Received: 22 February 2023 | Revised: 18 August 2023 | Accepted: 1 September 2023

DOI: 10.1002/sim.9904

RESEARCH ARTICLE

Statistics in Medicine WILEY

On the estimation of interval censored destructive negative binomial cure model

Jodi Treszoks | **Suvra Pal**

Department of Mathematics, University of Texas at Arlington, Arlington, Texas, USA

Correspondence
Suvra Pal, Department of Mathematics, University of Texas at Arlington, 411 S. Nedderman Drive, Arlington, TX 76019, USA.
Email: suvra.pal@uta.edu

Funding information
National Institute of General Medical Sciences, Grant/Award Number: R15GM150091

In this article, a competitive risk survival model is considered in which the initial number of risks, assumed to follow a negative binomial distribution, is subject to a destructive mechanism. Assuming the population of interest to have a cure component, the form of the data as interval-censored, and considering both the number of initial risks and risks remaining active after destruction to be missing data, we develop two distinct estimation algorithms for this model. Making use of the conditional distributions of the missing data, we develop an expectation maximization (EM) algorithm, in which the conditional expected complete log-likelihood function is decomposed into simpler functions which are then maximized independently. A variation of the EM algorithm, called the stochastic EM (SEM) algorithm, is also developed with the goal of avoiding the calculation of complicated expectations and improving performance at parameter recovery. A Monte Carlo simulation study is carried out to evaluate the performance of both estimation methods through calculated bias, root mean square error, and coverage probability of the asymptotic confidence interval. We demonstrate the proposed SEM algorithm as the preferred estimation method through simulation and further illustrate the advantage of the SEM algorithm, as well as the use of a destructive model, with data from a children's mortality study.

KEYWORDS
children's mortality, competing causes, interval censoring, SEM algorithm

1 | INTRODUCTION

While conventional survival analysis methods assume all individuals to be susceptible to the event of interest given sufficient follow-up time,[1-3] advancements in medical science have led to a proportion of patients responding favorably to treatment and experiencing no recurrence of disease until the end of a long follow-up time. These patients are described as recurrence-free survivors. An effective treatment for certain types of diseases (eg, cancer, heart disease) may result in some of these recurrence-free patients not presenting with recurrence for a sufficiently long period after the follow-up time due to the disease reaching an undetectable and harmless stage. These survivors are termed as cured or non-susceptible. The remaining patients are subject to recurrence of the disease and are described as susceptible. The population can then be viewed as a mixture of susceptible and non-susceptible patients. Survival models that allow for a proportion of the population to be cured are termed cure rate models. Cure models are instrumental in reliability and biomedical science and commonly used in cancer clinical trials.

The first cure rate model, known in the literature as the mixture cure rate model, was developed by Boag,[4] where the author proposed a cure component to represent the proportion of patients not susceptible to recurrence of disease.[5] In the mixture cure rate model, the population survival function of a time-to-event variable Y is given by

$$S_{\text{pop}}(y) = p_0 + (1 - p_0)S_{\text{susc}}(y),$$

where p_0 is the cure rate and $S_{\text{susc}}(\cdot)$ is a proper survival function of the susceptible patients only. Interested readers may refer to the research monographs by Maller and Zhou[6] and Peng and Yu[7] for a book-length account on mixture cure rate models. While the mixture cure rate model is the most used cure rate model, Chen et al[8] addressed several drawbacks of the mixture cure rate model in developing the promotion time cure rate model. The promotion time cure rate model introduces a latent competing risk variable, which follows a Poisson distribution, to describe the underlying biological process producing the failure time. In this case, the population survival function is given by

$$S_{\text{pop}}(y) = e^{-\eta(1-S(y))},$$

where η represents the mean number of competing risks and $S(\cdot)$ represents the common survival function corresponding to the promotion time of each competing risk. Rodrigues et al[9] proposed the flexible Conway Maxwell Poisson (COM-Poisson) cure rate model which includes the cure rate models proposed by Boag[4] and Chen et al[8] as special cases while allowing for both over-dispersion and under-dispersion. Rodrigues et al[10] extended the COM-Poisson model by introducing a natural destructive process to the original number of competing risks, while Borges et al[11] introduced to this destructive model a biological dependence between the initiated cells. The development of likelihood inference and finding maximum likelihood estimates (MLEs) in cure rate models has been well studied, largely making use of the expectation maximization (EM) algorithm. Interested readers may refer to Balakrishnan and Pal,[12-16] Pal and Balakrishnan,[17-20] Pal et al,[21] and Majakwara and Pal,[22] among others. A novel variation of the EM algorithm for the destructive weighted Poisson cure rate model was developed by Gallardo et al[23] This variation, which makes use of conditional distributions of the missing data, allows the complete log-likelihood function to be split into simpler functions which may be maximized independently.

In a practical scenario such that subjects are not monitored continuously but rather observed at discrete times, an observed lifetime may be said only to lie in an interval.[24] Though interval-censored data have been described extensively in the literature, limited studies deal with interval-censored data in a cure rate setup. Banerjee and Carlin[25] extended the model of Chen et al[8] to accommodate interval-censored data while considering spatial correlation and used Markov Chain Monte Carlo methods to develop Bayesian inference. Covariate effects were introduced to the mixture cure rate model under interval censoring and the EM algorithm was developed for this framework.[26] Pal and Balakrishnan[18] formulated the EM algorithm for the COM-Poisson cure rate model parameters with interval censoring; see also Wiangnak and Pal.[27] Most recently, an efficient EM algorithm was formulated for the destructive shifted Poisson cure rate model under interval censoring.[28]

In this article, we extend the destructive cure rate model developed by Rodrigues et al[10] to accommodate interval-censored data assuming initial competing risks follow a negative binomial distribution.[29] It is of particular interest for practitioners to obtain accurate and precise estimates for the cure probability of subjects, as a high probability of cure can inform a patient's continued treatment, allowing them to avoid further unnecessary and potentially harmful interventions. Similarly, a low probability of cure can help a clinician to decide on the need to develop adjuvant therapies. In the context of public health applications, a child may be susceptible to death due to a number of risk factors, such as poorly sanitized water or air pollution, poor nutrition of the mother while in utero, or inadequate access to medical facilities. These factors can be viewed as working together to produce a quantity of competing risks of death to a child. In this context, a cure to child mortality would involve reducing the competing risks through remedial measures such as improved access to healthcare and nutrition for child-bearing women. It is therefore of interest for practitioners to accurately assess the prognostic effect of demographic indicators when developing strategies to reduce risk of child mortality. While most established parameter estimation methods for the destructive negative binomial cure rate model with right censored data employ an EM algorithm with a profile likelihood approach (for the shape parameter of the negative binomial distribution) which may result in inaccurate and imprecise estimates for cure rate, we aim to identify an efficient estimation method that can provide accurate and precise cure estimates. To this end, we propose a novel stochastic variation of the EM algorithm, called the stochastic EM (SEM) algorithm, to find the MLEs of the destructive negative binomial cure rate model in the presence of interval censored data. To compare the SEM algorithm with the commonly used EM algorithm, we also develop the steps of the EM algorithm since such an algorithm does not exist in the context

of destructive negative binomial cure rate model with interval censored data. We show that the SEM algorithm easily avoids calculation of complicated conditional expectations and hence allows simultaneous maximization of all model parameters. We further show that the SEM algorithm results in more accurate and precise estimates of model parameters, specifically those that are related to the estimation of cure rate. In addition, through a real data analysis, we show the destructive mechanism when taken into account results in a better model fit.

The rest of this article is organized as follows: in Section 2, we present the destructive negative binomial cure rate model under interval censoring and define the complete and observed log-likelihood functions. In Section 3, the steps of the EM algorithm are formulated in detail, while Section 4 presents the steps of the SEM algorithm in detail and describes the advantages of the proposed SEM algorithm over the EM algorithm. A Monte Carlo simulation study is presented in Section 5 to compare the performances of the SEM and EM algorithms. In Section 6, the SEM and EM algorithms are applied to real data on children's mortality arising from the 2018 Nigeria Demographic Health Survey, illustrating both the advantage of the SEM algorithm and the practicality of using a destructive model in this context. Finally, in Section 7 we summarize and discuss directions for future research.

2 | DESTRUCTIVE NEGATIVE BINOMIAL CURE RATE MODEL UNDER INTERVAL CENSORING

Consider a practical scenario where an unobserved number of risk factors (also called competing risks or competing causes) compete to produce an event of interest (eg, death or recurrence of a disease). For example, several malignant cells are related to the occurrence of a cancerous tumor. However, the number of malignant cells produced by nature remains unobserved. Similarly, in the context of child mortality, while a number of adverse environmental factors can act jointly to produce a quantity of competing risks of death to a child, the quantity of fatal risks is not directly observable. Hence, we term malignant cells or quantity of fatal risks as competing risks.

Let M be a random variable denoting the unobserved number of initial competing risks. Since M is unobserved, we assume it to follow a negative binomial distribution with probability mass function (pmf)

$$P[M=m;\eta,\phi]=\frac{\Gamma(\phi^{-1}+m)}{\Gamma(\phi^{-1})m!}\left(\frac{\phi\eta}{1+\phi\eta}\right)^{m}(1+\phi\eta)^{-\phi^{-1}},\ m=0,1,2,\ \dots\ , \tag{1}$$

where $\eta>0,\ \phi>0$, and $\Gamma(\cdot)$ is the complete gamma function. We call (1) as a negative binomial distribution with parameters $r=\phi^{-1}$ and $p=\frac{\phi\eta}{1+\phi\eta}$. Suppose that an intervention takes place resulting in a quantity $D(\leq M)$ competing risks remaining active, each of which could bring about the event of interest with probability p. For example, after a patient goes through a chemotherapy or a radiation session, it is expected that a certain number of initial competing risks will be destroyed. Similarly, risk of death to a child may be decreased through environmental intervention. After the destructive process, only $D(\leq M)$ active competing risks are still capable of producing the event of interest. Consistent with the existing literature (see Rodrigues et al),[10] we assume a common activation probability for each risk factor within a subject. Later, to capture heterogeneity in the population we propose to vary the activation probabilities across subjects through incorporation of subject-related characteristics or covariates. To model the destruction process of risk factors, we can associate a Bernoulli random variable X_j, such that $P(X_j=1)=p$, to each competing risk M. Note that p denotes the activation probability. The remaining quantity of competing risks not destroyed can then be modeled as

$$D=\begin{cases}X_1+X_2+\ \dots\ +X_M, & \text{if } M>0,\\ 0, & \text{if } M=0.\end{cases}$$

Note that unlike the standard cure models that do not look at the destruction process of risk factors (see Pal),[29] the proposed approach allows a subject to be cured even in the presence of initial risk factors. For the cure rate model with the pmf of the number of competing risks as in (1), the distribution of active risks D is given by

$$P[D=d;\eta,\phi,p]=\frac{\Gamma(\phi^{-1}+d)}{\Gamma(\phi^{-1})d!}\left(\frac{\phi\eta p}{1+\phi\eta p}\right)^{d}(1+\phi\eta p)^{-\phi^{-1}},\ d=0,1,2,\ \dots\ ,M. \tag{2}$$

A detailed proof of (2) has been provided by Pal and Balakrishnan.[17]

Let W_a represent the time taken for the ath active risk to produce the event. Consistent with the existing literature and conditioned on D, the waiting times $W_a, a = 1, 2, \ldots, D$, are assumed to be identically and independently distributed with a common distribution function $F(t; \lambda)$, where λ is an unknown set of parameters, and are also independent of D. Because the number of active risks D and waiting times W_a associated with each given risk are latent variables, one typically only observes the minimum of $W_1, \ldots, W_D$. To accommodate the presence of a cured proportion, the time-to-event or lifetime is defined as

$$T = \min\{W_0, W_1, \ldots, W_D\},$$

where W_0 is such that $P[W_0 = \infty] = 1$. Utilizing the results of Rodrigues et al,[10] the population survival function can be expressed as

$$S_{\text{pop}}(t; \phi, \eta, p, \lambda) = P(T > t) = \left[1 + \phi\eta p F(t; \lambda)\right]^{-\phi^{-1}}. \tag{3}$$

For the sake of simplicity, we will use $F(\cdot)$ instead of $F(\cdot; \lambda)$. Similarly, we will use $S(\cdot)$ instead of $S(\cdot; \lambda)$, where $S(\cdot; \lambda) = 1 - F(\cdot; \lambda)$. Note that $S_{\text{pop}}(\cdot)$ is an improper survival function since $\lim_{y\to\infty} S_{\text{pop}} > 0$. From Equation (3), the long-term survival probability or the cure fraction is given by

$$p_0 = S_{\text{pop}}(\infty; \phi, \eta, p, \lambda) = \left[1 + \phi\eta p\right]^{-\phi^{-1}}. \tag{4}$$

We note that the parameters in the negative binomial cure rate model in Equation (1) possess biological interpretations. Parameter η is related to the mean number of initial competing risks, while ϕ accounts for the inter-individual variance in the quantity of initial competing risks. To capture heterogeneity in the population, we bring in the effects of covariates. For this purpose, we propose to use two sets of covariates, $\boldsymbol{z_1}$ and $\boldsymbol{z_2}$, where $\boldsymbol{z_1}$ is related to the initial number of competing risks through the parameter η and $\boldsymbol{z_2}$ is related to activation probability for each risk, with corresponding link functions

$$\eta = e^{\boldsymbol{z_1'}\boldsymbol{\beta_1}} \text{ and } p = \frac{e^{\boldsymbol{z_2'}\boldsymbol{\beta_2}}}{1 + e^{\boldsymbol{z_2'}\boldsymbol{\beta_2}}},$$

where $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ represent vectors of regression coefficients. We note that the destructive negative binomial cure rate model is not identifiable.[10,30] To circumvent this problem, we make sure that $\boldsymbol{z_1}$ and $\boldsymbol{z_2}$ share no common elements, and an intercept term is excluded either from $\boldsymbol{\beta_1}$ or from $\boldsymbol{\beta_2}$. Let $\boldsymbol{\psi} = (\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \lambda, \phi,)$ denote the vector of unknown parameters. Depending on the specific needs of an application, it is straightforward to also relate another set of covariates to a suitable component of λ.

2.1 | Form of data

We consider the scenario where true lifetimes are not exactly observed and are subject to interval censoring. Let T_i denote the ith subject's true failure time (unobserved), for $i = 1, 2, \ldots, n$, with n denoting the number of subjects in the study. To develop an interval censoring scheme, let us assume the ith subject is observed at times $\boldsymbol{Y_i} = (Y_{0(i)}, Y_{1(i)}, \ldots, Y_{j-1(i)}, Y_{j(i)})$, with $0 < Y_{0(i)} < Y_{1(i)} < \cdots < Y_{j(i)} < \infty$. In practice, observation occurs only until either the event of interest has occurred or the lifetime is right-censored. In the case that the lifetime is observed, the interval $(Y_{j-1(i)}, Y_{j(i)})$ is known to contain T_i, and we consider the lifetime to be interval-censored. If the ith subject's failure time did not occur prior to the last observation time, we consider the lifetime right-censored and may only conclude the lifetime is contained in the interval $(Y_{j(i)}, \infty)$. We denote the observed data as $\boldsymbol{D}_{\text{obs}} = (\boldsymbol{l}, \boldsymbol{r}, \boldsymbol{\delta})$, where $\boldsymbol{l} = (l_1, l_2, \ldots, l_n)$ such that

$$l_i = \begin{cases} Y_{j-1(i)}, & \text{if lifetime is interval-censored,} \\ Y_{j(i)}, & \text{otherwise,} \end{cases}$$

$\boldsymbol{r} = (r_1, r_2, \ldots, r_n)$ such that

$$r_i = \begin{cases} Y_{j(i)}, & \text{if lifetime is interval-censored,} \\ \infty, & \text{otherwise,} \end{cases}$$

$\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_n)$ such that

$$\delta_i = \begin{cases} 1, & \text{if lifetime is interval-censored,} \\ 0, & \text{otherwise.} \end{cases}$$

2.2 | Observed and complete likelihood functions

Based on the observed data $(\boldsymbol{l}, \boldsymbol{r}, \boldsymbol{\delta})$, the observed likelihood function under non-informative censoring is given by

$$L_o(\boldsymbol{\psi}|\boldsymbol{D}_{\text{obs}}) \propto \prod_{i=1}^{n} \{S_{\text{pop}}(Y_{j-1(i)}) - S_{\text{pop}}(Y_{j(i)})\}^{\delta_i} \{S_{\text{pop}}(Y_{j(i)})\}^{1-\delta_i},$$

while the observed log-likelihood function is given by

$$l_o(\boldsymbol{\psi}|\boldsymbol{D}_{\text{obs}}) \propto \sum_{i=1}^{n} \left[\delta_i \log \left\{S_{\text{pop}}(Y_{j-1(i)}) - S_{\text{pop}}(Y_{j(i)})\right\} + (1-\delta_i) \log S_{\text{pop}}(Y_{j(i)})\right]. \tag{5}$$

The complete data are denoted by $\boldsymbol{D}_{\text{comp}} = (\boldsymbol{l}, \boldsymbol{r}, \boldsymbol{\delta}, \boldsymbol{M}, \boldsymbol{D})$ which includes the observed data and the missing data, where $\boldsymbol{M} = (M_1, M_2, \ldots, M_n)$ and $\boldsymbol{D} = (D_1, D_2, \ldots, D_n)$ are the missing data. Following Yakovlev and Tsodikov,[31] the joint distribution of the complete data can be expressed as

$$f(l_i, r_i, \delta_i, m_i, d_i) = f(l_i, r_i, \delta_i | D_i = d_i) P(D_i = d_i | M = m_i) P(M_i = m_i). \tag{6}$$

The second and third terms in the product above are well defined, and by the results of Treszoks and Pal,[28] we have that

$$f(l_i, r_i, \delta_i | D_i = d_i) = \left[d_i \{S(Y_{j-1(i)}) - S(Y_{j(i)})\}\right]^{\delta_i} \{S(Y_{j(i)})\}^{d_i - \delta_i}. \tag{7}$$

Using (6) and (7), the complete data likelihood function may be defined as

$$\begin{aligned} L_c(\boldsymbol{\psi}|\boldsymbol{D}_{\text{comp}}) &= \prod_{i=1}^{n} f(l_i, r_i, \delta_i, m_i, d_i) \\ &= \prod_{i=1}^{n} f(Y_{j-1(i)}, Y_{j(i)}, \delta_i, m_i, d_i) \\ &= \prod_{i=1}^{n} S(Y_{j(i)})^{d_i - \delta_i} [d_i \{S(Y_{j-1(i)}) - S(Y_{j(i)})\}]^{\delta_i} \binom{m_i}{d_i} p_i^{d_i} (1-p_i)^{m_i - d_i} \\ &\quad \times \frac{\Gamma(\phi^{-1} + m_i)}{\Gamma(\phi^{-1}) m_i!} \left(\frac{\phi\eta}{1+\phi\eta}\right)^{m_i} (1+\phi\eta)^{-\phi^{-1}}. \end{aligned}$$

The complete data log-likelihood function can then be written as

$$l_c(\boldsymbol{\psi}|\boldsymbol{D}_{\text{comp}}) = l_c(\boldsymbol{\xi}) + l_c(\boldsymbol{\beta}_2) + l_c(\lambda) + K, \tag{8}$$

where

$$\begin{aligned} l_c(\boldsymbol{\xi}) &= \sum_{i=1}^{n} \left\{ \log[\Gamma(\phi^{-1} + m_i)] - \phi^{-1} \log(1 + \phi\eta_i) + m_i \log \frac{\phi\eta_i}{1+\phi\eta_i} \right\} - n \log[\Gamma(\phi^{-1})], \\ l_c(\boldsymbol{\beta}_2) &= \sum_{i=1}^{n} \left\{ d_i \log p_i + (m_i - d_i) \log(1-p_i) \right\}, \\ l_c(\lambda) &= \sum_{i=1}^{n} \left\{ (d_i - \delta_i) \log S(Y_{j(i)}) + \delta_i \log[S(Y_{j-1(i)}) - S(Y_{j(i)})] \right\}, \end{aligned}$$

$\boldsymbol{\xi} = (\boldsymbol{\beta_1}, \phi)$, and K is a constant independent of model parameters, given by

$$K = \sum_{i=1}^{n} \left\{ \delta_i \log d_i + \log\left(\frac{m_i!}{d_i!(m_i - d_i)!} \right) - \log(m_i!) \right\}. \tag{9}$$

3 | EM ALGORITHM

In this section, we present the construction of the EM algorithm to produce estimates of the parameters for the destructive negative binomial cure rate model under interval censoring. The first step in implementing the EM algorithm[32] requires taking the conditional expectation of a complete data log-likelihood function, such as (8), given some proposed parameter values and the observed data. In the second step, the conditional expected complete data log-likelihood function is maximized. While the traditional approach involves maximizing a complicated function consisting of numerous model parameters, this approach can lack both computational efficiency and robustness with respect to initial values. The implementation of the EM algorithm proposed here is motivated by the works of Gallardo et al[23] By using the conditional distributions of the missing data to decompose the conditional expectation of $l_c(\boldsymbol{\psi}|\boldsymbol{D_{comp}})$ into three simpler functions that can each be maximized independently, this formulation of the EM algorithm offers much greater efficiency.

Because taking the conditional expectation of the complete log-likelihood function in Equation (8) requires computing $E\left[\log\left\{\Gamma(M_i + \phi^{-1})\right\}|\boldsymbol{D_{obs}}; \boldsymbol{\psi}\right]$ which has no closed form, simultaneous maximization of all parameters in Equation (8) is challenging. A commonly used method to overcome this difficulty in computation is to use a profile likelihood approach, in which the EM algorithm is performed and the observed data log-likelihood function value is calculated for a fixed set of distinct admissible values of ϕ. The MLE of ϕ is the value for which the maximized log-likelihood function value is the maximum. We present the necessary expressions for the EM algorithm using this profile likelihood approach.

Let $\boldsymbol{\psi}^{(k)} = (\boldsymbol{\xi}^{(k)}, \boldsymbol{\beta}_2^{(k)}, \lambda^{(k)})$ be the estimate of $\boldsymbol{\psi}$ at the kth iteration, and let $Q(\boldsymbol{\psi}|\boldsymbol{\psi}^{(k)})$ denote the conditional expectation of $l_c(\boldsymbol{\psi}|\boldsymbol{D}_{\text{comp}})$ given the observed data and $\boldsymbol{\psi}^{(k)}$. Then by (8),

$$\begin{aligned} Q(\boldsymbol{\psi}|\boldsymbol{\psi}^{(k)}) &= E[l_c(\boldsymbol{\xi})|\boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}^{(k)}] + E[l_c(\boldsymbol{\beta}_2)|\boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}^{(k)}] + E[l_c(\lambda)|\boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}^{(k)}] + K^* \\ &= Q_1(\boldsymbol{\xi}|\boldsymbol{\psi}^{(k)}) + Q_2(\boldsymbol{\beta}_2|\boldsymbol{\psi}^{(k)}) + Q_3(\lambda|\boldsymbol{\psi}^{(k)}) + K^*, \end{aligned} \tag{10}$$

with

$$\begin{aligned} Q_1(\boldsymbol{\xi}|\boldsymbol{\psi}^{(k)}) &= \sum_{i=1}^{n} \left\{ \widetilde{M}_i^{(k)} \log \frac{\phi\eta_i}{1 + \phi\eta_i} - \frac{\log(1 + \phi\eta_i)}{\phi} \right\}, \\ Q_2(\boldsymbol{\beta}_2|\boldsymbol{\psi}^{(k)}) &= \sum_{i=1}^{n} \left\{ \widetilde{D}_i^{(k)} \log p_i + (\widetilde{M}_i^{(k)} - \widetilde{D}_i^{(k)}) \log(1 - p_i) \right\}, \\ Q_3(\lambda|\boldsymbol{\psi}^{(k)}) &= \sum_{i=1}^{n} \left\{ (\widetilde{D}_i^{(k)} - \delta_i) \log S(Y_{j(i)}) + \delta_i \log[S(Y_{j-1(i)}) - S(Y_{j(i)})] \right\}, \end{aligned}$$

where $\widetilde{D}_i^{(k)} = E(D_i|\boldsymbol{D}_{\text{obs}}, \boldsymbol{\psi}^{(k)})$, $\widetilde{M}_i^{(k)} = E\left(M_i|\boldsymbol{D}_{\text{obs}}, \boldsymbol{\psi}^{(k)}\right)$, and K^* is a constant independent of $\boldsymbol{\psi}$. The following results will be needed to compute the required conditional expectations.

Proposition 1. *For the cure rate model with the pmf of the number of initial competing risks as in Equation (1), the conditional distribution of $M_i - \delta_i$ given the observed data under interval censoring is given by*

$$M_i - \delta_i|\boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi} \sim \text{NB}\left(\phi^{-1} + \delta_i, \frac{\phi\eta_i[1 - p_i F(Y_{j(i)})]}{1 + \phi\eta_i} \right). \tag{11}$$

A proof of Proposition 1 is provided in Section A2.1 of the Supplemental Material.

Proposition 2. *For the cure rate model with the pmf of the number of active risks as in Equation (2), the conditional distribution of $D_i - \delta_i$ given the observed data under interval censoring is given by*

$$D_i - \delta_i|\boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi} \sim \text{NB}\left(\phi^{-1} + \delta_i, \frac{\phi\eta_i p_i S(Y_{j(i)})}{1 + \phi\eta_i p_i} \right). \tag{12}$$

A proof of Proposition 2 is provided in Section A2.2 of the Supplemental Material.

Applying the results of propositions 1 and 2, we can compute $\widetilde{D}_i^{(k)}$ and $\widetilde{M}_i^{(k)}$ as

$$\widetilde{D}_i^{(k)} = \delta_i + \frac{\eta_i^{(k)} p_i^{(k)} S(Y_{j(i)})(1+\delta_i\phi)}{1+\phi\eta_i^{(k)} p_i^{(k)} F(Y_{j(i)})},$$

$$\widetilde{M}_i^{(k)} = \delta_i + \frac{\left\{1 - p_i^{(k)} F(Y_{j(i)})\right\}\eta_i^{(k)}(1+\phi\delta_i)}{1+\phi\eta_i^{(k)} p_i^{(k)} F(Y_{j(i)})},$$

where $\eta_i = e^{z_{1i}'\beta_1}$ and $p_i = \frac{e^{z_{2i}'\beta_2}}{1+e^{z_{2i}'\beta_2}}$.

The next step is to maximize $Q_1(\boldsymbol{\xi}|\boldsymbol{\psi}^{(k)})$, $Q_2(\boldsymbol{\beta_2}|\boldsymbol{\psi}^{(k)})$, and $Q_3(\boldsymbol{\lambda}|\boldsymbol{\psi}^{(k)})$ independently with respect to $\boldsymbol{\xi}$, $\boldsymbol{\beta_2}$, and $\boldsymbol{\lambda}$, respectively. This can be done using readily available optimization routines in R such as the "optim()" function. In this regard, interested readers may also look at new optimization techniques studied by Pal and Roy.[33-35] The expectation and maximization steps are then repeated until a specified convergence criterion is satisfied. The steps of the EM algorithm can be summarized as follows:

Step 1 (Expectation step or E-step):

For $i = 1, \ldots, n$, compute $\widetilde{D}_i^{(k)}$ and $\widetilde{M}_i^{(k)}$ given parameter estimates $\boldsymbol{\psi}^{(k)}$ with a fixed value of ϕ.

Step 2 (Maximization step or M-step):

Given $\widetilde{\boldsymbol{D}}^{(k)} = \left(\widetilde{D}_1^{(k)}, \ldots, \widetilde{D}_n^{(k)}\right)$ and $\widetilde{\boldsymbol{M}}^{(k)} = \left(\widetilde{M}_1^{(k)}, \ldots, \widetilde{M}_n^{(k)}\right)$, find $\boldsymbol{\psi}$ that maximizes Equation (10) in relation to $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$, and $\boldsymbol{\lambda}$, respectively, to obtain an improved estimate $\boldsymbol{\psi}^{(k+1)}$.

Step 3 (Iterative step):

The E-step and M-step are repeated until a suitable convergence criterion is met. For this purpose, we use the relative difference in successive values of the estimates, $\left|\frac{\boldsymbol{\psi}^{(k+1)} - \boldsymbol{\psi}^{(k)}}{\boldsymbol{\psi}^{(k)}}\right|$, as stopping criterion with a tolerance value of 10^{-4}.

To apply the profile likelihood approach in the EM algorithm, we first select an initial grid of distinct admissable values of ϕ. We then employ the EM algorithm for each fixed value of ϕ and compute the observed log-likelihood for each estimate. If the log-likelihood is monotone decreasing (increasing), we decrease (increase) the grid and apply the EM algorithm to all prospective values until a maximum log-likelihood value is achieved. The value of ϕ which attains the maximized log-likelihood function value and corresponding estimates of other model parameters are taken as the MLEs.

3.1 | On the convergence of the EM algorithm

Following the approach of Wu,[36] for parameters $\boldsymbol{\psi} \in \boldsymbol{\Psi}$, we assume the following conditions hold:

(i) $\boldsymbol{\Psi}$ is a subset in d-dimensional Euclidean space $\mathbb{R}^d$.
(ii) $\boldsymbol{\Psi}_{\boldsymbol{\psi}^{(0)}} = \{\boldsymbol{\psi} \in \boldsymbol{\Psi} : L_o(\boldsymbol{\psi}) \geq L_o(\boldsymbol{\psi}^{(0)})\}$ is compact for any $L_o(\boldsymbol{\psi}^{(0)}) > -\infty$.
(iii) $L_o(\boldsymbol{\psi})$ is continuous in $\boldsymbol{\Psi}$ and differentiable in the interior of $\boldsymbol{\Psi}$.

By consequence of conditions (i)-(iii), we have $\{L_o(\boldsymbol{\psi}^{(p)})\}_{p\geq 0}$ is bounded above for any $\boldsymbol{\psi}^{(0)} \in \boldsymbol{\Psi}$. Taking this consequence in conjunction with the observation that any EM sequence $\{\boldsymbol{\psi}^{(p)}\}$ increases the likelihood, we conclude that $\left\{L_o(\boldsymbol{\psi}^{(p)})\right\}$ converges monotonically to some L_o^*. Further, we cite that the objective function $Q(\boldsymbol{\psi}'|\boldsymbol{\psi})$, as in Equation (10), is continuous in both $\boldsymbol{\psi}$ and $\boldsymbol{\psi}'$ for the case of a curved exponential family. Because the negative binomial distribution is a curved exponential family for fixed $r = \phi^{-1}$, we use the results of the following theorem developed by Wu,[36] which we state without further proof.

Proposition 3. *Suppose $Q(\boldsymbol{\psi}'|\boldsymbol{\psi})$ is continuous in both $\boldsymbol{\psi}$ and $\boldsymbol{\psi}'$. Then all the limit points of any instance $\{\boldsymbol{\psi}^{(p)}\}$ of the EM algorithm are stationary points of L_o and $L_o(\boldsymbol{\psi}^{(p)})$ converges monotonically to $L_o^* = L_o(\boldsymbol{\psi}^*)$ for some stationary point $\boldsymbol{\psi}^*$.*

4 | SEM ALGORITHM

While the EM algorithm as formulated will converge to a stationary point by proposition 3, a well-known drawback to the EM algorithm is that convergence to a global or even a local maximum is not guaranteed. As with other Newton-based

methods, such as the Newton Raphson method, the resulting point of convergence may be a saddle point close to the starting value rather than a maximum. Due to the stochastic nature of the SEM algorithm, it is free of this saddle point problem.[37] Because the pmf of the missing data given $\boldsymbol{D}_{\mathbf{obs}}$ and $\boldsymbol{\psi}$ (provided in Section A2.1 of the Supplemental Material) is positive for all $\boldsymbol{\psi} \in \boldsymbol{\Psi}$, a random sequence $\{\boldsymbol{\psi}^{(\boldsymbol{p})}\}$ produced by the SEM is an irreducible, homogeneous Markov chain. Further, if $\{\boldsymbol{\psi}^{(\boldsymbol{p})}\}$ is ergodic, then it converges to the unique stationary probability distribution of this Markov chain.[38] To verify convergence, it is suggested to inspect a trace plot of the sequence of estimates versus the iterations for random fluctuations.[39] SEM estimators are demonstrated to be efficient under some suitable regularity conditions. The SEM algorithm is also known to be insensitive to starting values and performs well for small sample sizes. Further, unlike the EM, the SEM does not require computation of complicated conditional expectations and hence allows simultaneous maximization of all parameters.[29,40] These benefits lead us to consider the SEM algorithm as the proposed estimation method for the destructive negative binomial cure model under interval censoring.

In the SEM algorithm, the expectation step (E-step) of the EM algorithm is replaced by a stochastic step (S-step), in which each missing datum in the complete log-likelihood function is replaced by a value randomly generated from the conditional distribution of the missing data given the observed data and current estimates of the parameters.[38,41] The S-step synthesizes a pseudo-complete data set, comprised of the observed data and randomly generated substitutes for the missing data, then the maximization step (M-step) involves maximizing the complete log-likelihood function based on the complete sample. Considering the complete data to be $\boldsymbol{D}_{\text{comp}} = (\boldsymbol{l}, \boldsymbol{r}, \boldsymbol{\delta}, \boldsymbol{M}, \boldsymbol{D})$, the development of the SEM algorithm makes use of Propositions 1 and 2 to randomly generate the missing data, $\boldsymbol{M}$ and $\boldsymbol{D}$, given the observed data and current parameter estimates.

4.1 | Steps of the SEM algorithm

Step 1 (Initial guess): Start with an initial guess of the parameter $\boldsymbol{\psi}^{(0)} = \left(\boldsymbol{\xi}^{(0)}, \boldsymbol{\beta}_2^{(0)}, \lambda^{(0)}\right)$ and the observed data $\boldsymbol{D}_{\mathbf{obs}}$.

Step 2: Stochastic step or S-step: Replace each missing datum, m_i and $d_i, i = 1, 2, \ldots, n$, in the complete data log-likelihood function $l_c(\boldsymbol{\psi}|\boldsymbol{D}_{\text{comp}})$ by a value randomly generated using the conditional distributions in Equations (11) and (12). Namely, generate $m_i - \delta_i$ from a negative binomial distribution with parameters $r = \phi^{-1} + \delta_i$ and $p = \frac{\phi\eta_i[1-p_iF(Y_{j(i)})]}{1+\phi\eta_i}$, where the parameters are evaluated at $\boldsymbol{\psi} = \boldsymbol{\psi}^{(0)}$. Similarly, generate $d_i - \delta_i$ from a negative binomial distribution with parameters $r = \phi^{-1} + \delta_i$ and $p = \frac{\phi\eta_i p_i S(Y_{j(i)})}{1+\phi\eta_i p_i}$, with parameters taking value $\boldsymbol{\psi} = \boldsymbol{\psi}^{(0)}$. Denote the generated values of m_i and d_i by $\widehat{m}_i^{(0)}$ and $\widehat{d}_i^{(0)}$, respectively, for all $i = 1, 2, \ldots, n$. Replace each unobserved m_i and d_i in $l_c(\boldsymbol{\psi}|\boldsymbol{D}_{\text{comp}})$ by $\widehat{m}_i^{(0)}$ and $\widehat{d}_i^{(0)}$, respectively, and denote the resulting function as

$$l_c\left(\boldsymbol{\psi}; \widehat{\boldsymbol{m}}^{(0)}, \widehat{\boldsymbol{d}}^{(0)}\right) = l_c\left(\boldsymbol{\xi}; \widehat{\boldsymbol{m}}^{(0)}\right) + l_c\left(\boldsymbol{\beta}_2; \widehat{\boldsymbol{m}}^{(0)}, \widehat{\boldsymbol{d}}^{(0)}\right) + l_c\left(\lambda; \widehat{\boldsymbol{d}}^{(0)}\right) + \widehat{K}^{(0)},$$

where

$$l_c\left(\boldsymbol{\xi}; \widehat{\boldsymbol{m}}^{(0)}\right) = \sum_{i=1}^{n}\left\{\log\left[\Gamma\left(\phi^{-1} + \widehat{m}_i^{(0)}\right)\right] - \phi^{-1}\log(1+\phi\eta_i) + \widehat{m}_i^{(0)}\log\frac{\phi\eta_i}{1+\phi\eta_i}\right\} - n\log(\Gamma(\phi^{-1})),$$

$$l_c\left(\boldsymbol{\beta}_2; \widehat{\boldsymbol{m}}^{(0)}, \widehat{\boldsymbol{d}}^{(0)}\right) = \sum_{i=1}^{n}\left\{\widehat{d}_i^{(0)}\log p_i + \left(\widehat{m}_i^{(0)} - \widehat{d}_i^{(0)}\right)\log(1-p_i)\right\},$$

$$l_c\left(\lambda; \widehat{\boldsymbol{d}}^{(0)}\right) = \sum_{i=1}^{n}\left\{\left(\widehat{d}_i^{(0)} - \delta_i\right)\log S(Y_{j(i)}) + \delta_i\log\left[S(Y_{j-1(i)}) - S(Y_{j(i)})\right]\right\},$$

and

$$\widehat{K}^{(0)} = \sum_{i=1}^{n}\left\{\delta_i\log\widehat{d}_i^{(0)} + \log\left(\frac{\widehat{m}_i!^{(0)}}{\widehat{d}_i!^{(0)}\left(\widehat{m}_i^{(0)} - \widehat{d}_i^{(0)}\right)!}\right) - \log\left(\widehat{m}_i!^{(0)}\right)\right\},$$

with $\widehat{\boldsymbol{m}}^{(0)}$ and $\widehat{\boldsymbol{d}}^{(0)}$ denoting the vectors of $\widehat{m}_i^{(0)}$ and $\widehat{d}_i^{(0)}$ values, respectively.

Step 3 (Maximization or M-Step): Maximize $l_c\left(\boldsymbol{\psi}; \widehat{\boldsymbol{m}}^{(0)}, \widehat{\boldsymbol{d}}^{(0)}\right)$ with respect to $\boldsymbol{\psi}$ to find an improved estimate of $\boldsymbol{\psi}$. This is a matter of maximizing $l_c\left(\boldsymbol{\xi}; \widehat{\boldsymbol{m}}^{(0)}\right)$ with respect to $\boldsymbol{\xi}$, $l_c\left(\boldsymbol{\beta_2}; \widehat{\boldsymbol{m}}^{(0)}, \widehat{\boldsymbol{d}}^{(0)}\right)$ with respect to $\boldsymbol{\beta_2}$, and $l_c\left(\lambda; \widehat{\boldsymbol{d}}^{(0)}\right)$ with respect to λ, independently. Denote the improved estimates of $\boldsymbol{\xi}$, $\boldsymbol{\beta_2}$, and λ by $\boldsymbol{\xi}^{(1)}, \boldsymbol{\beta}_2^{(1)}, \lambda^{(1)}$, respectively, where

$$\boldsymbol{\xi}^{(1)} = \arg\max_{\boldsymbol{\xi}} l_c\left(\boldsymbol{\xi}; \widehat{\boldsymbol{m}}^{(0)}\right), \quad \boldsymbol{\beta}_2^{(1)} = \arg\max_{\boldsymbol{\beta}_2} l_c\left(\boldsymbol{\beta}_2; \widehat{\boldsymbol{m}}^{(0)}, \widehat{\boldsymbol{d}}^{(0)}\right) \quad \text{and} \quad \lambda^{(1)} = \arg\max_{\lambda} l_c\left(\lambda; \widehat{\boldsymbol{d}}^{(0)}\right).$$

Since the missing data was replaced in the S-step, the M-step may be implemented by maximizing the complete data log-likelihood using standard optimization techniques such as the "Nelder-Mead" method readily available in R software.

Step 4 (Iterative step): Using the updated estimate $\boldsymbol{\psi}^{(1)} = (\boldsymbol{\xi}^{(1)}, \boldsymbol{\beta}_2^{(1)}, \lambda^{(1)})$ from Step 3, repeat steps 2 and 3 R times to obtain sequence $\boldsymbol{\psi}^{(k)}, k = 1, 2, \ldots, R$. While this sequence of estimates does not converge to a single point, the resulting Markov chain rapidly converges to a stationary distribution, given some regularity conditions are satisfied.[42]

Step 5 (Burn-in and MLE): The stationary distribution of estimates is achieved after a sufficiently long burn-in period of length r, and the MLE of $\boldsymbol{\psi}$ can be obtained by discarding the first r iterations. With the remaining $R - r$ iterations, the MLE may be obtained by averaging over the estimates (a method we will denote as "MLE (mean)") or by calculating the observed data log-likelihood function in Equation (5) for each $\boldsymbol{\psi}^{(k)}, k = r, 2, \ldots, R$, and taking $\boldsymbol{\psi}^{(k)}$ as the MLE for which the observed log-likelihood function is maximized (a method we will denote as "MLE (max)"). The length of the burn-in period is dependent on the form of the data, and as pointed out by Nielsen,[43] greater amounts of missing data can lead to longer necessary burn-in periods. However, with moderate missing data rates, a burn-in period of 100 iterations and an additional 1000 estimates are sufficient to estimate the parameters.[44,45] It is recommended to inspect a trace plot to validate the sufficiency of the burn-in period and adjust the length of the burn-in period as needed.

5 | SIMULATION STUDY

In this section, we evaluate the performance of the proposed EM and SEM algorithms. We first compare the efficacy of the proposed algorithms to direct maximization of the log-likelihood function (DM) with regards to parameter recovery. Demonstrating the proposed algorithms produce estimates with greater accuracy and efficiency than direct maximization, we further discriminate between the performance of the EM and SEM algorithms by comparing the resulting maximized log-likelihood values after the same generated datasets and initial values were used with both estimation methods. Finally, we consider the performance of the EM and SEM algorithms under model mis-specification. This empirical study partially mimics the real melanoma dataset which was used for illustrative purposes by Rodrigues et al[10] using covariates of treatment group (0: treatment, 1: placebo) and tumor thickness (in mm).

5.1 | Data generation

To simulate covariate data we first generated treatment group, denoted by z_1, from a Bernoulli distribution with probability of success 0.5. Noting that tumor thickness values in the melanoma data set range from 0.1 to 17.42 mm, we generated tumor thickness values, denoted by z_2, from a uniform U(0.1, 20) distribution. Parameter η is linked to treatment group only, using $\eta = e^{\beta_{11} z_1}$, and parameter p to tumor thickness only, using $p = \frac{\exp(\beta_{20} + \beta_{21} z_2)}{1 + \exp(\beta_{20} + \beta_{21} z_2)}$. Only one regression parameter corresponds to parameter η in order to avoid problems with identifiability in the sense of Li et al[30] Because a higher cure rate, and consequently a smaller value of η, is expected for the treatment group than the placebo group, a positive value of β_{11} is chosen, $\beta_{11} = 1$, to be consistent with this expectation. To select regression parameters for p, we propose a proportional increase in activation probability as tumor thickness increases and select values $\beta_{20} = -2$, $\beta_{21} = 0.1$ in accordance with this observation. We assume the waiting time W to follow a Weibull distribution with shape parameter $\frac{1}{\lambda_1}$ and scale parameter $\frac{1}{\lambda_2}$, and density function given by

$$f(t) = f(t; \lambda) = \frac{1}{\lambda_1 t} (\lambda_2 t)^{\frac{1}{\lambda_1}} e^{-(\lambda_2 t)^{\frac{1}{\lambda_1}}}, \quad t > 0, \lambda_1 > 0, \lambda_2 > 0. \tag{13}$$

Note that one can also use any other parametric distribution for W or choose to use a semi-parametric or a completely non-parametric model for W. Random censoring was introduced through censoring time C following exponential

distribution with rate α, where α can be chosen to achieve a desired censoring proportion. We chose α to be 0.05 which resulted in approximately 60% cured observations and 70% censored observations. To generate the observed data $(l_i, r_i, \delta_i), i = 1, 2, \ldots, n$, we execute the following steps:

Step 1: Generate censoring time C_i, competing risks $M_i \sim \text{NB}\left(\phi^{-1}, \frac{\phi\eta_i}{1+\phi\eta_i}\right)$, and damaged cells $D_i|M_i = m_i \sim \text{Bin}(m_i, p_i)$;
Step 2: If $D_i = 0$, then $(l_i, r_i, \delta_i) = (C_i, \infty, 0)$ and data generation is complete.
Step 3: If $D_i > 0$, generate times to event due to each non-eliminated risk, $W_j, j = 1, 2, \ldots, D_i$, from the considered Weibull distribution with parameter λ;
Step 4: Set $Y_i = \min\{W_1, W_2, \ldots, W_{D_i}\}$;
(a) If $Y_i > C_i$, then $(l_i, r_i, \delta_i) = (C_i, \infty, 0)$ and data generation is complete.
(b) If $Y_i \leq C_i$, set $\delta_i = 1$ and generate l_{1i} from U(0, 1) distribution and l_{2i} from U(0.2, 0.7) distribution. Construct intervals $(0, l_{1i}], (l_{1i}, l_{1i} + l_{2i}], \ldots, (l_{1i} + k \times l_{2i}, \infty], k = 1, 2, \cdots$, and select (l_i, r_i) that satisfies $l_i < Y_i \leq r_i$.

The choice of the lifetime parameter λ was obtained by equating the mean and variance of the underlying Weibull distribution to fixed values. For this purpose, we considered two different choices for the variance as 1.5 and 3 with a fixed mean value of 5, which yields two suitable choices for λ. We also choose two different true values of ϕ as 0.5 and 0.8. These specifications for λ and ϕ give way to four parameter settings which we denote as $\boldsymbol{\psi_1} = (1, -2, 0.1, 0.215, 0.183, 0.5)$, $\boldsymbol{\psi_2} = (1, -2, 0.1, 0.215, 0.183, 0.8)$, $\boldsymbol{\psi_3} = (1, -2, 0.1, 0.316, 0.179, 0.5)$, and $\boldsymbol{\psi_4} = (1, -2, 0.1, 0.316, 0.179, 0.8)$. Sample sizes of both $n = 200$ and $n = 300$ are used in order to observe the performance of the algorithms under small and moderate sample sizes. This study applies the algorithms proposed in Sections 3 and 4 to interval-censored data from the destructive negative binomial cure rate model, simulated using the parameters and data generation methods outlined above. The standard errors of the estimates are obtained by inverting the observed information matrix. All simulations are done using the R statistical software (version 4.2.1) and all results are based on $M = 250$ Monte Carlo runs. Computational codes for both EM and SEM algorithms are available in Section A1 of the Supplemental Material.

5.2 | Parameter estimation

To find an initial guess for the model parameters, we employ the following selection method: for a given parameter Γ, initial guess Γ_{init} is generated such that $\Gamma_{\text{init}} = \Gamma + \text{U}(0, 0.2)|\Gamma|$. To employ the profile likelihood approach to estimate ϕ in the EM algorithm, we select the initial grid for ϕ as $\{0.05, 0.1, \ldots, 2.05\}$. To implement the SEM algorithm, we first inspect trace plots for each parameter setting and identify the number of SEM iterations R of 1100 and burn-in period of $r = 100$ iterations to be sufficient. We also check both methods, MLE (mean) and MLE (max), for estimating model parameters. However, the MLE (mean) method is seen to produce large bias in the estimates for ϕ, irrespective of sample size. Consequently, we use the MLE (max) method to select parameter estimates for all SEM results. Now, the way the SEM algorithm is developed and the SEM estimates are obtained, that is, through the use of the MLE (max) method, it is ensured that the SEM estimates results in the highest log-likelihood value. Direct maximization of the observed log-likelihood function (DM) is used to produce parameter estimates as follows: we select initial guess for parameters as outlined above, then maximize (5) using the "optim()" function in R. We repeat this process 50 times to produce 50 sets of estimates, then choose as the MLE the set which yields the largest observed log-likelihood function value.

Tables 1-4 present the simulation results when the true value of $\boldsymbol{\psi}$ is taken as $\boldsymbol{\psi_1}, \boldsymbol{\psi_2}, \boldsymbol{\psi_3}$, and $\boldsymbol{\psi_4}$, respectively. We first note that for all parameter settings, SEM produces smaller biases and root mean square errors (RMSEs) of the parameters associated with the cure probability, that is, $\beta_{11}, \beta_{20}, \beta_{21}$, and ϕ. We also note that regardless of the true value of ϕ, the DM and EM approaches overestimate ϕ. Though the RMSEs of DM- and EM-produced ϕ estimates are high in all settings, this may be attributed to the relative flatness of the log-likelihood function with respect to ϕ, as shown in Figure 1 for sample size 200 and setting $\boldsymbol{\psi} = \boldsymbol{\psi_1}$. As compared to the SEM approach, larger biases and RMSEs of both DM- and EM-produced ϕ, $\boldsymbol{\beta_1}$, and $\boldsymbol{\beta_2}$ estimates indicate that the SEM algorithm performs better than both the EM algorithm and direct maximization of the observed log-likelihood with regard to both accuracy and precision. Further, the profile likelihood method used in the EM algorithm precludes the computation of standard error for ϕ estimates through the inversion of the observed information matrix, and the treatment of ϕ as fixed leads to underestimation of the standard error for other parameters. This underestimation is seen in the coverage probabilities of the asymptotic confidence intervals, where undercoverage is most prominent for the lifetime parameters λ_1 and λ_2. In all settings, the SEM approach produces empirical coverage probabilities that are close to the nominal level for the regression and lifetime parameters, with only slight undercoverage

TABLE 1 Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP) under setting $\boldsymbol{\psi = \psi_1}$.

Approach	n	Parameter	Estimate (SE)	Bias	RMSE	95% CP
DM	200	$\beta_{11} = 1$	1.150 (0.505)	0.150	0.542	0.960
		$\beta_{20} = -2$	−2.397 (0.840)	−0.397	1.610	0.968
		$\beta_{21} = 0.1$	0.254 (0.230)	0.154	1.025	0.940
		$\lambda_1 = 0.215$	0.202 (0.029)	−0.013	0.033	0.872
		$\lambda_2 = 0.183$	0.181 (0.011)	−0.002	0.011	0.884
		$\phi = 0.5$	1.224 (1.797)	0.724	1.988	0.896
	300	$\beta_{11} = 1$	1.126 (0.404)	0.126	0.440	0.936
		$\beta_{20} = -2$	−2.122 (0.504)	−0.122	0.572	0.976
		$\beta_{21} = 0.1$	0.137 (0.075)	0.037	0.122	0.960
		$\lambda_1 = 0.215$	0.207 (0.025)	−0.008	0.026	0.908
		$\lambda_2 = 0.183$	0.181 (0.010)	−0.002	0.009	0.924
		$\phi = 0.5$	1.018 (1.484)	0.518	1.619	0.932
EM	200	$\beta_{11} = 1$	1.086 (0.342)	0.086	0.412	0.944
		$\beta_{20} = -2$	−2.150 (0.593)	−0.150	0.752	0.956
		$\beta_{21} = 0.1$	0.150 (0.081)	0.050	0.249	0.948
		$\lambda_1 = 0.215$	0.211 (0.022)	−0.004	0.026	0.880
		$\lambda_2 = 0.183$	0.181 (0.006)	−0.002	0.010	0.784
		$\phi = 0.5$	1.039 (–)	0.539	1.473	–
	300	$\beta_{11} = 1$	1.032 (0.272)	0.032	0.325	0.920
		$\beta_{20} = -2$	−2.060 (0.439)	−0.060	0.409	0.976
		$\beta_{21} = 0.1$	0.120 (0.042)	0.020	0.073	0.944
		$\lambda_1 = 0.215$	0.209 (0.018)	−0.006	0.023	0.860
		$\lambda_2 = 0.183$	0.182 (0.005)	−0.001	0.008	0.824
		$\phi = 0.5$	0.866 (–)	0.366	1.214	–
SEM	200	$\beta_{11} = 1$	1.061 (0.471)	0.061	0.370	0.944
		$\beta_{20} = -2$	−2.132 (0.572)	−0.132	0.587	0.948
		$\beta_{21} = 0.1$	0.112 (0.069)	0.012	0.062	0.964
		$\lambda_1 = 0.215$	0.211 (0.031)	−0.004	0.023	0.964
		$\lambda_2 = 0.183$	0.182 (0.012)	−0.001	0.008	0.968
		$\phi = 0.5$	0.582 (1.581)	0.082	0.680	0.824
	300	$\beta_{11} = 1$	1.009 (0.351)	0.009	0.237	0.968
		$\beta_{20} = -2$	−2.053 (0.421)	−0.053	0.409	0.944
		$\beta_{21} = 0.1$	0.104 (0.049)	0.004	0.036	0.932
		$\lambda_1 = 0.215$	0.213 (0.024)	−0.002	0.021	0.928
		$\lambda_2 = 0.183$	0.184 (0.009)	0.001	0.006	0.960
		$\phi = 0.5$	0.431 (1.207)	−0.069	0.432	0.876

TABLE 2 Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP) under setting $\boldsymbol{\psi} = \boldsymbol{\psi}_2$.

Approach	n	Parameter	Estimate (SE)	Bias	RMSE	95% CP
DM	200	$\beta_{11} = 1$	1.106 (0.515)	0.106	0.500	0.972
		$\beta_{20} = -2$	−2.178 (0.728)	−0.178	0.843	0.992
		$\beta_{21} = 0.1$	0.160 (0.112)	0.060	0.196	0.964
		$\lambda_1 = 0.215$	0.209 (0.031)	−0.006	0.033	0.892
		$\lambda_2 = 0.183$	0.182 (0.012)	−0.001	0.010	0.916
		$\phi = 0.8$	1.338 (1.984)	0.538	2.094	0.912
	300	$\beta_{11} = 1$	1.113 (0.419)	0.113	0.472	0.932
		$\beta_{20} = -2$	−2.166 (0.536)	−0.166	0.662	0.972
		$\beta_{21} = 0.1$	0.155 (0.088)	0.055	0.201	0.960
		$\lambda_1 = 0.215$	0.207 (0.026)	−0.008	0.028	0.884
		$\lambda_2 = 0.183$	0.182 (0.010)	−0.001	0.010	0.876
		$\phi = 0.8$	1.309 (1.673)	0.509	1.966	0.912
EM	200	$\beta_{11} = 1$	1.052 (0.363)	0.052	0.467	0.884
		$\beta_{20} = -2$	−2.260 (0.637)	−0.260	0.929	0.936
		$\beta_{21} = 0.1$	0.156 (0.074)	0.056	0.219	0.912
		$\lambda_1 = 0.215$	0.208 (0.023)	−0.007	0.027	0.884
		$\lambda_2 = 0.183$	0.182 (0.007)	−0.001	0.010	0.816
		$\phi = 0.8$	1.408 (–)	0.608	2.061	–
	300	$\beta_{11} = 1$	1.041 (0.292)	0.041	0.371	0.904
		$\beta_{20} = -2$	−2.093 (0.491)	−0.093	0.691	0.948
		$\beta_{21} = 0.1$	0.129 (0.054)	0.029	0.175	0.900
		$\lambda_1 = 0.215$	0.212 (0.018)	−0.003	0.024	0.860
		$\lambda_2 = 0.183$	0.182 (0.005)	−0.001	0.008	0.816
		$\phi = 0.8$	1.241 (–)	0.441	1.494	–
SEM	200	$\beta_{11} = 1$	1.031 (0.473)	0.031	0.372	0.944
		$\beta_{20} = -2$	−2.120 (0.590)	−0.120	0.636	0.980
		$\beta_{21} = 0.1$	0.115 (0.070)	0.015	0.092	0.956
		$\lambda_1 = 0.215$	0.210 (0.031)	−0.005	0.025	0.936
		$\lambda_2 = 0.183$	0.184 (0.012)	0.001	0.008	0.956
		$\phi = 0.8$	0.779 (1.743)	−0.021	0.954	0.840
	300	$\beta_{11} = 1$	1.027 (0.386)	0.027	0.299	0.976
		$\beta_{20} = -2$	−2.092 (0.462)	−0.092	0.523	0.952
		$\beta_{21} = 0.1$	0.110 (0.058)	0.010	0.080	0.964
		$\lambda_1 = 0.215$	0.214 (0.026)	−0.001	0.021	0.968
		$\lambda_2 = 0.183$	0.184 (0.010)	0.001	0.007	0.956
		$\phi = 0.8$	0.773 (1.447)	−0.027	0.871	0.864

TABLE 3 Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP) under setting $\boldsymbol{\psi = \psi_3}$.

Approach	n	Parameter	Estimate (SE)	Bias	RMSE	95% CP
DM	200	$\beta_{11} = 1$	1.154 (0.507)	0.154	0.595	0.944
		$\beta_{20} = -2$	−2.220 (0.658)	−0.220	0.836	0.988
		$\beta_{21} = 0.1$	0.161 (0.103)	0.061	0.212	0.940
		$\lambda_1 = 0.316$	0.299 (0.043)	−0.017	0.050	0.860
		$\lambda_2 = 0.179$	0.177 (0.017)	−0.002	0.016	0.888
		$\phi = 0.5$	1.104 (1.854)	0.604	2.009	0.916
	300	$\beta_{11} = 1$	1.150 (0.397)	0.150	0.441	0.936
		$\beta_{20} = -2$	−2.133 (0.487)	−0.133	0.513	0.960
		$\beta_{21} = 0.1$	0.133 (0.070)	0.033	0.103	0.960
		$\lambda_1 = 0.316$	0.302 (0.036)	−0.014	0.037	0.916
		$\lambda_2 = 0.179$	0.177 (0.014)	−0.002	0.013	0.924
		$\phi = 0.5$	0.993 (1.459)	0.493	1.563	0.936
EM	200	$\beta_{11} = 1$	1.083 (0.342)	0.083	0.475	0.872
		$\beta_{20} = -2$	−2.126 (0.627)	−0.126	0.846	0.952
		$\beta_{21} = 0.1$	0.158 (0.076)	0.058	0.213	0.944
		$\lambda_1 = 0.316$	0.304 (0.032)	−0.012	0.042	0.848
		$\lambda_2 = 0.179$	0.176 (0.009)	−0.003	0.014	0.808
		$\phi = 0.5$	1.125 (–)	0.625	1.689	–
	300	$\beta_{11} = 1$	1.093 (0.279)	0.093	0.360	0.904
		$\beta_{20} = -2$	−2.116 (0.459)	−0.116	0.526	0.956
		$\beta_{21} = 0.1$	0.130 (0.047)	0.030	0.093	0.928
		$\lambda_1 = 0.316$	0.305 (0.026)	−0.011	0.036	0.832
		$\lambda_2 = 0.179$	0.176 (0.008)	−0.003	0.012	0.804
		$\phi = 0.5$	1.058 (–)	0.558	1.403	–
SEM	200	$\beta_{11} = 1$	1.094 (0.472)	0.094	0.395	0.980
		$\beta_{20} = -2$	−2.186 (0.603)	−0.186	0.651	0.960
		$\beta_{21} = 0.1$	0.132 (0.077)	0.032	0.100	0.960
		$\lambda_1 = 0.316$	0.301 (0.043)	−0.015	0.040	0.916
		$\lambda_2 = 0.179$	0.178 (0.016)	−0.001	0.013	0.936
		$\phi = 0.5$	0.878 (1.603)	0.378	1.195	0.816
	300	$\beta_{11} = 1$	1.061 (0.359)	0.061	0.279	0.956
		$\beta_{20} = -2$	−2.111 (0.439)	−0.111	0.512	0.924
		$\beta_{21} = 0.1$	0.117 (0.054)	0.017	0.069	0.936
		$\lambda_1 = 0.316$	0.307 (0.034)	−0.009	0.031	0.912
		$\lambda_2 = 0.179$	0.179 (0.013)	−0.000	0.009	0.944
		$\phi = 0.5$	0.688 (1.246)	0.188	0.871	0.832

TABLE 4 Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP) under setting $\boldsymbol{\psi} = \boldsymbol{\psi}_4$.

Approach	n	Parameter	Estimate (SE)	Bias	RMSE	95% CP
DM	200	$\beta_{11} = 1$	1.167 (0.525)	0.167	0.544	0.972
		$\beta_{20} = -2$	−2.351 (0.867)	−0.351	1.477	0.972
		$\beta_{21} = 0.1$	0.221 (0.154)	0.121	0.457	0.956
		$\lambda_1 = 0.316$	0.298 (0.044)	−0.018	0.048	0.872
		$\lambda_2 = 0.179$	0.177 (0.017)	−0.002	0.016	0.916
		$\phi = 0.8$	1.537 (1.954)	0.737	2.131	0.920
	300	$\beta_{11} = 1$	1.117 (0.426)	0.117	0.440	0.960
		$\beta_{20} = -2$	−2.160 (0.532)	−0.160	0.623	0.984
		$\beta_{21} = 0.1$	0.147 (0.084)	0.047	0.165	0.952
		$\lambda_1 = 0.316$	0.305 (0.039)	−0.011	0.039	0.912
		$\lambda_2 = 0.179$	0.176 (0.015)	−0.003	0.014	0.924
		$\phi = 0.8$	1.349 (1.795)	0.549	1.753	0.944
EM	200	$\beta_{11} = 1$	1.127 (0.368)	0.127	0.502	0.900
		$\beta_{20} = -2$	−2.222 (0.679)	−0.222	1.010	0.964
		$\beta_{21} = 0.1$	0.175 (0.098)	0.075	0.304	0.948
		$\lambda_1 = 0.316$	0.301 (0.032)	−0.015	0.048	0.796
		$\lambda_2 = 0.179$	0.176 (0.010)	−0.003	0.014	0.792
		$\phi = 0.8$	1.554 (–)	0.754	2.025	–
	300	$\beta_{11} = 1$	1.074 (0.293)	0.074	0.392	0.904
		$\beta_{20} = -2$	−2.091 (0.488)	−0.091	0.576	0.960
		$\beta_{21} = 0.1$	0.131 (0.053)	0.031	0.127	0.944
		$\lambda_1 = 0.316$	0.308 (0.027)	−0.008	0.035	0.864
		$\lambda_2 = 0.179$	0.177 (0.008)	−0.002	0.012	0.800
		$\phi = 0.8$	1.334 (–)	0.534	1.575	–
SEM	200	$\beta_{11} = 1$	1.092 (0.500)	0.092	0.379	0.980
		$\beta_{20} = -2$	−2.222 (0.681)	−0.221	0.720	0.965
		$\beta_{21} = 0.1$	0.165 (0.119)	0.065	0.269	0.950
		$\lambda_1 = 0.316$	0.305 (0.045)	−0.011	0.040	0.920
		$\lambda_2 = 0.179$	0.178 (0.017)	−0.001	0.014	0.910
		$\phi = 0.8$	1.242 (1.829)	0.442	1.544	0.825
	300	$\beta_{11} = 1$	1.090 (0.401)	0.090	0.363	0.976
		$\beta_{20} = -2$	−2.091 (0.504)	−0.091	0.566	0.976
		$\beta_{21} = 0.1$	0.114 (0.062)	0.014	0.075	0.976
		$\lambda_1 = 0.316$	0.311 (0.038)	−0.005	0.033	0.968
		$\lambda_2 = 0.179$	0.179 (0.015)	0.000	0.011	0.948
		$\phi = 0.8$	1.041 (1.561)	0.241	1.197	0.888

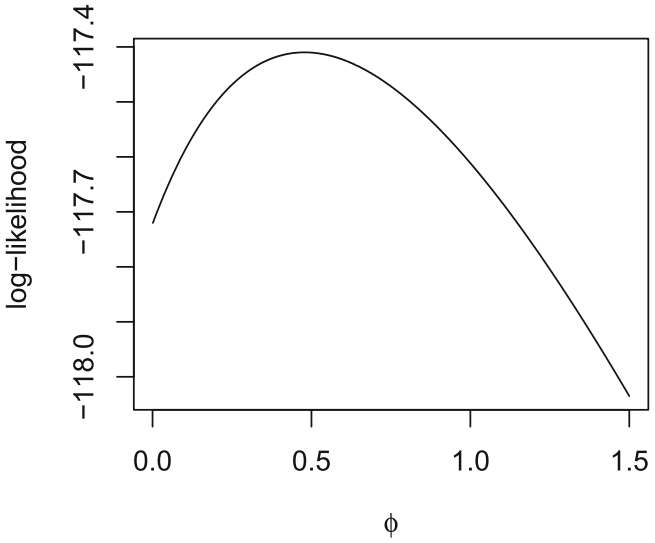

FIGURE 1 Profile likelihood plot for the parameter ϕ.

observed for ϕ. The standard errors and RMSEs decrease with an increase in sample size for all settings and estimation approaches, which is consistent with the large sample properties. Because of the advantages the SEM algorithm presents with respect to bias, standard error, and coverage probabilities, we consider the proposed SEM algorithm as the preferred algorithm.

5.3 | Estimation of cure rate

To illustrate the impact of accurate parameter estimation on the estimated cure rate, in Figure 2 we compare the estimated cure rates to the true cure rates for parameter setting $\boldsymbol{\psi}_1$ and $n = 200$. While estimated cure rates are relatively close to true cure rates for smaller values of tumor thickness, we see that both the DM- and EM-produced estimates, $\widehat{p}_0$(EM) and $\widehat{p}_0$(DM), respectively, lose accuracy for larger values of tumor thickness while the SEM-produced estimates ($\widehat{p}_0$(SEM)) remain closer to the true cure probabilities. In a clinical setting, tumor thickness is a prognostic factor of interest that may be expected to be associated with severity of illness. Significant underestimation of cure for patients with large tumor thickness may lead a practitioner to proceed with a more aggressive and potentially dangerous treatment than the patient's condition dictates. The values of $\widehat{p}_0$(EM) and $\widehat{p}_0$(DM) are consistently lower than $\widehat{p}_0$(SEM) for tumor thicknesses larger than 5 mm, with lowest accuracy observed in estimates for treatment group and larger tumor thickness values. The DM approach significantly underestimates cure, with relative error of $\widehat{p}_0$(DM) as much as 18.8% larger than the corresponding relative error of $\widehat{p}_0$(SEM). While differences in relative error of $\widehat{p}_0$(SEM) and $\widehat{p}_0$(EM) as large as 6.6%, as observed for a treatment patient with tumor thickness of 15 mm, may not constitute a drastic underestimation, neither can this underestimation be dismissed as negligible when considering potentially dangerous adjuvant therapy. Because the DM approach is seen to produce estimates with comparatively larger biases, SEs, and RMSEs than the SEM approach which consequently significantly underestimate cure probabilities, we focus the remaining analyses on comparison of the proposed EM and SEM algorithms.

5.4 | Comparison of maximized log-likelihood values

We further compare the performance of the EM and SEM algorithms by comparing the resulting maximized log-likelihood values after the same 250 datasets and initial values were used with both estimation methods. For each parameter setting, Table 5 presents the average maximized observed log-likelihood values ($\widehat{l}_o$) and the proportion of runs in which a given estimation method produced the greater maximized observed log-likelihood value (P_o), as well as the average maximized complete log-likelihood values ($\widehat{l}_c$) and the proportion of runs in which a given estimation method produced the greater maximized complete log-likelihood value (P_c).

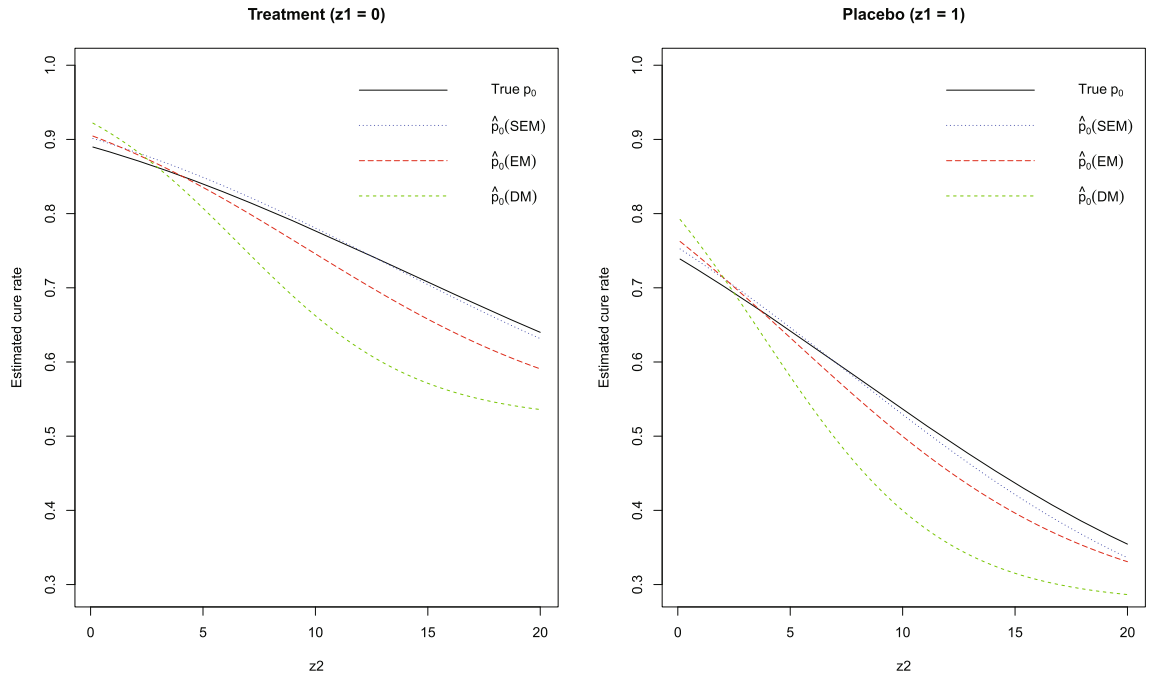

FIGURE 2 Plot of cure rate estimates against tumor thickness (z_2).

TABLE 5 Comparison of maximized log-likelihood values.

	EM		SEM		EM		SEM	
Parameter	$\widehat{l}_o$	P_o	$\widehat{l}_o$	P_o	$\widehat{l}_c$	P_c	$\widehat{l}_c$	P_c
$\psi = \psi_1$	−248.53	0.90	−248.69	0.10	−361.55	0.30	−336.73	0.70
$\psi = \psi_2$	−239.18	0.83	−239.41	0.17	−378.13	0.40	−345.07	0.60
$\psi = \psi_3$	−268.03	0.92	−268.15	0.08	−389.52	0.39	−389.15	0.61
$\psi = \psi_4$	−258.37	0.93	−258.51	0.07	−394.67	0.36	−390.42	0.64

While the EM algorithm both produced greater values of $\widehat{l}_o$ and P_o in all considered settings, the difference in $\widehat{l}_o$ produced by the two methods is relatively minor, being less than 0.3 in all cases. Although the complete log-likelihood function, $l_c(\cdot)$, would be a preferred measure of model fit, in practical applications $l_c(\cdot)$ cannot be computed due to the missing data in D_{comp}. Consequently, we discriminate between estimation methods in real data applications by comparing values of the observed log-likelihood function at the MLEs produced by each method. When simulating data, however, explicit values of the missing data (variables M and D) are generated and then used to generate the observed data. Hence, for this comparison we were able to retain the missing data values for the sole purpose of calculating $l_c(\cdot)$ using the complete data and MLEs. In all considered settings, the SEM algorithm produced greater values of $\widehat{l}_c$ and P_c, with more pronounced differences between $\widehat{l}_c$ produced by the two methods in several of the considered settings. Because $l_c(\cdot)$ is a function of the complete data, we conclude that the method which consistently maximizes this function, the SEM algorithm, is the preferred estimation method with regard to maximizing the log-likelihood function.

5.5 | Performance under model mis-specification

The proposed model utilizes a parametric framework for the lifetime distribution, where waiting time W is assumed to follow a Weibull distribution with density function given by (13). While the versatility afforded by the two-parameter Weibull distribution motivates its frequent usage to model lifetime data, it is of interest to consider the impact that mis-specifying

the lifetime distribution as Weibull may have on estimation. To this end, we assume the waiting time to follow a gamma distribution with shape parameter λ_1, scale parameter λ_2, and density function given by

$$f(t) = f(t; \lambda_1, \lambda_2) = \frac{1}{\lambda_2^{\lambda_1}\Gamma(\lambda_1)} t^{\lambda_1 - 1} e^{-t/\lambda_2}, \quad t > 0, \lambda_1 > 0, \lambda_2 > 0. \tag{14}$$

Considering the same variance and mean values as in Section 5.1 with the same specification for remaining parameters yields four parameter settings which we denote as $\widetilde{\boldsymbol{\psi}}_{\boldsymbol{1}} = (1, -2, 0.1, 16.667, 0.3, 0.5)$, $\widetilde{\boldsymbol{\psi}}_{\boldsymbol{2}} = (1, -2, 0.1, 16.667, 0.3, 0.8)$, $\widetilde{\boldsymbol{\psi}}_{\boldsymbol{3}} = (1, -2, 0.1, 8.333, 0.6, 0.5)$, and $\widetilde{\boldsymbol{\psi}}_{\boldsymbol{4}} = (1, -2, 0.1, 8.333, 0.6, 0.8)$. For each parameter setting, and for sample sizes $n = 200$ and $n = 300$, we follow data generation steps as in Section 5.1, using a gamma distribution with parameters (λ_1, λ_2) in Step 3. Both proposed algorithms (with Weibull lifetimes) are applied to $M = 250$ simulated data sets, with initial values generated as described in Section 5.2. Table A3.1 of the Supplemental Material reports the estimated survival probabilities, biases, and RMSEs at Q_1, Q_2, and Q_3, the first, second, and third quartile of the corresponding gamma distribution, respectively.

The proposed algorithm performs reasonably well under the described mis-specification, with slight overestimation of survival probabilities observed to decrease with increasing sample size for all considered parameter settings. The RMSEs also decrease with an increase in sample size for all settings and both estimation approaches, which is consistent with the large sample properties. The larger biases and RMSEs of EM-produced survival estimates support the continued preference of the SEM algorithm in the case of model mis-specification.

6 | ILLUSTRATIVE EXAMPLE

In this section, we demonstrate the performance of the proposed SEM algorithm using a real data on children's mortality obtained from the 2018 Nigeria Demographic and Health Survey (DHS), which examines the health of women of reproductive age (13-49 years) and their children. Because under-5 mortality rate is a leading indicator of the level of child health and socioeconomic development of a country, the analysis of child mortality data can inform both assessments of quality of life and the development of strategies to reduce risk of death.

The data set considered consists of 13 255 children under age 5 whose survival information is available along with covariate information. While the survival times of the children, reported by the mothers in retrospective interviews, are recorded as age at death in unit of days (for age less than 30 days), months (for age greater than 30 days but less than 24 months), or years (for age greater than 2 years), these stated ages are implicitly interval-censored due to age being recorded in integer values. The data contains numerous covariates of potential interest, however for this application we consider the mother's BMI and whether the child was delivered in the hospital (Hospital Delivery, 1 indicates child was delivered in a hospital and 0 otherwise) as covariates that are linked to model parameters η and p. A full description of the survey methodology can be obtained from the DHS program.[46] Denoting Hospital Delivery by z_1 and BMI by z_2, we consider five potential regression models:

- Model 1: $\eta = e^{\beta_{10}+\beta_{11}z_1}$, $p = \frac{\exp(\beta_{21}z_2)}{1+\exp(\beta_{21}z_2)}$, where p contains no intercept term.
- Model 2: $\eta = e^{\beta_{11}z_1}$, $p = \frac{\exp(\beta_{20}+\beta_{21}z_2)}{1+\exp(\beta_{20}+\beta_{21}z_2)}$, where η contains no intercept term.
- Model 3: $\eta = e^{\beta_{20}+\beta_{21}z_2}$, $p = \frac{\exp(\beta_{11}z_1)}{1+\exp(\beta_{11}z_1)}$, where p contains no intercept term.
- Model 4: $\eta = e^{\beta_{21}z_2}$, $p = \frac{\exp(\beta_{10}+\beta_{11}z_1)}{1+\exp(\beta_{10}+\beta_{11}z_1)}$, where η contains no intercept term.
- Model 5: $\eta = e^{\beta_0+\beta_1z_1+\beta_2z_2}$, $p = 1$.

We note that the first four models accommodate a destructive element, while the fifth model takes the activation probability of initial risks as $p = 1$, indicating that no destruction of initial risks may take place. In the context of infant mortality, it is hoped by a practitioner that increased access to healthcare and adequate nutrition for the mother will reduce a child's risk of death. Model 5 is included in this comparison to assess whether the inclusion of a destructive element facilitates greater maximization of the observed log-likelihood value, which may indicate that a preventive measure reduced the risk of mortality to the child. In order to identify the model best fitting the data, we compare AIC and BIC for

TABLE 6 Model discrimination.

	EM			SEM		
Model	**Obs log-lik**	**AIC**	**BIC**	**Obs log-lik**	**AIC**	**BIC**
1	−7740.3	15 490.6	15 528.0	−7735.0	15 479.9	15 517.4
2	−7735.0	15 479.9	15 517.4	−7734.8	15 479.6	15 517.1
3	−7766.7	15 543.3	15 580.8	−7741.5	15 493.1	15 530.5
4	−7789.1	15 588.1	15 625.6	−7741.9	15 493.8	15 531.3
5	−7985.0	15 979.9	16 017.4	−7802.9	15 615.8	15 653.3

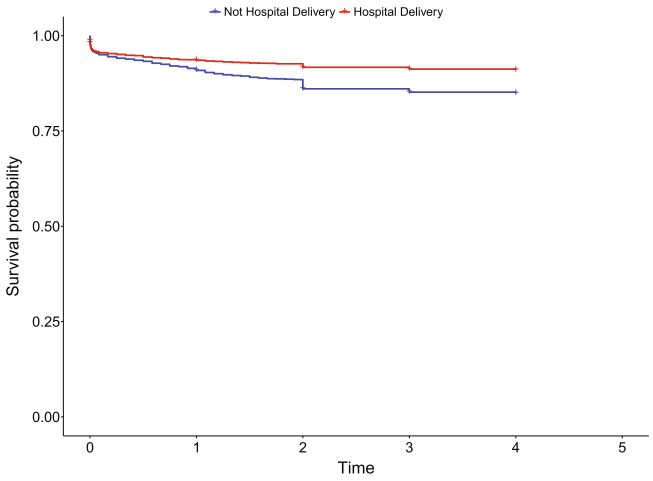

FIGURE 3 Kaplan-Meier plot of survival curves stratified by Hospital Delivery.

the above models by applying the proposed SEM algorithm and computing the observed log-likelihood function evaluated at MLEs. AIC and BIC values are reported in Table 6. For comparison purposes, the EM algorithm is applied to the above models and values of observed log-likelihood function, AIC, and BIC through using the EM algorithm are also reported in Table 6. We observe first that for each model, the SEM-produced estimates yield larger observed log-likelihood function values, and consequently smaller AIC and BIC values, than the EM-produced estimates. Further, the SEM-produced estimates for Model 2, which accommodates a destructive mechanism, produce the lowest AIC and BIC values, indicating that the effect of intervention may be captured through the destructive process. Consequently, we choose Model 2 as our working model. Kaplan-Meier curves stratified by Hospital Delivery, as shown in Figure 3, level off to non-zero proportions which supports the presence of a cure component in the data, and the shapes of the curves convey a similar relationship between Hospital Delivery and long term survival.

Table 7 presents the estimates and standard errors of the parameters of the working model using both the EM and SEM algorithms, as well as p-values of regression coefficients. To obtain the estimates using the SEM algorithm, we first conduct a preliminary study and decide to use 2000 iterations and consider the first 1000 iterations as burn-in. The standard errors of the estimates are obtained by inverting the observed information matrix. We note that while the sign of each regression coefficient estimate is the same regardless of approach, indicating the direction of covariate effect is consistent across approaches, the standard errors of SEM-produced estimates are smaller. Noting the consistency of sign across methods and the superior precision of SEM-produced estimates, we identify the SEM algorithm as the preferred estimation method and proceed with analysis of the results using only the SEM-produced estimates. The negative sign of the estimate for β_{11} is highly statistically significant, agreeing with both the stratified Kaplan-Meier plot in Figure 3 and previous findings observing a reduced risk of child mortality associated with hospital delivery.[47] While the previous analysis by Li et al,[47] which considered the 2003 Nigeria DHS child mortality data, failed to establish statistical significance of BMI, the estimate of $\beta_{21} < 0$ indicates that a mother's BMI is positively associated with a child's survival and is statistically significant at a 10% level of significance.

TABLE 7 Comparison of estimation results for the children's mortality data.

	EM			SEM		
Parameter	**Estimate**	**Standard error**	***p*−value**	**Estimate**	**Standard error**	***p*−value**
β_{11}	−0.432	0.064	< 0.001	−0.440	0.063	< 0.001
β_{20}	4.665	1.908	0.015	3.626	1.264	0.004
β_{21}	−0.095	0.049	0.054	−0.070	0.036	0.054
λ_1	3.382	0.244	–	3.375	0.163	–
λ_2	0.001	0.001	–	0.001	0.001	–
ϕ	0.010	–	–	0.039	1.546	–

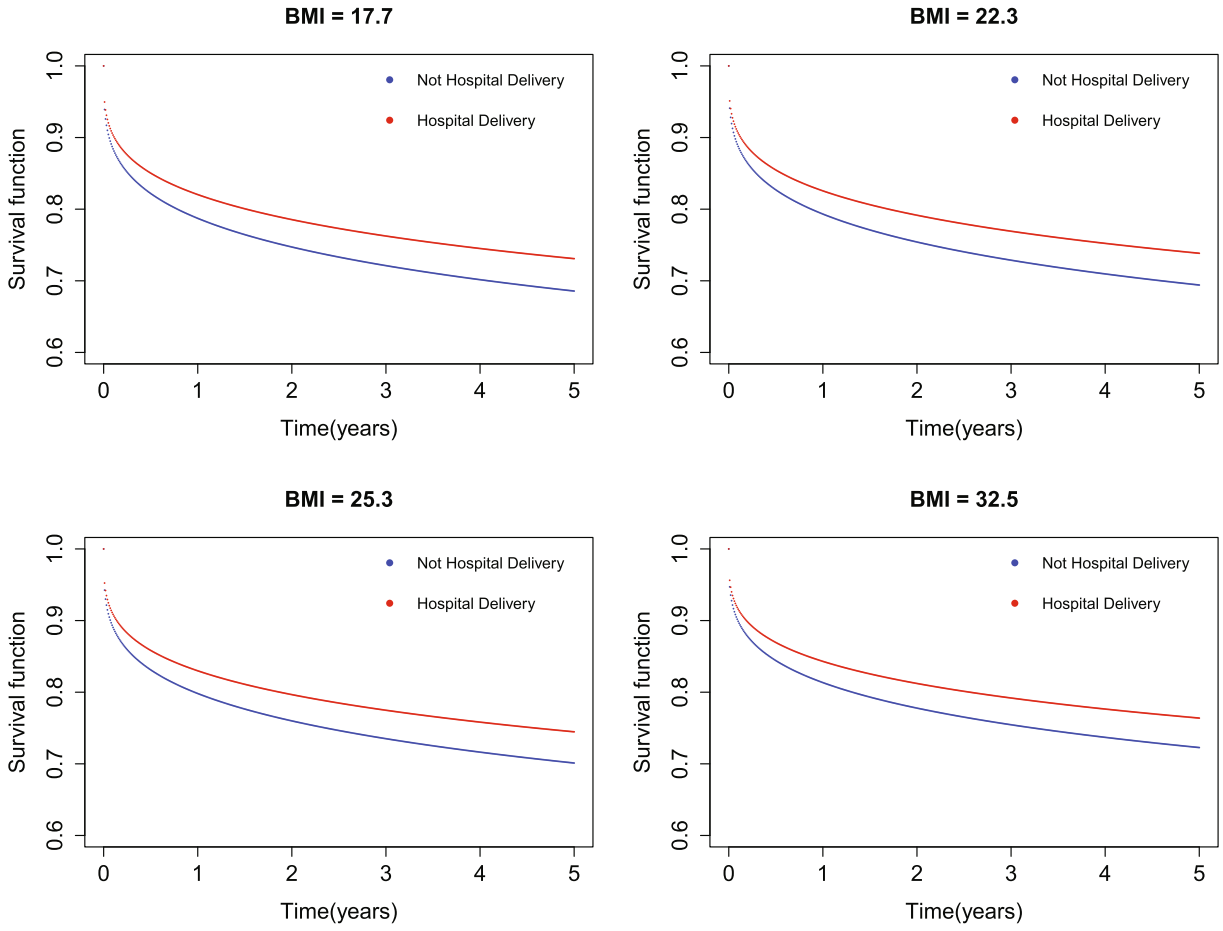

FIGURE 4 Predicted survival probabilities stratified by Hospital Delivery for children with different mother's BMI values.

Figure 4 shows the predicted survival probabilities for children whose mother's BMI is 17.7, 22.3, 25.3, and 32.5, which correspond to the 5th, 50th, 75th, and 95th percentiles, stratified by Hospital Delivery. Note that the survival probability for hospital-delivered children is higher across all values of BMI. It may be observed that the survival probability increases with higher BMI values by comparing the plots fixing BMI at 17.7 and 32.5. The effect of mother's BMI on survival is further conveyed in Figure 5 where the estimated cure rate is shown to increase in a nearly linear fashion as mother's BMI increases. Figure A4.1 of the Supplemental Material shows the evolution paths of parameter estimates in the SEM algorithm. After the burn-in period of 1000 iterations, the estimates are observed to oscillate without a discernible upward or downward trend.

We check the adequacy of the destructive negative binomial model by using the calculated normalized randomized quantile residuals[48]. Figure A4.2 of the Supplemental Material presents the quantile-quantile plot, where each point

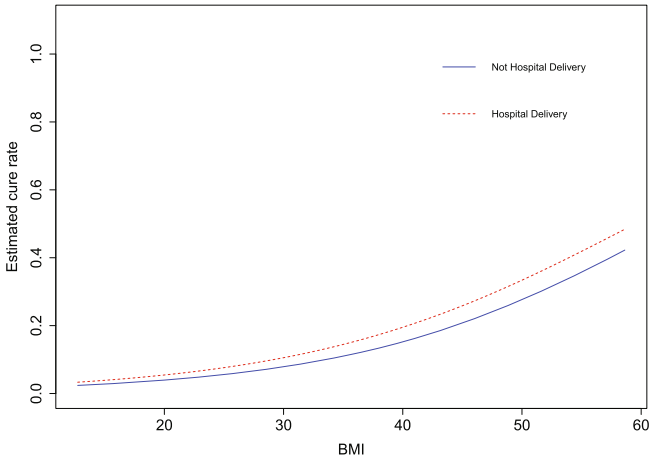

FIGURE 5 Cure rate against mother's BMI stratified by Hospital Delivery.

corresponds to the median of five sets of ordered residuals. The linearity in this plot suggests that the destructive negative binomial cure rate model with Weibull lifetimes provides a good fit to the children's mortality data. Finally, the Kolmogorov-Smirnov test for normality of residuals provides strong evidence for the normality of residuals, with a p-value of 0.996.

For interested readers, another application of the model and the estimation method to a data on smoking cessation is provided in Section A5 of the Supplemental Material.

7 | CONCLUSION AND FUTURE WORK

In this paper, we generalize the destructive negative binomial cure rate model to accommodate interval censoring. Motivated by the work of Gallardo et al,[23] we develop likelihood inference based on an implementation of the EM algorithm which splits the complete log-likelihood function into simpler functions to be maximized independently, as well as a novel implementation of the SEM algorithm. Through a Monte Carlo simulation study, we demonstrate that the proposed SEM method performs favorably when compared to the EM algorithm with respect to parameter recovery. In estimating parameters that are related to the cure probability (ie, $\phi, \boldsymbol{\beta_1}, \boldsymbol{\beta_2}$), the SEM algorithm provides greater accuracy and precision, yielding estimates with smaller biases and smaller RMSEs than the EM approach. This in turn results in more accurate estimates of the cured probabilities. While standard errors for other parameter estimates are comparable between the SEM and EM approaches, the coverage probabilities of the asymptotic confidence intervals produced with the SEM are consistently closer to the nominal level. For the EM and SEM estimation methods, both the standard error and RMSE of the estimators of the model parameters decrease with an increase in sample size. In the real data analysis, we first show through model discrimination that both the inclusion of a destructive process and the use of the SEM algorithm facilitate the greatest maximization of the observed log-likelihood function. Further, the proposed SEM algorithm produces parameter estimates that are consistent both with prior findings and with estimates produced by the EM algorithm.

While the objective of this study was to develop an efficient estimation algorithm for the destructive negative binomial cure rate model under interval censoring and perform a detailed analysis of its performance, a valuable future work will be to generalize this application to families of cure rate models; see Koutras and Milienos[49] and Milienos.[50] Considering a generalized family of lifetime distributions, for instance, the generalized gamma distribution, would allow us to perform model discrimination for the lifetime distribution under a fully parametric setup; see Wang and Pal.[51] Exploring the performance of the proposed SEM algorithm under semiparametric or nonparametric modeling for the lifetimes of susceptibles will provide greater insight as to the versatility of this estimation method. We hope to report on these problems in future manuscripts.

ACKNOWLEDGEMENTS

Our sincere thanks go to the Associate Editor and two anonymous reviewers for their constructive feedback on an earlier version of this article. The research reported in this publication was supported by the National Institute Of General

Medical Sciences of the National Institutes of Health under Award Number R15GM150091. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

DATA AVAILABILITY STATEMENT

The R codes for the data generation and the EM and SEM algorithms are available in Section A1 of the Supplemental Material.

ORCID


Suvra Pal https://orcid.org/0000-0001-9864-9489

REFERENCES

1. Pal S, Yu H, Loucks ZD, Harris IM. Illustration of the flexibility of generalized gamma distribution in modeling right censored survival data: analysis of two cancer datasets. *Ann Data Sci*. 2020;7(1):77-90.
2. Overman T, Pal S. Statistical tools and techniques in modeling survival data. In: Goldwyn E, Wootton A, Ganzell S, eds. *Mathematics Research for the Beginning Student Volume 2: Accessible Projects for Students After Calculus*. Switzerland: Birkhauser-Springer; 2022:75-99.
3. Hoang LQ, Pal S, Liu Z, Senkowsky J, Tang L. A time-dependent survival analysis for early prognosis of chronic wounds by monitoring wound alkalinity. *Int Wound J*. 2023;20(5):1459-1475.
4. Boag JW. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J R Stat Soc Ser B Stat Methodol*. 1949;11(1):15-53.
5. Aselisewine W, Pal S. On the integration of decision trees with mixture cure model. *Stat Med*. 2023. doi:10.1002/sim.9850
6. Maller RA, Zhou X. *Survival Analysis with Long-term Survivors*. New York: Wiley; 1996.
7. Peng Y, Yu B. *Cure Models: Methods, Applications, and Implementation*. Boca Raton, FL: CRC Press; 2021.
8. Chen MH, Ibrahim JG, Sinha D. A new Bayesian model for survival data with a surviving fraction. *J Am Stat Assoc*. 1999;94(447):909-919.
9. Rodrigues J, Castro M, Cancho VG, Balakrishnan N. COM–Poisson cure rate survival models and an application to a cutaneous melanoma data. *J Stat Plan Inference*. 2009;139(10):3605-3611.
10. Rodrigues J, Castro M, Balakrishnan N, Cancho VG. Destructive weighted Poisson cure rate models. *Lifetime Data Anal*. 2011;17:333-346.
11. Borges P, Rodrigues J, Balakrishnan N. Correlated destructive generalized power series cure rate models and associated inference with an application to a cutaneous melanoma data. *Comput Stat Data Anal*. 2012;56(6):1703-1713.
12. Balakrishnan N, Pal S. EM algorithm-based likelihood estimation for some cure rate models. *J Stat Theory Pract*. 2012;6:698-724.
13. Balakrishnan N, Pal S. Lognormal lifetimes and likelihood-based inference for flexible cure rate models based on COM-Poisson family. *Comput Stat Data Anal*. 2013;67:41-67.
14. Balakrishnan N, Pal S. An EM algorithm for the estimation of parameters of a flexible cure rate model with generalized gamma lifetime and model discrimination using likelihood- and information-based methods. *Comput Stat*. 2015;30(1):151-189.
15. Balakrishnan N, Pal S. Likelihood inference for flexible cure rate models with gamma lifetimes. *Commun Stat Theory Methods*. 2015;44(19):4007-4048.
16. Balakrishnan N, Pal S. Expectation maximization-based likelihood inference for flexible cure rate models with Weibull lifetimes. *Stat Methods Med Res*. 2016;25(4):1535-1563.
17. Pal S, Balakrishnan N. Destructive negative binomial cure rate model and EM-based likelihood inference under Weibull lifetime. *Stat Probab Lett*. 2016;116:9-20.
18. Pal S, Balakrishnan N. Likelihood inference for COM-Poisson cure rate model with interval-censored data and Weibull lifetimes. *Stat Methods Med Res*. 2017;26(5):2093-2113.
19. Pal S, Balakrishnan N. Expectation maximization algorithm for Box–Cox transformation cure rate model and assessment of model misspecification under Weibull lifetimes. *IEEE J Biomed Health Inform*. 2017;22(3):926-934.
20. Pal S, Balakrishnan N. Likelihood inference based on EM algorithm for the destructive length-biased Poisson cure rate model with Weibull lifetime. *Commun Stat Simul Comput*. 2018;47(3):644-660.
21. Pal S, Majakwara J, Balakrishnan N. An EM algorithm for the destructive COM-Poisson regression cure rate model. *Metrika*. 2018;81:143-171.
22. Majakwara J, Pal S. On some inferential issues for the destructive COM-Poisson-generalized gamma regression cure rate model. *Commun Stat Simul Comput*. 2019;48(10):3118-3142.
23. Gallardo DI, Bolfarine H, Lima AC. An EM algorithm for estimating the destructive weighted Poisson cure rate model. *J Stat Comput Simul*. 2016;86(8):1497-1515.
24. Pal S, Peng Y, Aselisewine W. A new approach to modeling the cure rate in the presence of interval censored data. *Comput Stat*. 2023. doi:10.1007/s00180-023-01389-7
25. Banerjee S, Carlin BP. Parametric spatial cure rate models for interval-censored time-to-relapse data. *Biometrics*. 2004;60(1):268-275.
26. Kim YJ, Jhun M. Cure rate model with interval censored data. *Stat Med*. 2008;27(1):3-14.
27. Wiangnak P, Pal S. Gamma lifetimes and associated inference for interval-censored cure rate model with COM–Poisson competing cause. *Commun Stat Theory Methods*. 2018;47(6):1491-1509.

28. Treszoks J, Pal S. A destructive shifted Poisson cure model for interval censored data and an efficient estimation algorithm. *Commun Stat Simul Comput*. 2022;1-15. doi:10.1080/03610918.2022.2067876
29. Pal S. A simplified stochastic EM algorithm for cure rate model with negative binomial competing risks: An application to breast cancer data. *Stat Med*. 2021;40(28):6387-6409.
30. Li CS, Taylor JM, Sy JP. Identifiability of cure models. *Stat Probab Lett*. 2001;54(4):389-395.
31. Tsodikov AD, Yakovlev AY, Asselain B. *Stochastic Models of Tumor Latency and their Biostatistical Applications*. Singapore: World Scientific; 1996.
32. McLachlan GJ, Krishnan T. *The EM Algorithm and Extensions*. Hoboken, NJ: John Wiley & Sons; 2007.
33. Pal S, Roy S. 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*. 2022;51(11):6866-6880.
34. Pal S, Roy S. On the estimation of destructive cure rate model: A new study with exponentially weighted Poisson competing risks. *Statistica Neerlandica*. 2021;75(3):324-342.
35. Pal S, Roy S. On the parameter estimation of Box-Cox transformation cure model. *Stat Med*. 2023;42(15):2600-2618. doi:10.1002/sim.9739
36. Wu CJ. On the convergence properties of the EM algorithm. *Ann Stat*. 1983;11(1):95–103.
37. Celeux G, Diebolt J. The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Comput Stat*. 1985;2:73-82.
38. Celeux G, Chauveau D, Diebolt J. Stochastic versions of the EM algorithm: An experimental study in the mixture case. *J Stat Comput Simul*. 1996;55(4):287-314.
39. Wei GC, Tanner MA. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *J Am Stat Assoc*. 1990;85(411):699-704.
40. Pal S, Barui S, Davies K, Mishra N. A stochastic version of the EM algorithm for mixture cure model with exponentiated weibull family of lifetimes. *J Stat Theory Pract*. 2022;16(3):48.
41. Diebolt J, Ip EH. A stochastic EM algorithm for approximating the maximum likelihood estimate. https://www.osti.gov/biblio/49148 1995.
42. Diebolt J, Celeux G. Asymptotic properties of a stochastic EM algorithm for estimating mixing proportions. *Stoch Model*. 1993;9(4):599-613.
43. Nielsen SF. The stochastic EM algorithm: estimation and asymptotic results. *Bernoulli*. 2000;6(3):457-489.
44. Marschner IC. On stochastic versions of the algorithm. *Biometrika*. 2001;88(1):281-286.
45. Ye Z, Ng HKT. On analysis of incomplete field failure data. *Ann Appl Stat*. 2014;8:1713-1727.
46. Demographic N. *Health Survey 2019. National Population Commission (NPC)*. Maryland, USA: NPC and ICF International; 2019.
47. Li S, Hu T, Zhao X, Sun J. A class of semiparametric transformation cure models for interval-censored failure time data. *Comput Stat Data Anal*. 2019;133:153-165.
48. Dunn PK, Smyth GK. Randomized quantile residuals. *J Comput Graph Stat*. 1996;5(3):236-244.
49. Koutras M, Milienos F. A flexible family of transformation cure rate models. *Stat Med*. 2017;36(16):2559-2575.
50. Milienos FS. On a reparameterization of a flexible family of cure models. *Stat Med*. 2022;41(21):4091-4111.
51. Wang P, Pal S. A two-way flexible generalized gamma transformation cure rate model. *Stat Med*. 2022;41(13):2427-2447.

SUPPORTING INFORMATION

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

How to cite this article: Treszoks J, Pal S. On the estimation of interval censored destructive negative binomial cure model. *Statistics in Medicine*. 2023;42(28):5113-5134. doi: 10.1002/sim.9904