{\xi_h} = \xi_h^* \pm z_{\alpha/2} \sqrt{\sigma_{\xi_h}^{2*}},$$

where  $z_{\alpha/2}$  is the  $\alpha/2$ -upper quantile of a standard normal variate.

### 2.5.2 | Credible interval for the incidence $p(\mathbf{x})$ and cure rate $1 - p(\mathbf{x})$

The incidence of the mixture cure model is a function of the latent vector  $\check{\beta}$  and hence an appropriate approach to derive credible intervals for  $p(\mathbf{x})$  is through using a “delta” method. In particular, let us consider the following differentiable function of the probability to be uncured  $\psi(\check{\beta} | \mathbf{x}) = \log(-\log p(\mathbf{x})) = \log(\log(1 + \exp(-\beta_0 - \mathbf{x}^\top \beta)))$ , where  $\mathbf{x}$  is a known profile of the covariate vector. The Laplace approximated posterior to vector  $\check{\beta}$  is known to be  $\tilde{p}_G(\check{\beta} | v^*, \mathcal{D}) = \mathcal{N}_{\dim(\check{\beta})}(\check{\beta}^*, \Sigma_{\check{\beta}}^*)$  with mean vector  $\check{\beta}^* = \check{\beta}^*(v^*)$  and covariance matrix  $\Sigma_{\check{\beta}}^* = \Sigma_{\check{\beta}}^*(v^*)$ . The delta method operates via a first-order Taylor expansion of  $\psi(\check{\beta} | \mathbf{x})$  around  $\check{\beta}^*$ :

$$\psi(\check{\beta} | \mathbf{x}) \approx \psi(\check{\beta}^* | \mathbf{x}) + (\check{\beta} - \check{\beta}^*)^\top \nabla \psi(\check{\beta} | \mathbf{x}) \Big|_{\check{\beta} = \check{\beta}^*}, \quad (6)$$

with gradient:

$$\begin{aligned} \nabla \psi(\check{\beta} | \mathbf{x}) &= \left( \frac{\partial \psi(\check{\beta} | \mathbf{x})}{\partial \beta_0}, \dots, \frac{\partial \psi(\check{\beta} | \mathbf{x})}{\partial \beta_p} \right)^\top \\ &= \frac{-(1 - p(\mathbf{x}))}{\log(1 + \exp(-\beta_0 - \mathbf{x}^\top \beta))} (1, \mathbf{x}^\top)^\top. \end{aligned}$$

Note that  $\psi(\check{\beta}|\mathbf{x})$  in (6) is still Gaussian as it is a linear combination of a random vector  $\check{\beta}$  that is a posteriori (approximated by a) Gaussian due to the Laplace approximation with mean  $\mathbb{E}(\psi(\check{\beta}|\mathbf{x})) \approx \psi(\check{\beta}^*|\mathbf{x})$  and covariance matrix  $\mathbb{V}(\psi(\check{\beta}|\mathbf{x})) \approx \nabla^\top \psi(\check{\beta}|\mathbf{x})|_{\check{\beta}=\check{\beta}^*} \Sigma_{\check{\beta}}^* \nabla \psi(\check{\beta}|\mathbf{x})|_{\check{\beta}=\check{\beta}^*}$ . This suggests to write the approximated posterior of  $\psi(\check{\beta}|\mathbf{x})$  as:

$$(\psi(\check{\beta}|\mathbf{x})|D) \sim \mathcal{N}_1 \left( \psi(\check{\beta}^*|\mathbf{x}), \nabla^\top \psi(\check{\beta}|\mathbf{x})|_{\check{\beta}=\check{\beta}^*} \Sigma_{\check{\beta}}^* \nabla \psi(\check{\beta}|\mathbf{x})|_{\check{\beta}=\check{\beta}^*} \right),$$

and so an approximate quantile-based  $(1 - \alpha) \times 100\%$  credible interval for  $\psi(\check{\beta}|\mathbf{x})$  is:

$$CI_{\psi(\check{\beta}|\mathbf{x})} = \psi(\check{\beta}^*|\mathbf{x}) \pm z_{\alpha/2} \sqrt{\nabla^\top \psi(\check{\beta}|\mathbf{x})|_{\check{\beta}=\check{\beta}^*} \Sigma_{\check{\beta}}^* \nabla \psi(\check{\beta}|\mathbf{x})|_{\check{\beta}=\check{\beta}^*}} \tag{7}$$

Multiplying the values in the interval (7) by  $\exp(-\exp(\cdot))$  gives us the desired credible interval for the incidence  $p(\mathbf{x})$ . If a credible interval for the cure rate  $1 - p(\mathbf{x})$  is required, simply use the transformation  $\psi(\check{\beta}|\mathbf{x}) = \log(-\log(1 - p(\mathbf{x}))) = \log(\log(1 + \exp(\beta_0 + \mathbf{x}^\top \boldsymbol{\beta})))$  with gradient  $\nabla \psi(\check{\beta}|\mathbf{x}) = (p(\mathbf{x})/\log(1 + \exp(\beta_0 + \mathbf{x}^\top \boldsymbol{\beta}))) (1, \mathbf{x}^\top)^\top$ .

### 2.5.3 | Credible interval for $S_0(\cdot)$ and $S_u(\cdot|\mathbf{z})$

Let us denote by  $t_q$  the  $q$ th quantile of the distribution of the survival time  $T$  at baseline by  $t_q = \inf\{t|S_0(t) \leq 1 - q\}$ . The “delta” method can also be used to compute an approximate credible interval for  $S_0(\cdot)$  at  $t_q$  by using a  $\log(-\log(\cdot))$  transformation  $\psi(\theta|t_q) = \log(-\log(S_0(t_q))) = \log\left(\sum_{j=1}^{j(t_q)} \exp(\theta^\top b(s_j))\Delta\right)$ . Starting from the Laplace approximated posterior  $\tilde{p}_G(\theta^*|v^*, D) = \mathcal{N}_{\dim(\theta)}(\theta^*, \Sigma_{\theta}^*)$ , one can show that the resulting credible interval for  $\psi(\theta|t_q)$  is:

$$CI_{\psi(\theta|t_q)} = \psi(\theta^*|t_q) \pm z_{\alpha/2} \sqrt{\nabla^\top \psi(\theta|t_q)|_{\theta=\theta^*} \Sigma_{\theta}^* \nabla \psi(\theta|t_q)|_{\theta=\theta^*}} \tag{8}$$

where  $\nabla \psi(\theta|t_q)|_{\theta=\theta^*}$  is the gradient of  $\psi(\theta|t_q)$  with respect to  $\theta$  evaluated at  $\theta^*$  and can be found in Gressani and Lambert<sup>21</sup>, Appendix C. Applying the inverse transformation  $\exp(-\exp(\cdot))$  to (8) yields the desired  $(1 - \alpha) \times 100\%$  credible interval for  $S_0(\cdot)$  at  $t_q$ .

The same approach is used to construct credible intervals for the survival function of the uncured  $S_u(t|\mathbf{z}) = S_0(t)^{\exp(\mathbf{z}^\top \boldsymbol{\gamma})}$  at  $t_q = \inf\{t|S_u(t|\mathbf{z}) \leq 1 - q\}$  for a given covariate profile  $\mathbf{z}$ . Applying the  $\log(-\log(\cdot))$  transform yields  $\psi(\theta, \boldsymbol{\gamma}|t_q, \mathbf{z}) = \mathbf{z}^\top \boldsymbol{\gamma} + \log\left(\sum_{j=1}^{j(t_q)} \exp(\theta^\top b(s_j))\Delta\right)$  with gradient:

$$\begin{aligned} \nabla \psi(\theta, \boldsymbol{\gamma}|t_q, \mathbf{z}) &= \left( \frac{\partial \psi(\theta, \boldsymbol{\gamma}|t_q, \mathbf{z})}{\partial \theta_1}, \dots, \frac{\partial \psi(\theta, \boldsymbol{\gamma}|t_q, \mathbf{z})}{\partial \theta_K}, \frac{\partial \psi(\theta, \boldsymbol{\gamma}|t_q, \mathbf{z})}{\partial \gamma_1}, \dots, \frac{\partial \psi(\theta, \boldsymbol{\gamma}|t_q, \mathbf{z})}{\partial \gamma_q} \right)^\top \\ &= \begin{pmatrix} \left(\sum_{j=1}^{j(t_q)} \exp(\theta^\top b(s_j))\Delta\right)^{-1} \sum_{j=1}^{j(t_q)} \exp(\theta^\top b(s_j)) b_1(s_j)\Delta \\ \vdots \\ \left(\sum_{j=1}^{j(t_q)} \exp(\theta^\top b(s_j))\Delta\right)^{-1} \sum_{j=1}^{j(t_q)} \exp(\theta^\top b(s_j)) b_K(s_j)\Delta \\ \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_q \end{pmatrix} \end{aligned}$$

The resulting credible interval for  $\psi(\theta, \boldsymbol{\gamma}|t_q, \mathbf{z})$  is:

$$CI_{\psi(\theta, \boldsymbol{\gamma}|t_q, \mathbf{z})} = \psi(\theta^*, \boldsymbol{\gamma}^*|t_q, \mathbf{z}) \pm z_{\alpha/2} \sqrt{\nabla^\top \psi(\theta, \boldsymbol{\gamma}|t_q, \mathbf{z})|_{\theta=\theta^*, \boldsymbol{\gamma}=\boldsymbol{\gamma}^*} \Sigma_{\theta, \boldsymbol{\gamma}}^* \nabla \psi(\theta, \boldsymbol{\gamma}|t_q, \mathbf{z})|_{\theta=\theta^*, \boldsymbol{\gamma}=\boldsymbol{\gamma}^*}} \tag{9}$$

where  $\Sigma_{\theta, \boldsymbol{\gamma}}^*$  is the covariance matrix of the vector  $(\theta^\top, \boldsymbol{\gamma}^\top)^\top$  obtained from the Laplace approximation. Finally, an  $\exp(-\exp(\cdot))$  transform on (9) gives a  $(1 - \alpha) \times 100\%$  credible interval for  $S_u(t|\mathbf{z})$  at  $t_q$  for a given covariate vector  $\mathbf{z}$ .



### 3 | BAYESIAN INFERENCE BASED ON MARKOV CHAIN MONTE CARLO

A classic alternative to the Laplacian-P-splines methodology for Bayesian inference in the mixture cure model is to rely on Markov chain Monte Carlo methods. In this section, we propose a fully stochastic algorithm to explore the joint posterior distribution  $p(\xi, \lambda, \delta | D)$ . The sampling scheme is based on a Metropolis-adjusted Langevin strategy to draw samples from the conditional posterior of the latent vector  $p(\xi | \lambda, D)$  given in (2). Such a strategy has already been proposed by Lambert and Eilers<sup>42</sup> in a proportional hazards model with time-varying regression coefficients and also later by Lambert and Eilers<sup>43</sup> for density estimation. The Gibbs sampler can then be used to directly simulate from the conditional posterior distributions of the remaining parameters  $\lambda$  and  $\delta$ . The proposed Metropolis-Langevin-within-Gibbs (MLWG) sampler complements the LPSMC approach and gives the end user the opportunity to choose between a totally sampling-free approach (as provided by LPSMC in Section 2) and a fully stochastic approach (as provided by MLWG) for inference. The conditional posterior of the roughness penalty parameter is given by:

$$\begin{aligned} p(\lambda | \xi, \delta, D) &\propto \lambda^{0.5(K+\varphi)-1} \exp(-0.5 (\xi^\top Q_\xi(\lambda) \xi + \lambda \varphi \delta)) \\ &\propto \lambda^{0.5(K+\varphi)-1} \exp(-0.5 (\xi^\top \{ \lambda \mathcal{P} + \mathcal{B} \} \xi + \lambda \varphi \delta)) \\ &\propto \lambda^{0.5(K+\varphi)-1} \exp(-\lambda 0.5 (\xi^\top \mathcal{P} \xi + \varphi \delta)), \end{aligned}$$

so that  $(\lambda | \xi, \delta, D) \sim \mathcal{G}(0.5(K + \varphi), 0.5 (\xi^\top \mathcal{P} \xi + \varphi \delta))$ , where  $\mathcal{P}$  is a square matrix of dimension  $\dim(\xi)$  with penalty matrix  $P$  on the upper-left block and  $\mathcal{B}$  a matrix with the same dimension having  $\zeta I_{(p+1)+q}$  on the lower-right block. The conditional posterior of the dispersion hyperparameter is shown to be  $(\delta | \xi, \lambda, D) \sim \mathcal{G}(0.5\varphi + a_\delta, 0.5\varphi\lambda + b_\delta)$ .

#### 3.1 | Metropolis-adjusted Langevin algorithm

The Metropolis-adjusted Langevin algorithm (MALA)<sup>44</sup> is used to explore the conditional posterior distribution  $p(\xi | \lambda, D)$ . Compared to a classic random-walk Metropolis algorithm, MALA was shown to guarantee a more efficient sampling scheme in the sense of improved mixing and convergence of the chains in the mixture cure model. Two main reasons explain this extra efficiency. First, the proposal distribution in MALA is constructed around Langevin diffusions that use the gradient of the (log) posterior target distribution, here  $\nabla \log p(\xi | \lambda, D) \doteq \nabla \ell(\xi; D) - Q_\xi(\lambda) \xi$ . Second, the covariance matrix of the proposal distribution is based on the covariance matrix of the Laplace approximation  $\Sigma^* := \Sigma_\xi^*(\lambda^*)$  that sets a desirable correlation structure between latent variables. Let  $\xi^{(m-1)}$  and  $\lambda^{(m-1)}$  denote the state of the chain for  $\xi$  and  $\lambda$  respectively at iteration  $(m - 1)$ . In MALA, the candidate latent vector at iteration  $m$  is a random draw from the following Gaussian proposal distribution:

$$\xi^{(\text{prop})} \sim \mathcal{N}_{\dim(\xi)}(\xi^{(m-1)} + 0.5\rho\Sigma^*\nabla \log p(\xi^{(m-1)} | \lambda^{(m-1)}, D), \rho\Sigma^*), \quad (10)$$

where  $\rho$  is a tuning parameter that is adaptively calibrated<sup>45,46</sup> to reach the optimal acceptance rate of 0.57.<sup>47</sup> The MLWG algorithm to explore  $p(\xi, \lambda, \delta | D)$  is summarized below.

---

#### Metropolis-Langevin-within-Gibbs algorithm to explore $p(\xi, \lambda, \delta | D)$

---

- 1: Set  $m = 0$  and fix initial values  $\xi^{(0)}$ ,  $\lambda^{(0)}$ ,  $\delta^{(0)}$  and  $\rho^{(0)}$ .
  - 2: **for**  $m = 1, \dots, M$  **do**
  - 3:    (*Metropolis-adjusted Langevin*)
  - 4:    Compute the Langevin diffusion:  $\mathcal{E}(\xi^{(m-1)}) = \xi^{(m-1)} + 0.5\rho^{(m-1)}\Sigma^*\nabla \log p(\xi^{(m-1)} | \lambda^{(m-1)}, D)$ .
  - 5:    Generate proposal:  $\xi^{(\text{prop})} \sim \mathcal{N}_{\dim(\xi)}(\mathcal{E}(\xi^{(m-1)}), \rho^{(m-1)}\Sigma^*)$ .
  - 6:    Compute the acceptance probability  $\pi(\xi^{(m-1)}, \xi^{(\text{prop})}) = \min\left\{1, \frac{p(\xi^{(\text{prop})} | \lambda^{(m-1)}, D) q(\xi^{(m-1)}, \xi^{(\text{prop})})}{p(\xi^{(m-1)} | \lambda^{(m-1)}, D) q(\xi^{(m-1)}, \xi^{(\text{prop})})}\right\}$ .
  - 7:    Draw  $u \sim \mathcal{U}(0, 1)$ .
  - 8:    **if**  $u \leq \pi$  **set**  $\xi^{(m)} = \xi^{(\text{prop})}$  (**accept**), **else**  $\xi^{(m)} = \xi^{(m-1)}$  (**reject**).
  - 9:    (*Gibbs sampler*)
  - 10:    Draw  $\lambda^{(m)} \sim \mathcal{G}(0.5(K + \varphi), 0.5 (\xi^{(m)\top} \mathcal{P} \xi^{(m)} + \varphi \delta^{(m-1)}))$ .
  - 11:    Draw  $\delta^{(m)} \sim \mathcal{G}(0.5\varphi + a_\delta, 0.5\varphi\lambda^{(m)} + b_\delta)$ .
  - 12:    (*Update tuning parameter*)
  - 13:    Set  $\rho^{(m)} = \hat{\kappa}^2 \left( \sqrt{\rho^{(m-1)}} + m^{-1} (\pi(\xi^{(m-1)}, \xi^{(\text{prop})}) - 0.57) \right)$ .
  - 14: **end for**
-

TABLE 1 Parameters for the incidence, latency, and censoring rate yielding different cure and censoring levels

Scenario	$\beta_0$	$\beta_1$	$\beta_2$	$\gamma_1$	$\gamma_2$	Cure	$\mu_c$	Censoring	Plateau
1	0.70	-1.15	0.95	-0.10	0.25	28.77%	0.16	48.56%	8.00%
2	1.25	-0.75	0.45	-0.10	0.20	21.20%	0.05	29.32%	14.21%

Note: The last column indicates the percentage of observations in the plateau located at the right tail of the Kaplan-Meier curve.

The kernel of the proposal density  $q(\xi^{(prop)}, \xi^{(m-1)})$  is given by:

$$\ker(q(\xi^{(prop)}, \xi^{(m-1)})) = \exp\left(-\frac{1}{2\varrho^{(m-1)}}(\xi^{(m-1)} - \xi^{(prop)} - 0.5\varrho^{(m-1)}\Sigma^* \mathcal{G}^{(prop)})^\top (\Sigma^*)^{-1} (\xi^{(m-1)} - \xi^{(prop)} - 0.5\varrho^{(m-1)}\Sigma^* \mathcal{G}^{(prop)})\right),$$

where  $\mathcal{G}^{(prop)} := \nabla \log p(\xi^{(prop)} | \lambda^{(m-1)}, \mathcal{D})$ . Defining  $\mathcal{G}^{(m-1)} := \nabla \log p(\xi^{(m-1)} | \lambda^{(m-1)}, \mathcal{D})$ , one can show that the ratio of proposal distributions in the computation of the acceptance probability (line 6 of the above algorithm) is:

$$\frac{q(\xi^{(prop)}, \xi^{(m-1)})}{q(\xi^{(m-1)}, \xi^{(prop)})} = \exp\left(-\frac{1}{2}(\mathcal{G}^{(prop)} + \mathcal{G}^{(m-1)})^\top \left(\xi^{(prop)} - \xi^{(m-1)} + \frac{\varrho^{(m-1)}}{4}\Sigma^* (\mathcal{G}^{(prop)} - \mathcal{G}^{(m-1)})\right)\right).$$

The function  $\mathbb{I}(\cdot)$  on line 13 of the MLWG algorithm is given by  $\mathbb{I}(z) = \varepsilon \mathbb{I}(z < \varepsilon) + z \mathbb{I}(\varepsilon \leq z \leq \mathcal{A}) + \mathcal{A} \mathbb{I}(z > \mathcal{A})$ , with  $\varepsilon = 10^{-4}$  and  $\mathcal{A} = 10^4$ , see also Lambert and Eilers<sup>43</sup>. The above MLWG algorithm is used to explore the posterior  $p(\xi, \lambda, \delta | \mathcal{D})$  and the generated samples (after burn-in) permit to compute posterior estimates of latent variables.

#### 4 | SIMULATION STUDY

To measure the statistical performance of the LPSMC methodology in a mixture cure model setting, we consider a numerical study where survival data is generated according to different cure and censoring rates. Generation of survival data for the  $i$ th subject is as follows. The incidence part is generated from a logistic regression function with two covariates  $p(\mathbf{X}_i) = 1/(1 + \exp(-\beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2}))$ , where  $X_{i1}$  is a standard normal variate and  $X_{i2} \sim \text{Bern}(0.5)$  is a Bernoulli random variable. The cure status is generated as a Bernoulli random variable with failure probability  $p(\mathbf{X}_i)$ , that is,  $B_i \sim \text{Bern}(p(\mathbf{X}_i))$ . Survival times  $T_i$  for the uncured subjects ( $B_i = 1$ ) are obtained from a Weibull Cox proportional hazards model and are truncated at  $\tau_0 = 8$ . The latency is given by  $S_u(t_i | \mathbf{Z}_i) = \exp(-\nu t_i^\kappa \exp(\gamma_1 Z_{i1} + \gamma_2 Z_{i2}))$ , with scale parameter  $\nu = 0.25$  and shape  $\kappa = 1.45$ . The covariate vector  $\mathbf{Z}_i = (Z_{i1}, Z_{i2})^\top$  is independent of  $\mathbf{X}_i$ . We assume that  $Z_{i1}$  follows a standard Gaussian distribution and  $Z_{i2} \sim \text{Bern}(0.4)$ . For the cured subject ( $B_i = 0$ ), the theoretically infinite survival times are replaced by a large value, say  $T_i = 20\,000$ . The censoring time  $C_i$  is independent of the vector  $(\mathbf{X}_i^\top, \mathbf{Z}_i^\top, T_i)^\top$  and is sampled from an exponential distribution with mean  $1/\mu_c$  (ie, density  $f(c_i) = \mu_c \exp(-\mu_c c_i)$ ) truncated at  $\tau_1 = 11$ . We consider samples of sizes  $n = 300$  and  $n = 600$  and simulate survival data with two scenarios for the coefficients  $\beta, \gamma$  and  $\mu_c$ , yielding different censoring and cure rates. In both scenarios (almost) all the observations in the plateau of the Kaplan-Meier<sup>48</sup> estimator are cured, such that the simulated data are representative of the practical real case scenarios for which mixture cure models are used. Table 1 provides a summary of the two considered scenarios.

We specify 15 cubic B-splines in the interval  $[0, t_u]$  with upper bound fixed at  $t_u = 11$  and a third order penalty to counterbalance model flexibility. In the bracketing algorithm, we use a step size  $\delta = 0.2$  to compute the (approximate) modal posterior log penalty value  $v^*$ . For each scenario, we simulate  $S = 500$  replications and compute the bias, empirical standard error (ESE), root mean square error (RMSE) and coverage probabilities for 90% and 95% (quantile-based) credible intervals. Results for LPSMC and the MLWG sampler with a chain of length 10 000 and a burn-in of 3000 are summarized in Table 2. The bias is negligible and the ESE and RMSE decrease with larger sample size, as expected. In addition, the estimated coverage probabilities are close to their respective nominal value in all scenarios. What can also be noted from Table 2 is the similarity between the LPSMC and MLWG results, showing that the Laplace approximation is quite accurate and competitive against a fully MCMC based approach even with a relatively long chain length. LPSMC is of course much faster than MLWG as it takes  $\approx 0.7$  s to fit the model in  $\mathbf{R}$  with an Intel Xeon E-2186M processor at

TABLE 2 Numerical results for  $S = 500$  replications of sample size  $n = 300$  and  $n = 600$  under two different cure-censoring scenarios

Scenario	Parameters	Mean	Bias	ESE	RMSE	CP <sub>90%</sub>	CP <sub>95%</sub>
Scenario 1 ( $n = 300$ )	$\beta_0 = 0.70$	0.721 (0.743)	0.021 (0.043)	0.249 (0.274)	0.250 (0.277)	91.0 (88.4)	96.4 (94.4)
	$\beta_1 = -1.15$	-1.182 (-1.230)	-0.032 (-0.080)	0.240 (0.255)	0.242 (0.267)	91.6 (89.8)	94.8 (94.2)
	$\beta_2 = 0.95$	0.954 (1.044)	0.004 (0.094)	0.390 (0.407)	0.390 (0.417)	90.6 (90.4)	94.2 (95.4)
	$\gamma_1 = -0.10$	-0.101 (-0.098)	-0.001 (0.002)	0.092 (0.094)	0.092 (0.094)	89.6 (87.8)	94.4 (94.8)
	$\gamma_2 = 0.25$	0.242 (0.234)	-0.008 (-0.016)	0.184 (0.188)	0.184 (0.188)	89.0 (89.2)	96.4 (94.8)
Scenario 1 ( $n = 600$ )	$\beta_0 = 0.70$	0.703 (0.715)	0.003 (0.015)	0.184 (0.181)	0.184 (0.181)	89.6 (90.0)	95.6 (95.0)
	$\beta_1 = -1.15$	-1.154 (-1.190)	-0.004 (-0.040)	0.167 (0.172)	0.167 (0.176)	90.2 (88.2)	95.0 (93.8)
	$\beta_2 = 0.95$	0.950 (0.978)	0.000 (0.028)	0.269 (0.280)	0.268 (0.281)	90.8 (88.4)	95.0 (93.8)
	$\gamma_1 = -0.10$	-0.102 (-0.100)	-0.002 (0.000)	0.064 (0.065)	0.064 (0.065)	89.6 (89.0)	94.8 (93.8)
	$\gamma_2 = 0.25$	0.254 (0.247)	0.004 (-0.003)	0.127 (0.123)	0.127 (0.123)	90.8 (91.4)	96.0 (95.2)
Scenario 2 ( $n = 300$ )	$\beta_0 = 1.25$	1.281 (1.275)	0.031 (0.025)	0.229 (0.239)	0.230 (0.240)	91.4 (91.0)	95.6 (94.2)
	$\beta_1 = -0.75$	-0.765 (-0.784)	-0.015 (-0.034)	0.182 (0.197)	0.183 (0.200)	91.0 (88.2)	95.4 (93.2)
	$\beta_2 = 0.45$	0.430 (0.480)	-0.020 (0.030)	0.330 (0.333)	0.330 (0.334)	89.0 (90.2)	94.8 (95.0)
	$\gamma_1 = -0.10$	-0.103 (-0.102)	-0.003 (-0.002)	0.074 (0.075)	0.074 (0.075)	88.8 (90.2)	94.4 (95.0)
	$\gamma_2 = 0.20$	0.194 (0.196)	-0.006 (-0.004)	0.151 (0.157)	0.151 (0.157)	87.2 (87.8)	94.8 (93.4)
Scenario 2 ( $n = 600$ )	$\beta_0 = 1.25$	1.248 (1.270)	-0.002 (0.020)	0.160 (0.164)	0.160 (0.165)	91.4 (90.8)	95.6 (95.8)
	$\beta_1 = -0.75$	-0.748 (-0.771)	0.002 (-0.021)	0.129 (0.132)	0.129 (0.133)	89.8 (88.6)	95.2 (94.4)
	$\beta_2 = 0.45$	0.459 (0.467)	0.009 (0.017)	0.223 (0.238)	0.223 (0.239)	91.0 (89.8)	95.6 (94.4)
	$\gamma_1 = -0.10$	-0.100 (-0.106)	0.000 (-0.006)	0.054 (0.052)	0.054 (0.052)	88.0 (89.6)	95.6 (94.0)
	$\gamma_2 = 0.20$	0.199 (0.195)	-0.001 (-0.005)	0.105 (0.111)	0.105 (0.111)	89.6 (87.8)	95.0 (92.6)

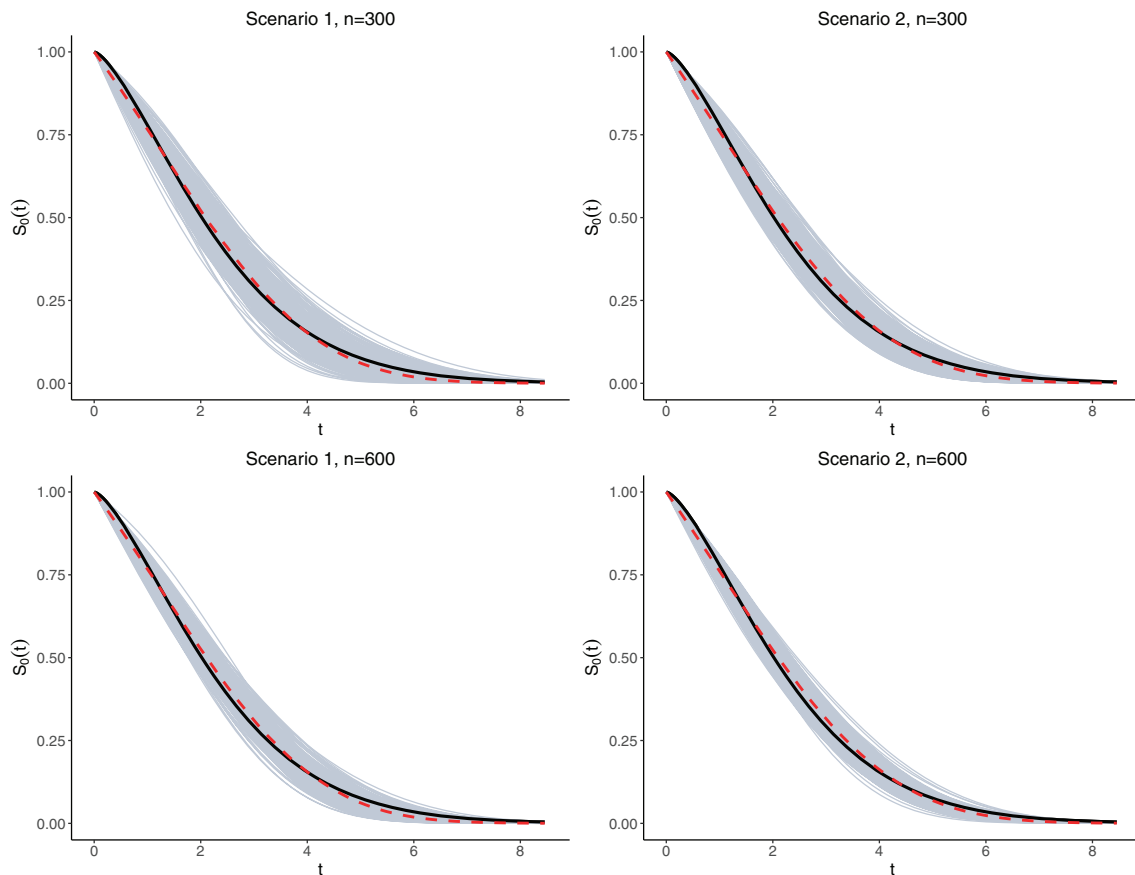
Note: Results in parenthesis are for the MLWG sampler with chain length 10 000 and burn-in 3000.

2.90GHz. Figure 2 shows the estimated baseline survival curves (gray), the target (solid) and the pointwise median of the 500 curves (dashed) for the different scenarios using the LPSMC approach.

Another performance measure for the incidence of the model is obtained by computing the average squared error (ASE) of  $\hat{p}(\mathbf{x})$  defined as  $ASE(\hat{p}) = M^{-1} \sum_{m=1}^M (\hat{p}(\mathbf{x}_m) - p(\mathbf{x}_m))^2$ . The latter quantity is computed on triplets of covariate values  $\mathbf{x}_m = (1, x_{m1}, x_{m2})^T$  for  $m = 1, \dots, M$ , where the set of couples  $\{(x_{m1}, x_{m2})\}_{m=1}^M$  equals the Cartesian product between an equidistant grid in  $[-1.50, 1.50]$  with step size 0.001 (for variable  $x_1$ ) and the set  $\{0, 1\}$  (for  $x_2$ ).

Boxplots for the ASE in the different scenarios are displayed in Figure 3. Coverage probability of the 90% and 95% (approximate) credible interval for the incidence  $p(\mathbf{x})$  (and cure rate  $1 - p(\mathbf{x})$ ) at the mean covariate profile  $\bar{\mathbf{x}} = (1, 0, 0.5)^T$  as computed from (7) have also been measured (with LPSMC) and are close to their nominal value in all scenarios. In Table 3, the performance of approximate credible intervals for the baseline survival and survival of the uncured for a mean covariate profile  $\bar{\mathbf{z}} = (0, 0.4)^T$  at selected quantiles of  $T$  is shown. Results are for  $S = 200$  replications of sample size  $n = 300$  with  $K = 20$  B-splines. In Scenario 1, there is a more pronounced difference in coverage performance between LPSMC and MLWG at quantiles  $t_{0.30}$ ,  $t_{0.40}$ , and  $t_{0.50}$ , where the MLWG sampler exhibits estimated coverage probabilities moderately below nominal coverage. The average credible interval widths and baseline bias  $S^{-1} \sum_s (\hat{S}_0(\cdot) - S_0(\cdot))$  (computed in the supplementary materials) at the selected quantiles  $t_q$  can be used to explain this difference. In Scenario 1, for  $t_{0.30}$ ,  $t_{0.40}$ , and  $t_{0.50}$ , the mean credible interval widths are roughly identical for LPSMC and the sampling approach, but the baseline bias (although negligible) is slightly higher for MLWG and thus implies a small loss in coverage. For the remaining quantiles, LPSMC and MLWG have comparable performance except for  $t_{0.80}$  and  $t_{0.90}$ , where both methods undercover but MLWG is closer to the nominal level in both scenarios.

The empirical computation time required for implementing LPSMC and MLWG (via the `mixcurelps` package) is assessed for different combinations of sample size  $n$  and B-spline dimension  $K$ . The real elapsed time averaged over  $S = 10$  replications in each  $(n, K)$ -pair is reported in Table 4. Even in relatively large sample settings ( $n = 1000$ ) with a large number of B-spline coefficients (such as  $K = 50$ ), LPSMC is able to fit the mixture cure model in a couple of seconds.



**FIGURE 2** Estimated baseline survival curves (gray) for  $S = 500$  replications under different scenarios with LPSMC. The black curve is the target baseline and the dashed curve is the pointwise median of the 500 gray curves

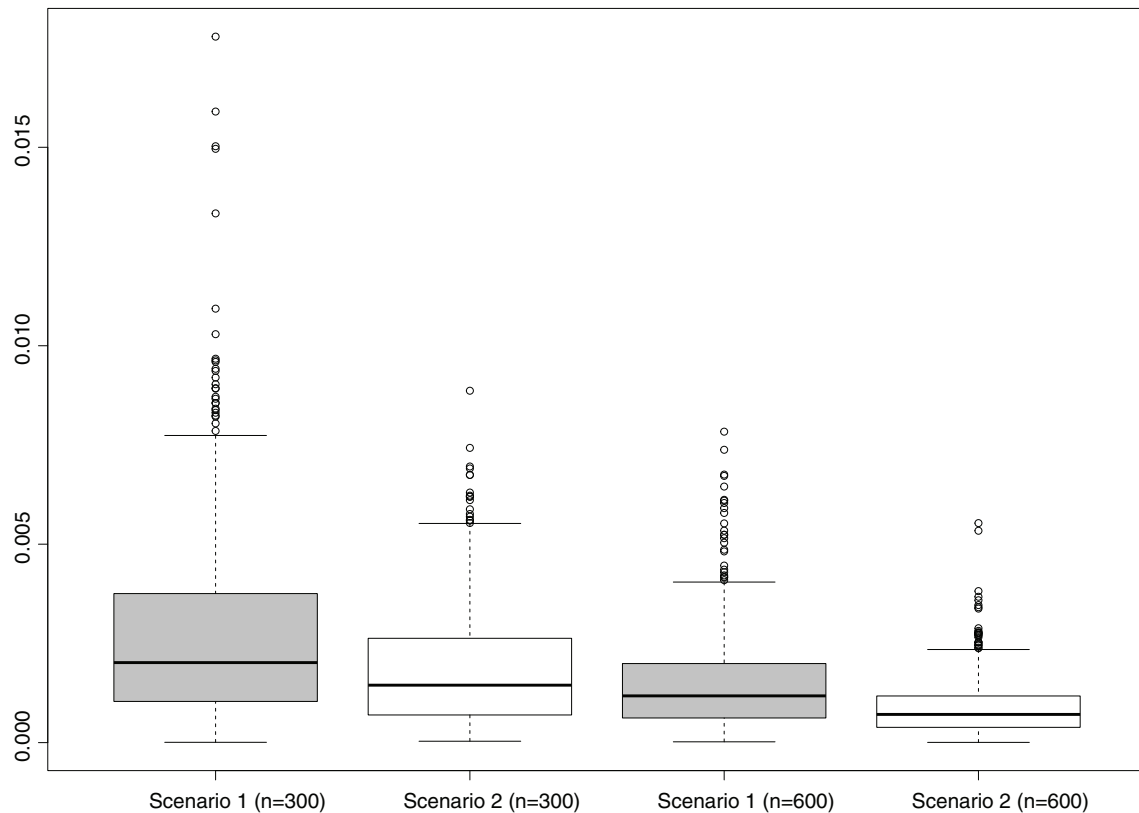
Computational times for LPSMC on denser grids of  $n$  and  $K$  respectively are reported in the supplementary materials and suggest that the underlying algorithms run in linear time  $\mathcal{O}(n)$  with respect to the sample size and in more than linear time with respect to the spline dimension.

## 5 | REAL DATA APPLICATIONS

### 5.1 | ECOG e1684 clinical trial

We start by applying the LPSMC methodology to a well-known dataset in the cure literature and compare the estimates with the ones obtained using the benchmark **smcure** package.<sup>49</sup> The data comes from the Eastern Cooperative Oncology Group (ECOG) phase III two-arm clinical trial with  $n = 284$  observations after removing missing data. The aim of the study was to assess whether Interferon alpha-2b (IFN) had a significant effect on relapse-free survival. There were 144 patients receiving the treatment and the remaining patients (140) belonged to the control group. The response of interest is relapse-free survival time measured in years. The following covariates enter both the incidence and latency part of the model.  $TRT$  is the binary variable indicating whether a patient received the IFN treatment ( $TRT = 1$ ) or not ( $TRT = 0$ ). Variable  $SEX$  indicates if a patient is female ( $SEX = 1$ ) or male ( $SEX = 0$ ). In total, the study involves 113 women and 171 men. Finally,  $AGE$  is a continuous covariate (centered to the mean) indicating the age of patients. Figure 4, shows the Kaplan-Meier estimated survival curve for the e1684 dataset. A plateau is clearly visible indicating the potential presence of a cure fraction, so that a mixture cure model is appropriate to fit this type of data.

Table 5 summarizes the estimation results for the e1684 dataset with LPSMC, MLWG with chain length 20 000 (burn-in of 10 000) and **smcure**. The estimated coefficients in the two parts of the model (incidence and latency) are of similar



**FIGURE 3** Boxplots of the ASE for the incidence in different scenarios

magnitude for all approaches. However, the computational times to fit the mixture cure model drastically differ depending on the method.

It takes approximately 125 s to fit the model with **smcure**, while LPSMC takes only  $\approx 0.5$  s. This speedup factor of  $\approx 250$  is mainly due to the bootstrap samples that **smcure** uses to compute estimates of the variance of estimated parameters. The LPSMC approach is completely sampling-free and intrinsically accounts for fast computation of credible intervals for the latent variables. Results for both **smcure** and LPSMC/MLWG show that treatment (TRT) has a significant and positive impact on the cure probability, while IFN treatment has no significant impact in the latency part. In other words, taking *AGE* and *SEX* into account, receiving the IFN treatment decreases the probability to be susceptible but will not postpone relapse among the uncured. Similar findings are reported in Corbière and Joly<sup>50</sup> and Legrand and Bertrand<sup>51</sup>. The e1684 dataset has also been analyzed by Lázaro et al<sup>24</sup> with the INLA methodology. For this dataset, they compared the runtimes of INLA with a Weibull baseline hazard (via modal Gibbs running 500 iterations) with an MCMC algorithm implemented in JAGS.<sup>52</sup> The authors reported computational times of 28 min for INLA and around 33 seconds per 1000 iterations for MCMC (55 minutes for three Markov chains with 100 000 iterations). This is substantially above the runtime that we measured with our MLWG sampler (around 3 seconds per 1000 iterations) which is approximately 10-fold faster.

## 5.2 | Breast cancer data

The second dataset used to illustrate the LPSMC methodology concerns  $n = 286$  patients with breast cancer studied in Wang et al<sup>53</sup>. Survival time (in days) is defined as the distant-metastasis-free survival (DMFS), that is, time to occurrence of distant metastases or death (whatever happens first) and the event of interest is distant-metastasis. The data is obtained from the **breastCancerVDX** package<sup>54</sup> on Bioconductor (<https://bioconductor.org/>). We consider two covariates entering simultaneously in the incidence and latency part, namely the age of patients, ranging between 26 and 83 years with a median of 52 years and the categorical variable Estrogen Receptor (*ER*) (with  $ER = 0$ : if  $\leq 10$  fmol per mg protein [77 patients] and  $ER = 1$  otherwise [209 patients]). We use  $K = 20$  cubic B-splines in  $[0, t_u]$ , with  $t_u = 5201$  days, the

**TABLE 3** Estimated coverage probability of 90% and 95% credible intervals for  $S_0(\cdot)$  and  $S_u(\cdot|\bar{z}) = (0, 0.4)^T$  with  $S = 200$  replications of sample size  $n = 300$  and  $K = 20$  B-splines at selected quantiles  $t_q$

	Nominal	Scenario	$t_{0.20}$	$t_{0.30}$	$t_{0.40}$	$t_{0.50}$	$t_{0.60}$	$t_{0.70}$	$t_{0.80}$	$t_{0.90}$
Baseline $S_0(\cdot)$	90 (LPSMC)	1	88.5	86.5	90.0	92.0	93.0	93.5	81.0	41.5
	90 (MLWG)	1	86.0	84.0	81.0	81.0	88.0	90.5	85.5	61.5
	95 (LPSMC)	1	93.5	94.0	94.5	94.5	95.0	95.5	90.0	49.0
	95 (MLWG)	1	92.0	88.5	88.5	90.5	93.5	95.0	92.5	74.5
	90 (LPSMC)	2	85.0	83.5	86.0	88.0	88.5	89.0	70.5	18.5
	90 (MLWG)	2	84.0	80.5	82.0	86.0	87.5	90.0	79.5	40.5
	95 (LPSMC)	2	93.0	92.5	91.0	94.5	94.5	92.0	79.0	30.5
	95 (MLWG)	2	93.0	90.0	90.0	92.0	95.5	94.0	87.0	52.5
Uncured $S_u(\cdot \bar{z})$	90 (LPSMC)	1	88.5	89.5	88.5	89.5	94.0	91.5	77.5	22.5
	90 (MLWG)	1	82.0	76.0	75.5	79.5	86.0	88.5	81.5	51.0
	95 (LPSMC)	1	96.0	93.5	95.0	96.5	98.0	97.0	85.5	34.0
	95 (MLWG)	1	89.0	83.5	84.5	87.0	93.0	95.5	90.0	60.5
	90 (LPSMC)	2	87.0	81.5	85.5	90.0	88.5	87.0	68.0	10.0
	90 (MLWG)	2	74.5	76.0	79.0	88.0	90.5	89.5	75.5	27.0
	95 (LPSMC)	2	91.5	90.0	91.0	93.0	95.0	93.0	78.0	17.5
	95 (MLWG)	2	85.0	83.5	89.0	92.5	96.0	94.0	82.0	42.5

Note: Results for the MLWG sampler with chain length 10 000 and burn-in 3000 are also shown.

**TABLE 4** Empirical computation times for LPSMC and MLWG (with chain length 500) for different combinations of sample size  $n$  and B-spline dimension  $K$

		$n = 100$	$n = 200$	$n = 500$	$n = 800$	$n = 1000$
LPSMC	$K = 15$	0.289	0.333	0.572	0.777	0.970
	$K = 20$	0.466	0.549	0.889	1.238	1.517
	$K = 30$	1.042	1.110	1.848	3.057	4.164
	$K = 50$	2.214	2.614	3.733	5.041	6.294
MLWG	$K = 15$	1.561	1.808	2.514	3.250	3.801
	$K = 20$	1.943	2.157	2.978	3.835	4.486
	$K = 30$	2.818	3.065	4.588	5.997	7.422
	$K = 50$	5.311	5.932	7.572	9.417	10.982

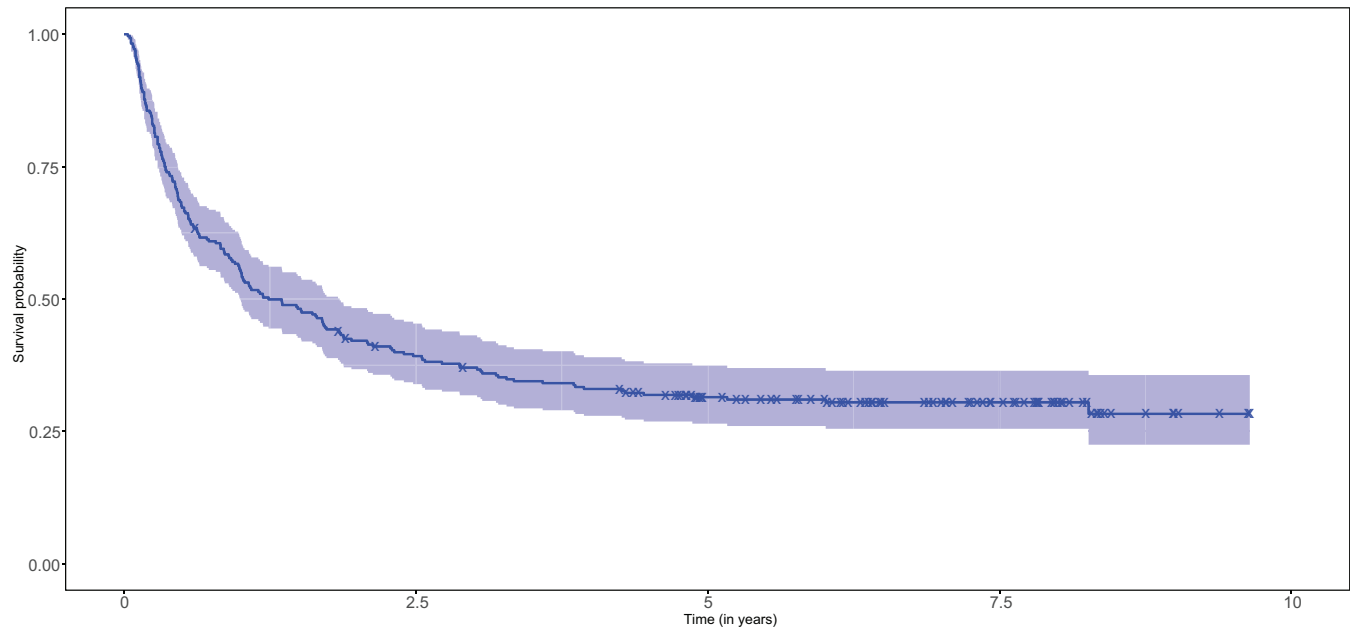
Note: Results correspond to average wall clock times (elapsed real times) in seconds over  $S = 10$  replications.

largest follow-up time. The Kaplan-Meier estimate of the data given in Figure 5 emphasizes the existence of a plateau and motivates the use of a mixture cure model.

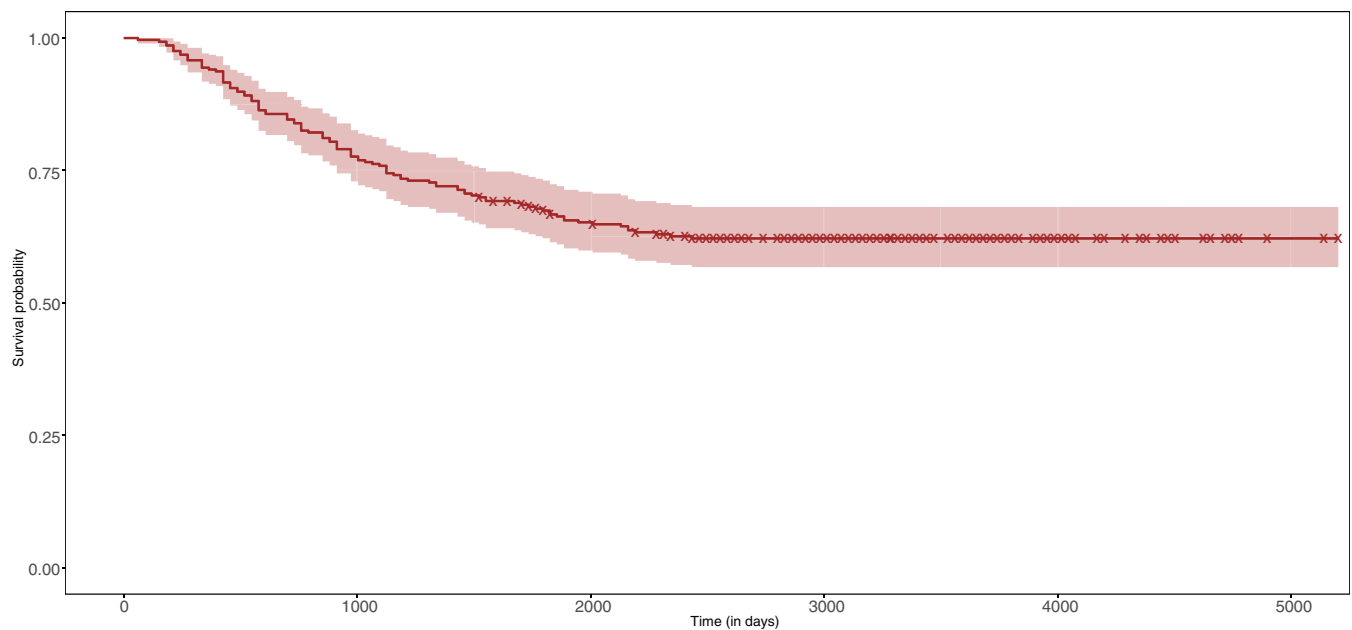
The estimated latent variables with LPSMC are reported in Table 6. We see that *AGE* and *ER* have no significant effect on the probability of being uncured (incidence) at the 10% level of significance. However, the latter two covariates significantly affect the survival of the uncured.

For instance, a subject that is susceptible to experience metastasis with  $ER = 1$  has a risk of experiencing the event that is  $1/\exp(-1.047) \approx 2.85$  times smaller than the risk with  $ER = 0$ . In Figure 6, the estimated survival for the susceptible subjects (and their associated 95% credible interval) is shown for two age categories (30 and 50 years) with  $ER = 1$ . The same data analysis is conducted with the MLWG sampler (see supplementary materials) using a chain of length 20 000 and a burn-in of 10 000. The trace plots of the regression coefficients (in the incidence and latency part respectively) and the roughness penalty parameter are provided together with the Geweke diagnostics.<sup>55</sup>





**FIGURE 4** Kaplan-Meier estimated survival for the e1684 ECOG dataset. A cross indicates a censored observation



**FIGURE 5** Kaplan-Meier estimated survival for the breast cancer data. A cross indicates a censored observation

### 5.3 | ZNA COVID-19 data

In a third application, we use LPSMC to investigate the impact of age on survival of COVID-19 patients. The data comes from the Ziekenhuis Netwerk Antwerpen (ZNA, Belgium), a network of hospitals in the province of Antwerp. The considered dataset has  $n = 3258$  patients entering hospitals between March 2020 and April 2021. The follow-up time is defined as the total number of days spent in hospital and receiving COVID-19 care. The outcome of interest is in-hospital death due to COVID-19 as indicated by a binary variable (1 = Dead; 0 = Alive). Among the 3258 patients, 461 (14.15%) experienced in-hospital death and the remaining 2797 (85.85%) are censored due to causes unrelated to the outcome of interest (uninformative censoring). The Kaplan-Meier curve is given in Figure 7 and highlights a wide plateau around 0.52,

TABLE 5 Estimation results for the e1684 dataset with LPSMC, MLWG, and smcure

Model	Parameter	Estimate	SD	CI 90%
LPSMC (incidence)	$\beta_0$ (intercept)	1.219	0.244	[0.819; 1.620]
	$\beta_1$ (SEX)	-0.061	0.284	[-0.528; 0.406]
	$\beta_2$ (TRT)	-0.567	0.281	[-1.029; -0.105]
	$\beta_3$ (AGE)	0.016	0.011	[-0.002; 0.034]
MLWG (incidence)	$\beta_0$ (intercept)	1.355	0.375	[0.884; 1.949]
	$\beta_1$ (SEX)	-0.062	0.329	[-0.596; 0.467]
	$\beta_2$ (TRT)	-0.567	0.325	[-1.092; -0.044]
	$\beta_3$ (AGE)	0.019	0.016	[-0.001; 0.044]
smcure (incidence)	$\beta_0$ (intercept)	1.365	0.329	[0.823; 1.906]
	$\beta_1$ (SEX)	-0.087	0.333	[-0.634; 0.461]
	$\beta_2$ (TRT)	-0.588	0.343	[-1.152; -0.025]
	$\beta_3$ (AGE)	0.020	0.016	[-0.006; 0.047]
LPSMC (latency)	$\gamma_1$ (SEX)	0.092	0.170	[-0.188; 0.371]
	$\gamma_2$ (TRT)	-0.137	0.169	[-0.415; 0.142]
	$\gamma_3$ (AGE)	-0.007	0.006	[-0.016; 0.003]
MLWG (latency)	$\gamma_1$ (SEX)	0.058	0.183	[-0.255; 0.353]
	$\gamma_2$ (TRT)	-0.170	0.188	[-0.484; 0.135]
	$\gamma_3$ (AGE)	-0.007	0.006	[-0.017; 0.004]
smcure (latency)	$\gamma_1$ (SEX)	0.099	0.175	[-0.189; 0.388]
	$\gamma_2$ (TRT)	-0.154	0.177	[-0.444; 0.137]
	$\gamma_3$ (AGE)	-0.008	0.007	[-0.019; 0.004]

Note: The first column indicates the model part, the second and third columns the parameter and its estimate. The fourth column is the (posterior) SD of the estimate and the last column the 90% confidence/credible interval.

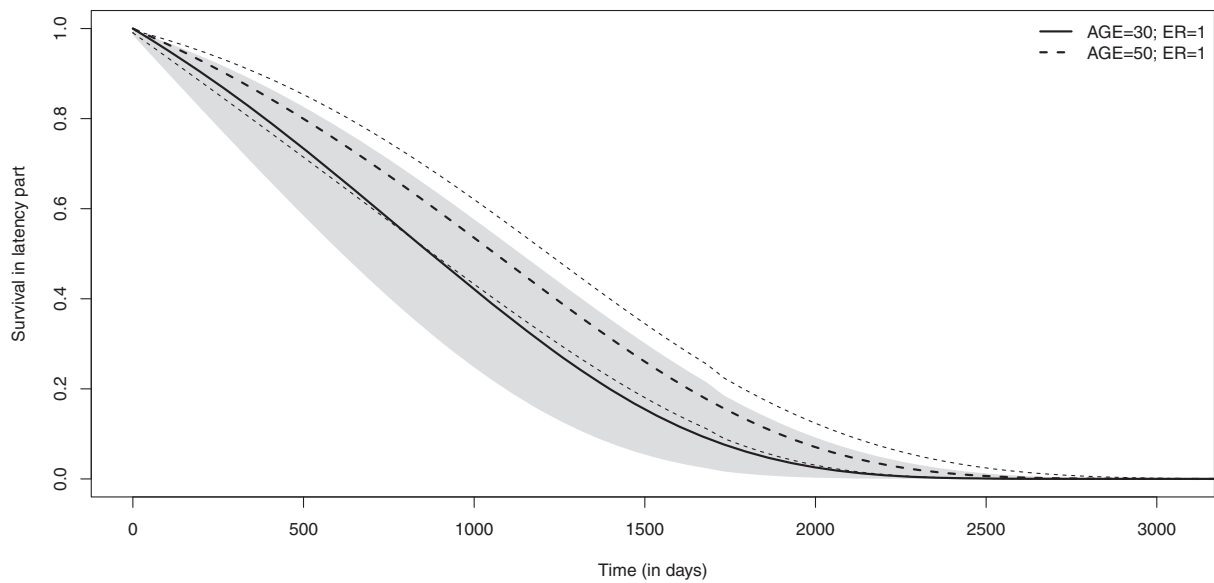
TABLE 6 Estimation results for the breast cancer data with LPSMC

Model	Parameter	Estimate	SD	CI 90%
(Incidence)	$\beta_0$ (intercept)	-0.013	0.576	[-0.959; 0.934]
	$\beta_1$ (AGE)	-0.012	0.010	[-0.028; 0.005]
	$\beta_2$ (ER)	0.181	0.281	[-0.280; 0.643]
(Latency)	$\gamma_1$ (AGE)	-0.016	0.008	[-0.030; -0.003]
	$\gamma_2$ (ER)	-1.047	0.245	[-1.450; -0.645]

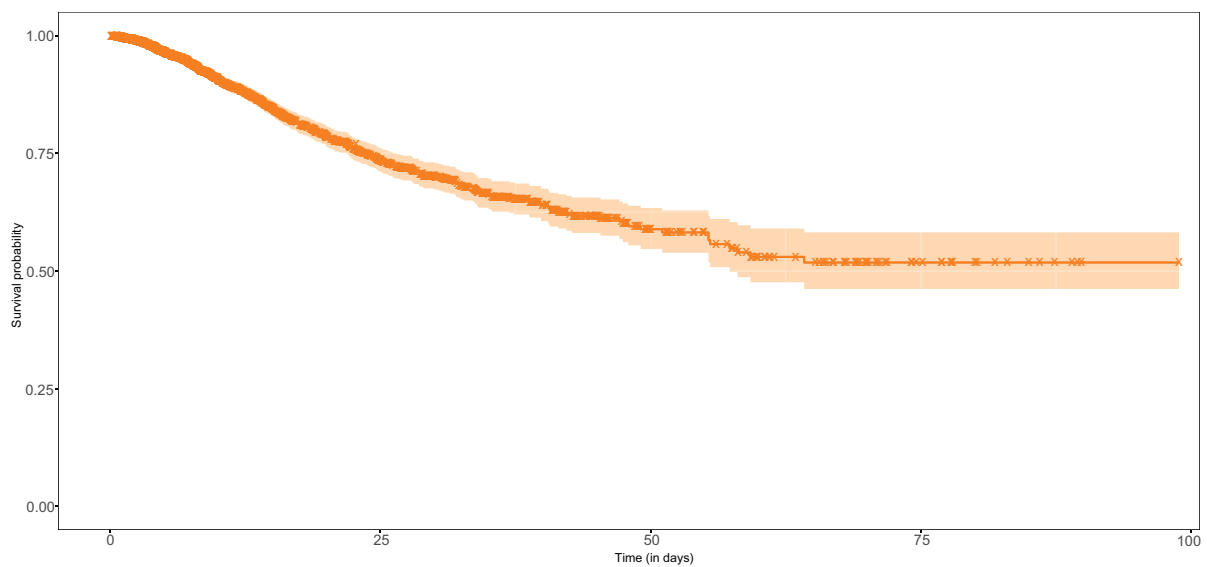
Note: The first column indicates the model part, the second and third columns the parameter and its posterior (mean) estimate. The fourth column is the posterior SD of the estimate and the last column the 90% credible interval.

suggesting the existence of a cure fraction and hence motivating a cure model analysis with LPSMC. The age of patients is taken as a covariate both in the incidence and latency part of the model. The mean age is 66 years and the youngest, respectively oldest patient is 1 and 103 year(s) old. For the LPSMC approach, we use  $K = 15$  B-splines between 0 and the largest follow-up time ( $t_u = 98.88$ ) along with a third order penalty and  $K = 10$  B-splines for the MLWG sampler.

Table 7 summarizes the estimation results for the model parameters with LPSMC and MLWG respectively (with a chain length of 15 000 and a burn-in of size 8000). We see that AGE has a positive and significant effect in the incidence part of the model. In other words, older patients have a larger probability of being uncured and hence a smaller probability to be cured from COVID-19. The estimated cure proportion for different AGE categories is shown in Table 8. In the latency



**FIGURE 6** Estimated survival of the susceptibles for two age categories and  $ER = 1$



**FIGURE 7** Kaplan-Meier estimated survival for the ZNA data. A cross indicates a censored observation

**TABLE 7** Estimation results for the ZNA data with LPSMC and MLWG

Model	Parameter	Estimate	SD	CI 95%
LPSMC (incidence)	$\beta_0$ (intercept)	-2.608	0.791	[-4.159; -1.057]
	$\beta_1$ (AGE)	0.031	0.010	[0.011; 0.051]
MLWG (incidence)	$\beta_0$ (intercept)	-2.370	0.780	[-3.850; -0.757]
	$\beta_1$ (AGE)	0.029	0.010	[0.009; 0.049]
LPSMC (latency)	$\gamma_1$ (AGE)	0.011	0.007	[-0.004; 0.025]
MLWG (latency)	$\gamma_1$ (AGE)	0.012	0.007	[-0.001; 0.027]

*Note:* The first column indicates the model part, the second and third columns the parameter and its posterior (mean) estimate. The fourth column is the posterior SD of the estimate and the last column the 95% credible interval.

TABLE 8 Estimation of the cure proportion for different age categories and associated 95% approximate credible interval

AGE	$1 - \hat{p}(x)$	CI 95%
20	0.879	[0.682; 0.958]
30	0.843	[0.658; 0.932]
50	0.742	[0.607; 0.837]
80	0.532	[0.466; 0.592]

part, *AGE* is positive and not significant at the 5% level and thus there is no significant effect of *AGE* on the survival of uncured patients.

## 6 | CONCLUSIONS

Approximate Bayesian inference methods are an interesting alternative to classic MCMC algorithms, especially when the latter require long computation times, as can be the case for models with complex likelihoods. In survival analysis, the mixture cure model is an interesting model class that allows for the existence of a cure fraction and hence goes beyond classic proportional hazards models for which the feature of long-term survivors is absent. In this article, we propose a new approach for fast Bayesian inference in the standard mixture cure model by combining the strength of Laplace approximations to (conditional) posterior distributions of latent variables and P-splines for flexible modeling of baseline smooth quantities. The attractiveness of the LPSMC approach lies in its completely sampling-free framework, with an analytically available gradient and Hessian of the log-likelihood that makes the approach extremely fast from a computational viewpoint. This computational advantage and the relatively straightforward possibility to derive (approximate) credible intervals for functions of latent variables makes it a promising tool for fast analysis of survival data with a cure fraction. Moreover, the proposed MLWG sampler brings a further layer of added value that complements the sampling-free LPSMC approach with a fully MCMC-based strategy for inference in the mixture cure model.

A possible interesting extension for future research would be to generalize the LPSMC approach at the level of the incidence with alternative specifications to the standard logistic regression setting. For instance, LPSMC could be extended to the single-index/Cox mixture cure model of Amico et al<sup>56</sup>. Another research path is to further develop the routines of the **mixcurelps** package and work on an article exclusively dedicated to the main features of the software. Finally, one can attempt to extend Laplacian-P-splines in cure models with frailties or in the context of competing risks survival data with a cure fraction.

## ACKNOWLEDGEMENTS

This project is funded by the European Union's Research and Innovation Action under the H2020 work programme, EpiPose (Grant number 101003688). The authors wish to thank the Ziekenhuis Netwerk Antwerpen for granting access to the COVID-19 hospitalization data.

## CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

## DATA AVAILABILITY STATEMENT

Simulation results in Section 4 and the ECOG and breast cancer data applications in Section 5 can be reproduced with the **mixcurelps** package<sup>27</sup> (see supplementary materials). The ZNA COVID-19 data is not publicly available for privacy reasons.

## ORCID

Oswaldo Gressani  <https://orcid.org/0000-0003-4152-6159>

## REFERENCES

1. Peng Y, Yu B. *Cure Models: Methods, Applications, and Implementation*. Boca Raton, FL: Chapman & Hall/CRC, Taylor & Francis Group; 2020.
2. Amico M, Van Keilegom I. Cure models in survival analysis. *Annu Rev Stat Appl*. 2018;5(1):311-342.
3. Boag JW. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J R Stat Soc Ser B Stat Methodol*. 1949;11(1):15-53.
4. Berkson J, Gage RP. Survival curve for cancer patients following treatment. *J Am Stat Assoc*. 1952;47(259):501-515.
5. Haybittle JL. A two-parameter model for the survival curve of treated cancer patients. *J Am Stat Assoc*. 1965;60(309):16-26.
6. Farewell VT. The combined effect of breast cancer risk factors. *Cancer*. 1977;40(2):931-936.
7. Farewell VT. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*. 1982;38(4):1041-1046.
8. Farewell VT. A model for a binary variable with time-censored observations. *Biometrika*. 1977;64(1):43-46.
9. Ghitany ME, Maller RA, Zhou S. Exponential mixture models with long-term survivors and covariates. *J Multivar Anal*. 1994;49(2):218-241.
10. Taylor JMG. Semi-parametric estimation in failure time mixture models. *Biometrics*. 1995;51(3):899-907.
11. Cox DR. Regression models and life-tables. *J R Stat Soc Ser B Stat Methodol*. 1972;34(2):187-202.
12. Kuk AYC, Chen C-H. A mixture model combining logistic regression with proportional hazards regression. *Biometrika*. 1992;79(3):531-541.
13. Peng Y, Dear KBG. A nonparametric mixture model for cure rate estimation. *Biometrics*. 2000;56(1):237-243.
14. Greenhouse JB, Wasserman L, Bayarri MJ. A practical, robust method for Bayesian model selection: a case study in the analysis of clinical trials. *Bayesian Robustness*. IMS Lecture Notes-Monograph Series. Vol 29. USA: Institute of Mathematical Statistics; 1996:41-62.
15. Chen WC, Hill BM, Greenhouse JB, Fayos JV. Bayesian analysis of survival curves for cancer patients following treatment. In: Bernardo JM, ed. *Proceedings of the 2nd Valencia International Meeting Bayesian Statistics*. Amsterdam: North-Holland; 1985.
16. Stangl DK. Modeling Heterogeneity in Multi-Center Clinical Trials Using Bayesian Hierarchical Survival Models. Doctoral dissertation. Department of Statistics, Carnegie Mellon University; 1991.
17. Stangl DK, Greenhouse JB. Assessing placebo response using Bayesian hierarchical survival models. *Lifetime Data Anal*. 1998;4:5-28.
18. Chen M-H, Ibrahim JG, Sinha D. A new Bayesian model for survival data with a surviving fraction. *J Am Stat Assoc*. 1999;94(447):909-919.
19. Yin G, Ibrahim JG. Cure rate models: a unified approach. *Can J Stat*. 2005;33(4):559-570.
20. Bremhorst V, Lambert P. Flexible estimation in cure survival models using Bayesian P-splines. *Comput Stat Data Anal*. 2016;93:270-284.
21. Gressani O, Lambert P. Fast Bayesian inference using Laplace approximations in a flexible promotion time cure model based on P-splines. *Comput Stat Data Anal*. 2018;124:151-167.
22. Yu B, Tiwari RC. A Bayesian approach to mixture cure models with spatial frailties for population-based cancer relative survival data. *Can J Stat*. 2012;40(1):40-54.
23. Martinez EZ, Achcar JA, Jácome AAA, Santos JS. Mixture and non-mixture cure fraction models based on the generalized modified Weibull distribution with an application to gastric cancer data. *Comput Methods Programs Biomed*. 2013;112(3):343-355.
24. Lázaro E, Armero C, Gómez-Rubio V. Approximate Bayesian inference for mixture cure models. *TEST*. 2020;29:750-767.
25. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J R Stat Soc Ser B Stat Methodol*. 2009;71(2):319-392.
26. Gómez-Rubio V. Mixture model fitting using conditional models and modal Gibbs sampling; 2017. <https://arxiv.org/abs/1712.09566>
27. Gressani O. A package for approximate Bayesian inference in mixture cure models with Laplacian-P-splines (version 1.1.2); 2021. <https://github.com/oswaldogressani/mixcurelps>
28. Eddelbuettel D, François R. Rcpp: seamless R and C++ integration. *J Stat Softw*. 2011;40(8):1-18.
29. Whittemore AS, Keller JB. Survival estimation using splines. *Biometrics*. 1986;42(3):495-506.
30. Rosenberg PS. Hazard function estimation using B-splines. *Biometrics*. 1995;51(3):874-887.
31. Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. *Stat Sci*. 1996;11(2):89-121.
32. Lang S, Brezger A. Bayesian P-splines. *J Comput Graph Stat*. 2004;13(1):183-212.
33. Lambert P. Smooth semiparametric and nonparametric Bayesian estimation of bivariate densities from bivariate histogram data. *Comput Stat Data Anal*. 2011;55(1):429-445.
34. Eilers PHC, Marx BD. *Practical Smoothing: The Joys of P-splines*. Cambridge, MA: Cambridge University Press; 2021.
35. Alston CL, Mengersen KL, Pettitt AN. *Case Studies in Bayesian Statistical Modelling and Analysis*. Hoboken, NJ: Wiley; 2012.
36. Jullion A, Lambert P. Robust specification of the roughness penalty prior distribution in spatially adaptive Bayesian P-splines models. *Comput Stat Data Anal*. 2007;51(5):2542-2558.
37. Sy JP, Taylor JMG. Estimation in a Cox proportional hazards cure model. *Biometrics*. 2000;56(1):227-236.
38. Tierney L, Kadane JB. Accurate approximations for posterior moments and marginal densities. *J Am Stat Assoc*. 1986;81(393):82-86.
39. Yoon JW, Wilson SP. Inference for latent variable models with many hyperparameters. Proceedings of the 58th World Statistical Congress of the International Statistical Institute; 2011; Dublin.
40. Gressani O, Lambert P. Laplace approximations for fast Bayesian inference in generalized additive models based on P-splines. *Comput Stat Data Anal*. 2021;154:107088.
41. Gómez-Rubio V, Rue H. Markov chain Monte Carlo with the integrated nested laplace approximation. *Stat Comput*. 2018;28:1033-1051.

42. Lambert P, Eilers PHC. Bayesian proportional hazards model with time-varying regression coefficients: a penalized Poisson regression approach. *Stat Med*. 2005;24(24):3977-3989.
43. Lambert P, Eilers PHC. Bayesian density estimation from grouped continuous data. *Comput Stat Data Anal*. 2009;53(4):1388-1399.
44. Roberts GO, Tweedie RL. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*. 1996;2(4):341-363.
45. Haario H, Saksman E, Tamminen J. An adaptive metropolis algorithm. *Bernoulli*. 2001;7(2):223-242.
46. Atchadé YF, Rosenthal JS. On adaptive Markov chain Monte Carlo algorithms. *Bernoulli*. 2005;11(5):815-828.
47. Roberts GO, Rosenthal JS. Optimal scaling of discrete approximations to Langevin diffusions. *J R Stat Soc Ser B Stat Methodol*. 1998;60(1):255-268.
48. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457-481.
49. Cai C, Zou Y, Peng Y, Zhang J. smcure: an R-package for estimating semiparametric mixture cure models. *Comput Methods Programs Biomed*. 2012;108(3):1255-1260.
50. Corbière F, Joly P. A SAS macro for parametric and semiparametric mixture cure models. *Comput Methods Programs Biomed*. 2007;85(2):173-180.
51. Legrand C, Bertrand A. *Cure Models in Cancer Clinical Trials*. Boca Raton, FL: Chapman & Hall/CRC, Taylor & Francis Group; 2019.
52. Plummer M. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing; Vol. 124, 2003; Vienna, Austria.
53. Wang Y, Klijn JGM, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet*. 2005;365(9460):671-679.
54. Schroeder M, Haibe-Kains B, Culhane A, Sotiriou C, Bontempi G, Quackenbush J. breastCancerVDX: gene expression datasets published by Wang et al. [2005] and Minn et al. [2007] (VDX); 2011. R package version 1.18.0.
55. Geweke J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM, eds. *Bayesian Statistics*. Vol 4. Oxford, UK: Oxford University Press; 1992:169-193.
56. Amico M, Van Keilegom I, Legrand C. The single-index/Cox mixture cure model. *Biometrics*. 2019;75(2):452-462.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Gressani O, Faes C, Hens N. Laplacian-P-splines for Bayesian inference in the mixture cure model. *Statistics in Medicine*. 2022;41(14):2602-2626. doi: 10.1002/sim.9373

## APPENDIX A. LAPLACE APPROXIMATION TO THE CONDITIONAL POSTERIOR OF THE LATENT VECTOR

Recall that the contribution of the  $i$ th unit to the log-likelihood is given by:

$$g_i(\xi) = \tau_i \left( \log p(\mathbf{x}_i) + \mathbf{z}_i^\top \boldsymbol{\gamma} + \boldsymbol{\theta}^\top b(t_i) - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \sum_{j=1}^{j(t_i)} \exp(\boldsymbol{\theta}^\top b(s_j)) \Delta \right) + (1 - \tau_i) \log \left( 1 - p(\mathbf{x}_i) + p(\mathbf{x}_i) \exp \left( - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \sum_{j=1}^{j(t_i)} \exp(\boldsymbol{\theta}^\top b(s_j)) \Delta \right) \right),$$

and using the definition of the population survival function, one has:

$$g_i(\xi) = \tau_i \left( \log p(\mathbf{x}_i) + \mathbf{z}_i^\top \boldsymbol{\gamma} + \boldsymbol{\theta}^\top b(t_i) - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \sum_{j=1}^{j(t_i)} \exp(\boldsymbol{\theta}^\top b(s_j)) \Delta \right) + (1 - \tau_i) \log (S_p(t_i | \mathbf{x}_i, \mathbf{z}_i)).$$



A second-order Taylor expansion of  $g_i(\xi)$  around an arbitrary point  $\xi^{(0)}$  is given by:

$$\begin{aligned}
 g_i(\xi) &\approx g_i(\xi^{(0)}) + (\xi - \xi^{(0)})^\top \nabla g_i(\xi)|_{\xi=\xi^{(0)}} + \frac{1}{2}(\xi - \xi^{(0)})^\top \nabla^2 g_i(\xi)|_{\xi=\xi^{(0)}} (\xi - \xi^{(0)}) \\
 &\approx c + \xi^\top \left( \nabla g_i(\xi)|_{\xi=\xi^{(0)}} - \nabla^2 g_i(\xi)|_{\xi=\xi^{(0)}} \xi^{(0)} \right) + \frac{1}{2} \xi^\top \nabla^2 g_i(\xi)|_{\xi=\xi^{(0)}} \xi,
 \end{aligned} \tag{A1}$$

where  $c$  is a constant that does not depend on  $\xi$ . The gradient  $\nabla g_i(\xi)|_{\xi=\xi^{(0)}}$  and Hessian  $\nabla^2 g_i(\xi)|_{\xi=\xi^{(0)}}$  are given by:

$$\nabla g_i(\xi)|_{\xi=\xi^{(0)}} = \begin{bmatrix} \frac{\partial}{\partial \theta_1} g_i(\xi) \\ \vdots \\ \frac{\partial}{\partial \theta_K} g_i(\xi) \\ \frac{\partial}{\partial \beta_0} g_i(\xi) \\ \vdots \\ \frac{\partial}{\partial \beta_p} g_i(\xi) \\ \frac{\partial}{\partial \gamma_1} g_i(\xi) \\ \vdots \\ \frac{\partial}{\partial \gamma_q} g_i(\xi) \end{bmatrix}_{\xi=\xi^{(0)}} \quad \nabla^2 g_i(\xi)|_{\xi=\xi^{(0)}} = \begin{bmatrix} \underbrace{\frac{\partial^2}{\partial \theta \partial \theta^\top} g_i(\xi)}_{K \times K} & \underbrace{\frac{\partial^2}{\partial \theta \partial \beta^\top} g_i(\xi)}_{K \times (p+1)} & \underbrace{\frac{\partial^2}{\partial \theta \partial \gamma^\top} g_i(\xi)}_{K \times q} \\ \underbrace{\frac{\partial^2}{\partial \beta \partial \theta^\top} g_i(\xi)}_{(p+1) \times K} & \underbrace{\frac{\partial^2}{\partial \beta \partial \beta^\top} g_i(\xi)}_{(p+1) \times (p+1)} & \underbrace{\frac{\partial^2}{\partial \beta \partial \gamma^\top} g_i(\xi)}_{(p+1) \times q} \\ \underbrace{\frac{\partial^2}{\partial \gamma \partial \theta^\top} g_i(\xi)}_{q \times K} & \underbrace{\frac{\partial^2}{\partial \gamma \partial \beta^\top} g_i(\xi)}_{q \times (p+1)} & \underbrace{\frac{\partial^2}{\partial \gamma \partial \gamma^\top} g_i(\xi)}_{q \times q} \end{bmatrix}_{\xi=\xi^{(0)}}.$$

To simplify the notation, we define the following quantities:

$$\begin{aligned}
 \omega_{0i} &:= \sum_{j=1}^{j(t_i)} h_0(s_j) \Delta, \\
 \omega_{0i}^k &:= \sum_{j=1}^{j(t_i)} h_0(s_j) b_k(s_j) \Delta, \\
 \omega_{0i}^{kl} &:= \sum_{j=1}^{j(t_i)} h_0(s_j) b_k(s_j) b_l(s_j) \Delta.
 \end{aligned}$$

### Gradient

We first derive the gradient and start with the partial derivatives with respect to the spline coefficients:

$$\frac{\partial}{\partial \theta_k} g_i(\xi) = \tau_i \left( b_k(t_i) - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \sum_{j=1}^{j(t_i)} h_0(s_j) b_k(s_j) \Delta \right) + \frac{(1 - \tau_i)}{S_p(t_i | \mathbf{x}_i, \mathbf{z}_i)} \frac{\partial}{\partial \theta_k} S_p(t_i | \mathbf{x}_i, \mathbf{z}_i).$$

Note that:

$$\frac{\partial}{\partial \theta_k} S_p(t_i | \mathbf{x}_i, \mathbf{z}_i) = -p(\mathbf{x}_i) \exp(\mathbf{z}_i^\top \boldsymbol{\gamma} - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \omega_{0i}) \omega_{0i}^k.$$

It follows that:

$$\frac{\partial}{\partial \theta_k} g_i(\xi) = \tau_i (b_k(t_i) - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \omega_{0i}^k) - \frac{(1 - \tau_i)}{S_p(t_i | \mathbf{x}_i, \mathbf{z}_i)} p(\mathbf{x}_i) \exp(\mathbf{z}_i^\top \boldsymbol{\gamma} - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \omega_{0i}) \omega_{0i}^k, \quad k = 1, \dots, K.$$

To obtain the derivatives with respect to the  $\tilde{\beta}$  coefficients, let us first compute:

$$\frac{\partial}{\partial \beta_m} p(\mathbf{x}_i) = \frac{\partial}{\partial \beta_m} \frac{\exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})}{1 + \exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})}$$

$$\begin{aligned}
 &= \frac{\partial}{\partial \beta_m} \frac{1}{1 + \exp(-\beta_0 - \mathbf{x}_i^\top \boldsymbol{\beta})} \\
 &= x_{im} (1 + \exp(-\beta_0 - \mathbf{x}_i^\top \boldsymbol{\beta}))^{-2} \exp(-\beta_0 - \mathbf{x}_i^\top \boldsymbol{\beta}) \\
 &= \frac{x_{im} \exp(-\beta_0 - \mathbf{x}_i^\top \boldsymbol{\beta})}{\left(\frac{1 + \exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})}{\exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})}\right)^2} \\
 &= x_{im} \frac{\exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})}{1 + \exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})} \frac{1}{1 + \exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})} \\
 &= x_{im} p(\mathbf{x}_i) (1 - p(\mathbf{x}_i)).
 \end{aligned}$$

It follows that:

$$\begin{aligned}
 \frac{\partial}{\partial \beta_m} g_i(\boldsymbol{\xi}) &= \tau_i x_{im} (1 - p(\mathbf{x}_i)) + \frac{(1 - \tau_i)}{S_p(t_i | \mathbf{x}_i, \mathbf{z}_i)} x_{im} p(\mathbf{x}_i) (1 - p(\mathbf{x}_i)) (\exp(-\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})) \omega_{0i} - 1) \\
 &\text{for } m = 0, \dots, p \text{ with } x_{i0} = 1.
 \end{aligned}$$

Derivatives with respect to the  $\boldsymbol{\gamma}$  coefficients are:

$$\frac{\partial}{\partial \gamma_s} g_i(\boldsymbol{\xi}) = \tau_i z_{is} (1 - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \omega_{0i}) - \frac{(1 - \tau_i)}{S_p(t_i | \mathbf{x}_i, \mathbf{z}_i)} p(\mathbf{x}_i) \exp(\mathbf{z}_i^\top \boldsymbol{\gamma} - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \omega_{0i}) \omega_{0i} z_{is}, \quad s = 1, \dots, q.$$

**Hessian**

To compute the Hessian, we only require the main diagonal and the upper triangular parts (Blocks 12, 13, and 23 below). The lower triangular part is obtained by symmetry.

- Block 11 :  $\frac{\partial^2}{\partial \theta_k \partial \theta_l} g_i(\boldsymbol{\xi}) \quad k = 1, \dots, K \quad l = 1, \dots, K.$
- Block 12 :  $\frac{\partial^2}{\partial \theta_k \partial \beta_m} g_i(\boldsymbol{\xi}) \quad k = 1, \dots, K \quad m = 0, \dots, p.$
- Block 13 :  $\frac{\partial^2}{\partial \theta_k \partial \gamma_s} g_i(\boldsymbol{\xi}) \quad k = 1, \dots, K \quad s = 1, \dots, q.$
- Block 22 :  $\frac{\partial^2}{\partial \beta_m \partial \beta_l} g_i(\boldsymbol{\xi}) \quad l = 0, \dots, p \quad m = 0, \dots, p.$
- Block 23 :  $\frac{\partial^2}{\partial \beta_m \partial \gamma_s} g_i(\boldsymbol{\xi}) \quad m = 0, \dots, p \quad s = 1, \dots, q.$
- Block 33 :  $\frac{\partial^2}{\partial \gamma_s \partial \gamma_v} g_i(\boldsymbol{\xi}) \quad s = 1, \dots, q \quad v = 1, \dots, q.$

Block 11 Let us first define the function:

$$f_i(\boldsymbol{\xi}) := \exp(\mathbf{z}_i^\top \boldsymbol{\gamma} - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \omega_{0i}) \omega_{0i}^k.$$

It follows that:

$$\begin{aligned}
 \frac{\partial^2}{\partial \theta_k \partial \theta_l} g_i(\boldsymbol{\xi}) &= -\tau_i \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \omega_{0i}^{kl} - (1 - \tau_i) p(\mathbf{x}_i) \left[ \frac{\partial f_i(\boldsymbol{\xi})}{\partial \theta_l} S_p(t_i | \mathbf{x}_i, \mathbf{z}_i) - f_i(\boldsymbol{\xi}) \frac{\partial S_p(t_i | \mathbf{x}_i, \mathbf{z}_i)}{\partial \theta_l} \right] S_p^{-2}(t_i | \mathbf{x}_i, \mathbf{z}_i) \\
 &\quad k, l = 1, \dots, K,
 \end{aligned} \tag{A2}$$

where

$$\frac{\partial f_i(\boldsymbol{\xi})}{\partial \theta_l} = -\exp(\mathbf{z}_i^\top \boldsymbol{\gamma} - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \omega_{0i}) \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \omega_{0i}^l \omega_{0i}^k + \exp(\mathbf{z}_i^\top \boldsymbol{\gamma} - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \omega_{0i}) \omega_{0i}^{kl}$$

$$= \exp(\mathbf{z}_i^\top \boldsymbol{\gamma} - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \omega_{0i}) \omega_{0i}^{kl} - f_i(\boldsymbol{\xi}) \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \omega_{0i}^l$$

and

$$\frac{\partial}{\partial \theta_l} S_p(t_i | \mathbf{x}_i, \mathbf{z}_i) = -p(\mathbf{x}_i) \exp(\mathbf{z}_i^\top \boldsymbol{\gamma} - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \omega_{0i}) \omega_{0i}^l.$$

Block 12

$$\frac{\partial^2}{\partial \theta_k \partial \beta_m} g_i(\boldsymbol{\xi}) = -(1 - \tau_i) f_i(\boldsymbol{\xi}) \left[ \frac{\partial p(\mathbf{x}_i)}{\partial \beta_m} S_p(t_i | \mathbf{x}_i, \mathbf{z}_i) - p(\mathbf{x}_i) \frac{\partial S_p(t_i | \mathbf{x}_i, \mathbf{z}_i)}{\partial \beta_m} \right] S_p^{-2}(t_i | \mathbf{x}_i, \mathbf{z}_i)$$

$$k = 1, \dots, K, m = 0, \dots, p,$$

with

$$\frac{\partial S_p(t_i | \mathbf{x}_i, \mathbf{z}_i)}{\partial \beta_m} = x_{im} p(\mathbf{x}_i) (1 - p(\mathbf{x}_i)) (\exp(-\exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \omega_{0i}) - 1).$$

Block 13

$$\frac{\partial^2}{\partial \theta_k \partial \gamma_s} g_i(\boldsymbol{\xi}) = -\tau_i \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) z_{is} \omega_{0i}^k - (1 - \tau_i) p(\mathbf{x}_i) \left[ \frac{\partial f_i(\boldsymbol{\xi})}{\partial \gamma_s} S_p(t_i | \mathbf{x}_i, \mathbf{z}_i) - f_i(\boldsymbol{\xi}) \frac{\partial S_p(t_i | \mathbf{x}_i, \mathbf{z}_i)}{\partial \gamma_s} \right] S_p^{-2}(t_i | \mathbf{x}_i, \mathbf{z}_i)$$

$$k = 1, \dots, K, s = 1, \dots, q,$$

with

$$\frac{\partial S_p(t_i | \mathbf{x}_i, \mathbf{z}_i)}{\partial \gamma_s} = -p(\mathbf{x}_i) \exp(\mathbf{z}_i^\top \boldsymbol{\gamma} - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \omega_{0i}) z_{is} \omega_{0i}.$$

$$\frac{\partial f_i(\boldsymbol{\xi})}{\partial \gamma_s} = f_i(\boldsymbol{\xi}) (z_{is} - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) z_{is} \omega_{0i}).$$

Block 22 Define the following function:

$$\tilde{f}_i(\boldsymbol{\xi}) = p(\mathbf{x}_i) (1 - p(\mathbf{x}_i)) (\exp(-\exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \omega_{0i}) - 1).$$

The second-order derivatives are:

$$\frac{\partial^2}{\partial \beta_m \partial \beta_l} g_i(\boldsymbol{\xi}) = -\tau_i x_{im} x_{il} p(\mathbf{x}_i) (1 - p(\mathbf{x}_i))$$

$$+ (1 - \tau_i) x_{im} \left[ \frac{\partial \tilde{f}_i(\boldsymbol{\xi})}{\partial \beta_l} S_p(t_i | \mathbf{x}_i, \mathbf{z}_i) - \tilde{f}_i(\boldsymbol{\xi}) \frac{\partial S_p(t_i | \mathbf{x}_i, \mathbf{z}_i)}{\partial \beta_l} \right] S_p^{-2}(t_i | \mathbf{x}_i, \mathbf{z}_i)$$

$$m, l = 0, \dots, p,$$

with

$$\frac{\partial \tilde{f}_i(\boldsymbol{\xi})}{\partial \beta_l} = x_{il} p(\mathbf{x}_i) (1 - p(\mathbf{x}_i)) (1 - 2p(\mathbf{x}_i)) (\exp(-\exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \omega_{0i}) - 1).$$

$$\frac{\partial S_p(t_i | \mathbf{x}_i, \mathbf{z}_i)}{\partial \beta_l} = x_{il} \tilde{f}_i(\boldsymbol{\xi}).$$

Block 23 Define the following function:

$$\check{f}_i(\xi) = \exp(-\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})\omega_{0i}) - 1.$$

The second-order partial derivative is:

$$\frac{\partial^2}{\partial \beta_m \partial \gamma_s} g_i(\xi) = (1 - \tau_i)x_{im}p(\mathbf{x}_i)(1 - p(\mathbf{x}_i)) \left[ \frac{\partial \check{f}_i(\xi)}{\partial \gamma_s} S_p(t_i | \mathbf{x}_i, \mathbf{z}_i) - \check{f}_i(\xi) \frac{\partial S_p(t_i | \mathbf{x}_i, \mathbf{z}_i)}{\partial \gamma_s} \right] S_p^{-2}(t_i | \mathbf{x}_i, \mathbf{z}_i)$$

$$m = 0 \dots, p, s = 1, \dots, q,$$

with

$$\frac{\partial \check{f}_i(\xi)}{\partial \gamma_s} = -\exp(\mathbf{z}_i^\top \boldsymbol{\gamma} - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma})\omega_{0i})z_{is}\omega_{0i}.$$

Block 33 Define the following function:

$$\ddot{f}_i(\xi) = \exp(\mathbf{z}_i^\top \boldsymbol{\gamma} - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma})\omega_{0i}).$$

The second-order partial derivative is:

$$\frac{\partial^2}{\partial \gamma_s \partial \gamma_v} g_i(\xi) = -\tau_i z_{is} z_{iv} \exp(\mathbf{z}_i^\top \boldsymbol{\gamma})\omega_{0i}$$

$$- (1 - \tau_i)p(\mathbf{x}_i)\omega_{0i}z_{is} \left[ \frac{\partial \ddot{f}_i(\xi)}{\partial \gamma_v} S_p(t_i | \mathbf{x}_i, \mathbf{z}_i) - \ddot{f}_i(\xi) \frac{\partial S_p(t_i | \mathbf{x}_i, \mathbf{z}_i)}{\partial \gamma_v} \right] S_p^{-2}(t_i | \mathbf{x}_i, \mathbf{z}_i)$$

$$s, v = 1, \dots, q,$$

with

$$\frac{\partial \ddot{f}_i(\xi)}{\partial \gamma_v} = \ddot{f}_i(\xi)z_{iv}(1 - \exp(\mathbf{z}_i^\top \boldsymbol{\gamma})\omega_{0i}).$$

Laplace approximation Define the short notation  $\sum_{i=1}^n \nabla g_i(\xi)|_{\xi=\xi^{(0)}} := \nabla g_{\xi^{(0)}}$  and  $\sum_{i=1}^n \nabla^2 g_i(\xi)|_{\xi=\xi^{(0)}} := \nabla^2 g_{\xi^{(0)}}$  and use (A1) to write the second-order Taylor expansion of the log-likelihood (omitting constant  $c$ ) as:

$$\ell(\xi; D) \approx \sum_{i=1}^n g_i(\xi)$$

$$\approx \xi^\top (\nabla g_{\xi^{(0)}} - \nabla^2 g_{\xi^{(0)}} \xi^{(0)}) + \frac{1}{2} \xi^\top \nabla^2 g_{\xi^{(0)}} \xi. \tag{A3}$$

Plugging the approximated log-likelihood (A3) into the conditional posterior of the latent vector yields:

$$\tilde{p}_G(\xi | \lambda, D) \propto \exp\left(-\frac{1}{2} \xi^\top (Q_\xi(\lambda) - \nabla^2 g_{\xi^{(0)}}) \xi + \xi^\top (\nabla g_{\xi^{(0)}} - \nabla^2 g_{\xi^{(0)}} \xi^{(0)})\right). \tag{A4}$$

The logarithm of (A4) is thus:

$$\log \tilde{p}_G(\xi | \lambda, D) \doteq -\frac{1}{2} \xi^\top (Q_\xi(\lambda) - \nabla^2 g_{\xi^{(0)}}) \xi + \xi^\top (\nabla g_{\xi^{(0)}} - \nabla^2 g_{\xi^{(0)}} \xi^{(0)}). \tag{A5}$$

Taking the gradient of (A5) and equating to the zero vector yields:

$$\xi^{(1)}(\lambda) = (Q_\xi(\lambda) - \nabla^2 g_{\xi^{(0)}})^{-1} (\nabla g_{\xi^{(0)}} - \nabla^2 g_{\xi^{(0)}} \xi^{(0)}).$$

The inverse of the negative Hessian of (A5) yields:

$$\Sigma_{\xi}^{(1)}(\lambda) = (Q_{\xi}(\lambda) - \nabla^2 g_{\xi^{(0)}})^{-1}.$$

Finally, the Laplace approximation to the conditional posterior latent vector is written as:

$$\tilde{p}_G(\xi|\lambda, \mathcal{D}) = \mathcal{N}_{\dim(\xi)}(\xi^{(1)}(\lambda), \Sigma_{\xi}^{(1)}(\lambda)).$$

One can iterate this in a Newton–Raphson type algorithm to obtain a Laplace approximation to the conditional latent vector  $\tilde{p}_G(\xi|\lambda, \mathcal{D}) = \mathcal{N}_{\dim(\xi)}(\xi^*(\lambda), \Sigma_{\xi}^*(\lambda))$ , where  $\xi^*(\lambda)$  and  $\Sigma_{\xi}^*(\lambda)$  denotes the mean vector and covariance matrix respectively towards which the Newton–Raphson algorithm has converged for a given value of  $\lambda$ .