Likelihood Inference for Flexible Cure Models with Interval Censored Data

by

JODI TRESZOKS

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2023

ACKNOWLEDGEMENTS

I would first like to express my gratitude to my doctoral supervisor, Dr. Suvra Pal, for sharing his expertise in survival analysis and his academic enthusiasm. Without his guidance, this work would not have been possible. I hope to emulate the care Dr. Pal shows for his students, and the passion and integrity he has for his work, in my future endeavors. I would like to thank the members of my dissertation committee, Dr. Andrzej Korzeniowski, Dr. Souvik Roy, Dr. Shan Sun-Mitchell, Dr. Xinlei Wang, and Dr. Dengdeng Yu, whose comments during the development of this thesis were invaluable. I also extend my thanks to Dr. Su, Lona Donnelly and the departmental staff, as well as the many professors I studied under in my time with the Department of Mathematics. Thank you to Professor Joe Heine, whose Pre-Calculus class gave me the confidence I needed to start down this path. I would like to thank many of my fellow graduate students for their kindness and friendship, with special thanks to Dr. Janessa Beach and Dr. Talon Johnson for the encouragement I've received from their faith and friendship.

Thank you to my friends and family who have supported and believed in me along the way. Finally, I wish to thank my husband Jon, for his love, encouragement, and perspective.

May 2023

ABSTRACT

Likelihood Inference for Flexible Cure Models with Interval Censored Data

Jodi Treszoks, Ph.D.

The University of Texas at Arlington, 2023

Supervising Professor: Dr. Suvra Pal

Models for survival data with a surviving fraction, known as cure rate models, play a vital role in survival analysis. Due to the improvement of intervening methodologies, some subjects are seen to be immune permanently. While cure rate models have been studied extensively in the recent literature with a standard assumption of right-censored data, in many applied settings, such as recidivism studies or medical studies where the event of interest is not immediately harmful, continuous observation of a subject is impracticable. We call lifetime data generated with discrete follow-up times as interval-censored.

In this thesis, we extend several existing cure models to accommodate interval censoring and develop efficient likelihood-based inference methods. We first extend two destructive cure rate models, in which a certain number of competing risks are reduced by a destructive mechanism, to the interval-censored setting. For the destructive shifted Poisson model under interval censoring, we propose an efficient expectation maximization (EM) algorithm by using the conditional distributions of the missing data to decompose the conditional expected complete log-likelihood function into simpler functions which are maximized independently and evaluate

the performance of this algorithm through Monte Carlo simulation and analysis of recidivism data. Likelihood inference for the interval-censored destructive negative binomial cure rate model is developed using the conditional distributions of the missing data in two distinct implementations of the EM algorithm and a variation of the EM algorithm called the stochastic EM (SEM) algorithm. We address the advantages of the recommended estimation method, the SEM algorithm, through simulation study and analysis of smoking cessation data. Finally, we use a semi-parametric framework to extend the Box-Cox transformation cure rate model to the interval-censored setting, formulate the EM algorithm, and present a comprehensive simulation study addressing its performance and an analysis of smoking cessation data.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Cure rate models

While conventional survival analysis methods assume all individuals to be susceptible to the event of interest given sufficient follow-up time [68, 83], 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 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 rate estimation is of particular interest to practitioners and is utilized in context of biomedical and financial applications, among others [1, 2, 3, 18, 46, 51, 98, 104]. Cure models are instrumental in reliability and biomedical science and are commonly used in cancer clinical trials.

1.2 Mixture cure rate model

The first cure rate model, known in the literature as the mixture cure rate model, was developed by Boag [15], where the author proposed a cure component to represent the proportion of patients not susceptible to recurrence of disease. In the mixture cure rate model, the population survival function of a time-to-event variable Y is given by

$$S_{\text{pop}}(y) = p_0 + (1 - p_0)S_{\text{susc}}(y), \tag{1.1}$$

where p_0 is the cure rate and $S_{\text{susc}}(\cdot)$ is a proper survival function of the susceptible patients only. For a book-length account on mixture cure rate model, interested readers may refer to the research monographs by Maller and Zhou [58] and Peng and Yu [85]. The mixture cure rate model is widely used, with parametric [20, 35, 36], semi-parametric [16, 34, 33, 46, 49, 66, 53, 67, 65, 92, 110], and non-parametric [21, 55, 99] inferential methods having been explored extensively.

1.3 Promotion time cure rate model

While the model in (1.1) is intuitive, it lacks a proportional hazards structure which is a desirable property to facilitate covariate analyses. The proportional hazards structure has been adopted in estimating cure rates using Bayesian [43, 44, 98, 111] as well as parametric [104] and semi-parametric [56, 54] approaches. The promotion time cure model is an alternative cure rate model which includes a proportional hazards structure, and is strongly motivated by biological considerations [23, 105]. Chen et al. proposed a Bayesian approach for a new cure rate model, known in the literature as the promotion time or Poisson cure rate model, where the event of interest (e.g., relapse of cancer) is related to a number of carcinogenic cells (competing risks) left active after an initial treatment [23]. Assuming the latent number of competing risks to follow a Poisson distribution, the population survival function for the promotion time cure rate model is given by

$$S_{\text{pop}}(y) = e^{-\eta(1-S(y))}, \tag{1.2}$$

with η representing the mean number of competing risks and $S(\cdot)$ representing the common survival function corresponding to the promotion time of each competing risk. Within the context of the promotion time cure rate model, varying assumptions about the parametric distribution of the competing risks variable appear in the literature

[13, 26, 38, 41, 87, 40], among others. Non-parametric [22, 103] and semi-parametric inference methods [14, 28, 45, 63] have also been developed using the promotion time cure model. A piecewise linear approximation to model the hazard functions of competing risks in the context of mixture and promotion time cure rate models was developed [5]. The parametric negative binomial cure model was extended by utilizing a piecewise exponential approximation for the progression time of each competing risk [42]. Novel estimation algorithms for cure models have recently been proposed and the advantages demonstrated with simulation and real data analyses [69, 80, 81]. Rodrigues et al. [90] proposed the Conway Maxwell Poisson (COM-Poisson) cure rate model which can account for both over-dispersion and under-dispersion and includes well-known cure rate models, such as (1.1) and (1.2), as special cases. Balakrishnan and Pal [6, 7, 8, 10] developed the steps of the expectation maximization (EM) algorithm for the determination of the maximum likelihood estimates (MLEs) for parameters of the COM-Poisson model based on non-informative right censored data and different parametric assumptions on the susceptible lifetime distribution. Alternative unified approaches considering both the mixture and promotion time cure rate models can be found in [24, 88, 107].

1.4 Destructive cure rate model

The destructive cure rate model, developed by Rodrigues et al., introduces to the COM-Poisson cure rate model a natural destructive process which imposes on the original number of competing risks [89]. Borges et al. extended the destructive cure rate model by incorporating a biological dependence between the initiated cells [17], and Gallardo et al. included the effect of bivariate random variable (U, V), where U is related to relapse times of the disease caused by non-destroyed cells and V is associated with the cure rate corresponding to a particular clinic under

study [37]. While [89] utilized direct maximization of the (observed) log-likelihood function for maximum likelihood estimation of the parameters of the destructive cure rate model, alternative estimation methods have been explored to address problems with convergence of the classical method depending on choice of initial parameter values [86]. 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 [6, 7, 8, 9, 10, 71, 72, 73, 74, 75, 76, 78, 57]. A novel variation of the EM algorithm for the destructive weighted Poisson cure rate model was developed by Gallardo et al. [37]. 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.

1.5 Transformation cure rate models

Many generalizations of the mixture and promotion time models utilizing transformations can be found in the literature [24, 48, 61, 96, 97, 100, 109]. By applying a reparameterization or a transformation to the population survival function, a resulting model may be seen to generalize popular models such as the mixture and promotion time models, as well as achieve greater realistic interpretability.

1.5.1 Box-Cox transformation cure rate model

Yin and Ibrahim proposed a class of cure rate models based on the Box-Cox transformation (BCT) which contains both the mixture and promotion time cure rate models as special cases [107]. A novel biological interpretation for the BCT model was proposed, along with maximum likelihood parameter estimation using a proportional hazards structure [84]. Non-parametric maximum likelihood estimation was developed for the BCT model with an added frailty term for multivariate survival

data [29]. Viewing the latent cured status as missing data, Pal and Balakrishnan developed an EM algorithm for the Box-Cox transformation cure rate model, with the assumption that the lifetimes follow a Weibull distribution [76]. Recently, Pal and Roy proposed an algorithm utilizing a non-linear conjugate gradient method for estimation of the BCT model parameters which was shown to produce more accurate and precise inference on the cure rate as compared to the EM algorithm [82].

1.6 Interval censoring

Though traditional survival models, as well as the cure rate models discussed previously, assume the data to be right censored, it is not uncommon in practice for data to be interval-censored. In a scenario where subjects are not monitored continuously but rather observed at discrete follow-up times, an observed lifetime may be said only to lie in an interval. The mixture cure rate model was extended to accommodate interval-censored data while considering spatial correlation, and Markov Chain Monte Carlo methods were used to develop Bayesian inference [11]. Covariate effects were introduced to the mixture cure rate model under interval censoring and the EM algorithm was developed for this framework [47]. The effects of covariates in both the mixture and promotion time cure rate models have been studied in the context of interval-censored data [50, 79, 93, 102, 108]. Pal and Balakrishnan [73] developed the steps of the EM algorithm for the determination of the MLEs of the COM-Poisson cure rate model parameters with interval censoring; see also Wiangnak and Pal [101].

1.6.1 Form of interval-censored data

Let T_i denote the i-th patient's true failure time (unobserved), for $i = 1, 2, \ldots, n,$, with n denoting the number of patients in the study. To develop an interval censoring

scheme, let us assume the i-th 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 a practical setting, 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 i-th subject's failure time did not occur prior to the last observation time, we consider the lifetime right-censored and can only conclude the lifetime is contained in the interval $(Y_{j(i)}, \infty)$. The observed data is denoted as $\boldsymbol{D}_{obs} = (\boldsymbol{l}, \boldsymbol{r}, \boldsymbol{\delta})$, 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

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

While the focus of this thesis is interval-censored data, it is important to note that both right- and left-censored data can be viewed as special cases of the general interval-censored form. This property supports the utility of developing inferential methods that accommodate interval-censored data, as these methods will be more broadly applicable to varying forms of data than the majority of existing methods which rely on an assumption of right censoring.

1.7 Thesis structure

This thesis aims to extend several existing cure models to accommodate interval censoring and develop corresponding likelihood-based inference methods. The rest of this thesis is organized as follows. In Chapter 2, we apply the destructive shifted Poisson cure rate model to interval-censored data, propose an efficient EM algorithm for this model, and examine the performance of the proposed estimation method through a simulation study and real data analysis. Chapter 3 presents the destructive negative binomial cure rate model under interval censoring and demonstrates the performance of the recommended estimation method, a variation of the EM algorithm called the stochastic EM (SEM) algorithm, as compared to two distinct implementations of the EM algorithm. The SEM algorithm is shown to be preferred both through simulation and analysis of a real data on smoking cessation. In Chapter 4, we use a semi-parametric framework to extend the Box-Cox transformation cure rate model to the interval-censored setting and formulate the EM algorithm. The performance of the proposed EM algorithm at parameter recovery is considered using both simultaneous maximization and profile likelihood techniques and compared to the performance of direct maximization of the log-likelihood function through a comprehensive simulation study, and the EM algorithm is applied to a real data on smoking cessation. Finally, in Chapter 5 we summarize and suggest areas for future related research.

CHAPTER 2

Destructive shifted Poisson cure rate model under interval censoring

2.1 Introduction

This chapter aims to extend the destructive cure rate model developed by Rodrigues et al. [89] to accommodate interval-censored data assuming initial competing risks follow a shifted Poisson distribution. The shifted Poisson distribution is preferable to the (standard) Poisson distribution for settings in which no cure component is present prior to a destructive process imposed by intervention. Applications exist in biomedical studies, such as the study of tumor recurrence in cancer patients, where all individuals are susceptible to the event of interest prior to treatment. In criminal recidivism applications, a group of inmates is susceptible to committing crime at time of arrest, however rehabilitative measures such as counseling or vocational training may take place during incarceration with the goal of preventing a relapse of crime. By using the identity weight function, the shifted Poisson distribution allows for modeling the scenario where at least one initial risk exists without introducing an additional parameter. Noting that both right-censored and left-censored data may be considered special cases of interval-censored data, this work provides a generalization of the destructive cure rate model. Further, the EM algorithm proposed by Gallardo et al. [37] is implemented in the interval-censored setting with a simulation study. This chapter is adapted from [95], previously published by *Communications in Statistics - Simulation and Computation*.

The rest of this chapter is organized as follows. In Section 2.2, we describe the destructive shifted Poisson cure rate model under interval censoring and present

the complete log-likelihood function. In Section 2.3, we develop the steps of the EM algorithm for the model in Section 2.2. Section 2.4 presents two Monte Carlo simulations, first evaluating the performance of the proposed EM algorithm in parameter recovery then comparing the performance of the EM algorithm to that of direct maximization of the observed log-likelihood function under three initial parameter settings. In Section 2.5, the EM algorithm is applied to real data from a study on crime recidivism.

2.2 Model formulation

Let us assume the unobserved number of competing risks M to follow a weighted Poisson distribution with probability mass function (pmf)

$$p_m = P[M = m; \theta, \phi] = \frac{w(m; \phi)p^*(m; \theta)}{E_\theta[w(M; \phi)]}, \quad m = 0, 1, 2, \ldots, \tag{2.1}$$

where $w(m; \phi)$ is a non-negative weight function with parameter ϕ, $p^*(m; \theta)$ is the pmf of a Poisson distribution with parameter $\theta > 0$, and $E_\theta[\cdot]$ implies that the expectation is taken with respect to $p^*(m; \theta)$. Taking $w(m; \phi) = m$, the form of (2.1) becomes

$$P[M = m; \theta] = \frac{me^{-\theta}\theta^m}{m!\theta} = \frac{e^{-\theta}\theta^{m-1}}{(m-1)!}, m = 1, 2, 3, \ldots, \tag{2.2}$$

thus $(M - 1)$ follows a Poisson distribution with parameter θ, and M is said to follow a shifted Poisson distribution. Now suppose that an intervention takes place resulting in a quantity $D(\leq M)$ competing risks remaining active, each of which could result in a detectable tumor with probability p. To each competing risk M we can associate a Bernoulli random variable X_j such that $P(X_j = 1) = p$. The remaining quantity of competing risks not destroyed can be modeled as

$$D = X_1 + X_2 + \ldots + X_M.$$

Proposition 2.2.1. For the cure rate model with the pmf of the number of competing risks as in (2.2), the distribution of D is given by

$$p_d = P(D = d; \theta, p) = \frac{e^{-\theta p}(\theta p)^d}{d!}[(1-p) + \frac{d}{\theta}], \quad d = 0, 1, 2, ..., M. \tag{2.3}$$

A proof of proposition 2.2.1 is provided in the Appendix A.1. We note that random variable D may be obtained as a sum of two random variables having Poisson and Bernoulli distributions with respective means θp and p, thus we will say $D \sim \text{Pois}(\theta p) + \text{Bern}(p)$.

Let W_a be the time taken for the ath active risk to produce a detectable disease (event of interest). The waiting times $W_a, a = 1, 2, \cdots$, are assumed to be identically and independently distributed with a common distribution function $F(t; \boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ is an unknown set of parameters, and are also independent of D. Because the number of active competing risks D and lifetime W_a associated with a given risk are latent variables, one generally only observes the time taken by the first active risk to produce the event. To accommodate the presence of a cured proportion, the time-to-event or lifetime is defined as

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

where W_0 is such that $P[W_0 = \infty] = 1$. After some algebra, the population survival function can be expressed as [89]

$$S_{\text{pop}}(t; \theta, p, \boldsymbol{\lambda}) = P(T > t) = \exp\{-\theta p F(t)\}(1 - pF(t; \boldsymbol{\lambda})).$$

For the sake of simplicity, we will use $F(\cdot)$ instead of $F(\cdot; \boldsymbol{\lambda})$. Similarly, we will use $S(\cdot)$ instead of $S(\cdot; \boldsymbol{\lambda})$. We note that $S_{\text{pop}}(\cdot)$ is an improper survival function with corresponding cure fraction $p_0 = S_{\text{pop}}(\infty; \theta, p, \boldsymbol{\lambda}) = e^{-\theta p}(1-p)$. To study the effects

of covariates, 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 and $\boldsymbol{z_2}$ is related to activation probability for non-destroyed risks, with corresponding link functions

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

where $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ represent vectors of regression coefficients. We note that $\boldsymbol{z_1}$ and $\boldsymbol{z_2}$ share no common elements, and an intercept term is excluded either from $\boldsymbol{\beta_1}$ or from $\boldsymbol{\beta_2}$ to avoid identifiability problems [52]. Let $\boldsymbol{\psi} = (\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \boldsymbol{\lambda})$ denote the vector of unknown parameters.

2.2.1 Form of data and likelihood function

We consider a scenario where the true lifetimes are not exactly observed and are subject to interval censoring. Adopting the form of data as described in Section 1.6.1, the observed data is denoted as $\boldsymbol{D}_{obs} = (\boldsymbol{l}, \boldsymbol{r}, \boldsymbol{\delta})$. Based on the observed data $(\boldsymbol{l}, \boldsymbol{r}, \boldsymbol{\delta})$, the observed likelihood function under non-informative censoring is given by

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

while the observed log-likelihood function is given by

$$l(\boldsymbol{\psi}|\boldsymbol{D}_{\text{obs}}) \propto \sum_{i=1}^{n} \left\{\delta_i \log\left[S_{\text{pop}}(Y_{j-1(i)}) - S_{\text{pop}}(Y_{j(i)})\right] + (1-\delta_i)\log S_{\text{pop}}(Y_{j(i)})\right\}. \quad (2.4)$$

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 [105], the joint distribution of the complete data can be expressed as

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

Noting that the second and third terms in the product above are well defined, we focus on the derivation of $f(l_i, r_i, \delta_i | D_i = d_i)$.

Proposition 2.2.2. For the cure rate model with the pmf of the number of active risks as in (2.3), the joint distribution of (l_i, r_i, δ_i) given $D_i = d_i$ under interval censoring is given by

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

A proof of proposition 2.2.2 is provided in the Appendix A.1.

Using proposition 2.2.2, the complete data likelihood function may be defined as

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

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

$$l_c(\boldsymbol{\psi}|\boldsymbol{D}_{\text{comp}}) = l_c(\boldsymbol{\beta}_1) + l_c(\boldsymbol{\beta}_2) + l_c(\boldsymbol{\lambda}) + K, \tag{2.5}$$

where

$$\begin{aligned} l_c(\boldsymbol{\beta}_1) &= \sum_{i=1}^{n} \big\{(m_i - 1)\log\theta_i - \theta_i\big\}, \\ l_c(\boldsymbol{\beta}_2) &= \sum_{i=1}^{n} \big\{d_i \log p_i + (m_i - d_i)\log(1 - p_i)\big\}, \\ l_c(\boldsymbol{\lambda}) &= \sum_{i=1}^{n} \big\{(d_i - \delta_i)\log S(Y_{j(i)}) + \delta_i \log[S(Y_{j-1(i)}) - S(Y_{j(i)})]\big\}, \end{aligned}$$

and K is a constant independent of model parameters, given by

$$K = \sum_{i=1}^{n} \left\{\delta_i \log d_i + \log \frac{m_i!}{d_i!(m_i - d_i)!} + \log[(m_i - 1)!]\right\}. \tag{2.6}$$

2.3 EM algorithm

In this section we present the construction of the EM algorithm to produce estimates of the parameters for the destructive shifted Poisson cure rate model under interval censoring. The first step in the implementation of the EM algorithm [60] requires taking the conditional expectation of a complete data log-likelihood function, such as (2.5), given some proposed parameter values. The second step requires maximizing the conditional expected complete data log-likelihood function. In the traditional approach, this may involve maximizing a complicated function that involves all model parameters. Such an approach is computationally less efficient and may lack robustness with respect to initial values, specifically in the presence of a large set of unknown parameters. The implementation of the EM algorithm proposed here is motivated by the work of Gallardo et al. [37] and uses the conditional distributions of both initial competing risks and active competing risks to decompose the conditional expectation of $l_c(\boldsymbol{\psi}|\boldsymbol{D}_{\text{comp}})$ into three simpler functions that can each be maximized independently. This makes the entire formulation of the EM algorithm much more efficient.

Let $\boldsymbol{\psi}^{(k)} = (\boldsymbol{\beta}_1^{(k)}, \boldsymbol{\beta}_2^{(k)}, \boldsymbol{\lambda}^{(k)})$ 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 (2.5),

$$\begin{aligned} Q(\boldsymbol{\psi}|\boldsymbol{\psi}^{(k)}) &= E\left[l_c(\boldsymbol{\beta}_1)|\boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}^{(k)}\right] + E[l_c(\boldsymbol{\beta}_2)|\boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}^{(k)}] + E[l_c(\boldsymbol{\lambda})|\boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}^{(k)}] + K^* \\ &= Q_1(\boldsymbol{\beta}_1|\boldsymbol{\psi}^{(k)}) + Q_2(\boldsymbol{\beta}_2|\boldsymbol{\psi}^{(k)}) + Q_3(\boldsymbol{\lambda}|\boldsymbol{\psi}^{(k)}) + K^*, \end{aligned}$$

with

$$Q_1(\boldsymbol{\beta}_1|\boldsymbol{\psi}^{(k)}) = \sum_{i=1}^{n} \left\{ \log\theta_i(\widetilde{M}_i^{(k)} - 1) - \theta_i \right\} \tag{2.7}$$

$$Q_2(\boldsymbol{\beta}_2|\boldsymbol{\psi}^{(k)}) = \sum_{i=1}^{n} \left\{ \widetilde{D}_i^{(k)} \log p_i + (\widetilde{M}_i^{(k)} - \widetilde{D}_i^{(k)}) \log(1-p_i) \right\} \tag{2.8}$$

$$Q_3(\boldsymbol{\lambda}|\boldsymbol{\psi}^{(k)}) \;=\; \sum_{i=1}^{n}\left\{(\widetilde{D}_i^{(k)}-\delta_i)\log S(Y_{j(i)})+\delta_i\log[S(Y_{j-1(i)})-S(Y_{j(i)})]\right\}, \quad (2.9)$$

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

Proposition 2.3.1. For the cure rate model with the pmf of the number of competing risks as in (2.2), the conditional distribution of $M_i-\delta_i$ given the observed data under interval censoring is given by

$$M_i-\delta_i|\boldsymbol{D}_{\text{obs}};\boldsymbol{\psi}\sim \text{Pois}\left(\theta_i[1-p_iF(Y_{j(i)})]\right)+\text{Bern}\left(\frac{1-p_iF(Y_{j(i)})}{1-p_iF(Y_{j(i)})+\frac{\delta_i}{\theta_i}}\right).$$

A proof of proposition 2.3.1 is provided in the Appendix A.1.

Proposition 2.3.2. For the cure rate model with the pmf of the number of active risks as in (2.3), the conditional distribution of $D_i-\delta_i$ given the observed data under interval censoring is given by

$$D_i-\delta_i|\boldsymbol{D}_{\text{obs}};\boldsymbol{\psi}\sim \text{Pois}(\theta_ip_iS(Y_{j(i)}))+\text{Bern}\left(\frac{p_iS(Y_{j(i)})}{1-p_iF(Y_{j(i)})+\frac{\delta_i}{\theta_i}}\right). \quad (2.10)$$

A proof of proposition 2.3.2 is provided in the Appendix A.1.

Applying the results of propositions 2.3.1 and 2.3.2, we can compute $\widetilde{D}_i^{(k)}, \widetilde{M}_i^{(k)}$ as

$$\begin{aligned}\widetilde{D}_i^{(k)} &= \theta_i[1-p_iF(Y_{j(i)})]+\frac{1-p_iF(Y_{j(i)})}{1-p_iF(Y_{j(i)})+\frac{\delta_i}{\theta_i}} \text{ and}\\ \widetilde{M}_i^{(k)} &= \delta_i+\theta_i[1-p_iF(Y_{j(i)})]+\frac{1-p_iF(Y_{j(i)})}{1-p_iF(Y_{j(i)})+\frac{\delta_i}{\theta_i}}-1,\end{aligned}$$

with $\theta_i = e^{\boldsymbol{z_{1i}}'\boldsymbol{\beta_1}}$ and $p_i = \frac{e^{\boldsymbol{z_{2i}}'\boldsymbol{\beta_2}}}{1+e^{\boldsymbol{z_{2i}}'\boldsymbol{\beta_2}}}$.

The next step is to maximize $Q_1(\boldsymbol{\beta}_1|\boldsymbol{\psi}^{(k)})$, $Q_2(\boldsymbol{\beta}_2|\boldsymbol{\psi}^{(k)})$, and $Q_3(\boldsymbol{\lambda}|\boldsymbol{\psi}^{(k)})$ independently with respect to $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$, 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 [80, 81]. 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 $\widetilde{M}_i^{(k)}$.

Step 2 (Maximization step or M-step): Given $\widetilde{\boldsymbol{D}}^{(\boldsymbol{k})} = (\widetilde{D}_1^{(k)}, \ldots, \widetilde{D}_n^{(k)})$ and $\widetilde{\boldsymbol{M}}^{(\boldsymbol{k})} = (\widetilde{M}_1^{(k)}, \ldots, \widetilde{M}_n^{(k)})$, find $\boldsymbol{\psi}^{(k)}$ that maximizes (2.7)-(2.9) 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}.

Note that in the work of Pal and Balakrishnan [75], an EM algorithm has been developed to estimate the parameters of the destructive length-biased Poisson cure rate model when the form of the data is right censored. In the construction of the EM algorithm, the authors considered the latent cured statuses to be the missing data. This resulted in one complicated objective function (containing all model parameters) to be maximized in the M-step of the EM algorithm. Such an approach is certainly not computationally efficient, and the computational complexity increases with an increase in the covariate dimension. In this thesis, we develop an EM algorithm for estimating the parameters of the destructive length-biased Poisson cure rate model when the form of the data is interval-censored. In this case, we assume the latent number of initial competing risks (M) and the latent number of active competing risks (D) to be the missing data. Interestingly, this approach, motivated by the work

of Gallardo et al. [37], results in an objective function that can be split into simpler functions. Each simple function can be maximized independently of the others since these simple functions do not share common model parameters. This makes the entire computation much more efficient and also suitable for incorporating more covariates.

2.4 Simulation study

This section presents the results of two simulation studies. The first study evaluates the performance of the proposed EM algorithm to recover parameter values, while comparison of the EM algorithm to direct maximization of the observed log-likelihood function is considered in the second study.

2.4.1 Parameter recovery

In this section, we evaluate the performance of the proposed EM algorithm to recover parameter values for a simulated data set. This empirical study partially mimics the real melanoma dataset which was used for illustrative purposes by Rodrigues et al. [89] using covariates of treatment group (1: treatment, 0: placebo) and tumor thickness (in mm).

2.4.1.1 Data generation

To simulate covariate data we first generated treatment group, denoted by z_1, from a Bern(0.5) distribution. 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 $\mathrm{U}(0.1, 20)$ distribution. We link parameter θ to treatment group only using $\theta = e^{\beta_{11} z_1}$ and parameter p to tumor thickness only, using $p = \frac{\exp(\beta_{20}+\beta_{21} z_2)}{1+\exp(\beta_{20}+\beta_{21} z_2)}$. Note that only one regression parameter corresponds to parameter θ in order to circumvent problems with identifiability in the sense of Li et al. [52]. A

higher cure rate, and consequently a smaller value of θ, is expected for the treatment group, while a lower cure rate and larger value of θ is expected for the placebo group. A negative value of β_{11} is consistent with this expectation, and we select $\beta_{11} = -0.5$. To select regression parameters for p, we posit that the activation probability will be proportionally higher for higher values of tumor thickness and select values $\beta_{20} = -1$, $\beta_{21} = 0.1$ in accordance with this observation. We shall 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

$$f(w) = f(w; \boldsymbol{\lambda}) = \frac{1}{\lambda_1 w}(\lambda_2 w)^{\frac{1}{\lambda_1}} e^{-(\lambda_2 w)^{\frac{1}{\lambda_1}}}, \quad w > 0, \lambda_1 > 0, \lambda_2 > 0. \tag{2.11}$$

Random censoring was introduced through censoring time C following exponential distribution with rate $\alpha = 0.05$. To generate interval-censored lifetime data $(l_i, r_i, \delta_i), i = 1, 2, \cdots, n$, we execute the following steps:

1. Generate censoring time C_i, competing risks $M_i \sim \text{Pois}(\theta) + 1$, and damaged cells $D_i | M_i = m_i \sim \text{Bin}(m, p)$;
2. If $D_i = 0$, then $(l_i, r_i, \delta_i) = (C_i, \infty, 0)$ and data generation is complete.
3. If $D_i > 0$, generate times to event due to each non-eliminated risk, $W_j, j = 1, 2, \cdots, D_i$, from the considered Weibull distribution with parameter $\boldsymbol{\lambda}$;
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 $\text{U}(0, 1)$ distribution and l_{2i} from $\text{U}(0.2, 0.7)$ distribution. Construct intervals $(0, l_{1i}], (l_{1i}, l_{1i} + l_{2i}], \cdots, (l_{1i} + k \times l_{2i}, \infty], k = 1, 2, \cdots$, and select (l_i, r_i) that satisfies $l_i < Y_i \leq r_i$.

The choice of the lifetime parameter $\boldsymbol{\lambda}$ 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 for the same fixed choice

of mean as 5, which gives us two possible choices of (λ_1, λ_2) as $(0.215, 0.183)$ and $(0.316, 0.179)$. Sample sizes of both $n = 100$ and $n = 200$ are used in order to observe the performance of the algorithm under small and moderate sample sizes. This study applies the EM algorithm proposed in Section 2.3 to interval-censored data from the destructive shifted Poisson cure rate model, simulated using the parameters and data generation methods outlined above. All simulations are done using the R statistical software (version 4.0.5) and all results are based on $M = 1000$ Monte Carlo runs. Computational codes for data generation and EM algorithm are available in the Appendix C.1.

2.4.1.2 Parameter estimation

To assess the performance of this method in parameter recovery, we report the empirical bias and root mean square error (RMSE) of the estimates, as well as the coverage probabilities (CP) of the asymptotic confidence intervals at 95% confidence level under the assumption that the estimators are asymptotically normally distributed. The results are presented in Table 2.1. Initial values of parameter estimates for a given Monte Carlo run were selected as follows: for a given parameter Γ, initial guess Γ_{init} is generated such that $\Gamma_{\text{init}} = \Gamma + \text{U}(0, 0.1)|\Gamma|$.

We observe that in all cases, the EM algorithm produces parameter estimates with high accuracy. The empirical coverage probabilities are close to the nominal level for all model parameters, and in all cases the bias and RMSE decrease with an increase in sample size, which is consistent with the large sample properties.

2.4.2 Comparison with direct maximization of log-likelihood function

In this study, we compare the performance of the proposed EM algorithm to that of direct maximization of the observed log-likelihood function using three different methods for selecting initial parameter estimates:

- $\psi_0 = (\beta_{11}, \beta_{20}, \beta_{21}, \lambda_1, \lambda_2)$
- $\psi_1 = (\beta_{11} + \epsilon_1|\beta_{11}|, \beta_{20} + \epsilon_2|\beta_{20}|, \beta_{21} + \epsilon_3|\beta_{21}|, \lambda_1 + \epsilon_4|\lambda_1|, \lambda_2 + \epsilon_5|\lambda_2|)$
- $\psi_2 = (\beta_{11} - \epsilon_1|\beta_{11}|, \beta_{20} - \epsilon_2|\beta_{20}|, \beta_{21} - \epsilon_3|\beta_{21}|, \lambda_1 - \epsilon_4|\lambda_1|, \lambda_2 - \epsilon_5|\lambda_2|)$,

where $\epsilon_i \sim \mathrm{U}(0, 0.5), i = 1, \cdots, 5$.

For each true and initial value parameter setting, $M = 1000$ convergent Monte Carlo runs were performed with $n = 100$ using both the proposed EM algorithm and direct maximization of the observed log-likelihood function (DM) with the "optim()" function in R. The median (Med) for each parameter under all settings was computed. To identify atypical estimate values, we compute Q_1, Q_3, and $IQR = Q_3 - Q_1$, the first quartile, third quartile, and interquartile range, respectively, and report the percentage of times that the estimate is outside the interval $(Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR)$, denoted by PT. Results are presented in Table 2.2. With the exception of median of estimates for parameter β_{11}, the medians of estimates from the EM algorithm are either comparable or closer to the true values than the medians of estimates obtained from direct maximization of the log-likelihood function. In comparing the PT values however, we see evidence that the atypical values obtained through the EM algorithm are consistently smaller, most notably when comparing percentage of atypical values in estimating β_{11}.

Table 2.1. Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP)

n	Parameter	Estimate (SE)	Bias	RMSE	95% CP
100	$\beta_{11} = -0.5$	−0.502 (1.135)	−0.002	0.580	0.957
	$\beta_{20} = -2$	−2.090 (0.556)	−0.090	0.563	0.967
	$\beta_{21} = 0.1$	0.103 (0.044)	0.003	0.045	0.952
	$\lambda_1 = 0.215$	0.210 (0.029)	−0.005	0.031	0.912
	$\lambda_2 = 0.183$	0.184 (0.008)	0.001	0.008	0.942
100	$\beta_{11} = -0.5$	−0.479 (1.076)	0.021	0.590	0.947
	$\beta_{20} = -2$	−2.095 (0.551)	−0.095	0.553	0.969
	$\beta_{21} = 0.1$	0.103 (0.043)	0.003	0.044	0.954
	$\lambda_1 = 0.316$	0.309 (0.042)	−0.007	0.043	0.934
	$\lambda_2 = 0.179$	0.180 (0.011)	0.001	0.011	0.939
200	$\beta_{11} = -0.5$	−0.568 (0.838)	−0.068	0.472	0.963
	$\beta_{20} = -2$	−2.030 (0.384)	−0.030	0.365	0.966
	$\beta_{21} = 0.1$	0.100 (0.030)	0.000	0.029	0.962
	$\lambda_1 = 0.215$	0.214 (0.021)	−0.001	0.019	0.954
	$\lambda_2 = 0.183$	0.183 (0.005)	0.000	0.006	0.944
200	$\beta_{11} = -0.5$	−0.532 (0.800)	−0.032	0.481	0.958
	$\beta_{20} = -2$	−2.032 (0.382)	−0.032	0.375	0.963
	$\beta_{21} = 0.1$	0.100 (0.030)	0.000	0.030	0.953
	$\lambda_1 = 0.316$	0.314 (0.030)	−0.002	0.031	0.946
	$\lambda_2 = 0.179$	0.179 (0.008)	0.000	0.008	0.948

Table 2.2. Comparison of the EM algorithm with direct maximization of the observed log-likelihood function (DM) for n=100

		ψ_0		ψ_1		ψ_2	
Approach	Parameter	Med	PT (%)	Med	PT (%)	Med	PT (%)
EM	$\beta_{11} = -0.5$	−0.680	0.0	−0.729	0.0	−0.770	0.0
	$\beta_{20} = -2$	−2.043	0.8	−2.059	1.7	−2.045	1.2
	$\beta_{21} = 0.1$	0.099	1.6	0.099	1.1	0.099	1.0
	$\lambda_1 = 0.215$	0.207	1.3	0.209	1.0	0.210	0.8
	$\lambda_2 = 0.183$	0.183	0.8	0.183	0.6	0.183	0.8
	$\beta_{11} = -0.5$	−0.624	0.0	−0.495	0.0	−0.705	0.0
	$\beta_{20} = -2$	−2.044	0.9	−2.081	1.4	−2.016	1.0
	$\beta_{21} = 0.1$	0.100	0.5	0.103	1.9	0.101	0.9
	$\lambda_1 = 0.316$	0.306	0.6	0.311	0.3	0.308	0.6
	$\lambda_2 = 0.179$	0.179	1.0	0.179	1.0	0.178	1.2
DM	$\beta_{11} = -0.5$	−0.402	8.0	−0.274	10.4	−0.502	8.9
	$\beta_{20} = -2$	−2.041	1.9	−2.025	1.6	−2.077	0.7
	$\beta_{21} = 0.1$	0.102	1.4	0.097	2.0	0.105	1.5
	$\lambda_1 = 0.215$	0.207	1.5	0.211	1.1	0.210	0.7
	$\lambda_2 = 0.183$	0.183	0.5	0.184	0.2	0.183	1.5
	$\beta_{11} = -0.5$	−0.340	8.1	−0.353	7.7	−0.461	11.5
	$\beta_{20} = -2$	−2.050	1.8	−2.003	1.4	−2.084	1.4
	$\beta_{21} = 0.1$	0.101	2.0	0.097	1.5	0.103	1.6
	$\lambda_1 = 0.316$	0.305	1.0	0.311	0.9	0.309	0.5
	$\lambda_2 = 0.179$	0.179	1.3	0.180	1.5	0.179	0.6

To compare the divergence rate of direct maximization of the observed log-likelihood function to the proposed EM algorithm, we compute the percentage of divergent runs out of $M = 1000$ Monte Carlo runs of each method using data simulated with true parameter values as in Section 2.4.1.1, sample sizes of both $n = 100$ and $n = 200$, and initial parameter estimates $\psi_0, \psi_1,$ and ψ_2. We note that parameter setting 1 is $\boldsymbol{\psi} = (-0.5, -2, 0.1, 0.215, 0.183)$ and parameter setting 2 is $\boldsymbol{\psi} = (-0.5, -2, 0.1, 0.305, 0.179)$. Additionally, in order to provide the best basis for comparison, the same simulated data sets and initial parameter estimates were used for both estimation methods under each sample size and parameter setting. Percentage of divergent runs, denoted by Div (%), are presented in Table 2.3. Direct maximization of the observed log-likelihood function is seen to be divergent in all settings, with the lowest divergent percentage of 16%, and percentages increasing to 40% and higher when initial estimates deviate from true parameter values. The EM algorithm, however, is observed to be convergent in all settings.

Table 2.3. Comparison of divergence of the EM algorithm with direct maximization of the observed log-likelihood function

		EM			DM		
		ψ_0	ψ_1	ψ_2	ψ_0	ψ_1	ψ_2
n	Parameter Setting	Div (%)	Div (%)	Div (%)	Div (%)	Div (%)	Div (%)
100	1	0	0	0	28.8	67.7	58.1
	2	0	0	0	30.4	60.8	47.9
200	1	0	0	0	17.3	61.8	50.7
	2	0	0	0	16	56.8	42.5

2.5 Illustration using data from a crime recidivism study

In this section, we apply the proposed EM algorithm to a real data on crime recidivism. The dataset, available in the "penPHcure" package in R software, includes 432 inmates who were released from Maryland state prisons from October 1971 to June 1973. Subjects were followed for one year after release, with the event of interest being rearrest. The aim of this study, conducted by Rossi et al. [91], was to investigate the relationship between the time to first arrest after release and some covariates observed during the follow-up period. The data contains continuous covariates age at release (Age) and number of prior convictions, and binary covariates race (black or other), marriage status at release, number of prior convictions, whether the subject received financial aid after release, whether the subject had full-time work experience prior to incarceration, and whether the individual was working full time during the observation period (EMP, 1:Employed and 0:Not Employed). It bears mentioning that while the employment status of some subjects was observed to vary over time, for the purpose of this analysis we take EMP as employment status during the last observation interval. To mimic the covariate configuration in Section 2.2 we considered regression models linking each potential continuous and binary covariate pair to model parameters θ and p. The proposed algorithm produced MLEs for models using each potential covariate pair, and results from all models are available from the author upon request. For illustrative purposes, we select one binary covariate, EMP, and one continuous covariate, Age, to examine the results of the proposed algorithm. Denoting EMP by z_1 and Age by z_2, this gives way to four potential regression models:

- Model 1.1: $\theta = e^{\beta_{10}+\beta_{11}z_1}, p = \frac{\exp(\beta_{21}z_2)}{1+\exp(\beta_{21}z_2)}$, where p contains no intercept term.
- Model 1.2: $\theta = e^{\beta_{11}z_1}, p = \frac{\exp(\beta_{20}+\beta_{21}z_2)}{1+\exp(\beta_{20}+\beta_{21}z_2)}$, where θ contains no intercept term.
- Model 2.1: $\theta = e^{\beta_{20}+\beta_{21}z_2}, p = \frac{\exp(\beta_{11}z_1)}{1+\exp(\beta_{11}z_1)}$, where p contains no intercept term.
- Model 2.2: $\theta = e^{\beta_{21}z_2}, p = \frac{\exp(\beta_{10}+\beta_{11}z_1)}{1+\exp(\beta_{10}+\beta_{11}z_1)}$, where θ contains no intercept term.

While the aim of this study is not model selection, we compare AIC and BIC values for the above models to identify the model that best fits the data by first selecting appropriate initial parameter estimates, applying the proposed EM algorithm to obtain MLEs, and computing the observed log-likelihood function (2.4) using MLEs. The process by which initial parameter estimates were selected is described in the Appendix B, and values for the observed log-likelihood function (Obs log-lik), AIC, and BIC are reported in Table 2.4. We choose the model producing the lowest AIC and BIC values, Model 1.1, as our working model. Kaplan-Meier curves stratified by EMP, as shown in Figure 2.1, level off to non-zero proportions which supports the presence of a cure component in the data. The stratified Kaplan-Meier curves do not intersect and demonstrate a trend of higher survival times for those subjects who were employed during the last observation interval.

Table 2.4. Model Discrimination

Model	Obs log-lik	AIC	BIC
1.1	-354.0	718.0	738.4
1.2	-356.6	723.3	743.6
2.1	-359.5	729.0	749.3
2.2	-355.1	720.3	740.6

Table 2.5 presents the estimates and standard errors of the parameters of the working model, as well as p-values and 95% asymptotic confidence intervals. The negative sign of the estimate for β_{11} agrees with the stratified Kaplan-Meier plot in Figure 2.1 and the predictor of EMP is significant at a 5% level of significance. The estimate of $\beta_{21} < 0$ indicates that older individuals tend to relapse later, but the predictor of Age fails to be significant at a 5% level of significance.

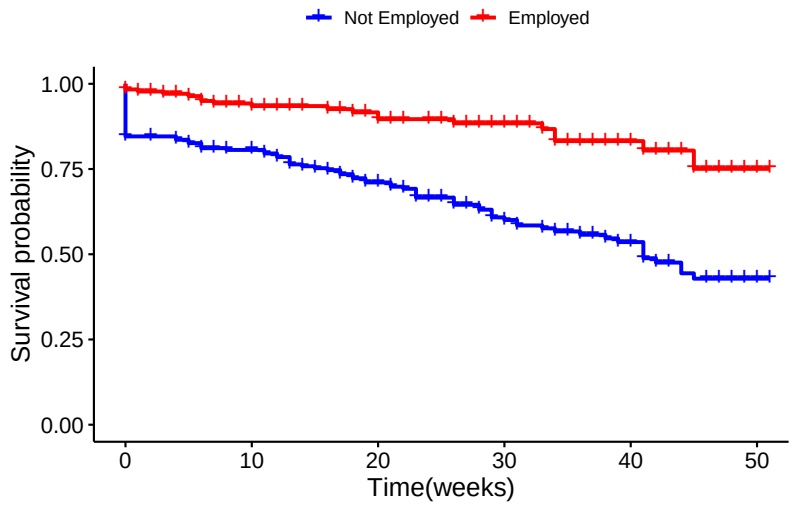

Figure 2.1. Kaplan-Meier plot of survival curves stratified by employment status.

Though the modeling framework differs from the structure adopted in this work, we find it interesting to compare the regression estimates reported herein with the analysis performed by Beretta and Heuchenne [12] using the methodology developed in R package "penPHcure." The semiparametric proportional hazards (PH) cure model of Sy and Taylor [94] was extended to accommodate time-varying covariates by using the covariate's time-weighted average over all observation periods and a PH cure model was fit to the crime recidivism data with all explanatory variables included in both the latency and incidence components. Confidence intervals created using the basic bootstrap method identified only one covariate, EMP, as statistically significant at a 5% level of significance. The regression coefficient for (time-varying) EMP was statistically significant in the latency component with a negative value which implies that among susceptible individuals, working full time after release is associated with a lower risk of rearrest. While the model as described in (2.2) does not separate the covariates' effects on the incidence and the latency, and the regression developed in this section does not consider the time-varying nature of EMP, our analysis is

consistent with the conclusion of Beretta and Heuchenne in finding EMP to have a statistically significant effect associated with a reduced incidence of rearrest.

Table 2.5. Estimation results corresponding to the working model for the crime recidivism data

Parameter	Estimate	Standard error	95% CI	$p-$value
β_{10}	2.024	0.625	(0.800, 3.249)	0.001
β_{11}	−1.500	0.500	(−2.481, −0.520)	0.003
β_{21}	−0.032	0.021	(−0.074, 0.009)	0.131
λ_1	0.757	0.071	(0.617, 0.897)	–
λ_2	0.009	0.003	(0.004, 0.014)	–

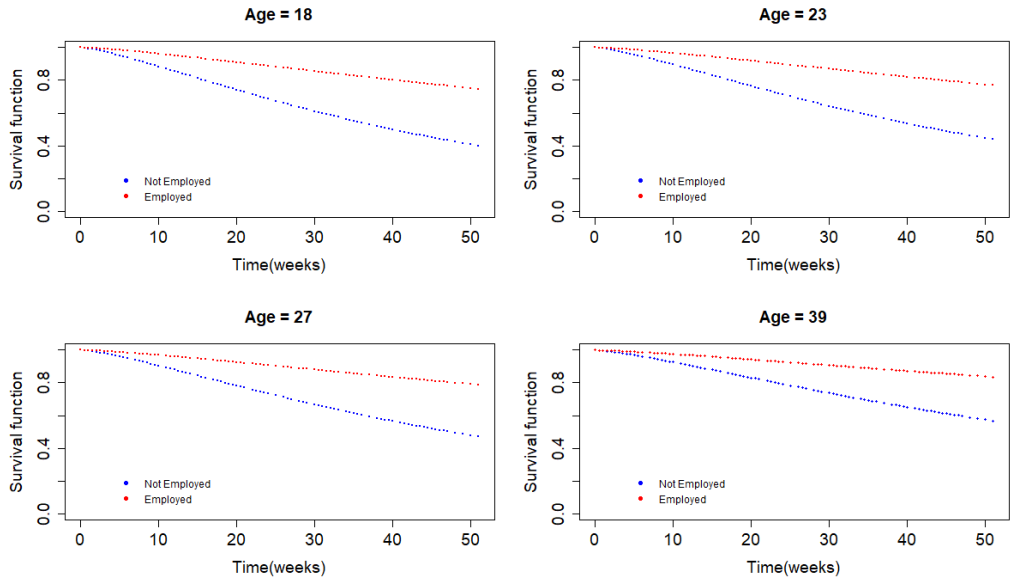

Figure 2.2. Predicted survival probabilities stratified by employment status for subjects of different ages.

Figure 2.2 shows the predicted survival probabilities for subjects with ages of 18, 23, 27 and 30.9 years, which correspond to the 5^{th}, 50^{th}, 75^{th}, and 95^{th} percentiles,

stratified by employment status. Note that the survival probability for employed subjects is higher across all values of age. While it may be seen by comparing the plots fixing age at 18 and 40.9 years that the survival probability is greater for greater values of age, this effect is more fully observed in Figure 2.3 where the cure rate is seen to increase in a nearly linear fashion as age increases.

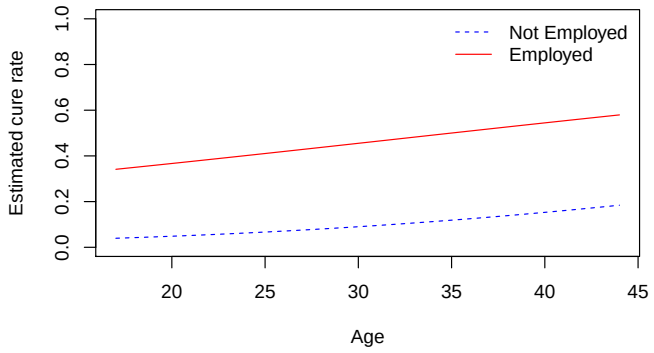

Figure 2.3. Cure rate against age stratified by employment status.

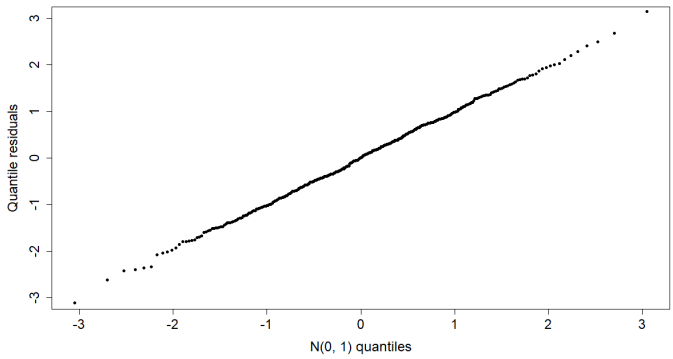

Figure 2.4. Q-Q plot of the normalized randomized quantile residuals.

We inspect the goodness-of-fit of our model by using the calculated normalized randomized quantile residuals [32]. Figure 2.4 presents the quantile-quantile plot, with each point corresponding to the median of five sets of ordered residuals. The linearity depicted in this plot supports the conclusion that the shifted Poisson destructive cure rate model with Weibull lifetimes provides a good fit to the crime recidivism data. Finally, the assumption of normality of residuals is tested using the Kolmogorov-Smirnov test. The resultant p-value of 0.9946 strongly supports the assumption of normality of residuals.

CHAPTER 3

Destructive negative binomial cure rate model under interval censoring

3.1 Introduction

In this chapter, we extend the destructive cure rate model developed by Rodrigues et al. [89] to accommodate interval-censored data assuming initial competing risks follow a negative binomial distribution; see also [69]. 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. While most established parameter estimation methods for the destructive negative binomial (DNB) 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 DNB 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 DNB cure rate model with interval-censored data. The EM algorithm is formulated first by using a profile likelihood approach for the estimation of shape parameter ϕ, and then using a mixture representation for the negative binomial

distribution, which is theoretically equivalent and offers the potential computational advantage of simultaneously maximizing all parameters. 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, as compared to both EM algorithms. 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 chapter is organized as follows. In Section 3.2, we present the destructive negative binomial cure rate model under interval censoring, define the form of the data and the complete and observed log-likelihood functions, and describe an alternative representation of the negative binomial distribution. In Section 3.3, the steps of the EM algorithm are formulated in detail for two distinct implementations of the EM algorithm, while Section 3.4 presents the steps of the SEM algorithm in detail and describes the advantages of the proposed SEM algorithm over both proposed EM algorithms. We present a three-way Monte Carlo simulation study in Section 3.5 to compare the performances of the SEM and EM algorithms. In Section 3.6, the SEM algorithm and preferred EM algorithm are applied to real data from a study on smoking cessation, illustrating both the advantage of the SEM algorithm and the practicality of using a destructive model in this context.

3.2 Model formulation

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 (e.g, death due to a disease 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. Hence, these malignant cells can be termed as competing risks.

Let M be a random variable denoting the initial number of competing risks. Assume M to follow a weighted Poisson distribution with probability mass function (pmf)

$$p_m = P[M = m; \theta, \phi] = \frac{w(m; \phi)p^*(m; \theta)}{E_\theta[w(M; \phi)]}, \quad m = 0, 1, 2, \ldots, \tag{3.1}$$

where $w(m; \phi)$ is a non-negative weight function with parameter ϕ, $p^*(m; \theta)$ is the pmf of a Poisson distribution with parameter $\theta > 0$, and $E_\theta[\cdot]$ implies that the expectation is taken with respect to $p^*(m; \theta)$. Taking $w(m; \phi) = \Gamma(m + \phi^{-1})$, the form of (3.1) becomes

$$P[M = m; \phi, \eta] = \frac{\Gamma(\phi^{-1} + m)}{m!\Gamma(\phi^{-1})} \left(\frac{\phi\eta}{1 + \phi\eta}\right)^m (1 + \phi\eta)^{-\phi^{-1}}, m = 0, 1, 2, \ldots, \tag{3.2}$$

Proposition 3.2.1. The weighted Poisson distribution of M in (3.2) with $\theta = \frac{\phi\eta}{1+\phi\eta}$ and $w(m; \phi) = \Gamma(m + \phi^{-1})$, is equivalent to consider

$$M \sim \text{NB}\left(\phi^{-1}, \frac{\phi\eta}{1 + \phi\eta}\right), \tag{3.3}$$

where $M \sim \text{NB}(r, p)$ conveys that M follows a negative binomial distribution with pmf $f(m; r, p) = \frac{\Gamma(m+r)}{m!\Gamma(m)} p^m (1 - p)^r$.

A proof of proposition 3.2.1 is provided in the Appendix A.2.

Suppose that an intervention takes place resulting in a quantity $D(\leq M)$ competing risks remaining active, each of which could bring about a detectable tumor 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. This results in only $D(\leq M)$ active malignant cells still capable of producing the tumor. Consistent with the existing literature [89], we assume a common activation probability for each risk factor within a patient. Later, to

capture the heterogeneity in the patient population, we propose to vary the activation probabilities across patients through incorporation of patient-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

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

Note that unlike the standard cure models that do not look at the destruction process of risk factors [69], the proposed approach allows a patient to be cured even in the presence of initial risk factors.

Proposition 3.2.2. For the cure rate model with the pmf of the number of competing risks as in (3.2), the distribution of active risks D is given by

$$P[D = d; \eta, \phi, p] = \frac{\Gamma(\phi^{-1} + d)}{\Gamma(\phi^{-1})d!} \left(\frac{\phi\eta p}{1 + \phi\eta p}\right)^d (1 + \phi\eta p)^{-\phi^{-1}}, d = 0, 1, 2, \ldots, M. \quad (3.4)$$

A detailed proof of proposition 3.2.2 is provided in the Appendix A.2. 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 $W_a, a = 1, 2, \cdots, D$, are assumed to be identically and independently distributed with a common distribution function $F(t; \boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ is an unknown set of parameters and is 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

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

where W_0 is such that $P[W_0 = \infty] = 1$. Utilizing the results of Rodrigues et al. [89], the population survival function can be expressed as

$$S_{\text{pop}}(t; \phi, \eta, p, \boldsymbol{\lambda}) = P(T > t) = [1 + \phi\eta p F(t; \lambda)]^{-\phi^{-1}}. \tag{3.5}$$

For the sake of simplicity, we will use $F(\cdot)$ instead of $F(\cdot; \boldsymbol{\lambda})$. Similarly, we will use $S(\cdot)$ instead of $S(\cdot; \boldsymbol{\lambda})$, where $S(\cdot; \boldsymbol{\lambda}) = 1 - F(\cdot; \boldsymbol{\lambda})$. Note that $S_{\text{pop}}(\cdot)$ is an improper survival function since $\lim_{t\to\infty} S_{\text{pop}} > 0$. From (3.5), the long-term survival probability or the cure fraction is given by

$$p_0 = S_{\text{pop}}(\infty; \phi, \eta, p, \boldsymbol{\lambda}) = [1 + \phi\eta p]^{-\phi^{-1}}. \tag{3.6}$$

We note that the parameters in the negative binomial cure rate model in (3.2) 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 the heterogeneity in patient 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 $\boldsymbol{z_1}$ and $\boldsymbol{z_2}$ share no common elements, and an intercept term must be excluded either from $\boldsymbol{\beta_1}$ or from $\boldsymbol{\beta_2}$ to circumvent identifiability problems [52]. Let $\boldsymbol{\psi} = (\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \boldsymbol{\lambda}, \phi)$ 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 $\boldsymbol{\lambda}$.

3.2.1 Form of data and likelihood function

We consider a scenario where the true lifetimes are not exactly observed and are subject to interval censoring. Adopting the form of data as described in Section 1.6.1, the observed data is denoted as $\boldsymbol{D}_{obs} = (\boldsymbol{l}, \boldsymbol{r}, \boldsymbol{\delta})$. Based on the observed data $(\boldsymbol{l}, \boldsymbol{r}, \boldsymbol{\delta})$, the observed likelihood function under non-informative censoring is given by

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

while the observed log-likelihood function is given by

$$l(\boldsymbol{\psi}|\boldsymbol{D}_{\text{obs}}) \propto \sum_{i=1}^{n} \left[\delta_i \text{log} \left\{S_{\text{pop}}(Y_{j-1(i)}) - S_{\text{pop}}(Y_{j(i)})\right\} + (1-\delta_i) \text{log} S_{\text{pop}}(Y_{j(i)})\right]. \quad (3.7)$$

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 [105], the joint distribution of the complete data can be expressed as

$$f(l_i, r_i, \delta_i, m_i, d_i) = f(l_i, r_i, \delta_i | D_i = d_i) P(D_i = d_i | M = m_i) P(M_i = m_i). \quad (3.8)$$

The second and third terms in the product above are well defined, and by proposition 2.2.2, we have that

$$f(l_i, r_i, \delta_i | D_i = d_i) = \left[d_i \{S(Y_{j-1(i)}) - S(Y_{j(i)})\}\right]^{\delta_i} \{S(Y_{j(i)})\}^{d_i - \delta_i}. \quad (3.9)$$

Using (3.8) and (3.9), the complete data likelihood function may be defined as

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

$$\frac{\Gamma(\phi^{-1}+m_i)}{\Gamma(\phi^{-1})m_i!}\left(\frac{\phi\eta_i}{1+\phi\eta_i}\right)^{m_i}(1+\phi\eta_i)^{-\phi^{-1}}.$$

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

$$l_c(\boldsymbol{\psi}|\boldsymbol{D}_{\text{comp}}) = l_c(\boldsymbol{\xi}) + l_c(\boldsymbol{\beta}_2) + l_c(\boldsymbol{\lambda}) + K, \tag{3.10}$$

where

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

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

$$K = \sum_{i=1}^{n}\left\{\delta_i\log d_i + \log\left(\frac{m_i!}{d_i!(m_i-d_i)!}\right) - \log(m_i!)\right\}. \tag{3.11}$$

3.2.2 Mixture representation of negative binomial distribution

Following the method employed by Gallardo et al. [39], we make use of an equivalent form to the distribution in (3.2) to present an alternative construction of the likelihood functions.

Proposition 3.2.3. The negative binomial distribution of M in (3.2) is equivalent to consider

$$M|U = u \sim \text{Poisson}(u) \text{ and } U \sim \text{Gamma}(\phi^{-1}, \phi\eta), \tag{3.12}$$

where Gamma(a,b) denotes the gamma distribution with density function $f(u;a,b) = b^{-a}u^{a-1}\mathrm{e}^{-\frac{u}{b}}/\Gamma(a)$.

A proof of proposition 3.2.3 is provided in the Appendix A.2.

Making use of the mixture representation, the complete data are then denoted by $\boldsymbol{D}_{\text{comp}} = (\boldsymbol{l}, \boldsymbol{r}, \boldsymbol{\delta}, \boldsymbol{M}, \boldsymbol{U}, \boldsymbol{D})$ which includes the observed data and the missing data,

where $\boldsymbol{M} = (M_1, M_2, \ldots, M_n)$, $\boldsymbol{U} = (U_1, U_2, \ldots, U_n)$, and $\boldsymbol{D} = (D_1, D_2, \ldots, D_n)$ are the missing data. The joint distribution of the complete data is expressed as

$$\begin{aligned} f(l_i, r_i, \delta_i, m_i, u_i, d_i) =& f(l_i, r_i, \delta_i | D_i = d_i) P(D_i = d_i | M_i = m_i) \times \\ & P(M_i = m_i | U_i = u_i) f_{U_i}(u_i; \eta_i, \phi). \end{aligned} \tag{3.13}$$

Using (3.9) and (3.13), the complete data likelihood function may be defined as

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

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

$$l_c(\boldsymbol{\psi}|\boldsymbol{D}_{\text{comp}}) = l_c(\boldsymbol{\xi}) + l_c(\boldsymbol{\beta}_2) + l_c(\boldsymbol{\lambda}) + K, \tag{3.14}$$

where

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

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

$$K = \sum_{i=1}^{n} \left\{ \delta_i \log d_i + \log \left(\frac{m_i!}{d_i!(m_i - d_i)!} \right) + m_i \log(u_i) - \log(m_i!) - u_i - \log(u_i) \right\}. \tag{3.15}$$

3.3 EM algorithm

In this section we present the construction of two distinct EM algorithms to produce estimates of the parameters for the DNB cure rate model under interval censoring. The first step in implementing the EM algorithm [60] requires taking the conditional expectation of a complete data log-likelihood function, such as (3.10) or (3.14), 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 two implementations of the EM algorithm proposed here are motivated by the works of Gallardo et al. [37, 39]. By using the conditional distributions of the missing data to decompose the conditional expectation of $l_c(\boldsymbol{\psi}|\boldsymbol{D}_{\text{comp}})$ into three simpler functions that can each be maximized independently, these formulation of the EM algorithm offer much greater efficiency.

3.3.1 EM algorithm with profile likelihood

Because taking the conditional expectation of the complete log-likelihood function in (3.10) requires computing $E\left[\log\left\{\Gamma(M_i+\phi^{-1})\right\}|\boldsymbol{D}_{\text{obs}};\boldsymbol{\psi}\right]$ which has no closed form, simultaneous maximization of all parameters in (3.10) 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 taken as the value for which the maximized log-likelihood function value is the maximum.

3.3.2 EM algorithm with simultaneous maximization

While estimation of the parameter ϕ is possible through the profile likelihood approach, this method can be computationally intensive as it requires the EM algorithm be performed for each permissible value of ϕ. As suggested by Gallardo et al. [39] we make use of proposition (3.2.3) to present an alternative construction of the EM algorithm which allows simultaneous maximization of all model parameters. We note that this representation introduces an additional latent variable, U, which, while lacking apparent biological interpretation, circumvents the issue of such calculations as $E[\Gamma(M_i + \phi^{-1})|\boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}]$ in (3.10), thus allowing ϕ to be estimated simultaneously.

We present the necessary expressions for the EM algorithms using profile likelihood (EM-PL) and simultaneous maximization (EM-SM) in parallel.

Let $\boldsymbol{\psi}^{(k)} = (\boldsymbol{\xi}^{(k)}, \boldsymbol{\beta}_2^{(k)}, \boldsymbol{\lambda}^{(k)})$ 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 (3.10) and (3.14),

$$\begin{aligned} Q(\boldsymbol{\psi}|\boldsymbol{\psi}^{(k)}) &= E[l_c(\boldsymbol{\xi})|\boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}^{(k)}] + E[l_c(\boldsymbol{\beta}_2)|\boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}^{(k)}] + E[l_c(\boldsymbol{\lambda})|\boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}^{(k)}] + K^* \\ &= Q_1(\boldsymbol{\xi}|\boldsymbol{\psi}^{(k)}) + Q_2(\boldsymbol{\beta}_2|\boldsymbol{\psi}^{(k)}) + Q_3(\boldsymbol{\lambda}|\boldsymbol{\psi}^{(k)}) + K^*, \end{aligned} \tag{3.16}$$

with

$$Q_1(\boldsymbol{\xi}|\boldsymbol{\psi}^{(k)}) = \begin{cases} \sum_{i=1}^n \left\{ \widetilde{M}_i^{(k)} \log \frac{\phi\eta_i}{1+\phi\eta_i} - \frac{\log(1+\phi\eta_i)}{\phi} \right\}, & \text{EM-PL} \\ \sum_{i=1}^n \left\{ \phi^{-1} \left[\widetilde{\log}(U_i^{(k)}) - \log\phi - \log\eta_i \right] - \frac{\widetilde{U}_i^{(k)}}{\phi\eta_i} - \log(\Gamma(\phi^{-1})) \right\}, & \text{EM-SM} \end{cases}$$

$$Q_2(\boldsymbol{\beta}_2|\boldsymbol{\psi}^{(k)}) = \sum_{i=1}^n \left\{ \widetilde{D}_i^{(k)} \log p_i + (\widetilde{M}_i^{(k)} - \widetilde{D}_i^{(k)}) \log(1-p_i) \right\},$$

$$Q_3(\boldsymbol{\lambda}|\boldsymbol{\psi}^{(k)}) = \sum_{i=1}^n \left\{ (\widetilde{D}_i^{(k)} - \delta_i) \log S(Y_{j(i)}) + \delta_i \log[S(Y_{j-1(i)}) - S(Y_{j(i)})] \right\},$$

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

Proposition 3.3.1. For the cure rate model with the pmf of the number of initial competing risks as in (3.2), the conditional distribution of $M_i - \delta_i$ given the observed data under interval censoring is given by

$$M_i - \delta_i|\boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi} \sim \text{NB}\left(\phi^{-1} + \delta_i, \frac{\phi\eta_i[1 - p_i F(Y_{j(i)})]}{1 + \phi\eta_i}\right). \tag{3.17}$$

A proof of proposition 3.3.1 is provided in the Appendix A.2.

Proposition 3.3.2. For the cure rate model with the pmf of the number of active risks as in (3.4), the conditional distribution of $D_i - \delta_i$ given the observed data under interval censoring is given by

$$D_i - \delta_i|\boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi} \sim \text{NB}\left(\phi^{-1} + \delta_i, \frac{\phi\eta_i p_i S(Y_{j(i)})}{1 + \phi\eta_i p_i}\right). \tag{3.18}$$

A proof of proposition 3.3.2 is provided in the Appendix A.2.

Proposition 3.3.3. For the cure rate model with the pmf of the number of competing risks as in (3.2), the conditional distributions of $M_i - \delta_i|U_i = u_i$ and U_i given $\boldsymbol{D}_{\text{obs}}$ are

$$M_i - \delta_i|U_i = u_i, \boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi} \sim \text{Poisson}(u_i[1 - p_i F(Y_{j(i)})])$$

and

$$U_i|\boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi} \sim \text{Gamma}\left(\phi^{-1} + \delta_i, \tfrac{\phi\eta_i}{1+\phi\eta_i p_i F(Y_{j(i)})}\right).$$

A proof of proposition 3.3.3 is provided in the Appendix A.2.

Applying the results of propositions 3.3.1, 3.3.2 and 3.3.3, we can compute $\widetilde{D}_i^{(k)}, \widetilde{M}_i^{(k)}, \widetilde{U}_i^{(k)}$ and $\widetilde{\log}\left(U_i^{(k)}\right)$ as

$$\begin{aligned}
\widetilde{D}_i^{(k)} &= \begin{cases} \delta_i + \frac{\eta_i^{(k-1)} p_i^{(k-1)} S(Y_{j(i)})(1+\delta_i\phi)}{1+\phi\eta_i^{(k-1)} p_i^{(k-1)} F(Y_{j(i)})}, & \text{for EM-PL} \\ \delta_i + \frac{\eta_i^{(k-1)} p_i^{(k-1)} S(Y_{j(i)})(1+\delta_i\phi^{(k-1)})}{1+\phi^{(k-1)}\eta_i^{(k-1)} p_i^{(k-1)} F(Y_{j(i)})}, & \text{for EM-SM} \end{cases} \\
\widetilde{M}_i^{(k)} &= \begin{cases} \delta_i + \frac{\left\{1-p_i^{(k-1)} F(Y_{j(i)})\right\}\eta_i^{(k-1)}(1+\phi\delta_i)}{1+\phi\eta_i^{(k-1)} p_i^{(k-1)} F(Y_{j(i)})}, & \text{for EM-PL} \\ \delta_i + \frac{\left\{(1-p_i^{(k-1)} F(Y_{j(i)})\right\}\eta_i^{(k-1)}(1+\phi^{(k-1)}\delta_i)}{1+\phi^{(k-1)}\eta_i^{(k-1)} p_i^{(k-1)} F(Y_{j(i)})}, & \text{for EM-SM} \end{cases} \\
\widetilde{U}_i^{(k)} &= \frac{(1+\phi^{(k-1)}\delta_i)\eta_i^{(k-1)}}{1+\phi^{(k-1)}\eta_i^{(k-1)} p_i^{(k-1)} F(Y_{j(i)})}, \text{ and} \\
\widetilde{\log}\left(U_i^{(k)}\right) &= \varphi(\phi^{-1(k-1)} + \delta_i) + \log(\phi^{(k-1)}) + \log(\eta_i^{(k-1)}) - \log(1+\phi^{(k-1)}\eta_i^{(k-1)} p_i^{(k-1)} F(Y_{j(i)})),
\end{aligned}$$

where $\varphi(\cdot)$ denotes the digamma function, $\eta_i = e^{\boldsymbol{z_{1i}}'\boldsymbol{\beta_1}}$, and $p_i = \frac{e^{\boldsymbol{z_{2i}}'\boldsymbol{\beta_2}}}{1+e^{\boldsymbol{z_{2i}}'\boldsymbol{\beta_2}}}$.

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 [80, 81]. 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):

EM-PL: 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.

EM-SM: For $i = 1, \ldots, n$, compute $\widetilde{D}_i^{(k)}, \widetilde{M}_i^{(k)}, \widetilde{U}_i^{(k)}$ and $\widetilde{\log}\left(U_i^{(k)}\right)$.

Step 2 (Maximization step or M-step):

EM-PL: Given $\widetilde{\boldsymbol{D}}^{(k)} = (\widetilde{D}_1^{(k)}, \ldots, \widetilde{D}_n^{(k)})$ and $\widetilde{\boldsymbol{M}}^{(k)} = (\widetilde{M}_1^{(k)}, \ldots, \widetilde{M}_n^{(k)})$, find $\boldsymbol{\psi}$ that

maximizes (3.16) in relation to $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$, and $\boldsymbol{\lambda}$, respectively, to obtain an improved estimate $\boldsymbol{\psi}^{(k+1)}$.

EM-SM: Given $\widetilde{\boldsymbol{D}}^{(k)} = (\widetilde{D}_1^{(k)}, \ldots, \widetilde{D}_n^{(k)})$, $\widetilde{\boldsymbol{M}}^{(k)} = (\widetilde{M}_1^{(k)}, \ldots, \widetilde{M}_n^{(k)})$, $\widetilde{\boldsymbol{U}}^{(k)} = (\widetilde{U}_1^{(k)}, \ldots, \widetilde{U}_n^{(k)})$, and $\widetilde{\log}\left(\boldsymbol{U}_{\boldsymbol{i}}^{(k)}\right) = (\widetilde{\log}(U_1^{(k)}), \ldots, \widetilde{\log}(U_n^{(k)}))$, find $\boldsymbol{\psi}^{(k)}$ that maximizes (3.16) in relation to $\boldsymbol{\xi}, \boldsymbol{\beta}_2$, 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-PL algorithm, we first select an initial grid of distinct admissible values of ϕ. We then employ the EM-PL 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-PL 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.

3.4 SEM algorithm

A well known drawback to the EM algorithm is that it does not guarantee convergence to a global or even a local maximum. 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. 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 [25, 69, 77]. These benefits lead us to consider the SEM algorithm as the proposed estimation method for the DNB 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. 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 3.3.1 and 3.3.2 to randomly generate the missing data, $\boldsymbol{M}$ and $\boldsymbol{D}$, given the observed data and current parameter estimates.

3.4.1 Steps of the SEM algorithm

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

Step 2: Stochastic step or S-step: Replace each missing datum, m_i and $d_i, i = 1, 2, \cdots, n$, 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 (3.17) and (3.18). 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 $i = 1, 2, \cdots, n$. 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

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

where

$$\begin{aligned}
l_c(\boldsymbol{\xi}; \widehat{\boldsymbol{m}}^{(0)}) &= \sum_{i=1}^{n} \left\{ \log\left[\Gamma(\phi^{-1} + \widehat{m_i}^{(0)})\right] - \phi^{-1}\log(1+\phi\eta_i) + \widehat{m_i}^{(0)} \log\frac{\phi\eta_i}{1+\phi\eta_i} \right\} - \\
&\quad n\log(\Gamma(\phi^{-1})), \\
l_c(\boldsymbol{\beta}_2; \widehat{\boldsymbol{m}}^{(0)}, \widehat{\boldsymbol{d}}^{(0)}) &= \sum_{i=1}^{n} \left\{ \widehat{d_i}^{(0)} \log p_i + (\widehat{m_i}^{(0)} - \widehat{d_i}^{(0)}) \log(1-p_i) \right\}, \\
l_c(\boldsymbol{\lambda}; \widehat{\boldsymbol{d}}^{(0)}) &= \sum_{i=1}^{n} \left\{ \left(\widehat{d_i}^{(0)} - \delta_i\right) \log S(Y_{j(i)}) + \delta_i \log[S(Y_{j-1(i)}) - S(Y_{j(i)})] \right\},
\end{aligned}$$

and

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

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

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

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 $\boldsymbol{\psi}^{(1)} = (\boldsymbol{\xi}^{(1)}, \boldsymbol{\beta}_2^{(1)}, \boldsymbol{\lambda}^{(1)})$ from Step 3, repeat steps 2 and 3 R times to obtain sequence $\boldsymbol{\psi}^{(k)}, k = 1, 2, \cdots, R$. 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 [30, 31].

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 (3.7) for each $\boldsymbol{\psi}^{(k)}, k = r, r+1, \cdots, R$, 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 [64], 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 [59, 106]. It is recommended to inspect a trace plot of the sequence of estimates versus the iterations to validate the sufficiency of the burn-in period and adjust the length of the burn-in period as needed.

3.5 Simulation Study

In this section, we evaluate the performance of the EM and SEM algorithms to recover parameter values for simulated data sets. This empirical study partially mimics the real melanoma dataset which was used for illustrative purposes by Rodrigues et al. [89] using covariates of treatment group (0: treatment, 1: placebo) and tumor thickness (in mm).

3.5.1 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. [52]. 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

$$f(w) = f(w; \boldsymbol{\lambda}) = \frac{1}{\lambda_1 w}(\lambda_2 w)^{\frac{1}{\lambda_1}} e^{-(\lambda_2 w)^{\frac{1}{\lambda_1}}}, \quad w > 0, \lambda_1 > 0, \lambda_2 > 0. \tag{3.19}$$

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 $(l_i, r_i, \delta_i), i = 1, 2, \cdots, n$, we execute the following steps:

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 $D_i|M_i = m_i \sim \text{Bin}(m_i, p_i)$;
2. If $D_i = 0$, then $(l_i, r_i, \delta_i) = (C_i, \infty, 0)$ and data generation is complete.
3. If $D_i > 0$, generate times to event due to each non-eliminated risk, $W_j, j = 1, 2, \cdots, D_i$, from the considered Weibull distribution with parameter $\boldsymbol{\lambda}$;
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 $(0, l_{1i}], (l_{1i}, l_{1i} + l_{2i}], \cdots, (l_{1i} + k \times l_{2i}, \infty], k = 1, 2, \cdots$, and select (l_i, r_i) that satisfies $l_i < Y_i \leq r_i$.

The choice of the lifetime parameter $\boldsymbol{\lambda}$ 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 $\boldsymbol{\lambda}$. 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 3.3 and 3.4 to interval-censored data from the destructive negative binomial cure rate model, simulated using the parameters and

data generation methods outlined above. 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 data generation, SEM and EM algorithms are available in the Appendix C.2.

3.5.2 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-PL algorithm, we select the initial grid for ϕ as $\{0.05, 0.1, \cdots, 2.05\}$. 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.

Tables 3.1, 3.2, 3.3, and 3.4 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 consider the performance of the EM-SM approach, which offers computational ease by using a mixture representation for the negative binomial distribution thus circumventing the estimation difficulties for parameter ϕ. First, regardless of the true value of ϕ, the EM-SM approach consistently and significantly underestimates ϕ, with a much larger bias, as well as SE and RMSE, as compared to the other two methods. The EM-SM method performs comparably in regards to the estimation of the remaining parameters, while also producing asymptotic confidence intervals close to the nominal level and demonstrating the large sample properties. However, accurate estimation of all parameters associated

Table 3.1. Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP) under setting $\boldsymbol{\psi} = \boldsymbol{\psi_1}$

n	Approach	Parameter	Estimate (SE)	Bias	RMSE	95% CP
200	EM-PL	$\beta_{11} = 1$	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	–
	EM-SM	$\beta_{11} = 1$	0.908 (0.442)	−0.092	0.311	0.973
		$\beta_{20} = -2$	−2.037 (0.518)	−0.037	0.466	0.989
		$\beta_{21} = 0.1$	0.091 (0.055)	−0.009	0.042	0.958
		$\lambda_1 = 0.215$	0.219 (0.032)	−0.006	0.023	0.990
		$\lambda_2 = 0.183$	0.186 (0.013)	0.003	0.007	0.998
		$\phi = 0.5$	0.045 (1.820)	-0.455	0.473	0.999
	SEM	$\beta_{11} = 1$	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	EM-PL	$\beta_{11} = 1$	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	–
	EM-SM	$\beta_{11} = 1$	0.910 (0.345)	−0.090	0.327	0.964
		$\beta_{20} = -2$	−2.041 (0.410)	−0.041	0.400	0.984
		$\beta_{21} = 0.1$	0.094 (0.044)	−0.006	0.033	0.984
		$\lambda_1 = 0.215$	0.217 (0.025)	0.002	0.018	0.980
		$\lambda_2 = 0.183$	0.186 (0.010)	0.003	0.006	1.000
		$\phi = 0.5$	0.100 (1.377)	−0.400	0.453	1.000
	SEM	$\beta_{11} = 1$	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 3.2. Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP) under setting $\boldsymbol{\psi} = \boldsymbol{\psi_2}$

n	Approach	Parameter	Estimate (SE)	Bias	RMSE	95% CP
200	EM-PL	$\beta_{11} = 1$	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	–
	EM-SM	$\beta_{11} = 1$	0.895 (0.492)	−0.105	0.334	0.971
		$\beta_{20} = -2$	−2.070 (0.568)	−0.070	0.628	0.982
		$\beta_{21} = 0.1$	0.090 (0.062)	−0.010	0.095	0.961
		$\lambda_1 = 0.215$	0.222 (0.036)	0.007	0.025	0.992
		$\lambda_2 = 0.183$	0.188 (0.015)	0.005	0.008	0.994
		$\phi = 0.8$	0.070 (2.215)	−0.730	0.753	1.000
	SEM	$\beta_{11} = 1$	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	EM-PL	$\beta_{11} = 1$	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	–
	EM-SM	$\beta_{11} = 1$	0.879 (0.365)	−0.121	0.296	0.964
		$\beta_{20} = -2$	−2.008 (0.417)	−0.008	0.379	0.996
		$\beta_{21} = 0.1$	0.086 (0.047)	−0.014	0.037	0.944
		$\lambda_1 = 0.215$	0.220 (0.027)	0.005	0.020	0.996
		$\lambda_2 = 0.183$	0.188 (0.011)	0.005	0.008	0.996
		$\phi = 0.8$	0.131 (1.643)	−0.669	0.706	1.000
	SEM	$\beta_{11} = 1$	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 3.3. Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP) under setting $\boldsymbol{\psi} = \boldsymbol{\psi_3}$

n	Approach	Parameter	Estimate (SE)	Bias	RMSE	95% CP
200	EM-PL	$\beta_{11} = 1$	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	–
	EM-SM	$\beta_{11} = 1$	0.984 (0.459)	−0.016	0.319	0.984
		$\beta_{20} = -2$	−2.059 (0.538)	−0.059	0.489	0.992
		$\beta_{21} = 0.1$	0.097 (0.060)	−0.003	0.047	0.968
		$\lambda_1 = 0.316$	0.318 (0.047)	0.002	0.036	0.980
		$\lambda_2 = 0.179$	0.182 (0.019)	0.003	0.011	0.992
		$\phi = 0.5$	0.265 (1.890)	−0.235	0.579	1.000
	SEM	$\beta_{11} = 1$	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	EM-PL	$\beta_{11} = 1$	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	–
	EM-SM	$\beta_{11} = 1$	0.950 (0.355)	−0.050	0.247	0.984
		$\beta_{20} = -2$	−2.007 (0.409)	−0.007	0.391	0.972
		$\beta_{21} = 0.1$	0.089 (0.044)	−0.011	0.034	0.964
		$\lambda_1 = 0.316$	0.321 (0.038)	0.005	0.026	0.992
		$\lambda_2 = 0.179$	0.183 (0.015)	0.004	0.009	1.000
		$\phi = 0.5$	0.132 (1.449)	−0.368	0.429	1.000
	SEM	$\beta_{11} = 1$	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 3.4. Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP) under setting $\boldsymbol{\psi} = \boldsymbol{\psi_4}$

n	Approach	Parameter	Estimate (SE)	Bias	RMSE	95% CP
200	EM-PL	$\beta_{11} = 1$	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	–
	EM-SM	$\beta_{11} = 1$	0.959 (0.508)	−0.041	0.317	0.988
		$\beta_{20} = -2$	−2.062 (0.569)	−0.062	0.553	0.992
		$\beta_{21} = 0.1$	0.094 (0.066)	−0.006	0.097	0.936
		$\lambda_1 = 0.316$	0.321 (0.053)	0.005	0.039	0.984
		$\lambda_2 = 0.179$	0.184 (0.023)	0.005	0.012	0.992
		$\phi = 0.8$	0.278 (2.387)	−0.522	0.737	0.996
	SEM	$\beta_{11} = 1$	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	EM-PL	$\beta_{11} = 1$	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	–
	EM-SM	$\beta_{11} = 1$	0.860 (0.369)	−0.140	0.319	0.928
		$\beta_{20} = -2$	−1.990 (0.456)	0.010	0.429	0.996
		$\beta_{21} = 0.1$	0.095 (0.053)	−0.005	0.056	0.960
		$\lambda_1 = 0.316$	0.319 (0.040)	0.003	0.028	0.964
		$\lambda_2 = 0.179$	0.184 (0.017)	0.005	0.011	0.988
		$\phi = 0.8$	0.289 (1.764)	−0.511	0.742	0.988
	SEM	$\beta_{11} = 1$	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

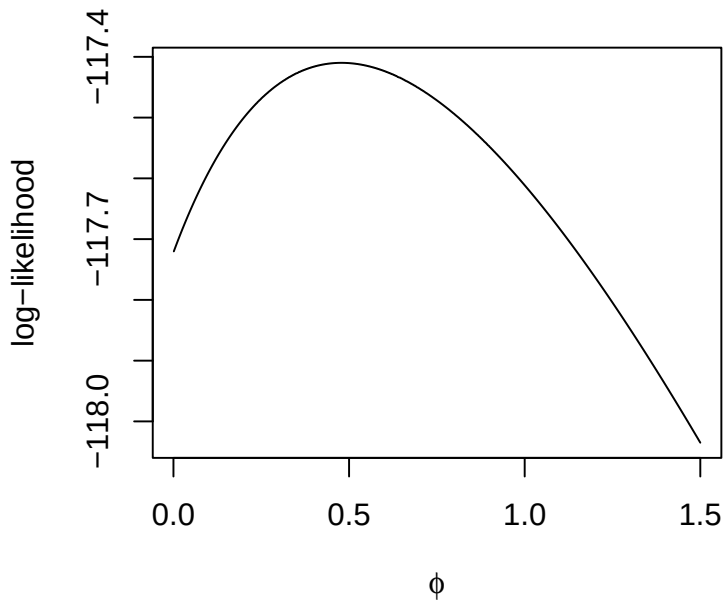

Figure 3.1. Profile likelihood plot for the parameter ϕ.

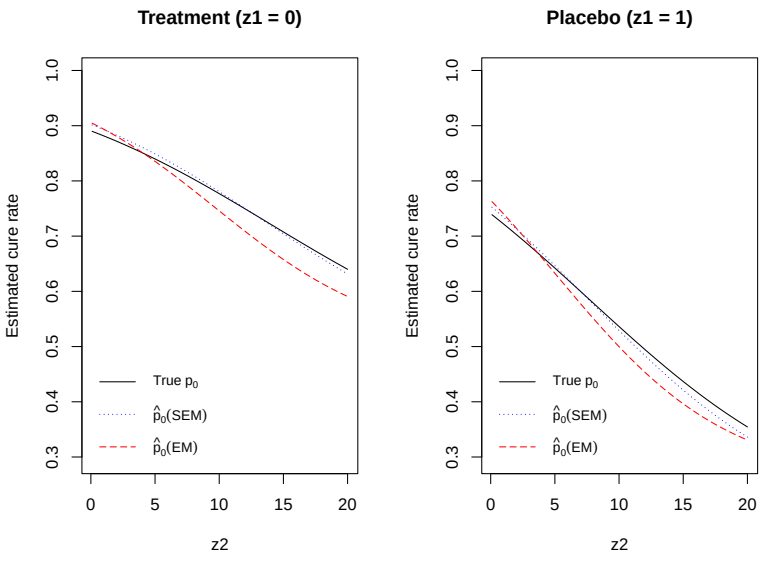

Figure 3.2. Plot of cure rate estimates against tumor thickness ($z2$).

with the cure probability is important in practice, as is the interpretability of the model. Because the addition of missing data variable U in the mixture representation of the negative binomial distribution lacks biological interpretation and provides no improvement in accuracy or efficiency for the parameter settings considered, we do not recommend the EM-SM algorithm as a viable inference method for the interval-censored DNB model. We note that while an EM algorithm was constructed using the mixture representation for the negative binomial distribution and applied to a right-censored real data [39], this work presents the first simulation study of the EM algorithm constructed as proposed in [39] using the mixture representation for the destructive negative binomial model. We restrict our further analysis to consideration of the EM-PL and SEM estimation methods.

First, we note that for all parameter settings, SEM produces smaller biases and root mean square errors (RMSEs) of the parameters associated with the cure probability, i.e., $\beta_{11}, \beta_{20}, \beta_{21}$, and ϕ, than the EM-PL approach. We also note that regardless of the true value of ϕ, the EM-PL approach overestimates ϕ. Though the RMSEs of EM-PL-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 3.1 for sample size 200 and setting $\boldsymbol{\psi} = \boldsymbol{\psi_1}$. As compared to the SEM approach, larger biases and RMSEs of EM-PL-produced ϕ, $\boldsymbol{\beta_1}$, and $\boldsymbol{\beta_2}$ estimates indicate that the SEM algorithm performs better than the EM-PL algorithm with regard to both accuracy and precision. Further, the profile likelihood method used in the EM-PL 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 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.

To illustrate the impact of accurate parameter estimation on the estimated cure rate, in Figure 3.2 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 the EM-PL-produced estimates ($\widehat{p}_0(\text{EM})$) lose accuracy for larger values of tumor thickness while the SEM-produced estimates ($\widehat{p}_0(\text{SEM})$) 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 $\widehat{p}_0(\text{EM})$ are consistently lower than $\widehat{p}_0(\text{SEM})$ for tumor thicknesses larger than 5 mm, with lowest accuracy observed in estimates for treatment group and larger tumor thickness values. While differences in relative error of $\widehat{p}_0(\text{SEM})$ and $\widehat{p}_0(\text{EM})$ 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.

3.6 Illustration using data from a smoking cessation study

In this section, we demonstrate the performance of the proposed SEM algorithm using a real data on smoking cessation. The study consists of 223 subjects who had attempted to quit smoking at least once during the study period of November 1986 to February 1989. For a full description of the data, see Murray et al. [62]. At the time of enrollment, subjects were randomly assigned to either a smoking intervention (SI) group (treatment group) or a usual care (UC) group (control group), which received no intervention. Subjects were monitored annually for a follow-up period of 5 consecutive years, with the event of interest being whether subjects resume smoking or not (relapse). The data consists of 65 (45 for SI and 20 for UC) out of 223 (169 for SI and 54 for UC) subjects experiencing relapse. For this application, we considered gender (GEN, 1:Female and 0:Male) and years smoking (DUR) as covariates of interest that are linked to model parameters η and p. We denote GEN by z_1 and DUR by z_2, and 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 smoking cessation, it is hoped by a practitioner that a medical or behavioral intervention will reduce a subject's propensity to re-engage in smoking. 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 an intervention reduced

the propensity to relapse. In order to identify the model best fitting the data, we compare AIC and BIC for 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 3.5. 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 3.5. 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 GEN, as shown in Figure 3.3, level off to non-zero proportions which supports the presence of a cure component in the data. While the stratified Kaplan-Meier curves intersect in the initial stage of the study ($t < 1$), the shapes of the curves convey a similar relationship between gender and long term survival.

Table 3.5. Model Discrimination

	EM			SEM		
Model	Obs log-lik	AIC	BIC	Obs log-lik	AIC	BIC
1	-213.3	438.6	459.0	-204.0	420.0	440.4
2	-202.9	417.8	438.2	-202.8	417.6	438.0
3	-205.9	423.8	444.2	-204.4	420.8	441.2
4	-203.4	418.8	439.2	-203.0	418.0	438.4
5	-206.2	424.3	444.8	-204.6	421.2	441.6

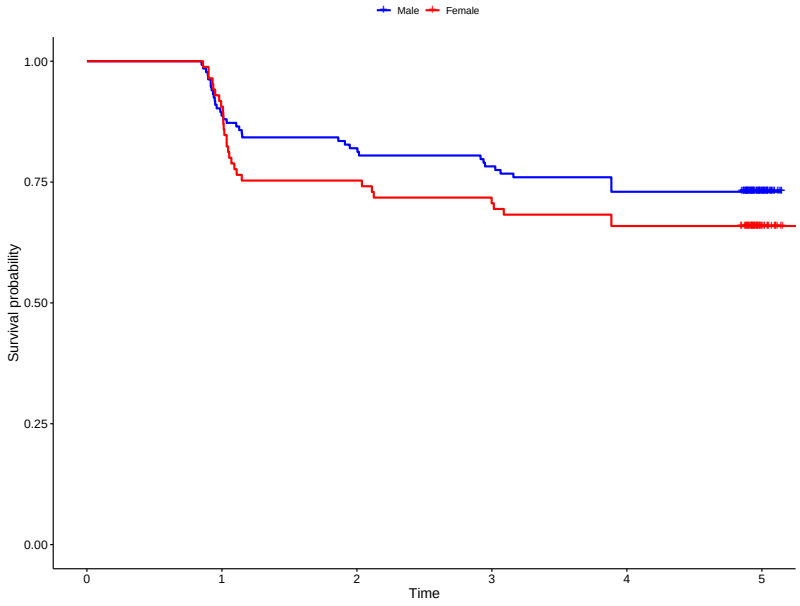

Figure 3.3. Kaplan-Meier plot of survival curves stratified by gender.

Table 3.6 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 1100 iterations and consider the first 100 iterations as burn-in. In computing the standard errors by inverting the observed information matrix, we observe that the second-order derivatives of the observed log-likelihood function are highly unstable, particularly with respect to the parameter ϕ. For this reason, we calculate and report the standard errors using the non-parametric bootstrap method and 200 bootstrap iterations for estimates produced by both SEM and EM algorithms. 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. While the positive sign of the estimate for β_{11} agrees with both the stratified Kaplan-Meier plot in Figure 3.3 and previous findings observing women as more likely to relapse than men [11], the predictor of GEN fails to be significant at a 5% level of significance. The estimate of $\beta_{21} < 0$ indicates that longer term smokers relapse later, which is consistent with the findings of Banerjee and Carlin [11]. Further, the significance of DUR at a 5% level of significance is consistent with prior analyses [73].

Table 3.6. Comparison of estimation results for the smoking cessation data

	EM			SEM		
Parameter	Estimate	Standard error	$p-$value	Estimate	Standard error	$p-$value
β_{11}	0.316	0.305	0.301	0.244	0.289	0.400
β_{20}	1.043	13.837	0.940	1.349	0.901	0.136
β_{21}	–0.058	0.352	0.870	–0.066	0.030	0.028
λ_1	0.389	0.030	–	0.405	0.024	–
λ_2	0.322	0.049	–	0.321	0.023	–
ϕ	1.200	1.561	–	0.587	0.083	–

Figure 3.4 shows the predicted survival probabilities for patients with smoking durations of 18, 31, 35 and 40.9 years, which correspond to the 5^{th}, 50^{th}, 75^{th}, and 95^{th} percentiles, stratified by gender. Note that the survival probability for males is higher across all values of duration. It may be observed that the survival probability increases for longer smoking duration by comparing the plots fixing duration at 18 and 40.9 years of smoking. The effect of smoking duration on survival is further conveyed in Figure 3.5 where the estimated cure rate is shown to increase in a nearly linear fashion as duration of smoking increases. Figure 3.6 shows the evolution paths

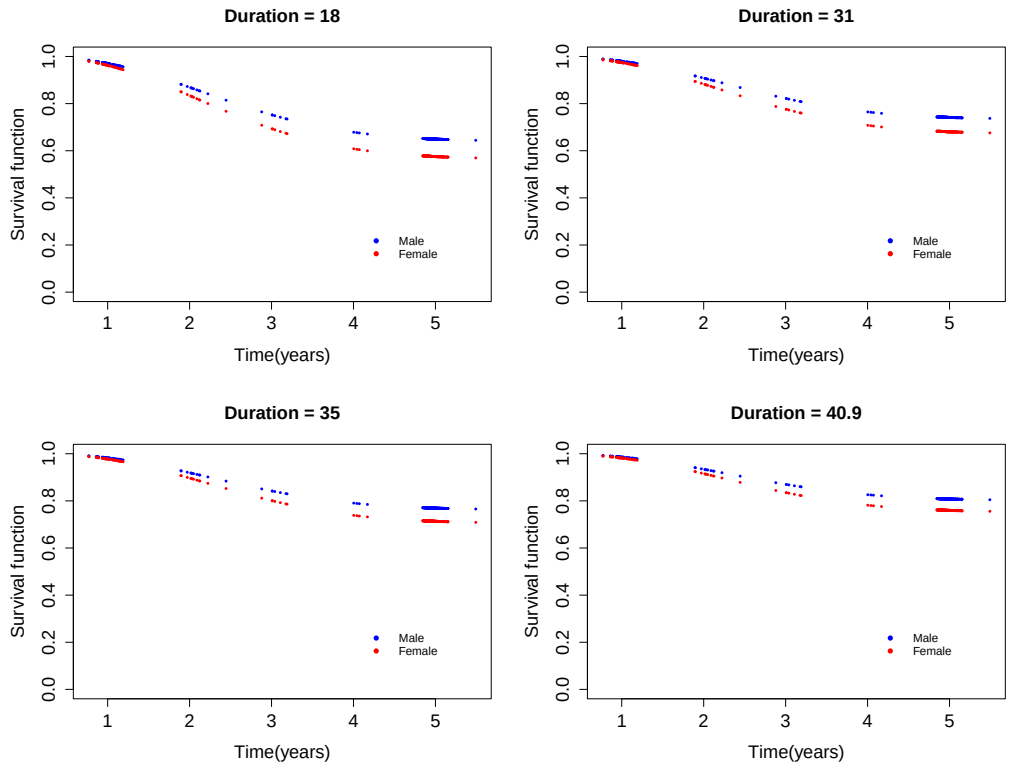

Figure 3.4. Predicted survival probabilities stratified by gender for patients with different durations of smoking.

of parameter estimates in the SEM algorithm. The estimates are observed to oscillate without any discernable upward or downward trend.

We check the adequacy of the DNB model by using the calculated normalized randomized quantile residuals [32]. Figure 3.7 presents the quantile-quantile plot, where each point 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 smoking cessation data. Finally, the Kolmogorov-Smirnov test for normality of residuals provides strong evidence for the normality of residuals, with a p-value of 0.986.

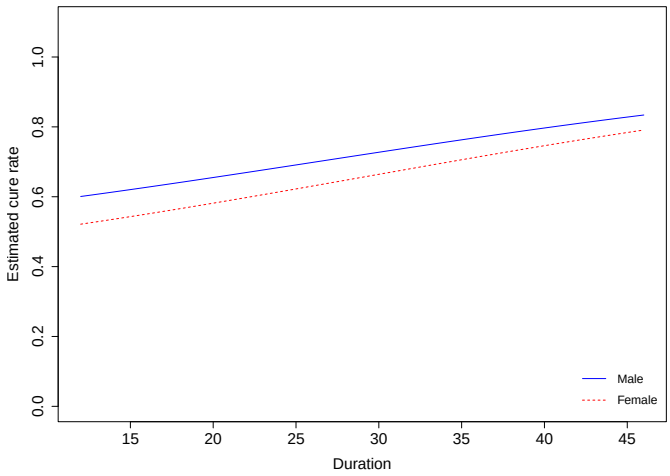

Figure 3.5. Cure rate against duration of smoking stratified by gender.

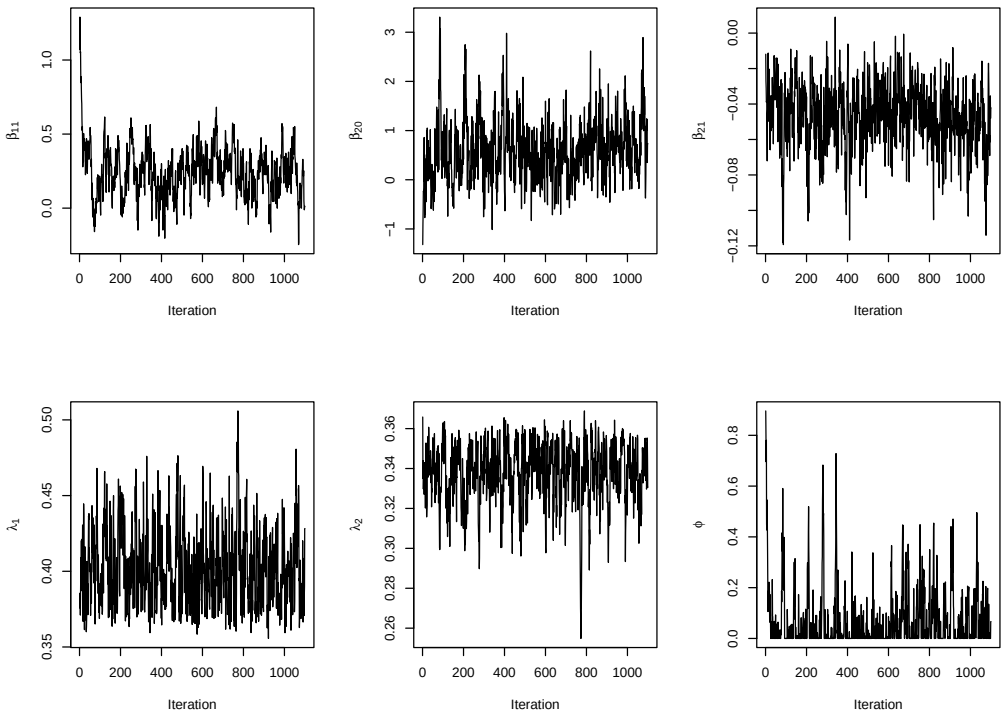

Figure 3.6. Parameter evolutions in the SEM algorithm.

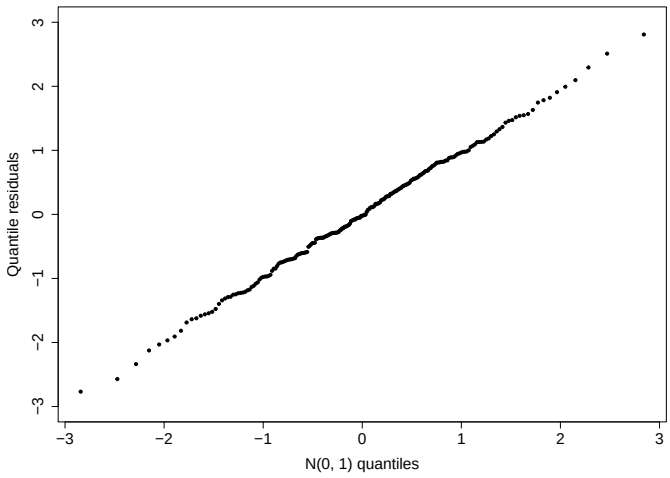

Figure 3.7. Q-Q plot of the normalized randomized quantile residuals.

CHAPTER 4

Box-Cox transformation cure rate model under interval censoring

4.1 Introduction

In this chapter, we extend the BCT cure rate model developed by Yin and Ibrahim [107] to interval-censored data. Assuming a semi-parametric setup, the main contribution of this work is twofold. First, we generalize the BCT model to accommodate interval-censored data. Further, the EM algorithm developed by Pal and Balakrishnan using a fully parametric framework [76] is implemented in the interval-censored setting, and an extensive simulation study demonstrates the performance of the proposed algorithm using simultaneous maximization of all parameters. Since the initial formulation of the EM algorithm for the BCT model used a profile likelihood technique, the proposition of simultaneous maximization introduces novelty and the potential for improved computational efficiency.

The rest of this chapter is organized as follows. In Section 4.2, we describe the BCT cure rate model under interval censoring and present the complete log-likelihood function. In Section 4.3, we develop the steps of the EM algorithm for the model in Section 4.2. Section 4.4 presents a Monte Carlo simulation study comparing the performance of the proposed EM algorithm, using both simultaneous maximization and profile likelihood techniques, and direct maximization of the observed log-likelihood function. In Section 4.5, the EM algorithm, using the preferred technique of simultaneous maximization, is applied to a real data from a study on smoking cessation.

4.2 Box-Cox transformation cure rate model

The Box-Cox transformation on a variable y, indexed by a transformation parameter α [19], is defined as

$$G(y,\alpha) = \begin{cases} \frac{y^{\alpha}-1}{\alpha}, & 0 < \alpha \leq 1 \\ \log(y), & \alpha = 0. \end{cases} \tag{4.1}$$

A general class of cure rate which unifies the mixture and promotion time cure models was proposed by applying a Box-Cox transformation to the population survival function, indexed by transformation parameter α [107]. Introducing a general covariate structure where $S_{\text{pop}}(t|\boldsymbol{x},\boldsymbol{z}))$ depends on set of covariates $\boldsymbol{x}$ and $\boldsymbol{z}$, the BCT cure rate model is defined as

$$G(S_{\text{pop}}(t|\boldsymbol{x},\boldsymbol{z}),\alpha) = -\theta(\alpha,\boldsymbol{z})F(t|\boldsymbol{x}),\ \alpha\ \epsilon\ [0,1] \tag{4.2}$$

where

$$\theta(\alpha,\boldsymbol{z}) = \begin{cases} \frac{\exp(\boldsymbol{z}'\boldsymbol{\beta})}{1+\alpha\exp(\boldsymbol{z}'\boldsymbol{\beta})}, & 0 < \alpha \leq 1 \\ \exp(\boldsymbol{z}'\boldsymbol{\beta}), & \alpha = 0 \end{cases} \tag{4.3}$$

and $F(\cdot)$ is a proper distribution function.

The BCT model has been studied with various modeling specifications. For example, [84] introduced a novel biological interpretation for the BCT model and assumed a proportional hazards structure to perform likelihood inference. Non-parametric maximum likelihood estimation was developed for the BCT model containing a frailty term for multivariate survival data [29]. A parametric set-up assuming a two-parameter Weibull distribution for $F(\cdot)$ was considered in the formulation of the EM algorithm for the BCT model [76]. In this work, we adopt a semi-parametric

framework by using the proportional hazards model in the latency. Accordingly, we incorporate set of covariates $\boldsymbol{x}$ through

$$F(t|\boldsymbol{x}) = 1 - S_0(t)^{\exp(\boldsymbol{\gamma}'\boldsymbol{x})} \tag{4.4}$$

where $S_0(\cdot)$ is the baseline survival function and $\boldsymbol{\gamma}$ represents the parameter vector associated with short-term survivors. While the proportional hazards structure allows for non-parametric estimation of the baseline survival $S_0(\cdot)$, for this work we will assume a one-parameter Weibull distribution for the baseline survival such that $S_0(t) = \exp(-t^{\frac{1}{\lambda}})$, for $\lambda > 0$. Applying (4.1) in the left hand side of (4.2), the population survival function is given by

$$S_{\text{pop}}(t|\boldsymbol{x}, \boldsymbol{z}) = \begin{cases} \{1 - \alpha\theta(\alpha, \boldsymbol{z})(1 - S_0(t)^{\exp(\boldsymbol{\gamma}'\boldsymbol{x})})\}^{1/\alpha}, & 0 < \alpha \leq 1 \\ \exp\{-\theta(0, \boldsymbol{z})(1 - S_0(t)^{\exp(\boldsymbol{\gamma}'\boldsymbol{x})})\}, & \alpha = 0. \end{cases} \tag{4.5}$$

The cure rate corresponding to the BCT cure rate model as in (4.5) is

$$\begin{aligned} p_0(\boldsymbol{x}, \boldsymbol{z}) = \lim_{t\to\infty} S_{\text{pop}}(t|\boldsymbol{x}, \boldsymbol{z}) &= \begin{cases} [1 - \alpha\theta(\alpha, \boldsymbol{z})]^{1/\alpha}, & 0 < \alpha \leq 1 \\ \exp\{-\theta(0, \boldsymbol{z})\}, & \alpha = 0. \end{cases} \\ &= \begin{cases} [1 + \alpha\exp(\boldsymbol{z}'\boldsymbol{\beta})]^{1/\alpha}, & 0 < \alpha \leq 1 \\ \exp\{-\exp(\boldsymbol{Z}'\boldsymbol{\beta})\}, & \alpha = 0. \end{cases} \end{aligned}$$

Note that (4.5) dictates constraints

$$0 \leq \alpha\theta(\alpha, \boldsymbol{z}) \leq 1 \text{ and } 0 \leq \exp\{-\theta(0, \boldsymbol{z})\} \leq 1, \tag{4.6}$$

(since $0 \leq F(t|\boldsymbol{x}) \leq 1$), which are automatically satisfied using (4.3). For the sake of simplicity, we will use θ, $F(\cdot)$, $S_{\text{pop}}(\cdot)$, and p_0 in place of $\theta(\alpha, \boldsymbol{z})$, $F(\cdot|\boldsymbol{x})$, $S_{\text{pop}}(\cdot|\boldsymbol{x}, \boldsymbol{z})$, and $p_0(\boldsymbol{x}, \boldsymbol{z})$, respectively.

4.2.1 Special cases

While all values such that $\alpha \geq 0$ satisfy the constraints in (4.6), we restrict the values of interest to $0 \leq \alpha \leq 1$ in order to examine intermediate modeling between the mixture and promotion time cure rate models. We present the special cases that occur at the boundaries of the interval of interest for α.

4.2.1.1 Mixture cure rate model: $\alpha = 1$

When $\alpha = 1$, $\theta(1, \boldsymbol{z}) = \frac{\exp(\boldsymbol{z}'\boldsymbol{\beta})}{1+\exp(\boldsymbol{z}'\boldsymbol{\beta})}$ and (4.5) reduces to

$$\begin{aligned} S_{\text{pop}}(t) &= 1 - \theta(1, \boldsymbol{z})F(t) \\ &= 1 - \frac{\exp(\boldsymbol{z}'\boldsymbol{\beta})}{1 + \exp(\boldsymbol{z}'\boldsymbol{\beta})}\{1 - S(t)\} \\ &= \frac{1}{1 + \exp(\boldsymbol{z}'\boldsymbol{\beta})} + \frac{\exp(\boldsymbol{z}'\boldsymbol{\beta})}{1 + \exp(\boldsymbol{z}'\boldsymbol{\beta})}S(t) \\ &= p_0 + (1 - p_0)S(t), \end{aligned}$$

which is the mixture model in (1.1) with $p_0 = \frac{1}{1+\exp(\boldsymbol{z}'\boldsymbol{\beta})}$, where $S(t) = 1 - F(t)$ denotes the proper survival function.

4.2.1.2 Promotion time cure model: $\alpha = 0$

When $\alpha = 0$, we have $\theta(0, \boldsymbol{z}) = \exp(\boldsymbol{z}'\boldsymbol{\beta})$ and (4.5) reduces to

$$S_{\text{pop}}(t) = \exp\{-\exp(\boldsymbol{z}'\boldsymbol{\beta})F(t)\},$$

which is the promotion time model in (1.2) with $\eta = \exp(\boldsymbol{z}'\boldsymbol{\beta})$.

4.2.2 Form of data and likelihood function

We consider a scenario where the true lifetimes are not exactly observed and are subject to interval censoring. Adopting the form of data as described in Section 1.6.1, the observed data is denoted as $\boldsymbol{D}_{obs} = (\boldsymbol{l}, \boldsymbol{r}, \boldsymbol{\delta})$.

Based on the observed data, the likelihood function under non-informative censoring is given by

$$L(\boldsymbol{\psi}) \propto \prod_{i=1}^{n} \{S_{\text{pop}}(l_i) - S_{\text{pop}}(r_i)\}^{\delta_i} \{S_{\text{pop}}(l_i)\}^{1-\delta_i}$$

$$= \prod_{\Delta_1} \{S_{\text{pop}}(l_i) - S_{\text{pop}}(r_i)\} \prod_{\Delta_0} S_{\text{pop}}(l_i)$$

where $\boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda, \alpha)$ denotes the unknown parameters, $\Delta_1 = \{i : \delta_i = 1\}$ corresponds to the set of observed lifetimes and $\Delta_0 = \{i : \delta_i = 0\}$ corresponds to the set of censored lifetimes.

4.3 EM algorithm

If the lifetime of a subject is interval-censored, then the cured status variable I takes the value one. However, if a subject's lifetime is right-censored, its value is unknown, and can be either zero or one. With the introduction of the missing data, we can estimate the MLE of $\boldsymbol{\psi}$ by the use of the EM algorithm, which incorporates the complete data likelihood function. For the BCT model as in (4.5), the survival function of the susceptibles is obtained as

$$\begin{aligned} S_{\text{susc}}(t|\boldsymbol{x}, \boldsymbol{z}) &= P[T > t | I = 1] \\ &= \frac{P[T > t] - P[I = 0]}{P[I = 1]} \\ &= \frac{S_{\text{pop}}(t) - p_0}{1 - p_0} \\ &= \begin{cases} \frac{\{1-\alpha\theta(1-S_0(t)^{\exp(\boldsymbol{\gamma}'\boldsymbol{x})})\}^{1/\alpha} - [1-\alpha\theta]^{1/\alpha}}{1-[1-\alpha\theta]^{1/\alpha}}, & 0 < \alpha \leq 1 \\ \frac{\exp\{-\theta(1-S_0(t)^{\exp(\boldsymbol{\gamma}'\boldsymbol{x})})\} - \exp\{-\theta\}}{1-\exp\{-\theta\}}, & \alpha = 0. \end{cases} \end{aligned} \tag{4.7}$$

We will denote $S_{\text{susc}}(\cdot|\boldsymbol{x}, \boldsymbol{z})$ by $S_{\text{susc}}(\cdot)$.

By using (4.7) and noting that any cure rate model can be represented as a mixture of cure and susceptible, the complete data likelihood function may be expressed as

$$L(\boldsymbol{\psi}) \propto \prod_{\Delta_1} \{S_{\text{pop}}(l_i) - S_{\text{pop}}(r_i)\} \prod_{\Delta_0} p_0^{1-I_i} \{(1-p_0) S_{\text{susc}}(l_i)\}^{I_i}$$

and the corresponding log-likelihood function is

$$l(\boldsymbol{\psi}) \propto \sum_{\Delta_1} \log\{S_{\text{pop}}(l_i) - S_{\text{pop}}(r_i)\} + \sum_{\Delta_0} (1-I_i) \log p_0 + \sum_{\Delta_0} I_i \log\{(1-p_0) S_{\text{susc}}(l_i)\}.$$

To implement the EM algorithm, in the E-step we replace unknown cure status I_i with w_i, the conditional expectation of I_i given the observed data and current parameter estimates, $\boldsymbol{\psi}^{(k)}$, where

$$\begin{aligned} w_i^{(k)} &= E(I_i | \boldsymbol{D}_{obs}; \boldsymbol{\psi}^{(k)}) \\ &= P[I_i = 1 | T_i > l_i; \boldsymbol{\psi}^{(k)}] \\ &= \frac{P[T_i > l_i | I_i = 1][P[I_i = 1]}{P[T_i > l_i]} \Big|_{\boldsymbol{\psi} = \boldsymbol{\psi}^{(k)}} \\ &= \frac{(1-p_0) S_{\text{susc}}(l_i)}{S_{\text{pop}}(l_i)} \Big|_{\boldsymbol{\psi} = \boldsymbol{\psi}^{(k)}}. \end{aligned}$$

The resulting conditional expectation of the complete data log-likelihood function is denoted by $Q(\boldsymbol{\psi}, w^{(k)})$, where

$$\begin{aligned} Q(\boldsymbol{\psi}, w^{(k)}) = \sum_{\Delta_1} \log\{S_{\text{pop}}(l_i) - S_{\text{pop}}(r_i)\} + \sum_{\Delta_0} (1 - w_i^{(k)}) \log p_0 + \\ \sum_{\Delta_0} w_i^{(k)} \log\{(1-p_0) S_{\text{susc}}(l_i)\}. \end{aligned} \tag{4.8}$$

In the maximization step, the function $Q(\boldsymbol{\psi}, w^{(k)})$ is maximized with respect to $\boldsymbol{\psi}$ to obtain a new estimate

$$\boldsymbol{\psi}^{(k+1)} = \arg\max_{\boldsymbol{\psi} \, \epsilon \, \boldsymbol{\Psi}} Q(\boldsymbol{\psi}, w^{(k)}).$$

The E-step and then M-step are then continued iteratively until a suitable convergence criterion is met to obtain the MLE of $\boldsymbol{\psi}$.

We can summarize the steps of the EM algorithm as follows:

Step 1 (Expectation step or E-step):

Given parameter estimates $\boldsymbol{\psi^{(k)}}$, for i in Δ_0 compute

$$w_i^{(k)} = \frac{(1-p_0)S_{\text{susc}}(l_i)}{S_{\text{pop}}(l_i)}|_{\boldsymbol{\psi}=\boldsymbol{\psi}^{(k)}} = \frac{S_{\text{pop}}(l_i)-p_0}{S_{\text{pop}}(l_i)}|_{\boldsymbol{\psi}=\boldsymbol{\psi}^{(k)}}.$$

Step 2 (Maximization step or M-step):

Given $\boldsymbol{w}^{(k)}$, find $\boldsymbol{\psi}$ that maximizes (4.8) 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}.

4.4 Simulation study

To determine the accuracy of the proposed model, we study the performance of the proposed model with simulated data. We consider the case where $\boldsymbol{x} = \boldsymbol{z}$ contains both a binary and continuous covariate.

4.4.1 Data generation

To generate the data for the i-th subject, we first generate covariates x_1 from a Bernoulli distribution taking values of 0 or 1 with probability 0.5 and x_2 from a uniform $U(0.1, 20)$ distribution. We assume a Weibull distribution for the baseline survival function such that the scale parameter is equal to 1, i.e. $S_0(t) = \exp(-t^{\frac{1}{\lambda}})$, where $\lambda > 0$. Random censoring is introduced through censoring time C following

exponential distribution with rate $\nu = 0.2$. Note that ν can be chosen to achieve a desired censoring proportion.

True parameter values were chosen to produce varying censored and cured proportions. Parameter settings chosen as $\boldsymbol{\psi_1} = (0.2, -1.4, 0.1, -1.2, 0.05, 0.215, \alpha)$ and $\boldsymbol{\psi_2} = (0.9, -0.6, -0.1, 1, -0.2, 0.316, \alpha)$, with α taking values $\{0, 0.5, 0.75, 1\}$, produced empirical censored and cured proportions (averaged over 100 Monte Carlo trials with sample size $n = 200$) as reported in Table 4.1. Choosing settings with empirical cured proportions varying from lowest proportion of 0.25 to highest proportion of 0.59 is useful in checking the algorithm's performance in scenarios where probability of cure may be relatively lower or higher. Sample sizes of both $n = 200$ and $n = 400$ are used in order to observe the performance of the algorithms under small and moderate sample sizes. This study applies the algorithms proposed in Section 4.3 to interval-censored data from the BCT cure rate model, simulated using the parameters and data generation methods outlined above. All simulations are done using the R statistical software (version 4.2.2) and all results are based on $M = 250$ Monte Carlo runs. Computational codes for data generation and EM algorithm are available in the Appendix C.3.

To generate the observed data $(l_i, r_i, \delta_i), i = 1, 2, \cdots, n$, we execute the following steps:

1. Generate censoring time C_i from exponential distribution with rate ν;
2. Given $X_1 = x_1$, $X_2 = x_2$, calculate θ, where
   (a) $\theta = \mathrm{e}^{\beta_0+\beta_1 x_1+\beta_2 x_2}$, if $\alpha = 0$
   (b) $\theta = \frac{\mathrm{e}^{\beta_0+\beta_1 x_1+\beta_2 x_2}}{1+\alpha \mathrm{e}^{\beta_0+\beta_1 x_1+\beta_2 x_2}}$, if $0 < \alpha \leq 1$
3. Calculate p_{0i}, where
   (a) $p_{0i} = \exp\{-\theta\}$ if $\alpha = 0$;
   (b) $p_{0i} = [1 - \alpha\theta]^{1/\alpha}$ if $0 < \alpha \leq 1$;

4. Generate a variable U_{1i} from a Uniform(0,1) distribution;
5. If $U_{1i} \leq p_{0i}$, then $(l_i, r_i, \delta_i) = (C_i, \infty, 0)$ and data generation is complete.
6. If $U_{1i} > p_{0i}$, generate another variable U_{2i} from a Uniform$(0, 1)$ distribution;
7. Time to event (T_i) is generated by setting (4.7) equal to U_{2i} and solving for T_i, namely
   (a) $T_i = \{-\mathrm{e}^{-\gamma_1 x_1 - \gamma_2 x_2} \log\left(1 + \theta \log\left[p_{0i} + (1 - p_{0i})U_{2i}\right]\right)\}^{\lambda}$ if $\alpha = 0$;
   (b) $T_i = \left\{-\mathrm{e}^{-\gamma_1 x_1 - \gamma_2 x_2} \log\left(\frac{\alpha\theta + [p_{0i} + U_{2i} - p_{0i}U_{2i}]^{\alpha} - 1}{\alpha\theta}\right)\right\}^{\lambda}$ if $0 < \alpha \leq 1$;
8. Set $Y_i = \min\{T_i, C_i\}$;
   (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$(0.1, 0.5)$ distribution. Construct intervals $(0, l_{1i}], (l_{1i}, l_{1i} + l_{2i}], \cdots, (l_{1i} + k \times l_{2i}, \infty], k = 1, 2, \cdots$, and select (l_i, r_i) that satisfies $l_i < Y_i \leq r_i$.

Table 4.1. Empirical cured proportions (Cure) and censored proportions (Cens) for chosen parameter settings

	ψ_1		ψ_2	
α	Cure	Cens	Cure	Cens
0	0.25	0.35	0.50	0.62
0.5	0.33	0.45	0.55	0.65
0.75	0.36	0.46	0.57	0.67
1	0.40	0.50	0.59	0.69

4.4.2 Parameter recovery

It is noted that the observed log-likelihood can be flat with respect to the transformation parameter α. Figure 4.1 presents the profile likelihood plots for a single dataset simulated using parameter setting $\boldsymbol{\psi} = \boldsymbol{\psi_1}$ and each possible value of α

considered, stratified by true value of α. In this setting, the profile likelihood is seen to be increasingly flat for larger values of α. Because flatness of the likelihood surface may produce convergence problems when maximizing all parameters simultaneously, we consider the performance of the proposed EM algorithm by using simultaneous maximization (EM-SM) as well as by using a profile likelihood technique (EM-PL) to estimate α.

To find an initial guess for the model parameters, we first employ the following selection method: for a given parameter Γ, we take 20% deviation off its true value and create the interval $(\Gamma - 0.2|\Gamma|, \Gamma + 0.2|\Gamma|)$. Then, we randomly sample a value of Γ from this interval which is used as the initial value (IV) to start off the iterative algorithm. Note that we restrict the selection of initial value of α to values that fall in the interval of interest, [0,1], by sampling from the intersection of the interval of interest and the interval constructed using 20% deviation as described above. To employ the profile likelihood approach to estimate α in the EM algorithm, we first select a set of possible values for α as $\{0, 0.05, 0.1, \cdots, 1\}$. Then, for each fixed value of α, the proposed EM algorithm is run and the log-likelihood value is calculated. The value of α which attains the maximized log-likelihood function value and corresponding estimates of other model parameters are taken as the MLEs. Tables 4.2, 4.3, 4.4, and 4.5 present the simulation results using both EM-PL and EM-SM techniques with α taking values 0, 0.5, 0.75, and 1, respectively. We first note that in all considered parameter settings, the proposed algorithm performs well when using both techniques in regards to biases, standard errors (SEs), and root mean square errors (RMSEs) of estimates. In all cases, the bias and RMSE decrease with an increase in sample size, which is consistent with the large sample properties. The transformation parameter α is estimated with greater accuracy using the EM-SM approach across all settings, as indicated by comparatively smaller biases.

Additionally, the EM-SM approach offers several computational advantages. Firstly, because the EM-SM approach requires only one convergent run of the EM algorithm to produce a set of parameter estimates whereas the EM-PL approach requires one convergent run of the EM algorithm for each permissible value of α, the computational time is expected to be lower using the EM-SM approach. For parameter setting $\boldsymbol{\psi} = \boldsymbol{\psi_1}$, $\alpha = 0$, and $n = 200$, one convergent run using the EM-SM approach took 1.89 seconds as compared to 10.97 seconds for one convergent run using the EM-PL approach, which clearly makes the EM-SM approach more computationally appealing. Secondly, because the profile likelihood method precludes the computation of standard error for α estimates through the inversion of the observed information matrix, the EM-SM technique offers the advantage of constructing confidence intervals for cure estimates without needing to produce estimates for the standard error of α through potentially computationally intensive methods such as bootstrapping. While the SEs and RMSEs of EM-PL-produced estimates are marginally smaller than those of EM-SM-produced estimates, the treatment of α 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 of some EM-PL-produced estimates is observed in all settings. The EM-SM approach produces empirical coverage probabilities that are close to the nominal level in all settings. Because the EM-SM approach provides reasonably accurate and efficient estimates with added computational advantages, the EM-SM approach is the preferred method to implement the proposed EM algorithm.

4.4.3 Comparison with direct maximization of the log-likelihood function

In considering the efficacy of the proposed EM algorithm in parameter estimation, we compare the performance of the proposed algorithm to direct maximization

Table 4.2. Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP) for parameter settings when $\alpha = 0$

		EM-PL				EM-SM			
n	Parameter	Estimate (SE)	Bias	RMSE	95% CP	Estimate (SE)	Bias	RMSE	95% CP
200	$\beta_0 = 0.2$	0.271 (0.238)	0.071	0.305	0.916	0.310 (0.320)	0.110	0.404	0.976
	$\beta_1 = -1.4$	−1.571 (0.287)	−0.171	0.432	0.900	−1.551 (0.432)	−0.190	0.514	0.972
	$\beta_2 = 0.1$	0.114 (0.025)	0.014	0.037	0.948	0.111 (0.032)	0.011	0.035	0.976
	$\gamma_1 = -1.2$	−1.255 (0.337)	−0.055	0.381	0.908	−1.215 (0.392)	−0.015	0.392	0.952
	$\gamma_2 = 0.05$	0.056 (0.023)	0.006	0.030	0.868	0.053 (0.033)	0.003	0.035	0.912
	$\lambda = 0.215$	0.214 (0.019)	−0.001	0.019	0.948	0.212 (0.021)	−0.003	0.020	0.968
	$\alpha = 0$	0.085 (–)	0.085	0.199	–	0.076 (0.192)	0.076	0.261	0.968
	$\beta_0 = 0.9$	1.147 (0.268)	0.247	0.485	0.772	1.052 (0.588)	0.152	0.583	0.968
	$\beta_1 = -0.6$	−0.691 (0.266)	−0.091	0.300	0.928	−0.680 (0.311)	−0.080	0.329	0.940
	$\beta_2 = -0.1$	−0.115 (0.025)	−0.015	0.033	0.888	−0.109 (0.035)	−0.009	0.036	0.964
	$\gamma_1 = 1$	1.031 (0.272)	−0.005	0.259	0.964	1.042 (0.277)	0.042	0.273	0.972
	$\gamma_2 = -0.2$	−0.202 (0.027)	−0.002	0.027	0.936	−0.205 (0.031)	−0.005	0.031	0.968
	$\lambda = 0.316$	0.311 (0.030)	−0.005	0.031	0.932	0.308 (0.033)	−0.008	0.032	0.928
	$\alpha = 0$	0.155 (–)	0.155	0.267	–	0.034 (0.432)	0.034	0.418	0.964
400	$\beta_0 = 0.2$	0.232 (0.162)	0.032	0.205	0.912	0.231 (0.197)	0.031	0.207	0.964
	$\beta_1 = -1.4$	−1.511 (0.194)	−0.111	0.265	0.904	−1.453 (0.272)	−0.053	0.274	0.972
	$\beta_2 = 0.1$	0.111 (0.017)	0.011	0.023	0.912	0.103 (0.021)	0.003	0.020	0.960
	$\gamma_1 = -1.2$	−1.247 (0.236)	−0.047	0.253	0.936	−1.259 (0.275)	−0.059	0.283	0.948
	$\gamma_2 = 0.05$	0.055 (0.016)	0.005	0.020	0.872	0.054 (0.023)	0.004	0.022	0.956
	$\lambda = 0.215$	0.216 (0.014)	0.001	0.015	0.940	0.211 (0.014)	−0.004	0.015	0.924
	$\alpha = 0$	0.063 (–)	0.063	0.131	–	0.026 (0.127)	0.026	0.129	0.992
	$\beta_0 = 0.9$	1.051 (0.178)	0.151	0.292	0.832	0.934 (0.407)	0.034	0.400	0.960
	$\beta_1 = -0.6$	−0.640 (0.180)	−0.040	0.177	0.952	−0.625 (0.215)	−0.025	0.215	0.952
	$\beta_2 = -0.1$	−0.109 (0.017)	−0.009	0.020	0.924	−0.102 (0.024)	−0.002	0.024	0.956
	$\gamma_1 = 1$	1.013 (0.189)	0.013	0.188	0.960	1.010 (0.192)	0.010	0.193	0.940
	$\gamma_2 = -0.2$	−0.199 (0.018)	0.001	0.019	0.944	−0.203 (0.022)	−0.003	0.022	0.956
	$\lambda = 0.316$	0.315 (0.022)	−0.001	0.021	0.976	0.310 (0.024)	−0.006	0.024	0.916
	$\alpha = 0$	0.105 (–)	0.105	0.195	–	0.012 (0.308)	0.012	0.294	0.956

Table 4.3. Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP) for parameter settings when $\alpha = 0.5$

		EM-PL				EM-SM			
n	Parameter	Estimate (SE)	Bias	RMSE	95% CP	Estimate (SE)	Bias	RMSE	95% CP
200	$\beta_0 = 0.2$	0.233 (0.322)	0.033	0.442	0.856	0.335 (0.532)	0.135	0.555	0.936
	$\beta_1 = -1.4$	−1.462 (0.322)	−0.062	0.473	0.820	−1.580 (0.603)	−0.180	0.657	0.932
	$\beta_2 = 0.1$	0.104 (0.028)	0.004	0.035	0.896	0.114 (0.050)	0.014	0.056	0.944
	$\gamma_1 = -1.2$	−1.220 (0.274)	−0.020	0.298	0.928	−1.234 (0.302)	−0.034	0.327	0.930
	$\gamma_2 = 0.05$	0.049 (0.014)	−0.001	0.024	0.740	0.055 (0.030)	0.005	0.031	0.890
	$\lambda = 0.215$	0.208 (0.019)	−0.007	0.022	0.876	0.210 (0.023)	−0.005	0.024	0.912
	$\alpha = 0.5$	0.503 (–)	0.003	0.384	–	0.660 (0.630)	0.160	0.668	0.890
	$\beta_0 = 0.9$	1.002 (0.336)	0.102	0.531	0.784	1.001 (0.651)	0.101	0.670	0.954
	$\beta_1 = -0.6$	−0.638 (0.313)	−0.038	0.321	0.964	−0.643 (0.356)	−0.043	0.383	0.932
	$\beta_2 = -0.1$	−0.106 (0.029)	−0.006	0.036	0.912	−0.106 (0.038)	−0.006	0.039	0.942
	$\gamma_1 = 1$	1.036 (0.270)	0.036	0.310	0.908	1.050 (0.278)	0.050	0.276	0.950
	$\gamma_2 = -0.2$	−0.209 (0.028)	−0.009	0.035	0.896	−0.209 (0.033)	−0.009	0.036	0.942
	$\lambda = 0.316$	0.303 (0.030)	−0.013	0.035	0.880	0.304 (0.034)	−0.012	0.037	0.902
	$\alpha = 0.5$	0.523 (–)	0.023	0.394	–	0.504 (0.655)	0.004	0.663	0.964
400	$\beta_0 = 0.2$	0.227 (0.226)	0.027	0.322	0.820	0.276 (0.367)	0.076	0.372	0.924
	$\beta_1 = -1.4$	−1.468 (0.226)	−0.068	0.390	0.728	−1.508 (0.407)	−0.108	0.398	0.954
	$\beta_2 = 0.1$	0.104 (0.020)	0.004	0.027	0.840	0.105 (0.030)	0.005	0.030	0.950
	$\gamma_1 = -1.2$	−1.229 (0.191)	−0.029	0.207	0.936	−1.222 (0.211)	−0.022	0.219	0.946
	$\gamma_2 = 0.05$	0.050 (0.010)	0.000	0.018	0.648	0.051 (0.021)	0.001	0.021	0.926
	$\lambda = 0.215$	0.210 (0.014)	−0.005	0.017	0.868	0.212 (0.016)	−0.003	0.016	0.944
	$\alpha = 0.5$	0.520 (–)	0.020	0.343	–	0.560 (0.407)	0.060	0.399	0.898
	$\beta_0 = 0.9$	0.932 (0.232)	0.032	0.409	0.680	0.938 (0.452)	0.038	0.465	0.948
	$\beta_1 = -0.6$	−0.618 (0.218)	−0.018	0.241	0.940	−0.621 (0.247)	−0.021	0.253	0.956
	$\beta_2 = -0.1$	−0.102 (0.020)	−0.002	0.025	0.884	−0.103 (0.026)	−0.003	0.027	0.946
	$\gamma_1 = 1$	0.995 (0.188)	−0.005	0.200	0.928	1.020 (0.190)	0.020	0.202	0.946
	$\gamma_2 = -0.2$	−0.203 (0.020)	−0.003	0.021	0.928	−0.206 (0.023)	−0.006	0.025	0.954
	$\lambda = 0.316$	0.311 (0.022)	−0.005	0.024	0.904	0.310 (0.024)	−0.006	0.026	0.924
	$\alpha = 0.5$	0.502 (–)	0.002	0.347	–	0.499 (0.452)	−0.001	0.463	0.964

Table 4.4. Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP) for parameter settings when $\alpha = 0.75$

		EM-PL				EM-SM			
n	Parameter	Estimate (SE)	Bias	RMSE	95% CP	Estimate (SE)	Bias	RMSE	95% CP
200	$\beta_0 = 0.2$	0.138 (0.329)	−0.062	0.412	0.848	0.363 (0.584)	0.163	0.659	0.920
	$\beta_1 = -1.4$	−1.354 (0.320)	0.046	0.438	0.804	−1.635 (0.627)	−0.235	0.673	0.958
	$\beta_2 = 0.1$	0.096 (0.028)	−0.004	0.034	0.872	0.118 (0.051)	0.018	0.056	0.950
	$\gamma_1 = -1.2$	−1.188 (0.271)	0.012	0.283	0.948	−1.251 (0.281)	−0.051	0.294	0.944
	$\gamma_2 = 0.05$	0.046 (0.013)	−0.004	0.021	0.796	0.057 (0.027)	0.007	0.030	0.886
	$\lambda = 0.215$	0.206 (0.019)	−0.009	0.024	0.852	0.212 (0.023)	−0.003	0.023	0.930
	$\alpha = 0.75$	0.604 (–)	−0.146	0.419	–	0.998 (0.734)	0.248	0.784	0.924
	$\beta_0 = 0.9$	0.863 (0.352)	−0.037	0.445	0.860	1.019 (0.692)	0.119	0.683	0.966
	$\beta_1 = -0.6$	−0.593 (0.328)	0.007	0.345	0.944	−0.615 (0.384)	−0.015	0.380	0.938
	$\beta_2 = -0.1$	−0.100 (0.030)	0.000	0.035	0.924	−0.108 (0.041)	−0.008	0.042	0.950
	$\gamma_1 = 1$	1.056 (0.274)	0.056	0.295	0.940	1.081 (0.277)	0.081	0.300	0.948
	$\gamma_2 = -0.2$	−0.213 (0.029)	−0.013	0.035	0.916	−0.208 (0.034)	−0.008	0.034	0.954
	$\lambda = 0.316$	0.301 (0.031)	−0.015	0.036	0.892	0.305 (0.034)	−0.011	0.035	0.910
	$\alpha = 0.75$	0.644 (–)	−0.106	0.385	–	0.811 (0.783)	0.061	0.745	0.982
400	$\beta_0 = 0.2$	0.157 (0.238)	−0.043	0.303	0.848	0.269 (0.399)	0.069	0.386	0.938
	$\beta_1 = -1.4$	−1.366 (0.231)	0.034	0.305	0.836	−1.503 (0.426)	−0.103	0.409	0.940
	$\beta_2 = 0.1$	0.099 (0.020)	−0.001	0.027	0.848	0.107 (0.033)	0.007	0.032	0.952
	$\gamma_1 = -1.2$	−1.213 (0.187)	−0.013	0.208	0.936	−1.209 (0.197)	−0.009	0.192	0.952
	$\gamma_2 = 0.05$	0.047 (0.009)	−0.003	0.018	0.672	0.052 (0.020)	0.002	0.019	0.932
	$\lambda = 0.215$	0.208 (0.014)	−0.007	0.016	0.900	0.212 (0.017)	−0.003	0.016	0.944
	$\alpha = 0.75$	0.672 (–)	−0.078	0.345	–	0.843 (0.514)	0.093	0.489	0.946
	$\beta_0 = 0.9$	0.840 (0.247)	−0.060	0.373	0.796	0.964 (0.470)	0.064	0.476	0.950
	$\beta_1 = -0.6$	−0.614 (0.230)	−0.014	0.252	0.932	−0.619 (0.264)	−0.019	0.271	0.940
	$\beta_2 = -0.1$	−0.097 (0.021)	0.003	0.025	0.892	−0.105 (0.028)	−0.005	0.027	0.952
	$\gamma_1 = 1$	1.021 (0.190)	0.021	0.195	0.944	1.019 (0.191)	0.019	0.188	0.954
	$\gamma_2 = -0.2$	−0.206 (0.020)	−0.006	0.023	0.916	−0.203 (0.023)	−0.003	0.024	0.956
	$\lambda = 0.316$	0.309 (0.023)	−0.007	0.025	0.920	0.310 (0.024)	−0.006	0.024	0.934
	$\alpha = 0.75$	0.648 (–)	−0.102	0.355	–	0.782 (0.522)	0.032	0.548	0.950

Table 4.5. Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP) for parameter settings when $\alpha = 1$

		EM-PL				EM-SM			
n	Parameter	Estimate (SE)	Bias	RMSE	95% CP	Estimate (SE)	Bias	RMSE	95% CP
200	$\beta_0 = 0.2$	0.118 (0.343)	−0.082	0.448	0.832	0.365 (0.625)	0.165	0.663	0.904
	$\beta_1 = -1.4$	−1.323 (0.328)	0.077	0.405	0.848	−1.602 (0.642)	−0.202	0.688	0.926
	$\beta_2 = 0.1$	0.090 (0.029)	−0.010	0.033	0.888	0.115 (0.054)	0.015	0.059	0.938
	$\gamma_1 = -1.2$	−1.209 (0.270)	−0.009	0.281	0.964	−1.228 (0.278)	−0.028	0.297	0.948
	$\gamma_2 = 0.05$	0.042 (0.013)	−0.008	0.019	0.848	0.055 (0.027)	0.005	0.031	0.874
	$\lambda = 0.215$	0.204 (0.019)	−0.001	0.024	0.876	0.210 (0.024)	−0.005	0.023	0.942
	$\alpha = 1$	0.721 (–)	−0.279	0.471	–	1.251 (0.889)	0.251	0.987	0.908
	$\beta_0 = 0.9$	0.693 (0.355)	−0.207	0.534	0.768	1.024 (0.714)	0.124	0.776	0.946
	$\beta_1 = -0.6$	−0.533 (0.332)	0.067	0.358	0.944	−0.637 (0.410)	−0.037	0.433	0.936
	$\beta_2 = -0.1$	−0.092 (0.030)	0.008	0.035	0.904	−0.106 (0.042)	−0.006	0.046	0.930
	$\gamma_1 = 1$	1.040 (0.278)	0.040	0.302	0.944	1.032 (0.277)	0.032	0.292	0.948
	$\gamma_2 = -0.2$	−0.216 (0.030)	−0.016	0.040	0.892	−0.207 (0.034)	−0.007	0.037	0.942
	$\lambda = 0.316$	0.302 (0.032)	−0.014	0.037	0.904	0.304 (0.034)	−0.012	0.035	0.908
	$\alpha = 1$	0.677 (–)	−0.323	0.513	–	1.089 (0.900)	0.089	0.962	0.966
400	$\beta_0 = 0.2$	0.094 (0.247)	−0.106	0.312	0.836	0.274 (0.426)	0.074	0.431	0.932
	$\beta_1 = -1.4$	−1.332 (0.237)	0.068	0.296	0.840	−1.485 (0.435)	−0.085	0.443	0.938
	$\beta_2 = 0.1$	0.096 (0.021)	−0.004	0.022	0.932	0.106 (0.035)	0.006	0.035	0.934
	$\gamma_1 = -1.2$	−1.200 (0.187)	0.000	0.175	0.960	−1.223 (0.192)	−0.023	0.205	0.930
	$\gamma_2 = 0.05$	0.044 (0.009)	−0.006	0.014	0.796	0.052 (0.019)	0.002	0.020	0.922
	$\lambda = 0.215$	0.208 (0.014)	−0.007	0.017	0.844	0.212 (0.017)	−0.003	0.018	0.948
	$\alpha = 1$	0.798 (–)	−0.202	0.357	–	1.097 (0.611)	0.097	0.608	0.930
	$\beta_0 = 0.9$	0.758 (0.259)	−0.142	0.374	0.808	0.968 (0.490)	0.068	0.487	0.962
	$\beta_1 = -0.6$	−0.582 (0.239)	0.018	0.244	0.940	−0.615 (0.280)	−0.015	0.283	0.932
	$\beta_2 = -0.1$	−0.095 (0.022)	0.005	0.024	0.924	−0.105 (0.029)	−0.005	0.029	0.962
	$\gamma_1 = 1$	1.019 (0.194)	0.019	0.207	0.944	1.020 (0.192)	0.020	0.200	0.940
	$\gamma_2 = -0.2$	−0.208 (0.021)	−0.008	0.023	0.928	−0.203 (0.024)	−0.003	0.024	0.962
	$\lambda = 0.316$	0.309 (0.023)	−0.007	0.025	0.904	0.310 (0.025)	−0.006	0.026	0.936
	$\alpha = 1$	0.772 (–)	−0.228	0.386	–	1.041 (0.596)	0.041	0.571	0.970

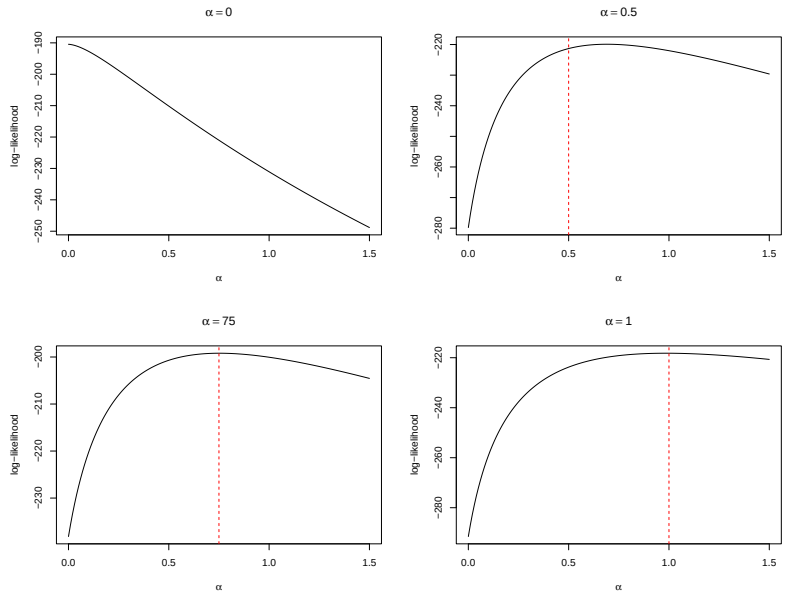

Figure 4.1. Profile likelihood plot for the transformation parameter α, $\boldsymbol{\psi} = \boldsymbol{\psi_1}$.

of the observed log-likelihood function using the "optim()" function in R. For this comparison, we use two different methods for selecting initial parameter estimates:

- $\text{IV} = \text{IV}_1$: For a given parameter Γ, take 20% deviation off its true value and create the interval $(\Gamma - 0.2|\Gamma|, \Gamma + 0.2|\Gamma|)$. Then, randomly sample a value from this interval as the initial value.
- $\text{IV} = \text{IV}_2$: For a given parameter Γ, ensure a $50\% - 75\%$ deviation off its true value by creating intervals $(\Gamma - 0.75|\Gamma|, \Gamma - 0.5|\Gamma|)$ and $(\Gamma + 0.5|\Gamma|, \Gamma + 0.75|\Gamma|)$ then randomly sampling a value from the union of these intervals as the initial value.

We restrict the selection of initial value of α to values that fall in the interval of interest, [0,1], by sampling from the intersection of the interval of interest and the intervals constructed using one of the methods described above.

4.4.3.1 Simultaneous maximization of all parameters

To address the value of the proposed estimation method, we compare the preferred implementation of the EM algorithm, EM-SM, with the direct maximization of the observed log-likelihood function using the "optim()" function in R to maximize all parameters simultaneously (DM-SM).

Tables 4.6, 4.7, 4.8, and 4.9 present the simulation results using both EM-SM and DM-SM techniques for all considered parameter settings and $n = 200$. We first note that in all settings, both estimation methods perform well with regard to biases, SEs, RMSEs, and coverage probabilities of estimates, with corresponding statistics produced by both methods differing by only a matter of hundredths in the vast majority of instances. Both approaches produce empirical coverage probabilities that are close to the nominal level in all settings. Even when the initial values deviate significantly from the true parameter values, as is the case when IV = IV_2, the estimates retain a comparable level of accuracy and efficiency to estimates produced using initial values that are closer to the true parameter values.

To further discriminate between the performance of the two methods, we compare the maximized log-likelihood values. For each parameter setting, Table 4.10 presents the average maximized log-likelihood values ($\widehat{l}_{\text{obs}}$) as well as the proportion of runs in which a given estimation method produced the greater maximized log-likelihood value ($P_{\max}$). We note that this comparison can be made equitably since, for each parameter setting, the same 250 datasets and initial values were used in applying both estimation methods. It can be seen that for all parameter settings, the value of $\widehat{l}_{\text{obs}}$ produced using the EM algorithm is greater than the corresponding value using DM-produced estimates. Further, for all parameter settings, the EM algorithm produced a greater maximized observed log-likelihood at least 98% of the

time. Because the EM algorithm produces comparable estimates with regards to accuracy and efficiency and consistently produces a greater value of the maximized log-likelihood function as compared to direct optimization, we prefer the proposed EM algorithm for the BCT cure model with interval-censored data.

4.4.3.2 Profile likelihood estimation of α

While the EM-SM estimation method is demonstrated to perform well in the specified parameter settings, we acknowledge that the true parameter settings encountered in real data can assume many forms. Because flatness of the likelihood with respect to transformation parameter α may cause difficulties implementing estimation algorithms using simultaneous maximization of all parameters, it is valuable to examine the performance of estimation methods which use profile likelihood to estimate α. Consequently, we compare the performance of the proposed EM algorithm using profile likelihood, EM-PL, to direct maximization of the observed log-likelihood function with a profile likelihood estimation of α (DM-PL) using the "optim()" function in R.

Tables 4.11, 4.12, 4.13, and 4.14 present the simulation results using both EM-PL and DM-PL techniques for all considered parameter settings and $n = 200$. In all settings, both estimation methods perform well with regard to biases, SEs, and RMSEs, with corresponding statistics produced by both methods differing by a matter of hundredths in all instances. The empirical coverage probabilities are below the nominal level in most cases, likely due to the underestimation of SEs and RMSEs resulting from treating α as fixed. Even when the initial values deviate more drastically from the true parameter values, as is the case when IV = IV_2, the estimates retain a comparable level of accuracy and efficiency to estimates produced using initial values that are closer to the true parameter values.

Table 4.6. Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP) for $\boldsymbol{\psi} = \boldsymbol{\psi_1}$ and IV = IV_1

	EM-SM				DM-SM			
Parameter	Estimate (SE)	Bias	RMSE	95% CP	Estimate (SE)	Bias	RMSE	95% CP
$\beta_0 = 0.2$	0.324 (0.319)	0.124	0.388	0.968	0.320 (0.319)	0.120	0.375	0.972
$\beta_1 = -1.4$	−1.569 (0.425)	−0.169	0.506	0.960	−1.559 (0.423)	−0.159	0.481	0.968
$\beta_2 = 0.1$	0.109 (0.032)	0.009	0.037	0.952	0.109 (0.032)	0.009	0.036	0.952
$\gamma_1 = -1.2$	−1.280 (0.383)	−0.080	0.396	0.928	−1.272 (0.387)	−0.072	0.381	0.940
$\gamma_2 = 0.05$	0.058 (0.032)	0.008	0.035	0.936	0.058 (0.032)	0.008	0.034	0.948
$\lambda = 0.215$	0.211 (0.020)	−0.004	0.021	0.924	0.211 (0.020)	−0.004	0.021	0.928
$\alpha = 0$	0.083 (0.197)	0.083	0.263	0.988	0.078 (0.200)	0.078	0.256	0.988
$\beta_0 = 0.2$	0.283 (0.498)	0.083	0.555	0.880	0.279 (0.533)	0.079	0.526	0.904
$\beta_1 = -1.4$	−1.545 (0.564)	−0.145	0.638	0.936	−1.546 (0.596)	−0.146	0.603	0.960
$\beta_2 = 0.1$	0.108 (0.044)	0.008	0.048	0.924	0.109 (0.045)	0.009	0.047	0.952
$\gamma_1 = -1.2$	−1.192 (0.305)	0.008	0.307	0.940	−1.195 (0.310)	0.005	0.297	0.944
$\gamma_2 = 0.05$	0.049 (0.029)	−0.001	0.030	0.908	0.050 (0.031)	0.000	0.028	0.940
$\lambda = 0.215$	0.211 (0.023)	−0.004	0.023	0.928	0.211 (0.023)	−0.004	0.023	0.944
$\alpha = 0.5$	0.567 (0.550)	0.067	0.596	0.876	0.572 (0.600)	0.072	0.544	0.924
$\beta_0 = 0.2$	0.284 (0.554)	0.084	0.581	0.900	0.283 (0.565)	0.083	0.556	0.924
$\beta_1 = -1.4$	−1.548 (0.594)	−0.148	0.648	0.952	−1.542 (0.603)	−0.142	0.611	0.964
$\beta_2 = 0.1$	0.111 (0.048)	0.011	0.052	0.948	0.111 (0.049)	0.011	0.050	0.952
$\gamma_1 = -1.2$	−1.230 (0.287)	−0.030	0.275	0.956	−1.232 (0.287)	−0.032	0.271	0.952
$\gamma_2 = 0.05$	0.053 (0.028)	0.003	0.029	0.908	0.054 (0.028)	0.004	0.028	0.916
$\lambda = 0.215$	0.210 (0.023)	−0.005	0.023	0.944	0.210 (0.023)	−0.005	0.023	0.944
$\alpha = 0.75$	0.848 (0.700)	0.098	0.726	0.920	0.858 (0.728)	0.108	0.683	0.928
$\beta_0 = 0.2$	0.305 (0.592)	0.105	0.597	0.916	0.266 (0.604)	0.066	0.550	0.924
$\beta_1 = -1.4$	−1.537 (0.620)	−0.137	0.663	0.932	−1.499 (0.630)	−0.099	0.611	0.952
$\beta_2 = 0.1$	0.111 (0.053)	0.011	0.061	0.940	0.108 (0.053)	0.008	0.056	0.952
$\gamma_1 = -1.2$	−1.246 (0.278)	−0.046	0.293	0.956	−1.243 (0.283)	−0.043	0.292	0.948
$\gamma_2 = 0.05$	0.053 (0.027)	0.003	0.028	0.912	0.052 (0.028)	0.002	0.026	0.948
$\lambda = 0.215$	0.211 (0.024)	−0.004	0.025	0.912	0.211 (0.024)	−0.004	0.025	0.916
$\alpha = 1$	1.187 (0.870)	0.187	0.921	0.932	1.127 (0.898)	0.127	0.824	0.932

Table 4.7. Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP) for $\boldsymbol{\psi} = \boldsymbol{\psi_2}$ and IV = IV_1

	EM-SM				DM-SM			
Parameter	Estimate (SE)	Bias	RMSE	95% CP	Estimate (SE)	Bias	RMSE	95% CP
$\beta_0 = 0.9$	1.018 (0.588)	0.118	0.565	0.980	1.025 (0.592)	0.125	0.516	0.988
$\beta_1 = -0.6$	−0.651 (0.310)	−0.051	0.296	0.968	−0.652 (0.312)	−0.052	0.290	0.972
$\beta_2 = -0.1$	−0.108 (0.035)	−0.008	0.034	0.960	−0.108 (0.035)	−0.008	0.032	0.984
$\gamma_1 = 1$	1.069 (0.281)	0.069	0.319	0.932	1.063 (0.280)	0.063	0.307	0.936
$\gamma_2 = -0.2$	−0.207 (0.031)	−0.007	0.034	0.940	−0.206 (0.031)	−0.006	0.033	0.936
$\lambda = 0.316$	0.306 (0.033)	−0.010	0.035	0.908	0.306 (0.033)	−0.010	0.035	0.904
$\alpha = 0$	0.023 (0.435)	0.023	0.402	0.980	0.042 (0.435)	0.042	0.361	0.984
$\beta_0 = 0.9$	1.010 (0.662)	0.110	0.660	0.952	0.993 (0.662)	0.093	0.591	0.956
$\beta_1 = -0.6$	−0.658 (0.362)	−0.058	0.367	0.928	−0.651 (0.361)	−0.051	0.350	0.944
$\beta_2 = -0.1$	−0.106 (0.038)	−0.006	0.038	0.940	−0.105 (0.038)	−0.005	0.036	0.952
$\gamma_1 = 1$	1.034 (0.277)	0.034	0.282	0.944	1.034 (0.276)	0.034	0.282	0.944
$\gamma_2 = -0.2$	−0.205 (0.033)	−0.005	0.032	0.968	−0.205 (0.033)	−0.005	0.031	0.964
$\lambda = 0.316$	0.308 (0.034)	−0.008	0.032	0.960	0.308 (0.034)	−0.008	0.032	0.960
$\alpha = 0.5$	0.528 (0.669)	0.028	0.687	0.988	0.528 (0.666)	0.028	0.600	0.988
$\beta_0 = 0.9$	1.019 (0.693)	0.119	0.765	0.944	1.016 (0.693)	0.116	0.671	0.968
$\beta_1 = -0.6$	−0.611 (0.386)	−0.011	0.409	0.944	−0.606 (0.388)	−0.006	0.394	0.968
$\beta_2 = -0.1$	−0.109 (0.041)	−0.009	0.047	0.924	−0.109 (0.041)	−0.009	0.044	0.944
$\gamma_1 = 1$	1.078 (0.280)	0.078	0.304	0.936	1.067 (0.279)	0.067	0.295	0.936
$\gamma_2 = -0.2$	−0.207 (0.034)	−0.007	0.034	0.956	−0.206 (0.034)	−0.006	0.032	0.952
$\lambda = 0.316$	0.305 (0.034)	−0.011	0.034	0.956	0.306 (0.035)	−0.010	0.033	0.956
$\alpha = 0.75$	0.828 (0.802)	0.078	0.849	0.964	0.861 (0.791)	0.111	0.716	0.976
$\beta_0 = 0.9$	1.022 (0.716)	0.122	0.768	0.956	1.016 (0.718)	0.116	0.691	0.976
$\beta_1 = -0.6$	−0.651 (0.411)	−0.051	0.410	0.972	−0.655 (0.411)	−0.055	0.401	0.960
$\beta_2 = -0.1$	−0.108 (0.042)	−0.008	0.046	0.932	−0.107 (0.042)	−0.007	0.043	0.956
$\gamma_1 = 1$	1.056 (0.282)	0.056	0.326	0.928	1.051 (0.281)	0.051	0.323	0.936
$\gamma_2 = -0.2$	−0.207 (0.035)	−0.007	0.037	0.964	−0.206 (0.034)	−0.006	0.037	0.968
$\lambda = 0.316$	0.309 (0.035)	−0.007	0.036	0.920	0.310 (0.035)	−0.006	0.037	0.932
$\alpha = 1$	1.034 (0.884)	0.034	0.916	0.956	1.037 (0.881)	0.037	0.778	0.972

Table 4.8. Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP) for $\boldsymbol{\psi} = \boldsymbol{\psi_1}$ and IV = IV_2

	EM-SM				DM-SM			
Parameter	Estimate (SE)	Bias	RMSE	95% CP	Estimate (SE)	Bias	RMSE	95% CP
$\beta_0 = 0.2$	0.359 (0.337)	0.159	0.488	0.956	0.373 (0.373)	0.173	0.476	0.964
$\beta_1 = -1.4$	−1.627 (0.445)	−0.227	0.568	0.976	−1.648 (0.475)	−0.248	0.541	0.980
$\beta_2 = 0.1$	0.110 (0.033)	0.010	0.034	0.992	0.112 (0.034)	0.012	0.035	0.972
$\gamma_1 = -1.2$	−1.222 (0.382)	−0.022	0.389	0.932	−1.210 (0.396)	−0.010	0.394	0.932
$\gamma_2 = 0.05$	0.058 (0.032)	0.008	0.035	0.936	0.058 (0.034)	0.008	0.037	0.908
$\lambda = 0.215$	0.212 (0.021)	−0.003	0.020	0.940	0.214 (0.021)	−0.001	0.021	0.956
$\alpha = 0$	0.103 (0.203)	0.103	0.318	0.984	0.121 (0.237)	0.121	0.313	0.984
$\beta_0 = 0.2$	0.103 (0.203)	0.103	0.318	0.984	0.121 (0.237)	0.121	0.313	0.984
$\beta_1 = -1.4$	−1.541 (0.574)	−0.141	0.580	0.944	−1.596 (0.635)	−0.196	0.569	0.976
$\beta_2 = 0.1$	0.114 (0.046)	0.014	0.050	0.960	0.118 (0.054)	0.018	0.052	0.972
$\gamma_1 = -1.2$	−1.234 (0.299)	−0.034	0.296	0.976	−1.245 (0.299)	−0.045	0.290	0.964
$\gamma_2 = 0.05$	0.054 (0.029)	0.004	0.030	0.892	0.058 (0.031)	0.008	0.029	0.916
$\lambda = 0.215$	0.208 (0.022)	−0.007	0.023	0.928	0.211 (0.024)	−0.004	0.022	0.932
$\alpha = 0.5$	0.641 (0.583)	0.141	0.623	0.908	0.739 (0.699)	0.239	0.642	0.952
$\beta_0 = 0.2$	0.291 (0.554)	0.091	0.542	0.936	0.343 (0.607)	0.143	0.524	0.952
$\beta_1 = -1.4$	−1.511 (0.586)	−0.111	0.595	0.940	−1.559 (0.628)	−0.159	0.583	0.976
$\beta_2 = 0.1$	0.108 (0.047)	0.008	0.049	0.952	0.112 (0.051)	0.012	0.049	0.960
$\gamma_1 = -1.2$	−1.220 (0.283)	−0.020	0.309	0.928	−1.238 (0.281)	−0.038	0.303	0.924
$\gamma_2 = 0.05$	0.052 (0.027)	0.002	0.028	0.932	0.056 (0.028)	0.006	0.028	0.932
$\lambda = 0.215$	0.208 (0.023)	−0.007	0.025	0.908	0.211 (0.024)	−0.004	0.025	0.932
$\alpha = 0.75$	0.872 (0.709)	0.122	0.783	0.892	0.988 (0.801)	0.238	0.769	0.932
$\beta_0 = 0.2$	0.301 (0.584)	0.101	0.670	0.908	0.235 (0.611)	0.035	0.624	0.908
$\beta_1 = -1.4$	−1.553 (0.614)	−0.153	0.753	0.884	−1.492 (0.631)	−0.092	0.714	0.900
$\beta_2 = 0.1$	0.109 (0.051)	0.009	0.058	0.936	0.106 (0.054)	0.006	0.053	0.940
$\gamma_1 = -1.2$	−1.219 (0.284)	−0.019	0.296	0.948	−1.216 (0.291)	−0.016	0.284	0.960
$\gamma_2 = 0.05$	0.051 (0.027)	0.001	0.030	0.900	0.049 (0.030)	−0.001	0.028	0.920
$\lambda = 0.215$	0.206 (0.023)	−0.009	0.027	0.900	0.206 (0.025)	−0.009	0.025	0.920
$\alpha = 1$	1.079 (0.834)	0.079	0.971	0.904	1.010 (0.955)	0.010	0.853	0.928

Table 4.9. Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP) for $\boldsymbol{\psi} = \boldsymbol{\psi_2}$ and IV = IV_2

	EM-SM				DM-SM			
Parameter	Estimate (SE)	Bias	RMSE	95% CP	Estimate (SE)	Bias	RMSE	95% CP
$\beta_0 = 0.9$	0.967 (0.581)	0.067	0.579	0.980	0.989 (0.588)	0.089	0.557	0.984
$\beta_1 = -0.6$	−0.631 (0.305)	−0.031	0.299	0.948	−0.631 (0.306)	−0.031	0.295	0.940
$\beta_2 = -0.1$	−0.105 (0.035)	−0.005	0.034	0.952	−0.106 (0.035)	−0.006	0.034	0.960
$\gamma_1 = 1$	1.039 (0.283)	0.039	0.291	0.948	1.041 (0.281)	0.041	0.301	0.944
$\gamma_2 = -0.2$	−0.209 (0.032)	−0.009	0.031	0.960	−0.207 (0.032)	−0.007	0.033	0.952
$\lambda = 0.316$	0.305 (0.033)	−0.011	0.034	0.920	0.307 (0.033)	−0.009	0.036	0.904
$\alpha = 0$	0.021 (0.447)	0.021	0.408	0.972	0.053 (0.460)	0.053	0.392	0.976
$\beta_0 = 0.9$	0.963 (0.660)	0.063	0.659	0.960	1.021 (0.677)	0.121	0.588	0.972
$\beta_1 = -0.6$	−0.648 (0.362)	−0.048	0.356	0.956	−0.655 (0.371)	−0.055	0.355	0.948
$\beta_2 = -0.1$	−0.104 (0.038)	−0.004	0.040	0.932	−0.106 (0.039)	−0.006	0.038	0.944
$\gamma_1 = 1$	1.046 (0.279)	0.046	0.269	0.976	1.043 (0.277)	0.043	0.274	0.964
$\gamma_2 = -0.2$	−0.207 (0.034)	−0.007	0.033	0.972	−0.204 (0.033)	−0.004	0.031	0.976
$\lambda = 0.316$	0.306 (0.034)	−0.010	0.033	0.948	0.308 (0.034)	−0.008	0.032	0.952
$\alpha = 0.5$	0.522 (0.690)	0.022	0.720	0.980	0.622 (0.709)	0.122	0.616	0.992
$\beta_0 = 0.9$	1.022 (0.678)	0.122	0.705	0.956	1.070 (0.691)	0.170	0.624	0.972
$\beta_1 = -0.6$	−0.625 (0.384)	−0.025	0.426	0.936	−0.624 (0.395)	−0.024	0.425	0.936
$\beta_2 = -0.1$	−0.107 (0.040)	−0.007	0.042	0.940	−0.109 (0.041)	−0.009	0.039	0.968
$\gamma_1 = 1$	1.020 (0.277)	0.020	0.296	0.940	1.021 (0.276)	0.021	0.296	0.944
$\gamma_2 = -0.2$	−0.204 (0.033)	−0.004	0.033	0.948	−0.201 (0.033)	−0.001	0.032	0.944
$\lambda = 0.316$	0.307 (0.034)	−0.009	0.039	0.900	0.309 (0.034)	−0.007	0.038	0.916
$\alpha = 0.75$	0.837 (0.766)	0.087	0.814	0.960	0.931 (0.787)	0.181	0.684	0.972
$\beta_0 = 0.9$	1.005 (0.720)	0.105	0.761	0.948	0.865 (0.716)	−0.035	0.594	0.992
$\beta_1 = -0.6$	−0.641 (0.408)	−0.041	0.397	0.960	−0.616 (0.398)	−0.016	0.380	0.960
$\beta_2 = -0.1$	−0.109 (0.043)	−0.009	0.042	0.956	−0.102 (0.042)	−0.002	0.036	0.960
$\gamma_1 = 1$	1.060 (0.283)	0.060	0.294	0.944	1.044 (0.285)	0.044	0.295	0.944
$\gamma_2 = -0.2$	−0.208 (0.035)	−0.008	0.037	0.932	−0.210 (0.035)	−0.010	0.037	0.944
$\lambda = 0.316$	0.307 (0.035)	−0.009	0.038	0.892	0.306 (0.035)	−0.010	0.038	0.900
$\alpha = 1$	1.016 (0.892)	0.016	0.982	0.964	0.873 (0.898)	−0.127	0.687	0.984

Table 4.10. Comparison of maximized observed log-likelihood values

		IV = IV$_1$				IV = IV$_2$			
		EM-SM		DM-SM		EM-SM		DM-SM	
		$\widehat{l}_{\text{obs}}$	$P_{\max}$	$\widehat{l}_{\text{obs}}$	$P_{\max}$	$\widehat{l}_{\text{obs}}$	$P_{\max}$	$\widehat{l}_{\text{obs}}$	$P_{\max}$
$\psi = \psi_1$	$\alpha = 0$	-197.60	0.98	-197.67	0.02	-196.14	0.99	-196.57	0.01
	$\alpha = 0.5$	-214.56	0.99	-214.64	0.01	-212.44	1	-212.80	0
	$\alpha = 0.75$	-213.99	1	-214.06	0	-213.73	1	-214.09	0
	$\alpha = 1$	-214.88	0.98	-214.97	0.02	-212.55	1	-213.03	0
$\psi = \psi_2$	$\alpha = 0$	-224.68	0.98	-224.77	0.02	-226.18	0.99	-226.62	0.01
	$\alpha = 0.5$	-216.63	0.98	-216.74	0.02	-215.20	1	-215.63	0
	$\alpha = 0.75$	-209.60	0.99	-209.74	0.01	-211.26	1	-211.60	0
	$\alpha = 1$	-204.38	0.98	-204.52	0.02	-203.34	0.98	-203.71	0.02

We further compare the performance of the two methods 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 4.15 presents the average maximized log-likelihood values ($\widehat{l}_{\text{obs}}$) as well as the proportion of runs in which a given estimation method produced the greater maximized log-likelihood value ($P_{\max}$). In all cases, the value of $\widehat{l}_{\text{obs}}$ produced using the EM-PL method is greater than the corresponding value using DM-PL-produced estimates. The EM-PL technique also produced a greater maximized observed log-likelihood in all settings, with the difference in proportions $P_{\max}$ being more pronounced when the initial values deviate more from the true parameter values. The EM-PL technique produced comparable estimates with regards to accuracy and efficiency and consistently produces a greater value of the maximized log-likelihood function as compared to the DM-SM technique. We note that the EM-SM method is the preferred method

according to the analysis in Section 4.4.2, however if computational difficulties arise in implementation, the EM-PL method is shown to be a viable option.

4.5 Illustration using data from a smoking cessation study

In this section, we demonstrate the performance of the preferred EM algorithm (EM-SM) using a real data on smoking cessation. The study consists of 223 subjects who attempted to quit smoking at least once during the study period of November 1986 to February 1989. A full description of the data may be found in Murray et al. [62]. Subjects were randomly assigned to either a smoking intervention (SI) group (treatment group) or a usual care (UC) group (control group), which received no intervention, at time of enrollment. Annual monitoring was conducted for a follow-up period of 5 consecutive years, with the event of interest being whether subjects resume smoking or not (relapse). The data consists of 65 (45 for SI and 20 for UC) out of 223 (169 for SI and 54 for UC) subjects experiencing relapse. Available covariates for each subject are gender (GEN, 1:Female and 0:Male), treatment group, years smoking (DUR), and average number of cigarettes smoked daily prior to attempting to quit smoking. To mimic the covariate configuration in Section 4.2 we consider GEN (x_1) and DUR (x_2) as covariates of interest. Kaplan-Meier curves stratified by GEN, as shown in Figure 2.1, level off to non-zero proportions which supports the presence of a cure component in the data. While the stratified Kaplan-Meier curves intersect in the initial stage of the study ($t < 1$), the shapes of the curves convey a similar relationship between gender and long term survival.

To apply the algorithm, we first select initial values for the parameters by using a grid-search method and selecting the set of values which maximize the log-likelihood function as initial values. Since the computational time of a grid search grows exponentially with the inclusion of additional parameters, we fix values of $\alpha = 0.5$

Table 4.11. Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP) for $\boldsymbol{\psi} = \boldsymbol{\psi_1}$ and IV = IV_1

	EM-PL				DM-PL			
Parameter	Estimate (SE)	Bias	RMSE	95% CP	Estimate (SE)	Bias	RMSE	95% CP
$\beta_0 =$ 0.2	0.311 (0.250)	0.111	0.445	0.896	0.310 (0.250)	0.110	0.435	0.896
$\beta_1 =$ −1.4	−1.645 (0.302)	−0.245	0.561	0.852	−1.645 (0.303)	−0.245	0.553	0.848
$\beta_2 =$ 0.1	0.116 (0.026)	0.016	0.035	0.932	0.116 (0.026)	0.016	0.034	0.928
$\gamma_1 =$ −1.2	−1.294 (0.340)	−0.094	0.393	0.920	−1.296 (0.339)	−0.096	0.390	0.920
$\gamma_2 =$ 0.05	0.059 (0.023)	0.009	0.033	0.816	0.060 (0.023)	0.010	0.033	0.824
$\lambda =$ 0.215	0.212 (0.019)	−0.003	0.020	0.932	0.212 (0.019)	−0.003	0.020	0.932
$\alpha =$ 0	0.112 (–)	0.112	0.248	–	0.112 (–)	0.112	0.240	–
$\beta_0 =$ 0.2	0.247 (0.319)	0.047	0.447	0.852	0.244 (0.319)	0.044	0.435	0.868
$\beta_1 =$ −1.4	−1.458 (0.319)	−0.058	0.482	0.792	−1.455 (0.318)	−0.055	0.468	0.800
$\beta_2 =$ 0.1	0.101 (0.028)	0.001	0.036	0.892	0.101 (0.028)	0.001	0.036	0.900
$\gamma_1 =$ −1.2	−1.215 (0.275)	−0.015	0.312	0.924	−1.212 (0.275)	−0.012	0.308	0.936
$\gamma_2 =$ 0.05	0.049 (0.014)	−0.001	0.025	0.716	0.049 (0.014)	−0.001	0.024	0.724
$\lambda =$ 0.215	0.210 (0.019)	−0.005	0.022	0.888	0.210 (0.019)	−0.005	0.022	0.892
$\alpha =$ 0.5	0.487 (–)	−0.013	0.396	–	0.487 (–)	−0.013	0.385	–
$\beta_0 =$ 0.2	0.117 (0.331)	−0.083	0.437	0.832	0.119 (0.332)	−0.081	0.430	0.836
$\beta_1 =$ −1.4	−1.377 (0.323)	0.023	0.435	0.832	−1.379 (0.323)	0.021	0.428	0.836
$\beta_2 =$ 0.1	0.100 (0.029)	0.000	0.037	0.840	0.100 (0.029)	0.000	0.037	0.840
$\gamma_1 =$ −1.2	−1.211 (0.272)	−0.011	0.305	0.932	−1.208 (0.272)	−0.008	0.306	0.928
$\gamma_2 =$ 0.05	0.045 (0.013)	−0.005	0.022	0.820	0.045 (0.013)	−0.005	0.021	0.828
$\lambda =$ 0.215	0.206 (0.019)	−0.009	0.024	0.868	0.206 (0.019)	−0.009	0.023	0.872
$\alpha =$ 0.75	0.612 (–)	−0.138	0.416	–	0.616 (–)	−0.134	0.403	–
$\beta_0 =$ 0.2	0.021 (0.339)	−0.179	0.431	0.840	0.024 (0.339)	−0.176	0.428	0.852
$\beta_1 =$ −1.4	−1.258 (0.323)	0.142	0.391	0.848	−1.258 (0.323)	0.142	0.386	0.860
$\beta_2 =$ 0.1	0.092 (0.029)	−0.008	0.034	0.884	0.092 (0.029)	−0.008	0.033	0.888
$\gamma_1 =$ −1.2	−1.193 (0.271)	0.007	0.278	0.960	−1.193 (0.271)	0.007	0.275	0.960
$\gamma_2 =$ 0.05	0.041 (0.013)	−0.009	0.022	0.768	0.042 (0.013)	−0.008	0.021	0.768
$\lambda =$ 0.215	0.207 (0.020)	−0.008	0.022	0.900	0.207 (0.020)	−0.008	0.022	0.900
$\alpha =$ 1	0.700 (–)	−0.300	0.496	–	0.705 (–)	−0.295	0.478	–

Table 4.12. Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP) for $\boldsymbol{\psi} = \boldsymbol{\psi_2}$ and IV = IV_1

	EM-PL				DM-PL			
Parameter	Estimate (SE)	Bias	RMSE	95% CP	Estimate (SE)	Bias	RMSE	95% CP
$\beta_0 = 0.9$	1.156 (0.275)	0.256	0.507	0.792	1.157 (0.275)	0.257	0.506	0.796
$\beta_1 = -0.6$	−0.689 (0.268)	−0.089	0.306	0.944	−0.689 (0.268)	−0.089	0.308	0.940
$\beta_2 = -0.1$	−0.113 (0.025)	−0.013	0.033	0.924	−0.113 (0.025)	−0.013	0.033	0.920
$\gamma_1 = 1$	1.024 (0.271)	0.024	0.294	0.940	1.026 (0.271)	0.026	0.296	0.940
$\gamma_2 = -0.2$	−0.198 (0.026)	0.002	0.028	0.952	−0.199 (0.026)	0.001	0.028	0.944
$\lambda = 0.316$	0.315 (0.031)	−0.001	0.032	0.940	0.315 (0.030)	−0.001	0.032	0.940
$\alpha = 0$	0.172 (–)	0.172	0.297	–	0.173 (–)	0.173	0.293	–
$\beta_0 = 0.9$	1.052 (0.343)	0.152	0.571	0.792	1.041 (0.341)	0.141	0.557	0.804
$\beta_1 = -0.6$	−0.677 (0.317)	−0.077	0.374	0.928	−0.672 (0.316)	−0.072	0.368	0.928
$\beta_2 = -0.1$	−0.109 (0.029)	−0.009	0.036	0.904	−0.109 (0.029)	−0.009	0.036	0.904
$\gamma_1 = 1$	1.068 (0.273)	0.068	0.314	0.936	1.071 (0.273)	0.071	0.317	0.936
$\gamma_2 = -0.2$	−0.207 (0.028)	−0.007	0.031	0.936	−0.208 (0.028)	−0.008	0.032	0.928
$\lambda = 0.316$	0.306 (0.031)	−0.010	0.033	0.924	0.306 (0.031)	−0.010	0.033	0.920
$\alpha = 0.5$	0.541 (–)	0.041	0.392	–	0.533 (–)	0.033	0.379	–
$\beta_0 = 0.9$	0.853 (0.348)	−0.047	0.463	0.836	0.850 (0.348)	−0.050	0.454	0.860
$\beta_1 = -0.6$	−0.620 (0.325)	−0.020	0.354	0.928	−0.618 (0.325)	−0.018	0.352	0.924
$\beta_2 = -0.1$	−0.099 (0.030)	0.001	0.032	0.924	−0.099 (0.030)	0.001	0.032	0.920
$\gamma_1 = 1$	1.051 (0.274)	0.051	0.279	0.952	1.053 (0.274)	0.053	0.277	0.948
$\gamma_2 = -0.2$	−0.207 (0.029)	−0.007	0.033	0.932	−0.207 (0.029)	−0.007	0.033	0.928
$\lambda = 0.316$	0.308 (0.032)	−0.008	0.033	0.920	0.308 (0.032)	−0.008	0.034	0.912
$\alpha = 0.75$	0.623 (–)	−0.127	0.419	–	0.623 (–)	−0.127	0.409	–
$\beta_0 = 0.9$	0.801 (0.364)	−0.099	0.479	0.828	0.796 (0.363)	−0.104	0.471	0.844
$\beta_1 = -0.6$	−0.569 (0.338)	0.031	0.365	0.932	−0.567 (0.337)	0.033	0.361	0.932
$\beta_2 = -0.1$	−0.099 (0.031)	0.001	0.035	0.920	−0.099 (0.031)	0.001	0.035	0.912
$\gamma_1 = 1$	1.011 (0.277)	0.011	0.304	0.952	1.011 (0.277)	0.011	0.306	0.952
$\gamma_2 = -0.2$	−0.210 (0.030)	−0.010	0.034	0.928	−0.210 (0.030)	−0.010	0.034	0.932
$\lambda = 0.316$	0.305 (0.032)	−0.011	0.035	0.920	0.305 (0.032)	−0.011	0.035	0.912
$\alpha = 1$	0.738 (–)	−0.262	0.448	–	0.732 (–)	−0.268	0.438	–

Table 4.13. Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP) for $\boldsymbol{\psi} = \boldsymbol{\psi_1}$ and IV = IV_2

	EM-PL				DM-PL			
Parameter	Estimate (SE)	Bias	RMSE	95% CP	Estimate (SE)	Bias	RMSE	95% CP
$\beta_0 = 0.2$	0.342 (0.266)	0.142	0.544	0.900	0.381 (0.268)	0.181	0.537	0.896
$\beta_1 = -1.4$	−1.652 (0.316)	−0.252	0.574	0.872	−1.692 (0.318)	−0.292	0.587	0.872
$\beta_2 = 0.1$	0.115 (0.026)	0.015	0.033	0.920	0.117 (0.027)	0.017	0.035	0.912
$\gamma_1 = -1.2$	−1.321 (0.343)	−0.121	0.410	0.872	−1.350 (0.335)	−0.150	0.433	0.848
$\gamma_2 = 0.05$	0.062 (0.023)	0.012	0.035	0.808	0.066 (0.022)	0.016	0.039	0.740
$\lambda = 0.215$	0.212 (0.019)	−0.003	0.022	0.912	0.214 (0.019)	−0.001	0.021	0.920
$\alpha = 0$	0.107 (–)	0.107	0.240	–	0.142 (–)	0.142	0.264	–
$\beta_0 = 0.2$	0.183 (0.313)	−0.017	0.398	0.872	0.179 (0.313)	−0.021	0.390	0.864
$\beta_1 = -1.4$	−1.435 (0.315)	−0.035	0.415	0.892	−1.437 (0.316)	−0.037	0.400	0.908
$\beta_2 = 0.1$	0.105 (0.028)	0.005	0.036	0.876	0.105 (0.028)	0.005	0.036	0.900
$\gamma_1 = -1.2$	−1.211 (0.275)	−0.011	0.338	0.900	−1.211 (0.274)	−0.011	0.335	0.908
$\gamma_2 = 0.05$	0.049 (0.014)	−0.001	0.026	0.724	0.049 (0.014)	−0.001	0.025	0.732
$\lambda = 0.215$	0.209 (0.019)	−0.006	0.021	0.904	0.210 (0.019)	−0.005	0.021	0.908
$\alpha = 0.5$	0.484 (–)	−0.016	0.380	–	0.487 (–)	−0.013	0.351	–
$\beta_0 = 0.2$	0.206 (0.311)	0.006	0.404	0.872	0.207 (0.313)	0.007	0.398	0.856
$\beta_1 = -1.4$	−1.404 (0.311)	−0.004	0.429	0.832	−1.411 (0.313)	−0.011	0.423	0.840
$\beta_2 = 0.1$	0.100 (0.027)	0.000	0.035	0.856	0.101 (0.028)	0.001	0.035	0.880
$\gamma_1 = -1.2$	−1.203 (0.275)	−0.003	0.311	0.940	−1.210 (0.274)	−0.010	0.307	0.936
$\gamma_2 = 0.05$	0.047 (0.014)	−0.003	0.024	0.748	0.048 (0.014)	−0.002	0.023	0.764
$\lambda = 0.215$	0.208 (0.019)	−0.007	0.022	0.900	0.209 (0.019)	−0.006	0.021	0.900
$\alpha = 0.75$	0.457 (–)	−0.293	0.483	–	0.471 (–)	−0.279	0.451	–
$\beta_0 = 0.2$	0.026 (0.340)	−0.174	0.434	0.824	0.014 (0.336)	−0.186	0.422	0.852
$\beta_1 = -1.4$	−1.275 (0.324)	0.125	0.397	0.828	−1.259 (0.320)	0.141	0.387	0.828
$\beta_2 = 0.1$	0.094 (0.029)	−0.006	0.036	0.840	0.093 (0.029)	−0.007	0.036	0.864
$\gamma_1 = -1.2$	−1.184 (0.269)	0.016	0.276	0.940	−1.183 (0.270)	0.017	0.274	0.944
$\gamma_2 = 0.05$	0.040 (0.013)	−0.010	0.021	0.784	0.040 (0.013)	−0.010	0.020	0.788
$\lambda = 0.215$	0.204 (0.019)	−0.011	0.024	0.844	0.203 (0.019)	−0.012	0.024	0.832
$\alpha = 1$	0.702 (–)	−0.298	0.486	–	0.683 (–)	−0.317	0.464	–

Table 4.14. Estimates, standard errors (SE), bias, root mean square error (RMSE), and 95% coverage probabilities (95% CP) for $\boldsymbol{\psi} = \boldsymbol{\psi_2}$ and IV = IV_2

	EM-PL				DM-PL			
Parameter	Estimate (SE)	Bias	RMSE	95% CP	Estimate (SE)	Bias	RMSE	95% CP
$\beta_0 = 0.9$	1.145 (0.275)	0.245	0.496	0.812	1.181 (0.282)	0.281	0.508	0.788
$\beta_1 = -0.6$	−0.689 (0.268)	−0.089	0.302	0.936	−0.704 (0.272)	−0.104	0.311	0.928
$\beta_2 = -0.1$	−0.113 (0.025)	−0.013	0.034	0.868	−0.114 (0.025)	−0.014	0.034	0.868
$\gamma_1 = 1$	1.047 (0.271)	0.047	0.301	0.924	1.047 (0.270)	0.047	0.303	0.912
$\gamma_2 = -0.2$	−0.202 (0.026)	−0.002	0.028	0.944	−0.201 (0.026)	−0.001	0.028	0.924
$\lambda = 0.316$	0.313 (0.030)	−0.003	0.032	0.904	0.314 (0.030)	−0.002	0.032	0.904
$\alpha = 0$	0.175 (–)	0.175	0.307	–	0.208 (–)	0.208	0.322	–
$\beta_0 = 0.9$	0.921 (0.325)	0.021	0.472	0.812	0.917 (0.325)	0.017	0.446	0.856
$\beta_1 = -0.6$	−0.627 (0.308)	−0.027	0.361	0.916	−0.620 (0.309)	−0.020	0.352	0.924
$\beta_2 = -0.1$	−0.102 (0.028)	−0.002	0.034	0.916	−0.102 (0.029)	−0.002	0.033	0.920
$\gamma_1 = 1$	1.053 (0.276)	0.053	0.278	0.952	1.050 (0.276)	0.050	0.277	0.956
$\gamma_2 = -0.2$	−0.210 (0.028)	−0.010	0.033	0.928	−0.209 (0.028)	−0.009	0.032	0.932
$\lambda = 0.316$	0.306 (0.031)	−0.010	0.035	0.916	0.307 (0.031)	−0.009	0.035	0.920
$\alpha = 0.5$	0.471 (–)	−0.029	0.375	–	0.476 (–)	−0.024	0.352	–
$\beta_0 = 0.9$	0.892 (0.357)	−0.008	0.461	0.824	0.867 (0.352)	−0.033	0.444	0.836
$\beta_1 = -0.6$	−0.622 (0.329)	−0.022	0.354	0.932	−0.610 (0.327)	−0.010	0.344	0.936
$\beta_2 = -0.1$	−0.103 (0.031)	−0.003	0.032	0.944	−0.102 (0.030)	−0.002	0.032	0.932
$\gamma_1 = 1$	1.052 (0.276)	0.052	0.281	0.956	1.053 (0.276)	0.053	0.281	0.956
$\gamma_2 = -0.2$	−0.210 (0.029)	−0.010	0.034	0.936	−0.211 (0.029)	−0.011	0.035	0.924
$\lambda = 0.316$	0.303 (0.032)	−0.013	0.038	0.876	0.303 (0.032)	−0.013	0.038	0.876
$\alpha = 0.75$	0.657 (–)	−0.343	0.513	–	0.632 (–)	−0.368	0.505	–
$\beta_0 = 0.9$	0.769 (0.360)	−0.131	0.494	0.800	0.756 (0.357)	−0.144	0.483	0.804
$\beta_1 = -0.6$	−0.587 (0.333)	0.013	0.336	0.944	−0.585 (0.332)	0.015	0.336	0.944
$\beta_2 = -0.1$	−0.099 (0.031)	0.001	0.033	0.940	−0.098 (0.031)	0.002	0.034	0.928
$\gamma_1 = 1$	1.012 (0.277)	0.012	0.293	0.924	1.014 (0.277)	0.014	0.296	0.924
$\gamma_2 = -0.2$	−0.213 (0.030)	−0.013	0.031	0.964	−0.214 (0.030)	−0.014	0.032	0.960
$\lambda = 0.316$	0.303 (0.032)	−0.013	0.033	0.896	0.303 (0.032)	−0.013	0.034	0.908
$\alpha = 1$	0.677 (–)	0.677	0.780	–	0.661 (–)	0.661	0.744	–

Table 4.15. Comparison of maximized observed log-likelihood values

		IV = IV_1				IV = IV_2			
		EM-PL		DM-PL		EM-PL		DM-PL	
		$\widehat{l}_{\text{obs}}$	$P_{\max}$	$\widehat{l}_{\text{obs}}$	$P_{\max}$	$\widehat{l}_{\text{obs}}$	$P_{\max}$	$\widehat{l}_{\text{obs}}$	$P_{\max}$
$\boldsymbol{\psi = \psi_1}$	$\alpha = 0$	-196.94	0.70	-196.98	0.30	-195.56	0.91	-195.78	0.09
	$\alpha = 0.5$	-214.12	0.78	-214.14	0.22	-213.90	0.95	-213.98	0.05
	$\alpha = 0.75$	-214.79	0.79	-214.80	0.21	-214.17	0.94	-214.25	0.06
	$\alpha = 1$	-213.81	0.77	-213.82	0.23	-214.78	0.95	-214.86	0.05
$\boldsymbol{\psi = \psi_2}$	$\alpha = 0$	-227.70	0.75	-227.71	0.25	-226.26	0.97	-226.38	0.03
	$\alpha = 0.5$	-216.26	0.79	-216.29	0.21	-215.51	0.95	-215.59	0.05
	$\alpha = 0.75$	-211.72	0.83	-211.74	0.17	-209.51	0.97	-209.60	0.03
	$\alpha = 1$	-206.99	0.90	-207.01	0.10	-205.34	0.97	-205.42	0.03

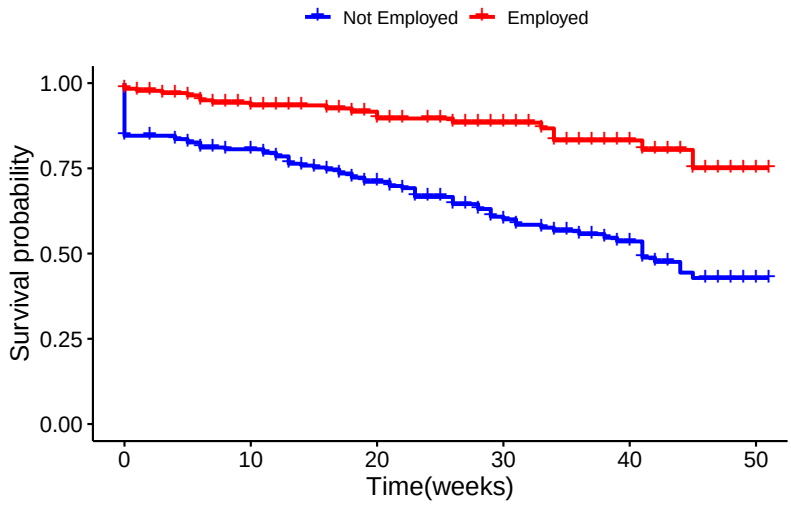

Figure 4.2. Kaplan-Meier plot of survival curves stratified by gender.

and $\lambda = 0.5$ then use a five-fold grid search to identify initial values of the remaining parameters. The grid for all parameters is chosen as $\{-1, -0.9, ...0.9, 1\}$, and the set of values which maximize the log-likelihood function is found to be interior to the boundary. We proceed with parameter estimation by applying both the EM-SM algorithm, which encompasses the general case where $\alpha \, \epsilon \, (0, 1]$ and performing the

EM-PL for only $\alpha = 0$. To discriminate between these two sets of estimates we compare the log-likelihood function values and select the set which produces the greater maximized log-likelihood function, corresponding to $\alpha = 0$, as the optimal estimates. We note that this decision is supported by a test of hypothesis [76, 84]. For the likelihood-ratio test ($H_0 : \alpha = 0$ vs $H_a : 0 < \alpha \leq 1$), the likelihood-ratio test statistic (Λ) turned out to be approximately equal to 0. Because $H_0 : \alpha = 0$ lies on the boundary of the parameter space, the null distribution of (Λ) is such that $P[\Lambda \leq \lambda] = 0.5 + 0.5P[\chi_1^2 \leq \lambda]$, where χ_1^2 denotes a chi-square variable with one degree of freedom. The resultant p-value is approximately 0.5 and hence we do not reject the assumption of $\alpha = 0$, i.e. a promotion time cure model, for the data. The subsequent analyses report parameter estimates corresponding to the preferred model when $\alpha = 0$.

Table 4.16 presents the estimates and standard errors of the parameters, as well as p-values and 95% asymptotic confidence intervals. While the positive sign of the estimates corresponding to covariate GEN, β_1 and γ_1, agree with both the stratified Kaplan-Meier plot in Figure 2.1 and previous findings observing women as more likely to relapse than men [11], the predictor of GEN fails to be significant at a 5% level of significance. The estimate of $\beta_2 > 0$ conveys a decrease in cure rate for longer term smokers, however this relation fails to be significant at a 5% level of significance. Regression parameter γ_2, relaying the effect of DUR on short-term survivors, is highly significant with $p < 0.001$. The effect of $\gamma_2 < 0$ is a decrease in survival probability associated with longer duration of smoking, and this result is consistent with previous analysis [79].

Figure 4.3 shows the predicted survival probabilities for patients with smoking durations of 18, 31, 35 and 40.9 years, which correspond to the 5^{th}, 50^{th}, 75^{th}, and 95^{th} percentiles, stratified by gender. Note that the survival probability for males is

Table 4.16. Estimation results for the smoking cessation data

Parameter	Estimate	Standard error	$p-$value
β_0	−1.136	0.599	0.059
β_1	0.153	0.305	0.616
β_2	0.010	0.024	0.672
γ_1	0.272	0.426	0.524
γ_2	−0.097	0.015	< 0.001
λ	0.484	0.060	–

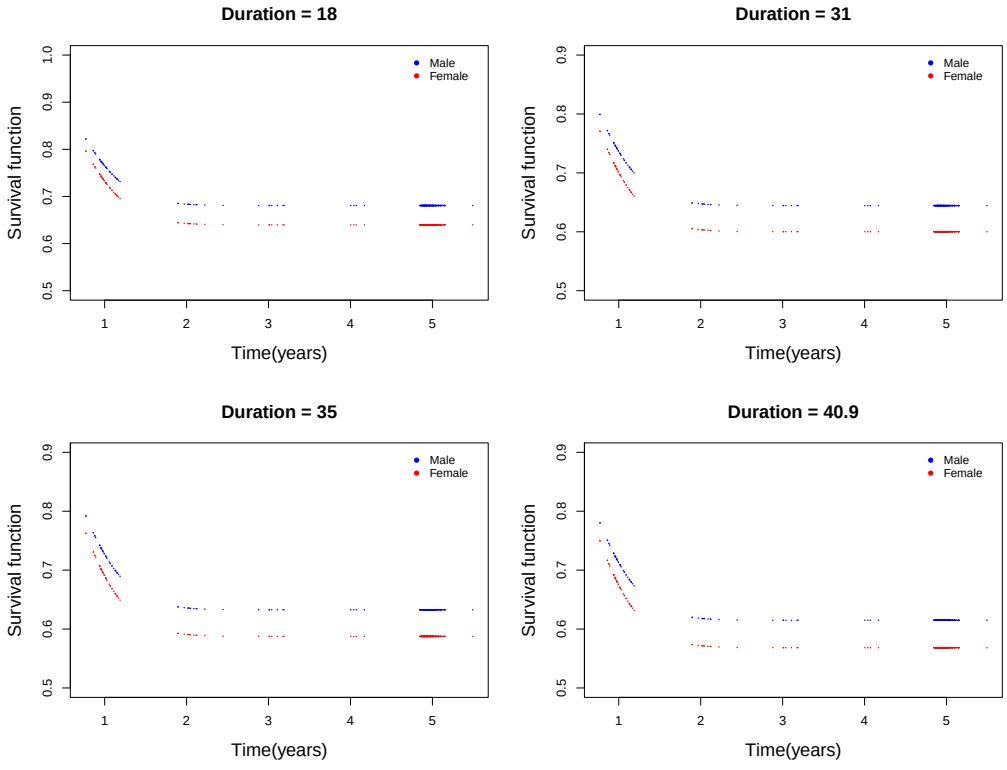

Figure 4.3. Predicted survival probabilities stratified by gender for patients with different durations of smoking.

higher across all values of duration. It may be observed that the survival probability decreases for longer smoking duration by comparing the plots fixing duration at 18 and 40.9 years of smoking. The effect of smoking duration on survival is further conveyed in Figure 4.4 where the estimated cure rate is shown to decrease in a nearly linear fashion as duration of smoking increases.

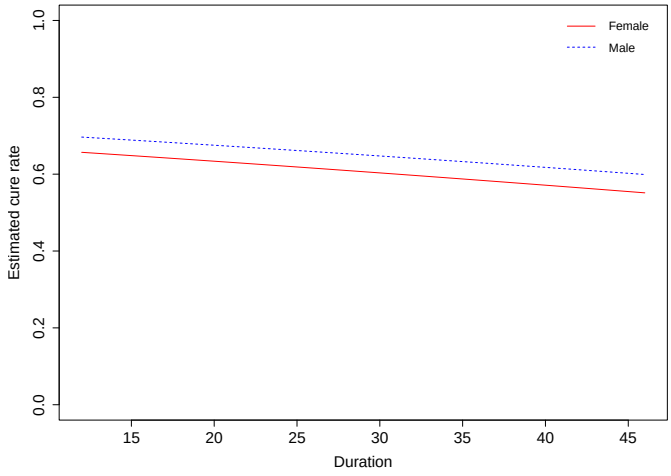

Figure 4.4. Cure rate against duration of smoking stratified by gender.

We check the adequacy of the promotion time cure model by using the calculated normalized randomized quantile residuals [32]. Figure 4.5 presents the quantile-quantile plot, where each point corresponds to the median of five sets of ordered residuals. The linearity in this plot suggests that the promotion time cure model with a Weibull baseline provides a good fit to the smoking cessation data. Finally, the Kolmogorov-Smirnov test for normality of residuals provides strong evidence for the normality of residuals, with a p-value of 0.9943.

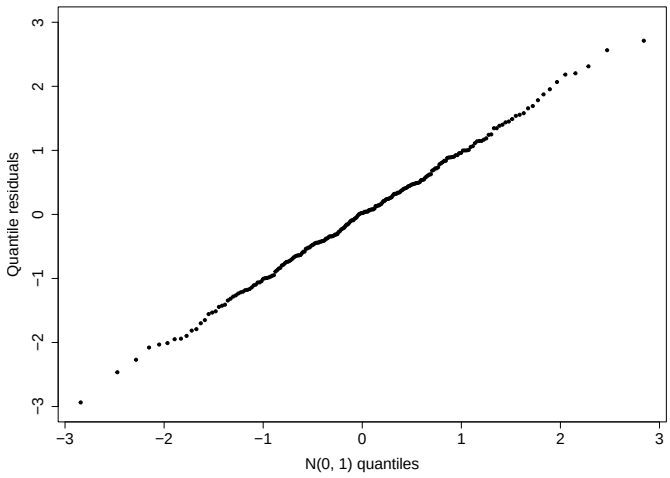

Figure 4.5. Q-Q plot of the normalized randomized quantile residuals.

CHAPTER 5

Concluding remarks

5.1 Summary of research

A primary objective in clinical research is to maximize the efficacy of medical interventions. Favorable response to treatment may result in a subject's subsequent immunity to a given disease in which case they are said to be cured, while those who do not respond favorably remain uncured. The purpose of the clinician is then to estimate the proportion of patients cured and to examine the causes for failure of treatment in the uncured patients. Estimating the cure proportion, as well as understanding the effects of prognostic factors, is important in understanding and improving trends in the survival of patients suffering from diseases such as cancer.

While the mixture cure rate model is the inaugural survival analysis model to accommodate a cure component, alternative cure rate models have been developed to encompass a wider variety of biological applications. The promotion time cure rate model facilitates modeling the presence of a latent quantity of competing risks, such as number of metastasis-competent tumor cells, and the destructive cure rate models allow for modeling the impact of a destructive process, such as chemotherapy, upon the competing risks. In many clinical applications in which a patient is not hospitalized, continuous observation is not possible. While these applications produce interval-censored data, the majority of research on cure rate models assume the data to be right-censored. In this thesis, we have extended several existing cure models to accommodate interval censoring. The main contribution of this work is the development of efficient likelihood-based inference methods for the interval-censored

cure rate models presented, noting that these methods are of broader utility in application since the form of interval-censored data encompasses both right- and left-censored data forms.

We first generalized the destructive shifted Poisson cure rate model to accommodate interval censoring and, motivated by the work of Gallardo et al. [37], developed likelihood inference based on an implementation of the EM algorithm which splits the complete log-likelihood function into simpler functions to be maximized independently. A detailed study was completed examining the performance of the EM algorithm for this model. Through Monte Carlo simulations, we demonstrated that the proposed method produces parameter estimates accurately and performs favorably when compared to direct maximization of the observed log-likelihood function with respect to both lower incidence of atypical estimates and higher rates of convergence. In the real data analysis, the proposed algorithm converged to parameter estimates that were consistent with prior findings.

We then extended the destructive negative binomial cure rate model to accommodate interval censoring. We developed likelihood inference based on two implementations of the EM algorithm which split 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 first demonstrated that the EM-SM algorithm, previously formulated for the DNB model with right-censored data but not studied through simulation prior to this work, consistently and significantly underestimates ϕ while providing no improvement in accuracy or efficiency for the parameter settings considered. We showed the proposed SEM method to perform favorably with simulated data when compared to both variations of the EM algorithm with respect to parameter recovery. In estimating parameters that are related to the cure probability (i.e., $\phi, \boldsymbol{\beta_1}, \boldsymbol{\beta_2}$), the SEM algorithm provided

greater accuracy and precision, yielding estimates with smaller biases and smaller RMSEs than the EM-PL approach. This in turn resulted in more accurate estimates of the cured probabilities. While standard errors for other parameter estimates were comparable between the SEM and EM-PL 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 showed 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 produced parameter estimates that are consistent both with prior findings and with estimates produced by the EM algorithm, and the greater precision of SEM-produced estimates allowed us to identify the statistical significance of covariate duration.

Finally, we extended the Box-Cox transformation cure rate model developed by Yin and Ibrahim [107] to interval-censored data using a semi-parametric framework for the latency. The EM algorithm for the Box-Cox transformation cure rate model was formulated in the interval-censored setting followed by an extensive simulation study examining the performance of the proposed algorithm using both simultaneous maximization of all parameters and profile likelihood to estimate transformation parameter α. The EM algorithm with simultaneous maximization, EM-SM, is shown to be the preferred method, providing comparably accurate and efficient estimates with added computational advantages. In applying the EM-SM algorithm to a real smoking-cessation data, we produced results consistent with prior analysis and illustrated the flexibility of the Box-Cox family of cure rate models in enabling us to

carry out a formal test of hypothesis to select the proper model within this family that provides the best fit for the data.

5.2 Future work

In this section, we explore some of the future research problems that could be considered as natural extensions of the work carried out in this thesis.

5.2.1 Extend alternative models

A natural continuation of the objective of this work in developing efficient estimation algorithms for cure rate models under interval censoring will be to generalize the extension of the interval-censored setting to families of cure rate models; see [48] and [61]. Within the context of destructive cure rate models, 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 [100].

5.2.2 Semi-parametric and non-parametric approaches

Exploring the performance of the proposed algorithms using semi-parametric or non-parametric modeling for the lifetimes of susceptibles will provide greater insight as to the versatility of these estimation methods. Adopting semi-parametric modeling, such as a piecewise linear approximation for the hazard functions of competing risks [5] or a piecewise exponential approximation for the progression time of each competing risk [42], would allow the avoidance of potentially inappropriate parametric assumptions.

5.2.3 Alternatives to standard model assumptions

The majority of survival analysis methods rely on the assumption of non-informative and independent censoring times, however these assumptions are not always practically appropriate. Relaxing the assumption of non-informative and independent censoring can help to reduce bias in parameter estimation. We could, for example, extend the definition of informative censoring used to construct likelihood inference for the mixture cure rate model [27] to the destructive cure rate or Box-Cox transformation cure rate model with informative censoring.

5.2.4 Novel estimation methods

The EM algorithm has been used extensively in likelihood inference for cure rate models due to the presence of latent variable(s). Recently, alternative estimation methods have been developed incorporating machine learning techniques in the context of cure models [53, 70, 79]. A non-linear conjugate gradient algorithm was developed for parameter estimation for the Box-Cox transformation cure model, and shown to perform better than the existing EM algorithm [82]. It will be of interest to apply such estimation methods in novel contexts to develop inference for models such as the interval-censored destructive cure models developed herein.

We hope to report on these problems in future works.

APPENDIX A

Appendix A

In this appendix, we present the proofs of the propositions used in Chapters 2 and 3 to construct the EM algorithms for the destructive shifted Poisson cure rate model and the EM and SEM algorithms for the destructive negative binomial cure rate model under interval censoring.

A.1 Proofs of the propositions used in Chapters 2

A.1.1 Proof of Proposition 2.2.1

Noting that $D|M = m \sim \text{Bin}(m, p)$ if $M > 0$ and $P(D = 0|M = 0) = 1$, the distribution of D may be derived as

$$\begin{aligned} P(D = d; \theta, p) &= \sum_{m=d}^{\infty} P(D|M = m)P(M = m) \\ &= \sum_{m=d}^{\infty} \binom{m}{d} p^d (1-p)^{m-d} \frac{e^{-\theta}\theta^{m-1}}{(m-1)!} \\ &= \frac{e^{-\theta}(p\theta)^d}{d!\theta} \sum_{m=0}^{\infty} \frac{m! \left[\theta(1-p)\right]^{m-d}}{(m-d)!(m-1)!} \\ &= \frac{e^{-\theta}(p\theta)^d}{d!\theta} \sum_{m=0}^{\infty} \frac{m \left[\theta(1-p)\right]^{m-d}}{(m-d)!}. \end{aligned}$$

Let $t = m - d$. Then

$$\begin{aligned} P(D = d; \theta, p) &= \frac{e^{-\theta}(p\theta)^d}{d!\theta} \left[\sum_{t=0}^{\infty} \frac{t \left[\theta(1-p)\right]^t}{t!} + d \sum_{t=0}^{\infty} \frac{\left[\theta(1-p)\right]^t}{t!} \right] \\ &= \frac{e^{-\theta}(p\theta)^d}{d!\theta} \left[\theta(1-p) \sum_{t=1}^{\infty} \frac{\left[\theta(1-p)\right]^{t-1}}{(t-1)!} + d \sum_{t=0}^{\infty} \frac{\left[\theta(1-p)\right]^t}{t!} \right] \\ &= \frac{e^{-\theta p}(\theta p)^d}{d!} \left[(1-p) + \frac{d}{\theta} \right], d = 0, 1, 2, ..., M. \end{aligned}$$

A.1.2 Proof of Proposition 2.2.2

Case 1: $\delta_i = 1$. Thus lifetime is interval-censored, and $(l_i, r_i) = (Y_{j-1(i)}, Y_{j(i)})$.

$f(l_i, r_i, \delta_i | D_i = d_i)$

$= f(Y_{j-1(i)}, Y_{j(i)}, \delta_i = 1 | D_i = d_i)$

$= P[\text{Exactly one of } W_a\text{'s has progression time in } (Y_{j-1(i)}, Y_{j(i)}) \text{ and all others} > Y_{j(i)}]$

$= \sum_{a=1}^{d_i} P[W_a \text{ has progression time in } (Y_{j-1(i)}, Y_{j(i)}) \text{ and all others} > Y_{j(i)}]$

$= \sum_{a=1}^{d_i} \{F(Y_{j(i)}) - F(Y_{j-1(i)})\}\{S(Y_{j(i)})\}^{d_i - 1}$

$= d_i \{F(Y_{j(i)}) - F(Y_{j-1(i)})\}\{S(Y_{j(i)})\}^{d_i - 1}.$

Case 2: $\delta_i = 0$. Then the lifetime is right-censored and the subject is either cured or susceptible. Thus the observed data is $(l_i, r_i, \delta_i) = (Y_{j(i)}, \infty, 0)$.

a) Suppose the patient is cured. Then $D_i = 0$ and $T = W_0 = \infty$.

$$\begin{aligned} f(Y_{j(i)}, \infty, \delta_i = 0 | D_i = d_i = 0) &= P(\text{survival past } Y_{j(i)} \text{ given } d_i = 0) \\ &= P(W_0 = \infty) \\ &= 1 \\ &= \{S(Y_{j(i)})\}^{d_i} \end{aligned}$$

b) Suppose the patient is susceptible. Then $d_i > 0$ and $T = \min\{W_1, \ldots, W_{d_i}\}$.

$$\begin{aligned} f(Y_{j(i)}, \infty, \delta_i = 0 | D_i = d_i) &= P(W_1 > Y_{j(i)}, W_2 > Y_{j(i)}, \ldots, W_{d_i} > Y_{j(i)}) \\ &= \{S(Y_{j(i)})\}^{d_i}. \end{aligned}$$

Hence, when $\delta_i = 0$, $f(Y_{j(i)}, \infty, \delta_i = 0 | D_i = d_i) = \{S(Y_{j(i)})\}^{d_i}$.

Finally, combining cases 1 and 2, we may write

$$\begin{aligned} f(l_i, r_i, \delta_i | D_i = d_i) &= f(Y_{j-1(i)}, Y_{j(i)}, \delta_i | D_i = d_i) \\ &= [d_i \{F(Y_{j(i)}) - F(Y_{j-1(i)})\}]^{\delta_i} \{S(Y_{j(i)})\}^{d_i - \delta_i} \\ &= [d_i \{S(Y_{j-1(i)}) - S(Y_{j(i)})\}]^{\delta_i} \{S(Y_{j(i)})\}^{d_i - \delta_i}. \end{aligned}$$

A.1.3 Proof of Proposition 2.3.1

Note that

$$P(D_i = d_i, M_i = m_i | \boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}) = \frac{f(l_i, r_i, \delta_i, d_i, m_i; \boldsymbol{\psi})}{f(l_i, r_i, \delta_i; \boldsymbol{\psi})}. \tag{A.1}$$

Using the results of propositions 2.2.1 and 2.2.2,

$$\begin{aligned}
f(l_i, r_i, \delta_i; \boldsymbol{\psi}) = & \sum_{d_i=\delta_i}^{\infty} f(l_i, r_i, \delta_i | D_i = d_i) P(D_i = d_i; \theta_i, p_i) \\
= & \sum_{d_i=\delta_i}^{\infty} [d_i \{S(Y_{j-1(i)}) - S(Y_{j(i)})\}]^{\delta_i} \{S(Y_{j(i)})\}^{d_i - \delta_i} \frac{e^{-\theta_i p_i} (\theta_i p_i)^{d_i}}{d_i!} \left[(1 - p_i) + \frac{d_i}{\theta_i}\right] \\
= & \left(\frac{S(Y_{j-1(i)}) - S(Y_{j(i)})}{S(Y_{j(i)})}\right)^{\delta_i} \frac{e^{-\theta_i p_i}}{\theta_i} \times \\
& \sum_{d_i=\delta_i}^{\infty} \frac{(\theta_i p_i S(Y_{j(i)}))^{d_i} [\theta_i (1 - p_i) + d_i]}{(d_i - \delta_i)!}.
\end{aligned} \tag{A.2}$$

Focusing only on the sum term from (A.2),

$$\begin{aligned}
& \sum_{d_i=\delta_i}^{\infty} \frac{(\theta_i p_i S(Y_{j(i)}))^{d_i} [\theta_i (1 - p_i) + d_i]}{(d_i - \delta_i)!} \\
& = \theta_i (1 - p_i) \sum_{d_i=\delta_i}^{\infty} \frac{(\theta_i p_i S(Y_{j(i)}))^{d_i}}{(d_i - \delta_i)!} + \sum_{d_i=\delta_i}^{\infty} \frac{d_i (\theta_i p_i S(Y_{j(i)}))^{d_i}}{(d_i - \delta_i)!} \\
& = \begin{cases}
\theta_i (1 - p_i) e^{\theta_i p_i S(Y_{j(i)})} + \theta_i p_i S(Y_{j(i)}) e^{\theta_i p_i S(Y_{j(i)})}, & \delta_i = 0 \\
\theta_i^2 p_i (1 - p_i) S(Y_{j(i)}) e^{\theta_i p_i S(Y_{j(i)})} + \theta_i p_i S(Y_{j(i)}) e^{\theta_i p_i S(Y_{j(i)})} \left[\theta_i p_i (S(Y_{j(i)}) + 1)\right], & \delta_i = 1
\end{cases} \\
& = e^{\theta_i p_i S(Y_{j(i)})} \left[\theta_i (1 - p_i) (\theta_i p_i S(Y_{j(i)}))^{\delta_i} + \theta_i p_i S(Y_{j(i)}) \left(\theta_i p_i S(Y_{j(i)}) + \delta_i\right)^{\delta_i}\right].
\end{aligned}$$

Now

$$\begin{aligned}
f(l_i, r_i, \delta_i; \boldsymbol{\psi}) = & \left(\frac{S(Y_{j-1(i)}) - S(Y_{j(i)})}{S(Y_{j(i)})}\right)^{\delta_i} \frac{e^{-\theta_i p_i (1 - S(Y_{j(i)}))}}{\theta_i} \times \\
& \left[\theta_i (1 - p_i) (\theta_i p_i S(Y_{j(i)}))^{\delta_i} + \theta_i p_i S(Y_{j(i)}) \left(\theta_i p_i S(Y_{j(i)}) + \delta_i\right)^{\delta_i}\right] \\
= & \left[\frac{\theta_i p_i (S(Y_{j-1(i)}) - S(Y_{j(i)}))}{S(Y_{j(i)})}\right]^{\delta_i} e^{-\theta_i p_i (F(Y_{j(i)}))} \times
\end{aligned}$$

$$\left[1-p_i+p_iS(Y_{j(i)})\left(1+\frac{\delta_i}{\theta_i p_i S(Y_{j(i)})}\right)^{\delta_i}\right]$$
$$= \left[\theta_i p_i(S(Y_{j-1(i)})-S(Y_{j(i)}))\right]^{\delta_i} e^{-\theta_i p_i(F(Y_{j(i)}))}\left[1-p_i+p_iS(Y_{j(i)})+\frac{\delta_i}{\theta_i}\right].$$

Let $\theta_i^* = \theta_i\left(1-p_iF(Y_{j(i)})+\frac{\delta_i}{\theta_i}\right), p_i^* = \frac{1-p_iF(Y_{j(i)})}{1-p_iF(Y_{j(i)})+\frac{\delta_i}{\theta_i}}, p_i^+ = \frac{p_iS(Y_{j(i)})}{1-p_iF(Y_{j(i)})}$.

$$P(D_i=d_i, M_i=m_i|D_{obs},\boldsymbol{\psi})$$
$$= \frac{[d_i\{S(Y_{j-1(i)})-S(Y_{j(i)})\}]^{\delta_i}\{S(Yj(i))\}^{d_i-\delta_i}\binom{m_i}{d_i}p_i^{d_i}(1-p_i)^{m_i-d_i}e^{-\theta_i}\frac{\theta_i^{m_i-1}}{(m_i-1)!}}{\left[\theta_i p_i(S(Y_{j-1(i)})-S(Y_{j(i)}))\right]^{\delta_i} e^{-\theta_i p_i(F(Y_{j(i)}))}\left[1-p_i+p_iS(Y_{j(i)})+\frac{\delta_i}{\theta_i}\right]}$$
$$= \frac{e^{-\theta_i(1-p_iF(Y_{j(i)}))}}{(m_i-\delta_i)!}(\theta_i(1-p_i))^{m_i}\binom{m_i-\delta_i}{d_i-\delta_i}\left(\frac{S(Y_{j(i)})p_i}{1-p_i}\right)^{d_i}\times$$
$$(\theta_i p_i S(Y_{j(i)}))^{-\delta_i}\frac{m_i}{\theta_i\left(1-p_iF(Y_{j(i)})+\frac{\delta_i}{\theta_i}\right)}. \tag{A.3}$$

Focusing on the last term in (A.3),

$$\frac{m_i}{\theta_i\left(1-p_iF(Y_{j(i)})+\frac{\delta_i}{\theta_i}\right)} = \frac{\delta_i+m_i-\delta_i}{\theta_i\left(1-p_iF(Y_{j(i)})+\frac{\delta_i}{\theta_i}\right)}$$
$$= \frac{\theta_i\left(1-p_iF(Y_{j(i)})+\frac{\delta_i}{\theta_i}\right)-\theta_i\left(1-p_iF(Y_{j(i)})\right)+m_i-\delta_i}{\theta_i\left(1-p_iF(Y_{j(i)})+\frac{\delta_i}{\theta_i}\right)}$$
$$= 1-\frac{1-p_iF(Y_{j(i)})}{1-p_iF(Y_{j(i)})+\frac{\delta_i}{\theta_i}}+\frac{m_i-\delta_i}{\theta_i\left(1-p_iF(Y_{j(i)})+\frac{\delta_i}{\theta_i}\right)}.$$

Now

$$P(D_i=d_i, M_i=m_i|\boldsymbol{D}_{obs},\boldsymbol{\psi}) =$$
$$\frac{(\theta_i(1-p_i))^{m_i}e^{-\theta_i(1-p_iF(Y_{j(i)}))}}{(m_i-\delta_i)!}\binom{m_i-\delta_i}{d_i-\delta_i}\left(\frac{S(Y_{j(i)})p_i}{1-p_i}\right)^{d_i}\times$$
$$(\theta_i p_i S(Y_{j(i)}))^{-\delta_i}\left(1-\frac{1-p_iF(Y_{j(i)})}{1-p_iF(Y_{j(i)})+\frac{\delta_i}{\theta_i}}+\frac{m_i-\delta_i}{\theta_i\left(1-p_iF(Y_{j(i)})+\frac{\delta_i}{\theta_i}\right)}\right). \tag{A.4}$$

Noting that

$$\theta_i^*p_i^* = \theta_i(1-p_iF(Y_{j(i)})),$$

$$1 - p_i^+ = \frac{1 - p_i}{1 - p_i F(Y_{j(i)})},$$

$$(1 - p_i^+)(\theta_i^* p_i^*) = \theta_i(1 - p_i),$$

$$p_i^+ \theta_i^* p_i^* = \theta_i p_i S(Y_{j(i)}),$$

we can write (A.4) in equivalent form

$$P(D_i = d_i, M_i = m_i | \boldsymbol{D}_{obs}; \boldsymbol{\psi})$$
$$= \underbrace{\frac{(\theta_i^* p_i^*)^{m_i - \delta_i} e^{-\theta_i^* p_i^*}}{(m_i - \delta_i)!} \left(1 - p_i^* + \frac{m_i - \delta_i}{\theta_*}\right)}_{M_i - \delta_i \sim \text{Pois}(\theta_i^* p_i^*) + \text{Bern}(p_i^*)} \underbrace{\binom{m_i - \delta_i}{d_i - \delta_i} (p_i^+)^{d_i - \delta_i} (1 - p_i^+)^{m_i - d_i}}_{D_i - \delta_i | M_i = m_i; \boldsymbol{D}_{\text{obs}} \sim \text{Bin}(m_i - \delta_i, p_i^+)}.$$

A.1.4 Proof of Proposition 2.3.2

Let $\theta_i^* = \theta_i(1 - p_i F(Y_{j(i)}) + \frac{\delta_i}{\theta_i}), p_i^* = \frac{p_i S(Y_{j(i)})}{1 - p_i F(Y_{j(i)}) + \frac{\delta_i}{\theta_i}}$.

Then

$$\begin{aligned}
P(D_i = d_i | \boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}) &= \frac{f(l_i, r_i, \delta_i, d_i; \theta_i, p_i, \phi, \boldsymbol{\lambda})}{f(l_i, r_i, \delta_i; \theta_i, p_i, \phi, \boldsymbol{\lambda})} \\
&= \frac{f(l_i, r_i, \delta_i | d_i) P(D_i = d_i; \theta_i, p_i)}{f(l_i, r_i, \delta_i; \theta_i, p_i, \phi, \boldsymbol{\lambda})} \\
&= \frac{\overbrace{\left[d_i \{S(Y_{j-1(i)}; \boldsymbol{\lambda}) - S(Y_{j(i)})\}\right]^{\delta_i} S(Yj(i))^{d_i - \delta_i}}^{\text{From Proposition (2.2.2)}} \frac{e^{-\theta_i p_i} (\theta_i p_i)^{d_i}}{d_i!} \left[(1 - p_i) + \frac{d_i}{\theta_i}\right]}{\left[\theta_i p_i (S(Y_{j-1(i)}) - S(Y_{j(i)}))\right]^{\delta_i} e^{-\theta_i p_i (F(Y_{j(i)}))} \left[1 - p_i + p_i S(Y_{j(i)}) + \frac{\delta_i}{\theta_i}\right]} \\
&= \frac{[\theta_i p_i S(Y_{j(i)})]^{d_i - \delta_i} e^{-\theta_i p_i S(Y_{j(i)})}}{(d_i - \delta_i)!} \left(\frac{1 - p_i + \frac{d_i}{\theta_i}}{1 - p_i F(Y_{j(i)}) + \frac{\delta_i}{\theta_i}}\right). \qquad \text{(A.5)}
\end{aligned}$$

Focusing only on the second term in (A.5),

$$\begin{aligned}
\frac{1 - p_i + \frac{d_i}{\theta_i}}{1 - p_i F(Y_{j(i)}) + \frac{\delta_i}{\theta_i}} &= \frac{\theta_i - \theta_i p_i + d_i}{\theta_i \left(1 - p_i F(Y_{j(i)}) + \frac{\delta_i}{\theta_i}\right)} \\
&= \frac{\theta_i \left(1 - p_i F(Y_{j(i)}) + \frac{\delta_i}{\theta_i}\right) - \theta_i p_i \left(1 - p_i F(Y_{j(i)}) + \frac{\delta_i}{\theta_i}\right) + d_i - \delta_i}{\theta_i \left(1 - p_i F(Y_{j(i)}) + \frac{\delta_i}{\theta_i}\right)} \\
&= \frac{\theta_i^* - \theta_i^* p_i^* + d_i - \delta_i}{\theta_i^*}
\end{aligned}$$

$$= 1 - p_i^* + \frac{d_i - \delta_i}{\theta_i^*}.$$

Now (A.5) may be written as

$$P(D_i = d_i | \boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}) = \frac{(\theta_i^* p_i^*)^{d_i - \delta_i} e^{-\theta_i^* p_i^*}}{(d_i - \delta_i)!} \left(1 - p_i^* + \frac{d_i - \delta_i}{\theta_i^*}\right). \tag{A.6}$$

A.2 Proofs of the propositions used in Chapters 3

A.2.1 Proof of Proposition 3.2.1

Selecting η such that $\eta = \frac{\theta}{\phi(1-\theta)}$, $\theta \neq 1$,

$$\begin{aligned} E_\theta \left[w(M;\phi)\right] &= e^{-\frac{\phi\eta}{1+\phi\eta}} \sum_{m=0}^{\infty} \frac{\Gamma\left(m + \phi^{-1}\right)}{m!} \left(\frac{\phi\eta}{1+\phi\eta}\right)^m \\ &= e^{-\frac{\phi\eta}{1+\phi\eta}} \Gamma(\phi^{-1})(1+\phi\eta)^{\phi^{-1}}. \end{aligned}$$

Hence,

$$\begin{aligned} P[M = m; \phi, \theta] &= \frac{\Gamma(m+\phi^{-1}) e^{-\theta} \theta^m}{m! \Gamma(\phi^{-1}) e^{-\frac{\phi\eta}{1+\phi\eta}} (1+\phi\eta)^{\phi^{-1}}} \\ &= \frac{\Gamma(m+\phi^{-1})}{m!\Gamma(\phi^{-1})} \left(\frac{\phi\eta}{1+\phi\eta}\right)^m (1+\phi\eta)^{-\phi^{-1}}, m = 0, 1, 2 \ldots, \end{aligned}$$

which is the pmf of a negative binomial distribution with $r = \phi^{-1}$ and $p = \frac{\phi\eta}{1+\phi\eta}$.

A.2.2 Proof of proposition 3.2.2

Noting that $D|M = m \sim \text{Bin}(m, p)$ if $m > 0$ and $P(D = 0|M = 0) = 1$,

$$\begin{aligned} P(D = d; \eta, p) &= \sum_{m=d}^{\infty} P(D|M = m) P(M = m) \\ &= \sum_{m=d}^{\infty} \binom{m}{d} p^d (1-p)^{m-d} \frac{\Gamma(m+\phi^{-1})}{m!\Gamma(\phi^{-1})} \left(\frac{\phi\eta}{1+\phi\eta}\right)^m (1+\phi\eta)^{-\phi^{-1}} \\ &= \frac{(1+\phi\eta)^{-\phi^{-1}} (p\theta)^d}{d!\Gamma(\phi^{-1})} \sum_{m=d}^{\infty} \frac{\Gamma(m+\phi^{-1}) \left[\theta(1-p)\right]^{m-d}}{(m-d)!}. \end{aligned}$$

Let $t = m - d$. Then

$$P(D = d; \theta, p) = \frac{(1+\phi\eta)^{-\phi^{-1}} (p\theta)^d}{d!\Gamma(\phi^{-1})} \sum_{t=0}^{\infty} \frac{\Gamma(t+d+\phi^{-1}) \left[\theta(1-p)\right]^t}{t!}$$

$$
\begin{aligned}
&= \frac{(1+\phi\eta)^{-\phi^{-1}}(p\theta)^d\Gamma(\phi^{-1}+d)}{d!\Gamma(\phi^{-1})}\sum_{t=0}^{\infty}\frac{\Gamma(t+d+\phi^{-1})\left[\theta(1-p)\right]^t}{\Gamma(\phi^{-1}+d)t!} \\
&= \frac{\Gamma(\phi^{-1}+d)}{d!\Gamma(\phi^{-1})}\left(\frac{\phi\eta p}{1+\phi\eta p}\right)^d\left(\frac{1}{1+\phi\eta p}\right)^{\phi^{-1}}, d=0,1,2,\ldots
\end{aligned}
$$

Thus $D \sim \text{NB}\left(\phi^{-1}, \frac{\phi\eta p}{1+\phi\eta p}\right)$.

A.2.3 Proof of Proposition 3.3.1

Note that

$$
P(D_i = d_i, M_i = m_i|\boldsymbol{D}_{\text{obs}};\boldsymbol{\psi}) = \frac{f(l_i, r_i, \delta_i, d_i, m_i;\boldsymbol{\psi})}{f(l_i, r_i, \delta_i;\boldsymbol{\psi})}. \tag{A.7}
$$

Using the results of propositions 3.2.1 and 3.2.2,

$$
\begin{aligned}
&f(l_i, r_i, \delta_i;\boldsymbol{\psi}) \\
&= \sum_{d_i=\delta_i}^{\infty} f(l_i, r_i, \delta_i|D_i = d_i)P(D_i = d_i;\theta_i, p_i) \\
&= \sum_{d_i=\delta_i}^{\infty}\left[d_i\{S(Y_{j-1(i)}) - S(Y_{j(i)})\}\right]^{\delta_i}\{S(Y_{j(i)})\}^{d_i-\delta_i}\frac{\Gamma(\phi^{-1}+d_i)}{\Gamma(\phi^{-1})d_i!}\left(\frac{\phi\eta_i p_i}{1+\phi\eta_i p_i}\right)^{d_i}\left(\frac{1}{1+\phi\eta_i p_i}\right)^{\phi^{-1}} \\
&= \left(\frac{S(Y_{j-1(i)}) - S(Y_{j(i)})}{S(Y_{j(i)})}\right)^{\delta_i}\left(\frac{1}{1+\phi\eta_i p_i}\right)^{\phi^{-1}}\frac{1}{\Gamma(\phi^{-1})}\sum_{d_i=\delta_i}^{\infty}\left(\frac{\phi\eta_i p_i S(Y_{j(i)})}{1+\phi\eta_i p_i}\right)^{d_i}\frac{\Gamma(\phi^{-1}+d_i)}{(d_i-\delta_i)!}.
\end{aligned} \tag{A.8}
$$

Focusing on the sum in (A.8):

If $\delta_i = 0$ then

$$
\sum_{d_i=0}^{\infty}\left(\frac{\phi\eta_i p_i S(Y_{j(i)})}{1+\phi\eta_i p_i}\right)^{d_i}\frac{\Gamma(\phi^{-1}+d_i)}{d_i!} = \Gamma(\phi^{-1})\left(\frac{1+\phi\eta_i p_i}{1+\phi\eta_i p_i F(Y_{j(i)})}\right)^{\phi^{-1}}, \tag{A.9}
$$

and (A.8) is equivalent to

$$
f(l_i, r_i;\delta_i = 0,\boldsymbol{\psi}) = (1+\phi\eta_i p_i F(Y_{j(i)}))^{-\phi^{-1}}. \tag{A.10}
$$

If $\delta_i = 1$ then, letting $q_i = d_i - 1$,

$$
\begin{aligned}
&\sum_{d_i=1}^{\infty}\left(\frac{\phi\eta_i p_i S(Y_{j(i)})}{1+\phi\eta_i p_i}\right)^{d_i}\frac{\Gamma(\phi^{-1}+d_i)}{(d_i-1)!}\\
&=\Gamma(\phi^{-1}+1)\left(\frac{\phi\eta_i p_i S(Y_{j(i)})}{1+\phi\eta_i p_i}\right)\sum_{q_i=0}^{\infty}\left(\frac{\phi\eta_i p_i S(Y_{j(i)})}{1+\phi\eta_i p_i}\right)^{q_i}\frac{\Gamma(\phi^{-1}+1+q_i)}{\Gamma(\phi^{-1}+1)q_i!}\\
&=\Gamma(\phi^{-1}+1)\left(\frac{\phi\eta_i p_i S(Y_{j(i)})}{1+\phi\eta_i p_i}\right)\left(\frac{1+\phi\eta_i p_i}{1+\phi\eta_i p_i F(Y_{j(i)})}\right)^{\phi^{-1}+1}, \qquad \text{(A.11)}
\end{aligned}
$$

and (A.8) is equivalent to

$$
f(l_i,r_i;\delta_i=1,\boldsymbol{\psi})=\eta_i p_i\{S(Y_{j-1(i)})-S(Y_{j(i)})\}\left[1+\phi\eta_i p_i F(Y_{j(i)})\right]^{-(\phi^{-1}+1)}. \quad \text{(A.12)}
$$

Combining (A.10) and (A.12),

$$
f(l_i,r_i,\delta_i;\boldsymbol{\psi})=\left(\eta_i p_i\{S(Y_{j-1(i)})-S(Y_{j(i)})\}\right)^{\delta_i}\left[1+\phi\eta_i p_i F(Y_{j(i)})\right]^{-(\phi^{-1}+\delta_i)}. \quad \text{(A.13)}
$$

Let $\phi_i^*=\phi^{-1}+\delta_i, \eta_i^+=\frac{\phi\eta_i\left(1-p_iF(Y_{j(i)})\right)}{1+\phi\eta_i}, p_i^+=\frac{p_iS(Y_{j(i)})}{1-p_iF(Y_{j(i)})}$.

$$
\begin{aligned}
&P(D_i=d_i,M_i=m_i|\boldsymbol{D}_{\text{obs}},\boldsymbol{\psi})\\
&=\frac{S(Y_{j(i)})^{d_i-\delta_i}[d_i\{S(Y_{j-1(i)})-S(Y_{j(i)})\}]^{\delta_i}\binom{m_i}{d_i}p_i^{d_i}(1-p_i)^{m_i-d_i}\frac{\Gamma(\phi^{-1}+m_i)}{m_i!\Gamma(\phi^{-1})}\left(\frac{\phi\eta_i}{1+\phi\eta_i}\right)^{m_i}(1+\phi\eta_i)^{-\phi^{-1}}}{\left(\eta_i p_i\{S(Y_{j-1(i)})-S(Y_{j(i)})\}\right)^{\delta_i}\left[1+\phi\eta_i p_i F(Y_{j(i)})\right]^{-(\phi^{-1}+\delta_i)}}\\
&=\frac{\Gamma(\phi^{-1}+\delta_i+m_i-\delta_i)}{\Gamma(\phi^{-1}+\delta_i)(m_i-\delta_i)!}\left(\frac{\phi\eta_i(1-p_iF(Y_{j(i)}))}{1+\phi\eta_i}\right)^{m_i-\delta_i}\left(\frac{1+\phi\eta_i p_iF(Y_{j(i)})}{1+\phi\eta_i}\right)^{\phi^{-1}+\delta_i}\times\\
&\quad\binom{m_i-\delta_i}{d_i-\delta_i}\left(\frac{p_iS(Y_{j(i)})}{1-p_iF(Y_{j(i)})}\right)^{d_i-\delta_i}\left(\frac{1-p_i}{1-p_iF(Y_{j(i)})}\right)^{m_i-d_i} \qquad \text{(A.14)}\\
&=\underbrace{\frac{\Gamma(\phi_i^*+m_i-\delta_i)}{\Gamma(\phi_i^*)(m_i-\delta_i)!}(\eta^+)^{m_i-\delta_i}(1-\eta_i^+)^{\phi_i^*}}_{M_i-\delta_i|\boldsymbol{D}_{\text{obs}}\sim\text{NB}(\phi_i^*,\eta_i^+)}\times\underbrace{\binom{m_i-\delta_i}{d_i-\delta_i}(p_i^+)^{d_i-\delta_i}(1-p_i^+)^{m_i-d_i}}_{D_i-\delta_i|M_i=m_i;\boldsymbol{D}_{\text{obs}}\sim\text{Bin}(m_i-\delta_i,p_i^+)}. \qquad \text{(A.15)}
\end{aligned}
$$

A.2.4 Proof of Proposition 3.3.2:

Let $\phi_i^*=\phi^{-1}+\delta_i, \eta_i^*=\frac{\phi\eta_i p_i S(Y_{j(i)})}{1+\phi\eta_i p_i}$.

Then

$$
P(D_i=d_i|\boldsymbol{D}_{\text{obs}};\boldsymbol{\psi})=\frac{f(l_i,r_i,\delta_i,d_i;\theta_i,p_i,\phi,\boldsymbol{\lambda})}{f(l_i,r_i,\delta_i;\theta_i,p_i,\phi,\boldsymbol{\lambda})}
$$

$$
\begin{aligned}
&= \frac{f(l_i, r_i, \delta_i | d_i) P(D_i = d_i; \theta_i, p_i)}{f(l_i, r_i, \delta_i; \theta_i, p_i, \phi, \boldsymbol{\lambda})} \\
&\overset{\text{From (3.9)}}{=} \frac{\overbrace{\left[d_i\{S(Y_{j-1(i)}) - S(Y_{j(i)})\}\right]^{\delta_i} S(Y_{j(i)})^{d_i - \delta_i}} \frac{\Gamma(\phi^{-1}+d_i)}{d_i!\Gamma(\phi^{-1})} \left(\frac{\phi\eta_i p_i}{1+\phi\eta_i p_i}\right)^{d_i} (1+\phi\eta_i p_i)^{-\phi^{-1}}}{\left(\eta_i p_i\{S(Y_{j-1(i)}) - S(Y_{j(i)})\}\right)^{\delta_i} \left[1+\phi\eta_i p_i F(Y_{j(i)})\right]^{-(\phi^{-1}+\delta_i)}} \\
&= \frac{\Gamma(\phi^{-1}+\delta_i+d_i-\delta_i)}{(d_i-\delta_i)!\Gamma(\phi^{-1}+\delta_i)} \left(\frac{\phi\eta_i p_i S(Y_{j(i)})}{1+\phi\eta_i p_i}\right)^{d_i-\delta_i} \left(\frac{1+\phi\eta_i p_i F(Y_{j(i)})}{1+\phi\eta_i p_i}\right)^{\phi^{-1}+\delta_i} \\
&= \underbrace{\frac{\Gamma(\phi_i^* + d_i - \delta_i)}{(d_i-\delta_i)!\Gamma(\phi_i^*)} (\eta_i^*)^{d_i-\delta_i} (1-\eta_i^*)^{\phi_i^*}}_{D_i - \delta_i | \boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi} \sim \text{NB}(\phi_i^*, \eta_i^*)}. \qquad \text{(A.16)}
\end{aligned}
$$

A.2.5 Proof of Proposition 3.3.3:

From (3.9) and (3.13),

$$
\begin{aligned}
&f(l_i, r_i, \delta_i, m_i, u_i, d_i; \boldsymbol{\psi}) \\
&= \sum_{d_i=0}^{m_i} f(l_i, r_i, \delta_i, m_i, u_i, d_i; \boldsymbol{\psi}) \\
&= \sum_{d_i=\delta_i}^{m_i} \left[d_i\{S(Y_{j-1(i)}) - S(Y_{j(i)})\}\right]^{\delta_i} S(Yj(i))^{d_i-\delta_i} \times \\
&\quad \binom{m_i}{d_i} p_i^{d_i} (1-p_i)^{m_i-d_i} \frac{u_i^{m_i} e^{-u_i}}{m_i!} f_{U_i}(u_i; \eta_i, \phi) \\
&= \left[\frac{S(Y_{j-1(i)}) - S(Y_{j(i)})}{S(Y_{j(i)})}\right]^{\delta_i} (1-p_i)^{m_i} \times \\
&\quad \frac{u_i^{m_i} e^{-u_i}}{m_i!} f_{U_i}(u_i; \eta_i, \phi) \sum_{d_i=\delta_i}^{m_i} d_i^{\delta_i} \binom{m_i}{d_i} \left[\frac{p_i S(Y_{j(i)})}{1-p_i}\right]^{d_i}.
\end{aligned}
$$

If $\delta_i = 0$, then

$$
\begin{aligned}
&f(l_i, r_i, \delta_i, m_i, u_i, d_i; \boldsymbol{\psi}) \\
&= f_{U_i}(u_i; \eta_i, \phi) \frac{u_i^{m_i} e^{-u_i}}{m_i!} \sum_{d_i=0}^{m_i} \binom{m_i}{d_i} [p_i S(Y_{j(i)})]^{d_i} (1-p_i)^{m_i-d_i} \\
&= f_{U_i}(u_i; \eta_i, \phi) \frac{u_i^{m_i} e^{-u_i}}{m_i!} \left[1 - p_i F(Y_{j(i)})\right]^{m_i} \times
\end{aligned}
$$

$$\underbrace{\sum_{d_i=0}^{m_i}\binom{m_i}{d_i}\left(\frac{p_iS(Y_{j(i)})}{1-p_iF(Y_{j(i)})}\right)^{d_i}\left(\frac{1-p_i}{1-p_iF(Y_{j(i)})}\right)^{m_i-d_i}}_{\text{sum of pmf equal to 1}}.$$

$$= f_{U_i}(u_i;\eta_i,\phi)\frac{u_i^{m_i}e^{-u_i}}{m_i!}\left[1-p_iF(Y_{j(i)})\right]^{m_i} \tag{A.17}$$

If $\delta_i = 1$, then

$$f(l_i, r_i, \delta_i, m_i, u_i, d_i; \boldsymbol{\psi})$$
$$= f_{U_i}(u_i;\eta_i,\phi)\frac{u_i^{m_i}e^{-u_i}}{m_i!}\frac{[S(Y_{j-1(i)})-S(Y_{j(i)})]}{S(Y_{j(i)})}\sum_{d_i=1}^{m_i} d_i\binom{m_i}{d_i}[p_iS(Y_{j(i)})]^{d_i}(1-p_i)^{m_i-d_i}.$$

Letting $q_i = d_i - 1$,

$$f(l_i, r_i, \delta_i, m_i, u_i, d_i; \boldsymbol{\psi})$$
$$= \frac{S(Y_{j-1(i)})-S(Y_{j(i)})}{S(Y_{j(i)})}f_{U_i}(u_i;\eta_i,\phi)\frac{u_i^{m_i}e^{-u_i}}{m_i!}m_ip_iS(Y_{j(i)})[1-p_iF(Y_{j(i)})]^{m_i-1}\times$$
$$\underbrace{\sum_{q_i=0}^{m_i-1}\binom{m_i-1}{q_i}[p_iS(Y_{j(i)})]^{q_i}\left(\frac{1-p_i}{1-p_iF(Y_{j(i)})}\right)^{m_i-1-q_i}}_{\text{sum of pmf equal to 1}}$$
$$= p_i(S(Y_{j-1(i)})-S(Y_{j(i)}))[1-p_iF(Y_{j(i)})]^{m_i-1}f_{U_i}(u_i;\eta_i,\phi)\frac{u_i^{m_i}e^{-u_i}}{(m_i-1)!}. \tag{A.18}$$

Combining (A.17) and (A.18),

$$f(l_i, r_i, \delta_i, m_i, u_i, d_i; \boldsymbol{\psi})$$
$$= \left[p_i(S(Y_{j-1(i)})-S(Y_{j(i)}))\right]^{\delta_i}[1-p_iF(Y_{j(i)})]^{m_i-\delta_i}f_{U_i}(u_i;\eta_i,\phi)\frac{u_i^{m_i}e^{-u_i}}{(m_i-\delta_i)!}. \tag{A.19}$$

Now

$$f(m_i, u_i|\boldsymbol{D}_{\text{obs}};\boldsymbol{\psi})$$
$$= \frac{f(l_i, r_i, \delta_i, m_i, u_i, d_i; \boldsymbol{\psi})}{f(l_i, r_i, \delta_i, ; \boldsymbol{\psi})}$$
$$= \frac{\left[p_i(S(Y_{j-1(i)})-S(Y_{j(i)}))\right]^{\delta_i}[1-p_iF(Y_{j(i)})]^{m_i-\delta_i}f_{U_i}(u_i;\eta_i,\phi)\frac{u_i^{m_i}e^{-u_i}}{(m_i-\delta_i)!}}{\left(\eta_ip_i\{S(Y_{j-1(i)})-S(Y_{j(i)})\}\right)^{\delta_i}\left[1+\phi\eta_ip_iF(Y_{j(i)})\right]^{-(\phi^{-1}+\delta_i)}}$$

$$= \frac{\left[u_i(1-p_iF(Y_{j(i)}))\right]^{m_i-\delta_i}}{(m_i-\delta_i)!}e^{-u_i(1-p_iF(Y_{j(i)}))}\frac{\left(\frac{1+\phi\eta_ip_iF(Y_{j(i)})}{\phi\eta_i}\right)^{\phi^{-1}+\delta_i}}{\Gamma(\phi^{-1}+\delta_i)}\times$$
$$u_i^{\phi^{-1}+\delta_i-1}\exp\left\{-\left(\frac{1+\phi\eta_ip_iF(Y_{j(i)})}{\phi\eta_i}\right)u_i\right\}.$$

Let $a_i^* = u_i(1-p_iF(Y_{j(i)}))$, $\eta_i^* = \frac{\phi\eta_i}{1+\phi\eta_ip_iF(Y_{j(i)})}$, $\phi_i^* = \phi^{-1}+\delta_i$. Then

$$f(m_i,u_i|\boldsymbol{D}_{\text{obs}};\boldsymbol{\psi}) = \underbrace{\frac{(a_i^*)^{m_i-\delta_i}}{(m_i-\delta_i)!}e^{a_i^*}}_{M_i-\delta_i|U_i=u_i,\boldsymbol{D}_{\text{obs}};\boldsymbol{\psi}\sim\text{Poisson}(a_i^*)} \times \underbrace{\frac{(\eta_i^*)^{-\phi_i^*}}{\Gamma(\phi_i^*)}u_i^{\phi_i^*-1}e^{-\frac{u_i}{\eta_i^*}}}_{U_i|\boldsymbol{D}_{\text{obs}};\boldsymbol{\psi}\sim\text{Gamma}(\phi_i^*,\eta_i^*)}. \tag{A.20}$$

APPENDIX B

Appendix B

In this appendix, we present the process by which initial parameter estimates were selected in implementing the EM algorithm for the destructive shifted Poisson cure rate model in the real data analysis in Section 2.5.

- **Model 1.1:** $\theta = e^{\beta_{10}+\beta_{11}z_1}, p = \frac{\exp(\beta_{21}z_2)}{1+\exp(\beta_{21}z_2)}$, where p contains no intercept term. Denote the cure rate of the i-th patient as q_{0i}.

$$\begin{aligned} q_{0i|z_1=0} = &\quad (1-p_i)\exp\left(-p_i e^{\beta_{10}}\right), \\ q_{0i|z_1=1} = &\quad (1-p_i)\exp\left(-p_i e^{\beta_{10}+\beta_{11}}\right). \end{aligned}$$

We obtain estimates for cure rates k_0, k_1 from the Kaplan-Meier plots stratified by EMP such that

$$\widehat{q}_{0i} = \begin{cases} k_0, z_1 = 0 \\ \\ k_1, z_1 = 1, \end{cases}$$

where $k_0, k_1 \in (0,1)$. We equate the cure rate formulas with the above estimates:

$$\begin{aligned} (1-p_i)\exp\left(-p_i e^{\beta_{10}}\right) &= k_0, \\ (1-p_i)\exp\left(-p_i e^{\beta_{10}+\beta_{11}}\right) &= k_1, \end{aligned}$$

which gives rise to the equations

$$\begin{aligned} \beta_{10} = &\quad \log\left[-\tfrac{1}{p_i}\log\left(\tfrac{k_0}{1-p_i}\right)\right], \\ \beta_{11} = &\quad \log\left[-\tfrac{1}{p_i}\log\left(\tfrac{k_1}{1-p_i}\right)\right] - \beta_{10}. \end{aligned}$$

We observe that in order for the outer log function to be defined in both equations,

$p_i < \min(1-k_0, 1-k_1)$. Naively choosing $\widehat{p}_i = \frac{\min(1-k_0,1-k_1)}{2}$, we can compute

$$\begin{aligned} \widehat{\beta}_{10} = &\quad \log\left[-\tfrac{1}{\widehat{p}_i}\log\left(\tfrac{k_0}{1-\widehat{p}_i}\right)\right], \\ \widehat{\beta}_{11} = &\quad \log\left[-\tfrac{1}{\widehat{p}_i}\log\left(\tfrac{k_1}{1-\widehat{p}_i}\right)\right] - \widehat{\beta}_{10}. \end{aligned}$$

An estimate for β_{21} may be obtained by taking $\widehat{p}_i$ as the cure rate for the mean age of observed lifetimes. Let $\bar{z}_{2_{obs}}$ denote the mean age given $\delta_i = 1$. Then $\widehat{\beta}_{21} = \frac{1}{\bar{z}_{2_{\text{obs}}}} \log(\frac{\widehat{p}_i}{1-\widehat{p}_i})$. Finally, we find initial estimates for lifetime parameters λ_1 and λ_2. For this purpose, we take a point estimate for each observed lifetime contained in interval (l_i, r_i), with $r_i < \infty$, as $t_{\text{est}_i} = \frac{l_i + r_i}{2}$, and find Weibull parameters $\widehat{\lambda}_1, \widehat{\lambda}_2$, such that the mean is equal to the mean of $\boldsymbol{t}_{\text{est}}$ and variance is equal to the mean of $\boldsymbol{t}_{\text{est}}{}^{\mathbf{2}}$.

- **Model 1.2:** $\theta = e^{\beta_{11} z_1}, p = \frac{\exp(\beta_{20} + \beta_{21} z_2)}{1 + \exp(\beta_{20} + \beta_{21} z_2)}$, where θ contains no intercept term. Denote the cure rate of the i-th patient as q_{0i}.

$$\begin{aligned} q_{0i|z_1=0} &= \quad (1-p_i)e^{-p_i}, \\ q_{0i|z_1=1} &= \quad (1-p_i)e^{-p_i e^{\beta_{11}}}. \end{aligned}$$

We obtain estimates for cure rates k_0, k_1 from the Kaplan-Meier plots stratified by age such that

$$\widehat{q}_{0i} = \begin{cases} k_0, z_1 = 0 \\ k_1, z_1 = 1, \end{cases}$$

where $k_0, k_1 \in (0, 1)$. We equate the cure rate formulas with the above estimates:

$$\begin{aligned} (1-p_i)e^{-p_i} &= k_0, \\ (1-p_i)e^{-p_i e^{\beta_{11}}} &= k_1. \end{aligned}$$

The first equation can be solved explicitly for p_i, which we will take as $p_{i_{|z_2=\text{Med}(z_2)}} = k_2$. Then $\widehat{\beta}_{11} = \log\left[-\frac{1}{k_2}\log\left(\frac{k_1}{1-k_2}\right)\right]$. It is reasonable to expect lower activation probability to be associated with higher values of age, thus to find suitable initial estimates for β_{20} and β_{21}, we propose an approximately

proportional relationship between z_2 and p_i, so that if $p_{i_{|z_2=\text{Med}(z_2)}} = k_2$, then $p_{i_{|z_2=Q_3(z_2)}} = \frac{Q_3(z_2)}{\text{Med}(z_2)}k_2$. Now

$$\begin{aligned} p_{i_{|z_2=\text{Med}(z_2)}} = & \quad \frac{\exp[\beta_{20}+\beta_{21}\text{Med}(z_2)]}{1+\exp[\beta_{20}+\beta_{21}\text{Med}(z_2)]} = k_2 \\ p_{i_{|z_2=Q_3(z_2)}} = & \quad \frac{\exp[\beta_{20}+\beta_{21}Q_3(z_2)]}{1+\exp[\beta_{20}+\beta_{21}Q_3(z_2)]} = \frac{Q_3(z_2)}{\text{Med}(z_2)}k_2 \end{aligned}$$

can be solved explicitly for β_{20} and β_{21} yielding initial estimates of

$$\begin{aligned} \widehat{\beta}_{21} = & \quad \frac{\log\left(\frac{k_2}{1-k_2}\right) - \log\left(\frac{\frac{Q_3(z_2)}{\text{Med}(z_2)}k_2}{1-\frac{Q_3(z_2)}{\text{Med}(z_2)}k_2}\right)}{\text{Med}(z_2)-Q_3(z_2)}, \\ \widehat{\beta}_{20} = & \quad \log\left(\frac{k_2}{1-k_2}\right) - \widehat{\beta}_{21}\text{Med}(z_2). \end{aligned}$$

Initial values for lifetime parameters are obtained as in Model 1.1.

- **Model 2.1:** $\theta = e^{\beta_{20}+\beta_{21}z_2}, p = \frac{\exp(\beta_{11}z_1)}{1+\exp(\beta_{11}z_1)}$, where p contains no intercept term. Denote the cure rate of the i-th patient as q_{0i}.

$$\begin{aligned} q_{0i|z_1=0} = & \quad \frac{1}{2}\exp\left[-\exp\left(\frac{\beta_{20}+\beta_{21}z_2}{2}\right)\right], \\ q_{0i|z_1=1} = & \quad (1-p_i)\exp\left[-p_i\exp\left(\beta_{20}+\beta_{21}z_2\right)\right]. \end{aligned}$$

We obtain estimates for cure rates k_0, k_1 from the Kaplan-Meier plots stratified by EMP such that

$$\widehat{q}_{0i} = \begin{cases} k_0, z_1 = 0 \\ k_1, z_1 = 1, \end{cases}$$

where $k_0, k_1 \in (0,1)$. We equate the cure rate formulas with the above estimates:

$$\begin{aligned} \frac{1}{2}\exp\left[-\exp\left(\frac{\beta_{20}+\beta_{21}z_2}{2}\right)\right] &= k_0, \\ \left(1-\frac{e^{\beta_{11}}}{1+e^{\beta_{11}}}\right)\exp\left[-\exp\left(\beta_{20}+\beta_{21}z_2\right)\frac{e^{\beta_{11}}}{1+e^{\beta_{11}}}\right] &= k_1. \end{aligned}$$

Some algebraic manipulations yield

$$\beta_{20}+\beta_{21}z_2 = \quad \log\left[-2\log(2k_0)\right], \tag{B.1}$$

$$\beta_{20} + \beta_{21} z_2 = \log\left[-\tfrac{1}{p_i}\log\left(\tfrac{k_1}{1-p_i}\right)\right]. \tag{B.2}$$

We note that (B.1) is only defined if $k_0 < 0.5$, so if the estimated cure rate from the Kaplan-Meier curve is greater than 0.5 we choose $k_0 = 0.49$ to proceed. (B.2) is only defined if $p_i < 1 - k_1$. To proceed, we choose the midpoint of $(0, 1-k_1)$ as an estimator for $p_i|_{z_1=1}$. Substituting this value into the logistic link function for p, we obtain initial estimate $\widehat{\beta}_{11} = \log\left(\frac{1-k_1}{1+k_1}\right)$. We naively set $z_2 = \text{Med}(z_2)$ in (B.2) and, since we could reasonably expect the number of initial cells to be lower with lower age, after observing that in our case $\log\left[-2\log(2k_0)\right] < \log\left[-\frac{2}{1-k_1}\log\left(\frac{2k_1}{1+k_1}\right)\right]$, we fix $z_2 = Q_1(z_2)$ in (B.1) and calculate initial estimates for β_{20} and β_{21} :

$$\begin{aligned}
\widehat{\beta}_{21} &= \frac{\log[-2\log(2k_0)] - \log\left[-\frac{2}{1-k_1}\log\left(\frac{2k_1}{1+k_1}\right)\right]}{\text{Med}(z_2) - Q_1(z_2)}, \\
\widehat{\beta}_{20} &= \log\left[-2\log(2k_0)\right] - \widehat{\beta}_{21}\text{Med}(z_2).
\end{aligned}$$

Initial values for lifetime parameters are obtained as in Model 1.1.

- **Model 2.2:** $\theta = e^{\beta_{21} z_2}, p = \frac{e^{\beta_{10}+\beta_{11} z_1}}{1+e^{\beta_{10}+\beta_{11} z_1}}$, where θ contains no intercept term. Denote the cure rate of the i-th patient as q_{0i}.

$$\begin{aligned}
q_{0i|z_1=0} &= (1 - p_{i|z_1=0}) \exp\left[-\exp\left(\beta_{21} z_2\right) p_{i|z_1=0}\right], \\
q_{0i|z_1=1} &= (1 - p_{i|z_1=1}) \exp\left[-\exp\left(\beta_{21} z_2\right) p_{i|z_1=1}\right].
\end{aligned}$$

We obtain estimates for cure rates k_0, k_1 from the Kaplan-Meier plots stratified by EMP such that

$$\widehat{q}_{0i} = \begin{cases} k_0, z_1 = 0 \\ k_1, z_1 = 1, \end{cases}$$

where $k_0, k_1 \in (0, 1)$. We equate the cure rate formulas with the above estimates:

$$(1 - p_{i|z_1=0}) \exp\left[-\exp\left(\beta_{21} z_2\right) p_{i|z_1=0}\right] = k_0,$$

$$(1 - p_{i|z_1=1}) \exp\left[- \exp\left(\beta_{21} z_2\right) p_{i|z_1=1} \right] = k_1,$$

which yields the equations

$$\begin{aligned} \beta_{21} = & \ \tfrac{1}{z_2} \log\left[- \tfrac{1}{p_{i|z_1=0}} \log(\tfrac{k_0}{1-p_{i|z_1=0}}) \right], \\ \beta_{21} = & \ \tfrac{1}{z_2} \log\left[- \tfrac{1}{p_{i|z_1=1}} \log(\tfrac{k_1}{1-p_{i|z_1=1}}) \right]. \end{aligned} \tag{B.3}$$

From the above equations we can see that $p_{i|z_1=0} \in (0, 1-k_0)$ and $p_{i|z_1=1} \in (0, 1-k_1)$. We choose estimates for p as the midpoints of these intervals, namely

$$\begin{aligned} \widehat{p}_{i|z_1=0} = & \ \tfrac{1-k_0}{2} \\ \widehat{p}_{i|z_1=1} = & \ \tfrac{1-k_1}{2}, \end{aligned}$$

which allows computation of initial estimates

$$\begin{aligned} \widehat{\beta}_{10} = & \ \log\left(\tfrac{1-k_0}{1+k_0}\right) \\ \widehat{\beta}_{11} = & \ \log\left(\tfrac{1-k_1}{1+k_1}\right) - \widehat{\beta}_{10}. \end{aligned}$$

Since we chose $p_{i|z_1=0}$ as the midpoint of the intervals of possible values, it is reasonable to set $z_2 = \text{Med}(z_2)$ in (B.3) to solve for initial estimate

$$\widehat{\beta}_{21} = \ \tfrac{1}{\text{Med}(z_2)} \log\left[- \tfrac{2}{1-k_0} \log(\tfrac{2k_0}{1-k_0}) \right].$$

Initial values for lifetime parameters are obtained as in Model 1.1.

APPENDIX C

Appendix C

In this appendix, we present the computational codes which were developed to generate data and implement the EM and SEM algorithms.

C.1 Computational code for interval-censored destructive length biased Poisson cure model

C.1.1 Data generation code

```
LR_int=function(y1,len1,l1){
  if(y1>0 & y1<=l1){
    a=c(.Machine$double.eps,l1)
  }else{
    k=as.integer((y1-l1)/len1)+1
    a=c(l1+((k-1)*len1),l1+(k*len1))
  }
  return(a)
}
sim_data=function(n,b11,b02,b12,g1,g2,delta){
  l=rep(NA,n)#left endpoint of lifetime interval
  r=rep(NA,n)#right endpoint of lifetime interval
  d=rep(NA,n)#censoring indicator
  x1 = rbinom(n,size=1,prob=0.5)
  x2 = runif(n,0.1,20)
  theta = exp(b11*x1)
  p = exp(b02+b12*x2)/(1+exp(b02+b12*x2))
  M=rep(NA,n)
  D=rep(NA,n)
```

```
C=rexp(n,rate=delta)
count=0
for(i in 1:n){
  M[i]=rpois(1,lambda=(theta[i]))
  D[i]=rbinom(1,size=M[i]+1,prob=p[i])
   if(D[i]==0){
     count=count+1
     l[i]=C[i]
     r[i]=1/0
     d[i]=0
   }else{
     y=min(rweibull(D[i],shape=1/g1,scale=1/g2))
     if(min(y,C[i])==C[i]){
       l[i]=C[i]
       r[i]=1/0
       d[i]=0
     }else{
       len=runif(1,0.2,0.7)
       li=runif(1,0,1)
       ans=LR_int(y,len,li)
       l[i]=ans[1]
       r[i]=ans[2]
       d[i]=1
     }
   }
```

```
  }#end of for
  data=data.frame(cbind(l,r,d,x1,x2,M,D))
  return(data)
}
```

C.1.2 EM algorithm code

This section presents the computational code to implement the EM algorithm for covariate configuration denoted by Model 1.1 in Section 2.5

```
EM.DLBP.IC1.1 = function(mydata,tol,maxit,b01.init,
                         b11.init,b12.init,g1.init,g2.init){
  obs_data = mydata[mydata$d==1,]
  cens_data = mydata[mydata$d==0,]
  L_obs = obs_data$l
  R_obs = obs_data$r
  L_cens = cens_data$l
  x1t = obs_data$x1
  x1c = cens_data$x1
  x2t = obs_data$x2
  x2c = cens_data$x2
  x1 = c(x1t,x1c)
  x2 = c(x2t,x2c)
  y = c(R_obs,L_cens)
  d = c(obs_data$d, cens_data$d)
  p.new = matrix(0,ncol=1,nrow=5)
  p.old = matrix(0,ncol=1,nrow=5)
```

```
p.old[1,1] = b01.init
p.old[2,1] = b11.init
p.old[3,1] = b12.init
p.old[4,1] = g1.init
p.old[5,1] = g2.init
count = 0
continue = TRUE
iter = 1
while(continue){
  ########Q function:E step: ###########
  theta.r = exp(p.old[1,1]+p.old[2,1]*x1)
  p.r = exp(p.old[3,1]*x2)/(1+exp(p.old[3,1]*x2))
  F.r = 1 - exp(-(p.old[5,1]*y)^(1/p.old[4,1]))
  S.r = exp(-(p.old[5,1]*y)^(1/p.old[4,1]))
  e.M = d + theta.r*(1-p.r*F.r) +
        (1-p.r*F.r)/(1-p.r*F.r+d/theta.r)
  e.D = theta.r*p.r*S.r + (p.r*S.r)/
        (1-p.r*F.r + d/theta.r) + d
  Q1 = function(par=c(b01,b11)){
    theta = exp(par[1]+par[2]*x1)
    #E-step (r-th iteration)
    dum.r = log(theta)*(e.M-1)
    result1 = sum(dum.r-theta)
    return(-result1)
  }#end of Q1 function
```

```
Q2 = function(par=c(b12)){
   p = exp(par[1]*x2)/(1+exp(par[1]*x2))
   #E-step (r-th iteration)
   dum.r = log(p)*(e.D)+log(1-p)*(e.M-e.D)
   result2 = sum(dum.r)
   return(-result2)
}#end of Q2 function
Q3 = function(par=c(g1,g2)){
   S = exp(-(par[2]*y)^(1/par[1]))
   #E-step (r-th iteration)
   for(i in 1:length(S)){
      if(S[i]==0){
         S[i] = .Machine$double.eps
      }
   }
   dum1.r = log(S)*(e.D-d)
   dum2.r = exp(-(par[2]*L_obs)^(1/par[1])) -
            exp(-(par[2]*R_obs)^(1/par[1]))
   dum3.r = log(dum2.r)
   result3 = sum(dum1.r) + sum(dum3.r)
   return(-result3)
}#end of Q3 function
####### M step ##########
beta1.new = optim(par=c(p.old[1,1],p.old[2,1]),
                  fn=Q1, method="Nelder-Mead")$par
```

```
  beta2.new = optim(par=c(p.old[3,1]),fn=Q2,
                    method="Brent",lower=-5,upper=5)$par
  lambda.new = optim(par=c(p.old[4,1],p.old[5,1]),
                    fn=Q3, method="Nelder-Mead")$par
  p.new = matrix(c(beta1.new,beta2.new,lambda.new))
  iter = iter+1
  continue = max(abs((p.new-p.old)/p.old))>tol&(iter<maxit)
  p.old[1,1] = p.new[1,1]
  p.old[2,1] = p.new[2,1]
  p.old[3,1] = p.new[3,1]
  p.old[4,1] = p.new[4,1]
  p.old[5,1] = p.new[5,1]
}#end of while
out = rep(NA,11)
# calculation of the standard error
std = function(param){
  b01 = param[1]
  b11 = param[2]
  b12 = param[3]
  g1 = param[4]
  g2 = param[5]
  theta.t = exp(b01+b11*x1t)
  theta.c= exp(b01+b11*x1c)
  p.t = exp(b12*x2t)/(1+exp(b12*x2t))
  p.c = exp(b12*x2c)/(1+exp(b12*x2c))
```

```
    F.c = 1 - exp(-(g2*L_cens)^(1/g1))
    F.tl = 1 - exp(-(g2*L_obs)^(1/g1))
    F.tr = 1 - exp(-(g2*R_obs)^(1/g1))
    Sp.c = exp(-theta.c*p.c*F.c)*(1-p.c*F.c)
    Sp.tl = exp(-theta.t*p.t*F.tl)*(1-p.t*F.tl)
    Sp.tr = exp(-theta.t*p.t*F.tr)*(1-p.t*F.tr)
    dum1 = rep(NA,length(Sp.tl))
    for(i in 1:length(dum1)){
      dum1[i] = max((Sp.tl[i]-Sp.tr[i]),.Machine$double.eps)
    }
    out1 = sum(log(Sp.c)) + sum(log(dum1))
    return(-out1)
  }
  hessmat1=hessian(std,c(p.new[1,1],p.new[2,1],p.new[3,1],
                   p.new[4,1],p.new[5,1]), method="Richardson")
  FI=solve(-1*hessmat1)
  std1 = sqrt(FI[1,1])
  std2 = sqrt(FI[2,2])
  std3 = sqrt(FI[3,3])
  std4 = sqrt(FI[4,4])
  std5 = sqrt(FI[5,5])
  out = c(p.new[1,1], p.new[2,1], p.new[3,1], p.new[4,1],
  p.new[5,1], std1, std2, std3, std4, std5, iter)
  return(out)
}
```

C.2 Computational code for interval-censored destructive negative binomial cure mode

C.2.1 Data generation code

This section presents the computational code to generate data with covariate configuration denoted as Model 2 in Section 3.6.

```
LR_int=function(y1,len1,l1){
  if(y1>0 & y1<=l1){
    a=c(.Machine$double.eps,l1)
  }else{
    k=as.integer((y1-l1)/len1)+1
    a=c(l1+((k-1)*len1),l1+(k*len1))
  }
  return(a)
}
sim_data_Wei=function(n,b11,b20,b21,g1,g2,delta,phi){
  l=rep(NA,n)#left endpoint of lifetime interval
  r=rep(NA,n)#right endpoint of lifetime interval
  d=rep(NA,n)#censoring indicator
  x1=rep(NA,n)#covariate: treatment group (1:trt, 0:placebo)
  x2=rep(NA,n)#covariate: tumor thickness
  x1 = rbinom(n,size=1,prob=0.5)
  x2 = runif(n,0.1,20)
  eta=rep(NA,n)
  theta=rep(NA,n)
  p=rep(NA,n)
```

```
D=rep(NA,n)
C=rexp(n,rate=delta)
count=0
for(i in 1:n){
  eta[i]=exp((b11*x1[i]))
  theta[i]=(phi*eta[i])/(1+phi*eta[i])
  p[i]=exp(b20+b21*x2[i])/(1+exp(b20+b21*x2[i]))
  m=rnbinom(1,size=(1/phi),prob=(1/(1+(phi*eta[i]))))
  if(m==0){
    D[i]=0
  }else{
    D[i]=rbinom(1,size=m,prob=p[i])
  }
  if(D[i]==0){
    count=count+1
    l[i]=C[i]
    r[i]=1/0
    d[i]=0
  }else{
    y=min(rweibull(D[i],shape=1/g1,scale=1/g2))
    if(min(y,C[i])==C[i]){
      l[i]=C[i]
      r[i]=1/0
      d[i]=0
    }else{
```

```
            len=runif(1,0.1,0.5)
            li=runif(1,0,1)
            ans=LR_int(y,len,li)
            l[i]=ans[1]
            r[i]=ans[2]
            d[i]=1
          }
        }
    }#end of for
    data=data.frame(cbind(l,r,d,x1,x2))
    return(data)
}
```

C.2.2 EM-PL algorithm code

This section presents the computational code to implement the EM algorithm formulated in Section 3.3 using a profile likelihood search to estimate ϕ and using the covariate configuration denoted as Model 2 in section 3.6.

```
EM.DNB=function(mydata,tol,maxit,increment,
                        b11,b20,b21,g1,g2,phi){
    obs_data=mydata[mydata$d==1,]
    cens_data=mydata[mydata$d==0,]
    L_obs=obs_data$l
    R_obs=obs_data$r
    L_cens=cens_data$l
    x1t=obs_data$x1
```

```
x1c=cens_data$x1
x2t=obs_data$x2
x2c=cens_data$x2
x1 = c(x1t,x1c)
x2 = c(x2t,x2c)
y = c(R_obs,L_cens)
d = c(obs_data$d, cens_data$d)
p.new=matrix(0,ncol=1,nrow=5)
p.old=matrix(0,ncol=1,nrow=5)
p.old[1,1]=b11 + runif(1,0,increment)*abs(b11)
p.old[2,1]=b20 + runif(1,0,increment)*abs(b20)
p.old[3,1]=b21 + runif(1,0,increment)*abs(b21)
p.old[4,1]=g1 + runif(1,0,increment)*abs(g1)
p.old[5,1]=g2 + runif(1,0,increment)*abs(g2)
count=0
continue = TRUE
iter=1
while(continue){
  ########Q function:E step: ###########
  eta.r = exp(p.old[1,1]*x1)
  p.r = exp(p.old[2,1] + p.old[3,1]*x2)/
        (1 + exp(p.old[2,1]+p.old[3,1]*x2))
  F.r = 1 - exp(-(p.old[5,1]*y)^(1/p.old[4,1]))
  S.r = exp(-(p.old[5,1]*y)^(1/p.old[4,1]))
  e.M = d + ((1+phi*d)*eta.r*(1-p.r*F.r))/
```

```
        (1+phi*eta.r*p.r*F.r)
e.D = d + ((1+phi*d)*eta.r*p.r*S.r)/(1+phi*eta.r*p.r*F.r)
Q1 = function(par=c(b11)){
   eta = exp(par[1]*x1)
   #E-step (r-th iteration)
   dum1.r = e.M*log((phi*eta)/(1+phi*eta))
   dum2.r = log(gamma(1/phi)*exp(-(phi*eta)/(1+phi*eta))*
          (1+phi*eta)^(1/phi))
   result1 = sum(dum1.r-dum2.r-(phi*eta)/(1+phi*eta))
   return(-result1)
}#end of Q1 function
Q2 = function(par=c(b20,b21)){
   p = exp(par[1]+par[2]*x2)/(1+exp(par[1]+par[2]*x2))
   #E-step (r-th iteration)
   dum.r = log(p)*(e.D)+log(1-p)*(e.M-e.D)
   result2 = sum(dum.r)
   return(-result2)
}#end of Q2 function
Q3 = function(par=c(g1,g2)){
   S = exp(-(par[2]*y)^(1/par[1]))
   #E-step (r-th iteration)
   for(i in 1:length(S)){
      if(S[i]==0 | is.nan(S[i])){
         S[i]=.Machine$double.eps
      }
```

```
  }
  dum1.r = log(S)*(e.D-d)
  dum2.r = exp(-(par[2]*L_obs)^(1/par[1])) -
                exp(-(par[2]*R_obs)^(1/par[1]))
  dum3.r = log(dum2.r)
  result3 = sum(dum1.r) + sum(dum3.r)
  return(-result3)
}#end of Q3 function
####### M step ##########
beta1.new = tryCatch({optim(par=p.old[1,1],fn=Q1,
                method="Brent",lower=-10,upper=10)$par
},error=function(e){
  beta1.new = c(0)
  return(beta1.new)
}
)
beta2.new = tryCatch({optim(par=c(p.old[2,1],p.old[3,1]),
                      fn=Q2, method="Nelder-Mead")$par
},error=function(e){
  beta2.new = c(0,0)
  return(beta2.new)
}
)
lambda.new=tryCatch({optim(par=c(p.old[4,1],p.old[5,1]),
                      fn=Q3, method="Nelder-Mead")$par
```

```
  },error=function(e){
    lambda.new = c(0,0)
    return(lambda.new)
  }
  )
  p.new = matrix(c(beta1.new,beta2.new,lambda.new))
  if(any(p.new==0)){
    continue = FALSE
    p.new = matrix(c(0,0,0,0,0))
    break
  }
  else{
    iter = iter+1
    continue=max(abs((p.new-p.old)/p.old))>tol&(iter<maxit)
    p.old[1,1]=p.new[1,1]
    p.old[2,1]=p.new[2,1]
    p.old[3,1]=p.new[3,1]
    p.old[4,1]=p.new[4,1]
    p.old[5,1]=p.new[5,1]
  }#end of else
}#end of while
out = rep(NA,12)
if(p.new[1,1]==0 & p.new[2,1]==0 & p.new[3,1]==0
   & p.new[4,1]==0 & p.new[5,1]==0){
  out = rep(0,12)
```

```
}else{
  std = function(param){
    b11 = param[1]
    b20 = param[2]
    b21 = param[3]
    g1 = param[4]
    g2 = param[5]
    eta.t = exp(b11*x1t)
    eta.c = exp(b11*x1c)
    theta.t = (phi*eta.t)/(1+phi*eta.t)
    theta.c = (phi*eta.c)/(1+phi*eta.c)
    p.t = exp(b20+b21*x2t)/(1+exp(b20+b21*x2t))
    p.c = exp(b20+b21*x2c)/(1+exp(b20+b21*x2c))
    F.c = 1 - exp(-(g2*L_cens)^(1/g1))
    F.tl = 1 - exp(-(g2*L_obs)^(1/g1))
    F.tr = 1 - exp(-(g2*R_obs)^(1/g1))
    Sp.c = (1 + phi*eta.c*p.c*F.c)^(-(1/phi))
    Sp.tl = (1 + phi*eta.t*p.t*F.tl)^(-(1/phi))
    Sp.tr = (1 + phi*eta.t*p.t*F.tr)^(-(1/phi))
    dum1=rep(NA,length(Sp.tl))
    for(i in 1:length(dum1)){
      dum1[i]=max((Sp.tl[i]-Sp.tr[i]),.Machine$double.eps)
    }
    out1 = sum(log(Sp.c)) + sum(log(dum1))
    return(out1)
```

```
    }
    hessmat1=hessian(std,c(p.new[1,1],p.new[2,1],
    p.new[3,1],p.new[4,1],p.new[5,1]),method="Richardson")
    FI=tryCatch({solve(-1*hessmat1)
    },error=function(e){
      FI=matrix(0,5,5)
      return(FI)
    }
    )
    std1 = sqrt(FI[1,1])
    std2 = sqrt(FI[2,2])
    std3 = sqrt(FI[3,3])
    std4 = sqrt(FI[4,4])
    std5 = sqrt(FI[5,5])
    out = c(p.new[1,1],p.new[2,1],p.new[3,1],p.new[4,1],
    p.new[5,1],phi,std1,std2,std3,std4,std5,iter)
  }#end of else
  return(out)
}
#profile likelihood approach to estimate phi
grid.phi = function(mydata,tol,maxit,increment,param){
  grid=seq(0.05, 2.05, by=0.05)
  continue=TRUE
  count=1
  while(continue){
```

```
phi.curr = grid[1]
phi.maxim = phi.curr
mle.curr = EM.DNB(mydata, tol, maxit, increment,
b11=param[1], b20=param[2], b21=param[3], g1=param[4],
g2=param[5], phi=phi.curr)
mle.maxim = mle.curr
log.maxim = std(mydata, param=mle.curr[1:6])
for(i in 2:length(grid)){
  phi.curr=grid[i]
  mle.curr = EM.DNB(mydata,tol,maxit,increment,
  b11=param[1],b20=param[2],b21=param[3],
  g1=param[4],g2=param[5],phi=phi.curr)
  log.curr=std(mydata,param=mle.curr[1:6])
  if(log.curr > log.maxim){
    phi.maxim = phi.curr
    log.maxim = log.curr
    mle.maxim = mle.curr
  }
}
if(phi.maxim==grid[1] | phi.maxim == grid[length(grid)]){
  if(phi.maxim==0.01){
    continue = FALSE
  }
  if(phi.maxim == grid[1]){
    count=count+1
```

```
      grid = grid-2
      if(grid[1] <= 0){
        grid = seq(0.01,1.01,by=0.05)
      }
    }
    if(phi.maxim == grid[length(grid)]){
      grid = grid+2
    }
    if(count>2){
      continue=FALSE
      count=1
    }
  }
  else{
    continue = FALSE
  }
 }
 return(mle.maxim)
}
#calculation of SE for EM algorithm estimates
std = function(mydata,param){
 b11 = param[1]
 b20 = param[2]
 b21 = param[3]
 g1 = param[4]
```

```
g2 = param[5]
phi = param[6]
obs_data = mydata[mydata$d==1,]
cens_data = mydata[mydata$d==0,]
L_obs = obs_data$l
R_obs = obs_data$r
L_cens = cens_data$l
x1t = obs_data$x1
x1c = cens_data$x1
x2t = obs_data$x2
x2c = cens_data$x2
x1 = c(x1t,x1c)
x2 = c(x2t,x2c)
y = c(R_obs,L_cens)
d = c(obs_data$d, cens_data$d)
eta.t = exp(b11*x1t)
eta.c = exp(b11*x1c)
p.t = exp(b20+b21*x2t)/(1+exp(b20+b21*x2t))
p.c = exp(b20+b21*x2c)/(1+exp(b20+b21*x2c))
F.c = 1 - exp(-(g2*L_cens)^(1/g1))
F.tl = 1 - exp(-(g2*L_obs)^(1/g1))
F.tr = 1 - exp(-(g2*R_obs)^(1/g1))
Sp.c = (1 + phi*eta.c*p.c*F.c)^(-(1/phi))
Sp.tl = (1 + phi*eta.t*p.t*F.tl)^(-(1/phi))
Sp.tr = (1 + phi*eta.t*p.t*F.tr)^(-(1/phi))
```

```
  dum1=rep(NA, length(Sp.tl))
  for(i in 1:length(dum1)){
    dum1[i] = max((Sp.tl[i]-Sp.tr[i]),.Machine$double.eps)
  }
  out1 = sum(log(Sp.c)) + sum(log(dum1))
  return(out1)
}
```

C.2.3 EM-SM algorithm code

This section presents the computational code to implement the EM algorithm formulated in Section 3.3 using the mixture representation of the negative binomial distribution and using the covariate configuration denoted as Model 2 in section 3.6.

```
EM.DNB.IC.MC=function(mydata, tol, maxit,
                      increment, b11, b20, b21, g1, g2, phi){
  obs_data=mydata[mydata$d==1,]
  cens_data=mydata[mydata$d==0,]
  L_obs=obs_data$l
  R_obs=obs_data$r
  L_cens=cens_data$l
  x1t=obs_data$x1
  x1c=cens_data$x1
  x2t=obs_data$x2
  x2c=cens_data$x2
  x1 = c(x1t, x1c)
  x2 = c(x2t, x2c)
```

```
y = c(R_obs,L_cens)
d = c(obs_data$d, cens_data$d)
p.new=matrix(0,ncol=1,nrow=6)
p.old=matrix(0,ncol=1,nrow=6)
p.old[1,1]=b11+increment*runif(1,0,1)*abs(b11)
p.old[2,1]=phi+increment*runif(1,0,1)*abs(phi)
p.old[3,1]=b20+increment*runif(1,0,1)*abs(b20)
p.old[4,1]=b21+increment*runif(1,0,1)*abs(b21)
p.old[5,1]=g1+increment*runif(1,0,1)*abs(g1)
p.old[6,1]=g2+increment*runif(1,0,1)*abs(g2)
count=0
continue = TRUE
iter=1
while(continue){
  ########Q function:E step: ###########
  phi.r = p.old[2,1]
  eta.r = exp(p.old[1,1]*x1)
  p.r = exp(p.old[3,1] + p.old[4,1]*x2)/
        (1 + exp(p.old[3,1]+p.old[4,1]*x2))
  F.r = 1 - exp(-(p.old[6,1]*y)^(1/p.old[5,1]))
  S.r = exp(-(p.old[6,1]*y)^(1/p.old[5,1]))
  e.M = d + ((1+phi.r*d)*eta.r*(1-p.r*F.r))/
        (1+phi.r*eta.r*p.r*F.r)
  e.D = d + ((1+phi.r*d)*eta.r*p.r*S.r)/
        (1+phi.r*eta.r*p.r*F.r)
```

```
e.U = ((1+phi.r*d)*eta.r)/(1+phi.r*eta.r*p.r*F.r)
e.logU = digamma(1/phi.r + d) + log(phi.r)
    +log(eta.r) - log(1+phi.r*eta.r*p.r*F.r)
Q1 = function(par=c(b11,phi)){
  eta = exp(par[1]*x1)
  phi = par[2]
  #E-step (r-th iteration)
  for(i in 1:length(phi)){
    if(phi[i]<=0 | is.nan(phi[i])){
      phi[i]=.Machine$double.eps
    }
  }
  dum1.r = (1/phi)*(e.logU - log(phi) -
  log(eta)) - e.U/(phi*eta)
  result1 = sum(dum1.r)-length(x1)*
            log(gamma(1/phi))
  return(-result1)
}#end of Q1 function
Q2 = function(par=c(b20,b21)){
  p = exp(par[1]+par[2]*x2)/(1+exp(par[1]+par[2]*x2))
  #E-step (r-th iteration)
  dum.r = log(p)*(e.D)+log(1-p)*(e.M-e.D)
  result2 = sum(dum.r)
  return(-result2)
}#end of Q2 function
```

```
Q3 = function(par=c(g1,g2)){
  S = exp(-(par[2]*y)^(1/par[1]))
  #E-step (r-th iteration)
  for(i in 1:length(S)){
    if(S[i]==0 | is.nan(S[i])){
      S[i]=.Machine$double.eps
    }
  }
  dum1.r = log(S)*(e.D-d)
  dum2.r = exp(-(par[2]*L_obs)^(1/par[1])) -
    exp(-(par[2]*R_obs)^(1/par[1]))
  dum3.r = log(dum2.r)
  result3 = sum(dum1.r) + sum(dum3.r)
  return(-result3)
}#end of Q3 function
####### M step ##########
beta1.new = tryCatch({optim(par=c(p.old[1,1],
    p.old[2,1]),fn=Q1, method="Nelder-Mead")$par
},error=function(e){
  beta1.new = c(0,0)
  return(beta1.new)
}
)
beta2.new = tryCatch({optim(par=c(p.old[3,1],
    p.old[4,1]),fn=Q2, method="Nelder-Mead")$par
```

```
},error=function(e){
  beta2.new = c(0,0)
  return(beta2.new)
}
)
lambda.new = tryCatch({optim(par=c(p.old[5,1],
    p.old[6,1]),fn=Q3, method="Nelder-Mead")$par
},error=function(e){
  lambda.new = c(0,0)
  return(lambda.new)
}
)
p.new = matrix(c(beta1.new,beta2.new,lambda.new))
if(p.new[1,1]==0 | p.new[2,1]==0 | p.new[3,1]==0 |
    p.new[4,1]==0 | p.new[5,1]==0 | p.new[6,1]==0){
  continue = FALSE
  p.new = matrix(c(0,0,0,0,0,0))
  break
}else{
iter = iter+1
continue= max(abs((p.new-p.old)/p.old))>tol&(iter<maxit)
p.old[1,1]=p.new[1,1]
p.old[2,1]=p.new[2,1]
p.old[3,1]=p.new[3,1]
p.old[4,1]=p.new[4,1]
```

```
    p.old[5,1]=p.new[5,1]
    p.old[6,1]=p.new[6,1]
    }#end of else
  }#end of while
  out = rep(NA,12)
  if(p.new[1,1]==0 & p.new[2,1]==0 & p.new[3,1]==0 &
       p.new[4,1]==0 & p.new[5,1]==0){
    out = rep(0,12)
  }else{
    std = function(param){
      b11 = param[1]
      phi = param[2]
      b20 = param[3]
      b21 = param[4]
      g1 = param[5]
      g2 = param[6]
      eta.t = exp(b11*x1t)
      eta.c = exp(b11*x1c)
      theta.t = (phi*eta.t)/(1+phi*eta.t)
      theta.c = (phi*eta.c)/(1+phi*eta.c)
      p.t = exp(b20+b21*x2t)/(1+exp(b20+b21*x2t))
      p.c = exp(b20+b21*x2c)/(1+exp(b20+b21*x2c))
      F.c = 1 - exp(-(g2*L_cens)^(1/g1))
      F.tl = 1 - exp(-(g2*L_obs)^(1/g1))
      F.tr = 1 - exp(-(g2*R_obs)^(1/g1))
```

```
    Sp.c = (1 + phi*eta.c*p.c*F.c)^(-(1/phi))
    Sp.tl = (1 + phi*eta.t*p.t*F.tl)^(-(1/phi))
    Sp.tr = (1 + phi*eta.t*p.t*F.tr)^(-(1/phi))
    dum1=rep(NA,length(Sp.tl))
    for(i in 1:length(dum1)){
      dum1[i]=max((Sp.tl[i]-Sp.tr[i]),.Machine$double.eps)
    }
    out1 = sum(log(Sp.c)) + sum(log(dum1))
    return(out1)
  }
  hessmat1=hessian(std,c(p.new[1,1],p.new[2,1],p.new[3,1],
  p.new[4,1],p.new[5,1],p.new[6,1]),method="Richardson")
  FI=solve(-1*hessmat1)
  std1 = sqrt(FI[1,1])
  std2 = sqrt(FI[2,2])
  std3 = sqrt(FI[3,3])
  std4 = sqrt(FI[4,4])
  std5 = sqrt(FI[5,5])
  std6 = sqrt(FI[6,6])
  out=c(p.new[1,1],p.new[2,1],p.new[3,1],p.new[4,1],
  p.new[5,1],p.new[6,1],std1,std2,std3,std4,std5,std6,iter)
 }#end of else
 return(out)
}
```

C.2.4 SEM algorithm code

This section presents the computational code to implement the SEM algorithm formulated in Section 3.4

```
SEM.DNB=function(mydata, burn.it, total.it,
                 increment, b11, b20, b21, g1, g2, phi){
  obs.data=mydata[mydata$d==1,]
  cens.data=mydata[mydata$d==0,]
  L.obs=obs.data$l
  R.obs=obs.data$r
  L.cens=cens.data$l
  x1t=obs.data$x1
  x1c=cens.data$x1
  x2t=obs.data$x2
  x2c=cens.data$x2
  x1 = c(x1t, x1c)
  x2 = c(x2t, x2c)
  y = c(R.obs, L.cens)
  d = c(obs.data$d, cens.data$d)
  #initializing the estimates for each iteration
  par.est=matrix(NA, nrow=total.it, ncol=6)
  #Step 1 start with initial guess for par
  # par.est is initial guess for par
  par.est[1,1]=b11 + runif(1,0,increment)*abs(b11)
  par.est[1,2]=b20 + runif(1,0,increment)*abs(b20)
  par.est[1,3]=b21 + runif(1,0,increment)*abs(b21)
```

```
par.est[1,4]=g1 + runif(1,0,increment)*abs(g1)
par.est[1,5]=g2 + runif(1,0,increment)*abs(g2)
par.est[1,6]=phi + runif(1,0,increment)*abs(phi)
#initializing final SEM estimates & SE
par.final=rep(NA,6)
std.b11=NA
std.b20=NA
std.b21=NA
std.g1=NA
std.g2=NA
std.phi=NA
#count allows for Step 4 (iterate R=total.it times)
count = 1
count.bad = 0
continue = TRUE
while(continue){
  #Step 2 replace missing m, d by random gen value
  M.gen = rep(0,length(y))
  D.gen = rep(0,length(y))
  eta.est = exp(par.est[count,1]*x1)
  p.est = exp(par.est[count,2]+par.est[count,3]*x2)/
  (1 + exp(par.est[count,2] + par.est[count,3]*x2))
  S.est = exp(-(par.est[count,5]*y)^(1/par.est[count,4]))
  F.est = 1 - S.est
  if(par.est[count,6]<=0|is.na(par.est[count,6])){
```

```
  par.est[count,6] = .Machine$double.eps
}
phi.est = par.est[count,6]
for(i in 1:length(y)){
  M.gen[i] = rnbinom(n=1, size = (phi.est^(-1)+d[i]),
  prob = ((1+phi.est*eta.est[i]*p.est[i]*F.est[i])/
  (1 + phi.est*eta.est[i])))+d[i]
  if(M.gen[i]==d[i]){
    D.gen[i]=d[i]
  }else{
     if(p.est[i]<=0){
     print(paste0(p.est[i],"invalid activation prob"))}
     if(M.gen[i]<d[i] | M.gen[i] < 1){
     print(paste0(M.gen[i],"invalid size for D"))}
     D.gen[i] = rbinom(n=1, size=M.gen[i]-d[i],
     prob = (p.est[i]*S.est[i]/
     (1-p.est[i]*F.est[i])))+d[i]
  }
}
#E-step (r-th iteration)
Q1 = function(par=c(b11,phi)){
  eta = exp(par[1]*x1)
  dum.r = suppressWarnings(par[2]*eta)
  for(i in 1:length(x1)){
     if(dum.r[i]<=0){
```

```
      dum.r[i]= .Machine$double.eps
    }
  }
  dum1.r = lgamma((1/par[2])+M.gen)
  dum2.r = (1/par[2])*log(1+dum.r)
  dum3.r = M.gen*(log((dum.r))-log(1+dum.r))
  dum4.r = length(x1)*lgamma(1/par[2])
  result1 = sum(dum1.r - dum2.r + dum3.r) - dum4.r
  for(i in 1:length(x1)){
    if(is.nan(dum1.r[i])|is.nan(dum2.r[i])|
    is.nan(dum3.r[i])){result1 = Inf}
  }
  return(-result1)
}#end of Q1 function
Q2 = function(par=c(b20,b21)){
  p = exp(par[1]+par[2]*x2)/(1+exp(par[1]+par[2]*x2))
  dum.r = log(p)*(D.gen)+log(1-p)*(M.gen-D.gen)
  result2 = sum(dum.r)
  return(-result2)
}#end of Q2 function
Q3 = function(par=c(g1,g2)){
  S = exp(-(par[2]*y)^(1/par[1]))
  for(i in 1:length(S)){
    if(S[i]==0 | is.nan(S[i])){
      S[i]=.Machine$double.eps
```

```
    }
  }
  dum1.r = log(S)*(D.gen - d)
  dum2.r = rep(NA, length(L.obs))
  for(i in 1:length(L.obs)){
    dum2.r[i] = max(0,exp(-(par[2]*L.obs[i])^(1/par[1]))
      - exp(-(par[2]*R.obs[i])^(1/par[1])))
  }
  dum3.r = log(dum2.r)
  result3 = sum(dum1.r) + sum(dum3.r)
  return(-result3)
}#end of Q3 function
####### M step ##########
#Step 3 maximize Q1,Q2,Q3 using M.gen,D.gen
beta1.phi.new = tryCatch({optim(par=c(par.est[count,1],
    par.est[count,6]),fn=Q1,method="Nelder-Mead")$par
},error=function(e){
  beta1.phi.new = c(0,0)
  return(beta1.phi.new)
}
)
beta2.new = tryCatch({optim(par=c(par.est[count,2],
    par.est[count,3]),fn=Q2,method="Nelder-Mead")$par
},error=function(e){
  beta2.new = c(0,0)
```

```
  return(beta2.new)
}
)
lambda.new = tryCatch({optim(par=c(par.est[count,4],
          par.est[count,5]),fn=Q3,method="Nelder-Mead")$par
},error=function(e){
  lambda.new = c(0,0)
  return(lambda.new)
}
)
par.est.optim = c(beta1.phi.new[1], beta2.new,
                  lambda.new, beta1.phi.new[2])
outlier = FALSE #check if param estimate is outlier
#check if par.est is suitable
if(0%in%par.est.optim | par.est.optim[4]<0 |
   par.est.optim[5]<0 | par.est.optim[6]<0){
  count.bad = count.bad + 1
  print(paste0(count.bad," bad runs"))
}else{
  par.est[count,] = par.est.optim
  count=min(count+1,total.it)
  par.est[count,] = par.est.optim
}
if(count.bad>5){
  count=total.it
```

```
    return(rep(0,12))
  }
  continue = count < total.it
}
#Step 5 calculate MLE (by log-lik method)
std = function(param){  # observed log-likelihood
  b11 = param[1]
  b20 = param[2]
  b21 = param[3]
  g1 = param[4]
  g2 = param[5]
  phi = param[6]
  eta.t = exp(b11*x1t)
  eta.c = exp(b11*x1c)
  theta.t = (phi*eta.t)/(1+phi*eta.t)
  theta.c = (phi*eta.c)/(1+phi*eta.c)
  p.t = exp(b20+b21*x2t)/(1+exp(b20+b21*x2t))
  p.c = exp(b20+b21*x2c)/(1+exp(b20+b21*x2c))
  F.c = 1 - exp(-(g2*L.cens)^(1/g1))
  F.tl = 1 - exp(-(g2*L.obs)^(1/g1))
  F.tr = 1 - exp(-(g2*R.obs)^(1/g1))
  Sp.c = (1 + phi*eta.c*p.c*F.c)^(-(1/phi))
  Sp.tl = (1 + phi*eta.t*p.t*F.tl)^(-(1/phi))
  Sp.tr = (1 + phi*eta.t*p.t*F.tr)^(-(1/phi))
  dum1=rep(NA,length(Sp.tl))
```

```
  for(i in 1:length(dum1)){
    dum1[i]=max((Sp.tl[i]-Sp.tr[i]),.Machine$double.eps)
  }
  out1 = sum(log(Sp.c)) + sum(log(dum1))
  return(out1)
}
log.lik = rep(NA, total.it-burn.it)
for(j in 1:length(log.lik)){
  log.lik[j]=std(param=par.est[burn.it+j,])
}
index = which.max(log.lik)
par.final = par.est[burn.it+index,]
hessmat.SEM = hessian(std, par.final, method="Richardson")
FI.SEM = tryCatch({ solve(-1*hessmat.SEM)
},error=function(e){
  FI.SEM = matrix(c(rep(0,36)),nrow=6,ncol=6)
  return(FI.SEM)
}
)
if(suppressWarnings(any(FI.SEM)==0|is.na(sqrt(FI.SEM[1,1]))
   | is.na(sqrt(FI.SEM[2,2])) | is.na(sqrt(FI.SEM[3,3])) |
   is.na(sqrt(FI.SEM[4,4])) | is.na(sqrt(FI.SEM[5,5])) |
   is.na(sqrt(FI.SEM[6,6])))){
  par.final=c(0,0,0,0,0,0)
  std.b11=0
```

```
    std.b20=0
    std.b21=0
    std.g1=0
    std.g2=0
    std.phi=0
  }else{
    std.b11 = sqrt(FI.SEM[1,1])
    std.b20 = sqrt(FI.SEM[2,2])
    std.b21 = sqrt(FI.SEM[3,3])
    std.g1 = sqrt(FI.SEM[4,4])
    std.g2 = sqrt(FI.SEM[5,5])
    std.phi = sqrt(FI.SEM[6,6])
  }
  result = c(par.final, std.b11, std.b20,
             std.b21, std.g1, std.g2, std.phi)
  return(result)
}#end of the SEM function
```

C.3 Computational code for interval-censored Box-Cox transformation cure model

C.3.1 Data generation for $\alpha \epsilon (0, 1]$

```
  LR_int=function(y1,len1,l1){
  if(y1>0 & y1<=l1){
    a = c(.Machine$double.eps,l1)
  }else{
    k = as.integer((y1-l1)/len1)+1
```

```
    a = c(l1+((k-1)*len1),l1+(k*len1))
  }
  return(a)
}
data_gen_BC=function(n,alpha,b0,b1,b2,g1,g2,lambda,cenrate){
  L=rep(NA,n)
  R=rep(NA,n)
  d=rep(NA,n)
  x1=rbinom(n=n,size=1,prob=0.5)
  x2=runif(n,min = 0.1,max = 20)
  phi=exp(b0+(b1*x1)+(b2*x2))/(1+
   (alpha*exp(b0+(b1*x1)+(b2*x2))))
  U=runif(n,min=0,max=1)
  C=rexp(n,rate=cenrate)
  p0 = (1-(alpha*phi))^(1/alpha)
  count.obs=0
  count.cure=0
  for(i in 1:n){
    if(U[i]<=p0[i]){
      L[i]=C[i]
      R[i]=Inf
      d[i]=0
      count.cure=count.cure+1
    }else{
      U1 = runif(1,min=0,max=1)
```

```
      y=(-exp(-(g1*x1[i])-(g2*x2[i]))*log(((alpha*phi[i])
      +((p0[i]+((1-p0[i])*U1))^alpha)-1)/
      (alpha*phi[i])))^lambda
      t=min(y,C[i])
      if(t==C[i]){
        L[i]=C[i]
        R[i]=Inf
        d[i]=0
      }else{
        len=runif(1,0.1,0.5)
        l=runif(1,0,1)
        ans=LR_int(t,len,l)
        L[i]=ans[1]
        R[i]=ans[2]
        d[i]=1
        count.obs = count.obs + 1
      }# end of inner else
    }# end of outer else
  }# end of for
  return(data.frame(L,R,d,x1,x2))
}
```

C.3.2 Data generation for $\alpha = 0$

```
data_0_BC=function(n,b0,b1,b2,g1,g2,lambda,cenrate){
  L=rep(NA,n)
```

```
R=rep(NA,n)
d=rep(NA,n)
x1 = rbinom(n=n, size=1, prob=0.5)
x2=runif(n,min = 0.1,max = 20)
for(i in 1:n){
  if(x2[i]<=0){
    x2[i] = .Machine$double.eps
  }
}
phi = exp(b0 + (b1*x1) + (b2*x2))
U = runif(n, min=0, max=1)
C = rexp(n, rate=cenrate)
p0 = exp(-phi)
count.obs = 0
count.cure = 0
for(i in 1:n){
  if(U[i]<=p0[i]){
    L[i] = C[i]
    R[i] = Inf
    d[i] = 0
    count.cure=count.cure+1
  }else{
    U1 = runif(1,min=0,max=1)
    y = (-exp(-(g1*x1[i])-(g2*x2[i]))*log(1+(exp(-b0-
        (b1*x1[i])-(b2*x2[i]))*log(p0[i]+
```

```
            ((1-p0[i])*U1)))))^lambda
        t = min(y,C[i])
        if(t==C[i]){
          L[i] = C[i]
          R[i] = Inf
          d[i] = 0
        }else{
          len = runif(1,0.1,0.5)
          l = runif(1,0,1)
          ans = LR_int(t,len,l)
          L[i] = ans[1]
          R[i] = ans[2]
          d[i] = 1
          count.obs = count.obs + 1
        }# end of inner else
      }# end of outer else
    }# end of for
    return(data.frame(L,R,d,x1,x2))
}
```

C.3.3 EM algorithm for interval-censored Box-Cox transformation cure model for $\alpha \epsilon (0, 1]$

```
BC_gen_EM_Wei=function(mydata,tol,maxit,increment,
                       alpha,b0,b1,b2,g1,g2,lambda){
  data_obs = mydata[mydata$d==1,]
```

```
data_cens = mydata[mydata$d==0,]
L_obs = data_obs$L
R_obs = data_obs$R
L_cens = data_cens$L
x1t = data_obs$x1
x1c = data_cens$x1
x2t = data_obs$x2
x2c = data_cens$x2
x1 = c(x1t,x1c)
x2 = c(x2t,x2c)
pnew = rep(0,7)
pold = rep(0,7)
pold[1] = runif(1,b0-increment*abs(b0),
          b0+increment*abs(b0))
pold[2] = runif(1,b1-increment*abs(b1),
          b1+increment*abs(b1))
pold[3] = runif(1,b2-increment*abs(b2),
          b2+increment*abs(b2))
pold[4] = runif(1,g1-increment*abs(g1),
          g1+increment*abs(g1))
pold[5] = runif(1,g2-increment*abs(g2),
          g2+increment*abs(g2))
pold[6] = runif(1,lambda-increment*abs(lambda),
          lambda+increment*abs(lambda))
if(alpha==0){
```

```
    pold[7]=increment*runif(1,0,1)
}else{pold[7]=min(alpha+increment*runif(1,0,1),1)}
continue = TRUE
iter = 1
while(continue){
  #Q function:E step:
  S0.L.obs = exp(-L_obs^(1/pold[6]))
  S0.R.obs = exp(-R_obs^(1/pold[6]))
  S0.L.cens = exp(-L_cens^(1/pold[6]))
  phi.cens.r = exp(pold[1]+pold[2]*x1c+pold[3]*x2c)/
  (1+pold[7]*exp(pold[1]+pold[2]*x1c+pold[3]*x2c))
  F.L.cens.r = 1-S0.L.cens^(exp(pold[4]*x1c+pold[5]*x2c))
  Sp.L.cens.r = (1-pold[7]*phi.cens.r*
                 (F.L.cens.r))^(1/pold[7])
  p0.r= (1-pold[7]*phi.cens.r)^(1/pold[7])
  Su.r = (Sp.L.cens.r-p0.r)/(1-p0.r)
  e.W = (Sp.L.cens.r-p0.r)/Sp.L.cens.r
  #E step
  Q = function(par=c(b0,b1,b2,g1,g2,lambda,alpha)){
    S0.L.obs.est = exp(-L_obs^(1/par[6]))
    S0.R.obs.est = exp(-R_obs^(1/par[6]))
    S0.L.cens.est = exp(-L_cens^(1/par[6]))
    phi.obs.est = (exp(par[1]+par[2]*x1t+par[3]*x2t))/
    (1+par[7]*exp(par[1]+par[2]*x1t+par[3]*x2t))
    phi.cens.est = (exp(par[1]+par[2]*x1c+par[3]*x2c))/
```

```
  (1+par[7]*exp(par[1]+par[2]*x1c+par[3]*x2c))
  F.L.obs.est = 1-S0.L.obs.est^
                (exp(par[4]*x1t+par[5]*x2t))
  F.R.obs.est = 1-S0.R.obs.est^
                (exp(par[4]*x1t+par[5]*x2t))
  F.L.cens.est = 1-S0.L.cens.est^
                 (exp(par[4]*x1c+par[5]*x2c))
  Sp.L.obs.est = (1-par[7]*phi.obs.est*
                 F.L.obs.est)^(1/par[7])
  Sp.R.obs.est = (1-par[7]*phi.obs.est*
                 F.R.obs.est)^(1/par[7])
  Sp.L.cens.est = (1-par[7]*phi.cens.est*
                   F.L.cens.est)^(1/par[7])
  p0.est = (1-par[7]*phi.cens.est)^(1/par[7])
  Su.est = (Sp.L.cens.est-p0.est)/(1-p0.est)
  for(i in 1:length(p0.est)){
    p0.est[i] = max(p0.est[i], .Machine$double.eps)
    Su.est[i] = max(Su.est[i], .Machine$double.eps)
  }
  dum = Sp.L.obs.est - Sp.R.obs.est
  dum1 = log(dum)
  dum2 = (1-e.W)*log(p0.est) + e.W*log((1-p0.est)*Su.est)
  result = sum(dum1)+sum(dum2)
  return(-result)
} #end of Q Function
```

```
  #M step
  pnew = tryCatch({optim(par=c(pold),fn=Q,
          method="Nelder-Mead")$par
  },error=function(e){
    pnew = c(0,0,0,0,0,0,0)
    return(pnew)
  }
  )
  if(0 %in% pnew){
    continue = FALSE
    pnew = c(0,0,0,0,0,0,0)
    break
  }else{
    iter = iter+1
    continue = max(abs((pnew-pold)/pold))>tol
    pold = pnew
  }#end of else
}#end of while
out = rep(NA,14)
#if error, don't calculate SE and return all zeros
if(0 %in% pnew){
  out = rep(0,15)
}else{
  std = function(param){
    b0 = param[1]
```

```
b1 = param[2]
b2 = param[3]
g1 = param[4]
g2 = param[5]
lambda = param[6]
alpha = param[7]
phi.t = exp(b0+b1*x1t+b2*x2t)/
        (1+alpha*exp(b0+b1*x1t+b2*x2t))
phi.c = exp(b0+b1*x1c+b2*x2c)/
        (1+alpha*exp(b0+b1*x1c+b2*x2c))
S0.L.cens = exp(-L_cens^(1/lambda))
S0.L.obs = exp(-L_obs^(1/lambda))
S0.R.obs = exp(-R_obs^(1/lambda))
F.c = 1 - S0.L.cens^(exp(g1*x1c+g2*x2c))
F.tl = 1 - S0.L.obs^(exp(g1*x1t+g2*x2t))
F.tr = 1 - S0.R.obs^(exp(g1*x1t+g2*x2t))
Sp.c = (1-alpha*phi.c*F.c)^(1/alpha)
Sp.tl = (1-alpha*phi.t*F.tl)^(1/alpha)
Sp.tr = (1-alpha*phi.t*F.tr)^(1/alpha)
dum1 = rep(NA,length(Sp.tl))
for(i in 1:length(dum1)){
  dum1[i] = max((Sp.tl[i]-Sp.tr[i]),
            .Machine$double.eps)
}
out1 = sum(log(Sp.c)) + sum(log(dum1))
```

```
  return(out1)
}
hessmat1 = hessian(std,c(pnew[1],pnew[2],pnew[3],pnew[4],
            pnew[5],pnew[6],pnew[7]),method="Richardson")
FI = tryCatch({solve(-1*hessmat1)
},error=function(e){
  FI=matrix(0,nrow=6,ncol=6)
}
)
if(any(is.na(FI))){
  out = c(pnew[1],pnew[2],pnew[3],pnew[4],pnew[5],
          pnew[6],pnew[7],0,0,0,0,0,0,0,0)
}else if(FI[1,1]<0 | FI[2,2]<0 | FI[3,3]<0|
        FI[4,4]<0 | FI[5,5]<0 | FI[6,6]<0 | FI[7,7]<0){
  out = c(pnew[1],pnew[2],pnew[3],pnew[4],pnew[5],
          pnew[6],pnew[7],0,0,0,0,0,0,0,0)
}else{
  std1 = sqrt(FI[1,1])
  std2 = sqrt(FI[2,2])
  std3 = sqrt(FI[3,3])
  std4 = sqrt(FI[4,4])
  std5 = sqrt(FI[5,5])
  std6 = sqrt(FI[6,6])
  std7 = sqrt(FI[7,7])
  std = std(pnew)
```

```
      out = c(pnew[1],pnew[2],pnew[3],pnew[4],pnew[5],
      pnew[6],pnew[7],std1,std2,std3,std4,std5,std6,std7,std)
    }
  }#end of else
  return(out)
}
```

C.3.4 EM algorithm for interval-censored Box-Cox transformation cure model for $\alpha = 0$

```
BC_0_EM_Wei=function(mydata,tol,maxit,
                     increment,b0,b1,b2,g1,g2,lambda){
  data_obs = mydata[mydata$d==1,]
  data_cens = mydata[mydata$d==0,]
  L_obs = data_obs$L
  R_obs = data_obs$R
  L_cens = data_cens$L
  x1t = data_obs$x1
  x1c = data_cens$x1
  x2t = data_obs$x2
  x2c = data_cens$x2
  x1 = c(x1t,x1c)
  x2 = c(x2t,x2c)
  pnew = rep(0,6)
  pold = rep(0,6)
  pold[1] = runif(1,b0-increment*abs(b0),
```

```
            b0+increment*abs(b0))
pold[2] = runif(1,b1-increment*abs(b1),
            b1+increment*abs(b1))
pold[3] = runif(1,b2-increment*abs(b2),
            b2+increment*abs(b2))
pold[4] = runif(1,g1-increment*abs(g1),
            g1+increment*abs(g1))
pold[5] = runif(1,g2-increment*abs(g2),
            g2+increment*abs(g2))
pold[6] = runif(1,lambda-increment*abs(lambda),
            lambda+increment*abs(lambda))
continue = TRUE
count = 0
iter = 1
while(continue){
  S0.L.obs = exp(-L_obs^(1/pold[6]))
  S0.R.obs = exp(-R_obs^(1/pold[6]))
  S0.L.cens = exp(-L_cens^(1/pold[6]))
  phi.cens.r = exp(pold[1]+pold[2]*x1c+pold[3]*x2c)
  F.L.cens.r = 1-S0.L.cens^(exp(pold[4]*x1c+pold[5]*x2c))
  Sp.L.cens.r = exp(-phi.cens.r*F.L.cens.r)
  p0.r = exp(-phi.cens.r)
  Su.r = (Sp.L.cens.r-p0.r)/(1-p0.r)
  e.W = (Sp.L.cens.r-p0.r)/Sp.L.cens.r
  #E step
```

```
Q = function(par=c(b0,b1,b2,g1,g2,lambda)){
  S0.L.obs.est = exp(-L_obs^(1/par[6]))
  S0.R.obs.est = exp(-R_obs^(1/par[6]))
  S0.L.cens.est = exp(-L_cens^(1/par[6]))
  phi.obs.est = exp(par[1]+par[2]*x1t+par[3]*x2t)
  phi.cens.est = exp(par[1]+par[2]*x1c+par[3]*x2c)
  F.L.obs.est = 1-S0.L.obs.est^
                (exp(par[4]*x1t+par[5]*x2t))
  F.R.obs.est = 1-S0.R.obs.est^
                (exp(par[4]*x1t+par[5]*x2t))
  F.L.cens.est = 1-S0.L.cens.est^
                 (exp(par[4]*x1c+par[5]*x2c))
  Sp.L.obs.est = exp(-phi.obs.est*F.L.obs.est)
  Sp.R.obs.est = exp(-phi.obs.est*F.R.obs.est)
  Sp.L.cens.est = exp(-phi.cens.est*F.L.cens.est)
  p0.est = exp(-phi.cens.est)
  Su.est = (Sp.L.cens.est-p0.est)/(1-p0.est)
  for(i in 1:length(p0.est)){
    p0.est[i] = max(p0.est[i], .Machine$double.eps)
    Su.est[i] = max(Su.est[i], .Machine$double.eps)
  }
  dum = Sp.L.obs.est - Sp.R.obs.est
  dum1 = log(dum)
  dum2 = (1-e.W)*log(p0.est) + e.W*log((1-p0.est)*Su.est)
  result = sum(dum1)+sum(dum2)
```

```
    return(-result)
  } #end of Q Function
  #M step
  pnew = opt$par
  pnew = optim(par=c(pold),fn=Q, method="Nelder-Mead")$par
  iter = iter+1
  continue = (max(abs((pnew-pold)/pold))>tol&iter<maxit)
  pold = pnew
}#end of while
out = rep(NA,14)
if(0 %in% pnew){
  out = rep(0,14)
}else{
  std = function(param){
    b0 = param[1]
    b1 = param[2]
    b2 = param[3]
    g1 = param[4]
    g2 = param[5]
    lambda = param[6]
    phi.t = exp(b0+b1*x1t+b2*x2t)
    phi.c = exp(b0+b1*x1c+b2*x2c)
    S0.L.cens = exp(-L_cens^(1/lambda))
    S0.L.obs = exp(-L_obs^(1/lambda))
    S0.R.obs = exp(-R_obs^(1/lambda))
```

```
  F.c = 1 - S0.L.cens^(exp(g1*x1c+g2*x2c))
  F.tl = 1 - S0.L.obs^(exp(g1*x1t+g2*x2t))
  F.tr = 1 - S0.R.obs^(exp(g1*x1t+g2*x2t))
  Sp.c = exp(-phi.c*F.c)
  Sp.tl = exp(-phi.t*F.tl)
  Sp.tr = exp(-phi.t*F.tr)
  dum1 = rep(NA,length(Sp.tl))
  for(i in 1:length(dum1)){
    dum1[i] = max((Sp.tl[i]-Sp.tr[i]),
                  .Machine$double.eps)
  }
  out1 = sum(log(Sp.c)) + sum(log(dum1))
  return(out1)
}
hessmat1 = hessian(std,c(pnew[1],pnew[2],pnew[3],
             pnew[4],pnew[5],pnew[6]),method="Richardson")
FI = solve(-1*hessmat1)
if(any(is.na(FI))){
  out = c(pnew[1],pnew[2],pnew[3],pnew[4],pnew[5],
            pnew[6],alpha,0,0,0,0,0,0,NA)
}else if(FI[1,1]<0 | FI[2,2]<0 | FI[3,3]<0|
         FI[4,4]<0 | FI[5,5]<0 | FI[6,6]<0){
  out = c(pnew[1],pnew[2],pnew[3],pnew[4],pnew[5],
            pnew[6],alpha,0,0,0,0,0,0,NA)
}else{
```

```
        std1 = sqrt(FI[1,1])
        std2 = sqrt(FI[2,2])
        std3 = sqrt(FI[3,3])
        std4 = sqrt(FI[4,4])
        std5 = sqrt(FI[5,5])
        std6 = sqrt(FI[6,6])
        std.mle = std(c(pnew[1],pnew[2],pnew[3],pnew[4],
                    pnew[5],pnew[6]))
        out = c(pnew[1],pnew[2],pnew[3],pnew[4],pnew[5],
              pnew[6],0,std1,std2,std3,std4,std5,std6,std.mle)
        }
    }#end of else
    return(out)
}
```

REFERENCES

[1] M. Amico, I. Van Keilegom, and B. Han, Assessing cure status prediction from survival data using receiver operating characteristic curves, *Biometrika*, **108** (2021), 727-740.

[2] J. Asano and A. Hirakawa, Assessing the prediction accuracy of a cure model for censored survival data with long-term survivors: application to breast cancer data, *Journal of Biopharmaceutical Statistics*, **26** (2017), 918-932.

[3] J. Asano, A. Hirakawa, and C. Hamada, Assessing the prediction accuracy of cure in the Cox proportional hazards cure model: an application to breast cancer data, *Pharmaceutical Statistics*, **13** (2014), 357-363.

[4] N. Balakrishnan and F.S. Milienos, On a class of non-linear transformation cure rate models, *Biometrical Journal*, **62** (2020), 1208-1222.

[5] N. Balakrishnan, M.V. Koutras, F.S. Milienos, and S. Pal, Piecewise linear approximations for cure rate models and associated inferential issues, *Methodology and Computing in Applied Probability*, **18** (2016), 937-966.

[6] N. Balakrishnan and S. Pal, EM algorithm-based likelihood estimation for some cure rate models, *Journal of Statistical Theory and Practice*, **6** (2012), 698-724.

[7] N. Balakrishnan and S. Pal, Lognormal lifetimes and likelihood-based inference for flexible cure rate models based on COM-Poisson family, *Computational Statistics & Data Analysis*, **67** (2013), 41-67.

[8] N. Balakrishnan and S. Pal, An EM algorithm for the estimation of parameters of a flexible cure rate model with generalized gamma lifetime and model discrimination

using likelihood- and information-based methods, *Computational Statistics*, **30** (2015), 151-189.

[9] N. Balakrishnan and S. Pal, Likelihood inference for flexible cure rate models with gamma lifetimes, *Communications in Statistics - Theory and Methods*, **44** (2015), 4007-4048.

[10] N. Balakrishnan and S. Pal, Expectation maximization-based likelihood inference for flexible cure rate models with Weibull lifetimes, *Statistical Methods in Medical Research*, **25** (2016), 1535-1563.

[11] S. Banerjee and B.P. Carlin, Parametric spatial cure rate models for interval-censored time-to-relapse data, *Biometrics*, **60** (2004), 268-275.

[12] A. Beretta and C. Heuchenne, penPHcure: Variable selection in proportional hazards cure model with time-varying covariates, *The R Journal*, **13** (2021), 116-129.

[13] J. Berkson and R.P. Gage, Survival curve for cancer patients following treatment, *Journal of the American Statistical Association*, **47** (1952), 501-515.

[14] A. Bertrand and C. Legrand, Inference in a survival cure model with mismeasured covariates using a simulation-extrapolation approach, *Biometrika*, **104** (2017), 31-50.

[15] J.W. Boag, Maximum likelihood estimates of the proportion of patients cured by cancer therapy, *Journal of the Royal Statistical Society Series B*, **11** (1949), 15-53.

[16] L. Bordes, D. Chauveau, and P. Vandekerkhove, A stochastic EM algorithm for a semiparametric mixture model, *Computational Statistics & Data Analysis*, **51** (2007), 5429-5443.

[17] P. Borges, J. Rodrigues, N. and Balakrishnan, Correlated destructive generalized power series cure rate models and associated inference with an application to a cutaneous melanoma data, *Computational Statistics & Data Analysis*, **56** (2012), 1703-1713.

[18] V. Boutchaktchiev, A Markov chain model for the cure rate of non-performing loans, (2018), arXiv:1805.11804.

[19] G.E.P. Box and D.R. Cox, An analysis of transformations (with discussion), *Journal of the Royal Statistical Society, Series B*, **26** (1964), 211-252.

[20] T. Chen, Predicting analysis times in randomized clinical trials with cancer immunotherapy, *BMC Medical Research Methodology*, **16** (2016), doi:10.1186/s12874-016-0117-3.

[21] T. Chen and P. Du, Mixture cure rate models with accelerated failures and nonparametric form of covariate effects, *Journal of Nonparametric Statistics*, **30** (2018), 216-237.

[22] T. Chen and P. Du, Promotion time cure rate model with nonparametric form of covariate effects, *Statistics in Medicine*, **37** (2018), 1625-1635.

[23] M-H. Chen, J.G. Ibrahim, and D. Sinha, A new Bayesian model for survival data with a surviving fraction, *Journal of the American Statistical Association*, **94** (1999), 909-919.

[24] F. Cooner, S. Banerjee, and D. Sinha, Flexible cure rate modeling under latent activation schemes, *Journal of the American Statistical Association*, **102** (2007), 560-572.

[25] K. Davies, S. Pal, and J.A. Siddiqua, Stochastic EM algorithm for generalized exponential cure rate model and an empirical study, *Journal of Applied Statistics*, **48** (2021), 2112-2135.

[26] M. de Castro, V.G. Cancho, and J. Rodrigues, A Bayesian long-term survival model parametrized in the cured fraction, *Biometrical Journal*, **51** (2009), 443-455.

[27] L.A. de Freitas and J. Rodrigues, Standard exponential cure rate model with informative censoring, *Communications in Statistics - Simulation and Computation*, **42** (2013), 8-23.

[28] F.N. Demarqui, D.K. Dey, R.H. Loschi, and E.A. Colosimo, Fully semiparametric Bayesian approach for modeling survival data with cure fraction, *Biometrical Journal*, **56** (2014), 198-218.

[29] G. Diao and G. Yin, A general transformation class of semiparametric cure rate frailty models, *Annals of the Institute of Statistical Mathematics*, **64** (2012), 959-989.

[30] J. Diebolt and G. Celeux, Asymptotic properties of a stochastic EM algorithm for estimating mixing proportions, *Stochastic Models*, **9** (1993), 599-613.

[31] J. Diebolt and E. Ip, Markov Chain Monte Carlo in Practice, *Springer, New York* (1996).

[32] P.K. Dunn and G.K. Smyth, Randomized quantile residuals, *Journal of Computational and Graphical Statistics*, **5** (1996), 236-244.

[33] X. Fan, M. Liu, K. Fang, Y. Huang, and S. Ma, Promoting structural effects of covariates in the cure rate model with penalization, *Statistical Methods in Medical Research*, **26** (2017), 2078-2092.

[34] H.B. Fang, G. Li, and J. Sun, Maximum likelihood estimation in a semiparametric logistic/proportional-hazards mixture model, *Scandinavian Journal of Statistics*, **32** (2005), 59-75.

[35] V.T. Farewell, The use of mixture models for the analysis of survival data with long-term survivors, *Biometrics*, **38** (1982), 1041-1046.

[36] V.T. Farewell, Mixture models in survival analysis: are they worth the risk?, *Canadian Journal of Statistics*, **14** (1986), 257-262.

[37] D.I. Gallardo, H. Bolfarine, and A.C. Pedroso-de Lima, An EM algorithm for estimating the destructive weighted Poisson cure rate model, *Journal of Statistical Computation and Simulation*, **86** (2016), 1497-1515.

[38] D.I. Gallardo, H.W. Gómez, and H. Bolfarine, A new cure rate model based on the Yule–Simon distribution with application to a melanoma data set, *Journal of Applied Statistics*, **44** (2017), 1153-1164.

[39] D.I. Gallardo, Y.M. Gómez, and M. de Castro, A note on the EM algorithm for estimation in the destructive negative binomial cure rate model, *Journal of Statistical Computation and Simulation*, **87** (2017), 2291-2297.

[40] D.I. Gallardo, Y.M. Gómez, and M. de Castro, A flexible cure rate model based on the polylogarithm distribution, *Journal of Statistical Computation and Simulation*, **88** (2018), 2137-2149.

[41] D.I. Gallardo, Y.M. Gómez, H.W. Gómez, and M. de Castro, On the use of the modified power series family of distributions in a cure rate model context, *Statistical Methods in Medical Research*, **29** (2019), 1831-1845.

[42] Y.M. Gómez, D.I. Gallardo, J. Leão, and V.F. Calsavara, On a new piecewise regression model with cure rate: diagnostics and application to medical data, *Statistics in Medicine*, **40** (2021), 6723-6742.

[43] Y. Gu, D. Sinha, and S. Banerjee, Analysis of cure rate survival data under proportional odds model, *Lifetime Data Analysis*, **17** (2011), 123-134.

[44] J.G. Ibrahim, M-H. Chen, and S. Kim, Bayesian variable selection for the Cox regression model with missing covariates, *Lifetime Data Analysis*, **14** (2008), 496-520.

[45] J.G. Ibrahim, M-H. Chen, and D. Sinha, Bayesian semiparametric models for survival data with a cure fraction, *Biometrics*, **57** (2001), 383-388.

[46] W. Jiang, H. Sun, and Y. Peng, Prediction accuracy for the cure probabilities in mixture cure model, *Statistical Methods in Medical Research*, **26** (2017), 2029-2041.

[47] Y. Kim and M. Jhun, Cure rate model with interval censored data, *Statistics in Medicine*, **27** (2008), 3-14.

[48] M.V. Koutros and F.S. Milienos, A flexible family of transformation cure rate models, *Statistics in Medicine*, **36** (2017), 2559-2575.

[49] A.Y.C Kuk and C.H. Chen, A mixture model combining logistic regression with proportional hazards regression, *Biometrika*, **79** (1992), 531-541.

[50] K.F. Lam, K.Y. Wong, and F. Zhou, A semiparametric cure model for interval-censored data, *Biometrical Journal*, **55** (2013), 771-788.

[51] J. Leão, V. Leiva, H. Saulo, and V. Tomazella, Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data, *Statistics in Medicine*, **37** (2018), 4421-4440.

[52] C.S. Li, J.M.G. Taylor, and J.P. Sy, Identifiability of cure models, *Statistics & Probability Letters*, **54** (2001), 389-395.

[53] P. Li, Y. Peng, P. Jiang, and Q. Dong, A support vector machine based semi-parametric mixture cure model, *Computational Statistics*, **35** (2020), 931-945.

[54] X. Liu, Y. Peng, D. Tu, and H. Liang, Variable selection in semiparametric cure models based on penalized likelihood, with application to breast cancer clinical trials, *Statistics in Medicine*, **31** (2012), 2882-2891.

[55] A. López-Cheda, R. Cao, M.A. Jácome, and I. Van Keilegom, Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models, *Computational Statistics & Data Analysis*, **105** (2017), 144-165.

[56] Y. Ma and G. Yin, Cure rate model with mismeasured covariates under transformation, *Journal of the American Statistical Association*, **103** (2008), 743-756.

[57] J. Majakwara and S. Pal, On some inferential issues for the destructive COM-Poisson-generalized gamma regression cure rate model, *Communications in Statistics - Simulation and Computation*, **48** (2019), 3118-3142.

[58] R.A. Maller and X. Zhou, Survival Analysis with Long-Term Survivors, *John Wiley & Sons Ltd.*, (1996).

[59] I.C. Marschner, On stochastic versions of the EM algorithm, *Biometrika*, **88** (2001), 281-286.

[60] G.J. McLachlan and T. Krishnan, The EM Algorithm and Extensions, *John Wiley & Sons, Hoboken, New Jersey*, 2nd ed, (2008).

[61] F.S. Milienos, On a reparameterization of a flexible family of cure models. *Statistics in Medicine*, **41** (2022), 4091-4111.

[62] R.P. Murray, N.R. Anthonisen, J.E. Connett, R.A. Wise, P.G. Lindgren, P.G. Greene, and M.A. Nides, Effects of multiple attempts to quit smoking and relapses to smoking on pulmonary function, *Journal of Clinical Epidemiology*, **51** (1998), 1317-1326.

[63] M.A. Nicolaie, J.M.G. Taylor, and C. Legrand, Vertical modeling: analysis of competing risks data with a cure fraction, *Lifetime Data Analysis*, **25** (2019), 1-25.

[64] S.F. Nielsen, The stochastic EM algorithm: Estimation and asymptotic results, *Bernoulli*, **6** (2000), 457-489.

[65] L.E. Nieto-Barajas and G. Yin, Bayesian semiparametric cure rate model with an unknown threshold, *Scandinavian Journal of Statistics*, **35** (2008), 540-556.

[66] Y. Niu, L. Song, Y. Liu, and Y. Peng, Modeling clustered long-term survivors using marginal mixture cure model, *Biometrical Journal*, **60** (2018), 780-796.

[67] Y. Niu, X. Wang, and Y. Peng, geecure: An R-package for marginal proportional hazards mixture cure models, *Computer Methods and Programs in Biomedicine*, **161** (2018), 115-124.

[68] T. Overman and S. Pal, Statistical tools and techniques in modeling survival data. In: *Mathematics Research for the Beginning Student Volume 2: Accessible Research Projects for First- and Second-Year College and Community College Students after Calculus*, Chapter 3 (Eds., E. Goldwyn, A. Wootton and S. Ganzell), pp.75-99. Birkhauser-Springer.

[69] S. Pal, A simplified stochastic EM algorithm for cure rate model with negative binomial competing risks: An application to breast cancer data, *Statistics in Medicine*, **40** (2021), 6387-6409.

[70] S. Pal and W. Aselisewine, A semi-parametric promotion time cure model with support vector machine, (2023), arXiv:2304.11815.

[71] S. Pal and N. Balakrishnan, Destructive negative binomial cure rate model and EM-based likelihood inference under Weibull lifetime, *Statistics & Probability Letters*, **116** (2016), 9-20.

[72] S. Pal and N. Balakrishnan, An EM type estimation procedure for the destructive exponentially weighted Poisson regression cure model under generalized gamma lifetime, *Journal of Statistical Computation and Simulation*, **87** (2017), 1107-1129.

[73] S. Pal and N. Balakrishnan, Likelihood inference for COM-Poisson cure rate model with interval-censored data and Weibull lifetimes, *Statistical Methods in Medical Research*, **26** (2017), 2093-2113.

[74] S. Pal and N. Balakrishnan, Likelihood inference for the destructive exponentially weighted Poisson cure rate model with Weibull lifetime and an application to melanoma data, *Computational Statistics*, **32** (2017), 429-449.

[75] S. Pal and N. Balakrishnan, Likelihood inference based on EM algorithm for the destructive length-biased Poisson cure rate model with Weibull lifetime, *Communications in Statistics - Simulation and Computation*, **47** (2018), 644-660.

[76] S. Pal and N. Balakrishnan, Expectation maximization algorithm for Box-Cox transformation cure rate model and assessment of model misspecification under Weibull lifetimes, *IEEE Journal of Biomedical and Health Informatics*, **22** (2018), 926-934.

[77] S. Pal, S. Barui, K. Davies, and N. Mishra, A stochastic version of the EM algorithm for mixture cure model with exponentiated Weibull family of lifetimes, *Journal of Statistical Theory and Practice*. **16** (2022), 48.

[78] S. Pal, J. Majakwara, and N. Balakrishnan, An EM algorithm for the destructive COM-Poisson regression cure rate model, *Metrika*, **81** (2018), 143-171.

[79] S. Pal, Y. Peng, S. Barui, and P. Wang, A support vector machine based cure rate model for interval censored data, (2022), arXiv:2109.01098v2.

[80] S. Pal and S. Roy, A new non-linear conjugate gradient algorithm for destructive cure rate model and a simulation study: illustration with negative binomial competing risks, *Communications in Statistics - Simulation and Computation*, **51** (2020), 6866-6880.

[81] S. Pal and S. Roy, On the estimation of destructive cure rate model: A new study with exponentially weighted Poisson competing risks, *Statistica Neerlandica*, **75** (2021), 324-342.

[82] S. Pal and S. Roy, On the parameter estimation of Box-Cox transformation cure model, *Statistics in Medicine* (2023), 1-19, doi: 10.1002/sim.9739.

[83] S. Pal, H. Yu, Z.D. Loucks, and I.M. Harris, Illustration of the flexibility of generalized gamma distribution in modeling right censored survival data: analysis of two cancer datasets, *Annals of Data Science*, **7** (2020), 77-90.

[84] Y. Peng and J. Xu, An extended cure model and model selection, *Lifetime Data Analysis*, **18** (2012), 215-233.

[85] Y. Peng and B. Yu, Cure Models: Methods, Applications and Implementation, *Chapman and Hall/CRC*, (2021).

[86] J. Rodrigues, V.G. Cancho, M. de Castro, and N. Balakrishnan, A Bayesian destructive weighted Poisson cure rate model and an application to a cutaneous melanoma data, *Statistical Methods in Medical Research*, **21** (2012), 585-597.

[87] J. Rodrigues, G.M. Cordeiro, V.G. Cancho, and N. Balakrishnan, Relaxed Poisson cure rate models, *Biometrical Journal*, **58** (2016), 397-415

[88] J. Rodrigues, V.G. Cancho, M. de Castro, and F. Louzada-Neto, On the unification of the long-term survival models, *Statistics and Probability Letters*, **79** (2009), 753-759.

[89] J. Rodrigues, M. de Castro, N. Balakrishnan, and V.G. Cancho, Destructive weighted Poisson cure rate models, *Lifetime Data Analysis*, **17** (2011), 333-346.

[90] J. Rodrigues, M. de Castro, V.G. Cancho, and N. Balakrishnan, COM-Poisson cure rate survival models and an application to a cutaneous melanoma data, *Journal of Statistical Planning and Inference*, **139** (2009), 3605-3611.

[91] P.H. Rossi, R.A. Berk, and K.J. Lenihan, Money, Work and Crime. *Academic Press, New York*, (1980).

[92] S.M.H. Rua and D.K. Dey, A Bayesian piecewise survival cure rate model for spatially clustered data, *Spatial and Spatio-temporal Epidemiology*, **29** (2019), 149-159.

[93] F. Shao, J. Li, S. Ma, and M-L.T. Lee, Semiparametric varying-coefficient model for interval censored data with a cured proportion, *Statistics in Medicine*, **33** (2014), 1700-1712.

[94] J.P. Sy and J.M.G. Taylor, Estimation in a Cox proportional hazards cure model, *Biometrics*, **56** (2000), 227-236.

[95] J. Treszoks and S. Pal, A destructive shifted Poisson cure model for interval censored data and an efficient estimation algorithm, *Communications in Statistics - Simulation and Computation*, doi: 10.1080/03610918.2022.2067876.

[96] A.D. Tsodikov, Semi-parametric models of long- and short-term survival: an application to the analysis of breast cancer survival in Utah by age and stage, *Statistics in Medicine*, **21** (2002), 895-920.

[97] A.D. Tsodikov, Semiparametric models: a generalized self-consistency approach, *Journal of the Royal Statistical Society: Series B*, **65** (2003), 759-774.

[98] A.D. Tsodikov, J.G. Ibrahim, and A.Y. Yakovlev, Estimating cure rates from survival data: an alternative to two-component mixture models, *Journal of the American Statistical Association*, **98** (2003), 1063-1078.

[99] L. Wang, P. Du, and H. Liang, Two-component mixture cure rate model with spline estimated nonparametric components, *Biometrics*, **68** (2012), 726-735.

[100] P. Wang and S. Pal, A two-way flexible generalized gamma transformation cure rate model, *Statistics in Medicine*, **41** (2022), 2427-2447.

[101] P. Wiangnak and S. Pal, Gamma lifetimes and associated inference for interval-censored cure rate model with COM-Poisson competing cause, *Communications in Statistics - Theory and Methods*, **47** (2018), 1491-1509.

[102] L. Xiang, X. Ma, and K.K.W. Yau, Mixture cure model with random effects for clustered interval-censored survival, *Statistics in Medicine*, **30** (2011), 995-1006.

[103] Y. Xie and Z. Yu, Promotion time cure rate model with a neural network estimated nonparametric component, *Statistics in Medicine*, **40** (2021), 3516-3532.

[104] A.Y. Yakovlev, A.B. Cantor, and J.J. Shuster, Parametric versus non-parametric methods for estimating cure rates based on censored survival data, *Statistics in Medicine*, **13** (1994), 983-986.

[105] A.Y. Yakovlev and A.D. Tsodikov, Stochastic models of tumor latency and their biostatistical applications, *World Scientific Publishers, Singapore*, (1996).

[106] Z. Ye and H.K.T. Ng, On analysis of incomplete field failure data, *Annals of Applied Statistics*, **8** (2014), 1713-1727.

[107] G. Yin and J.G. Ibrahim, Cure cate models: an unified approach, *The Canadian Journal of Statistics*, **33** (2005), 559-570.

[108] B. Yiqi, V.G. Cancho, D.K. Dey, N. Balakrishnan, and A.K. Suzuki, Power series cure rate model for spatially correlated interval-censored data based on generalized

extreme value distribution, *Journal of Computational and Applied Mathematics*, (2020), 364:112362.

[109] D. Zeng, G. Yin, and J.G. Ibrahim, Semiparametric transformation models for survival data with a cure fraction, *Journal of the American Statistical Association*, **101** (2006), 670-684.

[110] J. Zhang and Y. Peng, Semiparametric estimation methods for the accelerated failure time mixture cure model, *Journal of the Korean Statistical Society*, **41** (2012), 415-422.

[111] L. Zhao, D. Feng, E.L. Bellile, and J.M. Taylor, Bayesian random threshold estimation in a Cox proportional hazards cure model, *Statistics in Medicine*, **33** (2014), 650-661.

BIOGRAPHICAL STATEMENT

Jodi Treszoks was born in North Richland Hills, Texas. She received a Bachelor of Arts in Psychology from University of North Texas in 2009 and a Bachelor of Science in Mathematics from The University of Texas at Arlington in 2018. She has been a PhD student in Mathematics at The University of Texas at Arlington since August 2018. During her PhD education, she received the Academic Excellence Scholarship and was recognized as Outstanding Graduate Student Teacher. Her research interests focus on the use of statistical computing and survival analysis techniques in the context of biological and social science applications.