Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

A Model based on Mixture of Weibull Distributions for Depending Competing Risks Data in the Presence of Long-term Survivors, and it's Application to Malignant Melanoma Cancer Data

Ayon Ganguly[1] | **Farha Sultana**[2] | **Debasis Kundu**[3] | **Ayan Pal**[4]

[1]Department of Mathematics, Indian Institute of Technology Guwahati, Assam, India

[2]Department of Science and Mathematics, Indian Institute of Information Technology Guwahati, Assam, India

[3]Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Uttar Pradesh, India

[4]Department of Statistics, The University of Burdwan, West Bengal, India

Correspondence
Ayon Ganguly, Department of Mathematics, Indian Institute of Technology Guwahati, Assam, India.
Email: aganguly@iitg.ac.in

Abstract

The flexibility of finite mixture models makes them suitable candidates for analyzing survival data with complex, multimodal distributions. Such data is often available if the event of interest occurs due to multiple failure modes. Here, we explore modelling of competing risks time-to-event data with covariates in the presence of long-term survivors in the population using finite mixture models. The mixture cure rate model is used to describe the uncertainty in the population, where the susceptible part of the population is modelled using a finite mixture of Weibull distributions with different shape and scale parameters. Moreover, if information on covariates are available, the cure rate may be modelled using a binary regression model on the covariates. Here, we use the logistic function to relate covariates to the cure rate. The distribution corresponding to the susceptible part may also depend on covariates. To explore such dependency, we model the scale parameter of the Weibull distribution using covariates. Then, we discuss the classical parametric inference for the constructed model based on random and non-informative right-censored competing risks time-to-event data. An efficient method based on the expectation-maximization algorithm is proposed to estimate model parameters, thereby avoiding the complexity of directly maximizing the likelihood function. Additionally, a method for constructing confidence intervals for all model parameters is addressed. A simulation study is performed in the presence of two competing causes to investigate the finite sample properties of the proposed estimation methodologies. Finally, the methods are illustrated by analyzing a real data set on malignant melanoma cancer. Predicting the conditional survival function of an alive patient is of natural interest to an experimenter or medical researcher. A method for estimating such a conditional survival probability is also discussed.

KEYWORDS
Finite mixture model, Weibull distribution, cure rate model, competing risks, censoring, likelihood inference, EM algorithm, confidence interval

1 | INTRODUCTION

Investigators in medical studies often observe a group of individuals who are succumbing to a particular disease (such as cancer), alongside other risk factors (like heart disease or kidney disease). Under such circumstances, a natural interest of the experimenter is to look for a reliable estimate of the proportion of subjects (individuals/devices/ units) and the distribution of time-to-event of a subject experiencing the event of interest due to a particular mode in the presence of other modes. This problem is known as the competing risks problem in the statistical literature. A typical competing risks data set looks like

$$(t_1, \delta_1), (t_2, \delta_2), \ldots, (t_n, \delta_n),$$

 © 2025 Copyright Holder Name

where t_i denotes the time-to-event for the i–th subject, and δ_i denotes the mode due to which the i–th subject experienced the event of interest. When the event of interest associated with a subject is its failure, the time-to-event becomes its failure time. In this article, we use the terms 'time-to-event', 'lifetime', and 'failure time' interchangeably.

One of the popular ways to model competing risks data is the latent failure times approach.[1] In the latent failure times approach, the time-to-event is modeled using the random variable $\min\{T_1^*, T_2^*, \ldots, T_K^*\}$, where K is the total number of modes, and T_k^* is the random variable denoting the time-to-event due to k–th mode, $k = 1, 2, \ldots, K$.[2,3,4]

When the failure modes are not independent, a reasonable and flexible modelling approach for competing risks data is to consider the finite mixture models. The underlying time-to-event distribution may have a bimodal or multimodal density in the presence of multiple failure modes. Finite mixture distributions often become a very flexible family of distributions in this setup.[5,6,7] Let π_k be the probability that a subject experiences the event of interest due to k–th mode, $k = 1, 2, \ldots, K$. The density of the survival time of the subject is $\sum_{k=1}^{K} \pi_k f_k(t)$, where $f_k(t)$ is the density associated with T_k^*. In parametric setup, one assumes parametric forms of $f_k(\cdot)$'s, for example, exponential or Weibull densities, and proceed for inference.[6] However, it is observed that the complexity of the optimization problem grows with an increase in the number of modes. An efficient algorithm based on expectation-maximization (EM) method for estimating the model parameters of mixture of Weibull distributions may be used when the data are possibly right censored.[8] The efficiency of the algorithm is due to its' ability to reduce the dimension of the optimization problem.

In survival and reliability analysis, a general perception is that the subjects under study will, over time, experience the event of interest if the monitoring time window is reasonably large. However, with huge developments in medical science and technology, many subjects never experience the event of interest. Cure rate models are well known in statistical literature to deal with situations when the underlying population comprises of a mixture of subjects (called cured/ immunes/ long-term survivors) not experiencing the event of interest, and subjects (called non-cured/ susceptible) experiencing the event of interest.[9] The most popular approach to model data in the presence of cured subjects is the two-component mixture cure rate model.[10] The model consists of a component representing the immune proportion - also known as incidence - in the population and a distribution - also known as a latency distribution - describing the time to the occurrence of the event of interest. Denoting cure fraction in the population by $p_0 \in (0, 1)$, the overall survival function, denoted by $S(t)$, of the entire population is given by

$$S(t) = p_0 + (1 - p_0)S_0(t), \tag{1}$$

where $S_0(t)$ denotes the conditional survival function of the lifetime of a subject given that the subject is susceptible in the population. It is assumed that $S_0(\cdot)$ is a proper survival function in the sense that $\lim_{t\to\infty} S_0(t) = 0$ and $S_0(0) = 1$.

Several authors have studied cure rate models assuming that the members of the susceptible population experience the event of interest owing to one of the independent competing failure modes.[11,12,13,14,15] However, it is hard to justify the assumption of independence between different failure modes in many practical scenarios. For example, the progress of cancer may be accelerated by the presence of comorbidities like heart disease, lung disease, etc. Therefore, the assumption of independence between lifetime due to cancer and lifetime due to lung disease of a patient may not be quite plausible. Moreover, the assumption of independence may not be verified when information on covariate are not available. Two popular ways of modelling competing risks data in the presence of cured subjects using (1) are as follows. First, $S_0(\cdot)$ may be modelled using $\sum_{k=1}^{K} \pi_k S_k(\cdot)$.[16] Secondly, $S_0(\cdot)$ can be modelled using vertical modelling.[17]

In this article, we consider modelling and statistical analysis of dependent competing risks time-to-event data in the presence of long-term survivors. The effect of the existence of long-term survivors in the population is modelled using a mixture cure rate model. A finite mixture of Weibull models explains the uncertainty due to dependent competing failure modes in susceptible part of the population. Use of Weibull distributions brings flexibility into the model. Generally, the probability of cure or/and latency distribution for a particular subject depend on a set of covariate information. Here, we used a logistic link function to connect the covariates of a subject to its cure fraction. The dependency of latency distribution on covariates is modelled assuming log-linear relationship between scale parameters involved in latency distribution and covariates. An efficient method based on an EM algorithm is proposed to obtain likelihood inference in the above setup. In summary, the contributions of this manuscript are as follows:

- We propose a model for dependent competing risks data in the presence of long-term survivors using mixture model.
- The proposed model can be used quite effectively to accommodate covariate information in time-to-event competing risks data.

- We propose an efficient algorithm to perform likelihood inference of dependent competing risks data in the presence of long-term survivors.
- An extensive simulation study is performed to check the efficacy of the proposed model. It is found that the proposed method works quite satisfactorily.

The rest of the paper is organized as follows. In Section 2, we motivate the problem using a data set on malignant melanoma. In Section 3, the model for analyzing the data set is formulated using mixture cure rate model, where a finite mixture of Weibull distributions is used as the latency distribution. Based on the non-informative random right censored competing risks data, the likelihood function is constructed, and an efficient EM based algorithm is proposed to obtain the maximum likelihood estimators of the model parameters in Section 4. In Section 5, an extensive simulation study is carried out to validate our proposed methodology. The model proposed in Section 3 is extended in Section 6 by incorporating covariates. The real data set introduced in Section 2, is analyzed in Section 7 to illustrate our proposed model and methods. Finally, we conclude the paper in Section 8.

2 | MOTIVATING DATA SET

Melanoma is a kind of skin cancer that starts in the melanocytes, the cells responsible for skin color. The malignant melanoma cancer data set[18,19] (named `melanoma`), which was collected at Odense University Hospital by K.T. Drzewiecki during 1962 to 1977, on survival times of patients after operation for malignant melanoma is available in `boot` package[20] of `R` software. Two hundred and five patients were followed after the operation until they died or they left the study or the termination of the study in the year 1977. When they passed away, the number of days the patient survived after the operation, along with the cause (death due to melanoma or death due to any other causes) of death, was recorded. The minimum and maximum values of observed survival times are 10 and 5565 days, respectively. Out of 205 patients, 57 patients (about 28%) died due to melanoma cancer, 14 patients (about 7%) died due to other causes, and 134 patients (about 65%) were right censored. Moreover, information on the thickness of tumor (with mean 2.92 mm, standard deviation 2.92 mm, minimum 0.10 mm, maximum 17.42 mm), the status of ulcer (Present 44%, Absent 56%), age (in year) at operation (with mean 52.46 years, standard deviation 16.67 years, minimum 4 years, maximum 95 years), year of operation (between 1962 and 1977) and sex (Male 39%, Female 61%) for each of the patients are available. Note that information on the cause of death makes it a competing risks data. This data set was earlier analyzed utilizing vertical modeling framework.[21] Based on the data set, the Kaplan-Meier estimate of the distribution function of survival time is given in Figure 1. It shows that the Kaplan-Meier estimate levels off at a non-zero proportion; therefore, interest lies in estimating the cure proportion for each patient by modelling the data set using a cure rate model.

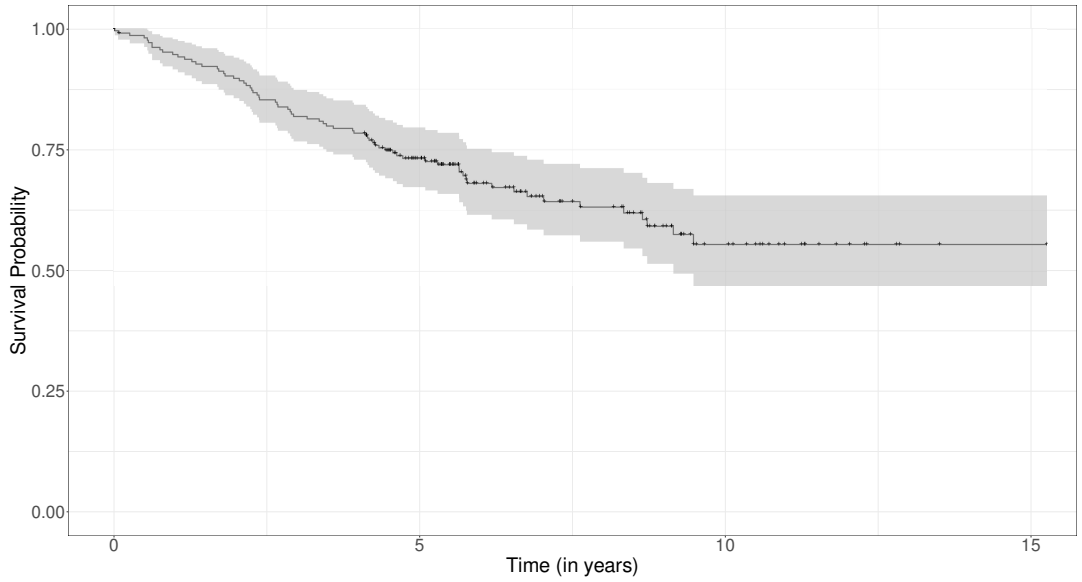

FIGURE 1 Kaplan-Meier Estimate of Survival Function

3 | MODEL FORMULATION

In this article, we assume that the population consists of two types of subjects: immune or cured and susceptible. A subject that is subject to the event of interest is referred to as susceptible. A subject that is not subject to the event of interest is referred to as immune or cured. Thus, the survival time of an immune subject is infinite. Let $\widetilde{T}$ denote the survival time of a subject. We introduce an indicator random variable $\widetilde{I}$, which takes a value of zero if the subject is cured and one if it is susceptible. Further, we assume that the probability that a randomly selected subject belongs to the immune group is $P(\widetilde{I} = 0) = p_0$; therefore, the probability that a randomly selected subject is susceptible is $P(\widetilde{I} = 1) = 1 - p_0$.

Consider a subject exposed to K mutually exclusive and exhaustive competing risks or failure modes. Thus, a susceptible subject fails due to one of the K modes. Let $\widetilde{N}$ denote the failure mode of a susceptible subject. Then, $\widetilde{N}$ is a discrete random variable taking values in the set $\{1, 2, \ldots, K\}$. Let $\pi_k = P(\widetilde{N} = k|\widetilde{I} = 1)$ be the probability of occurrence of the event of interest owing to the k–th mode, $k = 1, 2, \ldots, K$. Thus, $\sum_{k=1}^{K} \pi_k = 1$. Let the associated mode-specific conditional survival function be denoted by

$$S_k(t) = P(\widetilde{T} > t|\widetilde{N} = k, \widetilde{I} = 1), \quad \text{for} \quad k = 1, 2, \ldots, K.$$

It is assumed that $S_k(.)$ is a proper survival function in the sense that $S_k(0) = 1$ and $S_k(\infty) = 0$ for all $k = 1, 2, \ldots, K$. Let the corresponding mode-specific conditional cumulative distribution function (CDF) and probability density function (PDF) of the time-to-event be denoted by $F_k(\cdot)$ and $f_k(\cdot)$, respectively. Thus, given that a subject belongs to the susceptible group, the survival function of the time-to-event is given by

$$S_{mix}(t) = P(\widetilde{T} > t \mid \widetilde{I} = 1) = \sum_{k=1}^{K} \pi_k S_k(t). \tag{2}$$

The corresponding CDF and PDF are

$$F_{mix}(t) = \sum_{k=1}^{K} \pi_k F_k(t) \qquad \text{and} \qquad f_{mix}(t) = \sum_{k=1}^{K} \pi_k f_k(t),$$

respectively. It is clear from the above formulation that the total hazard function of a subject given that the subject is susceptible is given by

$$h(t) = \frac{\sum_{k=1}^{K} \pi_k f_k(t)}{\sum_{k=1}^{K} \pi_k S_k(t)}.$$

The cause specific hazard function for k-th mode given the subject is susceptible is

$$h_k(t) = \frac{\pi_k f_k(t)}{\sum_{j=1}^{K} \pi_j S_j(t)}$$

for $k = 1, 2, \ldots, K$.

It may be mentioned here that the survival functions $S_k(\cdot)$ are assumed to be independent in usual competing risks setup.[22] However, the survival function might be dependent in the mixture model formulation described above.[5,23] The vertical modelling is also a mixture distribution based model for analysis of competing risks data.[21] The main difference between the proposed model and vertical modelling is as follows. Note that the joint distribution of $\tilde{T}$ and $\tilde{N}$ can be factorized in two ways:

$$f_{\tilde{T},\tilde{N}|\tilde{I}}(t, n) = f_{\tilde{N}|\tilde{T},\tilde{I}}(n) f_{\tilde{T},\tilde{I}}(t) = f_{\tilde{T}|\tilde{N},\tilde{I}}(t) f_{\tilde{N}|\tilde{I}}(n),$$

where, f is used as a generic symbol for density function. The model is formulated using the first factorization in the vertical modelling, while we use the second factorization in building the proposed model.

Let us assume that the survival function $\widetilde{T}$ be denoted by $S_{pop}(\cdot)$. Then

$$S_{pop}(t) = P(\widetilde{T} > t) = P(\widetilde{T} > t|\widetilde{I} = 0)P(\widetilde{I} = 0) + P(\widetilde{T} > t|\widetilde{I} = 1)P(\widetilde{I} = 1) = p_0 + (1 - p_0)S_{mix}(t). \tag{3}$$

The associated CDF and PDF are

$$F_{pop}(t) = (1 - p_0)F_{mix}(t) \qquad \text{and} \qquad f_{pop}(t) = (1 - p_0)f_{mix}(t),$$

respectively.

In this manuscript, we further assume that the k–th mode-specific time-to-event follows Weibull distribution with shape parameter $\alpha_k > 0$ and scale parameter $\lambda_k > 0$, $k = 1, 2, \ldots, K$. The CDF of a Weibull distribution with shape parameter $\alpha > 0$ and $\lambda > 0$ is given by

$$F(t) = \begin{cases} 1 - e^{-\lambda t^{\alpha}} & \text{if } t > 0 \\ 0 & \text{elsewhere.} \end{cases} \tag{4}$$

From now on, we will denote Weibull distribution with CDF given in (4) by WE(α, λ). Note that the assumption of different shape and scale parameters for the individual mode-specific conditional distributions make the model a flexible one.

4 | LIKELIHOOD FUNCTION AND PARAMETER ESTIMATION

4.1 | Likelihood Function and Maximum Likelihood Estimators

Most of the survival or reliability data are right censored. This article considers a general framework of a random right censoring scheme. In a random sample of n subjects, let the random variables C_l and $\widetilde{T}_l$ denote the censoring time and the time-to-event, respectively, of the l–th subject, $l = 1, 2, \ldots, n$. In the presence of right censoring, the experimenter observes the actual time-to-event of a subject if the event of interest occurs before censoring time for the subject. If the event does not occur before censoring time for a subject, the actual time-to-event is not observed for the subject. When the actual time-to-event is observed, we assume that the failure mode is also observed. Therefore, the form of available data is as follows:

$$\{(t_l, \delta_l) : l = 1, 2, \ldots, n\},$$

where t_l and δ_l are the observed values of $T_l = \min(\widetilde{T}_l, C_l)$ and $\Delta_l = \widetilde{N} \times I(\widetilde{T}_l \leq C_l)$. Here, $I(\cdot)$ denotes an indicator function that takes value one if its argument is true and zero if it is false. Thus, $\delta_l = 0$ means the l–th subject is right censored and $\delta_l = k$ means the l–th subject experienced the event due to mode k, $k = 1, 2, \ldots, K$.

Based on n paired observations as given above, the likelihood function (up to a multiplicative constant) under non-informative random right censoring becomes

$$L(\boldsymbol{\theta}; \boldsymbol{t}, \boldsymbol{\delta}) \propto \prod_{l=1}^{n} \left[\left\{ \prod_{k=1}^{K} \{(1 - p_0)\pi_k f_k(t_l; \boldsymbol{\theta}_k)\}^{I(\delta_l = k)} \right\} \{S_{pop}(t_l; \boldsymbol{\theta})\}^{I(\delta_l = 0)} \right],$$

where $\boldsymbol{\theta} = (\alpha_1, \ldots, \alpha_K, \lambda_1, \ldots, \lambda_K, \pi_1, \ldots, \pi_K, p_0)$, $\boldsymbol{t} = (t_1, \ldots, t_n)$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)$ and $\boldsymbol{\theta}_k = (\alpha_k, \lambda_k)$. Therefore, the log-likelihood function (up to an additive constant) can be expressed as

$$l(\boldsymbol{\theta}) = \sum_{l=1}^{n} \left[\sum_{l=1}^{K} I(\delta_l = k) \{\ln(1 - p_0) + \ln \pi_k + \ln f_k(t_l; \boldsymbol{\theta}_k)\} + I(\delta_l = 0) \ln \{p_0 + (1 - p_0)S_{mix}(t_l; \boldsymbol{\psi})\} \right], \tag{5}$$

where $\boldsymbol{\psi} = (\alpha_1, \ldots, \alpha_K, \lambda_1, \ldots, \lambda_K, \pi_1, \ldots, \pi_K)$. The maximum likelihood estimators (MLEs) of model parameters can be found by maximizing log-likelihood function as given in (5) with respect to $\boldsymbol{\theta}$ over the parametric space

$$\boldsymbol{\Theta} = \left\{ \boldsymbol{\theta} : \alpha_k > 0,\ \lambda_k > 0,\ 0 < \pi_k < 1,\ k = 1, 2, \ldots, K,\ \sum_{k=1}^{K} \pi_k = 1,\ 0 < p_0 < 1 \right\}.$$

It is clear that finding MLEs of model parameters boils down to a $(3K + 1)$ dimensional constrained optimization problem, which is quite difficult even if K, the number of failure modes, has a moderate value. In the next subsection, we will propose an EM algorithm based technique that can efficiently overcome this issue. This technique will be referred to as MLE/EM algorithm in the rest of the article.

4.2 | Implementation of MLE/EM algorithm

Note that the indicator variable $\widetilde{I}$ takes the value 1 if the index of the subject belongs to the set $J_1 = \{l : \delta_l \neq 0\}$ and is latent if the index of the subject belongs to the set $J_0 = \{l : \delta_l = 0\}$. It is evident that a portion of the data is missing, and therefore, the EM algorithm[24,25] is a natural choice to obtain the MLEs of the unknown model parameters.

Let i_l denote the realized value of $\widetilde{I}$ for l-th subject. Of course, the values of i_l are not observed for all subjects. Denoting the complete data by

$$\{(t_l,\, \delta_l,\, i_l) : l = 1,\, 2,\, \ldots,\, n\},$$

which includes the observed and unobserved data, the complete data likelihood function (up to a multiplicative constant) can be expressed as

$$L_c(\boldsymbol{\theta}; \boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{i}) \propto \left[\prod_{l\in J_1}\prod_{k=1}^{K}\{(1-p_0)\pi_k f_k(t_l;\boldsymbol{\theta}_k)\}^{I(\delta_l=k)}\right]\left[\prod_{l\in J_0}\{p_0\}^{1-i_l}\{(1-p_0)S_{mix}(t_l;\boldsymbol{\psi})\}^{i_l}\right],$$

where, $\boldsymbol{i} = (i_1,\, i_2,\, \ldots,\, i_n)$. The corresponding complete data log-likelihood function (up to an additive constant) is then given by

$$\begin{aligned}\ln L_c(\boldsymbol{\theta}; \boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{i}) \;=\; & r\ln(1-p_0) + \sum_{l\in J_1}\sum_{k=1}^{K} I(\delta_l = k)\ln\pi_k + \sum_{l\in J_1}\sum_{k=1}^{K} I(\delta_l = k)\ln f_k(t_l;\boldsymbol{\theta}_k)\\ & + \sum_{l\in J_0}(1-i_l)\ln p_0 + \sum_{l\in J_0} i_l \ln(1-p_0) + \sum_{l\in J_0} i_l \ln S_{mix}(t_l;\boldsymbol{\psi}),\end{aligned}$$

where $r = |J_1|$ denotes the number of subjects experienced the event of interest due to any of K modes. Following EM-based algorithm to estimate parameters of mixture of Weibull distributions[8], here, we propose an algorithm to obtain MLEs of model parameters. Each iteration of the algorithm consists of two steps, *viz.*, the E-step (expectation step) and the M-step (maximization step) as described below.

E-Step: In the E-step, we compute the expectation of the complete data log-likelihood function with respect to the conditional distribution of the unobserved $\widetilde{I}_l$'s, given the observed data $\boldsymbol{O} = \{(t_l, \delta_l) : l = 1,\, 2,\, \ldots,\, n\}$ at the current values of the model parameters. Therefore, at the $(m+1)$–st step, we need to compute $E\left(\widetilde{I}_l|\boldsymbol{O}; \boldsymbol{\theta}^{(m)}\right)$ for all $l \in J_0$, where $\boldsymbol{\theta}^{(m)}$ denotes the parameter estimate at the m–th step. Note that, for $l \in J_0$, the conditional distribution of $\widetilde{I}_l$ given observed data is Bernoulli with success probability $P(\widetilde{I}_l = 1|T_l > t_l)$. Thus, for $l \in J_0$,

$$w_l^{(m+1)} = E\left(\widetilde{I}_l|\boldsymbol{O}; \boldsymbol{\theta}^{(m)}\right) = P\left(\widetilde{I}_l = 1|T_l > t_l; \boldsymbol{\theta}^{(m)}\right) = \left.\frac{(1-p_0)S_{mix}(t_l;\boldsymbol{\psi})}{p_0 + (1-p_0)S_{mix}(t_l,\boldsymbol{\psi})}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(m)}}. \tag{6}$$

Thus, at $(m+1)$–st step, the conditional expectation, denoted by $Q^{(m+1)}(\boldsymbol{\theta})$, of the complete data log-likelihood function can be found by replacing i_l's by $w_l^{(m+1)}$ for all $l \in J_0$ in the complete data log-likelihood function and is given by

$$\begin{aligned}Q^{(m+1)}(\boldsymbol{\theta}) \;=\; & r\ln(1-p_0) + \sum_{l\in J_1}\sum_{k=1}^{K} I(\delta_l = k)\ln\pi_k + \sum_{l\in J_1}\sum_{k=1}^{K} I(\delta_l = k)\ln f_k(t_l;\boldsymbol{\theta}_k)\\ & + \sum_{l\in J_0}(1-w_l^{(m+1)})\ln p_0 + \sum_{l\in J_0} w_l^{(m+1)}\ln(1-p_0) + \sum_{l\in J_0} w_l^{(m+1)}\ln S_{mix}(t_l;\boldsymbol{\psi}).\end{aligned}$$

M-Step: In the M-step, maximization of $Q^{(m+1)}(\boldsymbol{\theta})$ is performed with respect to $\boldsymbol{\theta}$ over the corresponding parametric space to obtain an improved estimate of $\boldsymbol{\theta}$ as

$$\widehat{\boldsymbol{\theta}}^{(m+1)} = \text{argmax}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Q^{(m+1)}(\boldsymbol{\theta}).$$

As $\sum_{k=1}^{K}\pi_k = 1$, the objective function at the $(m+1)$–st step becomes

$$\widetilde{Q}^{(m+1)}(\boldsymbol{\theta}) = Q^{(m+1)}(\boldsymbol{\theta}) - \mu\left(\sum_{k=1}^{K}\pi_k - 1\right), \tag{7}$$

where μ is the Lagrange's multiplier. Therefore, the improved estimate at $(m+1)$–st step can be obtained as

$$\widehat{\boldsymbol{\theta}}^{(m+1)} = \text{argmax}_{\boldsymbol{\theta}\in\widetilde{\boldsymbol{\Theta}}}\widetilde{Q}^{(m+1)}(\boldsymbol{\theta}),$$

where $\widetilde{\boldsymbol{\Theta}} = \{\boldsymbol{\theta} : 0 < \pi_k < 1,\ \alpha_k > 0,\ \lambda_k > 0,\ k = 1, 2, \ldots, K,\ 0 < p_0 < 1\}$.

At the $(m+1)$–st step, the improved estimate of p_0 can explicitly be obtained by taking a partial derivative of $\widetilde{Q}^{(m+1)}(\boldsymbol{\theta})$ with respect to p_0 and equating it to zero. The improved estimate of p_0 at $(m+1)$–st step is given by

$$\widehat{p}_0^{(m+1)} = \frac{n - r - \sum_{l\in J_0} w_l^{(m+1)}}{n}. \tag{8}$$

Now, taking the partial derivative of (7) with respect to π_k and equating it to zero, we get

$$\frac{\partial\widetilde{Q}^{(m+1)}(\boldsymbol{\theta})}{\partial\pi_k} = \frac{r_k}{\pi_k} + \sum_{l\in J_0} w_l^{(m+1)}\frac{S_k(t_l;\boldsymbol{\theta}_k)}{S_{mix}(t_l;\boldsymbol{\psi})} - \mu = 0, \tag{9}$$

where r_k is the number of subjects experiencing the event of interest due to k–th mode, $k = 1, 2, \ldots, K$. Multiplying (9) by π_k and then summing over k, we obtain the improved estimate of μ as

$$\widehat{\mu}^{(m+1)} = r + \sum_{l\in J_0} w_l^{(m+1)}.$$

The conditional probability that a surviving subject will experience the event of interest due to mode k given that it survived until time τ is given by

$$P(k|\tau) = \frac{\pi_k S_k(\tau;\boldsymbol{\theta}_k)}{S_{mix}(\tau;\boldsymbol{\psi})}. \tag{10}$$

Let $P^{(m+1)}(k|\tau)$ denote the value of $P(k|\tau)$ evaluated at $\boldsymbol{\theta}^{(m)}$. Then multiplying (9) by π_k and comparing it with (10), we obtain the improved estimate of π_k as

$$\widehat{\pi}_k^{(m+1)} = \frac{1}{r + \sum\limits_{l\in J_0} w_l^{(m+1)}}\left[r_k + \sum_{l\in J_0} w_l^{(m+1)} P^{(m+1)}(k|t_l)\right]. \tag{11}$$

Equating the partial derivatives of (7) with respect to λ_k and α_k to zero, we get

$$\frac{\partial\widetilde{Q}^{(m+1)}(\boldsymbol{\theta})}{\partial\lambda_k} = \sum_{l\in J_1} I(\delta_l = k)\frac{\partial \ln f_k(t_l;\boldsymbol{\theta}_k)}{\partial\lambda_k} + \sum_{l\in J_0} w_l^{(m+1)} P^{(m+1)}(k|t_l)\frac{\partial \ln S_k(t_l;\boldsymbol{\theta}_k)}{\partial\lambda_k} = 0, \tag{12}$$

and

$$\frac{\partial\widetilde{Q}^{(m+1)}(\boldsymbol{\theta})}{\partial\alpha_k} = \sum_{l\in J_1} I(\delta_l = k)\frac{\partial \ln f_k(t_l;\boldsymbol{\theta}_k)}{\partial\alpha_k} + \sum_{l\in J_0} w_l^{(m+1)} P^{(m+1)}(k|t_l)\frac{\partial \ln S_k(t_l;\boldsymbol{\theta}_k)}{\partial\alpha_k} = 0, \tag{13}$$

respectively. Solving (12) for λ_k, we obtain the improved estimator of λ_k in terms of α_k as

$$\widehat{\lambda}_k^{(m+1)}(\alpha_k) = \frac{r_k}{\sum_{l\in J_1} I(\delta_l = k)t_l^{\alpha_k} + \sum_{l\in J_0} w_l^{(m+1)} P^{(m+1)}(k|t_l)t_l^{\alpha_k}}. \tag{14}$$

Plugging in $\widehat{\lambda}_k^{(m+1)}(\alpha_k)$ from (14) into (13) and simplifying, we get

$$\frac{r_k}{\alpha_k} + \sum_{l\in J_1} I(\delta_l = k)\ln t_l - r_k\frac{\sum_{l\in J_1} I(\delta_l = k)t_l^{\alpha_k}\ln t_l + \sum_{l\in J_0} w_l^{(m+1)} P^{(m+1)}(k|t_l)t_l^{\alpha_k}\ln t_l}{\sum_{l\in J_1} I(\delta_l = k)t_l^{\alpha_k} + \sum_{l\in J_0} w_l^{(m+1)} P^{(m+1)}(k|t_l)t_l^{\alpha_k}} = 0. \tag{15}$$

Theorem 1. *Define, for* $k = 1, 2, \ldots, K$ *and* $\alpha > 0$,

$$g_k(\alpha) = \frac{r_k}{\alpha} + \sum_{l \in J_1} I(\delta_l = k) \ln t_l - r_k \frac{\sum_{l \in J_1} I(\delta_l = k) t_l^{\alpha} \ln t_l + \sum_{l \in J_0} w_l^{(m+1)} P^{(m+1)}(k|t_l) t_l^{\alpha} \ln t_l}{\sum_{l \in J_1} I(\delta_l = k) t_l^{\alpha} + \sum_{l \in J_0} w_l^{(m+1)} P^{(m+1)}(k|t_l) t_l^{\alpha}}.$$

Then, for all k, $g_k(\alpha)$ *is a strictly decreasing function of* α *and the equation* $g_k(\alpha) = 0$ *has exactly one solution in* $(0, \infty)$.

Proof. See Appendix A. □

Due to Theorem 1, the improved estimate of α_k, say $\widehat{\alpha}_k^{(m+1)}$, at $(m+1)$–st step can be found by solving (15) for α_k using any numerical method (for example bisection method or Newton-Rapshon method). Then replacing α_k by $\widehat{\alpha}_k^{(m+1)}$ in (14), one can compute $\widehat{\lambda}_k^{(m+1)}$. The E and M steps are continued iteratively until the desired convergence is obtained. One can use different criteria to check if the updated estimates obtained in two consecutive M steps are close to each other. Once they are sufficiently close, the iterative process may be terminated. Here, in this article, for fixed pre-assigned value of $\varepsilon > 0$, we stop the iterative process as soon as the following criterion is satisfied:

$$\sum_k \left| \widehat{\theta}_k^{(i)} - \widehat{\theta}_k^{(i+1)} \right| < \varepsilon, \tag{16}$$

where $\widehat{\theta}_k^{(i)}$ is the estimate of the k-th parameter at the end of the i-th implementation of MLE/EM algorithm. To be precise, the MLE/EM algorithm for computing the MLEs of the model parameters of a finite mixture of Weibull distributions involving competing risks, cure fraction, and random censoring can be summarized in the Algorithm 1.

Algorithm 1 MLE/EM Algorithm

Start with an initial guess of $p_0^{(0)}$, $\pi_k^{(0)}$, $\lambda_k^{(0)}$, $\alpha_k^{(0)}$ for $k = 1, 2, \ldots, K$.
Set $m = 0$.
repeat
  Calculate $P^{(m+1)}(k|t_l)$ using (10) for all $k = 1, 2, \ldots, K$ and $l \in J_0$.
  Calculate $w_l^{(m+1)}$ for $l \in J_0$ using (6).
  Compute the improved estimates of p_0 using (8).
  For $k = 1, 2, \ldots, K$, compute the improved estimates of π_k using (11).
  For $k = 1, 2, \ldots, K$, compute the improved estimates of α_k by solving (15).
  For $k = 1, 2, \ldots, K$, compute the improved estimates of λ_k using (14).
  Set $m = m + 1$.
until Convergence is achieved.

4.3 | Asymptotic Confidence Interval

The asymptotic confidence intervals of different model parameters can be obtained using asymptotic normality of MLEs. For this purpose, one needs the asymptotic variance-covariance matrix of the MLE of $\boldsymbol{\theta}$, which may be found by inverting the Fisher information matrix based on observed data. The missing value principle[26] can be used to obtain the Fisher information matrix. However, it requires the calculation of the conditional (on observed data) expectation of the complete-data information matrix and of the complete-data score statistics, which are quite difficult in many cases. In the case of independent and identically distributed sample, the observed data Fisher information matrix can be approximated by empirical observed information matrix.[27,25] The empirical observed Fisher information matrix for independent and identically distributed sample is given by

$$\boldsymbol{I}_e\left(\widehat{\boldsymbol{\theta}}\right) = \sum_{l=1}^{n} \boldsymbol{s}_l(\widehat{\boldsymbol{\theta}}) \boldsymbol{s}_l^T(\widehat{\boldsymbol{\theta}}),$$

where $\boldsymbol{s}_l(\boldsymbol{\theta})$ is the conditional (on observed data) expectation of the first derivative (with respect to parameters $\boldsymbol{\theta}$) of log-likelihood contribution of l-th complete data point.

The detailed calculation of different entries of $\boldsymbol{s}_l(\widehat{\boldsymbol{\theta}})$ is provided in Appendix B. Once the empirical Fisher information matrix $\boldsymbol{I}_e(\widehat{\boldsymbol{\theta}})$ is obtained, the standard error of MLE of θ_j (j-th entry of $\boldsymbol{\theta}$) may be approximated by the positive square-root of j-th diagonal of the inverse of $\boldsymbol{I}_e(\widehat{\boldsymbol{\theta}})$. Therefore, using asymptotic normality of MLE, a $100(1-\gamma)\%$ asymptotic confidence interval of θ_j is constructed as

$$\left(\widehat{\theta}_j - z_{\frac{\gamma}{2}}\sqrt{V_j},\, \widehat{\theta}_j + z_{\frac{\gamma}{2}}\sqrt{V_j}\right),$$

where $\widehat{\theta}_j$ denotes the MLE of θ_j, V_j is the j-th diagonal of the inverse of $\boldsymbol{I}_e(\widehat{\boldsymbol{\theta}})$ and z_γ denotes the upper γ-point of standard normal distribution.

5 | SIMULATION STUDY

In this section, we conduct a simple simulation experiment to investigate the efficacy of the methods proposed in the previous section. For simplicity, we assume two modes of occurrence of the event of interest. Although the algorithm is exclusively designed for two failure modes, it can easily be extended to accommodate more than two modes. A sample of size n is generated from the survival function given in (3) with $K = 2$ under a random censoring scheme. The random censoring scheme is implemented by generating censoring times from a uniform distribution. This procedure is carried out M times to obtain M simulated data sets each of size n.

TABLE 1 Simulation results for different choices of model parameters

n	Parameter	True Value	AE	ASE	CP	True Value	AE	ASE	CP
200	p_0	0.30	0.301	0.033	0.945	0.30	0.300	0.033	0.947
	π_1	0.40	0.399	0.041	0.951	0.40	0.399	0.041	0.948
	α_1	0.80	0.822	0.092	0.953	1.20	1.236	0.131	0.953
	λ_1	1.00	1.004	0.151	0.955	1.00	1.010	0.147	0.946
	α_2	0.60	0.609	0.053	0.954	1.50	1.530	0.132	0.949
	λ_2	1.50	1.406	0.147	0.951	1.50	1.417	0.144	0.948
400	p_0	0.30	0.300	0.023	0.952	0.30	0.300	0.023	0.950
	π_1	0.40	0.399	0.029	0.951	0.40	0.400	0.028	0.955
	α_1	0.80	0.809	0.063	0.950	1.20	1.218	0.092	0.948
	λ_1	1.00	1.009	0.106	0.952	1.00	1.007	0.103	0.952
	α_2	0.60	0.604	0.036	0.949	1.50	1.517	0.091	0.952
	λ_2	1.50	1.422	0.116	0.946	1.50	1.428	0.138	0.948
200	p_0	0.60	0.600	0.035	0.948	0.60	0.600	0.035	0.951
	π_1	0.65	0.649	0.052	0.956	0.65	0.650	0.053	0.952
	α_1	0.80	0.821	0.095	0.957	1.20	1.236	0.138	0.953
	λ_1	1.00	1.021	0.159	0.948	1.00	1.015	0.153	0.946
	α_2	0.60	0.634	0.098	0.952	1.50	1.589	0.242	0.948
	λ_2	1.50	1.412	0.153	0.943	1.50	1.428	0.150	0.947
400	p_0	0.60	0.599	0.025	0.949	0.60	0.600	0.024	0.953
	π_1	0.65	0.648	0.038	0.948	0.65	0.650	0.037	0.949
	α_1	0.80	0.812	0.065	0.952	1.20	1.215	0.096	0.945
	λ_1	1.00	1.008	0.111	0.949	1.00	1.011	0.107	0.952
	α_2	0.60	0.616	0.066	0.950	1.50	1.542	0.164	0.954
	λ_2	1.50	1.429	0.146	0.947	1.50	1.440	0.143	0.947

We obtain the MLEs of the model parameters for each simulated data set using the proposed MLE/EM algorithm. The standard errors of MLEs of model parameters are also obtained using the Fisher information matrix. As performance measures of the proposed MLE/EM algorithm, the average estimate (AE) and the average standard error (ASE) of model parameters are computed by taking the average of estimates and the standard errors, respectively, over M simulated data sets. We also compute

asymptotic 95% confidence intervals for each model parameter based on the normal approximation for each simulated data set. Finally, the corresponding coverage percentages (CP) for asymptotic 95% confidence intervals are computed.

For this simulation study, we consider different sample sizes ranging from moderate to large (n = 200 and 400), different choices of cured fractions (p_0 = 0.3, 0.6) and mixing weights (π_1 = 0.40, 0.65). For the parameters of the Weibull distributions, we take two choices for shape parameters, *viz.*, (α_1, α_2) = (0.8, 0.6) (decreasing failure rate) and (1.2, 1.5) (increasing failure rate) with the common scale parameters (λ_1, λ_2) = (1, 1.5). All results are presented in Table 1 based on M = 5000 simulated data sets. The following observations are quite clear from these tables. As expected, the AEs of each model parameter come closer to the true values with the increase in sample size. It is also observed that the corresponding ASE decreases as the sample size increases. These show empirical validation of the consistency of MLEs as estimators of the corresponding model parameters. The CPs converge to 0.95 with increased sample size, providing empirical evidence for asymptotic normality.

6 | MODEL WITH COVARIATES

In many survival data, information on different characteristics, called covariates or features, are available along with information on lifetimes. In such cases, it may be reasonable to assume that the probability of a subject being cured depends on a set of covariates corresponding to the subject. Binary regression models, for example logistic link function[28,29,30,31,32] or probit link function[33], may be used for this purpose. Suppose that on each subject, information on L covariates are recorded. For the l–th subject, l = 1, 2, ..., n, we propose to relate the cured fraction to the covariates $\boldsymbol{x}_l = (1, x_{l1}, x_{l2}, \ldots, x_{lL})$ by the logistic link function, *i.e.*,

$$p_0(\boldsymbol{x}_l;\, \boldsymbol{\beta}) = \frac{1}{1 + \exp(\boldsymbol{x}_l^T \boldsymbol{\beta})},$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_L)$ denotes the vector of regression coefficients.

In many studies, the analyst is interested in one of the modes of failure. For example, while analyzing melanoma data set described in Section 2, one may be interested in the distribution of survival time due to melanoma cancer and its dependence on a set of covariates. Here, we assume that for the l-th subject, l = 1, 2, ..., n, the scale parameter of the weibull distribution corresponding to the mode (say mode 1) of interest is related to covariates as

$$\lambda_{1l}(\boldsymbol{x};\, \boldsymbol{\eta}) = \exp\left(\boldsymbol{\eta}^T \boldsymbol{x}_l\right), \tag{17}$$

where $\boldsymbol{\eta} = (\eta_0, \eta_1, \ldots, \eta_L)$ is the vector of parameters.

Denote ($\alpha_1, \ldots, \alpha_K, \boldsymbol{\eta}, \lambda_2, \ldots, \lambda_K, \pi_1, \ldots, \pi_K$) by $\boldsymbol{\zeta}$ and let $\boldsymbol{\gamma} = (\boldsymbol{\zeta}, \boldsymbol{\beta})$. At the $(m+1)$–st step of MLE/EM algorithm, the expectation of complete data log-likelihood function can now be expressed as

$$\widetilde{Q}^{(m+1)}(\boldsymbol{\gamma}) = \widetilde{Q}_1^{(m+1)}(\boldsymbol{\beta}) + \widetilde{Q}_2^{(m+1)}(\boldsymbol{\zeta}),$$

where

$$\widetilde{Q}_1^{(m+1)}(\boldsymbol{\beta}) = \sum_{l \in J_1} \boldsymbol{x}_l^T \boldsymbol{\beta} + \sum_{l \in J_0} w_l^{(m+1)} \boldsymbol{x}_l^T \boldsymbol{\beta} - \sum_{l=1}^{n} \ln(1 + \exp(\boldsymbol{x}_l^T \boldsymbol{\beta})),$$

$$\widetilde{Q}_2^{(m+1)}(\boldsymbol{\zeta}) = \sum_{l \in J_1} \sum_{k=1}^{K} I(\delta_l = k) \ln \pi_k + \sum_{l \in J_1} I(\delta_l = 1) \left\{ \ln \alpha_1 + \boldsymbol{\eta}^T \boldsymbol{x}_l + (\alpha_1 - 1) \ln t_l - t_l^{\alpha_1} \exp\left(\boldsymbol{\eta}^T \boldsymbol{x}_l\right) \right\}$$
$$+ \sum_{l \in J_1} \sum_{k=2}^{K} I(\delta_l = k) \ln f_k(t_l; \boldsymbol{\theta}_k) + \sum_{l \in J_0} w_l^{(m+1)} \ln S_{mix}(t_l; \boldsymbol{\zeta}) - \mu \left(\sum_{k=1}^{K} \pi_k - 1 \right),$$

and

$$w_l^{(m+1)} = \left. \frac{\exp(\boldsymbol{x}_l^T \boldsymbol{\beta}) S_{mix}(t_l; \boldsymbol{\zeta})}{1 + \exp(\boldsymbol{x}_l^T \boldsymbol{\beta}) S_{mix}(t_l; \boldsymbol{\zeta})} \right|_{\boldsymbol{\gamma} = \widehat{\boldsymbol{\gamma}}^{(m)}} \quad \text{for } l \in J_0.$$

Here $\widehat{\boldsymbol{\gamma}}^{(m)}$ is the improved estimate of $\boldsymbol{\gamma}$ after m–th step. It is to be noted that $\widetilde{Q}^{(m+1)}(\boldsymbol{\gamma})$ can be separated into two constituent parts: $\widetilde{Q}_1^{(m+1)}(\boldsymbol{\beta})$ depending only on the vector of regression coefficients $\boldsymbol{\beta}$ and $\widetilde{Q}_2^{(m+1)}(\boldsymbol{\zeta})$ depending only on the parameter vector $\boldsymbol{\zeta}$. Thus, maximization of $\widetilde{Q}^{(m+1)}(\boldsymbol{\gamma})$ in the M-step can be carried out by maximizing $\widetilde{Q}_1^{(m+1)}(\boldsymbol{\beta})$ and $\widetilde{Q}_2^{(m+1)}(\boldsymbol{\zeta})$ separately with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\zeta}$, respectively. The improved estimate of $\boldsymbol{\beta}$ at the $(m+1)$–st step of MLE/EM algorithm can be found by maximizing $\widetilde{Q}_1^{(m+1)}(\boldsymbol{\beta})$ using a numerical methods. Following the method described in Section 4.2, the maximization of $Q_2^{(m+1)}(\boldsymbol{\zeta})$ can be performed by solving K systems of non-linear equations - one involving $(L+2)$ variables $(\alpha_1,\ \boldsymbol{\eta})$ and each one of the rests involving one variable α_k; $k = 2, 3, \ldots, K$. Once the MLE of $\boldsymbol{\gamma}$ is obtained, the MLE of cure fraction can be found using the invariance property of the MLE. Denoting the MLE of $\boldsymbol{\beta}$ by $\widehat{\boldsymbol{\beta}}$, the MLE of cure fraction for a subject with covariate $\boldsymbol{x}$ is given by

$$\widehat{p}_0(\boldsymbol{x}) = p_0(\boldsymbol{x};\, \widehat{\boldsymbol{\beta}}) = \frac{1}{1+\exp(\boldsymbol{x}^T\widehat{\boldsymbol{\beta}})}. \tag{18}$$

Similarly, the MLE of λ_1 corresponding to a subject with covariate $\boldsymbol{x}$ can be obtained as

$$\widehat{\lambda}_1(\boldsymbol{x}) = \exp\left(\boldsymbol{x}^T\widehat{\boldsymbol{\eta}}\right),$$

where $\widehat{\boldsymbol{\eta}}$ is the MLE of $\boldsymbol{\eta}$. Note that the model described here can be easily extended in the sense that covariates can be added to cause-specific distributions if more than one specific modes of failure are available.

One can construct asymptotic confidence intervals of model parameters following the method described in Section 4.3. The different entries of $\boldsymbol{s}_l(\widehat{\boldsymbol{\gamma}})$, the conditional (on observed data) expectation of the first derivative (with respect to parameters $\boldsymbol{\gamma}$) of log-likelihood contribution of l-th complete observation, are listed in Appendix C. Furthermore, the asymptotic confidence interval of the cure fraction or λ_1 can be obtained using δ-method. For example, the distribution of $\widehat{p}_0(\boldsymbol{x})$ can be approximated by a normal distribution with mean $p_0(\boldsymbol{x})$ and variance $\boldsymbol{a}'\boldsymbol{I}_e^{-1}(\widehat{\boldsymbol{\beta}})\boldsymbol{a}$ for large sample sizes, where $\boldsymbol{a} = \left(\frac{\partial h(\beta)}{\partial\beta_0}, \ldots, \frac{\partial h(\beta)}{\partial\beta_k}\right)\Big|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}}$ and $\boldsymbol{I}_e(\boldsymbol{\beta})$ is the empirical Fisher information matrix of $\boldsymbol{\beta}$.

7 | ANALYSIS OF MELANOMA DATA

In this section, we demonstrate an application of the proposed estimation methodology based on malignant melanoma cancer data introduced in Section 2. As mentioned, the data set, named `melanoma`, is taken from `R` package `boot`. The data consist of possibly right censored survival times (in days) of 205 patients after operation for malignant melanoma cancer along with information on the cause of death, tumor thickness, presence of ulcer, age at operation, year of operation, and gender for each patient. Among these covariates, age at operation, year of operation and tumor thickness are continuous covariates.

For computation purposes, we divide the survival times, which are in days, by 365; thus, convert the unit of survival times into years. We also divide age of operation by 10 and center year of operation at 1970 and then scale it by 10. To analyze this data set, we level the death due to malignant melanoma as mode 1 and death due to any other causes as mode 2. As comorbidities may accelerate the progress of cancer, it would not be appropriate to model the data set using independent latent failure times without considering a dependence mechanism between latent lifetimes corresponding to different modes. Here, we model the data set using a mixture cure-rate model, where the cure fraction is connected to all five covariates by a logistic link function, and the latency distribution is a proper mixture of two Weibull distributions. It is further assumed that the scale parameter of the Weibull distribution corresponding to melanoma cancer depends on all covariates through (17) with coefficients η_1, η_2, η_3, η_4, and η_5 correspond to tumor thickness, ulceration status, age at operation, year of operation, and sex, respectively. The regression coefficients in the logistic regression corresponding to tumor thickness, ulceration status, age at operation, year of option, and sex are denoted by β_1, β_2, β_3, β_4, and β_5, respectively. Finding meaningful initial guesses of model parameters is crucial for implementing MLE/EM algorithm. A method for the same is described in Appendix D. The initial guesses of different model parameters obtained using the method and are reported in Table 2.

The maximum likelihood estimates of different parameters are obtained by implementing MLE/EM algorithm. The E and M steps are executed iteratively till the condition given in (16) is satisfied with $\varepsilon = 10^{-10}$. The MLEs of different parameters are reported in Table 2. The trace plots for different parameters and the plot of log-likelihood function corresponding to MLE/EM iterations are given in Appendix E. These plots indicate stabilization of the updated estimates and show that the log-likelihood function increases over MLE/EM iterations. The 95% asymptotic confidence intervals of model parameters are also computed using the asymptotic normality of MLEs and are reported in the same table. As asymptotic 95% confidence intervals of regression

TABLE 2 Point and interval estimates of different model parameters based on malignant melanoma data when cure rate and scale parameter depend on covariates

Parameters	Initial Guess	MLE	95% Confidence Intervals
α_1	1.291	1.644	1.120 – 2.167
α_2	0.520	0.677	0.355 – 0.998
η_0 (Intercept)	–3.320	–4.447	–7.019 – –1.875
η_1 (Thickness)	–0.004	0.146	0.060 – 0.232
η_2 (Ulcer)	–0.021	1.162	–0.006 – 2.330
η_3 (Age)	0.006	–0.049	–0.327 – 0.230
η_4 (Year)	0.133	–1.787	–3.549 – –0.024
η_5 (Sex)	–0.066	0.731	–0.124 – 1.586
λ_2	0.029	0.089	–0.007 – 0.186
π_1	0.803	0.669	0.419 – 0.919
π_2	0.197	0.331	0.111 – 0.550
β_0 (Intercept)	0.000	–3.593	–7.555 – 0.368
β_1 (Thickness)	0.000	0.003	–0.372 – 0.377
β_2 (Ulcer)	0.000	1.696	–0.773 – 4.165
β_3 (Age)	0.000	0.698	0.077 – 1.319
β_4 (Year)	0.000	0.521	–4.553 – 5.595
β_5 (Sex)	0.000	1.544	–0.422 – 3.511

parameters η_2, η_3, η_5, β_0, β_1, β_2, β_4, and β_5 include zero, all these regression coefficients are statistically insignificant; the rest (η_0, η_1, η_4, and β_3) are statistically significant. Thus, tumor thickness and year of operation are associated with death due to melanoma cancer, and age at operation is associated with incidence. It may be mentioned here that in their study using vertical modelling, Nicolaie et al.[17] established that tumor thickness, ulceration status, and patient sex are significantly associated with melanoma mortality through the latency part of the model, ulceration and sex are associated with the incidence part.

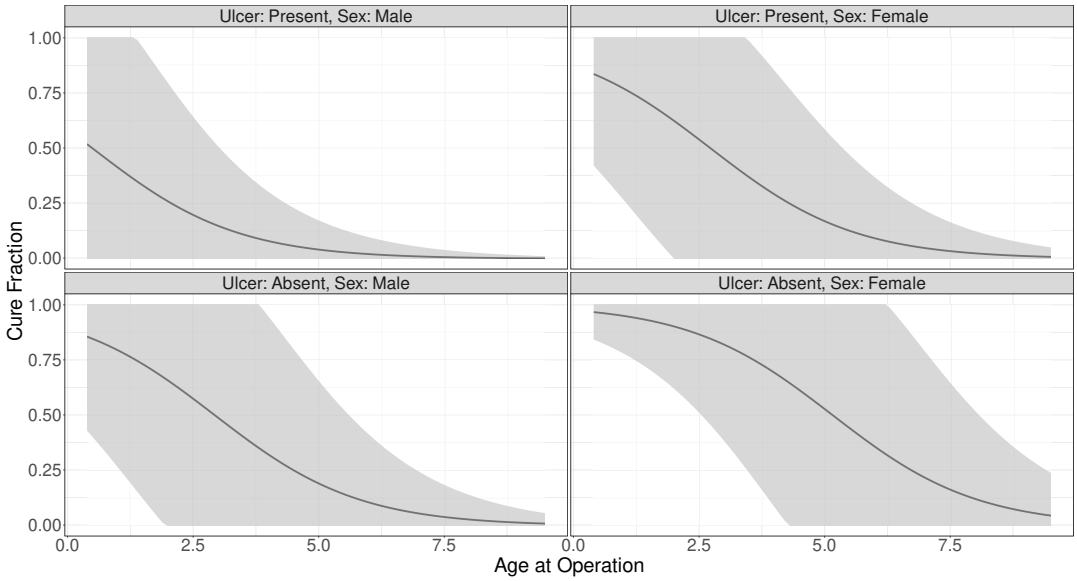

FIGURE 2 Plots of the estimated cure fraction with respect to age at operation for mean values of other continuous covariates

Note that ulceration status and sex are binary covariates having two levels each. For each of the four combinations of these two covariates, the estimated cure fractions are computed for different values of age at operation for mean values of other continuous covariates using (18) and plotted in Figure 2. The corresponding point-wise 95% confidence intervals of cure fraction are also computed using δ-method and shown as a gray shaded area in the same figure. This figure shows that for a fixed value of age at operation, the presence of ulcer reduces the probability of being cured after the operation across both sexes. On the other hand, the probability of a cure is higher for female patients than males, keeping other covariates fixed.

Figures 3 and 4 show the estimated conditional (on $\tilde{I} = 1$) and unconditional survival functions, calculated using (2) and (3), respectively, of time to any event for patients with mean values of the continuous covariates for different ulceration status and sex. Figure 5 shows the estimated conditional (on $\tilde{I} = 1$) cumulative incidence function of death due to melanoma for susceptible patients with mean values of the continuous covariates for different ulceration status and sex. The estimated unconditional incidence function of death due to melanoma for patients with same values of covariates are given in Figure 6. In these figures, the uncertainty of the estimates is depicted by the gray shaded corresponding 95% point-wise confidence intervals. Note that the conditional cumulative incidence function and unconditional incidence function of a cause, say cause 1, is defined by

$$F_1(t \mid \tilde{I} = 1, \boldsymbol{x}) = P\left(\tilde{T} \leq t \mid \tilde{N} = 1, \tilde{I} = 1, \boldsymbol{x}\right) = \pi_1 \left(1 - S_1(t; \boldsymbol{x})\right)$$

and

$$F_1(t \mid \boldsymbol{x}) = P\left(\tilde{T} \leq t \mid \tilde{N} = 1, \boldsymbol{x}\right) = \left(1 - p_0(\boldsymbol{x})\right) \pi_1 \left(1 - S_1(t; \boldsymbol{x})\right),$$

respectively. The estimated values of cure fractions for mean values of continuous covariates for different values of ulceration status and gender are reported in Table 3. As the estimated cure fraction for a male patient with ulcer and mean values of continuous covariates are quite small, the conditional and unconditional cumulative incidence functions of death due to melanoma are quite similar. However, cumulative incidence functions are not so similar for a female patient with no ulcer as the estimated cure fraction in this case is larger.

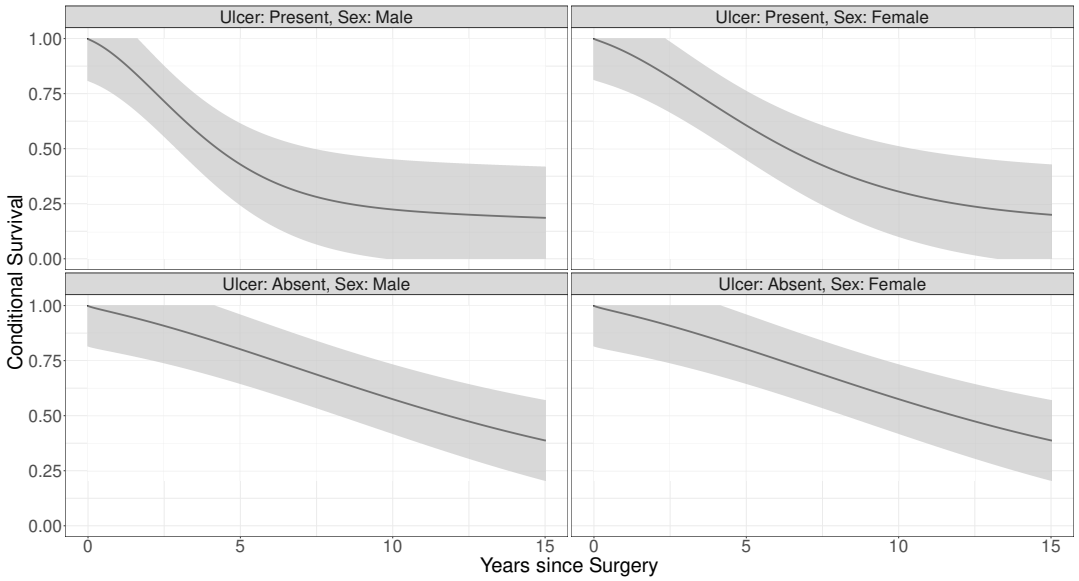

FIGURE 3 Plot of the estimated conditional (on $\tilde{I} = 1$) survival function of time to any event at mean values of continuous covariates

TABLE 3 The estimated cure fractions with mean values of continuous covariates for different ulceration status and gender

Ulceration Status	Sex	Estimated Cure Fraction
Present	Male	0.035
Present	Female	0.146
Absent	Male	0.166
Absent	Female	0.482

Predictions are of natural interest in survival analysis. For instance, the estimated survival probability at a future time may be important to the medical personnel for a right censored patient. Let the right censoring time for a patient be τ_R. We are interested

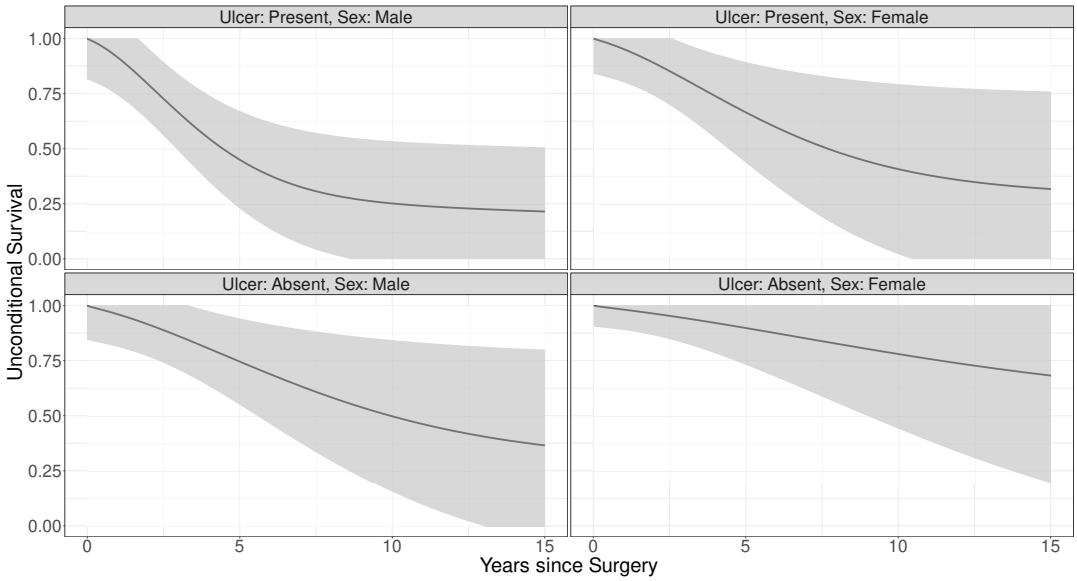

FIGURE 4 Plot of the estimated unconditional survival function of time to any event at mean values of continuous covariates

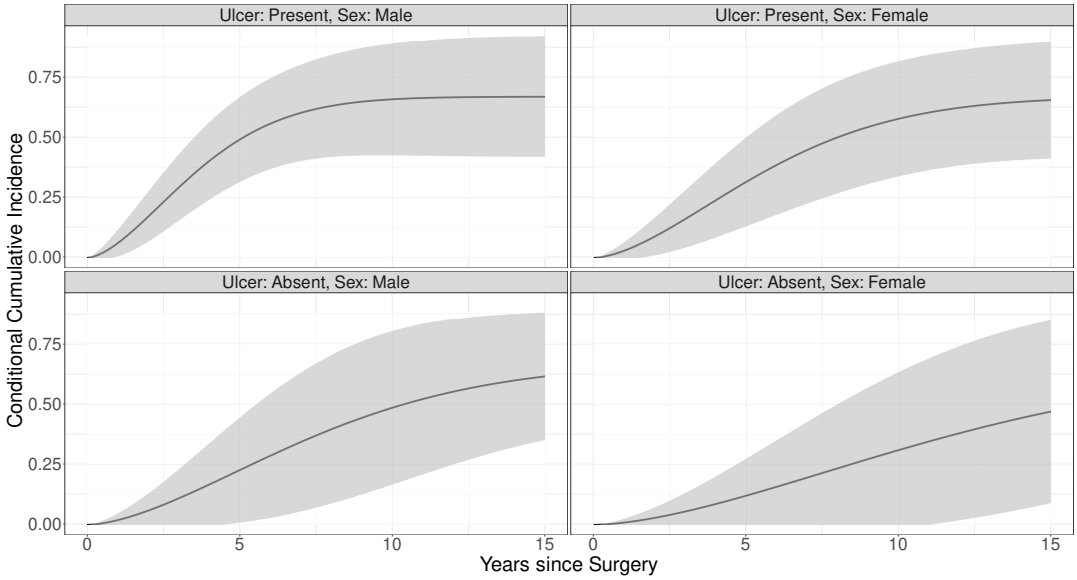

FIGURE 5 Plot of the conditional (on $\tilde{I} = 1$) cumulative incidence function for death due to melanoma at mean values of continuous covariates

TABLE 4 Right censoring times and the value of covariates for two selected patients

Patient Identity Number	Right censoring time	Tumor thickness	Ulcer status	Age	Year	Sex
55	4.266	1.29	Absent	2.6	–0.1	Female
161	8.712	1.29	Absent	2.5	0.3	Female

in estimating the conditional survival probability at a future time τ_f given that the patient has survived till τ_R and the patient belongs to the susceptible population, where $\tau_f > \tau_R$. Thus, denoting the lifetime of the patient by $\widetilde{T}$, the probability under

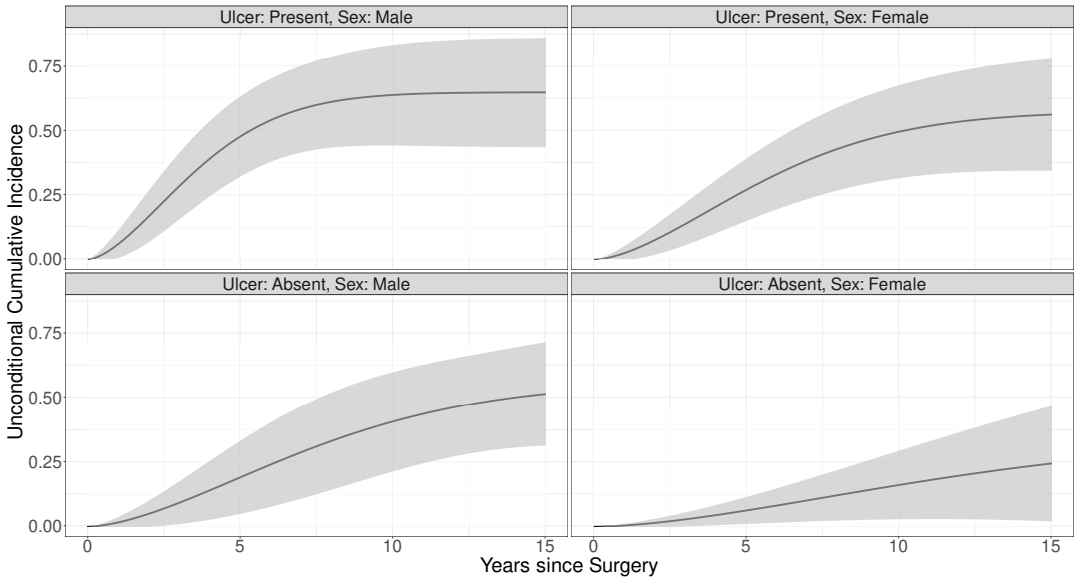

FIGURE 6 Plot of the estimated unconditional cumulative incidence function for death due to melanoma at mean values of continuous covariates

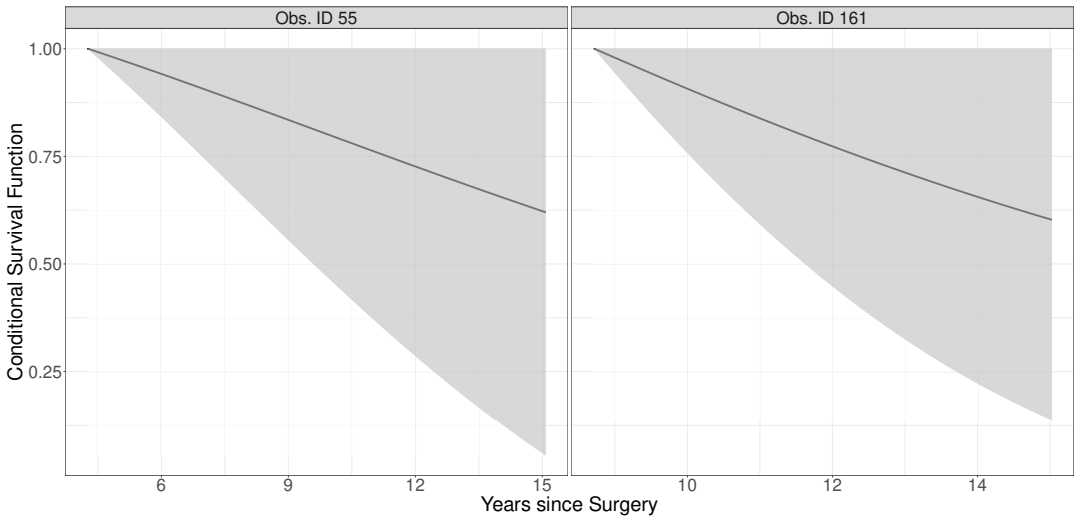

FIGURE 7 Plots of the estimated conditional survival functions for right censored patients with ID 55 and 161

consideration is

$$\kappa\left(\tau_R,\, \tau_f;\, \boldsymbol{\zeta}\right) = P\left(\widetilde{T} > \tau_f \mid \widetilde{T} > \tau_R \text{ and } \widetilde{I} = 1;\, \boldsymbol{\zeta}\right) = \frac{S_{mix}\left(\tau_f;\, \boldsymbol{\zeta}\right)}{S_{mix}\left(\tau_R;\, \boldsymbol{\zeta}\right)}.$$

Note that the above conditional survival probability does not depend of $\boldsymbol{\beta}$. Now, using the invariance property of MLE, the MLE of the above conditional probability is obtained by replacing the parameter $\boldsymbol{\zeta}$ by its' MLE $\widehat{\boldsymbol{\zeta}}$ and is given by

$$\widehat{\kappa}\left(\tau_R,\, \tau_f;\, \boldsymbol{\zeta}\right) = \kappa\left(\tau_R,\, \tau_f;\, \widehat{\boldsymbol{\zeta}}\right) = \frac{S_{mix}\left(\tau_f;\, \widehat{\boldsymbol{\zeta}}\right)}{S_{mix}\left(\tau_R;\, \widehat{\boldsymbol{\zeta}}\right)}.$$

The estimated conditional survival probabilities for patients with identity numbers 55 and 161 are computed using the above formula for different future times and are plotted in Figure 7. The corresponding point-wise 95% confidence intervals are obtained using δ-method and are also shown in the same figure as gray shaded region. The right censoring time, along with the

values of covariates are given in Table 4 for these two patients. Note that these two female patients without ulcer have same value of tumor thickness, however, the values of other covariates are close. The estimated scale parameters (of Weibull distributions) corresponding to melanoma cancer for these two patients are 0.007 and 0.015, respectively.

8 | CONCLUSION

In this paper, we consider an analysis of time-to-event data when the event of interest occurs due to one of multiple failure modes in the presence of long-term survivors and random censoring. We use a finite mixture of Weibull distributions to model the susceptible population. This modelling approach enjoys the flexibility of Weibull distribution on the one hand and can handle dependence among different modes of failure on the other hand. For mixture of K Weibull distributions, usual likelihood inference requires maximization of a $3K + 1$ dimensional optimization problem when no covariates are present. If covariates are present for all subjects, we propose to link cure rate parameter to covariates using a logistic function. We also propose to model the scale parameter of Weibull distribution corresponding to the mode of interest using a log-linear function of covariates. The dimension of the problem of finding MLEs of model parameters increased to $3K + 2L + 1$ when information on L covariates are incorporated in modelling. Here, we propose an efficient EM-based algorithm, which can efficiently estimate model parameters by reducing the dimension of the optimization problems involved. When no covariates are considered, one needs to solve K one-dimensional optimization problem in each iteration of the MLE/EM algorithm. If L covariates are involved in the model, one needs to solve $K + 1$ systems of non-linear equations - one involving $L + 1$ parameters, another involving $L + 2$ parameters and each of rests involving one parameter - in each iteration of the MLE/EM algorithm. Simulation results reveal that the proposed estimation procedure works well for different sets of parameters and different sample sizes. Finally, a real data set on malignant melanoma patients is analyzed. We also discuss a method of estimating survival probability at a future time of a right censored observation. Note that if more than one modes are of interest, one can easily extend the proposed model by connecting parameters of respective Weibull distributions to covariates through log-linear functions. An alternative way of modelling the same problem would be to use a $K + 1$ mixture distribution:

$$S(t) = p_0 + \sum_{k=1}^{K} p_k S_k(t),$$

where $p_0 + \sum_{k=1}^{K} p_k = 1$. The covariates can be related to mixing probabilities p_k's using a multinomial logistic regression. More works in this regards are needed and we are hoping to report the same in our future manuscript.

ACKNOWLEDGMENTS

The authors are grateful to the anonymous reviewers and the editors for their many constructive comments and useful suggestions that have helped us in improving the earlier version of the manuscript.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

REFERENCES

1. Cox DR. The analysis of exponentially distributed lifetimes with two types of failures. *Journal of the Royal Statistical Society. Series B (Methodological).* 1959;21:411–421.
2. Prentice RL, Kalbfleisch JD, Peterson AV, Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. *Biometrics.* 1978;34:541–554.
3. Lawless JF. *Statistical models and methods for lifetime data*. John Wiley & Sons, New York, 1982.
4. Crowder MJ. *Classical Competing Risks*. Chapman and Hall,London, 2001.
5. Babu GJ, Rao CR, Rao MB. Nonparametric estimation of specific occurrence/exposure rate in risk and survival analysis. *Journal of the American Statistical Association.* 1992;87(417):84–89.
6. Kundu D, Kannan N, Mazumdar M. Inference on risk rates based on mortality data under censoring and competing risks using parametric models. *Biometrical Journal.* 1992;34:315–328.
7. McLachlan GJ, Peel D. *Finite mixture models*. John Wiley & Sons, New York, 2004.
8. Jiang S, Kececioglu D. Maximum likelihood estimates, from censored data, for mixed-Weibull distributions. *IEEE Transactions on Reliability.* 1992;41(2):248–255.
9. Maller RA, Zhou X. *Survival analysis with long-term survivors*. John Wiley & Sons, New York, 1996.
10. Boag JW. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B (Methodological).* 1949;11:15–33.

11. Rodrigues J, Castro M, Balakrishnan N, Cancho VG. Destructive weighted Poisson cure rate models. *Lifetime data analysis.* 2011;17:333–346.
12. Balakrishnan N, Pal S. EM algorithm-based likelihood estimation for some cure rate models. *Journal of Statistical Theory and Practice.* 2012;6:698–724.
13. Balakrishnan N, Pal S. Lognormal lifetimes and likelihood-based inference for flexible cure rate models based on COM-Poisson family. *Computational Statistics & Data Analysis.* 2013;30:41–67.
14. Balakrishnan N, Pal S. An EM algorithm for the estimation of parameters of a flexible cure rate model with generalized gamma lifetime and model discrimination using likelihood-and information-based methods. *Computational Statistics.* 2015;30:151–189.
15. Wiangnak P, Pal S. Gamma lifetimes and associated inference for interval-censored cure rate model with COM Poisson competing cause. *Communications in Statistics-Theory and Methods.* 2018;47:1491–1509.
16. Chao EC. Gibbs sampling for long-term survival data with competing risks. *Biometrics.* 1998:350–366.
17. Nicolaie MA, Taylor JM, Legrand C. Vertical modeling: analysis of competing risks data with a cure fraction. *Lifetime data analysis.* 2019;25(1):1–25.
18. Drzewiecki K, Ladefoged C, Christensen H. Biopsy and prognosis for cutaneous malignant melanomas in clinical stage I. *Scandinavian journal of plastic and reconstructive surgery.* 1980;14(2):141–144.
19. Andersen PK, Borgan O, Gill RD, Keiding N. *Statistical models based on counting processes*. Springer Science & Business Media, 2012.
20. Canty A, Ripley B. boot: Bootstrap Functions. 2025. R package version 1.3-32doi: 10.32614/CRAN.package.boot
21. Nicolaie M, Houwelingen vHC, Putter H. Vertical modeling: a pattern mixture approach for competing risks modeling. *Statistics in medicine.* 2010;29(11):1190–1205.
22. Kalbfleisch JD, Prentice RL. *The statistical analysis of failure time data*. John Wiley & Sons, 2002.
23. Kundu D, Basu S. Analysis of incomplete data in presence of dependent competing risks. *Statistical Methods and Practice: Recent Advances.* 2003:313–322.
24. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological).* 1977;39(1):1–22.
25. McLachlan GJ, Krishnan T. *The EM algorithm and extensions*. 382. John Wiley & Sons, 2007.
26. Louis TA. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological).* 1982;44(2):226–233.
27. Meilijson I. A Fast Improvement to the EM Algorithm on its Own Terms. *Journal of the Royal Statistical Society. Series B (Methodological).* 1989;51(1):127–138.
28. Farewell VT. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics.* 1982:1041–1046.
29. Kuk AY, Chen CH. A mixture model combining logistic regression with proportional hazards regression. *Biometrika.* 1992;79(3):531–541.
30. Peng Y, Dear KB. A nonparametric mixture model for cure rate estimation. *Biometrics.* 2000;56(1):237–243.
31. Sy J, Taylor J. Standard errors for the Cox proportional hazards cure model. *Mathematical and computer modelling.* 2001;33(12-13):1237–1251.
32. Kannan N, Kundu D, Nair P, Tripathi R. The generalized exponential cure rate model with covariates. *Journal of Applied Statistics.* 2010;37(10):1625–1636.
33. Lam K, Fong DY, Tang O. Estimating the proportion of cured patients in a censored sample. *Statistics in medicine.* 2005;24(12):1865–1879.

□

APPENDIX

A PROOF OF THEOREM 1

Note that for all $k = 1, 2, \ldots, K$, the first derivative of $g_k(\cdot)$ with respect to α is

$$g_k'(\alpha) = -\frac{r_k}{2\alpha^2} - r_k \frac{\left\{\sum_{l=1}^{n} a_l t_l^{\alpha} (\ln t_l)^2\right\}\left\{\sum_{l=1}^{n} a_l t_l^{\alpha}\right\} - \left\{\sum_{l=1}^{n} a_l t_l^{\alpha} \ln t_l\right\}^2}{\left\{\sum_{l=1}^{n} a_l t^{\alpha}\right\}^2},$$

where

$$a_l = \begin{cases} I(\delta_l = k) & \text{if } l \in J_1 \\ w_l^{(m+1)} P^{(m+1)}(k|t_l) & \text{if } l \in J_0. \end{cases}$$

Now,

$$\left\{\sum_{l=1}^{n} a_l t_l^{\alpha}(\ln t_l)^2\right\}\left\{\sum_{l=1}^{n} a_l t_l^{\alpha}\right\} - \left\{\sum_{l=1}^{n} a_l t_l^{\alpha}\ln t_l\right\}^2 = \sum_{l=1}^{n}\sum_{m=1}^{n} a_l a_m t_l^{\alpha} t_m^{\alpha}(\ln t_l - \ln t_m)^2 \geq 0,$$

as $a_l \geq 0$ for all l. Therefore, we conclude that $g_k(\cdot)$ is an strictly increasing function for all $k = 1, 2, \ldots, K$. It is easy to see that $\lim_{\alpha\to 0} g_k(\alpha) = \infty$ and $\lim_{\alpha\to\infty} g_k(\alpha) < 0$ for all $k = 1, 2, \ldots, K$. Thus, $g_k(\alpha) = 0$ has exactly one solution in $(0, \infty)$.

B CALCULATION OF EMPIRICAL FISHER INFORMATION MATRIX

The log-likelihood contribution of l-th observation of complete data is

$$\ell_c = \sum_{k=1}^{K} I(\delta_l = k)\left\{\ln(1-p_0) + \ln\pi_k + \ln f_k(t_l;\,\boldsymbol{\theta}_k)\right\} + I(\delta_l = 0)\left\{(1-i_l)\ln p_0 + i_l\ln(1-p_0) + i_l\ln S_{mix}(t_l;\,\boldsymbol{\psi})\right\}.$$

Therefore, the components of $s_l(\boldsymbol{\theta})$ are

$$\begin{aligned}
\frac{\partial\ell_c}{\partial\alpha_k} &= I(\delta_l = k)\frac{\frac{\partial}{\partial\alpha_k}f_k(t_l;\,\boldsymbol{\theta}_k)}{f_k(t_l;\,\boldsymbol{\theta}_k)} + I(\delta_l = 0)w_l\pi_k\frac{\frac{\partial}{\partial\alpha_k}S_k(t_l;\,\boldsymbol{\theta}_l)}{S_k(t_l;\,\boldsymbol{\theta}_k)}; \quad k = 1, 2, \ldots, K,\\
\frac{\partial\ell_c}{\partial\lambda_k} &= I(\delta_l = k)\frac{\frac{\partial}{\partial\lambda_k}f_k(t_l;\,\boldsymbol{\theta}_k)}{f_k(t_l;\,\boldsymbol{\theta}_k)} + I(\delta_l = 0)w_l\pi_k\frac{\frac{\partial}{\partial\lambda_k}S_k(t_l;\,\boldsymbol{\theta}_l)}{S_k(t_l;\,\boldsymbol{\theta}_k)}; \quad k = 1, 2, \ldots, K,\\
\frac{\partial\ell_c}{\partial\pi_k} &= \frac{I(\delta_l = k)}{\pi_k} + I(\delta_l = 0)w_l\frac{S_k(t_l;\,\boldsymbol{\theta}_k)}{S_{mix}(t_l;\,\boldsymbol{\theta}_k)}; \quad k = 1, 2, \ldots, K,\\
\frac{\partial\ell_c}{\partial p_0} &= -\frac{1}{1-p_0}\sum_{k=1}^{K} I(\delta_l = k) + I(\delta_l = 0)\frac{1-p_0-w_l}{p_0(1-p_0)},
\end{aligned}$$

where

$$w_l = \frac{(1-p_0)S_{mix}(t_l;\,\boldsymbol{\psi})}{p_0 + (1-p_0)S_{mix}(t_l;\,\boldsymbol{\psi})}.$$

C CALCULATION OF EMPIRICAL FISHER INFORMATION MATRIX WITH COVARIATES

In the presence of covariates, the log-likelihood contribution of l-th observation of complete data is

$$\ell_c = \sum_{k=1}^{K} I(\delta_l = k)\left\{\boldsymbol{x}^T\boldsymbol{\beta} - \ln\left(1+e^{\boldsymbol{x}^T\boldsymbol{\beta}}\right) + \ln\pi_k + \ln f_k(t_l;\,\boldsymbol{\theta}_k)\right\} + I(\delta_l = 0)\left\{i_l\boldsymbol{x}^T\boldsymbol{\beta} - \ln\left(1+e^{\boldsymbol{x}^T\boldsymbol{\beta}}\right) + i_l\ln S_{mix}(t_l;\,\boldsymbol{\psi})\right\}.$$

Therefore, the components of $s_l(\boldsymbol{\gamma})$ are

$$\begin{aligned}
\frac{\partial\ell_c}{\partial\alpha_k} &= I(\delta_l = k)\frac{\frac{\partial}{\partial\alpha_k}f_k(t_l;\,\boldsymbol{\theta}_k)}{f_k(t_l;\,\boldsymbol{\theta}_k)} + I(\delta_l = 0)w_l\pi_k\frac{\frac{\partial}{\partial\alpha_k}S_k(t_l;\,\boldsymbol{\theta}_l)}{S_k(t_l;\,\boldsymbol{\theta}_k)}; \quad k = 1, 2, \ldots, K,\\
\frac{\partial\ell_c}{\partial\lambda_k} &= I(\delta_l = k)\frac{\frac{\partial}{\partial\lambda_k}f_k(t_l;\,\boldsymbol{\theta}_k)}{f_k(t_l;\,\boldsymbol{\theta}_k)} + I(\delta_l = 0)w_l\pi_k\frac{\frac{\partial}{\partial\lambda_k}S_k(t_l;\,\boldsymbol{\theta}_l)}{S_k(t_l;\,\boldsymbol{\theta}_k)}; \quad k = 1, 2, \ldots, K,\\
\frac{\partial\ell_c}{\partial\pi_k} &= \frac{I(\delta_l = k)}{\pi_k} + I(\delta_l = 0)w_l\frac{S_k(t_l;\,\boldsymbol{\theta}_k)}{S_{mix}(t_l;\,\boldsymbol{\theta}_k)}; \quad k = 1, 2, \ldots, K,\\
\frac{\partial\ell_c}{\partial\beta_m} &= \sum_{k=1}^{K} I(\delta_l = k)x_m\left(1 - \frac{e^{\boldsymbol{x}^T\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}^T\boldsymbol{\beta}}}\right) + I(\delta_l = 0)x_m\left(w_l - \frac{e^{\boldsymbol{x}^T\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}^T\boldsymbol{\beta}}}\right); \quad m = 0, 1, \ldots, L,
\end{aligned}$$

where

$$w_l = \frac{\exp\left(\boldsymbol{x}^T\boldsymbol{\beta}\right) S_{mix}(t_l;\, \boldsymbol{\psi})}{1 + \exp\left(\boldsymbol{x}^T\boldsymbol{\beta}\right) S_{mix}(t_l;\, \boldsymbol{\psi})}.$$

D A METHOD OF FINDING INITIAL GUESSES

For finding initial guesses of α_1 and η, we consider the patients died due to mode 1 as event and rest as right censored. Then the survival function of lifetime is estimated using Kaplan-Meier estimate, call it $\widehat{S}(\cdot)$. A linear regression for $\ln(-\ln \widehat{S}(\cdot))$ is fitted on log of corresponding event times and covariates. Finally, the estimated coefficient of log of event times may be taken as an initial guess of α_1. Estimates of other regression coefficients provide initial guesses of $\boldsymbol{\eta}$. Similarly, the initial guesses for α_2 and λ_2 are obtained considering death due to mode 2 as event and rest as right censored. The ratio of number of patients died due to mode 1 to the total number of deaths during the study period is used as initial guess for π_1. Similarly, the initial guess of π_2 is taken to be the ratio of number of patients died due to mode 2 to the total number of deaths. Finally, the initial guesses of parameters involved in logistic link function are taken to be zero.

E PLOTS

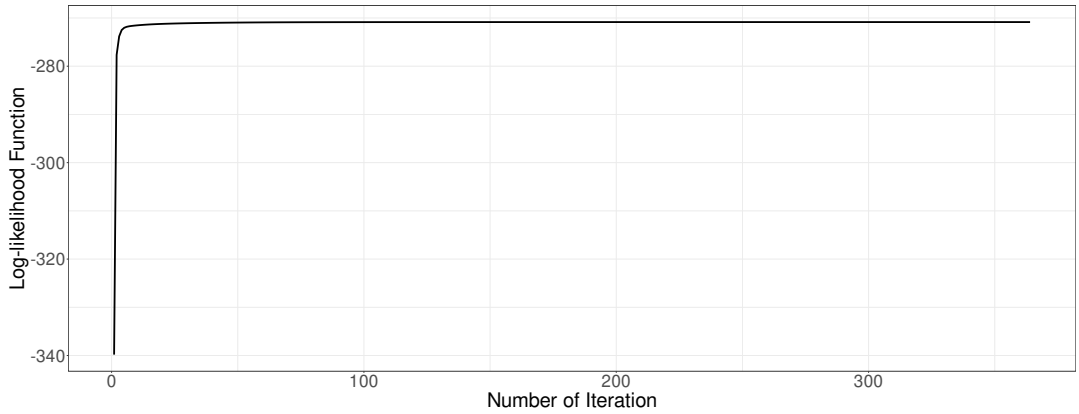

FIGURE E1 Plot of log-likelihood function with respect to EM iterations

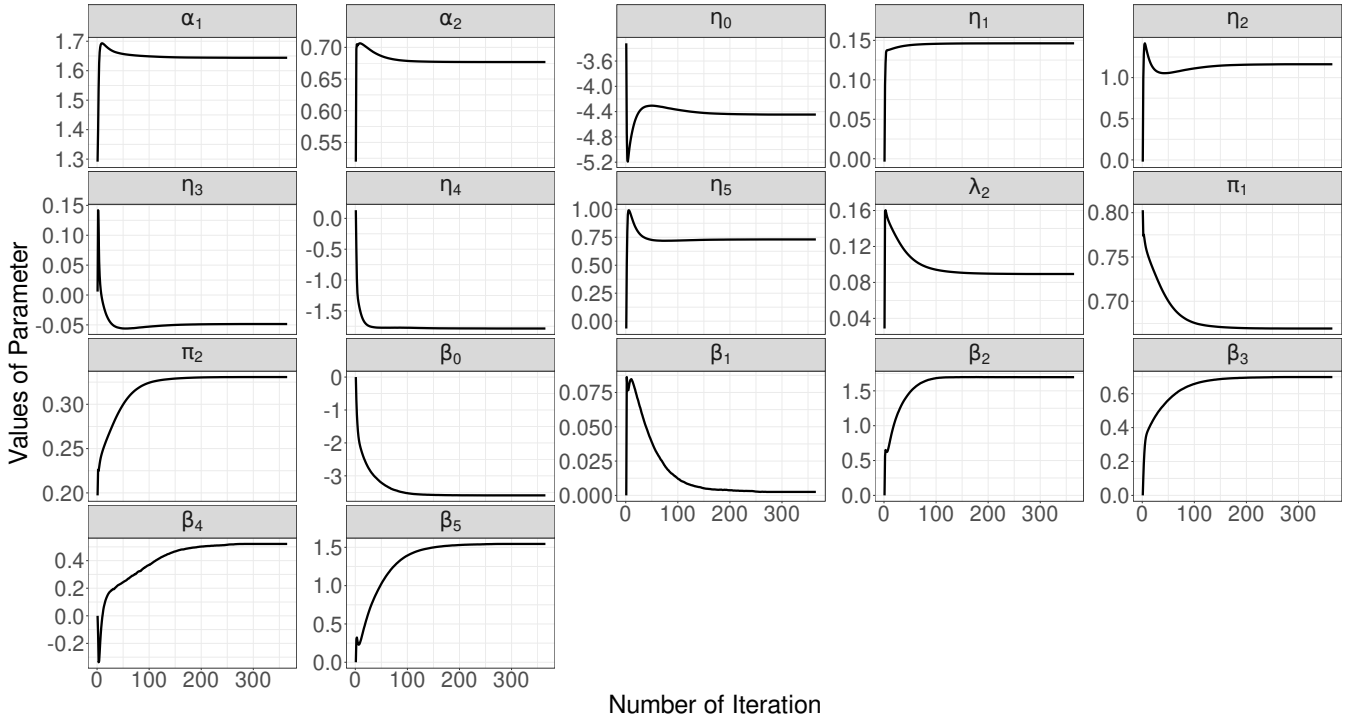

FIGURE E2 Plots of EM iterations for different parameters