ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistics in Society
Series A

*J. R. Statist. Soc.* A (2017)

# A latent promotion time cure rate model using dependent tail-free mixtures

Li Li and Ji-Hyun Lee

*University of New Mexico, Albuquerque, USA*

[Received November 2015. Final revision June 2016]

**Summary.** The paper extends the latent promotion time cure rate marker model of Kim, Xi and Chen for right-censored survival data. Instead of modelling the cure rate parameter as a deterministic function of risk factors, they assumed that the cure rate parameter of a targeted population is distributed over a number of ordinal levels according to the probabilities governed by the risk factors. We propose to use a mixture of linear dependent tail-free processes as the prior for the distribution of the cure rate parameter, resulting in a latent promotion time cure rate model. This approach provides an immediate answer to perhaps one of the most pressing questions 'what is the probability that a targeted population has high proportions (e.g. greater than 70%) of being cured?'. The approach proposed can accommodate a rich class of distributions for the cure rate parameter, while centred at gamma densities. The algorithms that are developed in this work allow the fitting of latent promotion time cure rate models with several survival models for metastatic tumour cells.

*Keywords*: Breast cancer data; Cure rate models; Latent cure rate; Right-censored survival data; Tail-free process

## 1. Introduction

Cure rate models have been increasingly used for modelling time-to-event data related to various types of cancers where a significant proportion of patients are freed of their diseases, i.e. 'cured' and thus are long-term survivors. One popular model for this type of data is the mixture cure rate model of Berkson and Gage (1952) which assumes that the survivor function for the entire population, denoted by $S(t)$, is $S(t) = \pi + (1 - \pi)\, S^*(t)$, where $S^*(t)$ is the survivor function for the non-cured group and $\pi$ is the fraction of the population being cured. Typically $\pi$ is related to covariates through a standard binomial regression. When the regression parameters are given improper priors, however, the joint posterior distribution of the parameters of the mixture model is improper and one resulting issue is that convergence of the Markov chain Monte Carlo (MCMC) algorithm is not assured. An alternative cure rate model proposed by Yakovlev *et al.* (1993) avoids this drawback and was later extensively studied by Chen *et al.* (1999) and Tsodikov *et al.* (2003). We refer to this model as the promotion time cure rate model, which can be derived as follows. Suppose that an individual in the population has $N$ metastatic tumour cells left at the beginning of the observation time and assume that $N$ follows a Poisson($\theta$) distribution. In addition, assume that the time to produce detectable metastatic disease for each tumour cell has a common distribution $F_T(t) = 1 - S_T(t)$. Following these biological assumptions, the time to observation of a relapse or death from the disease has the survival function

*Address for correspondence*: Li Li, Department of Mathematics and Statistics, University of New Mexico, 311 Terrace NE, Albuquerque, NM 87131, USA.
E-mail: llis@unm.edu

 0964–1998/17/180000

$$S(t|\theta) = P(T > t|\theta) = \exp\{-\theta F_T(t)\}. \quad (1)$$

The proportion of being cured in this model is $S(\infty|\theta) = \exp(-\theta)$ and the survival function for the 'non-cured' group is $S^*(t) = [\exp\{-\theta F_T(t)\} - \exp(-\theta)]/\{1 - \exp(-\theta)\}$. We refer to $\theta$ as the cure rate parameter and $\exp(-\theta)$ as the cure rate. When risk factors are present, the cure rate parameter $\theta$ has often been modelled by a deterministic function of the risk factors, e.g. $\theta = \exp(\mathbf{z}'\psi)$ where $\mathbf{z}$ denotes the related risk factors. Ibrahim *et al.* (2005) gave detailed technical notes for fitting such models through Bayesian approaches. When a hierarchy of different populations is present, for example, patients are treated by medical care strategies due to their clinical characteristics of the disease, the promotion time cure rate model has been extended to incorporate random effects; see Lopes and Bolfarine (2012) and Gallardo *et al.* (2016). The drawback of using a deterministic function of risk factors and random effects is that it assumes a constant cure rate for each risk group and hence does not account for the heterogeneities due to unobserved factors in the cure rates. Moreover, it is often of interest to investigate the predictive probabilities that are associated with each ordinal level of the cure rate for a specific group. Motivated by the drawbacks of using deterministic functions for the cure rate, Kim *et al.* (2009) proposed a latent cure rate marker model and applied the model to a prostate cancer patient data set. The cure rate in their model is assumed to follow a multinomial distribution where the probability for each ordinal level is linked to risk factors through logistic regressions. In this paper, we do not assume a fixed number of ordinal levels for the cure rate but consider a rich class of factor-dependent continuous distributions for the cure rate. By doing this, we do not need to specify or estimate the number of levels; instead, we can estimate the density of a cure rate, given the risk factors. To our knowledge, no work has been done for this case and we call the corresponding model the 'latent promotion time cure rate (LPTCR) model'. As it is difficult to specify a class of arbitrary factor-dependent distributions, we propose to use a Bayesian non-parametric prior, the mixture of linear-dependent tail-free processes (LDTFPs) prior (Jara and Hanson, 2011), for the distribution of the cure rate. For the distribution of time to produce detectable metastatic disease $F_T(\cdot)$ for each tumour cell, the method proposed can accommodate several survival models, including the proportional hazard (PH) model, the accelerated failure time (AFT) model and the proportional odds (PO) model. Denote the baseline distribution of the assumed survival model as $F_0(\cdot)$ on which we place a mixture of tail-free processes prior. Prominent success has been achieved in applications by Bayesian non-parametric priors, because of their flexibility in modelling unknown distributions; some other examples include the Dirichlet process (Ferguson, 1973), Polya tree priors (Lavine, 1992) and Dirichlet process mixtures (Escobar and West, 1995). Two main advantages of using tail-free mixtures for our model are that, firstly, tail-free mixture models are easily used to generalize existing parametric models in a robust fashion, e.g. tail-free processes have been used for PH modelling (Hanson, 2006), PO modelling (Hanson and Yang, 2007) and AFT modelling (Hanson and Johnson, 2002). This is practically appealing for us considering our complicated survival model. Secondly, the simplicity of the density expression for dependent tail-free mixtures also allows us to obtain the marginal survival distribution in a simple way by integrating out the covariate-dependent distribution for the cure rate. This also provides efficient MCMC updating and a unified approach to the PH, AFT and PO survival models. For the LPTCR model with dependent tail-free mixtures that we outlined above, we develop an MCMC algorithm for estimating the parameters. We demonstrate its usage in simulations and real data analysis. Compared with the algorithms in Kim *et al.* (2009) which were specifically developed for the PH model, our algorithms can fit all of PH, AFT and PO models.

The rest of this paper is constructed as follows. Section 2 describes the proposed model in

detail and outlines the LDTFP prior. Section 3 describes the Bayesian MCMC algorithm for fitting the model. Model comparison criteria are presented in Section 4.1. The model and its performance will be examined by simulations in Section 4.2 and data analyses in Section 5. We conclude the paper in Section 6.

## 2. Model development

### 2.1. General model

Denote $t_i$ as the observed event time for the $i$th subject and let $\delta_i$ be its right censoring indicator, i.e. $\delta_i = 1$ if $t_i$ is an observed failure time and $\delta_i = 0$ if it is right censored. Let $F_T(\cdot|\mathbf{x}_i)$ be a cumulative distribution function conditioning on a covariate vector $\mathbf{x}_i$. Let $\theta_i$ be the cure rate for the $i$th subject, and $\mathbf{z}_i$ be a covariate vector related to $\theta_i$. Given $\theta_i$ and $F_T(\cdot|\mathbf{x}_i)$, the survival function for the $i$th subject has the form

$$S_i(t|\mathbf{x}_i, \theta_i) = P(T_i > t|\mathbf{x}_i, \mathbf{z}_i) = \exp\{-\theta_i F_T(t_i|\mathbf{x}_i)\}. \tag{2}$$

The cure rate parameter $\theta_i$ can be assumed to be a deterministic scalar related to $\mathbf{z}_i$. For example, in Ibrahim *et al.* (2005), $\theta_i = \exp(\mathbf{z}_i'\psi)$; however, we model it as a continuous random variable with its distribution changing with $\mathbf{z}_i$. Let the density of $\theta$ given $\mathbf{z}_i$ be $f_\theta(\theta|\mathbf{z}_i)$ and its cumulative distribution function be $F_\theta(\theta|\mathbf{z}_i)$. Instead of focusing on one specific survival model for $F_T(t_i|\mathbf{x}_i)$ (Li and Taylor, 2002; Tsodikov *et al.*, 2003; Gu *et al.*, 2011), we consider all the following commonly used survival models: PH,

$$h_T(t_i) = h_0(t_i)\exp(\mathbf{x}_i'\gamma), \tag{3}$$

AFT,

$$h_T(t_i) = h_0\{t_i \exp(\mathbf{x}_i'\gamma)\}\exp(\mathbf{x}_i'\gamma), \tag{4}$$

PO,

$$\frac{F_T(t_i)}{1 - F_T(t_i)} = \frac{F_0(t_i)}{1 - F_0(t_i)}\exp(\mathbf{x}_i'\gamma), \tag{5}$$

where $h_0(\cdot)$ is the baseline hazard function and $F_0(\cdot)$ is the corresponding baseline cumulative distribution function. Given the distributions $F_\theta(\cdot|\mathbf{z}_i)$ and $F_T(\cdot|\mathbf{x}_i)$, the marginal survival function $S_i(t_i|\mathbf{x}_i, \mathbf{z}_i)$ is computed:

$$S_i(t|\mathbf{x}_i, \mathbf{z}_i) = \int_0^\infty \exp\{-\theta F_T(t_i|\mathbf{x}_i)\} f_\theta(\theta|\mathbf{z}_i)\mathrm{d}\theta \tag{6}$$

and the marginal density function $f_i(t|\mathbf{x}_i, \mathbf{z}_i)$ is

$$f_i(t|\mathbf{x}_i, \mathbf{z}_i) = \int_0^\infty \theta f_T(t|\mathbf{x}_i)\exp\{-\theta F_T(t_i|\mathbf{x}_i)\} f_\theta(\theta|\mathbf{z}_i)\mathrm{d}\theta. \tag{7}$$

In equation (7), $f_T(\cdot|\mathbf{x}_i)$ is the density function of $F_T(\cdot|\mathbf{x}_i)$. The full likelihood function given data $\mathcal{D} = \{t_i, \delta_i, \mathbf{x}_i, \mathbf{z}_i\}_{i=1}^n$ under the LPTCR model is

$$L = \prod_{i=1}^n f_i(t_i|\mathbf{x}_i, \mathbf{z}_i)^{\delta_i} S_i(t_i|\mathbf{x}_i, \mathbf{z}_i)^{1-\delta_i}. \tag{8}$$

We are interested in inference on the unknown distributions $f_\theta(\cdot|\mathbf{z}_i)$ and $F_T(\cdot|\mathbf{x}_i)$. As mentioned in Section 1, we assign an LDTFP prior on the distribution $F_\theta(\cdot|\mathbf{z}_i)$ and another independent mixture of tail-free processes prior on $F_0(\cdot)$. The non-parametric priors allow both

distributions to have tremendous flexibility while being centred at two parametric distributions: the gamma distribution for $F_\theta(\cdot|\mathbf{z}_i)$ and Weibull distribution for $F_0(\cdot)$. The centring feature, especially in the right-hand tails of the distributions, is appealing, as Tsodikov *et al.* (2003) suggested that, since typically few subjects are at risk in the tail of the survival curve after sufficient follow-up, there is a need to model the right-hand tail of the survival curve carefully and to allow the model to be more parametric in the tail, while also allowing the model to be non-parametric in other parts of the curve. This suggestion is achieved naturally by the LDTFP prior.

### 2.2. A summary of the linear dependent tail-free process prior

The following description outlines an LDTFP prior for $F(\cdot|\mathbf{w}_i)$ over domain $[0, \infty)$ where $\mathbf{w}_i$ is a $p$-dimensional numeric vector based on covariates (see the last paragraph of this section for more details). Denote $G_\phi$ as a parametric cumulative distribution function. Let $\Pi = \{\Pi_1, \Pi_2, \ldots\}$ be a sequence of partitions of the positive real numbers $\mathbb{R}^+$ and let $\Pi_j = \{B(\epsilon_1 \ldots \epsilon_j) : \epsilon_i \in \{0, 1\}, i = 1, \ldots, j\}$ be the partition at the $j$th level, e.g $\{B(0), B(1)\}$ at the first level, $\{B(00), B(01), B(10), B(11)\}$ at the second level, and so on. Each set in $\Pi_j$ is split into two sets in $\Pi_{j+1}$, i.e. $B(0) = B(00) \cup B(01)$. Following Lavine (1992), the sets are given by quantiles of the centring family: $B(\epsilon_1 \ldots \epsilon_j)$ is the interval $[G_\phi^{-1}(m/2^j), G_\phi^{-1}\{(m+1)/2^j\}]$ where $m$ is the base 10 representation of the binary number $\epsilon_1 \ldots \epsilon_j$. We also refer to $\Pi$ as the partition tree and $j = 1, 2, \ldots$ as the tree levels.

Define $F(A|\mathbf{w}_i)$ to be the probability of any set $A$ given the distribution $F(\cdot|\mathbf{w}_i)$. The LDTFP prior for $F(\cdot|\mathbf{w}_i)$ is constructed from the sequence of partitions $\Pi$ and associated pairwise conditional probabilities $(\pi(\epsilon_1 \ldots \epsilon_{j-1}0|\mathbf{w}_i), \pi(\epsilon_1 \ldots \epsilon_{j-1}1|\mathbf{w}_i))$, assuming that $\pi(\epsilon_1 \ldots \epsilon_{j-1}0|\mathbf{w}_i) = 1 - \pi(\epsilon_1 \ldots \epsilon_{j-1}1|\mathbf{w}_i) = F\{B(\epsilon_1 \ldots \epsilon_{j-1}0)|B(\epsilon_1 \ldots \epsilon_{j-1}), \mathbf{w}_i\}$. Let $\boldsymbol{\pi}(\epsilon|\mathbf{w}_i) = \{\pi(\epsilon_1 \ldots \epsilon_{j-1}0|\mathbf{w}_i), j = 1, 2, \ldots\}$. Further an LDTFP prior assumes that random probabilities in $\boldsymbol{\pi}(\epsilon|\mathbf{w}_i)$ are mutually independent, and random measure $F$ is related to the probabilities through the relations $F\{B_\phi(\epsilon_1 \ldots \epsilon_j)|\mathbf{w}_i\} = \Pi_{i=1}^{j} \pi(\epsilon_1 \ldots \epsilon_i|\mathbf{w}_i)$. Let $\epsilon 0$ represent $\epsilon_1 \ldots \epsilon_{j-1}0$ and assume that

$$\text{logit}\{\pi(\epsilon 0|\mathbf{w}_i)\} = \mathbf{w}_i' \boldsymbol{\nu}(\epsilon 0), \qquad \boldsymbol{\nu}(\epsilon 0) \sim N_p[\mathbf{0}, \{2n/(cj^2)\}(\mathbf{D}^{\mathrm{T}}\mathbf{D})^{-1}] \tag{9}$$

where $\mathbf{D}$ is the design matrix, i.e. $\mathbf{D} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n)^{\mathrm{T}}$, $\boldsymbol{\nu}(\epsilon 0)$ is a vector of regression parameters and $c$ is a precision parameter. The prior on $\boldsymbol{\nu}(\epsilon 0)$ is a modified version of Zellner's $g$-prior (Zellner, 1983), which takes the scale, location and correlation among the predictor variables into account to standardize and bound prior variability across representative predictor values. Lower values of $c$ in equation (9) allow mass of $F(\cdot|\mathbf{w}_i)$ to move easily from the centring distribution $G_\phi$. One common choice simply fixes $c$ at small values, e.g. $c = 1$ (Hanson, 2006).

The infinite number of levels in the partition tree $\Pi$ is usually capped by some fixed level $J$, which yields partitions up to level $J$, say $\Pi^J$. Furthermore, on partition sets $B(\epsilon_1 \ldots \epsilon_J) \in \Pi^J$ at level $J$ we assume that $F(\cdot|\mathbf{w}_i)$ follows the base measure $G_\phi$, i.e., for all measurable $A \subset B(\epsilon_1 \ldots \epsilon_J)$,

$$F\{A|B(\epsilon_1 \ldots \epsilon_J), \mathbf{w}_i\} = G_\phi(A)/G_\phi\{B(\epsilon_1 \ldots \epsilon_J)\}. \tag{10}$$

We use the notation $\text{LDTFP}^J(c, G_\phi, \mathbf{w}_i)$ for this finite LDTFP prior with cap $J$. For any $F(\cdot|\mathbf{w}_i) \sim \text{LDTFP}^J(c, G_\phi, \mathbf{w}_i)$, the survival function $S(t|\mathbf{w}_i)$ with respect to $F(\cdot|\mathbf{w}_i)$ is given by

$$S(t|\mathbf{w}_i) = p\{s(t)\}\{s(t) - 2^J G_\phi(t)\} + \sum_{j=s(t)+1}^{2^J} p(j), \tag{11}$$

where $s(t) = \lceil 2^J G_\phi(t) \rceil$ and $\lceil \cdot \rceil$ is the ceiling function. Here $p(j)$, $j = 1, \ldots, 2^J$, is defined as

$$p(j+1) = F\{B(\epsilon_1 \ldots \epsilon_J)|\mathbf{w}_i\} = \prod_{k=1}^{J} \pi(\epsilon_1 \ldots \epsilon_k|\mathbf{w}_i), \tag{12}$$

where $\epsilon_1 \ldots \epsilon_J$ is the base 2 representation of $j$. Formula (11) can be obtained from equations (10) and (12). By differentiating equation (11), the density corresponding to $F(\cdot|\mathbf{w}_i)$ is given by

$$f(t|\mathbf{w}_i) = 2^J p\{s(t)\} g_\phi(t), \tag{13}$$

where $g_\phi(\cdot)$ is the density corresponding to $G_\phi$.

On the basis of the notation that was developed above, we assign the following priors on $F_\theta(\cdot|\mathbf{z}_i)$ and $F_0(t)$ for the semiparametric LPTCR model:

$$F_\theta(\cdot|\mathbf{z}_i) \sim \text{LDTFP}^{J_\theta}(c_\theta, G_\phi, \mathbf{w}_i), \tag{14}$$

$$F_0(\cdot) \sim \text{LDTFP}^{J_T}(c_T, G_{\boldsymbol{\lambda}}, 1), \tag{15}$$

where $J_\theta$ and $J_T$ are two integers; $c_\theta$ and $c_T$ are two precision parameters and $\mathbf{w}_i$ is the vector $\mathbf{z}_i$ expanded by an additional intercept and other terms for the continuous variates. Suppose that $\mathbf{z}_i = z_i$ is a continuous covariate; Jara and Hanson (2011) compared models by using a polynomial expansion ($\mathbf{w}_i = (1, z_i, \ldots, z_i^k)$, $k = 1, 2$), a $B$-spline expansion ($\mathbf{w}_i = (1, \phi_1(z_i), \ldots, \phi_k(z_i))$) and a cosine basis expansion ($\mathbf{w}_i = (1, \kappa_1(z_i), \ldots, \kappa_k(z_i))$). We found that low order polynomial expansions fit well to our simulated data. Let $\boldsymbol{\beta}$ be the corresponding vector of regression parameters in equation (9) for $\text{LDTFP}^{J_\theta}(c_\theta, G_\phi, \mathbf{w}_i)$, i.e. $\boldsymbol{\beta}$ is the vectorized expression of the set $\{\boldsymbol{\beta}(\epsilon_1 \ldots \epsilon_j 0), j = 1, \ldots, J_\theta - 1\}$. Let $\boldsymbol{\eta}$ be the vector of intercepts for $\text{LDTFP}^{J_T}(c_T, G_{\boldsymbol{\lambda}}, 1)$, i.e. $\boldsymbol{\eta}$ is the vectorized expression of the set $\{\eta(\epsilon_1 \ldots \epsilon_j 0), j = 1, \ldots, J_T - 1\}$. Following the assumption in equation (9) for $\boldsymbol{\nu}(\epsilon 0)$, the following priors on $\boldsymbol{\beta}(\epsilon_1 \ldots \epsilon_{j-1} 0)$ and $\eta(\epsilon_1 \ldots \epsilon_{j-1} 0)$ are assumed:

$$\boldsymbol{\beta}(\epsilon_1 \ldots \epsilon_{j-1} 0) \sim N_p\{\mathbf{0}, (2n/j^2)(\mathbf{D}^{\mathrm{T}}\mathbf{D})^{-1}\}, \tag{16}$$

$$\eta(\epsilon_1 \ldots \epsilon_{j-1} 0) \sim N(0, 2/j^2), \tag{17}$$

where $\mathbf{D}$ is $(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n)^{\mathrm{T}}$.

### *2.3. Hyperparameters settings in the linear dependent tail-free process priors*

We have found that setting $c_\theta = 1$ and $c_T = 1$ provides enough flexibility while speeding up the convergence of MCMC algorithms. For a sensitivity analysis, we also implement the full conditional updating of $c_\theta$ and $c_T$ (Jara and Hanson, 2011) in our breast cancer data analysis. The full conditional distributions are included in Section 3. For the partition caps $J_T$ and $J_\theta$, a conservative rule of thumb was suggested in Jara and Hanson (2011) that $J \approx \log_2(n/N)$ where $n$ is the sample size and $N$ is typically 5–10. For our simulations and data analysis, we start from fixing $J_\theta = 3$ and $J_T = 5$. We suggest performing sensitivity analyses by comparing the goodness-of-fit statistics (see Section 4.1) for different partition caps.

Jara and Hanson (2011) suggested continuous priors on the parameters that are associated with the centring distributions, $\boldsymbol{\phi}$ and $\boldsymbol{\lambda}$, to smooth out the jumps in the densities. Let $\boldsymbol{\phi} = (\log(\phi_1), \log(\phi_2))'$ and $G_{\boldsymbol{\phi}}(\cdot)$ be the cumulative distribution function that is associated with gamma density ${\phi_2}^{\phi_1}/\Gamma(\phi_1) x^{\phi_1 - 1} \exp(-\phi_2 x)$. Let $\boldsymbol{\lambda} = (\log(\lambda_1), \log(\lambda_2))'$ and $G_{\boldsymbol{\lambda}}(t)$ be a Weibull cumulative distribution function where $G_{\boldsymbol{\lambda}}(t) = 1 - \exp\{-(t/\lambda_2)^{\lambda_1}\}$. Given a survival model for $F_T(t|\mathbf{x})$, we suggest fitting the following parametric LPTCR model to construct Gaussian priors for $\boldsymbol{\phi}$ and $\boldsymbol{\lambda}$:

$$S_i(t|\mathbf{x}_i,\theta_i)=\exp\{-\theta_i F_T(t|\mathbf{x}_i)\},$$
$$\theta_i\sim G_{\boldsymbol{\phi}}(\cdot),$$
$$F_0(t)=G_{\boldsymbol{\lambda}}(t).$$

The resulting model has the following marginal survival and density function respectively:

$$S_i(t|\mathbf{x}_i)=\left\{\frac{\phi_2}{\phi_2+F_T(t|\mathbf{x}_i)}\right\}^{\phi_1}, \tag{18}$$

$$f_i(t|\mathbf{x}_i)=\frac{\phi_1\phi_2^{\phi_1}f_T(t|\mathbf{x}_i)}{\{\phi_2+F_T(t|\mathbf{x}_i)\}^{\phi_1+1}}. \tag{19}$$

A practical procedure is as follows.

(a) Use the function `optim` in the R 'stats' package to maximize the likelihood (8) where the marginal survival and density function are in equations (18) and (19).
(b) Obtain maximum likelihood estimates $\boldsymbol{\mu}_{\boldsymbol{\phi}}$ for $\boldsymbol{\phi}$, $\boldsymbol{\mu}_{\boldsymbol{\lambda}}$ for $\boldsymbol{\lambda}$, the inverse information matrix $\mathbf{V}_{\boldsymbol{\phi}}$ associated with $\boldsymbol{\mu}_{\boldsymbol{\phi}}$ and the inverse information matrix $\mathbf{V}_{\boldsymbol{\lambda}}$ associated with $\boldsymbol{\mu}_{\boldsymbol{\lambda}}$.
(c) Place a Gaussian prior $N_2(\boldsymbol{\mu}_{\boldsymbol{\phi}},\mathbf{V}_{\boldsymbol{\phi}})$ on $\boldsymbol{\phi}$ and a Gaussian prior $N_2(\boldsymbol{\mu}_{\boldsymbol{\lambda}},\mathbf{V}_{\boldsymbol{\lambda}})$ on $\boldsymbol{\lambda}$. We can also obtain an estimate for the survival regression parameter $\boldsymbol{\gamma}$ which will be used as the MCMC starting values for the LPTCR model using linear dependent tail-free mixtures.

## 3. Markov chain Monte Carlo methods for implementing the latent promotion time cure rate model using dependent tail-free mixtures

Given the observed data $\mathcal{D}=\{(t_i,\delta_i,\mathbf{x}_i,\mathbf{z}_i)\}_{i=1}^n$ and the parameters that are related to the LPTCR model using linear dependent tail-free mixtures $\Theta=\{\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\eta},\boldsymbol{\lambda}\}$, the full likelihood is

$$L(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\eta},\boldsymbol{\lambda}|\mathcal{D})=\prod_{i=1}^n f_i(t_i|\mathbf{x}_i,\mathbf{z}_i)^{\delta_i}S_i(t_i|\mathbf{x}_i,\mathbf{z}_i)^{1-\delta_i}, \tag{20}$$

where

$$S_i(t|\mathbf{x}_i,\mathbf{z}_i)=2^{J_\theta}\sum_{l=1}^{2^{J_\theta}}p_\theta(l)\frac{\phi_2^{\phi_1}}{\Gamma(\phi_1)}\int_{G_{\boldsymbol{\phi}}^{-1}\{(l-1)/(2^{J_\theta})\}}^{G_{\boldsymbol{\phi}}^{-1}\{l/(2^{J_\theta})\}}\theta^{\phi_1-1}\exp[-\theta\{F_T(t|\mathbf{x}_i)+\phi_2\}]\,\mathrm{d}\theta, \tag{21}$$

$$f_i(t|\mathbf{x}_i,\mathbf{z}_i)=2^{J_\theta}\sum_{l=1}^{2^{J_\theta}}p_\theta(l)\frac{\phi_2^{\phi_1}}{\Gamma(\phi_1)}\int_{G_{\boldsymbol{\phi}}^{-1}\{(l-1)/(2^{J_\theta})\}}^{G_{\boldsymbol{\phi}}^{-1}\{l/(2^{J_\theta})\}}f_T(t|\mathbf{x}_i)\theta^{\phi_1}\exp[-\theta\{F_T(t|\mathbf{x}_i)+\phi_2\}]\,\mathrm{d}\theta, \tag{22}$$

and $p_\theta(1),\ldots,p_\theta(2^{J_\theta})$ are functions of $\boldsymbol{\beta}$, defined in equation (12). The joint posterior is given by

$$\pi(\Theta|\mathcal{D})\propto L(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\eta},\boldsymbol{\lambda}|\mathcal{D})\pi(\boldsymbol{\gamma})\pi(\boldsymbol{\beta})\pi(\boldsymbol{\phi})\pi(\boldsymbol{\eta})\pi(\boldsymbol{\lambda}), \tag{23}$$

where priors $\pi(\boldsymbol{\beta})$ and $\pi(\boldsymbol{\eta})$ are specified in equations (16) and (17) and priors $\pi(\boldsymbol{\phi})$ and $\pi(\boldsymbol{\lambda})$ are detailed in the last paragraph of Section 3. Suppose that the length of $\boldsymbol{\gamma}$ is $q$. We let $\pi(\boldsymbol{\gamma})$ be $N_q(\mathbf{0},4^2\mathbf{I}_q)$ which provides a weakly informative prior centred at zero. Parameters $\{\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\eta},\boldsymbol{\lambda}\}$ are updated by using random-walk Metropolis–Hastings algorithms (Tierney, 1994) based on the joint posterior (23). In each iteration of updating $\boldsymbol{\gamma}$ while the other parameters are fixed at their latest value $(\boldsymbol{\beta}^*,\boldsymbol{\phi}^*,\boldsymbol{\eta}^*,\boldsymbol{\lambda}^*)$, a new vector for $\boldsymbol{\gamma}$ is first sampled from the proposal distribution

$$\gamma' \sim N_q(\gamma^*, \mathbf{V})$$

where $\gamma^*$ is the last value for $\gamma$, and $\mathbf{V}$ is a covariance matrix tuned to obtain acceptance rates in the 20–50% range. The new value is accepted with probability

$$\frac{L(\gamma', \beta^*, \phi^*, \eta^*, \lambda^* | \mathcal{D})\pi(\gamma')}{L(\gamma^*, \beta^*, \phi^*, \eta^*, \lambda^* | \mathcal{D})\pi(\gamma^*)}.$$

The automatic tuning of $\mathbf{V}$ in Haario *et al.* (2005) is used in this paper. Specifically, let the sequence $\gamma^{(1)}, \gamma^{(2)}, \ldots$ be the states of the Markov chain for $\gamma$. When deciding the $t$th state $\gamma$, we sample $\gamma' \sim N_q(\gamma^{(t-1)}, \mathbf{V}^{(t)})$ with

$$\mathbf{V}^{(t)} = \begin{cases} \mathbf{V}^{(0)}, & t < t_0, \\ s\,\mathrm{var}(\gamma^{(1)}, \ldots, \gamma^{(t-1)}) + s_0 \mathbf{I}_q, & t > t_0, \end{cases}$$

where $s$ is recommended to be $2.4^2/q$, $s_0$ is a small constant, $\mathbf{V}^{(0)}$ is the initial variance of the proposal distribution and $\mathbf{I}_q$ is an identity matrix. Similar automatic tuning procedures apply to $\beta, \eta, \phi$ and $\lambda$.

As mentioned in Section 2.3, instead of fixing $c_T$ and $c_\theta$, we also allow both $c_T$ and $c_\theta$ to be random and to be updated through the full conditional distributions with conjugate gamma priors gamma$(a_T, b_T)$ and gamma$(a_\theta, b_\theta)$. The conditional posterior distributions are

$$\pi(c_T|\eta) \sim \Gamma\{(a_T + 2^{J_T - 1} - \tfrac{1}{2}), b_T + \sum_{\epsilon_1 \epsilon_2 .. \epsilon_j} \eta(\epsilon_1 \epsilon_2 \ldots \epsilon_j)^2 j^2/4\}$$

and

$$\pi(c_\theta|\beta) \sim \Gamma\{a_\theta + p(2^{J_\theta - 1} - \tfrac{1}{2}), b_\theta + \sum_{\epsilon_1 \epsilon_2 .. \epsilon_j} \beta(\epsilon_1 \epsilon_2 \ldots \epsilon_j)' \mathbf{D}' \mathbf{D} \beta(\epsilon_1 \epsilon_2 \ldots \epsilon_j) j^2/(4n)\}.$$

## 4. Evaluating the performance of the semiparametric latent promotion time cure rate models

### 4.1. Model comparison criteria

We compare models by using the log-pseudomarginal likelihood LPML (Geisser and Eddy, 1979), which is a measure of model predictive ability, and the deviance information criterion DIC (Spiegelhalter *et al.*, 2002), which is a Bayesian model selection criterion related to the Akaike information criterion. Both are easy to compute on the basis of the MCMC output. In what follows, we give the details of their definitions and computational algorithms.

For our model,

$$\mathrm{LPML} = \sum_{i=1}^{n} \log\{p(t_i|\mathbf{t}_{-i})\},$$

where $p(t_i|\mathbf{t}_{-i})$ is the predictive probability for $t_i$ based on the remaining data $\mathbf{t}_{-i}$. Following the computing algorithms that were suggested in Gelfand and Dey (1994), we have

$$p(t_i|\mathbf{t}_{-i}) = \left\{ \int \frac{1}{p(t_i|\Theta)} \mathrm{d}\Theta \right\}^{-1},$$
$$p(t_i|\Theta) = f_i(t_i|\mathbf{x}_i, \mathbf{z}_i)^{\delta_i} S_i(t_i|\mathbf{x}_i, \mathbf{z}_i)^{1-\delta_i}.$$

LPML is then estimated by

$$\text{LPML} = -\sum_{i=1}^{n} \log\left\{\frac{1}{s}\sum_{k=1}^{s}\frac{1}{p(t_i|\boldsymbol{\Theta}^{(k)})}\right\}, \tag{24}$$

where $\boldsymbol{\Theta}^{(k)} = \{\boldsymbol{\gamma}^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\phi}^{(k)}, \boldsymbol{\eta}^{(k)}, \boldsymbol{\lambda}^{(k)}\}$ are iterates from the MCMC output of all the parameters. When comparing two models, the model with greater LPML suggests a better fit of the data.

For our model,

$$\text{DIC} = 2E[D(\boldsymbol{\Theta}|y)] - D(\hat{\boldsymbol{\Theta}}),$$

where $D(\boldsymbol{\Theta}) = -2\log\{L(\boldsymbol{\Theta})\} + c$, $L(\boldsymbol{\Theta})$ is the likelihood in equation (20) and $c$ is a constant cancelled in model comparison. The effective number of parameters of the model is defined as $p_D = E[D(\boldsymbol{\Theta}|y)] - D(\hat{\boldsymbol{\Theta}})$. The conditional expectation $E[D(\boldsymbol{\Theta}|y)]$ is typically estimated by an average of $D(\boldsymbol{\Theta})$ over the posterior samples of $\boldsymbol{\Theta}$, and the $\hat{\boldsymbol{\Theta}}$ in $D(\hat{\boldsymbol{\Theta}})$ is commonly chosen as the posterior mean of $\boldsymbol{\Theta}$. When comparing two models, the model with smaller DIC suggests a better fit of the data.

### 4.2. Simulations

We simulate data from parametric LPTCR models to evaluate model performance with respect to sample size, different survival models for $F_T(\cdot|\mathbf{x})$ and different sampling distributions for a covariate $z_i$ that is related to the cure rate. The parametric LPTCR models that are considered here assume that

$$S_i(t|\mathbf{x}_i, \theta_i) = \exp\{-\theta_i F_T(t|\mathbf{x}_i)\},$$
$$\theta_i \sim F_\theta(\cdot|z_i),$$

where the baseline distribution $F_0(\cdot)$ is assumed to be a mixture of log-normal distributions: $0.5N\{\log(t), 1.5, 0.8^2\} + 0.5N\{\log(t), 0.5, 0.3^2\}$, $F_T(\cdot|\mathbf{x}_i)$ is related to $F_0(\cdot)$ through the PH, AFT and PO survival models, and $F_\theta(\cdot|z_i)$ is assumed to be a gamma distribution with $\exp(1+z_i)$ for the shape parameter and 2 for the scale parameter. We let $\mathbf{x}_i$ be bivariate where $x_{i1}$ takes values $-0.5$ and 0.5 with equal probabilities and $x_{i2}$ is sampled from $N(0,1)$. We set the true values of the regression parameters $\gamma_1$ and $\gamma_2$ as 0.5 and $-0.5$ respectively. To obtain an observation $t_i$, we firstly sample $\theta_i$ from gamma$\{\exp(1+z_i), 2\}$ and a random number $u$ from the uniform(0,1) distribution. If $u < \exp(-\theta_i)$ or $F_T^{-1}\{-\log(u)/\theta_i|\mathbf{x}_i\} > 15$, we set $t_i = 15$ and $\delta_i = 0$. Otherwise, we set $t_i = F_T^{-1}\{-\log(u)/\theta_i|\mathbf{x}_i\}$ and $\delta_i = 1$. Around 35–50% observations are right censored.

We consider two sample sizes, $n = 500$ and $n = 800$, and two sampling distributions for $z_i$, $N(0, 0.5^2)$ and Bernoulli(0.5). When fitting the samples to semiparametric LPTCR models, we consider two values for the tail-free partition cap $J_{\boldsymbol{\theta}}$ of the prior on $F_\theta(\cdot|z_i)$: 3 and 4. The combinations of the scenarios are displayed in Table 1. The semiparametric LPTCR models that we fit assume that

$$S_i(t|\mathbf{x}_i, \theta_i) = \exp\{-\theta_i F_T(t_i|\mathbf{x}_i)\},$$
$$\theta_i|z_i \sim F_\theta(\cdot|z_i),$$
$$F_\theta(\cdot|z_i) \sim \text{LDTFP}^{J_\theta}(1, G_{\boldsymbol{\phi}}, \mathbf{w}_i),$$
$$F_0(\cdot) \sim \text{LDTFP}^{J_T}(1, G_{\boldsymbol{\lambda}}, 1),$$

where $G_{\boldsymbol{\phi}}$ is a gamma distribution and $G_{\boldsymbol{\lambda}}$ is a Weibull distribution. We let $\mathbf{w}_i = (1, z_i, z_i^{2*}, z_i^{3*})$ for which $z_i^{2*}$ and $z_i^{3*}$ are centred values of $z_i^2$ and $z_i^3$. The starting values of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ are fixed at $\mathbf{0}$. Starting values for $\boldsymbol{\gamma}, \boldsymbol{\phi}$ and $\boldsymbol{\lambda}$ are obtained by fitting the parametric models to the data, which are detailed in Sections 2.2 and 2.3.

**Table 1.** Parameter estimates of the simulated data sets†

| *Model* | *n* | *z* | $J_\theta$ | *Parameter* | *Mean* | *Mean of standard errors* | *Standard deviation of means* | *Coverage probability* |
|---|---|---|---|---|---|---|---|---|
| PH | 500 | $z \sim N(0, 0.5^2)$ | 3 | $\gamma_1$ | 0.518 | 0.174 | 0.171 | 0.943 |
| | | | | $\gamma_2$ | −0.501 | 0.087 | 0.081 | 0.957 |
| PH | 800 | | | $\gamma_1$ | 0.490 | 0.137 | 0.137 | 0.953 |
| | | | | $\gamma_2$ | −0.495 | 0.069 | 0.070 | 0.943 |
| AFT | 500 | $z \sim N(0, 0.5^2)$ | 3 | $\gamma_1$ | 0.497 | 0.087 | 0.101 | 0.906 |
| | | | | $\gamma_2$ | −0.497 | 0.046 | 0.055 | 0.863 |
| AFT | | | 4 | $\gamma_1$ | 0.494 | 0.086 | 0.099 | 0.890 |
| | | | | $\gamma_2$ | −0.500 | 0.046 | 0.053 | 0.917 |
| PO | 500 | $z \sim N(0, 0.5^2)$ | 3 | $\gamma_1$ | 0.500 | 0.204 | 0.199 | 0.977 |
| | | | | $\gamma_2$ | −0.505 | 0.104 | 0.101 | 0.953 |
| PO | | $z \sim$ Bernoulli(0.5) | | $\gamma_1$ | 0.519 | 0.203 | 0.198 | 0.963 |
| | | | | $\gamma_2$ | −0.502 | 0.103 | 0.104 | 0.963 |

†True values $\gamma_1 = 0.5$ and $\gamma_2 = -0.5$.

For each scenario, we simulated 300 data sets and, for each simulated data set, we obtained a chain of 8000 iterates after a burn-in of 20 000 and thinning of every other five iterates. MCMC diagnostics are presented in section 2 of the on-line supplementary file. For each parameter in the vector $\boldsymbol{\gamma}$, we computed its posterior mean, posterior standard deviation and posterior 95% predictive interval. We summarize the results for the 300 data sets in Table 1 where 'Mean' represents the mean of the 300 posterior means, 'Mean of standard errors' the mean of the 300 posterior standard errors, 'Standard deviation of means' the standard deviations of the 300 posterior means and 'Coverage probability' the proportion of predictive intervals which cover the true values.

Table 1 shows that the posterior mean estimates are fairly close to the true values (within 1 standard deviation of the true values) and the coverage probabilities are close to the nominal level 95% across the majority of the models specified. When the sample size increases in the PH model group, the standard deviation estimates decrease but the coverage probabilities stay around the same. For the AFT model group, the coverage probabilities are a little lower than the nominal level. The main reason is that the approximation for the covariate-dependent distribution of the cure rate, using a dependent tail-free process, can be affected by sample size, censoring rate and fixed level of $J_\theta$. The tail-free prior shrinks the cure rate distribution towards a gamma distribution, which serves as a parametric guide, but also induces bias when there is not enough information from the data. Higher $J_\theta$ results in more flexible models and hence reduces bias in estimation. Table 1 also shows that higher $J_\theta$ for the AFT model yields better coverage probabilities. Additional simulation studies for the AFT model (which are not shown here) where the censoring rate is lower also have higher coverage probabilities. We suggest a sensitivity analysis with different $J_\theta$s in data analysis. Overall, setting $J_\theta = 3$ and $J_T = 5$ results in estimation with relatively low bias and good coverage probabilities for the simulations that are considered here.

We also plot the mean pointwise estimated survival or density function and its 95% predictive interval in Figs 1(a) and 1(b) for one scenario. The 'true' curves correspond to the true functions; the 'mean' line corresponds to the mean of the 300 posterior means; 'lower quantile' and 'upper quantile' correspond to the 2.5% and 97.5% quantiles of the 300 posterior means respectively. The estimated mean functions are fairly close to the true functions. An increase in $J_T$ (which is

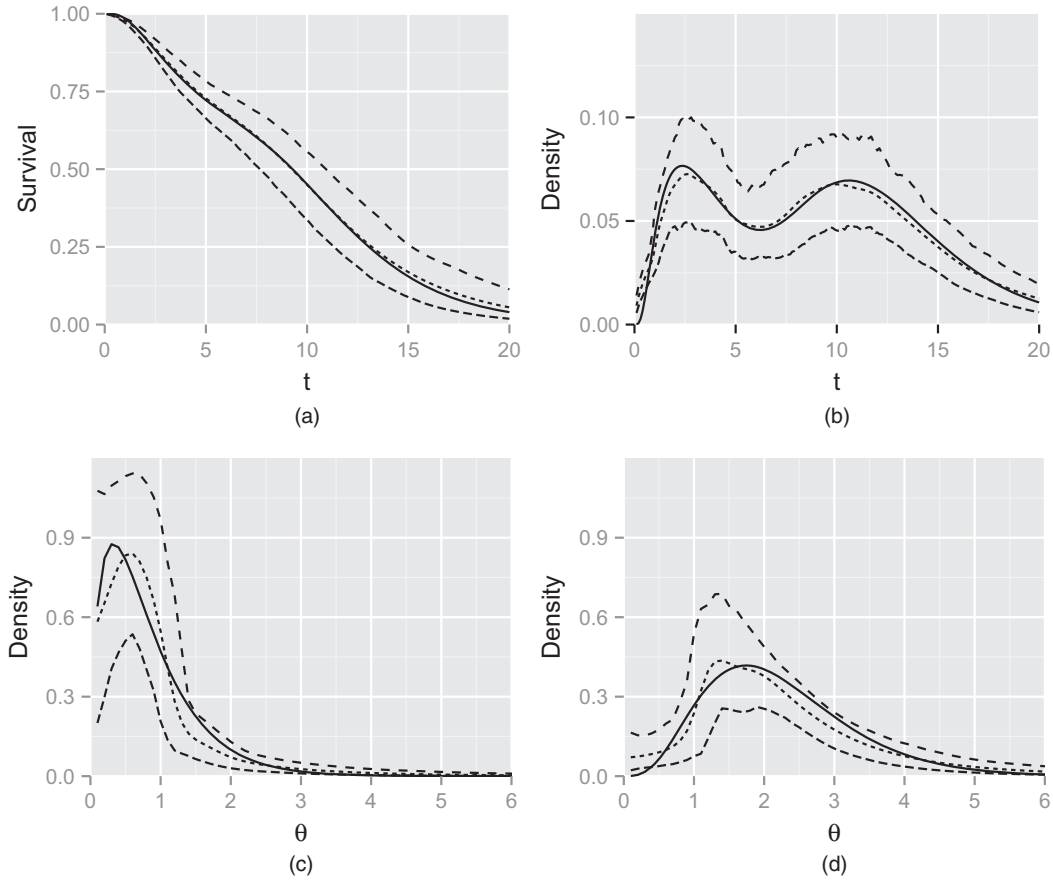


**Fig. 1.** (a) Survival and (b) density estimates of baseline $f_T(\cdot)$ and (c), (d) density estimates of $f_\theta(\theta|z)$ for the AFT model, $n = 500$, $f(\theta) = \text{gamma}\{\exp(1 + z_i), 2\}$, $z_i \sim N(0, 0.5^2)$, $J_T = 5$ and $J_\theta = 3$ (———, true; - - - - - -, mean; — — —, lower and upper quantiles): (a) $S_0(t) = 1 - F_0(t)$ estimates; (b) $f_0(t)$ estimates; (c) $f_\theta(\theta| - 0.5)$ estimates; (d) $f_\theta(\theta|0.5)$

not shown here) would also improve the estimations of survival and density functions. Figs 1(c) and 1(d) plot the estimated cure rate parameter distributions and their 95% predictive intervals given the value of $z$s. The true distributions are mostly covered by the predictive intervals. Less precision in the left-hand tail of each estimated $f_\theta(\cdot|z)$ would be expected due to less information for large cure rate probabilities $\exp(-\theta)$.

## 5. Application to New Mexico 'Surveillance, epidemiology, and end results' breast cancer patient data

We analyse a New Mexico 'Surveillance, epidemiology, and end results' (SEER) data set for women who were initially diagnosed with stage III breast cancer during 2000–2012. The recorded survival time is either time to death from breast cancer or the end of the observation period (right censored). More than 95% of the patients received surgery and, of those who received surgery, only 61% received additional radiation treatments. We focus on a two-group comparison in this study where one group includes the patients who received surgery only (41%) and the other group includes those who received radiation in addition to surgery (61%). 92.3% of the women

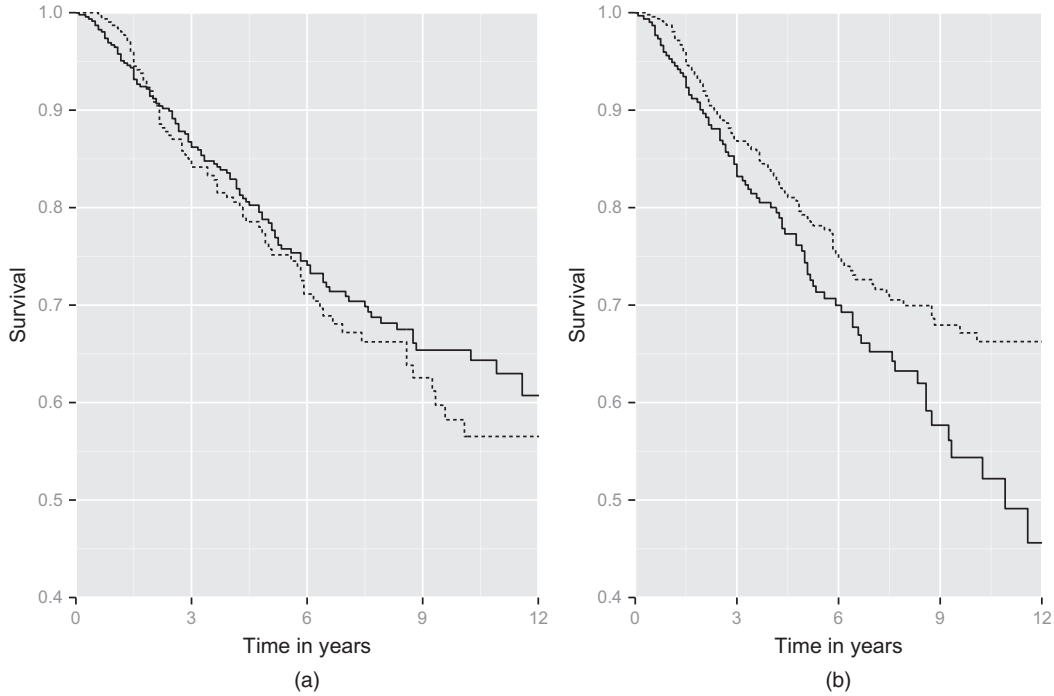


**Fig. 2.** Kaplan–Meier estimates of the survival functions for the New Mexico SEER stage III breast cancer data: (a) estimates by different ethnicity (———, non-Hispanic; - - - - - -, Hispanic); (b) estimates by treatment (———, surgery only; - - - - - -, surgery and radiation)

are either Hispanic white or non-Hispanic white, and other ethnicity groups were excluded in the analysis. About 41.2% of the patients in the data are Hispanic white (coded 1) and the final data size is $n = 801$. We consider standardized age at diagnosis ((age − mean of age)/(standard deviation of age)), ethnicity (Hispanic *versus* non-Hispanic) and an indicator for treatment (surgery only *versus* surgery plus radiation). Fig. 2 gives an illustration of how ethnicity and radiation affect the survival functions by using Kaplan–Meier estimates. Patients who received radiation in addition to surgery seem to live significantly longer than those who received only surgery. There is also some evidence that the non-Hispanic group has a better survival than the Hispanic group.

We consider LPTCR PH, AFT and PO models with $\mathbf{w}$ including an intercept, standardized age, ethnicity and an indicator for radiation. Using the same mixture of tail-free processes prior for $F_0(t)$, we fit models with $\theta = \exp(\mathbf{z}'\boldsymbol{\psi})$ where $\mathbf{z}$ includes an intercept, the indicator for ethnicity, standardized age and the indicator for radiation. We also fit basic PH, AFT and PO models which do not consider the cure rate fraction in the population. For the models above, we fix the mixture of tail-free prior partition caps $J_\theta = 3$ and $J_T = 5$. For each LPTCR model, we obtain a chain of 8000 iterates after a burn-in of 100 000 and thinning every other 50 iterates. For the other simpler models that we compare, we obtain a chain of 8000 iterates after a burn-in of 40 000 and thinning every other 20 iterates. The acceptance rates for all the parameters are between 20% and 80% and good convergence is achieved. See the on-line supplementary file for more diagnostic details. Table 2 illustrates the LPML- and the DIC-values for all the models. On the basis of the results in Table 2, LPTCR models improve both LPML and DIC over the models assuming $\theta = \exp(\mathbf{z}'\boldsymbol{\psi})$ (especially for PH and PO models) and the models which do not

**Table 2.** Model comparisons for the New Mexico SEER stage III breast cancer data†

| *Model for* $F_T(\cdot)$ | *Results for the following cure rate models:* | | | | | |
|---|---|---|---|---|---|---|
| | *LPTCR* | | $\theta = exp(\mathbf{z}'\psi)$ | | *No cure rate* | |
| | *LPML* | *DIC* | *LPML* | *DIC* | *LPML* | *DIC* |
| PH | −751 | 1482 (−0.4) | −757 | 1503 (8.5) | −760 | 1514 (6.5) |
| AFT | −749 | 1477 (3.0) | −750 | 1488 (11.6) | −755 | 1499 (5.3) |
| PO | −750 | 1477 (−5.5) | −754 | 1497 (6.7) | −759 | 1515 (8.4) |

†Values in parentheses are the effective numbers of parameters for DICs.

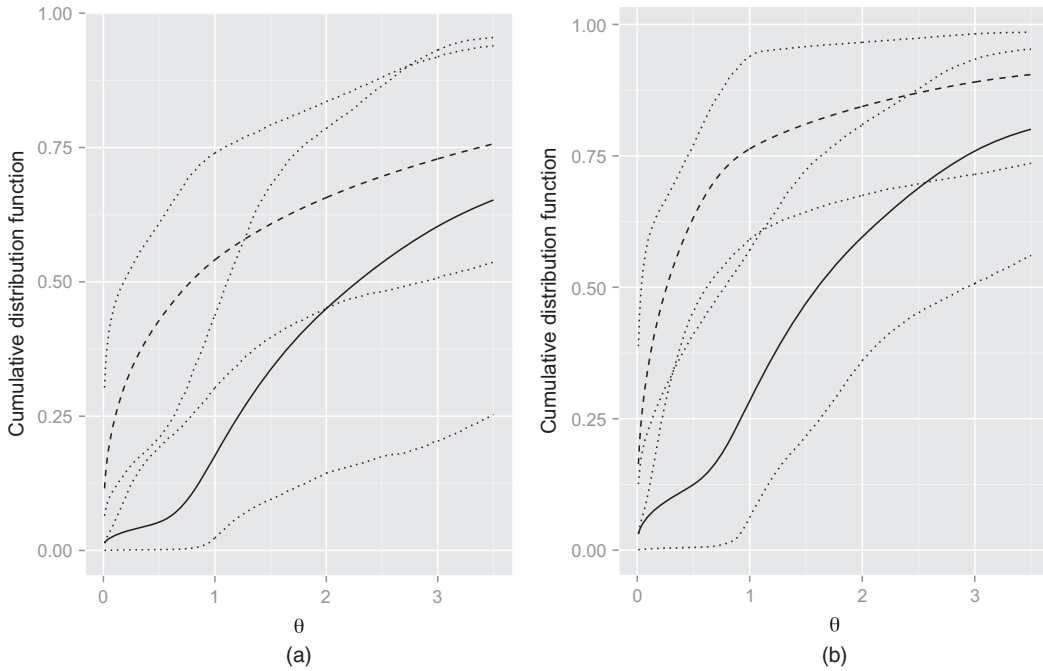


**Fig. 3.** Mean estimates (———, surgery; – – –, radiation and surgery) and 2.5% quantiles and 97.5% quantiles (· · · · · ·) of the cure rate distributions based on the AFT model for the New Mexico SEER stage III breast cancer data with age at diagnosis 57 years: (a) $F_\theta(\theta|\mathbf{z})$ estimates for the Hispanic white group; (b) $F_\theta(\theta|\mathbf{z})$ estimates for the non-Hispanic white group

assume a cure fraction in the population at all. Among the PH, AFT and PO models, the AFT models seem to yield a better fit consistently.

On the basis of the LPTCR AFT model with $J_\theta = 3$, we plot the estimated $F_\theta(\cdot|\mathbf{z})$ to compare the cure rate distribution for patients with and without radiation by ethnicity group in Fig. 3, while fixing age at diagnosis at its sample mean (57 years). Recall that the proportion of being cured for patients in a specific population is $\exp(-\theta)$. Lower $\theta$ indicates a higher cure rate probability. Fig. 3 shows that both Hispanic white and non-Hispanic white patients who received surgery and radiation have higher probabilities for high cure rates than those who

**Table 3.** Model fitting of the AFT models for the New Mexico SEER stage III breast cancer data

| *Parameter* | *LPTCR* | $\theta = exp(\mathbf{z}'\psi)$ | *No cure rate* |
|---|---|---|---|
| $\psi_1$ (intercept) | — | 0.076 (−0.337, 0.535) | — |
| $\psi_2$ (age) | — | 0.020 (−0.171, 0.212) | — |
| $\psi_3$ (ethnicity) | — | 0.597 (0.020, 1.363) | — |
| $\psi_4$ (radiation) | — | −0.962 (−1.389, −0.543) | — |
| $\gamma$ (age) | 0.278 (−0.036, 0.475) | 0.337 (0.213, 0.446) | 0.357 (0.222, 0.471) |
| $\gamma$ (ethnicity) | −0.301 (−0.797, 0.329) | −0.399 (−1.098, 0.235) | 0.254 (−0.125, 0.568) |
| $\gamma$ (radiation) | 0.516 (−0.231, 0.893) | 0.715 (0.452, 0.975) | −0.272 (−0.537, −0.064) |

received surgery only. The non-Hispanic group seems to have a distribution favouring small $\theta$ and hence favouring a higher cure rate. This also can provide a guide to practical interpretations. For example, from Fig. 3, given age 57 years at diagnosis and non-Hispanic ethnicity, the estimated probability of having a proportion of more than 70% being 'cured' is 0.132 for stage III breast cancer patients who have surgery only, whereas it is 0.560 for those who receive surgery and radiation, given age 57 years at diagnosis and Hispanic ethnicity. The estimated probability of having a proportion of more than 70% being cured is 0.056 for stage III breast cancer patients who have surgery only, whereas it is 0.382 for those who receive surgery and radiation.

For all the AFT models, the estimates of $\gamma$ and $\beta^*$ are presented in Table 3. An equivalent expression of the AFT model in equation (4) is $\log(T) = -\mathbf{x}'\gamma + T_0$ where $T_0$ is the error term. The LPTCR AFT model estimates that an increase of around 14 years in age at diagnosis significantly shortened the average time for each remaining tumour cell to produce detectable metastatic disease by 1.3 years. Meanwhile, ethnicity and radiation do not have a significant effect on those remaining tumour cells. On the basis of the AFT model without cure rate, an increase of around 14 years in age at diagnosis significantly shortened the average patient lifespan by 1.4 years whereas additional radiation significantly prolonged the average patient lifespan by 1.3 years.

For sensitivity analysis regarding the LPTCR AFT model, we first let $J_\theta = 4$ and obtain −749 for LPML and 1476 for DIC. Hence we conclude that $J_\theta = 3$ is sufficient for modelling the distribution of $F_\theta(\cdot|\mathbf{z})$. We also let $c_\theta$ and $c_T$ be assigned independent gamma priors gamma(1, 1), we obtain −749 for LPML and 1483 for DIC. Compared with the LPML for setting both precision parameters to 1, little difference is found. We also set $c_T$ to be a large value, resulting in a parametric Weibull model for the survival distribution, in the same spirit of the parametric AFT promotion time model in Yakovlev and Tsodikov (1996). We obtain −748 for LPML and 1490 for DIC. It suggests that a Weibull model for the survival distribution suffices for the LPTCR AFT model.

We also compare our LPTCR AFT model with the AFT model with ordinal regression for $\theta$ (Kim *et al.*, 2009). The number of discrete levels $G$ for $\theta$ is fixed at 3, 4 and 5. We further specify the prior means for $\theta$s as $\theta_{01} = 0.4, \theta_{02} = 0.7$ and $\theta_{03} = 1$ when $G = 3$, $\theta_{01} = 0.4, \theta_{02} = 0.7, \theta_{03} = 1$ and $\theta_{04} = 1.8$ when $G = 4$ and $\theta_{01} = 0.4, \theta_{02} = 0.7, \theta_{03} = 1, \theta_{04} = 1.8$ and $\theta_{05} = 2.5$ when $G = 5$. We note that $(0.4, 0.7, 1, 1.8)$ for $G = 4$ corresponds to the estimated $\theta$s for the four ethnicity-by-radiation groups based on the cure rate model in Table 2 where $\theta = \exp(\mathbf{z}'\psi)$. The hyperparameter $c_0$ is fixed at 2.5, which reflects a vague prior belief. Priors for the ordinal regression parameters $\phi_k$, $k = 1, \ldots, G$, are $N_4(\mathbf{0}, 3^2\mathbf{I}_4)$ and for the AFT regression parameter $\gamma$ is $N_3(\mathbf{0}, 5^2\mathbf{I}_3)$. For each model fit, we obtain 15 000 iterates, after a burn-in of 200 000 and thinning every other 50.

We obtain −765, −765 and −766 for LPML and 1484, 1480 and 1475 for DIC. The LPMLs are significantly lower than that of our LPTCR AFT model though the DICs suggest similar fits.

## 6. Conclusion

This paper proposes an LPTCR model for data where a proportion of subjects are cured and the cure rate has significant medical implications. The model proposed is more flexible than the latent cure rate marker model of Kim *et al*. (2009) by generalizing the discrete distribution of the cure rate parameter to a factor-dependent continuous distribution. We use a mixture of tail-free processes for both the distribution of the cure rate and the survival distribution of each metastatic tumour cell. The algorithms that were developed in this work allow us to fit LPTCR with several survival models for the tumour cell. Its flexibility is demonstrated through the simulations. We hope that the New Mexico SEER breast cancer data analysis provides some practical interpretations for physicians and patients. C++ code for implementing the algorithms has been written in the 'Rcpp' environment and can be called in R easily. The code is available on request.

## Acknowledgements

Ji-Hyun Lee's work was supported in part by National Cancer Institute grant P30CA11800-110. The authors thank Tim Hanson, Jiajia Zhang and Davis Bert for helpful discussions and Tawny Boyce who helped to extract and managed the SEER data.

## References

Berkson, J. and Gage, R. P. (1952) Survival curve for cancer patients following treatment. *J. Am. Statist. Ass.*, **47**, 501–515.

Chen, M. H., Ibrahim, J. G. and Sinha, D. (1999) A new Bayesian model for survival data with a surviving fraction. *J. Am. Statist. Ass.*, **94**, 909–919.

Escobar, M. D. and West, M. (1995) Bayesian density estimation and inference using mixtures. *J. Am. Statist. Ass.*, **90**, 577–588.

Ferguson, T. S. (1973) A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, **1**, 209–230.

Gallardo, D. I., Bolfarine, H. and Pedroso-De-Lima, A. C. (2016) Promotion time cure rate model with bivariate random effects. *Communs Statist. Simuln Computn*, **45**, 603–624.

Geisser, S. and Eddy, W. F. (1979) A predictive approach to model selection. *J. Am. Statist. Ass.*, **74**, 153–160.

Gelfand, A. E. and Dey, D. K. (1994) Bayesian model choice: asymptotics and exact calculations. *J. R. Statist. Soc.* B, **56**, 501–514.

Gu, Y., Sinha, D. and Banerjee, S. (2011) Analysis of cure rate survival data under proportional odds model. *Liftim. Data Anal.*, **17**, 123–134.

Haario, H., Saksman, E. and Tamminen, J. (2005) Componentwise adaptation for high dimensional MCMC. *Computnl Statist.*, **20**, 265–273.

Hanson, T. (2006) Inference for mixtures of finite Polya tree models. *J. Am. Statist. Ass.*, **101**, 1548–1565.

Hanson, T. and Johnson, W. (2002) Modelling regression error with a mixture of Polya trees. *J. Am. Statist. Ass.*, **97**, 1020–1033.

Hanson, T. and Yang, M. (2007) Bayesian semiparametric proportional odds models. *Biometrics*, **63**, 88–95.

Ibrahim, J. G., Chen, M. H. and Sinha, D. (2005) *Bayesian Survival Analysis*. Hoboken: Wiley.

Jara, A. and Hanson, T. (2011) A class of mixtures of dependent tailfree processes. *Biometrika*, **98**, 553–566.

Kim, S., Xi, Y. and Chen, M. H. (2009) A new latent cure rate marker model for survival data. *Ann. Appl. Statist.*, **3**, 1124–1146.

Lavine, M. (1992) Some aspects of Polya tree distributions for statistical modelling. *Ann. Statist.*, **20**, 1222–1235.

Li, C. S. and Taylor, J. M. (2002) A semi-parametric accelerated failure time cure model. *Statist. Med.*, **21**, 3235–3247.

Lopes, C. M. C. and Bolfarine, H. (2012) Random effects in promotion time cure rate models. *Computnl Statist. Data Anal.*, **56**, 75–87.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion), *J. R. Statist. Soc.* B, **64**, 583–639.

Tierney, L. (1994) Markov chains for exploring posterior distributions (with discussion). *Ann. Statist.*, **22**, 1701–1762.
Tsodikov, A. D., Ibrahim, J. G. and Yakovlev, A. Y. (2003) Estimating cure rates from survival data. *J. Am. Statist. Ass.*, **98**, 1063–1078.
Yakovlev, A. Y., Asselain, B., Bardou, V. J., Fourquet, A., Hoang, T., Rochefediere, A. and Tsodikov, A. D. (1993) A simple stochastic model of tumour recurrence and its application to data on premenopausal breast cancer. *Biomet. Anal. Donn. Spatio-temp.*, **12**, 66–82.
Yakovlev, A. Y. and Tsodikov, A. D. (1996) *Stochastic Models of Tumour Latency and Their Biostatistical Applications*, vol. 1. Singapore: World Scientific Publishers.
Zellner, A. (1983) Applications of Bayesian analysis in econometrics. *Statistician*, **32**, 23–34.

*Supporting information*
Additional 'supporting information' may be found in the on-line version of this article:

'Web-based supplementary materials for "Latent promotion time cure rate model using tailfree mixtures"'.