A Stochastic Version of the EM Algorithm for Mixture Cure Model with Exponentiated Weibull Family of Lifetimes

Suvra Pal1 · Sandip Barui2 · Katherine Davies3 · Nutan Mishra4

Accepted: 6 June 2022 / Published online: 7 July 2022
© Grace Scientific Publishing 2022

Abstract
In this paper, we propose a stochastic approximation to the well-studied expectation–maximization (EM) algorithm for finding the maximum likelihood (ML)-type estimates in situations where missing data arise naturally and a proportion of individuals are immune to the event of interest. A flexible family of three parameter exponentiated Weibull (EW) distributions is assumed to characterize lifetimes of the non-immune individuals as it accommodates both monotone (increasing and decreasing) and non-monotone (unimodal and bathtub) hazard functions. To evaluate the performance of the proposed algorithm, an extensive simulation study is carried out under various parameter settings. Using likelihood ratio tests, we also carry out model discrimination within the EW family of distributions. Furthermore, we study the robustness of the proposed algorithm with respect to outliers in the data and the choice of initial values to start the algorithm. In particular, we show that our proposed algorithm is less sensitive to the choice of initial values when compared to the EM algorithm. For illustration, we analyze a real survival data on cutaneous melanoma. Through this data, we illustrate the applicability of the likelihood ratio test toward rejecting several well-known lifetime distributions that are nested within the wider class of the proposed EW distributions.

Keywords Incomplete data · Robustness · Optimization · Model discrimination · Melanoma

1 Department of Mathematics, University of Texas at Arlington, Arlington, TX, USA
2 Quantitative Methods and Operations Management Area, Indian Institute of Management, Kozhikode, Kerala, India
3 Department of Mathematical and Computational Sciences, University of Toronto Mississauga, Mississauga, Canada
4 Department of Mathematics and Statistics, University of South Alabama, Mobile, AL, USA
1 Introduction

Immune or cured individuals in the context of survival analysis refers to subjects who would not encounter the event of interest under study, e.g., death due to a disease or relapse of a condition or getting admitted to a hospital. Consequently, the observed lifetimes for the immune individuals would always concur with the length of the study. Hence, the immune individuals would be indiscernible from the censored yet non-immune or susceptible individuals. Ordinary survival analysis techniques ignore the presence of the fraction of individuals who are cured, commonly known as the cure fraction or cure rate. Therefore, several modified modeling techniques (known as cure rate models) to analyze time to event data marked by the presence of cure fraction have been studied over the years (see [5, 10, 43, 46, 54]). Cure rate models have been applied extensively on cancer survival data for cancers with relatively better prognosis (e.g., melanoma, breast cancer, leukemia, and prostate cancer), recidivism studies, engineering reliability, and defaulting on a loan in credit risk assessment studies.

The mixture cure rate model, also called the Bernoulli cure rate model, is probably the most widely used cure rate model; see the recently published monograph by Peng and Yu [52]. Under the mixture cure rate model, the overall population lifetime \( Y \) is defined as

\[
Y = \eta Y_s + (1 - \eta) Y_c,
\]

where \( Y_s \) denotes the survival time for any susceptible individual, \( Y_c = \infty \) denotes the survival time for any cured individual and \( \eta \) is a random variable taking the value 1 or 0 depending on whether an individual is susceptible or immuned, respectively. The model in (1) can be further represented by

\[
S_p(y) = (1 - \pi_0) S_s(y) + \pi_0,
\]

where \( S_p(.) \) and \( S_s(.) \) are the respective survival functions corresponding to \( Y \) and \( Y_s \), and \( \pi_0 = P(\eta = 0) \) is the cure rate. Interestingly, Meeker [33] used the mixture cure model to study reliability of integrated circuits but called the model as limited failure population model. The mixture cure rate model has been explored in detail with various assumptions and extensions. For nonparametric approaches to cure rate estimation, interested readers may refer to López-Cheda et al. [29] and Amico et al. [1]. Furthermore, for new approaches to mixture cure rate parameter estimation, one may refer to the recent works of Musta et al. [36] and Patilea and Van Keilegom [50].

An alternative representation of the cure rate model, namely, the promotion time cure rate model was suggested by Yakovlev et al. [60] and was later investigated by Chen et al. [19]. In a promotion time cure rate model, the case of 0 or 1 risks (causes) is extended to a more general case. In particular, a random variable \( M \) is introduced to denote the number of competing causes and, in this particular case, it is distributed as Poisson. Furthermore, letting \( Y_j, j = 1, \ldots, M, \) denote the promotion time or survival time corresponding to the \( j \)-th cause, the overall population survival function
where given \( M = m, Y_j, j = 1, \ldots, m \) are independently and identically distributed with a common survival function \( S(.) \), and \( \tilde{g}(.) \) is the probability generating function of \( M \). Note that, in (3), \( M \) is unobserved, \( Y_j, j = 1, \ldots, M \), are independent of \( M \), and \( Y = \min\{Y_0, Y_1, \ldots, Y_M\} \) is the actual lifetime of an individual with \( P(Y_0 = \infty) = 1 \).

In cancer studies, competing causes may refer to the tumor cells that can potentially metastasize and cause detectable cancer. Several authors [19] have assumed \( M \) to follow a Poisson distribution, whereas [2, 5, 7, 31, 44, 47, 54] have modeled \( M \) by a flexible Conway–Maxwell (COM) Poisson distribution. When \( M \) is assumed to follow a Poisson distribution with mean \( \theta \), \( S_p(y) \) in (3) reduces to

\[
S_p(y) = e^{-\theta F(y)}
\]

and the cure rate is given by \( \pi_0 = e^{-\theta} \), where \( F(\cdot) = 1 - S(\cdot) \).

The survival function \( S_s(y) \) in (2) or \( S(y) \) in (3) for any susceptible individual could be modeled and estimated by both parametric and nonparametric methods. From the statistical literature, positive valued continuous distributions like Weibull, gamma, generalized gamma and log normal distributions have been applied to model \( Y_s \) or \( Y_j \) (refer [6, 8, 9]). Semi-parametric generalizations to the model by assuming proportional hazards structure for \( Y_s \) or \( Y_j \) have been discussed by [19, 51, 55], whereas a class of semi-parametric transformation models have been studied by [30, 62, 63], among others. Applications of piecewise constant and linear functions to estimate the baseline hazard function under proportional hazards model were discussed by [4, 62].

Missing data arise in many forms. While it is common, especially in the modeling of lifetimes, to face missing covariates, censoring can give rise to other forms of missing information. In the analysis of data with cure fraction, incompleteness in data comes in two folds. Firstly, if censored, the information on the actual survival time of an individual is missing. Secondly, the information on the cured status is also missing for all individuals whose lifetimes are censored. Therefore, parameter estimation may be challenging for the cure rate models. Several methods of estimating the model parameters as well as the baseline hazard or survival functions have been implemented, including ordinary maximum likelihood (ML) estimation, Monte Carlo approximation of a marginal likelihood, profile likelihood, restricted nonparametric ML estimation [57], unbiased estimating equations [11, 30] and projected nonlinear conjugate gradient technique-based estimation [48, 49]. The expectation–maximization (EM) algorithm is also a commonly used estimation algorithm in the context of cure rate models. Taylor [56] developed an EM algorithm for mixture cure model where a Kaplan–Meier-type approach was used to model the latency part of the model. Sy and Taylor [55] developed an EM algorithm for joint estimation of the incidence and latency regression parameters in mixture cure model using the nonparametric form of the likelihood. Kuk and Chen [28] developed an EM algorithm to estimate the baseline survival function of their proposed semi-parametric mixture cure model and proposed to estimate the regression parameters by maximizing a Monte Carlo approximation of the marginal likelihood function. Peng and Dear [51] developed an EM algorithm for a nonparametric mixture cure model where they estimated the baseline survival using a Breslow-type estimator. Balakrishnan and Pal [5]
first developed an EM algorithm for the Conway–Maxwell Poisson cure model that includes the mixture cure model as a special case; see also [3, 6, 8, 10, 27]. Very recently, Davies et al. [20] have introduced a stochastic version of the EM algorithm, called the stochastic EM algorithm (SEM), in the context of cure rate models where \( Y_j \) is modeled by a generalized exponential distribution for every \( j = 1, \ldots, M \). Interested readers may also refer to Pal [40] for a computationally efficient SEM algorithm developed in the context of cure rate model with negative binomial competing risks.

For the mixture cure model, \( M = 0 \) or \( 1 \), and \( Y = \min\{Y_0, Y_1\} \). In this manuscript, our main contribution is in the development of a SEM algorithm to find the estimates of the parameters of the mixture (Bernoulli) cure rate model. In this regard, we propose to model the lifetime \( Y_1 \) by the flexible family of exponentiated Weibull (EW) distributions, which has not been studied before in the context of cure rate models. Being introduced by Celeux and Diebolt [16], the SEM algorithm has been designed to precisely estimate parameters in cases where the log-likelihood function has multiple stationary points, and the EM algorithm does not guarantee convergence to the significant local maxima. Unlike the EM algorithm, the SEM technique is less sensitive to the initial parameter choices, and the implementation is less cumbersome since it does not involve derivation of explicit expected values [15, 18]. In fact, we show in this paper that our proposed SEM algorithm is more robust to the choice of initial values when compared to the EM algorithm. This is the main motivation behind the development of the SEM algorithm.

The probability density function (pdf) of \( Y_1 \), under the assumption of EW distribution, is expressed as:

\[
f_1(y_1) = f_1(y_1; \alpha, k, \lambda) = \left(\frac{\alpha k}{\lambda}\right) \left(\frac{y_1}{\lambda}\right)^{k-1} e^{-\left(y_1/\lambda\right)^k} \left[1 - e^{-\left(y_1/\lambda\right)^k}\right]^{\alpha-1},
\]

where \( y_1 > 0 \) is the support of the distribution, \( \alpha > 0 \) and \( k > 0 \) are the shape parameters, and \( \lambda > 0 \) denotes the scale parameter characterizing the distribution. The EW distribution has been introduced by Mudholkar and Srivastava [35] as an extension to the Weibull distribution by considering an additional shape parameter to the model. As pointed out by Mudholkar and Hutson [34] and Khan [26], modeling failure times by an EW distribution is parsimonious as it accommodates both monotone increasing \((k\alpha \geq 1, k \geq 1)\) or decreasing \((k\alpha \leq 1, k \leq 1)\), and non-monotone unimodal \((k\alpha > 1, k < 1)\) or bathtub shaped \((k\alpha < 1, k > 1)\) hazard functions. Moreover, EW encompasses many well-known lifetime distributions as special cases, e.g., exponential \((\alpha = k = 1)\), Rayleigh \((\alpha = 1, k = 2)\), Weibull \((\alpha = 1)\), generalized or exponentiated exponential \((k = 1)\), and Burr type X \((k = 2)\) distributions. As a result, one can carry out hypotheses tests and model discrimination to validate if the sub-models fit better. Furthermore, EW serves as an alternative to the generalized gamma distribution, which is known to accommodate both monotone and non-monotone hazard functions.

The remainder of this manuscript is arranged in the following manner. We provide our model descriptions for the mixture cure model and basic properties of the EW distribution in Sect. 2. Section 3 deals with the structure of the observed data and development of the likelihood function. In Sect. 4, we discuss the implementation of the SEM algorithm for estimating the model parameters and their standard errors. An
extensive simulation study with carefully chosen parameter settings is carried out in Sect. 5 to examine the robustness and accuracy of the proposed estimation technique. A model discrimination using likelihood-based criterion is performed to assess the flexibility of the EW distribution and the performance of the likelihood ratio test to correctly identify the true distribution. In Sect. 6, the flexibility of the proposed model and the performance of the estimation method are further substantiated based on real-life data collected from a malignant melanoma study. Finally, we provide some concluding remarks and scope of future research in Sect. 7.

2 Model Descriptions

2.1 Exponentiated Weibull Lifetime Distribution

We assume the lifetime of the susceptible individuals to follow an EW distribution. Hence, the cumulative distribution function (cdf), survival function and hazard function of the susceptible lifetime \( Y_1 \) have the following forms:

\[
F_s(y_1) = F_s(y_1; \alpha, k, \lambda) = \left[ 1 - e^{-\left(\frac{y_1}{\lambda}\right)^k} \right]^\alpha, \quad (5)
\]

\[
S_s(y_1) = S_s(y_1; \alpha, k, \lambda) = 1 - \left[ 1 - e^{-\left(\frac{y_1}{\lambda}\right)^k} \right]^\alpha, \quad (6)
\]

and

\[
h_s(y_1) = h_s(y_1; \alpha, k, \lambda) = \frac{\left( \frac{ak}{\lambda} \right) \left( \frac{y_1}{\lambda} \right)^k - 1}{\left[ 1 - e^{-\left(\frac{y_1}{\lambda}\right)^k} \right]^\alpha - 1}, \quad (7)
\]

respectively, where \( y_1 > 0, \alpha > 0, k > 0 \) and \( \lambda > 0 \). One interesting interpretation of the EW distribution is in the area of reliability. If there are \( n \) components in a parallel system and the lifetimes of the components are independently and identically distributed as EW, then, the system lifetime also follows an EW distribution. As pointed out by Nadarajah et al. [37], EW finds applications in a wide variety of problems, e.g., modeling extreme value data on water discharge arising due to river floods, data on optimal accelerated life test plans under type I censoring, firmware system failures, software release times, fracture toughness of materials, bus motor failures and number of ozone peaks, among others. From Mudholkar and Srivastava [35] and Mudholkar and Hutson [34], we note that

(a) If \( \alpha = k = 1 \), the hazard rate is constant;
(b) If \( \alpha = 1 \), the hazard rate is increasing for \( k > 1 \) and decreasing for \( k < 1 \);
(c) If \( k = 1 \), the hazard rate is increasing for \( \alpha > 1 \) and decreasing for \( \alpha < 1 \).

Additionally, the combinations of the two shape parameters, as presented in Table 1, render various shapes to the hazard function.
### Table 1  Hazard rate pattern across various parameter values

| $k$ | $\alpha$ | Hazard rate                        |
|-----|---------|------------------------------------|
| 1   | $< 1$   | Constant (exponential)             |
|     | $> 1$   | Monotonic (Weibull)                |
| $< 1$ | $< 1$ | Decreasing                         |
| $> 1$ | $> 1$ | Increasing                         |
| $> 1$ | $< 1$ | Bathtub or increasing              |
| $< 1$ | $> 1$ | Unimodal or decreasing             |

The general expression for the $q$-th-order raw moment for a random variable $Y_1$ following the EW distribution has been derived by [39], which is given by

$$E(Y_1^q) = \begin{cases} 
\frac{\alpha \lambda^q \Gamma \left( \frac{q}{k} + 1 \right)}{\sum_{i=0}^{\alpha-1} \left( \frac{\alpha-1}{i} \right)} (-1)^i (i + 1)^{-\frac{q}{k}-1}, & \text{if } \alpha \in N, \\
\frac{\alpha \lambda^q \Gamma \left( \frac{q}{k} + 1 \right)}{\sum_{i=0}^{\infty} \frac{(\alpha-1)(\alpha-2)...(\alpha-i)}{i!} (-1)^i (i + 1)^{-\frac{q}{k}-1}}, & \text{if } \alpha \notin N, \text{ for } q = 0, 1, 2, ... \end{cases}$$

(8)

where $N$ denotes the set of natural numbers.

### 2.2 Mixture (Bernoulli) Cure Rate Model

On assuming the number of competing causes $M$ to follow a Bernoulli distribution, i.e., there is either a single cause that can result in an event of interest or there is no cause resulting in a cure, the probability mass function (pmf) of $M$ can be expressed as:

$$p(m; \nu) = P(M = m; \nu) = \left( \frac{1}{1 + \nu} \right)^{1-m} \left( \frac{\nu}{1 + \nu} \right)^m, m = 0, 1,$$

(9)

where $\nu > 0$. The survival function of the random variable $Y = \min\{Y_0, Y_1\}$, also referred to as the population survival function, can be obtained by combining (3) and (6) and is given by

$$S_p(y) = S_p(y; \nu, \alpha, k, \lambda) = P(Y > y)$$

$$= \sum_{m=0}^{1} P(Y > y|M = m) P(M = m; \nu)$$

$$= p(0; \nu) + S_s(y) p(1; \nu)$$

$$= \left( \frac{1}{1 + \nu} \right) + \left\{ 1 - \left[ 1 - e^{-\nu/\lambda} \right]^k \right\} \left( \frac{\nu}{1 + \nu} \right).$$

(10)
Further, note that
\[
\lim_{y \to \infty} S_p(y) = \left(\frac{1}{1 + \nu}\right) = \pi_0
\]
(11)
is the cure rate or cure probability of any individual in the population. Hence, the population density function can be derived from (10) as:
\[
f_p(y) = f_p(y; \nu, \alpha, k, \lambda) = -\frac{dS_p(y)}{dy} = \left(\frac{\nu}{1 + \nu}\right)\left(\frac{\alpha ky^{k-1}}{\lambda^k}\right) e^{-(y/\lambda)^k} \left[1 - e^{-(y/\lambda)^k}\right]^{\alpha-1}.
\]
(12)

3 Form of the Data and Likelihood Function

The right censoring scheme is considered in our study. For \(i = 1, \ldots, n\) with \(n\) denoting the sample size, let \(Y_i\) and \(C_i\), respectively, denote the actual survival time and the censoring time for individual \(i\). Let \(\delta_i = I(Y_i \leq C_i)\) be the censoring indicator and \(T_i = \min\{Y_i, C_i\}\) be the observed lifetime for the \(i\)-th individual. Therefore, the observed survival data is represented in the form of a triplet denoted by \(\{(t_i, \delta_i, x^*_i) : i = 1, \ldots, n\}\), where \(t_i\) is a realization of \(T_i\) and \(x^*_i = (x_{i1}, \ldots, x_{id})^T \in \mathbb{R}^d\) is the \(d\)-dimensional covariate vector specific to the \(i\)-th subject. Let \(x_i = (1, x^*_i)^T \in \mathbb{R}^{d+1}\).

We further denote \(X = (x_1, \ldots, x_n)^T \in \mathbb{R}^{(d+1) \times n}\), \(\delta = (\delta_1, \ldots, \delta_n)^T \in \mathbb{R}_\delta\) and \(t = (t_1, \ldots, t_n)^T \in \mathbb{R}^n_{\geq 0}\), where \(\mathbb{R}_\delta = \{(\delta_1, \ldots, \delta_n) : \delta_i = 0, 1 \forall i = 1, \ldots, n\}\).

In order to associate the effect of covariates to the cure rate for every \(i = 1, \ldots, n\), we use log-linear function \(\nu_i = e^{x_i^T \beta}\) to link the parameter \(\nu > 0\) with the covariate vector \(x_i\), where \(\beta = (\beta_0, \beta_1, \ldots, \beta_d)^T\) is the respective \((d + 1)\)-dimensional vector of regression parameters. Note that this readily implies that the cure rate is linked to the covariates through a logistic function, i.e., we have
\[
\pi_{0i} = \frac{1}{1 + \nu_i} = \frac{1}{1 + e^{x_i^T \beta}}.
\]
(13)

We define \(\theta = (\beta^T, \alpha, k, \lambda)^T \in \Theta \subset \mathbb{R}^{d+4}\) as the unknown parameter vector and \(\Theta\) as the parameter space. Therefore, the likelihood function \(L_O(\theta; t, \delta, X)\) based on the observed data is given by
\[
L_O(\theta; t, \delta, X) \propto \prod_{i=1}^n \left\{ f_p(t_i; \theta, \delta_i, x^*_i) \right\}^{\delta_i} \left\{ S_p(t_i; \theta, \delta_i, x^*_i) \right\}^{1-\delta_i},
\]
(14)
where \(S_p(\cdot; \theta, \delta_i, x^*_i)\) and \(f_p(\cdot; \theta, \delta_i, x^*_i)\) denote the respective population density and survival functions for individual \(i\), and can be obtained from (10) and (12), respectively.
with some notation adjustments. Hence, the observed data log-likelihood function is expressed as:

\[
\begin{align*}
\ell_O(\theta; t, \delta, X) &= \log L_O(\theta; t, \delta, X) \\
&= \text{constant} + \sum_{i=1}^{n} \delta_i \log f_p(t_i; \theta, \delta_i, x_i^0) \\
&\quad + \sum_{i=1}^{n} (1 - \delta_i) \log S_p(t_i; \theta, \delta_i, x_i^0).
\end{align*}
\] (15)

Let us define \( \Delta_1 = \{i : \delta_i = 1\} \), \( \Delta_0 = \{i : \delta_i = 0\} \), \( n_1 = |\Delta_1| \) and \( F_w(t_i; k, \lambda) = 1 - e^{-(t_i/\lambda)^k} \) for \( i = 1, \ldots, n \).

From (10), (12), (15) and using \( \nu_i = e^{x_i^T \beta}, i = 1, \ldots, n \), the log-likelihood function for the mixture cure rate model takes the following form:

\[
\begin{align*}
\ell_O(\theta; t, \delta, X) &= \text{constant} + \sum_{i=1}^{n} \delta_i \log \left[ \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}} \left( \frac{\alpha k t_i^{k-1}}{\lambda^k} \right) e^{-(t_i/\lambda)^k} \left\{ 1 - e^{-(t_i/\lambda)^k} \right\}^{\alpha-1} \right] \\
&\quad + \sum_{i=1}^{n} (1 - \delta_i) \log \left[ \frac{1 + e^{x_i^T \beta}}{1 + e^{x_i^T \beta}} \left\{ 1 - \left[ 1 - e^{-(t_i/\lambda)^k} \right]^{\alpha} \right\} \right] \\
&= \text{constant} + n_1 (\log \alpha + \log k - k \log \lambda) \\
&\quad + \sum_{i \in \Delta_1} \left[ x_i^T \beta - \log (1 + e^{x_i^T \beta}) + (k - 1) \log t_i \\
&\quad - \left( \frac{t_i}{\lambda} \right)^k + (\alpha - 1) \log F_w(t_i; k, \lambda) \right] \\
&\quad + \sum_{i \in \Delta_0} \left[ \log \left\{ 1 + e^{x_i^T \beta} \left[ 1 - F_w(t_i; k, \lambda)^\alpha \right] \right\} - \log \left( 1 + e^{x_i^T \beta} \right) \right].
\end{align*}
\] (16)

The expressions of the first-order and second-order derivatives of \( \ell_O(\theta; t, \delta, X) \) with respect to \( \theta \) are presented in Section A1 of the supplementary material. These expressions would allow interested researchers to directly maximize \( \ell_O(\theta; t, \delta, X) \) to obtain an estimate of \( \theta \). However, the presence of missing data (due to censoring) strongly motivates us to develop algorithms that can handle such missingness of data. In this paper, we derive both EM and SEM algorithms but we focus on SEM because of the reasons already specified (and empirically proven later in Sect. 5.4). For interested readers, the steps involved in the development of the EM algorithm are presented in Appendix. Note that our EM algorithm is developed under the assumption of the flexible EW distribution to model the lifetime of the susceptible subjects and hence is different from the existing EM algorithms in the literature.
4 Development of the SEM Algorithm

As defined in Sect. 1, let $\eta_i = 1$ if an individual is not cured and $\eta_i = 0$ if an individual is cured, for $i = 1, \ldots, n$. It can be seen that $\eta_i = 1$ for $i \in \Delta_1$ and $\eta_i$ is unknown (hence, is missing) for $i \in \Delta_0$. The data we observe is partial, and hence, the problem can be treated as an incomplete data problem. Therefore, the EM or EM-like algorithms can be applied for the ML or ML-type estimation of $\theta$. Note that the convergence rate of the EM algorithm depends on factors such as choice of initials parameter values and the flatness of likelihood surface. Furthermore, for likelihood surfaces characterized by several stationary points including saddle points, the EM algorithm does not guarantee convergence to the significant local maxima. Moreover, analytical steps in deriving conditional expectation involve computation of integrals which is often intensive, complex, and in some cases, intractable. In our considered modeling framework, computation of the conditional expectations is not complicated. However, the EM may be quite sensitive to the choice of initial values, which motivates the development of an alternate estimation algorithm, i.e., the SEM algorithm.

The SEM algorithm works on the idea of simulating pseudo values to replace the missing values. The SEM comprises two steps, namely, the S-step and the M-step. The S-step involves generating a pseudo sample from the conditional distribution of the missing data given the observed information and current parameter values. The M-step involves finding the parameter value which maximizes the complete data log-likelihood function based on the pseudo sample \[15, 17\]. The random generation of values to impute missing data allows the SEM algorithm to overcome the problem of getting trapped in an insignificant local maxima or saddle point \[14, 15\]. A discussion on the asymptotic properties based on a mixture model reveals that the sequence of estimates generated by the SEM algorithm converges to a stationary Gaussian distribution whose mean is the consistent ML estimator of the mixing proportion \[21\].

Define $\tilde{H}_1 = \{i : \eta_i = 1\}$ and $\tilde{H}_0 = \{i : \eta_i = 0\}$. Note that $\tilde{H}_0$ is unobserved and $\tilde{H}_1$ is only partially observed. Hypothetically, assuming that we completely observe $\tilde{H}_0$ and $\tilde{H}_1$, then, for any individual $i \in \tilde{H}_0$, $Y_i > C_i$ and the contribution by $i$ to the likelihood function would be through the cure rate $\pi_0(x_i^*; \beta)$. Again, for any $i \in \tilde{H}_1$, $Y_i > C_i$ or $Y_i \leq C_i, T_i = \min\{Y_i, C_i\}$ and contribution to the likelihood function by $i$ would be through the population density function $f_p(t_i; \theta, \delta_i, x_i^*)$. For the latter, the information on the actual lifetime is missing if the individual is right censored, and observed when not censored. Therefore, we would stochastically generate both cured status $\eta_i$ and subject’s actual lifetime $y_i^*$, and hence, generate pseudo data of the form $\{(y_i^*, \delta_i, x_i^*, \eta_i) : i = 1, \ldots, n\}$.

To implement the SEM algorithm, the complete data likelihood and log-likelihood functions are expressed as

$$
\tilde{L}_C(\theta; y^*, \delta, X, \eta) \propto \prod_{i=1}^{n} \left\{ f_p(y_i^*; \theta, \delta_i, x_i^*) \right\}^{\eta_i} \left\{ \pi_0(x_i^*; \beta) \right\}^{1-\eta_i} \quad (17)
$$
and

\[
\tilde{l}_C(\theta; y^*, \delta, X, \eta) = \text{constant} + \sum_{i=1}^{n} \eta_i \log \left\{ f_p(y_i^*; \theta, \delta_i, x_i^*) \right\} \\
+ (1 - \eta_i) \log \left\{ \pi_0(x_i^*; \beta) \right\},
\]

respectively, where \( y_i^* \) denotes the actual lifetime generated stochastically for \( i \in \Delta_0 \), and \( y^* = (y_1^*, \ldots, y_n^*)^T \). For the mixture cure rate model, (18) becomes

\[
\tilde{l}_C(\theta; y^*, \delta, X, \eta) = \text{constant} + \sum_{i=1}^{n} \eta_i \log \left\{ \frac{\lambda}{\lambda'} \right\}^k \\
+ (\alpha - 1) \sum_{i=1}^{n} \eta_i \log \left\{ 1 - e^{-\left(y_i^*/\lambda\right)^k} \right\} \\
- \sum_{i=1}^{n} (1 - \eta_i) \log \left( 1 + e^{x_i^T \beta} \right).
\]

(19)

4.1 Steps Involved in the SEM Algorithm

Start the iterative process for the SEM algorithm with a reasonable initial choice \( \theta^{(0)} = (\beta^{(0)}, \alpha^{(0)}, k^{(0)}, \lambda^{(0)})^T \) of the parameter \( \theta \). For some pre-defined \( R \in \mathbb{Z}^+ \) and \( r = 0, 1, \ldots, R \), assume \( \theta^{(r)} = (\beta^{(r)}, \alpha^{(r)}, k^{(r)}, \lambda^{(r)})^T \) as the estimate of the parameter \( \theta \) for the \( r \)-th step. The steps below permit the computation of the ML-type estimate of \( \theta \) by applying the SEM algorithm.

1. **S-Step** There are two sub-steps to be followed in the stochastic step of the implementation.

   A. **Generating cure status** \( \eta_i^{(r+1)} \) for \( i = 1, \ldots, n \):
      (i) For \( i \in \Delta_1 \), \( \eta_i^{(r+1)} = 1 \).
      (ii) For \( i \in \Delta_0 \), generate \( \eta_i^{(r+1)} \) from a Bernoulli distribution with conditional probability of success \( p_{s,i}^{(r+1)} \) given by

   \[
   p_{s,i}^{(r+1)} = P \left\{ \eta_i^{(r+1)} = 1 \ \left| \ (\theta^{(r)}, Y_i > t_i, x_i^*, i \in \Delta_0) \right. \right\} \\
   = 1 - \frac{\pi_0(x_i^*; \beta^{(r)})}{S_p(t_i; \theta^{(r)}, \delta_i, x_i^*)}.
   \]

   B. **Generating actual lifetime** \( y_i^*(r+1) \) for \( i = 1, \ldots, n \):
      (i) For \( i \in \Delta_1 \), \( y_i^*(r+1) = t_i \) is the actual lifetime.
(ii) For \( i \in \Delta_0 \) and if \( \eta_i^{(r+1)} = 0 \) from step 1A., \( y_i^{*(r+1)} = \infty \) since the individual is cured with respect to the event of interest.

(iii) For \( i \in \Delta_0 \) and if \( \eta_i^{(r+1)} = 1 \) from 1A., we only observe the censoring time \( t_i = c_i \) since the actual lifetime \( Y_i > t_i \). Hence, actual lifetime \( y_i^{*(r+1)} \) is generated from a truncated EW distribution with density \( g(.; \theta^{(r)}, \delta_i, x_i^*) \), where

\[
g\left(y_i^{*(r+1)}; \theta^{(r)}, \delta_i, x_i^*\right) = \frac{f_p\left(y_i^{*(r+1)}; \theta^{(r)}, \delta_i, x_i^*\right)}{S_p\left(t_i; \theta^{(r)}, \delta_i, x_i^*\right)}, \quad y_i^{*(r+1)} > t_i \text{ with } \delta_i = 0. \tag{21}\]

Let \( G(.; \theta^{(r)}, \delta_i, x_i^*) \) denote the cdf corresponding to \( g(.; \theta^{(r)}, \delta_i, x_i^*) \). It can be noted that \( G(.; \theta^{(r)}, \delta_i, x_i^*) \) is not a proper cdf as

\[
G\left(y_i^{*(r+1)}; \theta^{(r)}, \delta_i, x_i^*\right) = 1 - \frac{S_p\left(y_i^{*(r+1)}; \theta^{(r)}, \delta_i, x_i^*\right)}{S_p\left(t_i; \theta^{(r)}, \delta_i, x_i^*\right)} \tag{22}\]

and

\[
\lim_{y_i^{*(r+1)} \to \infty} G\left(y_i^{*(r+1)}; \theta^{(r)}, \delta_i, x_i^*\right) = 1 - \frac{\pi_0\left(x_i^*; \beta^{(r)}\right)}{S_p\left(t_i; \theta^{(r)}, \delta_i, x_i^*\right)} = b_i^{(r+1)}, \tag{23}\]

where \( b_i^{(r+1)} = 1 \) only if \( \pi_0\left(x_i^*; \beta^{(r)}\right) = 0 \). In this case, two schemes could be followed for generating \( y_i^{*(r+1)} \).

(a) Generate \( u_i^{(r+1)} \) randomly from uniform \( (0, b_i^{(r+1)}) \) and take an inverse transformation to find \( y_i^{*(r+1)} = G^{-1}\left(u_i^{(r+1)}; \theta^{(r)}, \delta_i, x_i^*\right) \).

(b) Generate \( m_i^{(r+1)} \) using \( p\left(m_i^{(r+1)}; e^{x_i^T \beta^{(r)}}\right) \) given in (9), i.e., from a Bernoulli distribution with success probability \( \left\{ \frac{e^{x_i^T \beta^{(r)}}}{1 + e^{x_i^T \beta^{(r)}}} \right\} \). If \( m_i^{(r+1)} = 1 \), then simulate \( y_i^{*(r+1)} \) from the pdf \( g(.; \theta^{(r)}, \delta_i, x_i^*) \), which is the pdf of a truncated EW distribution given in (21), truncated at \( t_i \).
2. **M-Step** Once the pseudo complete data \( \left\{ (y_i^{*(r+1)}, \delta_i, x_i^*, \eta_i^{*(r+1)}) : i = 1, \ldots, n \right\} \) is obtained, find the updated estimate by

\[
\theta^{(r+1)} = \left( \beta^{(r+1)}, \alpha^{(r+1)}, k^{(r+1)}, \lambda^{(r+1)} \right)^T = \arg \max \hat{L}_C(\theta; y^{*(r+1)}, \delta, X, \eta^{(r+1)}),
\]

where \( y^{*(r)} = (y_1^{*(r)}, \ldots, y_n^{*(r)})^T \) and \( \eta^{*(r)} = (\eta_1^{*(r)}, \ldots, \eta_n^{*(r)})^T \). The maximization part can be carried out using multidimensional unconstrained optimization methods such as Nelder–Mead simplex search algorithm or quasi Newton methods such as BFGS algorithm (see [24]). These algorithms are available in statistical software R version 4.0.3 under General Purpose Optimization package called `optimr()`.

3. Repeat steps 1 and 2 a certain number of times, say \( R \), to obtain a sequence of estimates \( \{\theta^{(r)}\}_{r=1}^R \). As pointed out by Diebolt and Celeux [21], the sequence

\[
\lim_{R \to \infty} \{\theta^{(r)}\}_{r=1}^R
\]

does not converge pointwise, and hence, the implementation of the SEM algorithm will not result in the consistent ML estimator. However, the ergodic Markov chain \( \{\theta^{(r)}\}_{r=1}^R \) generated by the implementation of the SEM algorithm converges to a normal distribution. It was further established by [21] that the mean of the normal distribution is the consistent ML estimate of \( \theta \) under some mild technical assumptions. Based on this result and arguments provided by Celeux et al. [15] and Davies et al. [20], the SEM estimate \( \hat{\theta}_{SEM} \) may be obtained by one of the following two approaches:

(a) Calculate the SEM estimate by

\[
\hat{\theta}_{SEM} = (R - R^*)^{-1} \sum_{r=R^*+1}^R \theta^{(r)},
\]

where iterations \( r = 1, \ldots, R^* \) represent “burn-in” or “warm-up” period to reach the stationary regime, and the estimates \( \theta^{(r)}, r = 1, \ldots, R^* \) are discarded. Marschner [32] indicated that a point estimate of \( \theta \) can be calculated by taking average over the estimates obtained from iterations of the SEM algorithm after sufficiently long burn-in period. Both Marschner [32] and Ye et al. [61] used first 100 iterations of the algorithm as the burn-in period, and considered additional 900–1000 iterations for obtaining the SEM estimates. However, it is recommended to do a trace plot of the sequence of estimates against iteration numbers to examine the trend in the behavior of the estimates, and thereby, choosing an appropriate burn-in period.
(b) Carry out \( R^* \) iterations as “warm-up” and derive the sequence \( \left\{ \theta^{(r)} \right\}_{r=1}^{R^*} \) by implementing the SEM algorithm. Find

\[
\hat{\theta}_{SEMI} = \arg \max_{\{\theta^{(r)}, r=1,\ldots,R^*\}} l_O (\theta; t, \delta, X).
\]  

(26)

By taking \( \hat{\theta}_{SEMI} \) as the starting value, the EM algorithm is implemented to derive the ML estimate \( \hat{\theta} \) (see [15, 17]).

Note that approach (b) above requires the development of both SEM and EM algorithms and hence may not be a preferred approach to calculate the estimates. On the other hand, approach (a) above may result in under-estimation of the variances of the estimators, see [22]. In fact, in our model fitting study, as presented in Sect. 5.1, we have encountered the problem with under-estimated variances. The variances did improve when the sample size is very large and when the cure proportions are very small. In this paper, we propose to take each \( \theta^{(r)}, r = R^* + 1, \ldots, R \), and evaluate the observed data log-likelihood function. Then, we take that \( \theta^{(r)} \) as the estimate of \( \theta \) for which the log-likelihood function value is the maximum.

It is also important to discuss the differences between the SEM algorithm and the Monte Carlo EM (MCEM) algorithm [12, 13]. In the MCEM algorithm, the E-step, i.e., the conditional expectation of the missing data, is approximated by the Monte Carlo mean based on, say, \( N \) samples drawn from the conditional distribution of the missing data. The resulting conditional expected log-likelihood function is maximized (M-step) to obtain an improved set of estimates. Then, the E-step and M-step are repeated iteratively until some convergence criterion is achieved. On the other hand, in the SEM algorithm, the E-step is replaced by just a single draw from the conditional distribution of the missing data, followed by the M-step. Then, we repeat the E-step and M-step a fixed number, say, \( R \) times. Since the SEM is based on drawing one sample from the conditional distribution of the missing data along with \( R \) iterations, whereas the MCEM is based on drawing multiple samples (may be taken as 500 or 1000) in each iteration to approximate the conditional mean, it is clear that the MCEM is computationally more expensive than the proposed SEM. This has been recently empirically proven by [40], where it was shown that the time taken by MCEM is roughly 5 to 6 times the time taken by SEM; see Table 9 in [40]. In addition, it was also shown that the MCEM results in the coverage probabilities to go beyond the nominal level; see Table 8 in [40]. Given these recent findings, the proposed SEM is considered to be the preferred algorithm even though it is possible to develop the MCEM in our setting.

5 Simulation Study

5.1 Model Fitting

To assess the performance of the proposed SEM algorithm in the context of mixture cure model with EW lifetimes, we carry out an extensive simulation study. For this
purpose, we mimic the cutaneous melanoma data (analyzed later in Sect. 6) with the nodule category (taking values 1, 2, 3 and 4) as the only covariate in our application. As can be seen from the results in Sect. 6, the cure rate is monotonically decreasing with nodule category. So, along the same lines, in this simulation study we include a covariate effect \( x \) in the form of \( x = j \) for \( j = 1, 2, 3, 4 \), and consider the cure rate to be decreasing in the covariate. From hereon in, we refer to the observations associated with covariate value \( j \) as belonging to group \( j \). We also link the cure rate \( \pi_0 \) to the covariate \( x \) through the relation
\[
\pi_0(x, \beta) = \frac{1}{1 + e^{\beta_0 + \beta_1 x}} \quad \text{for} \quad j = 1, 2, 3, 4.
\]
In order to determine the values of the regression parameters, two cure rates need to be fixed. With this purpose, we fix the values of \( \pi_0(x = 1, \beta) \) (for group 1) and \( \pi_0(x = 4, \beta) \) (for group 4) as \( \pi_{01} \) and \( \pi_{04} \), respectively. This results in the following expressions for the regression parameters \( \beta_0 \) and \( \beta_1 \):
\[
\beta_1 = \frac{1}{3} \left[ \log \left( \frac{1}{\pi_{04}} - 1 \right) - \log \left( \frac{1}{\pi_{01}} - 1 \right) \right],
\beta_0 = \log \left( \frac{1}{\pi_{01}} - 1 \right) - \beta_1.
\]
(27)

Using (27), the cure rates for groups 2 and 3 can be easily calculated as \( \pi_{02} = \frac{1}{1 + e^{\beta_0 + 2\beta_1}} \) and \( \pi_{03} = \frac{1}{1 + e^{\beta_0 + 3\beta_1}} \), respectively.

For sample sizes \( n = 200 \) and 400, we vary the lifetime distribution parameters, cure rates and censoring proportions. For cure rates, we consider two levels, which we refer to as “High” and “Low.” Within our study, in the high setting, we fix groups 1 and 4’s cure rates as 0.50 and 0.20, respectively, and in the low setting, we fix them as 0.40 and 0.10, respectively. Finally, as mentioned in Sect. 3, we allow for observations to be right censored. In order to incorporate this mechanism, we fix the overall censoring proportion for each group \( (p_j, j = 1, 2, 3, 4) \). In the high setting, these are fixed as \((0.65,0.50,0.40,0.30)\) and in the low setting, \((0.50,0.40,0.30,0.20)\). With these values, for each group, realized censoring times can be generated by assuming they follow an exponential distribution with rate parameter \( \gamma \), which, for fixed censoring proportion \( p \) and cure rate \( \pi_0 \), can be found by solving the following equation:
\[
\frac{p - \pi_0}{1 - \pi_0} = P[Y > C | M > 0] = \frac{P[Y > C, M = 1]}{P[M = 1]} = \frac{1}{1 - \pi_0} \int_0^\infty S(x) \gamma e^{-\gamma x} \, dx,
\]
where under the mixture cure rate model in (9), \( 1 - \pi_0 = \frac{\nu}{1+\nu} \). Note that \( S(\cdot) \) is the survival function of the EW distribution, as defined in (6). From here, assuming that various cure rates and censoring rates have been predetermined for each group, the following steps are followed to generate the observed lifetime, \( T \), under our model. First, a value of \( M \) is generated from a Bernoulli distribution with \( P[M = 1] = 1 - \pi_0 \) and with it, a censoring time \( C \) is generated from an exponential distribution with rate

\( \gamma \).
Table 2  Parameter settings for simulation study

| Lifetime parameter | n   | Cure rate | $\beta_0$ | $\beta_1$ | $\pi_{01}$ | $\pi_{02}$ | $\pi_{03}$ | $\pi_{04}$ |
|--------------------|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Setting 1          | 200 | Low       | -0.192    | 0.597     | 0.400     | 0.268     | 0.168     | 0.100     |
|                    | 400 | Low       | -0.192    | 0.597     | 0.400     | 0.268     | 0.168     | 0.100     |
|                    | 200 | High      | -0.462    | 0.462     | 0.500     | 0.386     | 0.284     | 0.200     |
|                    | 400 | High      | -0.462    | 0.462     | 0.500     | 0.386     | 0.284     | 0.200     |
| Setting 2          | 200 | Low       | -0.192    | 0.597     | 0.400     | 0.268     | 0.168     | 0.100     |
|                    | 400 | Low       | -0.192    | 0.597     | 0.400     | 0.268     | 0.168     | 0.100     |
|                    | 200 | High      | -0.462    | 0.462     | 0.500     | 0.386     | 0.284     | 0.200     |
|                    | 400 | High      | -0.462    | 0.462     | 0.500     | 0.386     | 0.284     | 0.200     |
| Setting 3          | 200 | Low       | -0.192    | 0.597     | 0.400     | 0.268     | 0.168     | 0.100     |
|                    | 400 | Low       | -0.192    | 0.597     | 0.400     | 0.268     | 0.168     | 0.100     |

Parameter $\gamma$. If $M = 0$, this means there is no risk and the true lifetime is infinite with respect to the event of interest and so in this case, the observed lifetime is $T = C$. If $M = 1$, there is risk and so a true lifetime, $Y$, from the EW distribution is generated with parameters $(\alpha, k, \lambda)$ and in this case, the observed lifetime is simply $T = \min\{Y, C\}$. Finally, if $T = Y$, the right censoring indicator, $\delta$, is taken as 1, otherwise, it is taken as 0.

For the parameters of the lifetime distribution $Y$, we consider three different parameter settings as 1: $(\alpha, \lambda, k) = (2, 1.5, 1)$, 2: $(1,1.5,2)$ and 3: $(1,0.5,1.5)$. For these choices of lifetime parameters, we consider different combinations of cure rates and sample sizes, resulting in 12 different settings. These settings are presented in Table 2. For a given sample of observations, once the parameter estimates are obtained, with the goal to construct confidence intervals, we numerically approximate the Hessian matrix and then invert it to estimate the standard errors of the estimators. As to be seen in the tables, this will allow for the calculation of associated coverage probabilities. For each parameter setting, we generate $K = 500$ samples using Monte Carlo simulation. Note that within the SEM algorithm, we choose $R = 1500$ runs and use the first 500 as burn-in. To find the initial values of model parameters, we randomly choose a value in the parameter space within 10% of the true value. In Table 3, and in Tables A2.1 and A2.2 in the supplementary material, we summarize the performance of the SEM algorithm in estimating the model parameters. The tables include the estimates (and standard errors), bias, root mean square error (RMSE), and two coverage probabilities (90% and 95%). We first observe that as $n$ increases, with everything else fixed, the bias and RMSE both decrease, and coverage probabilities improve. Subsequently, in Table 4, and in Tables A2.3 and A2.4 in the supplementary material, we summarize the corresponding results for the estimation of cure rates. Across all parameter settings, it is evident that the estimates of the cure rates are consistently unbiased.
Table 3 SEM estimation results of model parameters when \((\alpha, \lambda, k) = (2, 1.5, 1)\) and the initial guess is close to the true parameter values

| n   | Parameters | Estimates (SE) | Bias   | RMSE  | 90% CP  | 95% CP  |
|-----|------------|----------------|--------|-------|---------|---------|
|     | \(\beta_0\) = -0.192 | -0.258 (0.447) | -0.066 | 0.480 | 0.874   | 0.936   |
| 200 | \(\beta_1\) = 0.597   | 0.641 (0.204)  | 0.043  | 0.224 | 0.884   | 0.938   |
|     | \(\alpha = 2\)      | 2.138 (1.383)  | 0.138  | 1.021 | 0.870   | 0.902   |
|     | \(\lambda = 1.5\)   | 1.590 (0.710)  | 0.090  | 0.567 | 0.898   | 0.954   |
|     | \(k = 1\)           | 1.084 (0.333)  | 0.084  | 0.295 | 0.948   | 0.982   |
|     | \(\beta_0 = -0.192\) | -0.205 (0.313) | -0.013 | 0.328 | 0.884   | 0.932   |
|     | \(\beta_1 = 0.597\)  | 0.609 (0.141)  | 0.012  | 0.144 | 0.898   | 0.950   |
| 400 | \(\alpha = 2\)      | 2.190 (0.927)  | 0.190  | 0.881 | 0.868   | 0.910   |
|     | \(\lambda = 1.5\)   | 1.511 (0.471)  | 0.011  | 0.466 | 0.884   | 0.930   |
|     | \(k = 1\)           | 1.028 (0.207)  | 0.028  | 0.219 | 0.894   | 0.948   |
|     | \(\beta_0 = -0.462\) | -0.453 (0.438) | -0.009 | 0.451 | 0.896   | 0.952   |
|     | \(\beta_1 = 0.462\)  | 0.468 (0.175)  | 0.006  | 0.175 | 0.908   | 0.952   |
| 200 | \(\alpha = 2\)      | 2.016 (1.469)  | 0.016  | 1.056 | 0.808   | 0.842   |
|     | \(\lambda = 1.5\)   | 1.708 (0.794)  | 0.208  | 0.686 | 0.908   | 0.946   |
|     | \(k = 1\)           | 1.154 (0.402)  | 0.154  | 0.394 | 0.968   | 0.984   |
|     | \(\beta_0 = -0.462\) | -0.463 (0.307) | -0.001 | 0.320 | 0.892   | 0.942   |
|     | \(\beta_1 = 0.462\)  | 0.462 (0.123)  | 0.000  | 0.126 | 0.898   | 0.936   |
| 400 | \(\alpha = 2\)      | 2.144 (1.079)  | 0.144  | 0.907 | 0.858   | 0.906   |
|     | \(\lambda = 1.5\)   | 1.547 (0.545)  | 0.047  | 0.487 | 0.900   | 0.950   |
|     | \(k = 1\)           | 1.047 (0.246)  | 0.047  | 0.232 | 0.932   | 0.974   |

\(CP\) coverage probability, \(SE\) standard error

5.2 Robustness Study with Respect to the Presence of Outliers

In this section, we study the performance of the SEM algorithm when there are outliers present in the data. We consider a scenario where the generated data contains 5% outliers. For this purpose, we generate 95% of the data with true parameter setting as follows: \((\beta_0, \beta_1, \alpha, \lambda, k) = (-0.462, 0.462, 1, 1.5, 2)\), which corresponds to \((\pi_{01}, \pi_{04}) = (0.5, 0.2)\) and \((p_1, p_2, p_3, p_4) = (0.65, 0.50, 0.40, 0.30)\). The remaining 5% of the data are outliers and are generated from \((\beta_0, \beta_1, \alpha, \lambda, k) = (-0.192, 0.597, 1, 1, 0.3)\), which corresponds to \((\pi_{01}, \pi_{04}) = (0.4, 0.1)\) and \((p_1, p_2, p_3, p_4) = (0.50, 0.40, 0.30, 0.20)\). As far as the lifetime distribution is concerned, the true parameter setting results in a mean of 1.329, whereas the setting to generate outliers results in a mean of 9.260. Then, for the entire data, we use the SEM algorithm to estimate the true model parameters \((\beta_0, \beta_1, \alpha, \lambda, k) = (-0.462, 0.462, 1, 1.5, 2)\) and the true cure rates \((\pi_{01}, \pi_{02}, \pi_{03}, \pi_{04}) = (0.5, 0.386, 0.284, 0.2)\). Based on 500 Monte Carlo simulations, we present the estimation results of model parameters in Table 5 and of the cure rates in Table 6. From Table 5, it is clear that the presence of outliers results in biased estimates, which is more pronounced for the lifetime parameters and specifically for the parameter \(k\). This is certainly due to the choice of the
Table 4  SEM estimation results of cure rates when \((\alpha, \lambda, k) = (2, 1.5, 1)\) and the initial guess is close to the true parameter values

| Parameter | Estimates (SE) | Bias  | RMSE  | 90% CP  | 95% CP  |
|-----------|----------------|-------|-------|---------|---------|
| \(n_{01} = 0.400\) | 0.408 (0.066) | 0.008 | 0.070 | 0.870   | 0.924   |
| \(n_{02} = 0.268\) | 0.266 (0.039) | -0.002 | 0.040 | 0.902   | 0.944   |
| \(n_{03} = 0.168\) | 0.164 (0.039) | -0.004 | 0.041 | 0.866   | 0.920   |
| \(n_{04} = 0.100\) | 0.099 (0.038) | -0.001 | 0.041 | 0.818   | 0.886   |
| \(n_{01} = 0.400\) | 0.401 (0.047) | 0.001 | 0.050 | 0.870   | 0.940   |
| \(n_{02} = 0.268\) | 0.267 (0.028) | -0.001 | 0.029 | 0.876   | 0.940   |
| \(n_{03} = 0.168\) | 0.167 (0.028) | -0.001 | 0.028 | 0.878   | 0.948   |
| \(n_{04} = 0.100\) | 0.101 (0.028) | 0.001 | 0.028 | 0.876   | 0.926   |
| \(n_{01} = 0.500\) | 0.496 (0.071) | -0.004 | 0.074 | 0.874   | 0.926   |
| \(n_{02} = 0.386\) | 0.383 (0.045) | -0.004 | 0.047 | 0.874   | 0.916   |
| \(n_{03} = 0.284\) | 0.281 (0.045) | -0.003 | 0.045 | 0.880   | 0.946   |
| \(n_{04} = 0.200\) | 0.201 (0.055) | 0.001 | 0.056 | 0.882   | 0.944   |
| \(n_{01} = 0.500\) | 0.500 (0.050) | 0.000 | 0.052 | 0.876   | 0.926   |
| \(n_{02} = 0.386\) | 0.387 (0.032) | 0.001 | 0.032 | 0.888   | 0.936   |
| \(n_{03} = 0.284\) | 0.285 (0.032) | 0.001 | 0.032 | 0.908   | 0.948   |
| \(n_{04} = 0.200\) | 0.203 (0.040) | 0.003 | 0.040 | 0.910   | 0.930   |

*CP* coverage probability, *SE* standard error

parameters using which the outliers were generated. Note that there is also a significant under-coverage that can be noticed for the lifetime parameters. The increase in sample size helps in the reduction of the standard errors and the RMSEs. It also helps in the reduction of bias for all model parameters except for the parameter \(k\). From Table 6, we note that the estimates of cure rates contain little bias when compared to the results in Sect. 5.1 where there were no outliers. A slight under-coverage is also noticed for some cure rates. Finally, the performances of the SEM and EM algorithms are similar in the presence of outliers.

### 5.3 Model Discrimination

As mentioned in Sects. 1 and 2, the EW family of distributions includes many well-known lifetime distributions. Consequentlly, it makes sense to carry out a model discrimination study across the sub-models through the general EW distribution. The idea is to evaluate the performance of the likelihood ratio test in discriminating among the sub-models. For this purpose, we choose the setting with “Low” cure rates and a sample of size 400. The EW scale parameter \(\lambda\) is chosen to be 2.5.

Data from the mixture cure rate model are generated with the lifetimes coming from the five special cases (true models) of the EW family, namely, exponential \((\alpha = 1, k = 1)\), Rayleigh \((\alpha = 1, k = 2)\), Weibull \((\alpha = 1, k = 1.5)\), generalized exponential \((\alpha = 2, k = 1)\) and Burr type X \((\alpha = 2, k = 2)\) distributions. For data generated from every true model, all five sub-models are fitted and parameter estimation is carried out
Table 5 SEM estimation results of model parameters in the presence of outliers in the data

| n  | Parameters | Estimates (SE) | Bias   | RMSE  | 90% CP  | 95% CP  |
|----|------------|----------------|--------|-------|---------|---------|
| 200| $\beta_0 = -0.462$ | -0.422 (0.452) | 0.040  | 0.464 | 0.904   | 0.952   |
|    | $\beta_1 = 0.462$   | 0.473 (0.172)  | 0.011  | 0.177 | 0.894   | 0.942   |
| 400| $\alpha = 1$       | 1.119 (0.343)  | 0.119  | 0.776 | 0.528   | 0.622   |
|    | $\lambda = 1.5$    | 1.573 (0.272)  | 0.073  | 0.453 | 0.610   | 0.722   |
|    | $k = 2$           | 1.750 (0.367)  | -0.250 | 0.845 | 0.456   | 0.506   |
|    | $\beta_0 = -0.462$ | -0.443 (0.316) | 0.020  | 0.311 | 0.894   | 0.952   |
|    | $\beta_1 = 0.462$   | 0.473 (0.120)  | 0.011  | 0.119 | 0.912   | 0.948   |
| 200| $\alpha = 1$       | 1.105 (0.213)  | 0.105  | 0.592 | 0.430   | 0.504   |
|    | $\lambda = 1.5$    | 1.528 (0.181)  | 0.028  | 0.370 | 0.530   | 0.608   |
|    | $k = 2$           | 1.576 (0.193)  | -0.424 | 0.748 | 0.256   | 0.326   |

CP coverage probability, SE standard error

Table 6 SEM estimation results of cure rates in the presence of outliers in the data

| n  | Parameter | Estimate (SE) | Bias   | RMSE  | 90% CP  | 95% CP  |
|----|-----------|---------------|--------|-------|---------|---------|
| 200| $\pi_{01} = 0.500$ | 0.488 (0.074) | -0.012 | 0.077 | 0.866   | 0.938   |
|    | $\pi_{02} = 0.368$ | 0.373 (0.046) | -0.013 | 0.050 | 0.796   | 0.926   |
|    | $\pi_{03} = 0.284$ | 0.272 (0.042) | -0.012 | 0.046 | 0.818   | 0.916   |
|    | $\pi_{04} = 0.200$ | 0.192 (0.050) | -0.008 | 0.054 | 0.828   | 0.920   |
| 400| $\pi_{01} = 0.500$ | 0.493 (0.052) | -0.007 | 0.053 | 0.844   | 0.954   |
|    | $\pi_{02} = 0.368$ | 0.377 (0.033) | -0.009 | 0.035 | 0.682   | 0.938   |
|    | $\pi_{03} = 0.284$ | 0.275 (0.030) | -0.009 | 0.032 | 0.724   | 0.932   |
|    | $\pi_{04} = 0.200$ | 0.193 (0.036) | -0.007 | 0.037 | 0.788   | 0.934   |

CP coverage probability, SE standard error

by applying the SEM algorithm. In particular, we carry out following hypothesis tests corresponding to the five sub-models:

- Exponential: $H_0 : \alpha = k = 1$ vs. $H_1$: at least one inequality in $H_0$;
- Rayleigh: $H_0 : \alpha = 1, k = 2$ vs. $H_1$: at least one inequality in $H_0$;
- Weibull: $H_0 : \alpha = 1$ vs. $H_1 : \alpha \neq 1$;
- Generalized exponential: $H_0 : k = 1$ vs. $H_1 : k \neq 1$;
- Burr type X: $H_0 : k = 2$ vs. $H_1 : k \neq 2$.

Let $\hat{l}$ and $\hat{l}_0$ denote the unrestricted maximized log-likelihood value and the maximized log-likelihood value obtained under $H_0$, respectively. Then, by Wilk’s theorem, $\Lambda = -2(\hat{l} - \hat{l}_0) \sim \chi^2_{q^*}$ asymptotically under $H_0$ where $\chi^2_{q^*}$ represents a chi-squared distribution with $q^*$ degrees of freedom and $q^*$ denotes the difference in the number of parameters estimated to obtain $\hat{l}$ and $\hat{l}_0$. The $p$ values for the tests are compared against a significance level of 0.05 to decide whether to reject $H_0$ or not. Proportion of rejections of $H_0$ based on 1000 Monte Carlo runs are reported in Table 7 for every combination of the true and fitted models.
Table 7 Rejection rates of the true models based on the likelihood ratio test

| Fitted model       | True model (λ = 2.5) | Exponential | Rayleigh  | Weibull (k = 1.5) | Generalized exponential (α = 2) | Burr type X (α = 2) |
|--------------------|----------------------|-------------|-----------|-------------------|-------------------------------|-------------------|
| Exponential        | 0.044                | 1.000       | 1.000     | 1.000             | 1.000                         | 1.000             |
| Rayleigh           | 1.000                | 0.050       | 1.000     | 1.000             | 1.000                         | 1.000             |
| Weibull            | 0.052                | 0.040       | 0.068     | 0.542             | 0.596                         | 0.596             |
| Generalized exponential | 0.054            | 0.964       | 0.578     | 0.048             | 0.950                         | 0.950             |
| Burr type X        | 0.904                | 0.042       | 0.330     | 0.956             | 0.046                         |                   |

Bold values represent the observed significance levels
The observed significance level corresponding to every true lifetime distribution is close to the nominal level or significance level of 0.05. This implies that the chi-squared distribution provides a good approximation to the null distribution of the likelihood ratio test statistic. Next, we observe that when the true lifetime model is either exponential or Rayleigh, the rejection rates for the fitted Weibull model are 0.052 and 0.040, respectively. These rejection rates are close to 0.05 because both exponential and Rayleigh are contained within the Weibull distribution. On the other hand, when the true lifetime is generalized exponential or Burr type X, the rejection rates for the fitted Weibull lifetime are 0.542 and 0.596, respectively. These rejection rates are moderate because the Weibull distribution doesn’t accommodate the generalized exponential or Burr type X distributions as special cases. Based on some of the high rejection rates, we can conclude that the likelihood ratio test can discriminate between the following models: exponential and Rayleigh, Burr type X and exponential, Burr type X and generalized exponential, and generalized exponential and Rayleigh. As such, for a given data, there is a necessity to employ the likelihood ratio test for choosing the correct sub-model, if possible. If none of the sub-models provide an adequate fit, the proposed EW model should be used.

5.4 Comparison of the SEM Algorithm with the EM Algorithm

In this section, we compare the proposed SEM algorithm with the EM algorithm through the robustness of these algorithms with respect to the choice of initial values. Specifically, we study cases where the initial guesses of the model parameters are far away from the true values. For this purpose, for each model parameter, we provide an initial guess that differs from its true value by at least 50% and by at most 75%. Then, we run the SEM and EM algorithms using the same choice of initial values to make sure that the comparison between the two algorithms is fair. In Table 8, we present the percentage of divergent samples based on 500 Monte Carlo runs for different parameter settings. It is easy to see that for any considered parameter setting, the divergence percentage corresponding to the SEM algorithm is much less when compared to the EM algorithm. This clearly shows that the EM algorithm is sensitive to the choice of initial values, whereas the SEM algorithm is more robust. This, certainly, is a big advantage of the SEM algorithm and, hence, the SEM algorithm can be considered a preferred algorithm over the EM algorithm. It is interesting to note that when the true lifetime parameters are as considered in either setting 1 or setting 3, the percentage of divergent samples decreases with an increase in sample size. However, this is not true when the true lifetime parameters are as in setting 2. Similarly, for lifetime parameters as in settings 1 and 3, and for the SEM algorithm, the divergence percentages are smaller when the true cure rates are low. In this regard, for the EM algorithm, the divergence percentages are smaller for low cure rates, irrespective of the lifetime parameters.

It is also of our interest to compare the divergence rates of SEM and EM algorithms in terms of proportion of missing data. For this comparison, the initial values of the model parameters are chosen to be close to the true values. Noting that missing data arise due to censoring, we study both high and low censoring cases. In Table 9, we
Table 8  Divergence rates (in %) of the SEM and EM algorithms when the initial values deviate from the true values by at least 50% and at most 75%

| n   | Cure rate | Lifetime | Divergence % |
|-----|-----------|----------|--------------|
|     |           | SEM      | EM           |
| 200 | High      | Setting 1| 6.800 43.600 |
| 400 | High      | Setting 1| 0.800 37.600 |
| 200 | Low       | Setting 1| 4.800 39.000 |
| 400 | Low       | Setting 1| 0.000 38.200 |
| 200 | High      | Setting 2| 5.255 25.839 |
| 400 | High      | Setting 2| 6.083 25.816 |
| 200 | Low       | Setting 2| 5.422 24.397 |
| 400 | Low       | Setting 2| 7.012 23.780 |
| 200 | High      | Setting 3| 4.282 35.138 |
| 400 | High      | Setting 3| 0.565 29.378 |
| 200 | Low       | Setting 3| 1.200 32.200 |
| 400 | Low       | Setting 3| 0.712 28.632 |

Table 9  Divergence rates (in %) of the SEM and EM algorithms in terms of proportion of missing data for n = 200

| Censoring rate | Lifetime | Divergence % |
|               |         | SEM | EM |
| High           | Setting 1| 0.500  9.000  |
| Low            | Setting 1| 0.000  4.500  |
| High           | Setting 2| 0.500  1.000  |
| Low            | Setting 2| 0.000  0.000  |
| High           | Setting 3| 1.500  4.500  |
| Low            | Setting 3| 0.000  3.000  |

present the divergence rates for n = 200. For other sample sizes, the observations are similar and hence are not reported. It is clear that in all scenarios the divergence rate corresponding to the SEM algorithm is less than the divergence rate corresponding to the EM algorithm.

6 Analysis of Cutaneous Melanoma Data

Data description: An illustration of our proposed model with EW lifetime and proposed estimation technique is presented in this section. Motivated by an example provided in Ibrahim et al. [25] which showed influences of cure fraction, we consider the data set on cutaneous melanoma (a type of malignant skin cancer) studied by the Eastern Cooperative Oncology Group (ECOG) where the patients were observed for the period between 1991 and 1995. The objective of the study was to assess the efficacy of the postoperative treatment with high dose of interferon alpha-2b drug to prevent recurrences of the cancer. Observed survival time (t, in years) representing either exact lifetime or censoring time, censoring indicator (δ = 0, 1) and nodule
Table 10 Descriptive summary of the observed survival times \((t)\) categorized by censoring indicator \((\delta)\) and nodule category \((x)\) corresponding to the melanoma data

| \(\delta\) | Measure | \(x = 1\) | \(x = 2\) | \(x = 3\) | \(x = 4\) |
|---------|---------|-------|-------|-------|-------|
|         | \(n\) (%) | 36 (8.633) | 53 (12.709) | 41 (9.832) | 55 (13.189) |
| Uncensored | Mean | 2.227 | 1.777 | 1.687 | 1.551 |
| \((\delta = 1)\) | SD | 0.889 | 1.006 | 1.248 | 1.114 |
| 185 (44.365%) | Median | 2.185 | 1.599 | 1.544 | 1.268 |
| | Min | 0.767 | 0.285 | 0.148 | 0.170 |
| | Max | 4.263 | 4.613 | 5.739 | 5.969 |
|         | \(n\) (%) | 75 (17.985) | 84 (20.143) | 46 (11.031) | 27 (6.474) |
| Censored | Mean | 4.265 | 4.217 | 4.312 | 4.609 |
| \((\delta = 0)\) | SD | 1.209 | 1.192 | 1.167 | 1.008 |
| 232 (55.635%) | Median | 4.208 | 4.112 | 4.474 | 5.043 |
| | Min | 0.791 | 1.325 | 1.139 | 2.927 |
| | Max | 7.012 | 6.976 | 6.623 | 6.045 |
|         | \(n\) (%) | 111 (26.618) | 137 (32.853) | 87 (20.863) | 82 (19.664) |
| Combined | Mean | 3.604 | 3.273 | 3.075 | 2.558 |
| 417 (100%) | SD | 1.467 | 1.636 | 1.782 | 1.801 |
| | Median | 3.537 | 3.387 | 3.047 | 1.966 |
| | Min | 0.767 | 0.285 | 0.148 | 0.170 |
| | Max | 7.012 | 6.976 | 6.623 | 6.045 |

\(n\) sample size, \(SD\) standard deviation, \(Min\) minimum, \(Max\) maximum

category \((x = 1, 2, 3, 4)\) based on tumor thickness are selected as the variables of interest for demonstrating the performance of our model. There are 427 observations in the data set; each observation corresponds to a patient in the study with respective nodule category information. Analysis is performed based on 417 patients’ data due to missing information on tumor thickness for the remaining 10 patients. Nodule category is taken as the only covariate for our illustration. A descriptive summary of the observed survival time categorized by censoring indicator and nodule category is presented in Table 10.

Assignment of initial parameter values Let us define \(\pi_{0x}, x = 1, 2, 3, 4\), to be the cure rate for the \(x\)-th nodule category. As indicated before and as can also be seen from Table 10, \(x = 1\) represents the group which is likely to have the best prognosis (i.e., highest cure rate), whereas \(x = 4\) represents the group likely to have the worst prognosis (i.e., lowest cure rate). Further, the censoring rates for \(x = 1\) and \(x = 4\) are 0.675 and 0.329, respectively. Using the monotone nature of the logistic-link function and assuming all censored individuals are cured, we obtain initial estimates of \(\beta_0\) and \(\beta_1\) by simultaneously solving

\[
\pi_{01} = \left[1 + \exp(\beta_0 + \beta_1)\right]^{-1} = 0.675
\]
\[
\pi_{04} = \left[1 + \exp(\beta_0 + 4\beta_1)\right]^{-1} = 0.329.
\]
Hence, the initial estimates $\beta(0)_0$ and $\beta(0)_1$ are $-1.212$ and $0.481$, respectively. On the other hand, the initial estimates for the EW lifetime parameters $\alpha$, $\lambda$ and $k$ are obtained in two steps. In the first step, we consider (6) and note that

$$\psi(t_i; \alpha, \lambda, k) = \log \left( - \log \left[ 1 - \left( 1 - S_s(t_i; \alpha, \lambda, k) \right)^{1/\alpha} \right] \right) = k \log t_i - k \log \lambda \tag{29}$$

is linear in $t_i$, where $i \in \Delta_1$. Hence, fixing $\alpha = \alpha_0$, ordinary least square estimates $\lambda_0$ of $\lambda$ and $k_0$ of $k$ are obtained by fitting a simple linear regression model with $\hat{\psi}(t; \alpha, \lambda, k)$ as the response and $t$ as the predictor. Here, $\psi(t_i; \alpha, \lambda, k) = \log \left( - \log \left[ 1 - \left( 1 - \hat{S}_s(t_i; \alpha, \lambda, k) \right)^{1/\alpha} \right] \right)$ and $\hat{S}_s(t_i; \alpha, \lambda, k)$ is the Kaplan–Meier estimate of the survival function evaluated at $t_i$ for the $i$-th individual with $i \in \Delta_1$.

In our case, $\alpha_0$ is chosen as 2. The Kaplan–Meier plots of the survival probabilities for the four nodule categories are presented in Fig. 1. In the second step, using (4), we define a likelihood function $L_s(\alpha, \lambda, k; t^*)$ as

$$L_s(\alpha, \lambda, k; t^*) = \prod_{i \in \Delta_1} f_s(t_i; \alpha, k, \lambda)$$

$$= \prod_{i \in \Delta_1} \left( \frac{\alpha k}{\lambda} \right) \left( \frac{t_i}{\lambda} \right)^{k-1} e^{-\left( t_i/\lambda \right)^k} \left[ 1 - e^{-\left( t_i/\lambda \right)^k} \right]^{\alpha-1}, \tag{30}$$

where $t^* = \{ (t_i; i \in \Delta_1) \}$. From here, $\log L_s(\alpha, \lambda, k; t^*)$ is then maximized with respect to $\alpha$, $\lambda$ and $k$ using numerical optimization routine in R with $\alpha_0$, $\lambda_0$ and $k_0$ as initial parameter guesses. Finally, the ML estimates of $\alpha$, $\lambda$ and $k$ are obtained as $\alpha^{(0)} = 1.983$, $\lambda^{(0)} = 1.326$ and $k^{(0)} = 1.214$. Hence, $\theta^{(0)} = (\beta_0^{(0)}, \beta_1^{(0)}, \alpha^{(0)}, \lambda^{(0)}, k^{(0)}) = (-1.212, 0.481, 1.983, 1.326, 1.214)$ is taken as the initial parameter guess for starting the iterative processes involved in the SEM algorithm.
Table 11  Estimates, standard errors and 95% confidence intervals of the model parameters and cure rates corresponding to the cutaneous melanoma data

| Parameter | Estimate | SE   | 95% CI          |
|-----------|----------|------|-----------------|
| \( \beta_0 \) | -1.117   | 0.283| (-1.672, -0.562) |
| \( \beta_1 \) | 0.495    | 0.113| (0.274, 0.716)  |
| \( \alpha \)  | 5.009    | 4.704| (0.000, 14.232) |
| \( \lambda \)  | 0.623    | 0.657| (0.000, 1.914)  |
| \( k \)       | 0.688    | 0.308| (0.083, 1.295)  |
| Cure rate  |          |      |                 |
| \( \pi_{01} \) | 0.6507   | 0.045| (0.563, 0.739)  |
| \( \pi_{02} \) | 0.5318   | 0.041| (0.451, 0.612)  |
| \( \pi_{03} \) | 0.4091   | 0.048| (0.315, 0.503)  |
| \( \pi_{04} \) | 0.2967   | 0.058| (0.184, 0.410)  |

\( SE \) standard error, \( CI \) confidence interval

Model fitting: As discussed in Sect. 4, model parameters are estimated by the SEM method. Point estimate, standard error (SE), and 95% confidence interval (CI) are displayed in Table 11 for both model parameters and cure rates of all nodules categories. The standard errors of the cure rate estimates are estimated using the delta method. No overlap is observed between the confidence intervals for \( \pi_{01} \) and \( \pi_{04} \) suggesting that cure rates for these groups are significantly different. Figure 2 presents plots corresponding to the overall population survival function \( \hat{S}_p (t; \theta) \), where

\[
\hat{S}_p (t_i; \theta) = \left[ \frac{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_i} \left\{ 1 - \left[ 1 - e^{-\left(\frac{t_i}{\hat{\lambda}}\right)^{\frac{1}{\hat{\alpha}}}} \right]^{\hat{\alpha}} \right\}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}} \right]
\]

evaluated at observed \( t_i \) for \( i = 1, \ldots, n \). The plot shows similar pattern as that of the Kaplan–Meier plot in Fig. 1. It can be seen that overall survival probability plots level off to points much higher than 0 (even when patients were followed up for more than 6 years), therefore, strongly indicating the presence of significant cure fractions.

Burn-in period: For the real data set, 10,000 iterations of the SEM algorithm are carried out. For each iteration, the SEM estimate for each parameter is plotted against iteration index and these are presented in Fig. 3. It is observed that all parameters show similar random behavior around the horizontal line with no discernible pattern, except for \( \alpha \). The plot for \( \alpha \) though doesn’t show upward, downward, or any other obvious pattern, yet the variability around the middle horizontal line is large and doesn’t show any obvious diminishing trend. This explains the large standard error that we have obtained corresponding to \( \alpha \). The middle horizontal lines correspond to the parameter values which return the maximized log-likelihood value after a burn-in period of 5000 iterations. The random oscillation with almost constant variance around the horizontal line indicates convergence of the SEM estimates to a stationary distribution. However, large variability in the estimates of \( \alpha \) is a concern and just taking the average over
the iterations after the burn-in period results in under-estimated variance. So, it is reasonable to consider the parameter estimates of the SEM algorithm as the one which return the maximized log-likelihood value after the burn-in period (see [38]).

*Model discrimination:* The cutaneous melanoma data set is further analyzed by fitting all nested sub-models of the EW lifetime distribution. The parameter estimates and corresponding standard errors are presented in Table 12. To verify the appropriateness of fitting EW lifetime distribution to the melanoma data under mixture cure rate setup, maximized log-likelihood ($\hat{l}$) values are calculated for all sub-models and formal hypotheses tests are carried out to test whether the sub-models deviate significantly from the model with EW lifetime distribution. By using the Wilk’s theorem, i.e.,

$$-2(\hat{l}_{EW} - \hat{l}_{sub}) \overset{a.s.}{\sim} \chi^2_\zeta,$$

where $\hat{l}_{EW}$ and $\hat{l}_{sub}$ are the respective maximized log-likelihood values under the EW model (alternative model) and sub-model (null model), and $\zeta$ is the difference in the number of parameters estimated, respective $p$ values for all sub-models are obtained (Table 12). The $p$ values indicate that all nested models, except the one fitted with generalized exponential distribution, are significantly different from the EW model, and hence are rejected. Further, Akaike information criterion

$$AIC = -2\hat{l}_{fit} + 2q$$

values for each fitted model are also presented in the same table where $\hat{l}_{fit}$ is the maximized log-likelihood value under the fitted model and $q$ denotes the number of parameters estimated. AIC values suggest that the generalized exponential (AIC = 1036.642) model provides the best fit. Hence, for the considered cutaneous melanoma

![Fig. 2 Overall population survival probability plot estimated pointwise by the SEM technique and categorized by nodule category](image-url)
data, the EW lifetime distribution reduces to the generalized exponential distribution. Note the closeness of the generalized exponential model to the EW model based on the AIC values.

We also check for the goodness-of-fit or adequacy of the proposed model using the estimated normalized randomized quantile residuals [23], where the residuals are estimated using the SEM estimates. Figure 4 presents the quantile–quantile (QQ) plot, which clearly suggests that the proposed mixture cure model with EW lifetime distribution provides an adequate fit to the cutaneous melanoma data. We also test for the normality of residuals using the Kolmogorov–Smirnov test. The $p$ value corresponding to the test turns out to be 0.934, which provides very strong evidence for the normality of residuals.

Comparison with semi-parametric piecewise linear mixture cure model: Our proposed mixture cure model with exponentiated Weibull lifetime distribution for the
| Measure     | Fitted models |
|-------------|---------------|
|             | EW | Exp | Ral | W  | GE  | Burr |
| \( \hat{\beta}_0 (SE(\hat{\beta}_0)) \) | -1.117 (0.283) | -0.936 (0.341) | -1.200 (0.260) | -1.154 (0.265) | -1.151 (0.269) | -1.168 (0.263) |
| \( \hat{\beta}_1 (SE(\hat{\beta}_1)) \) | 0.495 (0.113) | 0.607 (0.159) | 0.460 (0.100) | 0.463 (0.103) | 0.478 (0.105) | 0.460 (0.102) |
| \( \hat{\alpha} (SE(\hat{\alpha})) \) | 5.009 (4.704) | – | – | – | 2.441 (0.316) | 0.764 (0.074) |
| \( \hat{\lambda} (SE(\hat{\lambda})) \) | 0.623 (0.657) | 3.281 (0.549) | 2.218 (0.097) | 2.217 (0.135) | 1.239 (0.141) | 2.512 (0.181) |
| \( \hat{k} (SE(\hat{k})) \) | 0.688 (0.308) | – | – | 1.616 (0.105) | – | – |
| \( \hat{\ell} \) | -513.839 | -535.037 | -523.897 | -517.593 | -514.321 | -519.932 |
| AIC | 1037.678 | 1076.074 | 1053.794 | 1043.186 | 1036.642 | 1047.864 |
| \( p \) value | – | \( 6.221 \times 10^{-10} \) | \( 4.284 \times 10^{-5} \) | 0.023 | 0.617 | 0.002 |

\( \hat{\ell} \) maximized log-likelihood value, SE standard error, EW exponentiated Weibull, Exp exponential, Ral Rayleigh, W Weibull, GE generalized exponential, AIC Akaike Information Criterion
susceptible individuals is further compared (through AIC value and running time) with the mixture cure model where the susceptible lifetimes are estimated semi-parametrically. For this purpose, we approximate the hazard function of the susceptible lifetime by a piecewise linear function. A piecewise linear approximation (PLA) of the common hazard function \( h_s(t) \) of susceptible subjects is based on three factors, namely, the number of lines to be fitted \( (L^*) \), the chosen set of cut points \( \{\tau_0, \tau_1, \ldots, \tau_{L^*}\} \) related to \( L^* \) and the set of initial estimates \( \{\phi_0, \phi_1, \ldots, \phi_{L^*}\} \) of the hazard function at the cut points. Note that \( \tau_0 \leq \tau_1 \leq \cdots \leq \tau_{L^*} \) and \( \lim_{t \to \tau_l} h_s(t) = \phi_l \) for \( l = 0, 1, \ldots, L^* \) for ensuring continuity at the cut points [4]. Therefore, for \( t \geq 0 \),

\[
h_s(t) = h_s(t; \phi_0, \phi_1, \ldots, \phi_{L^*}) = \sum_{l=1}^{L^*} (\nu_l + \mu_l t) I_{[\tau_{l-1}, \tau_l]}(t), \tag{31}
\]

where \( \mu_l = \frac{\phi_l - \phi_{l-1}}{\tau_l - \tau_{l-1}} \) is the slope, \( \nu_l = \phi_l - \mu_l \tau_l \) is the intercept, and \( I_{[\tau_{l-1}, \tau_l]}(t) = 1 \) only if \( t \in [\tau_{l-1}, \tau_l] \), and is 0 elsewhere, for \( l = 0, 1, \ldots, L^* \). For \( l = 0, 1, \ldots, L^* \), \( \mu_l \) and \( \nu_l \) are obtained such that \( h_s(t) \geq 0 \) for any \( t \geq 0 \). The cure model under this setup is characterized by parameter vector \( (\beta_0, \beta_1, \phi_0, \phi_1, \ldots, \phi_{L^*})^T \). The estimate of the hazard function may vary depending on the choice of cut points \( \tau_l \) and corresponding \( \phi_l \) for \( l = 0, 1, \ldots, L^* \). Following [4], for our analysis, \( L^* \) is taken from \( \{1, 2, \ldots, 6\} \) indicating up to six lines are used for the hazard function approximation. For \( l = 0, 1, \ldots, (L^*-1) \), \( \tau_l \) is chosen as suitable sample quantile of the uncensored lifetimes \( t_i \) for \( i \in \Delta_1 \), and \( \tau_{L^*} = \max\{t_i; i = 1, \ldots, n\} \). Then, for \( l = 0, 1, \ldots, L^* \), initial estimates of \( \phi_l \) is obtained by applying a kernel-based hazard estimates (using \texttt{muhaz} function in R) and approximating hazard values at the cut points by interpolation,
Table 13 A comparison of mixture cure model with EW susceptible lifetime distribution and PLA-based susceptible lifetimes for the cutaneous melanoma data

| Model       | \( \hat{l} \)     | AIC         | Running time (in h) |
|-------------|-------------------|-------------|---------------------|
| PLA \((L^* = 1)\) | -521.7350        | 1051.4700  | 0.4307              |
| PLA \((L^* = 2)\) | -513.5740        | 1037.1480  | 0.8429              |
| PLA \((L^* = 3)\) | -514.3467        | 1040.6934  | 1.3346              |
| PLA \((L^* = 4)\) | -512.2277        | 1038.4554  | 2.2316              |
| PLA \((L^* = 5)\) | -512.2756        | 1040.5512  | 3.4786              |
| PLA \((L^* = 6)\) | -511.5842        | 1041.1684  | 9.3056              |
| EW (GE)     | -514.3210        | 1036.6420  | 0.0308              |

\( \hat{l} \) maximized log-likelihood value, EW exponentiated Weibull, \( L^* \) number of lines in the PLA

and \( \tau_0 = 0 \) with \( h_s(t) = v_{L^*} + \mu_{L^*}t \) for \( t \geq \tau_{L^*} \). We present a comparison of our proposed cure model with EW lifetime and cure model with PLA of the susceptible hazard function in Table 13 using the cutaneous melanoma data, where the parameters in both models are estimated by the SEM algorithm.

Results in Table 13 are based on 1000 iterations of the SEM algorithm with a burn-in period of 500 iterations. Larger iterations and burn-in periods are not considered since they don’t improve the estimation results and also avoid excessive computation time for the PLA-based models (Table 13). It can be observed that on increasing the number of cut points and lines for the PLA, the maximized log-likelihood value consistently increases in general. But, on comparing the AIC values, we note that the PLA cure model with \( L^* = 2 \) provides the lowest value of AIC. When compared to the proposed EW model, and noting that the EW model reduced to a generalized exponential model, the AIC value for the EW (GE) model turns out to be even smaller. Thus, the proposed mixture cure model with flexible EW lifetime distribution provides a better fit to the melanoma data when compared to the PLA-based semi-parametric mixture cure model. Also note that the other advantage of the EW model over the semi-parametric PLA-based model is with respect to the computation time (system running time in hours). As \( L^* \) increases, the computation time for the PLA model almost increases exponentially (0.4307 h. for \( L^* = 1 \) and 9.3056 hrs. for \( L^* = 6 \)). For the PLA model \((L^* = 2)\) with minimum AIC, the computation time is almost 27 times higher than the EW (GE) model. Moreover, to apply the PLA models, the cut points and corresponding initial hazard estimates should be chosen carefully to get optimum results. These factors provide justification in favor of applying a fully parametric model such as the one proposed in this paper, specifically when the parametric model is reasonably flexible and general.

7 Concluding Remarks

The main contribution of this manuscript is the development of the SEM algorithm in the context of mixture cure rate model when the lifetimes of the susceptible individuals
are modeled by the EW family of distributions. While we chose to demonstrate the performance of the SEM algorithm using the wider class of EW distributions, the proposed SEM algorithm can be easily extended to other flexible choices of lifetime distributions; see Pal and Balakrishnan [42] and Wang and Pal [58]. Different approaches of computing the estimates under the SEM framework have been discussed. An extensive Monte Carlo simulation study demonstrates the accuracy of the SEM algorithm in estimating the unknown model parameters. When compared with the well-known EM algorithm, we have shown that the proposed SEM algorithm is more robust to the choice of initial values than the EM algorithm. This can be seen as an advantage of the SEM algorithm over the EM algorithm. A detailed model discrimination study using the likelihood ratio test clearly shows that different sub-distributions of the EW distribution can be easily discriminated. Hence, blindly assuming a distribution for the lifetime is not recommended. Through the real cutaneous melanoma data, we have illustrated the flexibility of the proposed EW distribution. In this regard, we have seen that the assumption of the EW distribution allows formal tests of hypotheses to be performed to select the generalized exponential distribution as the best fitted distribution. In particular, we have seen that all other special cases of the EW distribution get rejected. When compared to the PLA-based semi-parametric mixture cure model, we have seen that our proposed parametric model with EW distribution provides a better fit in terms of smaller AIC value. As potential future works, we can consider more complicated cure models such as the ones that look at the elimination of risk factors after an initial treatment [41, 45, 53] and investigate the performance of the SEM algorithm. We can also think of extending the current framework, as studied in this paper, to accommodate interval censored data, as opposed to the commonly used right censored data [44, 59]. We are currently looking at some of these open problems and we hope to report our findings in future manuscripts.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42519-022-00274-8.

Acknowledgements The authors express their thanks to the Guest Editor and three anonymous reviewers for their careful reviews and useful comments and suggestions on an earlier version of this manuscript which led to this improved version.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Appendix: Development of the EM Algorithm

To implement the EM algorithm, we define the complete data likelihood function as:

$$L_C(\theta; t, \delta, X, \eta) \propto \prod_{i \in \Delta_1} f_p(t_i; \theta, \delta_i, x_i^*) \times \prod_{i \in \Delta_0} \left\{ \pi_0(x_i^*; \beta) \right\}^{1-\eta_i} \left\{ S_p(t_i; \theta, \delta_i, x_i^*) - \pi_0(x_i^*; \beta) \right\}^{\eta_i},$$

(32)
where $\eta = (\eta_1, \ldots, \eta_n)^T$ and $\pi_0(x_i^*; \beta) = \left\{1 + e^{x_i^T \beta}\right\}^{-1}$ is the cure rate. Equivalently, the expression for the complete data log-likelihood function is obtained as:

$$l_C(\theta; t, \delta, X, \eta) = \text{constant} + \sum_{i \in \Delta_1} \log f_p(t_i; \theta, \delta_i, x_i^*) + \sum_{i \in \Delta_0} (1 - \eta_i) \log \pi_0(x_i^*; \beta)$$

$$+ \sum_{i \in \Delta_0} \eta_i \log \left\{S_p(t_i; \theta, \delta_i, x_i^*) - \pi_0(x_i^*; \beta)\right\}. \quad (33)$$

For the mixture cure rate model, the expression given in (33) takes the following form:

$$l_C(\theta; t, \delta, X, \eta) = \text{constant} + n_1 (\log \alpha + \log k - k \log \lambda) + (k - 1) \sum_{i \in \Delta_1} \log t_i$$

$$- \sum_{i \in \Delta_1} \left(\frac{t_i}{\lambda}\right)^k + \sum_{i \in \Delta_1} (\alpha - 1) \log \left\{1 - e^{-(t_i/\lambda)^k}\right\}$$

$$+ \sum_{i \in \Delta_1} x_i^T \beta - \sum_{i=1}^n \log \left(1 + e^{x_i^T \beta}\right)$$

$$+ \sum_{i \in \Delta_0} \eta_i e^{x_i^T \beta} + \sum_{i \in \Delta_0} \eta_i \log \left\{1 - \left[1 - e^{-(t_i/\lambda)^k}\right]^\alpha\right\}. \quad (34)$$

**Steps Involved in the EM Algorithm**

Begin the iterative process by considering an initial estimate $\theta^{(0)} = (\beta^{(0)}, \alpha^{(0)}, k^{(0)}, \lambda^{(0)})^T$ of $\theta$. The choice of $\theta^{(0)}$ requires justifications based on background knowledge and some sample real-life data. For $r = 1, 2, \ldots$, let $\theta^{(r)}$ be the estimate of $\theta$ at the $r$-th step of the iteration. Then, $\theta^{(r+1)}$ is obtained using the following steps:

1. **E-Step** Find the conditional expectation $Q(\theta; \theta^{(r)}) = E\left\{l_C(\theta; t, \delta, X, \eta) | (\theta^{(r)}, t, \delta, X)\right\}$, which is given by

$$Q(\theta; \theta^{(r)}) = \text{constant} + n_1 (\log \alpha + \log k - k \log \lambda) + (k - 1) \sum_{i \in \Delta_1} \log t_i$$

$$- \sum_{i \in \Delta_1} \left(\frac{t_i}{\lambda}\right)^k + \sum_{i \in \Delta_1} (\alpha - 1) \log \left\{1 - e^{-(t_i/\lambda)^k}\right\}$$

$$+ \sum_{i \in \Delta_1} x_i^T \beta - \sum_{i=1}^n \log \left(1 + e^{x_i^T \beta}\right)$$

$$+ \sum_{i \in \Delta_0} E\left\{\eta_i | (\theta^{(r)}, t, \delta, X)\right\} e^{x_i^T \beta}$$

$$+ \sum_{i \in \Delta_0} E\left\{\eta_i | (\theta^{(r)}, t, \delta, X)\right\} \log \left\{1 - \left[1 - e^{-(t_i/\lambda)^k}\right]^\alpha\right\}. \quad (35)$$
where

\[
E \{ \eta_i \mid (\theta^{(r)}, t, \delta, X) \} = P \{ \eta_i = 1 \mid (\theta^{(r)}, t, \delta, X) \}
\]

\[
= P \left\{ \eta_i = 1 \mid (\theta^{(r)}, Y_i > t_i, x_i^*, i \in \Delta_0) \right\}
\]

\[
= \frac{P \left\{ Y_i > t_i \mid (\eta_i = 1, \theta^{(r)}, x_i^*, i \in \Delta_0) \right\} P \{ \eta_i = 1 \mid (\theta^{(r)}, x_i^*, i \in \Delta_0) \}}{P \left\{ Y_i > t_i \mid (\theta^{(r)}, x_i^*, i \in \Delta_0) \right\}}
\]

\[
= \frac{S_p \left( t_i; \theta^{(r)}, \delta_i, x_i^* \right) - \pi_0 \left( x_i^*; \beta^{(r)} \right)}{S_p \left( t_i; \theta^{(r)}, \delta_i, x_i^* \right)}
\]

\[
= 1 - \frac{\pi_0 \left( x_i^*; \beta^{(r)} \right)}{S_p \left( t_i; \theta^{(r)}, \delta_i, x_i^* \right)}.
\]

(36)

2. **M-Step** Find

\[
\theta^{(r+1)} = \left( \beta^{(r+1)}, \alpha^{(r+1)}, k^{(r+1)}, \lambda^{(r+1)} \right)^T = \arg \max_{\theta} Q \left( \theta; \theta^{(r)} \right).
\]

(37)

The maximization step can be carried out using multidimensional unconstrained optimization methods such as Nelder–Mead simplex search algorithm or quasi Newton methods such as BFGS algorithm. These algorithms are available in statistical software R version 4.0.3 under General Purpose Optimization package called optimr()

3. **Convergence** Check if the stopping or convergence criterion for the iterative process is met. For our analysis, we consider that the EM algorithm has converged to a local maxima if

\[
\max_{1 \leq k' \leq d+4} \left| \frac{\theta_{k'}^{(r+1)} - \theta_{k'}^{(r)}}{\theta_{k'}^{(r)}} \right| < \epsilon,
\]

(38)

where \( \theta_{k'}^{(r)} \) and \( \theta_{k'}^{(r+1)} \) are the \( k' \)-th component of \( \theta^{(r)} \) and \( \theta^{(r+1)} \), respectively, and \( \epsilon \) is a tolerance such as 0.001.

If the condition in (38) is satisfied, then the iterative process is stopped and \( \theta^{(r)} \) is considered as the ML estimate of \( \theta \).

**References**

1. Amico M, Van Keilegom I, Legrand C (2019) The single-index/Cox mixture cure model. Biometrics 75:452–462
2. Balakrishnan N, Barui S, Milienos F (2017) Proportional hazards under Conway–Maxwell–Poisson cure rate model and associated inference. Stat Methods Med Res 26(5):2055–2077
3. Balakrishnan N, Barui S, Milienos FS (2022) Piecewise linear approximations of baseline under proportional hazards based COM-Poisson cure models. Commun Stat Simul Comput. https://doi.org/10.1080/03610918.2022.2032157
4. Balakrishnan N, Koutras M, Milienos F (2016) Piecewise linear approximations for cure rate models and associated inferential issues. Methodol Comput Appl Probab 18(4):937–966
5. Balakrishnan N, Pal S (2012) EM algorithm-based likelihood estimation for some cure rate models. J Stat Theory Pract 6:698–724
6. Balakrishnan N, Pal S (2013) Lognormal lifetimes and likelihood-based inference for flexible cure rate models based on COM-Poisson family. Comput Stat Data Anal 67:41–67
7. Balakrishnan N, Pal S (2014) COM-Poisson cure rate models and associated likelihood-based inference with exponential and Weibull lifetimes. In: Frenkel I, Karagrigoriou A, Lisnianski A, Kleyner A (eds) Applied reliability engineering and risk analysis: probabilistic models and statistical inference applied reliability engineering and risk analysis: probabilistic models and statistical inference. Wiley, Chichester, pp 308–348
8. Balakrishnan N, Pal S (2015) 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. Comput Stat 30:151–189
9. Balakrishnan N, Pal S (2015) Likelihood inference for flexible cure rate models with gamma lifetimes. Commun Stat Theory Methods 44(19):4007–4048
10. Balakrishnan N, Pal S (2016) Expectation maximization-based likelihood inference for flexible cure rate models with Weibull lifetimes. Stat Methods Med Res 25:1535–1563
11. Barui S, Grace YY (2020) Semiparametric methods for survival data with measurement error under additive hazards cure rate models. Lifetime Data Anal 26(3):421–450
12. Bedair KF, Hong Y, Al-Khalidi HR (2021) Copula-frailty models for recurrent event data based on Monte Carlo EM algorithm. J Stat Comput Simul 91(17):3530–3548
13. Bedair KF, Hong Y, Li J, Al-Khalidi HR (2016) Multivariate frailty models for multi-type recurrent event data and its application to cancer prevention trial. Comput Stat Data Anal 10(11):61–173
14. Cariou C, Chehdi K (2008) Unsupervised texture segmentation/classification using 2-d autoregressive modeling and the stochastic expectation-maximization algorithm. Pattern Recogn Lett 29(7):905–917
15. Celeux G, Chauveau D, Diebolt J (1996) Stochastic versions of the EM algorithm: an experimental study in the mixture case. J Stat Comput Simul 55(4):287–314
16. Celeux G, Diebolt J (1985) The EM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. Comput Stat Q 2:73–82
17. Celeux G, Diebolt J (1992) A stochastic approximation type EM algorithm for the mixture problem. Stoch Int J Probab Stoch Process 41(1–2):119–134
18. Chauveau D (1995) A stochastic EM algorithm for mixtures with censored data. J Stat Plan Inference 46(1):1–25
19. Chen M-H, Ibrahim JG, Sinha D (1999) A new Bayesian model for survival data with a surviving fraction. J Am Stat Assoc 94:909–919
20. Davies K, Pal S, Siddiqua JA (2021) Stochastic EM algorithm for generalized exponential cure rate model and an empirical study. J Appl Stat 48:2112–2135
21. Diebolt J, Celeux G (1993) Asymptotic properties of a stochastic EM algorithm for estimating mixing proportions. Stoch Model 9(4):599–613
22. Diebolt J, Ip EH (1995) A stochastic EM algorithm for approximating the maximum likelihood estimate. https://www.osti.gov/biblio/49148
23. Dunn PK, Smyth GK (1996) Randomized quantile residuals. J Comput Graph Stat 5:236–244
24. Fletcher R (2013) Practical methods of optimization. Wiley, New York
25. Ibrahim JG, Chen M-H, Sinha D (2005) Bayesian survival analysis. Wiley, New York
26. Khan SA (2018) Exponentiated Weibull regression for time-to-event data. Lifetime Data Anal 24(2):328–354
27. Kosovlic N, Barui S (2022) A hard EM algorithm for prediction of the cured fraction in survival data. Comput Stat 37:817–835
28. Kuk AJ, Chen C-H (1992) A mixture model combining logistic regression with proportional hazards regression. Biometrika 79:531–541
29. López-Cheda A, Cao R, Van Jácome MA, Keilegom I (2017) Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models. Comput Stat Data Anal 105:144–165
30. Lu W, Ying Z (2004) On semiparametric transformation cure models. Biometrika 91:2331–343
31. Majakwara J, Pal S (2019) On some inferential issues for the destructive COM-Poisson-generalized gamma regression cure rate model. Commun Stat Simul Comput 48(10):3118–3142
32. Marschner IC (2001) Miscellanea on stochastic versions of the algorithm. Biometrika 88(1):281–286
33. Meeker WQ (1987) Limited failure population life tests: application to integrated circuit reliability. Technometrics 29(1):51–65
34. Mudholkar GS, Hutson AD (1996) The exponentiated Weibull family: some properties and a flood data application. Commun Stat Theory Methods 25(12):3059–3083
35. Mudholkar GS, Srivastava DK (1993) Exponentiated Weibull family for analyzing bathtub failure-rate data. IEEE Trans Reliab 42(2):299–302
36. Musta E, Patilea V, Van Keilegom I (2020) A presmoothing approach for estimation in mixture cure models. arXiv:2008.05338
37. Nadarajah S, Cordeiro GM, Ortega EM (2013) The exponentiated Weibull distribution: a survey. Stat Pap 54(3):839–877
38. Nielsen SF (2000) The stochastic EM algorithm: estimation and asymptotic results. Bernoulli 6(3):457–489
39. Pal M, Ali MM, Woo J (2006) Exponentiated Weibull distribution. Statistica (Bologna) 66(2):139–147
40. Pal S (2021) A simplified stochastic EM algorithm for cure rate model with negative binomial competing risks: an application to breast cancer data. A simplified stochastic EM algorithm for cure rate model with negative binomial competing risks: An application to breast cancer data. Stat Med 40:6387–6409
41. Pal S, Balakrishnan N (2016) Destructive negative binomial cure rate model and EM-based likelihood inference under Weibull lifetime. Stat Probab Lett 116:9–20
42. Pal S, Balakrishnan N (2017) An EM type estimation procedure for the destructive exponentially weighted Poisson regression cure model under generalized gamma lifetime. J Stat Comput Simul 87(6):1107–1129
43. Pal S, Balakrishnan N (2017) Expectation maximization algorithm for Box–Cox transformation cure rate model and assessment of model misspecification under Weibull lifetimes. IEEE J Biomed Health Inform 22:926–934
44. Pal S, Balakrishnan N (2017) Likelihood inference for COM-Poisson cure rate model with interval-censored data and Weibull lifetimes. Stat Methods Med Res 26:2093–2113
45. Pal S, Balakrishnan N (2017) Likelihood inference for the destructive exponentially weighted Poisson cure rate model with Weibull lifetime and an application to melanoma data. Comput Stat 32:429–449
46. Pal S, Balakrishnan N (2018) Likelihood inference based on EM algorithm for the destructive length-biased Poisson cure rate model with Weibull lifetime. Commun Stat Simul Comput 47:644–660
47. Pal S, Majakwara J, Balakrishnan N (2018) An EM algorithm for the destructive COM-Poisson regression cure rate model. Metrika 81(2):143–171
48. Pal S, Roy S (2020) A new non-linear conjugate gradient algorithm for destructive cure rate model and a simulation study: illustration with negative binomial competing risks. Commun Stat Simul Comput. https://doi.org/10.1080/03610918.2020.1819321
49. Pal S, Roy S (2021) On the estimation of destructive cure rate model: a new study with exponentially weighted Poisson competing risks. Stat Neerl 75(3):324–342
50. Patilea V, Van Keilegom I (2020) A general approach for cure models in survival analysis. Ann Stat 48:2323–2346
51. Peng Y, Dear KB (2000) A nonparametric mixture model for cure rate estimation. Biometrics 56(1):237–243
52. Peng Y, Yu B (2021) Cure models: methods, applications and implementation. Chapman and Hall/CRC, London
53. Rodrigues J, de Castro M, Balakrishnan N, Cancho VG (2011) Destructive weighted Poisson cure rate models. Lifetime Data Anal 17:333–346
54. Rodrigues J, de Castro M, Cancho VG, Balakrishnan N (2009) COM-Poisson cure rate survival models and an application to a Cutaneous Melanoma data. J Stat Plan Inference 139:3605–3611
55. Sy JP, Taylor JM (2000) Estimation in a Cox proportional hazards cure model estimation in a cox proportional hazards cure model. Biometrics 56:227–236
56. Taylor JMG (1995) Semi-parametric estimation in failure time mixture models. Biometrics 51:899–907
57. Tsodikov A, Ibrahim J, Yakovlev A (2003) Estimating cure rates from survival data. J Am Stat Assoc 98:1063–1078
58. Wang P, Pal S (2022) A two-way flexible generalized gamma transformation cure rate model. Stat Med 41(13):2427–2447
59. Wiangnak P, Pal S (2018) Gamma lifetimes and associated inference for interval-censored cure rate model with COM-Poisson competing cause. Commun Stat Theory Methods 47(6):1491–1509
60. Yakovlev AY, Tsodikov AD, Asselain B (1996) Stochastic models of tumor latency and their biostatistical applications (1). World Scientific, Singapore
61. Ye Z, Ng HKT (2014) On analysis of incomplete field failure data. Ann Appl Stat 8(3):1713–1727
62. Yin G, Ibrahim JG (2005) Cure rate models: a unified approach. Can J Stat 33(4):559–570
63. Zeng D, Yin G, Ibrahim JG (2006) Semiparametric transformation models for survival data with a cure fraction. J Am Stat Assoc 101:670–684

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.