

Article

A Non-Mixture Cure Model for Right-Censored Data with Fréchet Distribution

Durga H. Kotal ^{*,†} and Lianfen Qian [‡] 

Department of Mathematical Sciences, Florida Atlantic University, Boca Raton, FL 33431, USA; lqian@fau.edu

* Correspondence: kutaldh@wfu.edu; Tel.: +1-561-419-4604

† Durga H. Kotal currently is a visiting assistant professor at Wake Forest University, NC, USA.

‡ Lianfen Qian is a distinguished guest professor of Wenzhou University, Wenzhou, China.

Received: 29 September 2018; Accepted: 7 November 2018; Published: 15 November 2018



Abstract: This paper considers a non-mixture cure model for right-censored data. It utilizes the maximum likelihood method to estimate model parameters in the non-mixture cure model. The simulation study is based on Fréchet susceptible distribution to evaluate the performance of the method. Compared with Weibull and exponentiated exponential distributions, the non-mixture Fréchet distribution is shown to be the best in modeling a real data on allogeneic marrow HLA-matched donors and ECOG phase III clinical trial e1684 data.

Keywords: Non-mixture model; Fréchet distribution; Right-censored survival data; Maximum likelihood method

1. Introduction

The cure fraction models are broadly used for analyzing survival data. In the literature, there are two major models to fit survival data with a cure fraction. The first one is the mixture cure rate model, also known as the standard cure rate model. This model was initially introduced by Boag [1] in 1949 and further developed by Berkson and Gage [2] in 1952 and later extensively studied by other authors. In this model, it is assumed that a certain proportion of the population is cured and the remaining is not cured. In order to estimate the cure fraction, parametric, semi-parametric and non-parametric methods have been studied by several authors. Farewell [3] in 1982 used a Weibull distribution for uncured subjects and a logistic regression for cure probability. Goldman [4] in 1984 discussed parametric survivorship analyses using maximum likelihood estimation and the likelihood ratio test. Taylor [5] in 1995 proposed the semi-parametric method to a mixture model using the logistic regression for an incidence part and the Kaplan–Meier for the latency part. Peng and Dear [6] in 2000 investigated the non-parametric approach to a mixture model to estimate parameter of interest in the model using EM algorithm, marginal likelihood approach, and multiple imputations. They also extended to Cox’s proportional hazards cure model. Kuk and Chen [7] in 1992 also proposed the semi-parametric cure model using the logistic regression for cure probability and the proportional hazard regression model for failure time. Zhang and Peng [8] in 2009 proposed a mixture cure model where the covariate effects on the proportion of cure and the distribution of the failure time of uncured patients are separately modeled. Patilea and Van Keilegom [9] in 2017 proposed a general approach to estimate a mixture cure model for random right censoring data. They worked with a parametric model for the cure proportion and the conditional probability for the uncured subjects. Moreover, Kim and Jhun [10] in 2008 investigated interval censored data based on a mixture cure model. They derived the likelihood in interval censored data based on an approximate likelihood approach suggested by Goetghebeur and Ryan [11] in 2000.

The second one is a non-mixture cure rate model, also known as the bounded cumulative hazard model and promotion time cure model. In cancer study, this model was developed based on the assumption that the number of cancer cells that remain active after cancer treatment and that may grow slowly and produce a detectable cancer, which assumed to follow a Poisson distribution. This model was first proposed by Yakovlev et al. [12] in 1993 and was further discussed by Chen et al. [13] in 1999. The semi-parametric approaches of estimation for survival data with a cure fraction have been discussed by Chen et al. [14] in 2001. Tsodikov et al. [15] in 2003 provided a review of existing methodology of statistical inference based on the non-mixture model. They have highlighted that there are the distinct advantages of the non-mixture cure model: the non-mixture cure model has proportional hazard model structure, the non-mixture cure model presents a much more biologically meaningful interpretation of the results of the data analysis and the non-mixture cure model is easy in computations due to its simple structure for the survival function which can provide a certain technical advantage when developing maximum likelihood estimation procedures. Herring and Ibrahim [16] in 2002 studied the parametric estimation of random effects for non-ignorable missing covariates in a non-mixture cure model. Uddin et al. [17,18] in 2006 approached both non-parametric and parametric methods in a non-mixture model for uncensored data. Liu et al. [19] in 2009 investigated the semi-parametric non-mixture cure model for interval censored data using the EM method. Lopes et al. [20] in 2012 studied both Bayesian and classical approaches to long-term survivors and random effects based on a non-mixture cure model.

In this paper, we considered a generalized extreme value type II distribution for long-term survivors. The generalized extreme value distribution was introduced by Jenkinson [21] in 1955 and its sub-models are widely used for modeling extreme events. The generalized extreme value distribution sub-models are Gumbel, Fréchet and Weibull distributions. An extreme value distributions play an important role in statistics. Ramos et al. [22] in 2017, discussed some properties of the long-term Fréchet distribution. We study a long-term survival non-mixture cure model with Fréchet distribution for uncured population. The Fréchet distribution is one of the flexible distribution for survival analysis due to its properties. Its probability density function is skewed to the right and it has also increasing and decreasing hazard rate function depending on the values of a shape parameter. A simulation study is conducted to show the performance of the maximum likelihood estimators. Moreover, we compare the our proposed model with other established models using the real data-sets of leukemia for 46 allogeneic marrow transplantation patients presented by Kersey et al. [23] in 1987 and ECOG phase III clinical trial e1684 available in R package *smcure*.

The rest of the article is organized as follows. In Sections 2 and 3 we derive the likelihood function of a mixture cure model and a non-mixture cure model respectively. In Section 4 we introduce the Fréchet distribution and we derive the likelihood function of the non-mixture Fréchet distribution. In Section 5 we conduct the simulation study of the non-mixture Fréchet distribution. In Section 6 we present the leukemia dataset and melanoma dataset to compare the mixture cure model and the non-mixture cure model with different susceptible distributions and finally the conclusion is presented in Section 7.

2. Likelihood Function of a Mixture Cure Model

A mixture cure model refers to a class of model for survival data with long-term survivors when some of them will not develop the event of interest. In a mixture cure model, the population consists of two types of subjects: uncured (susceptible) subjects who experience the event of interest and cured (non-susceptible) subjects who will never experience it. Those who are not going to develop the event of interest are referred to as “cured subjects”, or “long-term survivors”.

Let T denote the survival time of a subject and let Δ be the cure indicator with $\Delta = 0$ when the subject is cured and $\Delta = 1$ when the subject is uncured. Let p be the proportion of subjects that are cured, and $1 - p$ be the proportion of subjects that are not cured. That is, $P(\Delta = 1) = 1 - p$ and

$P(\Delta = 0) = p$. So, $P(T \leq t | \Delta = 0) = 0$ and $P(T \leq t | \Delta = 1) = F(t)$, the cumulative distribution function of uncured subjects. Then the cumulative distribution function of the overall population T is

$$F_T(t) = P(T \leq t) = P(T \leq t | \Delta = 0)P(\Delta = 0) + P(T \leq t | \Delta = 1)P(\Delta = 1) \\ = (1 - p)F(t)$$

and the survival function of T is

$$S_T(t) = 1 - F_T(t) = 1 - (1 - p)F(t) = p + (1 - p)S(t), \tag{1}$$

where $S(t)$ denotes a proper survival function for the uncured patients. Notice that $S_T(t)$ is not proper. The corresponding density function of T is

$$f_T(t) = (1 - p)f(t), \tag{2}$$

where $f(t)$ is the probability density function of uncured patients.

Suppose y_i be the right-censored survival time for subject i , then $y_i = \min(T_i, C_i)$, where T_i be the failure time of the i th subject and C_i be the right-censored variable of the i th subject, $i = 1, \dots, n$. The observed survival time of i th subject is y_i and the censoring indicator is δ_i , where $\delta_i = I(T_i \leq C_i)$, for $i = 1, \dots, n$. Then the likelihood function of the mixture model is

$$L_m = \prod_{i=1}^n f_T^{\delta_i}(y_i) S_T^{1-\delta_i}(y_i) = \prod_{i=1}^n [(1 - p)f(y_i)]^{\delta_i} [p + (1 - p)S(y_i)]^{1-\delta_i}$$

and the log-likelihood function is

$$l_m \equiv \ln(L_m) = \ln(1 - p) \sum_{i=1}^n \delta_i + \sum_{i=1}^n \delta_i \ln(f(y_i)) + \sum_{i=1}^n (1 - \delta_i) \ln(p + (1 - p)S(y_i)). \tag{3}$$

3. Likelihood Function of a Non-Mixture Cure Model

In this section, we introduce a non-mixture cure model developed for survival data with long-term survivors in cancer study. Let N be the number of cancer cells for a patient after cancer treatment. Since the number of cancer cells may grow rapidly and may later produce a detectable cancer disease, the number of cancer cells, N , is assumed to have a Poisson distribution with mean λ . Let Z_j be the random time for the j th cancer cell to produce a detectable cancer mass. Then, the time to relapse of cancer can be defined by the random variable T such that $T = \min\{Z_j, j = 1, 2, \dots, N\}$. Furthermore, we assume that Z_j are independently and identically distributed with distribution and survival functions $F(\cdot)$ and $S(\cdot)$, respectively. Then the survival function of T is given by:

$$S_T(t) = P[\text{No cancer by time } t] \\ = P[N = 0] + P[Z_1 > t, Z_2 > t, \dots, Z_N > t, N \geq 1] \\ = e^{-\lambda} + \sum_{N=1}^{\infty} S^N(t) \frac{\lambda^N}{N!} e^{-\lambda} = e^{-\lambda + \lambda S(t)} = e^{-\lambda F(t)} = p^{F(t)},$$

where $p = e^{-\lambda}$ is the probability of cure or cure fraction in this model since

$$\lim_{t \rightarrow \infty} S_T(t) = p = e^{-\lambda} > 0,$$

which implies that $S_T(\cdot)$ is an improper survival function. The cumulative distribution, density and hazard functions of T are given below respectively,

$$F_T(t) = 1 - p^{F(t)}, \quad f_T(t) = -\ln(p)f(t)p^{F(t)}, \quad h_T(t) = -\ln(p)f(t).$$

Since $S_T(t)$ is not proper survival function, then we can derive the following theorem:

Theorem 1. Let $0 \leq p \leq 1$. For $T \sim F_T(\cdot) = 1 - p^{F(\cdot)}$ and $F(\cdot)$ is a strictly increasing proper distribution function, then $Y = F_T(T)$ has an improper uniform distribution over $[0, 1 - p]$.

Proof of Theorem 1. Notice that as $t \rightarrow \infty$, $p^{F(t)} \rightarrow p$. So,

$$\lim_{t \rightarrow \infty} F_T(t) = \lim_{t \rightarrow \infty} (1 - p^{F(t)}) = 1 - p.$$

Since $F(\cdot)$ is strictly increasing which implies that $F_T(\cdot)$ is also strictly increasing, it has an inverse, denoted by $F_T^{-1}(\cdot)$. Then for $0 \leq y \leq 1 - p$,

$$F_Y(y) = P(Y \leq y) = P(F_T(T) \leq y) = P(T \leq F_T^{-1}(y)) = F_T(F_T^{-1}(y)) = y.$$

Therefore, $Y = F_T(T)$ has an improper uniform distribution over $[0, 1 - p]$.

Case-1: If $p = 0$, this theorem is equivalent to the Inverse Transform Method.

Case-2: If $p = 1$, the result is not realistic for real life. \square

We now consider the right-censored life time data and assume there are n patients under cancer study. Let y_i refer to the right-censored survival time for individual i and $T_i = \min\{Z_{ij}, j = 1, 2, \dots, N_i\}$ where C_i be the right-censored variables. Then we observe $y_i = \min(T_i, C_i)$ and the censoring indicator $\delta_i = I(T_i \leq C_i)$. The likelihood function of (y_1, \dots, y_n) is

$$L = \prod_{i=1}^n h_T^{\delta_i}(y_i) S_T(y_i) = \prod_{i=1}^n [-\ln(p) f(y_i)]^{\delta_i} p^{F(y_i)}$$

and the log-likelihood function is

$$\ln(L) = \sum_{i=1}^n \delta_i \ln(-\ln(p)) + \sum_{i=1}^n \delta_i \ln f(y_i) + \sum_{i=1}^n \ln(p) F(y_i). \tag{4}$$

4. Fréchet Susceptible Distribution

We used a generalized extreme value distribution as the susceptible distribution, i.e., the type II Fréchet distribution. The class of extreme value distributions essentially consists of three types of extreme value distributions: type I (Gumbel distribution), type II (Fréchet distribution) and type III (Weibull distribution). The extreme value distribution is useful in modeling and measuring the events which occur with very low probability. Moreover, extreme value distributions are widely used in finance, insurance, risk management, economics, material sciences and many other subjects which dealing with extreme events.

More specifically, we considered the Fréchet distribution, which was named after the French mathematician Maurice Fréchet (1878–1973), to fit the survival time of uncured individuals. The Fréchet distribution is also known as inverse Weibull distribution under some parametric constraints. The probability density function of Fréchet distribution is

$$f(y) = \frac{\alpha}{\beta} \left(\frac{\beta}{y}\right)^{\alpha+1} e^{-\left(\frac{\beta}{y}\right)^\alpha}, \quad y \geq 0$$

where $\alpha > 0$ is the shape parameter and $\beta > 0$ is the scale parameter. The survival and hazard functions are respectively

$$S(y) = 1 - e^{-\left(\frac{\beta}{y}\right)^\alpha} \quad \text{and} \quad h(y) = \frac{\alpha}{\beta} \left(\frac{\beta}{y}\right)^{\alpha+1} e^{-\left(\frac{\beta}{y}\right)^\alpha} \left[1 - e^{-\left(\frac{\beta}{y}\right)^\alpha}\right]^{-1}, \quad y \geq 0.$$

Denote $\theta = (\alpha, \beta, p)'$. Substituting the above functions into (4) of the non-mixture model with Fréchet susceptible distribution, we obtain the log-likelihood:

$$l(\theta) = \ln L(\theta) = \sum_{i=1}^n \delta_i \ln[-\ln(p)] + \sum_{i=1}^n \ln(p) e^{-\left(\frac{\beta}{y_i}\right)^\alpha} + \sum_{i=1}^n \delta_i \left[\ln(\alpha) + \alpha \ln(\beta) - (\alpha + 1) \ln(y_i) - \left(\frac{\beta}{y_i}\right)^\alpha \right]. \tag{5}$$

The maximum likelihood estimator (MLE) of $\theta = (\alpha, \beta, p)'$ is one that maximizes $l(\theta)$, denoted by $\hat{\theta}$. The score functions are given below by taking partial derivatives of log-likelihood function (5) with respect to θ :

$$\begin{aligned} \frac{\partial l(\theta)}{\partial \alpha} &= \sum_{i=1}^n \delta_i \left[\frac{1}{\alpha} + \ln(\beta) - \ln(y_i) - \left(\frac{\beta}{y_i}\right)^\alpha \ln\left(\frac{\beta}{y_i}\right) \right] - \ln(p) \sum_{i=1}^n e^{-\left(\frac{\beta}{y_i}\right)^\alpha} \left(\frac{\beta}{y_i}\right)^\alpha \ln\left(\frac{\beta}{y_i}\right) \\ \frac{\partial l(\theta)}{\partial \beta} &= \frac{\alpha}{\beta} \left\{ \sum_{i=1}^n \delta_i \left[1 - \left(\frac{\beta}{y_i}\right)^\alpha \right] - \ln(p) \sum_{i=1}^n e^{-\left(\frac{\beta}{y_i}\right)^\alpha} \left(\frac{\beta}{y_i}\right)^\alpha \right\} \\ \frac{\partial l(\theta)}{\partial p} &= \frac{1}{p \ln(p)} \sum_{i=1}^n \delta_i + \frac{1}{p} \sum_{i=1}^n e^{-\left(\frac{\beta}{y_i}\right)^\alpha} \end{aligned}$$

If the log-likelihood function has a global maximizer, then the MLE is the solution of the score equations. Since it is a non-linear system, numerical solution is computed by the Newton-Raphson method.

5. Simulation Study

We conducted simulation studies to examine the performance of the maximum likelihood estimator of $\theta = (\alpha, \beta, p)'$ in finite samples. The right-censored survival times were generated by using the inverse transform method to the survival function $S_T(t) = p^{F(t)}$. We used the following algorithm to simulate a sample of size n from the non-mixture Fréchet distribution with right-censored data:

- Step 1: Generate a simple sample of u_1, \dots, u_n from $\sim U(0, 1)$.
- Step 2: Suppose p is a cure fraction. The random survival time can be calculated from the equation

$$t_i = \beta \left[-\ln \left(\frac{\ln(1 - u_i)}{\ln(p)} \right) \right]^{-\frac{1}{\alpha}},$$

if $u_i \leq 1 - p$, otherwise t_i is infinity.

- Step 3: Generate the simple sample of the censoring times c_1, \dots, c_n from a Fréchet distribution and adjust the parameters of the Fréchet distribution to obtain the desired censoring rates.
- Step 4: The right-censored data is obtained from the minimum of censoring time and survival time. That is $y_i = \min(t_i, c_i)$, $\delta_i = I(t_i \leq c_i)$, $i = 1, 2, \dots, n$.
- Step 5: The observed data-set is $D = \{(y_i, \delta_i), i = 1, \dots, n\}$.
- Step 6: Maximize the likelihood function with respect to θ to obtain $\hat{\theta}$. The standard optimization method, `optim()` in R is used.

We consider various simulation settings with different proportion of cure fractions and different censoring rates. The censoring variable follows a Fréchet distribution with parameter $\alpha > 0$ and $\beta > 0$, where the value of α and β would be adjusted to get the desired censoring rate in the right-censored survival data.

In this simulation study, we are interested in the bias, standard error, root mean square error, and 95% confidence interval as the performance measures. The bias is the difference between the expected value of an estimator and true parameter as given by $\text{bias} = E(\hat{\theta}) - \theta$ and the standard error is a measure of the dispersion of the values in the sampling distribution, which is a statistical term

that measures the accuracy and is defined by the standard deviation divided by the square root of sample size. The mean square error (MSE) of an estimator is the expected squared deviation of the estimator of a parameter from the true parameter. The root mean square error (RMSE) is the squared root of MSE, where $MSE(\hat{\theta}) = (bias)^2 + var(\hat{\theta})$. The confidence intervals tell you about the parameter of interest and not about the values of the estimated parameter.

The simulation results are based on 500 replications with the sample sizes 100, 200, 300, and 500 for each parameter setting. Results are presented in Tables 1–4, which show the values of mean, bias, standard error (SE), root mean square error (RMSE), and 95% confidence interval of MLE. The simulation results suggest that the proposed method has a good performance overall. The averages of the estimates are very close to respective true parameter values in all different settings of simulations. However, the simulation results for cure fractions of both 1% and 2% with lower censoring rates perform better than the higher censoring rates. The biases of estimates are small, and the estimates of the standard error (SE), and root mean square error (RMSE) of all examined parameters decreased with increasing sample size for all settings. The length of the confidence interval of each parameter also decreased with increasing sample size for all different setting. The estimates of all examined parameters perform better for low levels of censoring than high levels of censoring. Moreover, the estimates of all examined parameters performed well for low rates of cure proportion compared to high rates of cure proportion. Finally, we can conclude that the MLE works well for the non-mixture cure model with Fréchet susceptible distribution.

The overall censoring rate of Table 1 is about 7.78%, while Table 2 is about 10.34%. Moreover, θ_0 is the true parameter vector, CR the censoring rate, SE the standard error, RMSE a root mean square error and CI the confidence interval.

Table 1. Summary statistics of a non-mixture model with 1% cure fraction for right-censored data.

	θ_0	Mean	Bias	SE	RMSE	95% CI
<i>n</i> = 100 and CR 7.71%						
α	0.50	0.5100	0.0100	0.0880	0.0886	(0.338, 0.683)
β	1.00	1.4130	0.4130	1.4900	1.5462	(−1.491, 4.126)
<i>p</i>	0.01	0.0146	0.0046	0.0151	0.0158	(−0.015, 0.044)
<i>n</i> = 200 and CR 7.68%						
α	0.50	0.5023	0.0023	0.0620	0.0621	(0.381, 0.624)
β	1.00	1.1822	0.1822	0.6277	0.6536	(−0.048, 2.412)
<i>p</i>	0.01	0.0119	0.0019	0.0104	0.0106	(−0.008, 0.032)
<i>n</i> = 300 and CR 7.81%						
α	0.50	0.5019	0.0019	0.0499	0.0499	(0.404, 0.599)
β	1.00	1.1022	0.1022	0.4343	0.4462	(0.251, 1.953)
<i>p</i>	0.01	0.0117	0.0017	0.0083	0.0085	(−0.004, 0.028)
<i>n</i> = 500 and CR 7.82%						
α	0.50	0.5003	0.0003	0.0371	0.0371	(0.428, 0.573)
β	1.00	1.0588	0.0588	0.2879	0.2938	(0.495, 1.623)
<i>p</i>	0.01	0.0108	0.0008	0.0061	0.0062	(−0.001, 0.023)

Table 2. Summary statistics of non-mixture model with 2% cure fraction for right-censored data.

	θ_0	Mean	Bias	SE	RMSE	95% CI
<i>n</i> = 100 and CR 10.27%						
α	0.50	0.5110	0.0110	0.0895	0.0902	(0.336, 0.686)
β	1.00	1.3538	0.3538	1.1763	1.2284	(−0.952, 3.659)
<i>p</i>	0.02	0.0253	0.0053	0.0225	0.0231	(−0.019, 0.069)
<i>n</i> = 200 and CR 10.24%						
α	0.50	0.5019	0.0019	0.0637	0.0637	(0.377, 0.627)
β	1.00	1.1954	0.1954	0.6277	0.6574	(−0.035, 2.426)
<i>p</i>	0.02	0.0218	0.0018	0.0160	0.0161	(−0.009, 0.053)

Table 2. Cont.

	θ_0	Mean	Bias	SE	RMSE	95% CI
<i>n</i> = 300 and CR 10.39%						
α	0.50	0.5019	0.0019	0.0511	0.0511	(0.401, 0.602)
β	1.00	1.1035	0.1035	0.4269	0.4393	(0.267, 1.940)
p	0.02	0.0218	0.0018	0.0133	0.0134	(−0.004, 0.048)
<i>n</i> = 500 and CR 10.37%						
α	0.50	0.4992	−0.0008	0.0379	0.0379	(0.425, 0.573)
β	1.00	1.0694	0.0694	0.2898	0.2980	(0.502, 1.637)
p	0.02	0.0204	0.0004	0.0097	0.0097	(0.001, 0.039)

The overall censoring rates of Table 3 is about 14.36%, and Table 4 is about 18.55%. Moreover, θ_0 is the true parameter vector, CR the censoring rate, SE the standard error, RMSE a root mean square error and CI the confidence interval.

Table 3. Summary statistics of non-mixture model with 1% cure fraction for right-censored data.

	θ_0	Mean	Bias	SE	RMSE	95% CI
<i>n</i> = 100 and CR 14.26%						
α	0.50	0.5119	0.0119	0.0955	0.0962	(0.325, 0.694)
β	1.00	1.5112	0.5112	1.7149	1.7895	(−1.850, 4.872)
p	0.01	0.0163	0.0063	0.0181	0.0192	(−0.019, 0.052)
<i>n</i> = 200 and CR 14.34%						
α	0.50	0.5034	0.0034	0.0679	0.0680	(0.370, 0.636)
β	1.00	1.2270	0.2270	0.7577	0.7909	(−0.258, 2.712)
p	0.01	0.0130	0.0030	0.0123	0.0127	(−0.011, 0.037)
<i>n</i> = 300 and CR 14.41%						
α	0.50	0.5034	0.0034	0.0555	0.0556	(0.395, 0.612)
β	1.00	1.1274	0.1274	0.5458	0.5605	(0.058, 2.197)
p	0.01	0.0127	0.0027	0.0103	0.0106	(−0.007, 0.032)
<i>n</i> = 500 and CR 14.42%						
α	0.50	0.5001	0.0001	0.0399	0.0399	(0.422, 0.578)
β	1.00	1.0726	0.0726	0.3189	0.3271	(0.448, 1.697)
p	0.01	0.0111	0.0011	0.0072	0.0073	(−0.003, 0.025)

Table 4. Summary statistics of non-mixture model with 2% cure fraction for right-censored data.

	θ_0	Mean	Bias	SE	RMSE	95% CI
<i>n</i> = 100 and CR 18.49%						
α	0.50	0.5162	0.0162	0.1001	0.1014	(0.320, 0.712)
β	1.00	1.4419	0.4419	1.5639	1.6251	(−1.623, 4.507)
p	0.02	0.0285	0.0085	0.0276	0.0289	(−0.026, 0.083)
<i>n</i> = 200 and CR 18.55%						
α	0.50	0.5040	0.0040	0.0702	0.0703	(0.366, 0.692)
β	1.00	1.2361	0.2361	0.8332	0.8660	(−0.396, 2.869)
p	0.02	0.0237	0.0037	0.0193	0.0197	(−0.014, 0.062)
<i>n</i> = 300 and CR 18.63%						
α	0.50	0.5035	0.0035	0.0588	0.0589	(0.388, 0.618)
β	1.00	1.1442	0.1442	0.5999	0.6170	(−0.032, 2.320)
p	0.02	0.0236	0.0036	0.0171	0.0175	(−0.009, 0.057)
<i>n</i> = 500 and CR 18.53%						
α	0.50	0.4981	−0.0019	0.0409	0.0409	(0.418, 0.578)
β	1.00	1.0961	0.0961	0.3576	0.3703	(0.395, 1.796)
p	0.02	0.0206	0.0006	0.0115	0.0115	(−0.001, 0.043)

Figure 1 shows the relationship between the estimated bias against the censoring rate.

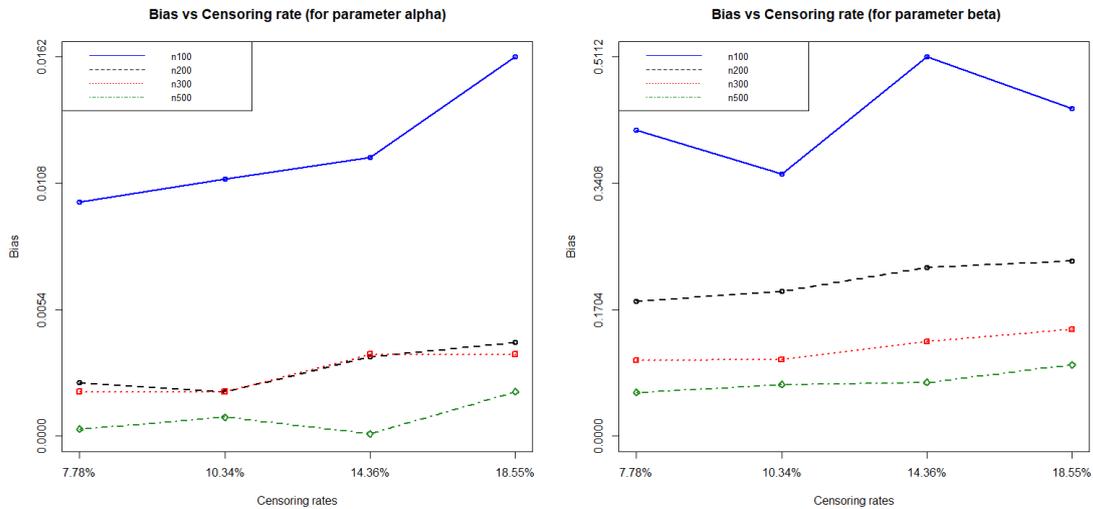


Figure 1. The relationship between the bias and the censoring rate.

6. A Real Data Analysis

In this section, we would like to compare non-mixture and mixture models with different distributions. We use different susceptible assessments: the negative value of the log-likelihood function, the Akaike’s information Criterion (*AIC*), the corrected Akaike’s information Criterion (*AIC_c*), and the Bayesian Information Criterion (*BIC*). The (*AIC*) is defined by $AIC = -2\ln L(\theta) + 2k$, where $\theta = (\alpha, \beta, p)'$ is the vector of unknown parameters included in the model, $L(\theta)$ is the likelihood function and where k is the number of free parameters in the model. The (*AIC_c*) is defined by

$$AIC_c = -2\ln L(\theta) + 2k \frac{n}{n - k - 1},$$

where n is the sample size. The *BIC* is defined by $BIC = k\ln(n) - 2L(\theta)$. A lower value of selection criterion indicates a better fit.

We consider the data-set of 46 patients with an HLA-matched donor-received allogeneic marrow presented by Kersey [23] in 1987. Out of the 46 patients, 12 with allogeneic transplantation died during their observed periods. Those are the censored observations (26%). Using this data, we compare the non-mixture and mixture models with Fréchet, two parameters Weibull and two parameters Exponentiated Exponential (EE) susceptible distributions using the maximum likelihood method. The density function of Weibull distribution is defined by

$$f(y) = \frac{\alpha}{\beta} \left(\frac{y}{\beta}\right)^{\alpha-1} e^{-\left(\frac{y}{\beta}\right)^\alpha}, \quad y \geq 0$$

and the density function of Exponentiated Exponential distribution is defined by

$$f(y) = \alpha\beta(1 - e^{-\beta y})^{(\alpha-1)} e^{-\beta y}, \quad y \geq 0.$$

The notations α , β , and p are same as to the Fréchet distribution.

Tables 5 and 6 show the maximum likelihood estimates of the parameters and their standard errors of the above three susceptible distributions with both non-mixture and mixture models for allogeneic marrow transplanted (leukemia) data, respectively. Table 7 shows the values of $-\log L$, *AIC*, *AIC_c* and *BIC* of three susceptible distributions using both mixture and non-mixture models for

leukemia data. Table 7 indicates that Fréchet Distribution is the best within both non-mixture and mixture cure models.

Table 5. Results of non-mixture cure models with leukemia data.

Parameter	Fréchet		Weibull		EE	
	Estimate	SE	Estimate	SE	Estimate	SE
α	0.430	0.134	1.005	0.149	1.045	0.220
β	1.526	1.781	1.221	0.356	0.863	0.318
p	0.068	0.092	0.233	0.069	0.234	0.069

Table 6. Results of mixture cure models with leukemia data.

Parameter	Fréchet		Weibull		EE	
	Estimate	SE	Estimate	SE	Estimate	SE
α	0.608	0.138	0.902	0.132	0.910	0.211
β	0.363	0.162	0.769	0.171	1.184	0.329
p	0.054	0.150	0.239	0.068	0.242	0.067

Table 7. Model comparison by information criteria for leukemia data.

Model	Non-Mixture Model				Mixture Model			
	$-\log L$	AIC	AIC_c	BIC	$-\log L$	AIC	AIC_c	BIC
Fréchet	47.366	100.73	101.30	99.72	47.371	100.75	101.32	99.73
Weibull	48.340	102.68	103.25	101.67	48.752	103.50	104.70	102.49
EE	48.319	102.64	103.21	101.63	48.937	103.87	104.45	102.86

We also compared the Kaplan–Meier survival curves with the parametric curves. The models offer a better fit when they are closer to each other. The top left corner figure of Figure 2 is the Kaplan–Meier survival curve of the allogenic marrow transplanted data. The top right corner of Figure 2 is the Kaplan–Meier estimated curve overlaid with the estimated survival curves using Fréchet non-mixture and mixture models. The bottom left corner of Figure 2 is the Kaplan–Meier estimated curve overlaid with the estimated survival curves using Weibull non-mixture and mixture models. The bottom right corner of Figure 2 is the Kaplan–Meier estimated curve with survival curve of non-mixture and mixture models of Exponentiated Exponential Distribution. The curve of non-mixture Fréchet distribution is closer to the Kaplan–Meier survival curve in comparison to all other distributions, and the mixture model is a little bit over estimating the survival rate.

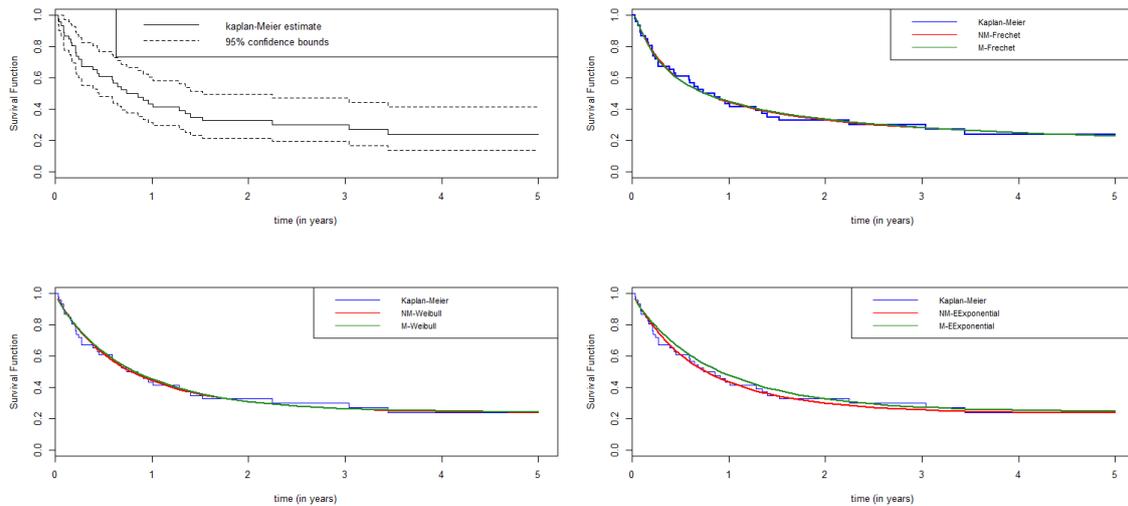


Figure 2. Fitted survival curves of the leukemia (Allogenic) data.

The second data-set we considered is the melanoma data without covariates from the Eastern Cooperative Oncology Group (ECOG) phase III clinical trial e1684 available in R package *smcure*. In this trial, we found that a total of 69% patients are censored. This trial, e1684, was a two-arm clinical trial comparing high-dose interferon alpha-2b (IFN) regimen to observation. There were a total of 286 patients enrolled in the study, collected from 1984 to 1990, and the study was published in 1996. After deleting the missing data, we used a total of 284 observations without covariates for analyzing our models.

The statistical summaries under Fréchet susceptible distribution, two parameters Weibull, and Exponentiated Exponential susceptible distributions are presented in Tables 8 and 9. Table 10 shows the values of $-\log L$, AIC , AIC_c and BIC of different distributions with both mixture and non-mixture models for melanoma data. From the result of Table 10, we conclude that the Fréchet distribution non-mixture and mixture cure models are better fit, since $-\log L$, AIC , AIC_c and BIC are smaller when compared to the other models.

Table 8. Results of a non-mixture cure model for melanoma data.

Parameter	Fréchet		Weibull		EE	
	Estimate	SE	Estimate	SE	Estimate	SE
α	0.574	0.059	1.014	0.058	1.089	0.096
β	0.963	0.240	1.566	0.153	0.695	0.089
p	0.191	0.040	0.293	0.028	0.293	0.028

Table 9. Results of a mixture cure model for melanoma data.

Parameter	Fréchet		Weibull		EE	
	Estimate	SE	Estimate	SE	Estimate	SE
α	0.732	0.061	0.913	0.053	0.947	0.091
β	0.440	0.058	1.076	0.094	0.866	0.092
p	0.201	0.041	0.299	0.028	0.301	0.028

Table 10. Models comparison by information criteria for melanoma data.

Model	Non-Mixture Model				Mixture Model			
	$-\log L$	AIC	AIC_c	BIC	$-\log L$	AIC	AIC_c	BIC
Fréchet	367.546	741.09	741.18	742.45	367.356	740.72	740.80	742.07
Weibull	378.405	762.81	762.90	764.17	381.396	768.79	768.88	770.15
EE	377.970	761.94	762.03	763.30	382.549	771.10	771.18	772.46

The Kaplan–Meier estimate of survival curve and fitted survival curves of the Fréchet, Weibull and Exponentiated Exponential distributions for the melanoma data are given in Figure 3. The top left corner figure of Figure 3 is the Kaplan–Meier survival curve of the melanoma data. The top right corner of Figure 3 is the Kaplan–Meier estimated curve overlaid with the estimated survival curves using Fréchet non-mixture and mixture models. The bottom left corner of Figure 3 is the Kaplan–Meier estimated curve overlaid with the estimated survival curves using Weibull non-mixture and mixture models. The bottom right corner of Figure 3 is the Kaplan–Meier estimated curve with survival curve of non-mixture and mixture model of Exponentiated Exponential distribution. The curve of non-mixture Fréchet distribution is closer to the Kaplan–Meier survival curve in comparison to all other distributions, and the mixture model is a little bit over estimating the survival rate.

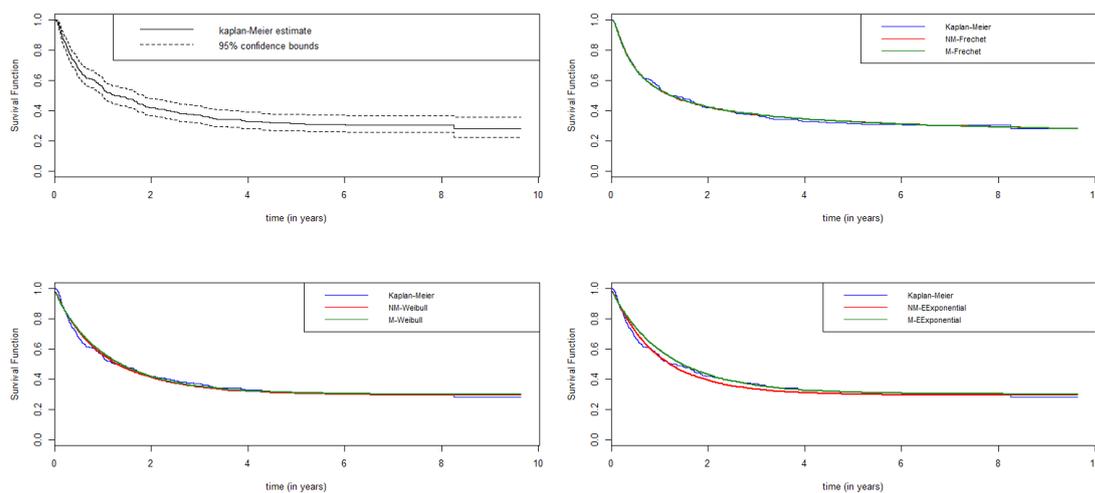


Figure 3. Fitted survival curves of the melanoma data.

7. Conclusions

In this article, we proposed a non-mixture cure model with Fréchet susceptible distribution for right-censored data. Finite sample properties of the maximum likelihood estimator of the model parameters are empirically illustrated by a simulation study under various parameter settings. The results of the simulation studies show that the proposed estimator has a good performance and is not sensitive to model parameters, but is a little bit sensitive to the censoring rate. Moreover, we compared the different susceptible distributions with both non-mixture and mixture models using real data-sets. We found that the proposed non-mixture Fréchet distribution model is the best one in comparison to other models for modeling real data regarding allogeneic marrow HLA-matched donors and the ECOG phase III clinical trial, e1684.

Author Contributions: L.Q. provided the theoretical derivation and D.H.K. did the simulation under L.Q. mentorship. Both authors contributed equally to the final version of the manuscript.

Funding: This research is partially supported by the Natural Science Foundations of Zhejiang Province (LY17A010012) and the OURI Curriculum Grant from Florida Atlantic University, USA.

Acknowledgments: The authors are grateful to the two referees for their insightful comments and suggestions that have led to significant improvements of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Boag, J.W. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J. R. Stat. Soc. Ser. B* **1949**, *11*, 15–53.
2. Berkson, J.; Gage, R.P. Survival curve for cancer patients following treatment. *J. Am. Stat. Assoc.* **1952**, *47*, 501–512. [[CrossRef](#)]
3. Farewell, V.T. The Use of Mixture Models for the Analysis of Survival Data with Long-Term survivors. *J. Int. Biom. Soc.* **1982**, *38*, 1041–1046. [[CrossRef](#)]
4. Goldman, A.I. Survivorship analysis when cure is a possibility: A monte carlo study. *J. Stat. Med.* **1984**, *3*, 153–163. [[CrossRef](#)]
5. Taylor, J.M. Semi-Parametric Estimation in Failure Time Mixture Models. *J. Int. Biom. Soc.* **1995**, *51*, 899–907. [[CrossRef](#)]
6. Peng, Y.; Dear, K.B. A Non-parametric Mixture model for Cure Rate Estimation. *J. Int. Biom. Soc.* **2000**, *56*, 237–243.
7. Kuk, A.Y.C.; Chen, C.H. A Mixture Model Combining Logistic Regression with Proportional Hazards Regression. *Biometrika* **1992**, *79*, 531–541. [[CrossRef](#)]
8. Peng, Y.; Zhag, J. Accelerated Hazards Mixture Cure Model. *Lifetime Data Anal.* **2009**, *15*, 455–467.
9. Patilea, V.; Keilegom, I.V. A General Approach for Cure Models in survival Analysis. *arXiv* **2017**, arXiv:1701.03769.
10. Kim, Y.; Jhun, M. Cure rate model with interval censored data. *Lifetime Data Anal.* **2008**, *27*, 3–14. [[CrossRef](#)] [[PubMed](#)]
11. Goetghebeur, E.; Ryan, L. Semiparametric Regression Analysis of Interval-Censored Data. *Biometrics* **2000**, *56*, 1139–1144. [[CrossRef](#)] [[PubMed](#)]
12. Klebanov, L.B.; Rachev, S.T.; Yakovlev, A.Y. A Stochastic model of radiation carcinogenesis: latent time distributions and their properties. *J. Math. Biosci.* **1993**, *113*, 51–75. [[CrossRef](#)]
13. Chen, M.H.; Ibrahim, J.G.; Sinha, D. A new bayesian model for survival data with a survival function. *J. Am. Stat. Assoc.* **1999**, *94*, 909–919. [[CrossRef](#)]
14. Ibrahim, J.G.; Chen, M.H.; Sinha, D. Bayessian Semiparametric Models for survival data with a Cure Fraction. *J. Int. Biom. Soc.* **2001**, *57*, 383–388.
15. Tsodikov, A.D.; Ibrahim, J.G.; Yakovlev, A.Y. Estimating Cure Rate from Survival Data: An Alternative to Two-component Mixture Models. *J. Am. Stat. Assoc.* **2003**, *98*, 1063–1078. [[CrossRef](#)] [[PubMed](#)]
16. Herring, A.H.; Ibrahim, J.G. Maximum likelihood estimation in random effects cure rate models with nonignorable missing covariates. *Biostatistics* **2002**, *3*, 387–405. [[CrossRef](#)] [[PubMed](#)]
17. Uddin, M.T.; Sen, A.; Noor, M.S.; Islam, M.N.; Chowdhury, Z.I. An Analytical Approach on Non-parametric Estimation of Cure Rate Based on Uncensored Data. *J. Appl. Sci.* **2006**, *6*, 1258–1264.
18. Uddin, M.T.; Islam, M.N.; Ibrahim, Q.I.U. An Analytical Approach on Cure Rate Estimation Based on Uncensored Data. *J. Appl. Sci.* **2006**, *6*, 548–552.
19. Liu, H.; Shen, Y. A Semiparametric Regression Cure Model for Interval-Censored Data. *J. Am. Stat. Assoc.* **2009**, *104*, 1168–1178. [[CrossRef](#)] [[PubMed](#)]
20. Lopes, C.M.C.; Bolfarine, H. Random effects in promotion time cure models. *Comput. Stat. Data Anal.* **2012**, *56*, 75–87. [[CrossRef](#)]
21. Jenkinson, A.F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Q. J. R. Meteorol. Soc.* **1955**, *81*, 158–171. [[CrossRef](#)]

22. Ramos, P.L.; Nascimento, D.; Louzada, F. The long-term Fréchet Distribution: Estimation, Properties and Its Application. *Biom. Biostat. Int. J.* **2017**, *6*, 00170. [[CrossRef](#)]
23. Kersy, J.H.; Weisdorf, D.; Nesbit, M.E.; LeBien, T.W.; Wood, W.G. Comparison of autologous and allogenic bone marrow transplantation for treatment of high-risk refractory acute lymphoblastic leukemia. *N. Engl. J. Med.* **1987**, *317*, 461–467. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).