RESEARCH ARTICLE

Editorial Process: Submission:06/30/2023 Acceptance:12/18/2023

Non-Mixture Cure Model Estimation in Bladder Cancer Patients: A Novel Approach with Exponentiated Weibull Exponential Distribution

Mohamed Elamin Omer[1,2]*, Mohd Mustafa[2], Norhaslinda Ali[2], Nur Haizum Abd Rahman[2]

Abstract

Objective: Cure models are frequently used in survival analysis to account for a cured fraction in the data. When there is a cure rate present, researchers often prefer cure models over parametric models to analyse the survival data. These models enable the ability to define the probability distribution of survival durations for patients who are at risk. Various distributions can be considered for the survival times, such as Exponentiated Weibull Exponential (EWE), Exponential Exponential (EE), Weibull and lognormal distribution. The objective of this research is to choose the most appropriate distribution that accurately represents the survival times of patients who have not been cured. This will be accomplished by comparing various non-mixture cure models that are based on the EWE distribution with its sub-distributions, and distributions distinct from those belonging to the EWE distribution family. **Material and Methods:** A sample of 85 patients diagnosed with superficial bladder tumours was selected to be used in fitting the non-mixture cure model. In order to estimate the parameters of the suggested model, which takes into account the presence of a cure rate, censored data, and covariates, we utilized the maximum likelihood estimation technique using R software version 3.5.7. **Result:** Upon conducting a comparison of various parametric models fitted to the data, both with and without considering the cure fraction and without incorporating any predictors, the EE distribution yields the lowest AIC, BIC, and HQIC values among all the distributions considered in this study, (1191.921/1198.502, 1201.692/1203.387, 1195.851/1200.467). Furthermore, when considering a non-mixture cure model utilizing the EE distribution along with covariates, an estimated ratio was obtained between the probabilities of being cured for placebo and thiotepa groups (and its 95% confidence intervals) were 0.76130 (0.13914, 6.81863). **Conclusion:** The findings of this study indicate that EE distribution is the optimal selection for determining the duration of survival in individuals diagnosed with bladder cancer.

Keywords: Non-mixture cure models- exponentiated Weibull exponential distribution- maximum likelihood estimation

Asian Pac J Cancer Prev, **24 (12)**, 4167-4177

Introduction

Bladder cancer is a type of cancer that develops in the tissues of the bladder, the organ responsible for storing urine. It is particularly prevalent among older adults and ranks as one of the most common forms of cancer. Several causes and risk factors have been identified, including smoking, which poses the most significant risk. The probability of developing bladder cancer is at least three times higher in smokers than in non-smokers. Occupational exposure to certain chemicals like aromatic amines found in dyes, paints, and solvents also increases the risk. Bladder cancer is more commonly observed in older adults and affects men more frequently than women. Chronic bladder inflammation resulting from urinary tract infections or irritation may elevate the risk, as does previous cancer treatment involving radiation or specific chemotherapy drugs.

Although the absence of a functional National Cancer Registry in Sudan makes it challenging to provide precise cancer burden data, GLOBOCAN's estimate suggests that cancer ranks as the second most common cause of mortality among the Sudanese population. In 2020, there were 598 newly diagnosed bladder cancer cases and 409 fatalities attributed to bladder cancer (International Agency for Research on Cancer, 2020).

Based on the projections of the American Cancer Society (2023), it is expected that, in the United States, there is an estimated number of 82,290 newly diagnosed cases of bladder cancer. Of these cases, around 62,420 will affect men, while about 19,870 will occur in women. Additionally, it is estimated that there will be around

*[1]Department of Mathematics, College of Science, Sudan University of Science and Technology, Khartoum, Sudan. [2]Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, UPM Serdang, Serdang 43400, Selangor, Malaysia. *For Correspondence: elamin_omer@yahoo.com*

16,710 deaths from bladder cancer, with approximately 12,160 occurring in male and the remaining in female. It is worth noting that both the incidence of new bladder cancer cases and the mortality rates associated with the disease have been declining in recent years. While bladder cancer ranks as the fourth most common cancer in men, it is less prevalent among women (National Cancer Institute, 2023).

Treatment options for bladder cancer vary depending on the stage and severity of the disease. Surgery is a common approach and may involve transurethral resection of bladder tumour (TURBT), partial cystectomy (removing a portion of the bladder), or radical cystectomy (removing the entire bladder). Intravesical therapy is another option, which entails administering medications directly into the bladder. High-energy radiation is employed in radiation therapy to eradicate cancer cells, while chemotherapy employs drugs to kill cancer cells or impede their growth. Immunotherapy is an emerging treatment modality that stimulates the immune system to identify and combat cancer cells, offering a promising approach for bladder cancer patients.

It is crucial to understand that treatment approaches may vary, and individualized treatment plans are developed by a team of healthcare experts who consider the unique circumstances of each patient's case.

In the field of survival analysis, it is common for researchers to utilize Kaplan-Meier estimates or the log-rank test, as well as semi-parametric models like the Cox proportional hazards model. Alternatively, analysts may also consider parametric models based on well-known distributions, taking into account covariates (Cox, 1972). In the realm of cancer research, the utilization of the Weibull distribution is widespread for its flexibility in modelling survival data (Bradburn et al., 2003b). It offers several benefits, including the ability to capture various hazard rate patterns such as decreasing, constant, and increasing hazards over time. This flexibility is particularly useful when analysing cancer survival data, as the hazard rates can exhibit different trends throughout the disease progression. Researchers rely on the Weibull distribution to estimate survival probabilities, evaluate the influence of covariates on survival outcomes, and compare survival rates among different patient groups or treatments. The Weibull distribution's versatility and capacity to accommodate diverse hazard rate patterns make it a valuable tool in cancer research. However, medical research datasets often require more sophisticated parametric models. Consequently, to address this challenge, numerous authors in the existing literature have introduced novel classes of parametric distributions that build upon the Weibull distribution. These include the exponentiated Weibull distribution proposed by Mudholkar and Srivastava (1993), the generalized modified Weibull distribution introduced by Carrasco et al., (2008), the log-beta Weibull distribution presented by (Ortega et al., 2013), the EWE distribution suggested by Elgarhy et al., (2017), and the X-Gamma inverse Weibull distribution proposed by Ibrahim and Almetwally (2021). These new distributions have been developed to better capture the characteristics of medical data and provide improved modelling approaches.

Survival data analysis, including studies focused on cancer, often involves scenarios where a subset of the population does not encounter the specific event of interest. In these instances, patients are commonly split into two categories: those who encountered the event under examination and thus faced potential risks, and those who did not encounter it and therefore were not vulnerable. These individuals are categorized as either healed or protected through immunization, indicating that they have a reduced probability of experiencing the event compared to the exposed group. In such circumstances, when the survival time distribution for susceptible patients is defined, researchers typically favour cure fraction models over parametric models. Cure fraction models have a significant role in analysing survival data involving long-term survivors, and they are considered an expanded version of conventional survival models. These models have been a subject of research since the 1940s. There are two primary categories of cure fraction models: mixture cure models and non-mixture cure models. In the case of the mixture cure model, the assumption is that the population can be divided into two groups: the cured (unsusceptible) and the uncured (susceptible). The concept of the mixture cure model was initially proposed by Boag (1949) and it was further developed by Berkson and Gag (1952) after three years. Numerous researchers have extensively investigated the mixture cure model, including (Farewell, 1982; Goldman, 1984; Kuk and Chen, 1992; Maller and Zhou, 1996; Peng and Dear, 2000), and (Patilea and Van Keilegom, 2017), among others. Chen et al., (1999) argued that the mixture cure model does not validate the proportional hazard property for the entire population. In contrast, the non-mixture cure model was proposed to uphold this property for all observations while allowing direct inference of the predictors influencing the probability of being cured (Klebanov et al., 1993; Tsodikov, 1998). Tsodikov et al., (2003) in their analysis of the previous study on the bounded cumulative hazard model using statistical inference, highlighted the simplicity of computations, meaningful biological interpretation, and flexible structure of the bounded cumulative hazard model for the survival function, which could provide technical advantages in designing maximum likelihood estimation procedures. The literature has proposed various approaches to model the non-mixture cure model, as documented in (Uddin et al., 2006; Liu and Shen, 2009; Bremhorst and lambert, 2016; Kutal and Qian, 2018).

The presence of individuals who are considered cured within a sample dataset is often illustrated by a Kaplan-Meier curve, which exhibits a plateau with dense censoring at the right end (Corbière et al., 2009). Various statistical methods have been proposed by several authors to model the proportion of cured individuals. For further exploration, interested readers can refer to works such as those by (Maller and Zhou, 1992; Lu, 2010; López-Cheda et al., 2017; Omer et al., 2021). Additionally, some authors have suggested the use of maximum likelihood estimation techniques, as seen in works by (Farewell, 1982; Ghitany and Maller, 1992; Sy and Taylor, 2000), and others.

The objective of this research was to select a

parametric distribution that best fit the survival times of uncured patients considering the bladder cancer dataset.

Materials and Methods

Bladder cancer data

In this article, we considered a bladder cancer dataset obtained from the Veterans Administration Cooperative Urological Research Group. The dataset includes 85 patients who initially had superficial bladder tumors when they enrolled in the trial. These tumors were surgically removed via transurethral procedures, and the patients were randomly assigned to two groups: the placebo or control group (Group 1), which comprised 47 patients, and the thiotepa group, which included 38 patients. The primary focus of the analysis was the time to cancer recurrence, serving as the event of interest. Notably, about 67% of the data were censored observations.

Thiotepa is a type of anticancer medication known as an alkylating agent. It works by impeding the proliferation and division of cancerous cells. When used intravesical, meaning it is directly administered into the bladder, it can effectively treat superficial bladder tumors. Due to limited data beyond the fourth recurrence, only the first four recurrence times are included in the analysis. Each recurrence time is measured from the start of the patient's treatment. Byar (1980) suggests that one of the methods to assess the effectiveness of thiotepa is to analyze the tumor recurrence times of patients from both treatment groups. The dataset is structured in the competing risks format, following the description provided in the paper by Wei et al., (1989). This dataset is available in the 'survival' package of the R software (Terry M Therneau, 2023).

Non-mixture cure model

Non-mixture cure models, also referred to as bounded cumulative hazard (BCH) models and promotion time cure models, are statistical tools employed in survival analysis to accommodate the possibility of a cure or long-term survival for all individuals in a population. Unlike mixture cure models that assume two distinct groups (susceptible and cured), non-mixture cure models treat the entire population as potentially curable. Various methods have been proposed to formulate non-mixture cure models, including parametric and nonparametric approaches. Parametric models assume a specific distribution for the non-cured fraction, such as the exponential or Weibull distribution, and estimate the parameters through maximum likelihood estimation. Nonparametric methods, on the other hand, do not rely on strong distributional assumptions and estimate the survival function using the Kaplan-Meier approach.

Non-mixture cure models have found applications in diverse research domains, including cancer studies, where the objective is to estimate the probability of long-term survival or cure for patients. These models provide insights into the factors associated with cure and facilitate understanding the effectiveness of treatments and predicting long-term outcomes.

In this paper, we assumed BCH model, which was developed under the assumption that the distribution of the remaining active cancer cells following treatment has a Poisson distribution. Yakovlev et al., (1993) made the initial proposal for this model, which was subsequently elaborated upon by Chen et al., (1999). Tsodikov et al., (2003) studied the non-mixture cure model and accentuated its advantages, including its proportional hazard model structure, meaningful interpretation of results from a biological perspective, and simplified computational procedures enabled by its straightforward survival function structure in maximum likelihood estimation techniques. The research conducted by Herring and Ibrahim (2002) focused on the parametric estimation of random effects within a non-mixture cure model, specifically addressing the presence of non-ignorable missing covariates. Uddin et al., (2006) investigated both non-parametric and parametric approaches within a non-mixture model to analyse uncensored data.

Let T represents the time at which the event occurs and let $0 < \pi < 1$ be a portion of cured patients. Moreover, for uncured individuals, presume that $S_{uc}(t)$ and $f_{uc}(t)$ denote survival and probability density functions, respectively. Therefore, the population survival function of the BCH model is

$$S_{pop}(t) = \pi^{F_{uc}(t)} = \exp\left[\ln(\pi) F_{uc}(t)\right] \quad (1)$$

where= $F_{uc}(t)$=1-$S_{uc}(t)$ is the cumulative distribution function for susceptible subjects.

Assuming a random sample (t_j, δ_j) of size n, j=1, ..., n, the j[th] individual's contribution to the likelihood function can be expressed as follows:

$$\begin{aligned} L_j &= \left[h(t_j)\right]^{\delta_j} S(t_j) \\ &= \left\{-\left[\ln(\pi)\right] f_{uc}(t_j)\right\}^{\delta_j} \exp\left[\ln(\pi) F_{uc}(t_j)\right], \end{aligned}$$

where in this context, the censoring indicator variable δ_j, is defined as follows:

$$\delta_j = \begin{cases} 1, & \text{uncensored} \\ 0, & \text{otherwise.} \end{cases}$$

The Exponentiated Weibull Exponential (EWE) Distribution

Consider the bladder cancer data, and let the susceptible individuals' survival times follow the EWE distribution, which was introduced by Elgarhy et al., (2017). The basis for this distribution lies in the exponentiated Weibull-G family, as proposed by Hassan and Elgarhy (2016) and is referred to as $X \sim EWE\ (\varepsilon, \alpha, \beta, \lambda)$.

The probability density function (pdf) of the EWE distribution for a random variable T can be expressed as follows:

$$f_{ewe}(t) = \frac{\varepsilon\alpha\beta\lambda\left[\exp(\lambda t)-1\right]^{\beta-1}\exp\left(-\left\{\alpha\left[\exp(\lambda t)-1\right]^{\beta}-\lambda t\right\}\right)}{\left(1-\exp\left\{-\alpha\left[\exp(\lambda t)-1\right]^{\beta}\right\}\right)^{1-\varepsilon}}, \quad t>0, \quad (2)$$

where the shape parameters ε and β are positive values, while the scale parameters α and λ are also positive quantities.

The survival and hazard functions of EWE distribution take respectively the following forms

$$S(t)=1-\left(1-\exp\left\{-\alpha\left[\exp(\lambda t)-1\right]^{\beta}\right\}\right)^{\varepsilon},$$

and

$$h(t)=\frac{\varepsilon\alpha\beta\lambda\left[\exp(\lambda t)-1\right]^{\beta-1}\exp\left(-\left\{\alpha\left[\exp(\lambda t)-1\right]^{\beta}-\lambda t\right\}\right)}{\left(1-\exp\left\{-\alpha\left[\exp(\lambda t)-1\right]^{\beta}\right\}\right)^{1-\varepsilon}\left\{1-\left(1-\exp\left\{-\alpha\left[\exp(\lambda t)-1\right]^{\beta}\right\}\right)^{\varepsilon}\right\}}.$$

The EWE distribution is a generalized form that encompasses several special cases, which are outlined below (Elgarhy et al., 2017).

- Exponentiated Exponential Exponential (EEE) distribution: in the case where β=1, Equation (2) takes the form

$$f_{eee}(t)=\frac{\varepsilon\alpha\lambda\exp\left(-\left\{\alpha\left[\exp(\lambda t)-1\right]-\lambda t\right\}\right)}{\left(1-\exp\left\{-\alpha\left[\exp(\lambda t)-1\right]\right\}\right)^{1-\varepsilon}},\quad t>0,$$

which is the pdf of EEE distribution. Moreover, if we substitute β=2 in the same equation, we have

$$f_{ere}(t)=\frac{2\varepsilon\alpha\lambda\left[\exp(\lambda t)-1\right]\exp\left(-\left\{\alpha\left[\exp(\lambda t)-1\right]^{2}-\lambda t\right\}\right)}{\left(1-\exp\left\{-\alpha\left[\exp(\lambda t)-1\right]^{2}\right\}\right)^{1-\varepsilon}},\quad t>0,$$

which represents the pdf of Exponentiated Rayleigh Exponential (ERE) distribution.

- Weibull Exponential (WE) distribution: for ε =*1*, the density function of the EWE distribution reduces to

$$f_{we}(t)=\alpha\beta\lambda\left[\exp(\lambda t)-1\right]^{\beta-1}\exp\left(-\left\{\alpha\left[\exp(\lambda t)-1\right]^{\beta}-\lambda t\right\}\right),\ t>0. \quad (3)$$

Equation (3) represents the pdf of WE distribution, which was proposed by Oguntunde et al., (2015).

- Exponential Exponential (EE) distribution: if we let ε =*1* and β=1 in Equation (2), we obtain the density function

$$f_{ee}(t)=\alpha\lambda\exp\left(-\left\{\alpha\left[\exp(\lambda t)-1\right]-\lambda t\right\}\right),\quad t>0,$$

which is the pdf of EE distribution. If β=2 , in addition to ε =*1*, this case aligns with Rayleigh Exponential (RE) distribution. The pdf of the RE distribution takes the form

$$f_{re}(t)=2\alpha\lambda\left[\exp(\lambda t)-1\right]\exp\left(-\left\{\alpha\left[\exp(\lambda t)-1\right]^{2}-\lambda t\right\}\right),\ t>0$$

The log - likelihood function

Consider the non-mixture cure model in Equation (1), and let θ= (π, ε, α, β, λ) then the likelihood function for θ is given by

$$L(\boldsymbol{\theta})=\prod_{j=1}^{n}\left[-\ln(\pi)f_{ewe}(t_j)\right]^{\delta_j}\pi^{F_{ewe}(t_j)},$$

and the Log-likelihood function can be expressed as

$$l(\boldsymbol{\theta})=\sum_{j=1}^{n}\delta_j\ln\left[-\ln(\pi)\right]+\sum_{j=1}^{n}\delta_j\ln\left[f_{ewe}(t_j)\right]+\sum_{j=1}^{n}\ln(\pi)F_{ewe}(t_j), \quad (4)$$

where f_{ewe} *(t)* and F_{ewe} *(t)* are probability and cumulative density functions for EWE distribution, respectively.

Predictors Influence

For the BCH model (1), we presume that the cure fraction π, and the scale parameter α, can be connected to a vector of predictors v= (v_1,..., v_m) by replacing π and α in the Equation (4) by:

$$\pi_j=1/\left\{1+\exp\left[-\left(\eta_\circ+\eta_1 v_{1j}+\eta_2 v_{2j}+...+\eta_m v_{mj}\right)\right]\right\} \quad (5)$$

and

$$\alpha_j=\exp(\gamma_\circ+\gamma_1 v_1+...+\gamma_m v_{mj}), \quad (6)$$

where, (η_0, η_1,..., η_m) and (γ_0, γ_1, ..., γ_m) are two sets of coefficients to be determined.

For the bladder data, we use the variable (rx: Treatment) as a covariate. In order to assess the influence of the kind of therapy on the probability of being cured, we adopt the subsequent model:

$$\pi_j=1/\left\{1+\exp\left[-(\eta_\circ+\eta_1 rx_j)\right]\right\},$$

where rx_j represents the treatment, a value of 1 corresponds to the placebo treatment and 2 corresponds to the thiotepa treatment. The parameter η_1 is associated with the impact of the treatment on the cure fraction. If the 95% confidence intervals for η_1 include zero, it can be inferred that there is no evidence to support the notion that the therapy has any influence. Moreover, we can establish a connection between the type of treatment and the predictor within the scale control parameter α by replacing α by α_j= exp (γ_0+$\gamma_1 rx_j$) in the Equation (4). Therefore, the parameter γ_1 is related to the effect of the treatment on the appearance and location of the survival curve.

Model selection

In order to assess the comparison of BCH model based on different distributions, three information criteria were employed: the Akaike Information Criteria (AIC) introduced by Akaike (1974), the Bayesian Information Criteria (BIC) suggested by Schwarz (1978), and the Hannan-Quinn Information Criteria (HQIC) proposed by Hannan and Quinn (1979). A lower value of these information criteria indicates a more favourable fit for the model.

The definitions of AIC, BIC, and HQIC are provided below:

AIC= -2 ln [*L(θ)*]+2*h*
BIC=-2 ln [*L(θ)*]+*h*ln (*m*)

$HQIC$= -2 ln [$L(\theta)$]+2$h$ln [ln (m)]

In the given context, The AIC, BIC, and HQIC are calculated using the likelihood function L(θ), as their foundation, where the model's number of free parameters is denoted by h, while m represents the total number of observations.

Results

Figure 1 (a) displays the survival function estimates using the Kaplan-Meier method, specifically analysing data related to bladder cancer. The curve shows a plateau on the right side, indicating a relatively stable survival rate. The height of this plateau is approximately 0.537. Figure 1 (b) displays the estimated survival curves for the placebo and thiotepa groups. These curves show that both groups reach steady plateaus after a certain period of follow-up. For the placebo group, this occurs around 51 months, while for the thiotepa group, it happens at approximately 47 months. These plateaus indicate a relatively constant survival rate for each group beyond those time points.

The estimated parameters of the BCH model are presented in Table 1 obtained by using maximum likelihood estimation method, considering the assumption of the EWE distribution and its specific cases. This table also provides information such as the standard error (SE) for each estimated parameter, as well as the AIC, BIC, and HQIC values. From Table 1, it becomes evident that the AIC values associated with the EE distribution are the smallest among all the cases considered, with a value of 1195.791. Furthermore, upon comparing these models, it is notable that the EEE and EE distributions provide the smallest BIC and HQIC values. Figure 2 presents a more illustrative representation of the model fitting procedure using various probability distributions for the bladder cancer data. It displays the Kaplan-Meier estimators for the survival function alongside the corresponding expected values obtained from the BCH models for each suggested distribution. This visualization aims to provide a clearer understanding of the model fitting process, building upon the findings presented in Table 1. Figure 3 presents the survival functions and corresponding hazard functions obtained from fitting the BCH model using the EWE distribution and its sub-models, as specified in Table 1. Panel (a) of Figure 3 displays the survival functions, while panel (b) showcases the risk functions. Interestingly, in panel (b) of Figure 3, it can be observed that the survival curves generated by the models using the EE and EEE distributions closely resemble the curves estimated by the Kaplan-Meier method. This suggests that these models provide the best fit to the data when compared to the other distributions considered.

Table 2 exhibits the results obtained from the models utilizing the EWE distribution and its sub-models, excluding the probability of being cured π. It is evident from this table that models excluding the cure fraction yield higher values for AIC, BIC, and HQIC compared to models that incorporate the cure fraction (as observed in Tables 1 and 2).

Using the data in hand, with the covariate (type of therapy) included in the probability of surviving, we compare the non-mixture models with EE, Fréchet (FR), and Generailized Modified Weibull (GMW) susceptible distributions using the maximum likelihood method. The density function of Fréchet distribution is defined by,

$$f(t)=\frac{\alpha}{\beta}\left(\frac{\beta}{t}\right)^{\alpha+1} e^{-\left(\frac{\beta}{t}\right)^{\alpha}}, \qquad t>0,$$

where $\alpha > 0$ is the shape parameter and $\beta > 0$ is the scale parameter; and the density function of GMW distribution is defined by

$$f(t)=\frac{\alpha\beta t^{\gamma-1}(\gamma+\lambda t)\exp\left[\lambda t-\alpha t^{\gamma}\exp(\lambda t)\right]}{\left\{1-\exp\left[-\alpha t^{\gamma}\exp(\lambda t)\right]\right\}^{1-\beta}}, \qquad t>0,$$

where $\alpha > 0$, $\gamma \geq 0$, $\lambda \geq 0$, $\beta > 0$.

The maximum likelihood estimations of the parameters, 95% confidence intervals and the values of the three information criteria for the above three susceptible distributions with a non-mixture model for the bladder dataset are shown in Table 3. In these models, the predictor is connected to the cure fraction. Table 3 indicates that EE distribution is the best distribution in comparison with FR

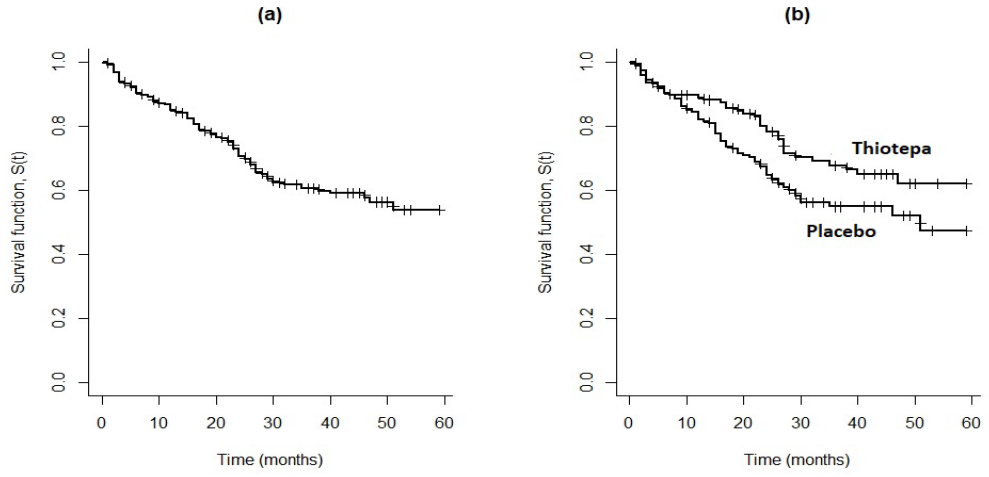

Figure 1. (a) Overall Survival Function Obtained by Kaplan-Meier Technique for the Bladder Cancer Data. (b) Curves of survival functions for each type of treatments.

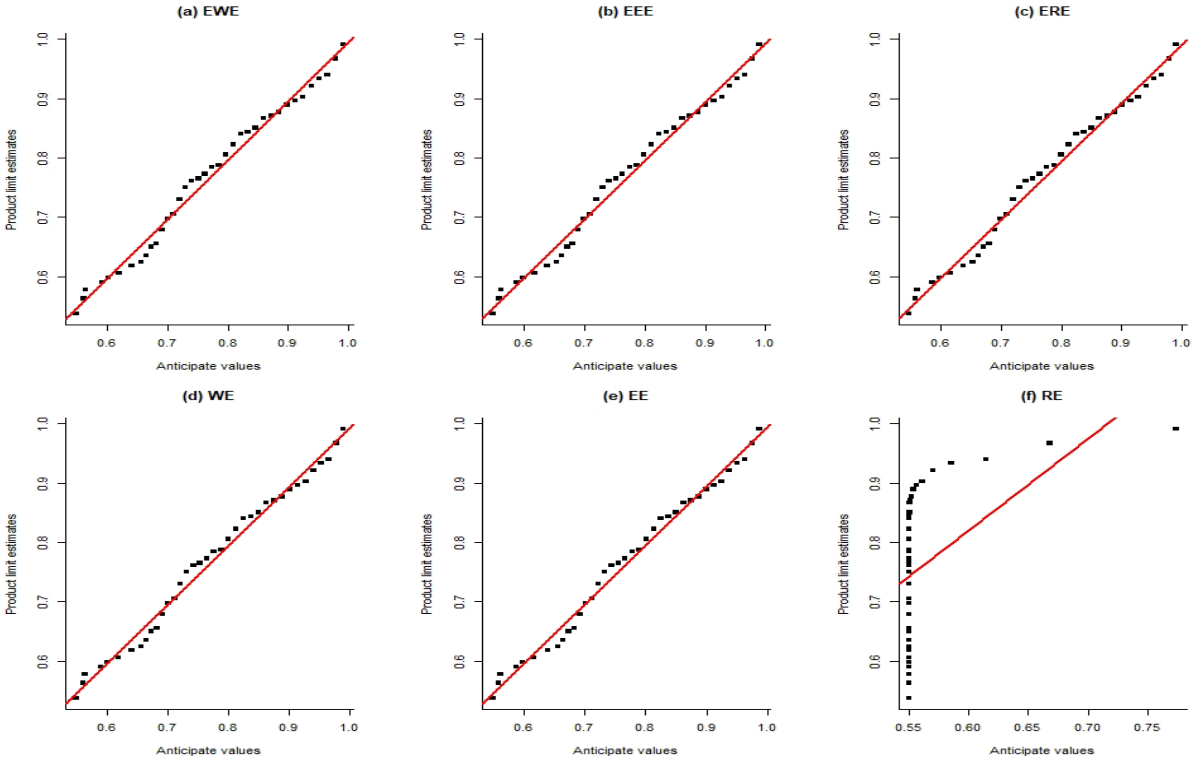

Figure 2. A Comparison between Kaplan-Meier Estimates of the Survival Function and the Corresponding Anticipated Values Obtained from the Non-mixture Models for Various Probability Distributions (as indicated in Table 1, results). The diagonal red lines in the graph indicate a complete agreement between the product-limit estimates and the anticipated values.

Table 1. Results of Maximum Likelihood Estimation for Non-mixture Cure Model with EWE Distribution and Sub-Distributions, excluding Predictors for Bladder Tumor Data.

Model	Parameter	Estimate	SE	AIC	BIC	HQIC
EWE	ε	47.7624	5.1572	1197.211	1209.424	1202.123
	α	3.1579	0.26332			
	β	0.1209	0.02299			
	λ	0.1084	0.04843			
	π	0.5145	0.05764			
EEE	ε	1.1159	0.21992	1197.489	1207.26	1201.42
	α	0.8033	0.89527			
	λ	0.0301	0.02301			
	π	0.5282	0.05442			
ERE	ε	0.5444	0.054	1197.514	1207.284	1201.444
	α	4.6669	1.5589			
	λ	0.0106	0.0023			
	π	0.5311	0.047			
WE	ε	0.8686	0.9674	1197.541	1212.857	1201.471
	β	1.1062	0.16265			
	λ	0.0265	0.02102			
	π	0.5273	0.05318			
EE	α	0.495	0.22361	1195.791	1203.119	1198.738
	λ	0.0395	0.01297			
	π	0.5344	0.04505			
RE	α	205.1849	0.0002	1226.276	1233.604	1229.224
	λ	0.0027	2.985			
	π	0.55	0.0356			

SE, Standard error; AIC, Akaike Information Criteria; BIC, Bayesian Information Criteria; HQIC, Hannan-Quinn Information Criteria

Table 2. Results of Maximum Likelihood Estimation for Non-mixture Cure Model with EWE Distribution and Sub-distributions, excluding both Cure Fraction and Predictors for Bladder Tumor Data

Model	Parameter	Estimate	SE	AIC	BIC	HQIC
EWE	ε	72.29727	4.45773	1197.356	1207.127	1201.286
	α	4.76487	1.23364			
	β	0.1254	0.01764			
	λ	0.01028	0.01664			
EEE	ε	1.002	0.1045	1200.579	1207.907	1203.526
	α	0.1907	2.621			
	λ	0.00068	0.00015			
ERE	ε	0.435	0.0394	1204.275	1211.603	1207.222
	α	26.327	0.7506			
	λ	0.0017 0	0.0002			
WE	ε	40.951	3.846	1200.389	1207.717	1203.336
	β	0.9827	0.0858			
	λ	0.000242	0.00011			
EE	α	20.807	3.434	1198.502	1203.387	1200.467
	λ	0.00046	0.00007			
RE	α	228.0904	2.9939	1300.96	1305.845	1302.925
	λ	0.0012	0.0001			

SE, Standard error; AIC, Akaike Information Criteria; BIC, Bayesian Information Criteria; HQIC, Hannan-Quinn Information Criteria

and GMW distributions. The values of π_0 and π_1, included in Table 3, were calculated using the formulas $\pi_0=1/\{1+\exp[-(\eta_0+\eta_1)]\}$ and $\pi_1= 1/\{1+\exp[-(\eta_0+2\eta_1)]\}$. These values represent the cure fractions for patients who received placebo treatment and thiotepa therapy, respectively. Table 4 presents the outcomes obtained from the BCH model utilizing the EE distribution, where a covariate is included in both the cure rate π and α. According to Table 4, the estimated cure rates for the placebo and thiotepa patient groups are 0.46675 and 0.61720, respectively.

Figure 4 illustrates the death risk functions derived from the EE non-mixture cure model, where a covariate related to the type of treatment is incorporated into both the cure fraction π and the scale parameter α (as presented in Table 3). The graph in Figure 4 demonstrates that the

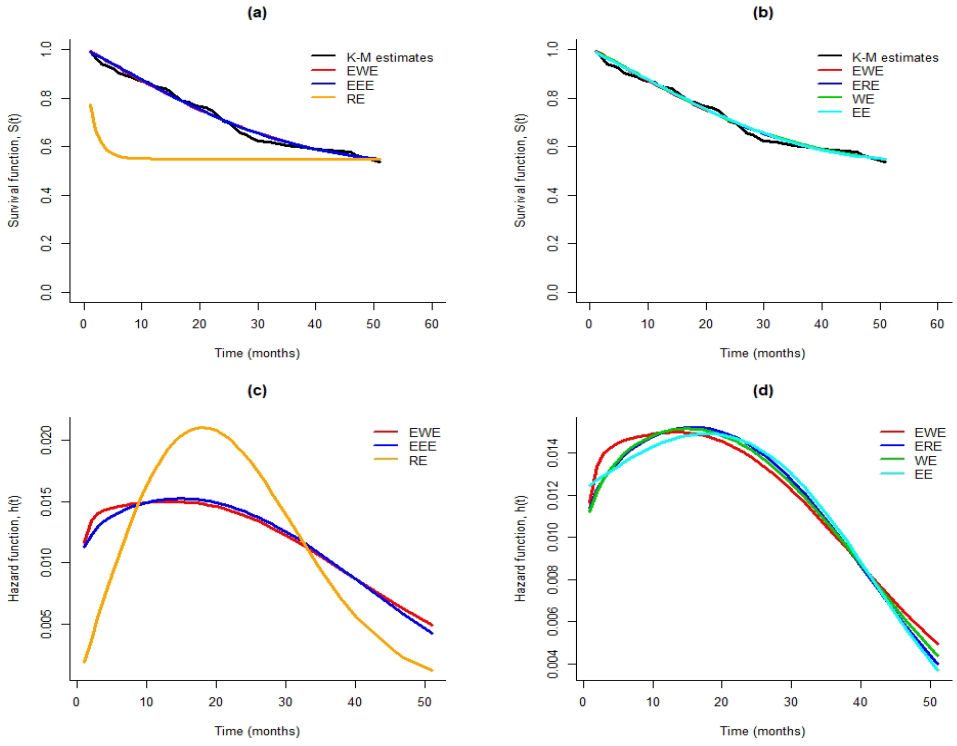

Figure 3. Panels (a) and (b) Display the Fitted Survival Curves Derived from the BCH Model Using the EWE Distribution and Its Sub-distributions for Bladder Cancer Data. The corresponding hazard functions are depicted in panels (c) and (d). In order to enable easy comparisons, all plots present curves based on the EWE distribution

Table 3. Maximum Likelihood Estimates, Considering the Non-Mixture Model with the EE, FR, and GMW Distributions where the Covariate is Including in Cure Fraction π for the Bladder Cancer data.

Model	Parameter	Estimate	95% Confidence interval	AIC	BIC	HQIC
EE	α	0.47932	(0.05423, 0.90440)	1191.921	1201.692	1195.851
	λ	0.03948	(0.01374, 0.06523)			
	η_0	-0.78307	(-1.63424, 0.06811)			
	η_1	0.63038	(0.11562, 1.14514)			
	π_0	0.4619	(0.03667, 0.77087)			
	π_1	0.6172	(0.19733, 0.91360)			
FR	α	0.37055	(0.33982, 0.40129)	1193.983	1203.754	1197.913
	β	169.6564	(163.6185, 175.6943)			
	η_0	-4.50417	(-6.37051, -2.63783)			
	η_1	1.24731	(0.13213, 2.36248)			
	π_0	0.037081	(0.000003, 0.4316)			
	π_1	0.1182	(0.00222, 0.88965)			
GMW	α	2.22249	(1.47529, 2.96970)	1194.323	1217.864	1200.218
	β	40.8064	(6.58439, 75.02841)			
	γ	0.12658	(0.067864, 0.18530)			
	λ	0.00694	(-0.00257, 0.01645)			
	η_0	-1.0944	(-2.61451, 0.42571)			
	η_1	0.66673	(0.01252, 1.32094)			
	π_0	0.39468	(0.06901, 0.85153)			
	π_0	0.55948	(0.06982, 0.95554)			

AIC, Akaike Information Criteria; BIC, Bayesian Information Criteria; HQIC, Hannan-Quinn Information Criteria

hazard functions exhibit a continuous decline over time, starting from approximately 20 months after the initiation of treatment.

Discussion

The main objective of this research is to choose an appropriate probability distribution for the survival times of individuals with bladder cancer who are susceptible to the disease. In order to accomplish this objective, we suggested a BCH model that is based on an EWE distribution and its sub-distributions. This model enhances the range of distributions commonly employed in the analysis of survival data. The EWE distribution offers flexibility in accommodating different types of hazard rate functions, including cases where failure rates follow a bathtub-shaped pattern.

Cure fraction models are developed with the purpose of estimating the probability of an individual being cured. When there are no individuals who have been cured, cure models can be simplified into traditional survival models. There are two primary types of cure models, the first being the mixture cure model, which is commonly used when analysing data that involves long-term survivors. In this model, the population is seen as a combination of individuals who have been cured and those who have not. The advantage of the mixture cure model is that it allows for covariates to have different effects on the cured individuals and the survival times of individuals who are susceptible to the disease. This means that

Table 4. Results of Maximum Likelihood Estimation for Non-mixture Cure Model Based on the EE Distribution with a Covariate Including in both the Cure Fraction π and Scale Parameter α for Bladder Tumor Data.

Parameter	Estimate	95% Confidence interval	AIC	BIC	HQIC
γ_0	-0.65711	(-1.9885, 0.67426)	1193.894	1206.107	1198.806
γ_1	-0.06024	(-0.78640, 0.66591)			
λ	0.03971	(0.01395, 0.06548)			
η_0	-0.72674	(-1.79997, 0.34649)			
η_1	0.59355	(-0.08498, 1.27209)			
π_0	0.46675	(0.13182, 0.83460)			
π_1	0.6131	(0.12240, 0.94738)			
π_0/π_1	0.7613	(0.13914, 6.81863)			

AIC, Akaike Information Criteria; BIC, Bayesian Information Criteria; HQIC, Hannan-Quinn Information Criteria

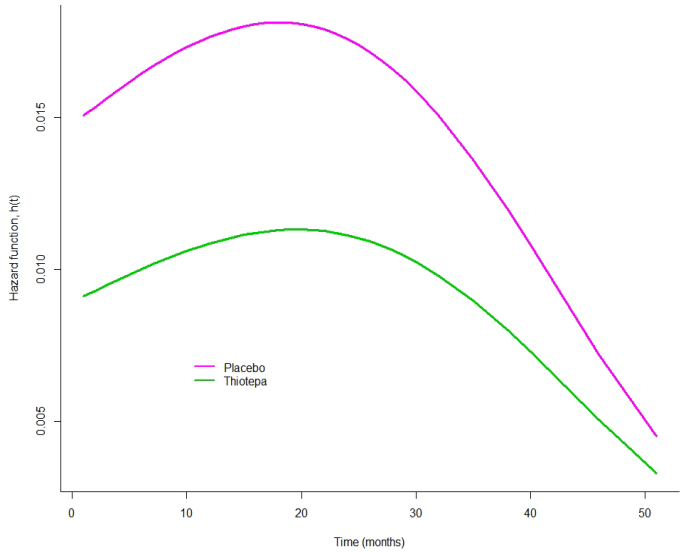

Figure 4. Hazards Functions Derived from the Non-mixture Cure Model with the EE Distribution where a covariate (specifically, the type of treatment) is connected both to the probability of being cured, π, and the scale parameter α.

various covariates can be considered in the two parts of the model (incidence and latency), and the impact of the same covariate (s) on both components can be evaluated. This characteristic sets the mixture cure model apart from other cure models.

In contrast, the mixture cure model lacks the ability to validate the proportional hazard functions property. Additionally, it lacks a clear biological interpretation, particularly when considering cancer recurrence. The other type of cure model, often referred to as the promotion time cure model or non-mixture cure model, assumes that following the initial cancer treatment, there may be residual cancer cells within the patient's body that can gradually develop over time and eventually lead to a detectable cancer relapse. In certain cases, there may be mathematical relationships between the mixture cure model and the promotion time cure model.

The Kaplan-Meier survival curves depicted in this study suggest that using survival models that do not account for the rate of cured individuals π would not be suitable for analysing the bladder dataset. Furthermore, the graph demonstrates that the probability of being cured for the thiotepa group is higher compared to the survival probability in the placebo group. Additionally, the curves exhibit consistent plateaus at the later stages, indicating the presence of individuals who are not susceptible to the disease in both treatment groups.

The bladder data analysis in this study revealed that BCH models utilizing the EE and ERE distributions exhibit a better fit, as they closely match the observed values. Furthermore, the findings indicate that the estimated cure proportion π obtained from cure models based on the EWE, EEE, and WE distributions is lower than the value of π shown in Figure 1 (a), whereas the RE distribution overestimates this value. By employing models based on the ERE and EE distributions, more precise estimates for the parameter π can be obtained, providing additional evidence of a good fit when assuming the BCH model based on the EE distribution. Moreover, the values of π_0 and π_1 obtained by the non-mixture cure model based on EE distribution (results of Table 3) are the closest values to the cure proportions for the placebo and thiotepa groups as shown in Figure 1 (b), and this means that we have an extra proof of a best fit when assuming EE distribution.

The hazard function curves derived from the BCH model based on the EE and ERE distributions exhibit a remarkable similarity. These curves suggest that there is a significant risk of death immediately following the transplant period. Subsequently, this risk gradually decreases until the end of the observation period.

In this study, when comparing Table 1 and Table 2, it is observed that the AIC, BIC, and HQIC values obtained from models that do not exclude the cure rate π are lower than the values obtained from models that exclude π. This finding aligns with expectations and further confirms that the BCH model is highly convenient for analysing the data under the study.

The findings of this study indicate that patients who received thiotepa treatment have a higher cure rate compared to those who received a placebo treatment. This suggests that thiotepa has a significant impact on the cure fraction across all models and improves the probability of being cured. These results align with previous studies conducted by Wei et al., (1989), Lin et al., (1998), and Sun and Wei (2000).

Furthermore, in different approaches to the present study, Zhao and Sun (2011) and Baetschmann and Winkelmann (2013) found significant variations in the recurrence rates of bladder tumours among the three treatment groups (placebo, pyridoxine, and thiotepa). Notably, the thiotepa treatment demonstrated a tumour reducing effect. These findings align with the results obtained in the current study, providing further support and consistent.

An intriguing observation is that the 95% confidence intervals for π_0/π_1 (0.13914, 6.81863) encompass the value of 1, indicating a lack of evidence for differences

in population cure rates between patients treated with placebo and thiotepa. Additionally, the 95% confidence intervals for η_1 (-0.08498, 1.27209) contain zero.

It is important to highlight that the 95% confidence intervals for γ_1 (-0.78640, 0.66591) in Table 3 include zero, suggesting a lack of substantial evidence for any significant alteration in the appearance or location of population survival functions.

The results of the current study reveal that the risk of mortality is higher in the time around 18 months after treatment for patients who received placebo therapy. In contrast, for patients treated with thiotepa therapy, the hazard of death is higher around 19.16 months after treatment. Following these peak points, the risk of death gradually decreases until the conclusion of the study period.

In conclusion, our analysis of bladder cancer patients data indicates that the EE and ERE distributions outperform the EWE, EEE, WE, and RE distributions. Among them, the EE distribution exhibits a slightly better fit compared to the ERE distribution. Additionally, the thiotepa treatment significantly improves the proportion of patients who are considered immune. Hence, using parametric models that include a cure fraction and specify a distribution for the survival time of susceptible individuals is a suitable approach for analysing survival data involving long-term survivors. These models enable the estimation of meaningful measures, such as the fractions of immune individuals and the mean survival time, which can be easily interpreted by practitioners and healthcare professionals.

The conducted study has brought attention to topic that would benefit from further research, specifically the analysis of bladder cancer data. It could be worthwhile to explore alternative generalized distributions using Bayesian inference techniques. By incorporating an oncologist's expertise and prior knowledge about the expected fraction of cured patients into the prior distribution of the cure rate parameter, π, more precise inferences can be obtained.

Author contributions Statement

Mohamed Elamin Omer was involved in in conceptualizing and designing the study, conducting data analyses, interpreting the results, and drafting the initial version of the manuscript. Mohd Shafie Mustafa contributed to the conception and design of the study, performed the data analyses, interpreted the results, and participated in manuscript writing. Norhaslinda Ali and Nur Haizum Abd Rahman were involved in the study's conception.

Acknowledgments

Ethical conduct of research

The study's dataset (bladder) is openly available in R software 4.1.2-Package 'survival'.

Conflict of Interest

The authors have no known conflict of interest.

References

Abu Bakar MR, Salah KA, Ibrahim NA, Haron K (2008). Cure Fraction, Modelling and Estimating in a Population-Based Cancer Survival Analysis. *Malaysian J Math Sci*, **2**, 113-34.

American Cancer Society (2023). Cancer Facts and Figures. Retrieved from https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2023/2023-cancer-facts-and-figures.pdf.

Akaike H (1974). A New Look at the Statistical Model Identification. *IEEE Trans Automat Contr*, **6**, 716-23.

Berkson J, Gage RP (1952). Survival curves for cancer patients following treatment. *J Am Stat Assoc*, **47**, 501-15.

Boag JW (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J R Stat Soc Series B Stat Methdol*, **11**, 15-44.

Bradburn MJ, Clark TG, Love SB, Altman DG (2003a). Survival Analysis Part II: Multivariate data analysis - An introduction to concepts and methods. *Br J Cancer*, **89**, 431-36.

Bradburn MJ, Clark TG, Love SB, Altman DG (2003b). Survival Analysis Part III: Multivariate data analysis - Choosing a model and assessing its adequacy and fit. *Br J Cancer*, **89**, 605 -11.

Bremhorst V, Lambert P (2016). Flexible estimation in cure survival models using Bayesian P-splines. *Comput Stat Data Anal*, **93**, 270-84.

Byar DP (1980). The Veterans Administration Study of Chemoprophylaxis for Recurrent Stage I Bladder Tumours: Comparisons of Placebo, Pyridoxine, and Topical Thiotepa. In 'Bladder Tumours and Other Topics in Urological Oncology', Plenum, New York, pp 363-70.

Carrasco JMF, Ortega EMM, Cordeiro GM (2008). A generalized modified Weibull distribution for lifetime modeling. *Comput Stat Data Anal*, **53**, 450-62.

Chen M, Ibrahim JG, Sinha D (1999). A new bayesian model for survival data with a surviving fraction. *J Am Stat Assoc*, **94**, 909-19.

Corbière F, Commenges D, Taylor JMG, Joly P (2009). A penalized likelihood approach for mixture cure models. *Stat Med*, **28**, 510-24.

Cox DR (1972). Regression Models and Life-Tables. *J R Stat Soc Series B Stat Methdol*, **34**, 187-202.

Elgarhy M, Shakil M, Kibria BMG (2017). Exponentiated Weibull-Exponential distribution with applications. *AAM Int J*, **12**, 710-25.

Farewell VT (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, **38**, 1041-46.

Ghitany ME, Maller RA (1992). Asymptotic results for exponential mixture models with long-term survivors. Stat: *J Theor Appl Stat*, **23**, 321-36.

Goldman AI (1984). Survivorship analysis when cure is a possibility: A Monte Carlo study. *Stat Med*, **3**, 153-63.

Hannan EJ, Quinn BG (1079). The determination of the order of an autoregression. *J R Stat Soc Ser B Methodol*, **41**, 190-5.

Hassan AS, Elgarhy M (2016 b). A New Family of Exponentiated Weibull-Generated Distributions. *IJMAA*, **4**, 135-48.

Herring AH, Ibrahim JG (2002). Maximum likelihood estimation in random effects cure rate models with nonignorable missing Covariates. *Biosts*, **3**, 387- 405.

Ibrahim GM, Almetwally EM (2021). The New Extension of Inverse Weibull Distribution with Applications of Medicine Data. *SJFCSR*, **2**, 577-98.

International Agency for Research on Cancer. Sudan fact sheets. Source: Globocan 2020. 2020. Available online: https://gco.iarc.fr/today/data/factsheets/populations/729-sudan-fact-sheets.pdf (accessed on 11 October 2023).

Klebanov LB, Rachev ST, Yakovolev AY (1993). A stochastic model of radiation carcinogenesis: Latent time distributions and their properties. *Math Biosci*, **113**, 51-75.
Kuk AYC, Chen CH (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, **79**, 531-41.
Kutal DH, Qinan L (2018). A non-mixture cure model for right-censored data with Fréchet Distribution. *Stats*, **1**, 176-88.
Lin DY, Wei LJ, Ying Z (1998). Accelerated failure time models for counting processes. *Biometrika*, **85**, 605-18.
Liu H, Shen Y (2009). A semiparametric regression cure model for interval-censored data. *J Am Stat Assoc*, **104**, 1168-78.
López-Cheda A, Cao R, Jácome MA, Keilegom IV (2017). Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models. *Comput Stat Data Anal*, **105**, 144-65.
Lu W (2010). Efficient estimation for an accelerated failure time model with cure fraction. *Stat Sin*, **20**, 661-74.
Maller RA, Zhou S (1992). Estimating the proportion of immunes in a censored sample. *Biometrika*, **79**, 731-9.
Maller RA, Zhou X (1996). Survival Analysis with Long-Term Survivors; Wiley, New York, NY, USA, ISBN 0-471-96201-5.
Mudholkar GS, Srivastava DK (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Trans Reliab*, **42**, 299-302.
National Cancer Institute (2023). Retrieved from https://www.cancer.gov/types/bladder.
Oguntunde PE, Balogun OS, Okagbue HI, Bishop SA (2015). The Weibull-exponential distribution: Its properties and applications. *J Appl Sci*, **15**, 1305-11.
Omer ME, Abu Bakar M, Adam MB, Mustafa MS (2021). Utilization of a Mixture Cure Rate Model based on the Generalized Modified Weibull Distribution for the Analysis of Leukemia Patients. *Asian Pac J Cancer Prev*, **22**, 1044-53.
Ortega EMM, Cordeiro GM, Kattan MW (2013). The log-beta Weibull regression model with application to predict recurrence of prostate cancer. *Stat Pap*, **54**, 113-32.
Patilea V, Keilegom IV (2017). A general approach for cure models in survival analysis. *Ann Statist*, **48**, 2323-46.
Peng Y, Dear KBG, Carriere KC (2001). Testing for the presence of cured patients: a simulation study. *Stat Med*, **20**, 1783-96.
Schwarz G (1978). Estimating the dimensions of a model. *Ann Stat*, **6**, 461- 4.
Sun J, Wei LJ (2002). Regression analysis of panel count data with covariate-dependent observation and censoring times. *J R Statist Soc B*, **62**, 293-302.
Sy JP, Taylor JMG (2000). Estimation in a Cox Proportional Hazards Cure Model. *Biometrics*, **56**, 227-36.
Therneau T (2023). A Package for Survival Analysis in R. R package version 3.5.7, https://CRAN.R-project.org/package=survival.
Tsodikov A (1998). A proportional hazards model taking account of long-term survivors. *Biom*, **54**, 1508-16.
Tsodikov AD, Ibrahim JG, Yakovlev AY (2003). Estimating cure rates from survival data: An alternative to two-component mixture models. *J Am Stat Assoc*, **98**, 1063-78.
Uddin MT, Sen A, Noor MS, Islam MN, Chowdhury ZI (2006). An analytical approach on non-parametric estimation of cure rate based on uncensored data. *J Appl Sci*, **6**, 1258-64.
Wei LJ, Lin DY, Weissfeld L (1989). Regression Analysis of Multivariate Incomplete Failure Time Data by Modelling Marginal Distributions. *J Am Stat Assoc*, **84**, 1065-73.
Yakovlev AY, Asselain B, Bardou VJ, et al (1993). A simple stochastic model of tumour recurrence and its application to data on premenopausal breast cancer. In Biometrie et analyse de donne'es spatio-temporelles, 12, pp 66-83.
Yamaguchi K (1992). Accelerated failure-time regression models with a regression model of surviving fraction: An application to the analysis of 'permanent employment' in Japan. *J Am Stat Assoc*, **87**, 284-92.