Thailand Statistician
January 2025; 23(1): 41-63
http://statassoc.or.th
Contributed paper

Non-mixture Cure Fraction Model in the Presence of Right Censored Survival Data and Covariates Based on Nadarajah-Haghighi Distribution with Applications to Medical Data

Yakubu Aliyu [a,b]* and Umar Usman [a]
[a] Department of Mathematics, Usmanu Danfodiyo University, Sokoto, Nigeria
[b] Department of Statistics, Ahmadu Bello University, Zaria, Nigeria
*Corresponding author; e-mail: aliyuyakubu40@gmail.com/yakubualiyu@abu.edu.ng

Received: 12 September 2021
Revised: 18 July 2022
Accepted: 4 September 2022

Abstract

In survival analysis it is assumed that every individual in the study population will eventually experience the event of interest if followed-up for a long period of time. However, there are some occasions in which a proportion of the population will never experience the event of interest. Hence, cure fraction models are used in modelling such type of data. The present paper introduces the Nadarajah-Haghighi distribution in the presence of cure fraction, right censored data and covariates. Comprehensive statistical properties of the model were explored. Inferences for the proposed model were obtained under the maximum likelihood and Bayesian approaches. Simulation study was provided in order to ascertain the performance of the maximum likelihood estimates. Illustrations of the proposed methodology were made by considering medical data sets using maximum likelihood and Bayesian methods. Results of the applications showed that the proposed methodology is a good competitor.

Keywords: Nadarajah-Haghighi distribution, long-term non-mixture model, censoring and Colorectal cancer.

1. Introduction

Cure fraction models are widely used in modelling survival data with a surviving fraction. Cure fraction models are used in modelling time-to-event data from a population in which the population consists of sampling units that are non-susceptible to the occurrence of the event of interest, which is caused by the improvement in the field of medicine. There are basically two major types of cure fraction models in the literature: mixture and non-mixture cure fraction models. The mixture cure fraction model also known as the standard cure fraction model was first introduced by Boag (1949) and further developed by Berkson and Gage (1952). The model was later studied extensively by several authors including Farewell (1986); Meeker (1987); Gamel et al. (1990); Ng and McLachlan (1998); Peng et al. (1998); Sy and Taylor (2000); Shao and Zhou (2004); Kannan et al. (2010); Achcar et al. (2012); Mazucheli et al. (2013); Usman et al. (2021) among many others. The mixture cure fraction model is the most popular type of cure fraction model. In the mixture cure fraction model, the study population assumed that a certain fraction of the population is non-susceptible to the event of interest while the remaining fraction is susceptible. Parametric, semi-parametric and

non-parametric statistical methods have been used by different researchers in order to estimate the proportion of non-susceptible.

On the other hand, the non-mixture cure fraction model also referred to as the promotion time cure model or bounded cumulative hazard model was first introduced by Yakovlev et al. (1993) and further discussed by Yakovlev et al. (1996); Chen et al. (1999); Tsodikov et al. (2003). According to Chen et al. (1999) the non-mixture cure fraction model was motivated by the underlying biological mechanism and was developed based on the assumption that the number of cancer cells that remain active after cancer treatment follows poison distribution. The non-mixture cure fraction model was shown to have distinct advantages over the mixture cure fraction model. For instance, it is easy to compute due to its simple structure for the survival function which provides certain technical advantage when developing maximum likelihood estimation procedures, it has proportional hazard model structure and it presents a much more biologically meaningful interpretation of the results of the data analysis Chen et al. (1999); Uddin et al. (2006). It is important to note that the non-mixture model has a mathematical relationship with the standard cure fraction model. That is, the non-mixture model can be written as the standard cure model and vice versa Uddin et al. (2006). Uddin et al. (2006) estimate the parameters of the non-mixture cure fraction model using maximum likelihood estimation procedure considering the data to be uncensored. A non-parametric maximum likelihood method of estimation was also considered by Uddin et al. (2006) to estimate the parameters of the non-mixture cure fraction model assuming uncensored data. Liu and Shen (2009) developed a semi-parametric maximum likelihood estimation procedure for the non-mixture model for interval censored time-to-event data. Lopes and Bolfarine (2012) estimate the parameters of the non-mixture cure model via classical and non-classical method of estimations. Herring and Ibrahim (2002) introduced a parametric method for estimating the parameters of the non-mixture model for a non-ignorable missing covariates.

In the present paper, a non-mixture cure fraction model called Nadarajah-Haghighi non-mixture cure fraction model was studied. The model was introduced in the presence of right censored data and covariates. Some of the statistical properties of the model were studied and applications of the model to some medical data sets were provided. The rest of the paper is organized as follows: In Section 2, we introduced the Nadarajah-Haghighi non-mixture cure fraction model, Section 3 study the statistical properties of the introduced model. Simulation study and real life applications are given in Sections 4 and 5 respectively and we finally conclude in Section 6.

2. Model

Nadarajah and Haghighi (2011) introduced an extension of the well-known exponential distribution called Nadarajah-Haghighi exponential distribution. The distribution was shown to serve as an alternative to the Weibull, gamma and generalized exponential distributions in modeling survival data. The probability density function (pdf), cumulative distribution function (cdf) and survival function for the Nadarajah-Haghighi exponential (NH) distribution with parameters ϕ and ψ are respectively given by

$$f_u\left(t/\phi,\psi\right) = \phi\psi\left(1+\phi t\right)^{\psi-1} exp\left(1-\left(1+\phi t\right)^{\psi}\right), \quad (1)$$

$$F_u\left(t/\phi,\psi\right) = 1-exp\left(1-\left(1+\phi t\right)^{\psi}\right), \quad (2)$$

and

$$S_u\left(t/\phi,\psi\right) = exp\left(1-\left(1+\phi t\right)^{\psi}\right), \quad (3)$$

where $\phi > 0$ and $\psi > 0$ are the scale and shape parameters, respectively. Similar to the Weibull, gamma and generalized exponential distributions, the hazard rate function of the NH distribution could be constant, decreasing or increasing. Also, the hazard rate function of the NH distribution reduces to the hazard rate function of the exponential distribution when the shape parameter assumes the value one.

The non-mixture cure fraction model defined an asymptote for the cumulative hazard and for the cure fraction. The non-mixture model is developed as follows:

Let N be the number of cancer cells for an individual after treatment. Assume that the number of cancer cells is Poisson distributed with parameter μ since the number of cancer cells may grow rapidly and produce a detectable cancer disease. Also, let Z_k denote the random time for the kth cancer cell to produce a detectable cancer mass. Assuming Z_k are independently and identically distributed (iid) with a common distribution function and survival function ($F(t)$ and $S(t)$). Assume further, that Z_k are independent of N. Then, the time to relapse of cancer is defined by the random variable $T = \min\{Z_k,\ 0 \leqslant k \leqslant N\}$, where $P(Z_0 = \infty) = 1$. Hence, the survival function of T is given by

$$\begin{aligned} S(t) &= P(\text{ number of cancer by time } t) \\ &= P(N = 0) + P(Z_1 > t,\ Z_2 > t, \ldots, Z_N > t, N \geqslant 1) \\ &= \exp(-\mu) + \sum_{N=1}^{\infty} S^N(t) \frac{\mu^N}{N!} \exp(-\mu) \\ &= \exp(-\mu + \mu S_u(t)) \\ &= \exp(-\mu F_u(t)) \\ S(t) &= p^{F_u(t)}, \end{aligned} \tag{4}$$

where $p = \exp(-\mu)$ is the proportion of non-susceptibles that lies in the interval $[0, 1]$. The corresponding cdf, pdf and hazard function are given by

$$F(t) = 1 - p^{F_u(t)}, \tag{5}$$

$$f(t) = -\log(p) f_u(t) p^{F_u(t)} \tag{6}$$

and

$$h(t) = -\log(p) f_u(t), \tag{7}$$

respectively, where $S_u(t)$ is the survival function for the susceptible group. Consequently, considering the Nadarajah-Haghighi distribution, the survival function of the Nadarajah-Haghighi non-mixture cure fraction model is

$$S(t/\phi, \psi, p) = p^{1-exp\left(1-(1+\phi t)^{\psi}\right)}. \tag{8}$$

The corresponding probability density function, cumulative distribution function and hazard function for the model are

$$f(t/\phi, \psi, p) = -\phi\psi \log(p) (1+\phi t)^{\psi-1} exp\left(1-(1+\phi t)^{\psi}\right) p^{1-exp\left(1-(1+\phi t)^{\psi}\right)} \tag{9}$$

$$F(t/\phi, \psi, p) = 1 - p^{1-exp\left(1-(1+\phi t)^{\psi}\right)} \tag{10}$$

and

$$h(t/\phi, \psi, p) = -\log(p) \phi\psi (1+\phi t)^{\psi-1} exp\left(1-(1+\phi t)^{\psi}\right) \tag{11}$$

respectively, where $\phi > 0, \psi > 0$ and $0 < p < 1$, ϕ is scale parameter, ψ is shape parameter and p is the proportion of non-susceptible. It should be mentioned that, $S(t)$ and $F(t)$ in Equations (8) and (10) respectively are improper survival and cumulative distribution functions since $\lim_{t\to\infty} S(t) = p < 1$ and $\lim_{t\to\infty} F(t) = 1 - p < 1$. Also, the survival function in Equation (8) reduces to the survival function of the exponential non-mixture cure fraction model when the shape parameter ψ assume the value one.

2.1. Maximum likelihood estimation of the NHNMCF model

Consider a random sample of lifetimes (t_i, δ_i), for $i = 1, 2, \ldots, n$ under the assumption of right censored lifetime. The likelihood function of (t_i, δ_i), $i = 1, 2, \ldots, n$ is defined by

$$L(\Theta/t, \delta) = \prod_{i=1}^{n} f(t_i)^{\delta_i} [S(t_i)]^{1-\delta_i} = \prod_{i=1}^{n} h(t_i)^{\delta_i} S(t_i)$$

substituting Equations (8) and (11), the likelihood function becomes

$$\prod_{i=1}^{n} \left[-\log(p)\, \phi\psi (1+\phi t)^{\psi-1} exp\left(1-(1+\phi t)^{\psi}\right)\right]^{\delta_i} \cdot p^{1-exp\left(1-(1+\phi t)^{\psi}\right)} \tag{12}$$

and taking natural logarithm gives the full log-likelihood function as

$$\begin{aligned} \ell(\Theta/t, \delta) &= z\log(-\log(p)) + z\log(\phi) + z\log(\psi) + (1-\psi)\sum \delta_i B_i^{\frac{1}{\psi}} \\ &\quad + z - \sum \delta_i B_i + n\log(p) - \log(p)\sum C_i, \end{aligned} \tag{13}$$

where $z = \sum \delta_i$, $B_i = (1+\phi t_i)^{\psi}$ and $C_i = exp(1-B_i)$.

Differentiating (13) partially with respect to ϕ, ψ and p gives the score function as

$$\frac{\partial \ell(\Theta/t, \delta)}{\partial \phi} = \frac{z}{\phi} - (\psi-1)\sum \delta_i t_i B_i^{-\frac{1}{\psi}} - \psi \sum \delta_i t_i D_i + \psi \log(p) \sum C_i D_i t_i \tag{14}$$

$$\frac{\partial \ell(\Theta/t, \delta)}{\partial \psi} = \frac{z}{\psi}\sum \delta_i \log(B_i) - \frac{1}{\psi}\sum \delta_i B_i \log(B_i) + \frac{\log(p)}{\psi}\sum B_i C_i \log(B_i) \tag{15}$$

and

$$\frac{\partial \ell(\Theta/t, \delta)}{\partial p} = \frac{z}{p\log(p)} + \frac{n}{p} - \frac{1}{p}\sum C_i. \tag{16}$$

The MLE of p can easily be obtain algebraically as

$$\widehat{p}(\phi, \psi) = exp\left(\frac{z}{\sum C_i - n}\right). \tag{17}$$

However, the estimates of $\widehat{\phi}$ and $\widehat{\psi}$ can be obtain by substituting $\widehat{p}(\phi, \psi)$ into Equations (14) and (15) and numerical methods can be used to solve for ϕ and ψ. Statistical package such as fitdist package in R can be used in maximizing these equations. In order to find interval estimates and to test for hypothesis, the observed information matrix $I(\Lambda)$ is used. This is given by

$$I(\Lambda) = -\begin{pmatrix} V_{\psi\psi} & V_{\psi\phi} & V_{\psi p} \\ & V_{\phi\phi} & V_{\phi p} \\ & & V_{pp} \end{pmatrix},$$

where the elements of the matrix are obtained as: $V_{\psi\psi} = \frac{\partial^2 \ell}{\partial \psi^2}, V_{\phi\phi} = \frac{\partial^2 \ell}{\partial \phi^2}, V_{pp} = \frac{\partial^2 \ell}{\partial p^2}, V_{\phi\psi} = \frac{\partial^2 \ell}{\partial \phi \partial \psi}, V_{\psi p} = \frac{\partial^2 \ell}{\partial \psi \partial p}$ and $V_{\phi p} = \frac{\partial^2 \ell}{\partial \phi \partial p}$.

These elements are given as

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \psi^2} &= -\frac{1}{\psi^2}\sum \delta_i B_i \left(\log\left(B_i\right)\right)^2 + \frac{\log\left(p\right)}{\psi^2}\sum B_i C_i \left(\log\left(B_i\right)\right)^2 - \frac{log\left(p\right)}{\psi^2}\sum \left(B_i \log\left(B_i\right)\right)^2 C_i, \\
\frac{\partial^2 \ell}{\partial \phi^2} &= -\frac{z}{\phi^2} + \left(\psi - 1\right)\sum \delta_i t_i^2 B_i^{-\frac{2}{\psi}} - \psi\left(\psi - 1\right)\sum \delta_i t_i^2 B_i^{\frac{\psi-1}{\psi}} \\
&\quad + \psi \log\left(p\right)\sum t_i^2 B_i^{\frac{\psi-2}{\psi}} C_i \left[\psi\left(1 - B_i\right) - 1\right], \\
\frac{\partial^2 \ell}{\partial p^2} &= -\frac{z}{p^2 \log\left(p\right)} - \frac{z}{p^2 \left(\log\left(p\right)\right)^2} - \frac{n}{p^2} + \frac{1}{p^2}\sum C_i, \\
\frac{\partial^2 \ell}{\partial \phi \partial \psi} &= \log\left(p\right)\sum C_i D_i t_i \left[1 + \log\left(B_i\right) + B_i \log\left(B_i\right)\right] - \sum \delta_i t_i \left[B_i^{-\frac{1}{\psi}} + D_i + D_i \log\left(B_i\right)\right], \\
\frac{\partial^2 \ell}{\partial p \partial \psi} &= \frac{1}{\psi p}\sum B_i \log\left(B_i\right) C_i, \\
\frac{\partial^2 \ell}{\partial p \partial \phi} &= \frac{\psi}{p}\sum C_i D_i t_i.
\end{aligned}$$

The variances of the parameters ψ, ϕ and p are the diagonal elements of $I(\Lambda)^{-1}$ while the off-diagonal elements are the covariances. The asymptotic distribution of $\sqrt{n}\left(\widehat{\Lambda} - \Lambda\right)$ is multivariate normal $N_3\left(0, J\left(\widehat{\Lambda}\right)^{-1}\right)$, where $J\left(\widehat{\Lambda}\right)$ is the total observed information matrix evaluated at $\widehat{\Lambda}$. The asymptotic $100\left(1-\tau\right)\%$ confidence interval for the parameters ϕ, ψ and p are $\widehat{\phi} \pm Z_{\frac{\tau}{2}}\sqrt{var\left(\widehat{\phi}\right)}$, $\widehat{\psi} \pm Z_{\frac{\tau}{2}}\sqrt{var\left(\widehat{\psi}\right)}$ and $\widehat{p} \pm Z_{\frac{\tau}{2}}\sqrt{var\left(\widehat{p}\right)}$ respectively, where $Z_{\frac{\tau}{2}}$ is the $100\left(1-\tau\right)\%$ quantile of the standard normal distribution.

2.2. Bayesian technique

In this section, Bayesian method of estimation based on Markov Chain Monte Carlo (MCMC) technique was considered. The methodology was used to get the approximate posterior summaries of the parameters of the NHNMCF model.

Consider the NHNMCF model, assume the vector of unknown parameters be Φ. For the Bayesian method not assuming covariates, let the prior density for the parameters ψ, ϕ and p respectively be $\pi(\psi), \pi(\phi)$ and $\pi(p)$ such that

$$\pi(\psi) = \frac{1}{\Gamma(a) b^{a-1}} \psi^{a-1} e^{-\frac{\psi}{b}}, \tag{18}$$

$$\pi(\phi) = \frac{1}{\Gamma(c) d^{c-1}} \phi^{c-1} e^{-\frac{\phi}{d}}, \tag{19}$$

$$\pi(p) = \frac{1}{B(e,f)} p^{e-1} (1-p)^{f-1}, \tag{20}$$

where a, b, c, d, e and f are hyper-parameters, $B(e,f)$ is referred to as beta function which is defined as $B(e,f) = \frac{\Gamma(e)\Gamma(f)}{\Gamma(e+f)}$. The hyper-parameters are assumed to be specified and non-negative. Prior independence among the parameters is also assumed. Hence, the joint prior distribution is given by

$$\pi(\Phi) = \omega \psi^{a-1} \phi^{c-1} p^{e-1} (1-p)^{f-1} e^{-(\frac{\psi}{b} + \frac{\phi}{d})}, \tag{21}$$

where $\omega = \frac{1}{\Gamma(a)\Gamma(c) b^{a-1} d^{c-1} B(e,f)}$. The joint posterior density of the model parameters ψ, ϕ and p is obtain as the product of the joint prior distributions in (21) and the likelihood function in (12). This

is given as

$$\pi(\Phi/t,\delta) \propto (-\ell n(p))^m \psi^{m+a-1}\phi^{m+c-1}p^{e-1}(1-p)^{f-1} \times \prod(1+\phi t_i)^{\delta_i(\psi-1)} \times exp\left(\sum\left(1-(1+\phi t_i)^{\psi}\right)(\delta_i+\ell n(p)) - \frac{\psi}{b} - \frac{\phi}{d}\right). \quad (22)$$

However, the prior distribution of the proportion of non-susceptible is assumed to have parameters $e = 1$ and $f = 1$ since $0 < p < 1$, hence, $\pi(p) = 1$ and the joint posterior density reduces to

$$\pi(\Phi/t,\delta) \propto (-\ell n(p))^m \psi^{m+a-1}\phi^{m+c-1} exp\left(\sum\left(1-(1+\phi t_i)^{\psi}\right)(\delta_i+\ell n(p)) - \frac{\psi}{b} - \frac{\phi}{d}\right) \times \prod(1+\phi t_i)^{\delta_i(\psi-1)}. \quad (23)$$

Observe that, the joint posterior distribution in (22) becomes more complex when covariates are present and may not be possible to compute the estimates of the parameters analytically. Hence, MCMC simulation techniques can be used to generate samples from the joint posterior distribution. A great computational simplification in simulating these samples can be achieved using the OpenBUGS software, where only the distribution of the data and the prior distributions of the parameters are required.

In the presence of r covariates $\mathbf{x} = (x_1, x_2, \ldots, x_r)'$ affecting the parameters of the NHNMCF model, a link function for the parameters ϕ, ψ and p is assumed. That is,

$$\log(\phi) = \phi_0 + \phi_1 x_i + \cdots + \phi_r x_r,$$
$$\log(\psi) = \psi_0 + \psi_1 x_i + \cdots + \psi_r x_r,$$

and

$$\log\left(\frac{p_i}{1-p_i}\right) = \eta_0 + \eta_1 x_i + \cdots + \eta_r x_r,$$

for ϕ, ψ and p respectively. To be specific, assume the NHNMCF model with shape parameter ψ, scale parameter ϕ and a cure fraction parameter p are affected by the presence of two covariate x_{1i} and x_{2i} for $i = 1, 2, \cdots, n$, then the link function

$$\log(\phi) = \phi_0 + \phi_1 x_{1i} + \phi_2 x_{2i},$$
$$\log(\psi) = \psi_0 + \psi_1 x_{1i} + \psi_2 x_{2i},$$

and

$$\log\left(\frac{p_i}{1-p_i}\right) = \eta_0 + \eta_1 x_{1i} + \eta_2 x_{2i},$$

are assumed for the scale parameter ϕ, shape parameter ψ and cure fraction parameter p respectively. All inferences considering covariate effect are obtain by replacing ϕ, ψ and p with these link functions. Furthermore, in the Bayesian analysis for the models with covariates, normal prior distribution is assumed for the effect of covariates. That is, $N\left(g, h^2\right)$ is assumed for each ψ_k, ϕ_k and p_k $k = 1, \ldots r$ present in the model, where g and h are known mean and standard deviation for each parameter present.

3. Statistical Properties

In this section, statistical properties of the model such as quantile function, median and moments of the NHNMCF model were discussed.

3.1. Quantile function

To obtain random realizations from a given model, the quantile function can be employed. The quantile function for the NHNMCF model is given by

$$Q(u) = \frac{1}{\phi}\left[1 - log\left[1 - \frac{log(1-u)}{log(p)}\right]\right]^{\frac{1}{\psi}} - \frac{1}{\phi}, \tag{24}$$

where u is a random number generated from uniform distribution within the interval $(0, 1)$. The first, second and third quantiles of the NHNMCF model are obtain by letting $u = 0.25, 0.50$ and 0.75 respectively. For instance, the median is obtain by letting $u = 0.50$. This gives

$$Q_2 = \frac{1}{\phi}\left[1 - log\left[\frac{log(2p)}{log(p)}\right]\right]^{\frac{1}{\psi}} - \frac{1}{\phi}. \tag{25}$$

Observe that the quantile function of the NHNMCF model is in closed form. Hence, the inverse transform method can easily be used in simulating random realizations from this model.

3.2. Moments

Following Ibrahim et al. (2001), the pdf of the NHNMCF model in (9) can be written in the mixture form as

$$f(t) = -\log(p)\, e^{-\theta F_u(t)} f_u(t) \tag{26}$$

where $p = exp(-\theta)$ is the proportion of non-susceptible. Hence, the pdf of the NHNMCF model can be written as

$$f(t) = -\phi\psi p \log(p)\, exp\left(\theta e^{\left(1-(1+\phi t)^{\psi}\right)}\right)(1+\phi t)^{\psi-1} e^{\left(1-(1+\phi t)^{\psi}\right)}$$

using the relation $e^{\theta t} = \sum_{j=0}^{\infty} \frac{\theta^j t^j}{j!}$ yields

$$f(t) = -\phi\psi p \log(p) \sum_{j=0}^{\infty} \frac{\theta^j}{j!}(1+\phi t)^{\psi-1} exp\left(\left(1-(1+\phi t)^{\psi}\right)(j+1)\right).$$

Hence, the r^{th} moment of the NHNMCF model is as follows

$$\begin{aligned} E(T^r) &= -\phi\psi p \log(p) \sum_{j=0}^{\infty} \frac{\theta^j}{j!} \int_0^{\infty} t^r (1+\phi t)^{\psi-1} e^{j+1} \\ &\quad \exp\left(-(j+1)(1+\phi t)^{\psi}\right)(j+1)\, dt \end{aligned} \tag{27}$$

let $m = (1+\phi t)^{\psi}$, then $E(T^r)$ becomes

$$E(T^r) = \frac{-p\log(p)}{\phi^r} \sum_{j=0}^{\infty} \frac{\theta^j}{j!} e^{j+1} \int_1^{\infty} \left(m^{\frac{1}{\psi}} - 1\right)^r e^{-(j+1)m} dm \tag{28}$$

but $\left(m^{\frac{1}{\psi}} - 1\right)^r = (-1)^r \left(1 - m^{\frac{1}{\psi}}\right)^r$ and applying binomial expansion to $\left(1 - m^{\frac{1}{\psi}}\right)^r$ yields

$$E(T^r) = \frac{-p\log(p)}{\phi^r} \sum_{j=0}^{\infty} \frac{\theta^j}{j!} e^{j+1} \sum_{k=0}^{r} \frac{(-1)^{r+k}\, {}^rC_k}{(j+1)^{\frac{k}{\psi}+1}} \int_{j+1}^{\infty} m^{\frac{k}{\psi}} e^{-m} dm.$$

Hence, the rth moment of the NHNMCF model is

$$E(T^r) = \frac{-p\log(p)}{\phi^r} \sum_{j=0}^{\infty} \frac{(-\log(p))^j}{j!} e^{j+1} \sum_{k=0}^{r} \frac{(-1)^{r+k}\, {}^rC_k}{(j+1)^{\frac{k}{\psi}+1}} \Gamma\left(\frac{k}{\psi}+1, j+1\right), \tag{29}$$

where $\Gamma\left(\frac{k}{\psi}+1, j+1\right)=\int_{j+1}^{\infty} m^{\frac{k}{\psi}} e^{-m} dm$ is the complementary incomplete gamma function. To obtain the first moment, let $r=1$ in (29). This gives

$$\begin{aligned}
E(T) &= \frac{-p\log(p)}{\phi}\sum_{j=0}^{\infty}\frac{(-\log(p))^j}{j!}e^{j+1}\sum_{k=0}^{1}\frac{(-1)^{1+k}\,{}^{1}C_k\,\Gamma\left(\frac{k}{\psi}+1, j+1\right)}{(j+1)^{\frac{k}{\psi}+1}} \\
&= \frac{-p\log(p)}{\phi}\sum_{j=0}^{\infty}\frac{(-\log(p))^j}{j!}e^{j+1}\left[\frac{-\Gamma(1, j+1)}{(j+1)}+\frac{\Gamma\left(\frac{1}{\psi}+1, j+1\right)}{(j+1)^{\frac{1}{\psi}+1}}\right] \\
&= \frac{-p\log(p)}{\phi}\sum_{j=0}^{\infty}\frac{(-\log(p))^j}{j!}\left[-\frac{-1}{(j+1)}+\frac{e^{j+1}\Gamma\left(\frac{1}{\psi}+1, j+1\right)}{(j+1)^{\frac{1}{\psi}+1}}\right]
\end{aligned}$$

since $\Gamma(1, j+1)=\int_{j+1}^{\infty} m^0 e^{-m} dm = e^{-(j+1)}$. The second moment is obtain as

$$\begin{aligned}
E\left(T^2\right) &= \frac{-p\log(p)}{\phi^2}\sum_{j=0}^{\infty}\frac{(-\log(p))^j}{j!}e^{j+1}\sum_{k=0}^{2}\frac{(-1)^{2+k}\,{}^{2}C_k}{(j+1)^{\frac{k}{\psi}+1}}\Gamma\left(\frac{k}{\psi}+1, j+1\right) \\
&= \frac{-p\log(p)}{\phi^2}\sum_{j=0}^{\infty}\frac{(-log(p))^j}{j!}e^{j+1}\left[\frac{\Gamma(1, j+1)}{(j+1)}-\frac{2\Gamma\left(\frac{1}{\psi}+1, j+1\right)}{(j+1)^{\frac{1}{\psi}+1}}\right. \\
&+ \left.\frac{\Gamma\left(\frac{2}{\psi}+1, j+1\right)}{(j+1)^{\frac{2}{\psi}+1}}\right] \\
&= \frac{-p\log(p)}{\phi^2}\sum_{j=0}^{\infty}\frac{(-\log(p))^j}{j!}\left[\frac{1}{(j+1)}-\frac{2e^{j+1}\Gamma\left(\frac{1}{\psi}+1, j+1\right)}{(j+1)^{\frac{1}{\psi}+1}}\right. \\
&+ \left.\frac{e^{j+1}\Gamma\left(\frac{2}{\psi}+1, j+1\right)}{(j+1)^{\frac{2}{\psi}+1}}\right].
\end{aligned}$$

Following the same procedure, the third and fourth moments are given by

$$\begin{aligned}
E\left(T^3\right) &= \frac{-p\log(p)}{\phi^3}\sum_{j=0}^{\infty}\frac{(-\log(p))^j}{j!}e^{j+1}\sum_{k=0}^{3}\frac{(-1)^{3+k}\,{}^{3}C_k}{(j+1)^{\frac{k}{\psi}+1}}\Gamma\left(\frac{k}{\psi}+1, j+1\right) \\
&= \frac{-p\log(p)}{\phi^3}\sum_{j=0}^{\infty}\frac{(-\log(p))^j}{j!}e^{j+1}\left[\frac{-\Gamma(1, j+1)}{(j+1)}+\frac{3\Gamma\left(\frac{1}{\psi}+1, j+1\right)}{(j+1)^{\frac{1}{\psi}+1}}\right. \\
&- \left.\frac{3\Gamma\left(\frac{2}{\psi}+1, j+1\right)}{(j+1)^{\frac{2}{\psi}+1}}+\frac{\Gamma\left(\frac{3}{\psi}+1, j+1\right)}{(j+1)^{\frac{3}{\psi}+1}}\right]
\end{aligned}$$

and

$$\begin{aligned}
E\left(T^4\right) &= \frac{-p\log(p)}{\phi^4}\sum_{j=0}^{\infty}\frac{(-\log(p))^j}{j!}e^{j+1}\sum_{k=0}^{4}\frac{(-1)^{4+k}\,{}^{4}C_k}{(j+1)^{\frac{k}{\psi}+1}}\Gamma\left(\frac{k}{\psi}+1, j+1\right) \\
&= \frac{-p\log(p)}{\phi^4}\sum_{j=0}^{\infty}\frac{(-\log(p))^j}{j!}e^{j+1}\left[\frac{\Gamma(1, j+1)}{(j+1)}-\frac{4\Gamma\left(\frac{1}{\psi}+1, j+1\right)}{(j+1)^{\frac{1}{\psi}+1}}\right. \\
&+ \left.\frac{6\Gamma\left(\frac{2}{\psi}+1, j+1\right)}{(j+1)^{\frac{2}{\psi}+1}}-\frac{4\Gamma\left(\frac{3}{\psi}+1, j+1\right)}{(j+1)^{\frac{3}{\psi}+1}}+\frac{\Gamma\left(\frac{4}{\psi}+1, j+1\right)}{(j+1)^{\frac{4}{\psi}+1}}\right],
\end{aligned}$$

respectively.

4. Simulation Study

Simulation studies was considered in this section in order to ascertain the performance of the NHNMCF model. The performance of the MLE of the parameters ϕ, ψ and p of the NHNMCF model discussed in section 2 were checked using simulation studies. The simulation study was carried out using the following algorithm:

1. Generate a random sample $u_i \sim U(0,1)$ for $i = 1, 2, \cdots, n$.
2. For the cure fraction parameter p, return $t = \frac{1}{\phi}\left[1 - log\left[1 - \frac{log(1-u)}{log(p)}\right]\right]^{\frac{1}{\psi}} - \frac{1}{\phi}$ for $u_i < 1 - p$ otherwise t_i is infinity.
3. Generate a sample of the censoring times $c_i \sim NH(\psi, \phi, p)$ for $i = 1, 2, \cdots, n$.
4. Compute $z_i = min\left(t_i, c_i\right), \delta_i = I\left(t_i \leq c_i\right), i = 1, 2, \cdots, n$
5. The observed data set $D = \{(z_i, \delta_i), i = 1, 2, \cdots, n\}$ are realizations from the NHNMCF model with right censoring.

Table 1 Summary statistics with 0.05 non-susceptible proportion value

n	parameters	estimates	bias	SE	MSE
		$\phi = 1.5,\ \psi = 0.5$			
100	ϕ	1.5094	0.0094	0.8727	1.4943
	ψ	0.6599	0.4099	0.3792	0.4682
	p	0.0495	-0.0005	0.0327	0.002
200	ϕ	1.5032	0.0032	0.5784	1.1147
	ψ	0.5560	0.3060	0.1884	0.2298
	p	0.0500	0.0000	0.0228	0.0012
300	ϕ	1.4986	-0.0014	0.4475	0.4196
	ψ	0.5335	0.2835	0.1359	0.1351
	p	0.0502	0.0002	0.0183	0.0007
400	ϕ	1.4974	-0.0026	0.3826	0.3113
	ψ	0.5264	0.2764	0.1111	0.1163
	p	0.0500	0.0000	0.0157	0.0005
500	ϕ	1.4984	-0.0016	0.3391	0.2344
	ψ	0.5189	0.2689	0.0956	0.0978
	p	0.0501	0.0001	0.0104	0.0004
		$\phi = 0.75,\ \psi = 0.25$			
100	ϕ	0.6929	-0.0571	0.3997	0.3907
	ψ	0.5551	0.3051	0.2630	0.6936
	p	0.0542	0.0042	0.0550	0.0044
200	ϕ	0.6835	-0.0665	0.2689	0.1994
	ψ	0.5390	0.2890	0.1609	0.8084
	p	0.0501	0.0001	0.0399	0.0034
300	ϕ	0.6824	-0.0676	0.2189	0.1511
	ψ	0.4249	0.2749	0.1211	0.6923
	p	0.0483	-0.0017	0.0328	0.0023
400	ϕ	0.6871	-0.0629	0.1914	0.1270
	ψ	0.3140	0.2640	0.1010	0.6667
	p	0.0475	-0.0025	0.0290	0.0017
500	ϕ	0.6942	-0.0558	0.1706	0.1072
	ψ	0.2846	0.2346	0.0830	0.5567
	p	0.0470	-0.0030	0.0267	0.0016

Samples of different sizes viz: $n = 100, n = 200, n = 300, n = 400$ and $n = 500$ from the NHNMCF model were generated assuming some arbitrary values for the shape, scale and proportion of non-susceptible parameters. Furthermore, the censoring indicator was assumed to follow the NH distribution.

All simulation results were replicated 5000 times. These results were obtained considering the maximum likelihood estimators. Additionally, bias, standard error (SE) and mean square error (MSE) were the performance measures used in assessing the performances of the estimates.

Table 2 Summary statistics with 0.10 non-susceptible proportion value

n	parameters	estimates	bias	SE	MSE
		$\phi = 1.5, \psi = 0.5$			
100	ϕ	1.5176	0.0176	0.8256	1.5437
	ψ	0.6262	0.3762	0.3074	0.3804
	p	0.0994	-0.0006	0.0436	0.0039
200	ϕ	1.5053	0.0053	0.5420	0.6388
	ψ	0.5456	0.2956	0.1555	0.1736
	p	0.0998	-0.0002	0.0299	0.0018
300	ϕ	1.5021	0.0021	0.4321	0.3878
	ψ	0.5275	0.2775	0.1134	0.1075
	p	0.1001	0.0001	0.0242	0.0012
400	ϕ	1.4944	-0.0056	0.3692	0.2795
	ψ	0.5212	0.2712	0.0948	0.0933
	p	0.1001	0.0001	0.0208	0.0009
500	ϕ	1.5021	0.0021	0.3291	0.2205
	ψ	0.5149	0.2649	0.0822	0.0848
	p	0.1000	0.0000	0.0186	0.0007
		$\phi = 0.75,\ \psi = 0.25$			
100	ϕ	0.7088	-0.0412	0.4125	0.3969
	ψ	0.5262	0.2762	0.2289	0.6883
	p	0.0958	-0.0042	0.0768	0.0114
200	ϕ	0.7113	-0.0387	0.2810	0.2058
	ψ	0.4647	0.2147	0.1399	0.5427
	p	0.0927	-0.0073	0.0566	0.0055
300	ϕ	0.7128	-0.0372	0.2285	0.146
	ψ	0.3535	0.2035	0.1109	0.5322
	p	0.0912	-0.0088	0.0490	0.0013
400	ϕ	0.7131	-0.0369	0.1965	0.117
	ψ	0.3312	0.1812	0.0914	0.4515
	p	0.0919	-0.0081	0.0427	0.004
500	ϕ	0.7174	-0.0326	0.1748	0.096
	ψ	0.3138	0.1638	0.0897	0.6257
	p	0.0927	-0.0073	0.0389	0.0027

Table 1 gives the simulation result based on 5000 replications for the parameter settings $\phi = 1.5, \psi = 0.5, p = 0.05$ and $\phi = 0.75, \psi = 0.25, p = 0.05$ while, Table 2 gives the simulation results for 5000 replications for the parameter settings $\phi = 1.5, \psi = 0.5, p = 0.10$ and $\phi = 0.75, \psi = 0.25, p = 0.10$ and finally, Table 3 gives the simulation results for the settings $\phi = 1.5, \psi = 0.5, p = 0.30$ and $\phi = 0.75, \psi = 0.25, p = 0.30$.

These tables give the mean estimates of the parameters together with bias, SE and MSE of the estimates. On average, the estimates gets closer to the true parameter values as sample size increases for all the parameters in the different settings.

Table 3 Summary statistics with 0.30 non-susceptible proportion value

n	parameters	estimates	bias	SE	MSE
		$\phi = 1.5,\ \psi = 0.5$			
100	ϕ	1.5080	0.0080	0.8625	1.7506
	ψ	0.6477	0.3977	0.2916	0.5928
	p	0.2999	-0.0001	0.0619	0.0079
200	ϕ	1.4992	-0.0008	0.5736	0.7285
	ψ	0.5519	0.3019	0.1402	0.1434
	p	0.3003	0.0003	0.0428	0.0037
300	ϕ	1.5045	0.0045	0.4617	0.4535
	ψ	0.5288	0.2788	0.1040	0.1025
	p	0.2996	-0.0004	0.0349	0.0025
400	ϕ	1.5158	0.0158	0.4016	0.3445
	ψ	0.5192	0.2692	0.0869	0.0892
	p	0.3002	0.0002	0.0302	0.0018
500	ϕ	1.5085	0.0085	0.3559	0.2661
	ψ	0.5165	0.2665	0.0764	0.0839
	p	0.3000	0.0000	0.0270	0.0015
		$\phi = 0.75,\ \psi = 0.25$			
100	ϕ	0.7829	0.0329	0.5156	0.6758
	ψ	0.3413	0.0913	0.2097	0.4862
	p	0.2741	-0.0259	0.1515	0.0527
200	ϕ	0.7773	0.0273	0.3463	0.2756
	ψ	0.2791	0.0291	0.1178	0.0611
	p	0.2820	-0.0180	0.1132	0.0597
300	ϕ	0.7759	0.0259	0.2794	0.1781
	ψ	0.2598	0.0098	0.0878	0.0215
	p	0.2835	-0.0165	0.0887	0.0223
400	ϕ	0.7722	0.0222	0.2386	0.1285
	ψ	0.2571	0.0071	0.0739	0.0144
	p	0.2876	-0.0124	0.0745	0.0132
500	ϕ	0.7713	0.0213	0.2121	0.1006
	ψ	0.2539	0.0039	0.0642	0.0097
	p	0.2906	-0.0094	0.0641	0.0114

As expected also, the estimates of the bias, SE and MSE for each of the examined parameter value in the different simulation settings gets closer to zero as sample size increases. Hence, these results show that based on the method of estimation, the NHNMCF model has a good performance overall.

5. Applications

In this section, some real life data sets were used in demonstrating the applicability of the proposed NHNMCF model. Three medical data sets were used: Diabetic retinotophy data, colorectal cancer data and Gastric cancer data. The Diabetic retinotophy data and colorectal cancer data sets were used in demonstrating the applicability of the proposed methodology in comparison to some competing models. On the other hand, the Gastric cancer data was used in comparing between the fits of the proposed methodology and that of Nadarajah-Haghighi mixture cure fraction (NHMCF) model.

5.1. Diabetic retinopathy data

The diabetic retinopathy data was presented by Huster et al. (1989). The study consists of follow-up times for 197 diabetic patients that were under the age of 60 years. The main purpose of the study is to assess the efficacy of photocoagulation treatment for proliferative retinopathy. The eye of each patient was randomized to laser treatment and the other eye received no treatment. The main event of interest is severe visual loss in each eye. Let T be random variable for the time up to visual loss for the left eye. Death, dropout and termination were the causes of censoring in the study.

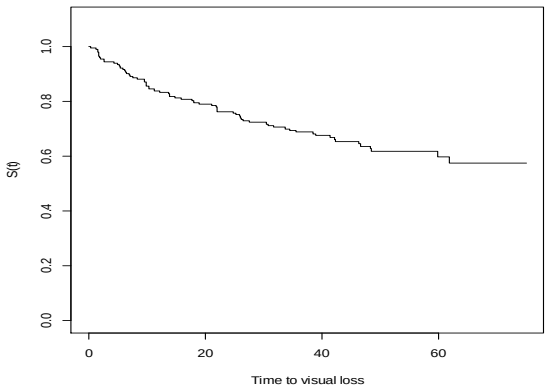

Figure 1 Kaplan-Meier survival curve of the Diabetic retinotophy data

We first assessed the presence of long-term survivors in the data. This is done by plotting Kaplan-Meier survival curve of the data. Figure 1 gives the Kaplan-Meier survival curve of the data. It is observed from the Kaplan-Meier survival curve in Figure 1 that after about 60-months follow-up, some patients have not experienced any recurrence after treatments. That is, the curve level off at a value between 0.56 and 0.60.

According to Corbière et al. (2009); Martinez et al. (2013), the presence of long-term survivors in a data set is seen whenever the Kaplan-Meier survival curve level-off. We therefore conclude that, there is presence of long-term survivors in the data. We then fitted the data to the NHNMCF model in the presence of cure fraction and right censoring and compared its performance with the fits of Rayleigh non-mixture cure fraction (RNMCF), Weibull non-mixture cure fraction (WNMCF), generalized exponential non-mixture cure fraction (GENMCF), modified Weibull non-mixture cure fraction (MWNMCF) and generalized modified Weibull non-mixture cure fraction (GMWNMCF) models.

Assuming whether the covariates: type of treatment and age have effect on the parameters of the model, five(5) different models were fitted. The first, second and third models assumed that the covariates (type of treatment and age) have effect on ϕ, ψ, and p respectively. While the fourth and fifth models assumed that the covariates have effect on ϕ & p and ψ & p respectively.

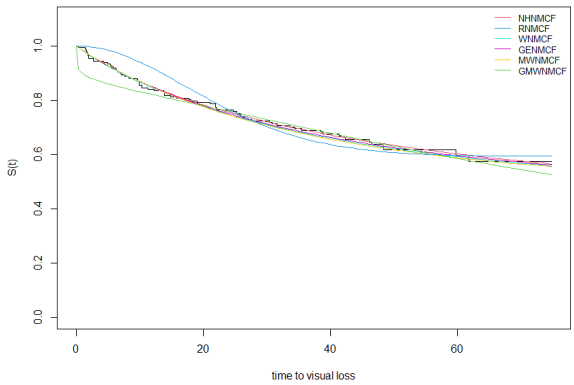

Figure 2 Kaplan-Meier survival curve overlaid with the fits of the MLE estimates for the Diabetic retinotophy data

Table 4 Maximum likelihood estimates for models with cure fraction - Diabetic Retinotophy data

Model	parame-ters	estimate	SE	95% CI	AIC	BIC	CAIC
NHN	ϕ	0.0377	0.0220	(-0.0054,0.0808)	778.7998	788.6494	772.9242
MCF	ψ	0.2909	0.5058	(-0.7005,1.0291)			
	p	0.2214	0.5321	(-0.8215,1.2643)			
RNM	ϕ	19.8360	1.8116	(16.2853,23.3867)	820.4488	827.0152	816.5107
CF	p	0.5935	0.0394	(0.5162,0.6708)			
WNM	ψ	0.9497	0.1663	(0.6237,1.2757)	779.2234	789.073	773.3478
CF	ϕ	0.0227	0.0127	(-0.0021,0.0476)			
	p	0.4563	0.1969	(0.0705,0.8422)			
GEN	ψ	0.9555	0.2021	(-0.4034,0.4415)	779.1828	789.0324	773.3072
MCF	ϕ	0.0191	0.0176	(0.9423,0.9688)			
	p	0.4727	0.1656	(0.4233,0.5221)			
MWN	ϕ	1.0558	0.0344	(0.9884,1.1232)	781.1994	794.3322	773.4077
MCF	ψ	0.0026	0.0058	(-0.0139,0.0087)			
	α	0.0237	0.0135	(-0.0028,0.0502)			
	p	0.4718	0.1117	(0.2529,0.6907)			
GMWN	ϕ	0.0082	0.0018	(2.4131,3.912)	878.7926	895.2086	869.1067
MCF	ψ	3.1625	0.3824	(0.0046,0.0117)			
	α	2.5380	0.6465	(1.2708,3.8052)			
	γ	0.0340	0.0168	(0.0011,0.0668)			
	p	0.3649	0.0587	(0.2498,0.4800)			

Table 4 gives the MLE results for the fitted NHNMCF model together with the fits of RNMCF, WNMCF, GENMCF, MWNMCF and GMWNMCF models. Standard error and 95% confidence interval (95% CI) were given. The information criteria: AIC, BIC and CAIC for the fitted non-mixture cure fraction models were also provided. The information criteria values for these fitted non-mixture cure fraction models indicates that the NHNMCF model is the best fitted model. The Kaplan-Meier survival curve of the fitted diabetic retinopathy data was compared with the fits of the aforementioned models. Careful observation of the curves in Figure 2 also showed that the NHNMCF model is closer to the Kaplan-Meier curves in comparison to all the fitted models.

In analyzing the diabetic retinopathy data via the non-classical method of estimation, gamma priors were assumed for both ϕ and ψ while, beta prior was assumed for the cure fraction parameter. To be specific, $\phi \sim Gamma(1,1), \psi \sim Gamma(1,1)$ and $p \sim Beta(1,1)$. We further assumed prior independence among the parameters included in the model. MCMC technique was applied in obtaining posterior summaries from the joint posterior distribution. We generated 1,100,000 samples for each parameter of interest from the posterior distribution. However, the first 100,000 samples were discarded as burn-in-samples in order to reduced the effect of initial values. To have an uncorrelated value, samples were taken at every 100th sample. Hence, all posterior summaries of interest were based on 10,000 samples.

Table 5 Posterior summaries assuming models with cure fraction - Diabetic Retinotophy data

Model	Parameters	median	sd	95% CrI	DIC
NHNMCF	ϕ	0.0368	0.0273	(0.0089, 0.1125)	766.8
	ψ	0.4250	0.4528	(0.1072, 1.7640)	
	p	0.3439	0.1555	(0.0329, 0.5858)	
RNMCF	ψ	0.0026	0.0004	(0.0018, 0.0034)	821.3
	p	0.5926	0.0387	(0.5146, 0.6651)	
WNMCF	ϕ	0.0212	0.0101	(0.0074, 0.0462)	777.7
	ψ	0.9206	0.1231	(0.7065, 1.1900)	
	p	0.4399	0.1380	(0.0864, 0.6110)	
MWNMCF	α	0.0171	0.0088	(0.0016, 0.0351)	778.3
	ϕ	0.0088	0.0100	(0.0003, 0.0372)	
	ψ	0.6302	0.3447	(0.0371, 1.2760)	
	p	0.4965	0.0949	(0.2474, 0.6194)	
GMWNMCF	α	0.1563	0.1613	(0.0116, 0.6158)	773.4
	ϕ	1.9460	1.1170	(0.7859, 5.0450)	
	γ	0.5213	0.2011	(0.2586, 1.0430)	
	ψ	0.0057	0.0058	(0.0002, 0.0212)	
	p	0.5115	0.1229	(0.1552, 0.6359)	
GENMCF	ϕ	1.0020	0.1673	(0.7330, 1.3820)	778.0
	ψ	0.0218	0.0109	(0.0042, 0.0456)	
	p	0.4933	0.1147	(0.1667, 0.6251)	

Table 5 gives the posterior summaries of the fits of NHNMCF, RNMCF, WNMCF, MWNMCF, GMWNMCF and GENMCF models. The table gives the estimates of the posterior median, standard deviation, 95% credible interval (95% CrI) and DIC value for each fitted model. In this method also, the DIC value for the NHNMCF model shows that the proposed NHNMCF model is a strong competitor.

We further compared the fits of the NHNMCF, RNMCF, WNMCF, GENMCF, MWNMCF and GMWNMCF parametric curves together with the Kaplan-Meier survival curve. Careful observation of the curves in Figure 3 showed that, the curve for the NHNMCF model is the most closet to the Kaplan-Meier survival curve in comparison to the fits of other parametric curves.

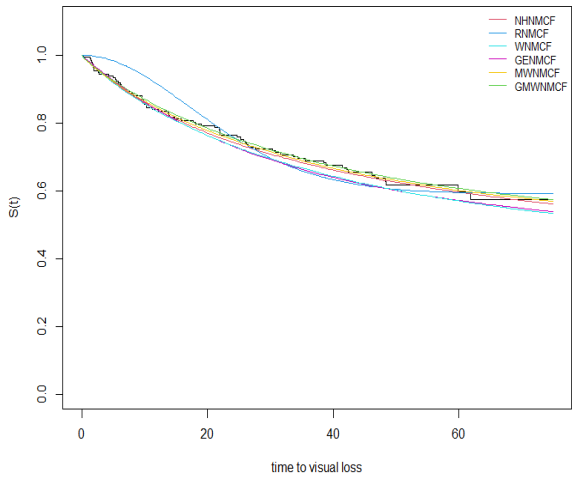

Figure 3 Kaplan-Meier survival curve overlaid with the fits of the Bayesian estimates for the Diabetic retinotophy data

Table 6 Maximum likelihood estimates for models assuming covariates - Diabetic retinopathy data

Model	Parameters	estimate	SE	95% CI	AIC	BIC	CAIC
Model I	ϕ_0	-3.9943	1.1198	(-6.1891,-1.7996)	759.7190	776.1350	750.0331
	ϕ_1	0.0087	0.0108	(-0.0125,0.0299)			
	ϕ_2	-1.6376	0.3764	(-2.3755,-0.8998)			
	ψ	0.1743	1.4712	(-2.7092,3.0579)			
	p	0.0033	0.1503	(-0.2912,0.2979)			
Model II	ϕ	0.0162	0.0246	(-0.0319,0.0643)	758.0024	774.4184	748.3165
	ψ_0	-0.1306	1.7408	(-3.5425,3.2813)			
	ψ_1	0.0080	0.0095	(-0.0107,0.0267)			
	ψ_2	-1.3705	0.3143	(-1.9865,-0.7545)			
	p	0.2797	0.2259	(-0.163,0.7225)			
Model III	ϕ	0.0529	0.0279	(-0.0017,0.1076)	759.4818	775.8978	749.7959
	ψ	0.3900	0.1965	(0.0048,0.7752)			
	η_0	-0.8378	0.7288	(-2.2662,0.5906)			
	η_1	-0.0145	0.0139	(-0.0417,0.0126)			
	η_2	1.9727	0.4751	(1.0416,2.9039)			
Model IV	ϕ_0	-2.2138	0.9103	(-3.9979,-0.4297)	767.8086	790.7910	754.4012
	ϕ_1	-0.0331	0.0263	(-0.0847,0.0185)			
	ϕ_2	-0.2771	1.1346	(-2.5008,1.9466)			
	ψ	0.2718	0.0932	(0.089,0.4545)			
	η_0	-0.9813	1.2460	(-3.4235,1.4608)			
	η_1	-0.0395	0.0356	(-0.1093,0.0303)			
	η_2	2.2989	1.1720	(0.0017,4.596)			
Model V	ϕ	0.0080	0.0027	(0.0026,0.0133)	763.8370	786.8194	750.4296
	ψ_0	-0.6293	0.5164	(-1.6414,0.3829)			
	ψ_1	-0.0112	0.0150	(-0.0406,0.0182)			
	ψ_2	-1.7809	0.3956	(-2.5563,-1.0055)			
	η_0	-2.3417	1.3723	(-5.0313,0.3479)			
	η_1	-0.1274	0.0984	(-0.3201,0.0654)			
	η_2	-2.4522	2.0084	(-6.3886,1.4843)			

Considering the type of treatment and age as covariates, five different models were fitted as mentioned earlier. Table 6 gives the results of the fits for these models via MLE method of estimation. The AIC, BIC and CAIC values from this table showed that model II is the best fitted model.

Table 7 Posterior summaries for models assuming covariates - Diabetic retinopathy data

Model	Parameters	median	sd	95% CrI	DIC
model I	ϕ_0	-3.0880	0.6618	(-4.4190, -1.8330)	758.7
	ϕ_1	-0.0012	0.0119	(-0.0255, 0.0211)	
	ϕ_2	-1.5800	0.3727	(-2.3390, -0.8700)	
	ψ	0.3746	0.2884	(0.1133, 1.2210)	
	p	0.2073	0.1182	(0.0191, 0.4454)	
model II	ϕ	0.02586	0.02187	(0.0031, 0.0848)	753.6
	ψ_0	-0.6193	0.8071	(-1.8990, 1.3710)	
	ψ_1	0.0071	0.0093	(-0.0114, 0.0249)	
	ψ_2	-1.3300	0.3032	(-1.9640, -0.7690)	
	p	0.2329	0.1175	(0.0263, 0.4548)	
model III	ϕ	0.0301	0.0235	(0.0072, 0.0949)	746
	ψ	0.5768	0.5278	(0.1755, 2.2120)	
	n_0	-0.9317	0.7329	(-2.7650, 0.1076)	
	n_1	-0.0093	0.0144	(-0.0384, 0.0182)	
	n_2	1.8450	0.4474	(1.0390, 2.8180)	
model IV	ϕ_0	-2.5730	0.7296	(-4.0580, -1.2260)	747
	ϕ_1	-0.0294	0.0206	(-0.0657, 0.0149)	
	ϕ_2	-0.5972	0.9592	(-2.2640, 1.2250)	
	ψ	0.3900	0.3123	(0.0980, 1.2700)	
	n_0	-0.4616	0.7824	(-2.2520, 0.7869)	
	n_1	-0.0478	0.0721	(-0.2682, 0.0088)	
	n_2	1.3900	1.0560	(-1.2250, 2.9040)	
model V	ϕ	0.0274	0.0228	(0.0037, 0.0898)	734
	ψ_0	-0.5648	0.8626	(-2.1300, 1.2750)	
	ψ_1	-0.0142	0.0203	(-0.0477, 0.0301)	
	ψ_2	-1.2490	0.7615	(-2.2890, 0.6526)	
	n_0	-0.8354	0.9896	(-3.1010, 0.7354)	
	n_1	-0.0533	0.1543	(-0.4810, 0.0324)	
	n_2	0.0923	1.4200	(-2.6970, 2.6140)	

On the other hand, the procedure mentioned earlier in this section for the Bayesian method was also followed in this case. However, we further assumed normal prior distribution (N(1, 0.5)) for the regression parameters. Table 7 gives the posterior summaries for the fitted models. The DIC value of each fitted model is also provided. Unlike the result of the MLE method, the best fitted model using the Bayesian method of estimation is the model that considered covariates effect on ψ and p.

5.2. Colorectal cancer data

Colorectal cancer has been ranked the third most commonly diagnosed malignancy Naishadham et al. (2011) and the second and third most frequent cancer in women and men respectively Naishadham et al. (2011); Magaji et al. (2014). It is also ranked the fourth leading cause of cancer related death in the world Magaji et al. (2014, 2017). Medical records of 166 patients diagnosed of colorectal cancer who underwent treatment in University of Malaya medical center (UMMC) between January 2001 and December 2010 were obtained. The record included 86 patients who were treated by surgery alone and 80 patients who were treated by surgery and chemotherapy/radiotherapy. The

survival time was defined to be the time from the date of commencement of treatment to death, loss to follow-up or end of the study.

Table 8 Maximum likelihood estimates for models with cure fraction - colorectal cancer data

Model	Parameters	estimate	SE	95% CI	AIC	BIC	CAIC
NHN	ϕ	0.0721	0.0478	(-0.0216,0.1658)	408.0324	417.3684	402.1805
MCF	ψ	0.9430	0.2479	(0.4571,0.558)			
	p	0.0150	0.0025	(0.0124,0.0199)			
RNM	ϕ	2.7348	0.2631	(2.2193,3.2504)	489.7586	495.9826	485.8322
CF	p	0.1339	0.0404	(0.0549,0.213)			
WNM	ψ	0.9537	0.0904	(0.7765,1.1308)	408.4104	417.7464	402.5585
CF	ϕ	0.0678	0.0421	(-0.0147,0.1503)			
	p	0.0114	0.0286	(-0.0447,0.0675)			
GEN	ψ	0.9373	0.1311	(-0.178,0.3261)	408.5790	417.9150	402.7272
MCF	ϕ	0.0741	0.1201	(0.8392,1.0354)			
	p	0.0299	0.1241	(-0.3324,0.3921)			
MWN	ϕ	3.5742	0.9864	(1.6408,5.5076)	412.9170	425.3650	405.1654
MCF	ψ	0.0003	0.0004	(-0.0004,0.0010)			
	α	0.1663	0.0451	(0.0779,0.2546)			
	p	0.1792	0.0780	(0.0263,0.3321)			
GMWN	ϕ	0.0755	0.0462	(0.1352,3.5796)	411.8884	427.4483	402.2634
MCF	ψ	1.8574	0.8787	(-0.0151,0.1660)			
	α	0.3674	0.2921	(-0.2051,0.9400)			
	γ	0.4897	0.1956	(0.1064,0.8730)			
	p	0.0066	0.0140	(-0.0209,0.0340)			

Table 9 Posterior summaries for models with cure fraction - colorectal cancer data

Model		median	sd	95% CrI	DIC
NHN	ϕ	0.1327	0.1030	(0.0331, 0.4216)	400.4
MCF	ψ	0.8887	0.7126	(0.2744, 3.0640)	
	p	0.0672	0.0573	(0.0034, 0.2142)	
RNM	ψ	0.1386	0.0256	(0.0934, 0.1930)	489.7
CF	p	0.1417	0.0398	(0.0729, 0.2275)	
WNM	ϕ	0.1201	0.0448	(0.0542, 0.2283)	408.3
CF	ψ	0.9609	0.0903	(0.7956, 1.1490)	
	p	0.0733	0.0577	(0.0046, 0.2188)	
MWN	α	0.0960	0.0513	(0.0096, 0.2070)	408.5
MCF	ϕ	0.0311	0.0414	(0.0010, 0.1516)	
	ψ	0.8851	0.4151	(0.2070, 1.9580)	
	p	0.0975	0.0597	(0.0112, 0.2379)	
GMWN	α	0.1962	0.2040	(0.0134, 0.7711)	403.9
MCF	ϕ	1.2700	0.7994	(0.5015, 3.5500)	
	γ	0.6557	0.3495	(0.2713, 1.6280)	
	ψ	0.0880	0.0541	(0.0093, 0.2176)	
	p	0.0968	0.0584	(0.0075,0.2261)	
GEN	ϕ	0.9606	0.1042	(0.7754, 1.1810)	408.4
MCF	ψ	0.1125	0.0520	(0.0396, 0.2395)	
	p	0.0779	0.0569	(0.0053, 0.2178)	

Table 10 Maximum likelihood estimates for models assuming covariates - colorectal cancer data

Model	Parameters	estimates	SE	95% CI	AIC	BIC	CAIC
Model I	ϕ_0	-2.7257	1.3794	(-5.4292, -0.0221)	405.7418	418.1898	397.9902
	ϕ_1	0.5809	0.2654	(0.0608, 1.1010)			
	ψ	0.7783	1.8583	(-2.8639, 4.4206)			
	p	0.0133	0.0821	(-0.1477, 0.1742)			
Model II	ϕ	0.1013	0.1242	(-0.1420, 0.3446)	406.2462	418.6942	398.4946
	ψ_0	-0.8561	3.3429	(-7.4081, 5.6960)			
	ψ_1	0.4785	0.2342	(0.0195, 0.9375)			
	p	0.0043	0.0620	(-0.1172, 0.1259)			
Model III	ϕ	0.0949	0.1190	(-0.1382, 0.3281)	406.3150	418.7630	398.5634
	ψ	0.4075	0.4240	(-0.4236, 1.2386)			
	η_1	-6.1142	1.4079	(-8.8737, -3.3547)			
	η_2	-3.3766	0.6998	(-4.7483, -2.0050)			
Model IV	ϕ_0	-2.8583	3.1862	(-9.1031, 3.3864)	407.3478	422.9078	397.7228
	ϕ_1	1.1899	3.7591	(-6.1778, 8.5576)			
	ψ_0	-0.1299	3.1202	(-6.2454, 5.9856)			
	ψ_1	-0.5867	3.3879	(-7.2269, 6.0535)			
	p	0.0106	0.0271	(-0.0426, 0.0638)			
Model V	ϕ_0	-2.8280	0.6095	(-4.0226, -1.6334)	407.3490	422.9089	397.7240
	ϕ_1	1.4681	0.1975	(1.0811, 1.8551)			
	ψ	0.4258	0.2211	(-0.0077, 0.8592)			
	η_1	-8.5804	1.7960	(-12.1004, -5.0604)			
	η_2	4.4249	1.7253	(1.0435, 7.8064)			
Model VI	ϕ	0.1113	0.1401	(-0.1633, 0.3859)	408.2252	423.7851	398.6002
	ψ_0	-2.2207	1.0823	(-4.3420, -0.0994)			
	ψ_1	1.3957	0.3751	(0.6606, 2.1309)			
	η_1	-19.6073	1.9508	(-23.4309, -15.7838)			
	η_2	11.8385	3.4740	(5.0297, 18.6473)			
Model VII	ϕ_0	-13.7274	0.4424	(-14.5945, -12.8602)	408.7100	427.3819	397.2383
	ϕ_1	12.2391	0.9591	(10.3593, 14.1189)			
	ψ_0	9.5010	0.6265	(8.2731, 10.7289)			
	ψ_1	-10.6311	0.7854	(-12.1704, -9.0918)			
	η_1	-13.6472	7.4300	(-28.2097, 0.9153)			
	η_2	7.6831	5.2795	(-2.6645, 18.0306)			

Different statistical methods have been used to analyzed medical information of patients suffering from colorectal cancer. These include: Magaji et al. (2014) provides an analysis on colorectal cancer patients who underwent treatment in the University of Malaya Medical Center from 2001 to 2010, the rates of survival and its predictors among colorectal cancer patients in Malaysia was studied by Magaji et al. (2017), also in Malaysia, survival analysis and prognostic factors for colorectal cancer patients was studied by Hassan et al. (2016). While Ghazali (2018) modelled the survival time and incidence for colorectal cancer patients. In Thai, Kittrongsiri et al. (2020) assess the overall and stage-specific colorectal cancer survival and identify the prognostic factors among the patients.

Table 11 Posterior summaries for models assuming covariates considering colorectal cancer data

Model	Parameters	median	sd	95% CrI	DIC
Model I	ϕ_0	-1.9940	0.5706	(-3.1740, -0.9393)	404.4
	ϕ_1	0.5724	0.2764	(0.0468, 1.1320)	
	ψ	0.6410	0.4620	(0.2158, 1.9620)	
	p	0.0599	0.0513	(0.0031, 0.1931)	
Model II	ϕ	0.0863	0.0925	(0.0094, 0.3540)	398.6
	ψ_0	-0.0124	0.8654	(-1.3500, 2.1020)	
	ψ_1	0.4841	0.2279	(0.0362, 0.9288)	
	p	0.0677	0.0492	(0.0050, 0.1871)	
Model III	ϕ	0.1376	0.1041	(0.0370, 0.4273)	392.5
	ψ	0.9089	0.6960	(0.3085, 3.0150)	
	η_0	-1.8630	0.7464	(-3.6210, -0.7426)	
	η_1	-1.2810	0.6196	(-2.6380, -0.2058)	
Model IV	ϕ_0	-2.1030	0.7105	(-3.6330, -0.8077)	405.2
	ϕ_1	0.2096	0.8870	(-1.6320, 1.8330)	
	ψ_0	-0.3088	0.6708	(-1.5240, 1.0840)	
	ψ_1	0.3073	0.7662	(-1.0560, 1.9790)	
	p	0.0695	0.0506	(0.0049, 0.1918)	
Model V	ϕ_0	-1.8300	0.5886	(-3.0330, -0.7250)	404.2
	ϕ_1	0.2581	0.4489	(-0.6129, 1.1320)	
	ψ	0.7828	0.5142	(0.2983, 2.2740)	
	η_0	-1.8730	0.7768	(-3.6370, -0.6270)	
	η_1	-0.7385	0.8858	(-2.5530, 0.9680)	
Model VI	ϕ	0.0942	0.0944	(0.0110, 0.3586)	395.3
	ψ_0	0.2015	0.8385	(-1.1130, 2.1700)	
	ψ_1	0.1355	0.3973	(-0.6419, 0.9123)	
	η_0	-1.9340	0.8143	(-3.8190, -0.6799)	
	η_1	-0.9873	1.0020	(-3.1130, 0.8886)	
Model VII	ϕ_0	-2.0330	0.7082	(-3.5480, -0.7837)	405.8
	ϕ_1	0.0765	0.9077	(-1.8440, 1.7320)	
	ψ_0	-0.0881	0.6804	(-1.2900, 1.3760)	
	ψ_1	0.0821	0.8194	(-1.4380, 1.8160)	
	η_0	-1.9130	0.8273	(-3.8430, -0.6273)	
	η_1	-1.0190	1.0110	(-3.1210, 0.9041)	

The Kaplan-Meier survival curve of the data was plotted and it was observed that, the curve level-off after about 3,000 days follow-up at a value close to 0.15, this suggest the presence of cure fraction in the data. Hence, the data was used in fitting the proposed NHNMCF model and compared its performance with that of RNMCF, WNMCF, GENMCF, MWNMCF and GMWNMCF models using MLE and Bayesian methods of estimation in the presence of right censoring and covariates. Table 8 gives the MLE estimates for the colorectal cancer data. The SE, 95% CI, AIC, BIC and CAIC values of the fitted models were also given. From the information criteria values, it was observed that the NHNMCF model fits the colorectal cancer data better than the RNMCF, WNMCF, MWNMCF, GMWNMCF and GENMCF models.

On the other hand, the Bayesian posterior summaries were given in Table 9. The DIC value for the NHNMCF model is the least among the DICs of other fitted non-mixture cure fraction models. Hence, the fit of NHNMCF model to the colorectal cancer data is more efficient than the fits of RNMCF, WNMCF, MWNMCF, GMWNMCF and GENMCF models.

We further analyzed the data taking type of treatment as covariate. Tables 10 and 11 gives

respectively the MLE estimates and Bayesian posterior summaries of the NHNMCF model taking type of treatment as covariate. Seven models were fitted: model I, II and III assumed that the covariate have effect on ϕ, ψ and p respectively. The models assuming covariate effect on ϕ and ψ, ϕ and p, ψ and p are respectively models IV, V and VI while model VII assumed that the covariate have effect on all of the parameters of the NHNMCF model. The AIC, BIC and CAIC values of the fitted models indicates that overall, the model that assumed covariate effect on ϕ is the best model. While the fits of the Bayesian method indicates that the best fitted model is the model that assumed covariate effects on the cure fraction parameter.

5.3. Gastric cancer data

Gastric cancer is the fifth leading cause of cancer related death despite its decrease in incidence and mortality in the world Ferlay et al. (2019); Talebi et al. (2020). However, the 5-year survival rate following all type of resections has increased significantly. According to Zare et al. (2013), the best treatment of gastric cancer at the initial stage is surgery. Radiotherapy and chemotherapy are used as renewed treatment, if necessary. However, most gastric cancer patients are diagnosed at a stage when common treatments such as gastrectomy, chemotherapy or radiotherapy may not be effective in increasing the survivorship of patients Sadighi et al. (2005). Retrospective study of patients with gastric adenocarcinoma who underwent curative resection between January 2002 to December 2007 at the Barretos Cancer Hospital was conducted by Jácome et al. (2013). The study consists of two hundred and one (201) patients with different clinical stages of gastric adenocarcinoma. One hundred and twenty-five (125) patients received resection only while the remaining seventy-six (76) received adjuvan chemoradiotherapy (CRT). The event of interest was defined as the time (in months) from the date of surgery until death. The data was used in comparing between the performances of the NHMCF and NHNMCF models.

Overall, 53.2% of the survival time of the patients are censored. While 50.4% and 57.9% of those patients treated with surgery alone and CRT respectively are censored.

Table 12 Maximum likelihood estimates of Gastric cancer data

Model	Parameter	estimate	SE	95% CI	AIC	BIC	CAIC
NHMCF	ϕ	0.0359	0.0137	(0.0090, 0.0628)	900.0038	909.9137	894.0647
	ψ	1.3184	0.5279	(0.2836, 1.0706)			
	p	0.4079	0.0903	(0.2309, 0.5848)			
NHNMCF	ϕ	0.0007	0.0001	(0.0004, 0.0010)	898.6282	908.5381	892.6891
	ψ	0.5420	5.3575	(43.7495, 10.5011)			
	p	0.4700	0.0520	(0.3684, 0.5723)			

Table 13 Posterior summary results of Gastric cancer data

Model	Parameter	median	sd	95% CrI	DIC
NHMCF	ϕ	0.0241	0.0310	(0.0045, 0.1145)	829.6
	ψ	1.9510	2.6770	(0.4153, 10.3800)	
	p	0.4214	0.1218	(0.0597, 0.5346)	
NHNMCF	ϕ	0.0100	0.0161	(0.0013, 0.0633)	812.7
	ψ	3.0460	6.0340	(0.3358, 23.0100)	
	p	0.3911	0.1269	(0.0583, 0.5314)	

Tables 12 and 13 respectively gives the MLE estimates and the Bayesian posterior summaries of the gastric cancer data. From these fits, the non-mixture model fits the data better than the mixture model. That is, The NHNMCF model is more efficient than the NHMCR model since it has the lowest AIC, BIC CAIC and DIC values. This confirms the results of Coelho-Barros et al. (2017) and Kutal and Qian (2018).

6. Conclusions

In survival analysis, the population of study may contain individuals that are susceptible to the event of interest and those individuals that are non-susceptible. To analyze data of such type, the cure fraction models are used. In this article, a non-mixture cure fraction model was studied using the NH distribution. The model was studied in the presence of right censoring and covariates. We studied some statistical properties of the model and estimated its parameters via MLE and Bayesian methods of estimation. Simulation study was carried out under different parameter settings. The results showed that, the proposed methodology has a good performance. Furthermore, two medical data sets were used in comparing the performance of the proposed methodology with the fits of RNMCR, WNMCR, GENMCR, MWNMCR and GMWNMCR models. The results of the fits showed that the proposed methodology fits the data better than the other models. The data sets were then fitted to the proposed methodology in the presence of covariate. For the diabetic retinotophy data, it was found that, age and type of treatment have effect on the shape parameter when MLE was used while it was observed that the covariates have effect on the shape and cure fraction parameters when the data was fitted using Bayesian method. On the other hand, the results for the colorectal cancer data showed that the type of treatment have effect on the scale parameter when the data was fitted via MLE while the covariate was shown to have effect on the shape parameter when Bayesian method.
was used. Finally, gastric cancer data was used to compare between the performance of the mixture and non-mixture cure fraction models. It was found that, the non-mixture model is more efficient than the mixture cure fraction model.

Acknowledgments

The authors want to thank the editor and the anonymous reviewers for their valuable comments and suggestions which have led to improvement in this paper.

References

Achcar JA, Coelho-Barros EA, Mazucheli J. Cure fraction models using mixture and non-mixture models. Tatra Mt Math Publ. 2012; 51(1): 1-9.

Berkson J, Gage RP. Survival curve for cancer patients following treatment. J Am Stat Assoc. 1952; 47(259): 501-515.

Boag JW. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. J Roy Stat Soc B Met. 1949; 11(1): 15-53.

Chen M, Ibrahim JG, Sinha D. A new Bayesian model for survival data with a surviving fraction. J Am Stat Assoc. 1999; 94(447): 909-919.

Coelho-Barros EA, Achcar JA, Mazucheli J. Cure Rate Models Considering The Burr XII Distribution in Presence of Covariate. 0.0 J Stat Theory App. 2017; 16(2): 150-164.

Corbière F, Commenges D, Taylor JMG, Joly P. A penalized likelihood approach for mixture cure models. Stat Med. 2009; 28(3): 510-524.

Farewell VT. Mixture models in survival analysis: Are they worth the risk?. Can J Stat. 1986; 14(3): 257-262.

Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, Znaor A, Bray F. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer. 2019; 144(8): 1941-1953.

Gamel JW, McLean IW, Rosenberg SH. Proportion cured and mean log survival time as functions of tumour size. Stat Med. 1990; 9(8): 999-1006.

Ghazali AK. Modelling of survival and incidence for colorectal cancer in Malaysia. PhD [dissertation]. Lancaster University; 2018.

Hassan MRA, Suan MAM, Soelar SA, Mohammed NS, Ismail I, Ahmad F. Survival analysis and prognostic factors for colorectal cancer patients in Malaysia. Asian Pac J Cancer Prev. 2016; 17(7): 3575-3581.

Herring AH, Ibrahim JG. Maximum likelihood estimation in random effects cure rate models with nonignorable missing covariates. Biostatistics. 2002; 3(3): 387-405.

Huster WJ, Brookmeyer R, Self SG. Modelling paired survival data with covariates. Biometrics. 1989; 45(1): 145-156.

Ibrahim JG, Chen M, Sinha D. Bayesian semiparametric models for survival data with a cure fraction. Biometrics. 2001; 57(2): 383-388.

Jácome AAA, Wohnrath DR, Neto CS, Fregnani JHTG, Quinto AL, Oliveira A TT, Vazquez VL, Fava G, Martinez EZ, Santos JS. Effect of adjuvant chemoradiotherapy on overall survival of gastric cancer patients submitted to D2 lymphadenectomy. Gastric Cancer. 2013; 16(2): 233-238.

Kannan N, Kundu D, Nair P, Tripathi RC. The generalized exponential cure rate model with covariates. J Appl Stat. 2010; 37(10): 1625-1636.

Kittrongsiri K, Wanitsuwan W, Prechawittayakul P, Sangroongruangsri S, Cairns J, Chaikledkaew U. Survival analysis of colorectal cancer patients in a Thai hospital-based cancer registry. Expert Rev Gastroenterol Hepatol. 2020; 14(4): 291-300.

Kutal D, Qian L. A Non-Mixture Cure Model for Right-Censored Data with Fréchet Distribution. Stats. 2018; 1(1): 176-188.

Liu H, Shen Y. A semiparametric regression cure model for interval-censored data. J Am Stat Assoc. 2009; 104(487): 1168-1178.

Lopes CMC, Bolfarine H. Random effects in promotion time cure rate models. Comput Stat Data An. 2012; 56(1): 75-87.

Magaji BA, Moy FM, Roslani AC, Law CW. Descriptive epidemiology of colorectal cancer in University Malaya Medical Centre, 2001 to 2010. Asian Pac J Cancer Prev. 2014; 15(15): 6059-6064.

Magaji BA, Moy FM, Roslani AC, Law CW. Survival rates and predictors of survival among colorectal cancer patients in a Malaysian tertiary hospital. BMC Cancer. 2017; 17(1): 1-8.

Martinez EZ, Achcar JA, Jácome AAA, Santos J. Mixture and non-mixture cure fraction models based on the generalized modified Weibull distribution with an application to gastric cancer data. Comput Meth Prog Bio. 2013; 112(3): 343-355.

Mazucheli J, Coelho-Barros EA, Achcar JA. The exponentiated exponential mixture and non-mixture cure rate model in the presence of covariates. Comput Meth Prog Bio. 2013; 112(1): 114-124.

Meeker WQ. Limited failure population life tests: application to integrated circuit reliability. Technometrics. 1987; 29(1): 51-65.

Nadarajah S, Haghighi F. An extension of the exponential distribution. Statistics. 2011; 45(6): 543-558.

Naishadham D, Lansdorp-Vogelaar I, Siegel R, Cokkinides V, Jemal A. State disparities in colorectal cancer mortality patterns in the United States. Cancer Epidemiol. Biomarkers Prev. 2011; 20(7): 1296-1302.

Ng SK, McLachlan GJ. On modifications to the long-term survival mixture model in the presence of competing risks. J Stat Comput Sim. 1998; 61(1-2): 77-96.

Peng Y, Dear KBG, Denham JW. A generalized F mixture model for cure rate estimation. Stat Med. 1998; 17(8): 813-830.

Sadighi S, Raafat J, Mohagheghi MA, Meemary F. Gastric carcinoma: 5 year experience of a single institute. Asian Pac J Cancer Prev. 2005; 6(2): 195-6.

Shao Q, Zhou X. A new parametric model for survival data with long-term survivors. Stat Med. 2004; 23(22): 3525-3543

Sy JP, Taylor JMG. Estimation in a Cox proportional hazards cure model. Biometrics. 2000; 56(1): 227-236.

Talebi A, Mohammadnejad A, Akbari A, Pourhoseingholi MA, Doosti H, Moghimi-Dehkordi B, Agah S, Bahardoust M. Survival analysis in gastric cancer: a multi-center study among Iranian patients. BMC Surg. 2020; 20(1): 1-8.

Tsodikov AD, Ibrahim JG, Yakovlev AY. Estimating cure rates from survival data: an alternative to two-component mixture models. J Am Stat Assoc. 2003; 98(464): 1063-1078.

Uddin MT, Islam MN, Ibrahim QIU. An Analytical approach on cure rate estimation based on uncensored data. J Appl Sci. 2006; 6(3): 548-552.

Uddin MT, Sen A, Noor MS, Islam MN, Chowdhury ZI. An analytical approach on non-parametric estimation of cure rate based on uncensored data. J Appl Sci. 2006; 6(6): 1258-1264.

Usman U, Suleiman S, Arkilla BM, Aliyu Y. Nadarajah-Haghighi Model for Survival Data With Long Term Survivors in the Presence of Right Censored Data. Pakistan J Stat Oper Res. 2021; 17(3): 695-709.

Yakovlev AY, Asselain B, Bardou VJ, Fourquet A, Hoang T, Rochefediere A, Tsodikov AD. A simple stochastic model of tumor recurrence and its application to data on premenopausal breast cancer. Biometrie et analyse de donnees spatio-temporelles. 1993; 12: 66-82.

Yakovlev AY, Tsodikov AD, Asselain B. Stochastic models of tumor latency and their biostatistical applications. Singapore: World Scientific; 1996.

Zare A, Mahmoodi M, Mohammad K, Zeraati H, Hosseini M, Naieni KH. Survival analysis of patients with gastric cancer undergoing surgery at the iran cancer institute: a method based on multi-state models. Asian Pac J Cancer Prev. 2013; 14(11): 6369-6373.