

## Retraction Notice

Title of retracted article: **Statistical Analysis of a Competing Risks Model with Weibull Sub-Distributions**

Author(s): Ammar M. Sarhan, Awad I. El- Gohary, Ahlam H. Tolba

\* Corresponding author. Email: a.hamdy6@yahoo.com

Journal: Applied Mathematics (AM)  
 Year: 2017  
 Volume: 8  
 Number: 11  
 Pages (from - to): 1671 - 1683  
 DOI (to PDF): <http://dx.doi.org/10.4236/am.2017.811121>  
 Paper ID at SCIRP: 7403757  
 Article page: <http://www.scirp.org/journal/PaperInformation.aspx?PaperID=80695>  
 Retraction date: 2018-06-25

### Retraction initiative (multiple responses allowed; mark with X):

- All authors  
 Some of the authors:  
 Editor with hints from  Journal owner (publisher)  
 Institution:  
 Reader:  
 Other:

Date initiative is launched: yyyy-mm-dd

### Retraction type (multiple responses allowed):

- Unreliable findings  
 Lab error  Inconsistent data  Analytical error  Biased interpretation  
 Other:  
 Irreproducible results  
 Failure to disclose a major competing interest likely to influence interpretations or recommendations  
 Unethical research  
 Fraud  
 Data fabrication  Fake publication  Other:  
 Plagiarism  Self plagiarism  Overlap  Redundant publication \*  
 Copyright infringement  Other legal concern:  
 Editorial reasons  
 Handling error  Unreliable review(s)  Decision error  Other:  
 Other:

### Results of publication (only one response allowed):

- are still valid.  
 were found to be overall invalid.

### Author's conduct (only one response allowed):

- honest error  
 academic misconduct  
 none (not applicable in this case – e.g. in case of editorial reasons)

\* Also called duplicate or repetitive publication. Definition: "Publishing or attempting to publish substantially the same work more than once."

**History**

Expression of Concern:

- yes, date: yyyy-mm-dd
- no

Correction:

- yes, date: yyyy-mm-dd
- no

**Comment:**

The following article has been retracted due to the authors' strong personal reason about the indexing databases of Applied Mathematics.

This article has been retracted to straighten the academic record. In making this decision the Editorial Board follows [COPE's Retraction Guidelines](#). Aim is to promote the circulation of scientific research by offering an ideal research publication platform with due consideration of internationally accepted standards on publication ethics. The Editorial Board would like to extend its sincere apologies for any inconvenience this retraction may have caused.

Editor guiding this retraction: Prof. Chris Cannings (EIC of AM).

# Statistical Analysis of a Competing Risks Model with Weibull Sub-Distributions

Ammar M. Sarhan<sup>1</sup>, Awad I. El-Gohary<sup>2</sup>, Ahlam H. Tolba<sup>2</sup>

<sup>1</sup>Department of Mathematics and Statistics, Faculty of Science, Dalhousie University, Halifax, Canada

<sup>2</sup>Department of Mathematics, Faculty of Science, Mansoura University, Mansoura, Egypt

Email: ammar.sarhan@dal.ca, elgohary0@yahoo.com, a.hamdy6@yahoo.com

**How to cite this paper:** Sarhan, A.M., El-Gohary, A.I. and Tolba, A.H. (2017) Statistical Analysis of a Competing Risks Model with Weibull Sub-Distributions. *Applied Mathematics*, 8, 1671-1683. <https://doi.org/10.4236/am.2017.811121>

**Received:** October 11, 2017

**Accepted:** November 26, 2017

**Published:** November 29, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Statistical inference for a competing risks model using Weibull sub-distributions is discussed in this paper. Both maximum likelihood and the Bayesian procedures are applied to report the point and interval estimations of all model parameters and some of its reliability measures. Complete analysis of a real data set is performed to show the applicability of the studied model.

## Keywords

Maximum Likelihood Estimator, Bayesian Method, Reliability Analysis, Survival Data

## 1. Introduction

In several applications in reliability analysis, medical research and biological situation there may be more than one cause of failure competing for the event. The event can be either death or recover from a certain disease (risk), see [1] [2] [3] [4]. In the literatures, the model that can be applied to deal with such situations is called competing risks model.

As examples of situations that require a competing risks model: 1) a patient may die from one of several causes such as cancer, heart disease, complications of diabetes, etc., and 2) in AIDS study, if the endpoint is time to opportunistic infection (relapse), the death and infection may be viewed as competing risks during the treatment of death and loss to follow-up as censoring, see [2] [5].

The available data, in the competing risks models, consist of the times to event and an indicator of either the cause of failure or censored. One of the interesting points in competing risks models is to report on the distribution of the time to failure from one risk in the presence of all other risks, which is called as the relative risk.

Competing risks models are studied by several authors based on both parametric and non-parametric setup, [6] [7]. It is assumed, in the parametric setup, that the risks follow a variety of lifetime distributions. Examples of such distributions include gamma, Weibull, exponential and generalized exponential, [4] [8] [9] [10] [11] [12]. The non-parametric setup does not consider a specific life time distribution. The analysis of the non-parametric version of competing risks models is investigated by several authors, such as [13] [14] [15].

The Weibull distribution has been frequently used in reliability analysis, see [16] and [17]. Weibull distribution is flexible (its hazard function can be constant, decreasing or increasing) and therefore can be used quite effectively to analyze a variety of real data sets. Various properties of Weibull distribution are discussed by [1] [12]. [18] discussed statistical inferences of a competing risks model assuming exponential causes and Weibull causes of failure with the same shape parameter. They used the maximum likelihood method when the cause of failure is either known or unknown.

In this paper, we discuss statistical inferences of a competing risks model when risks follow Weibull distributions with both unknown scale and shape parameters when the data can be censored. We will use maximum likelihood and Bayes methods. In Bayesian approach, we will consider that the unknown parameters are independent and follow gamma prior distributions. The rest of the paper is organized as follows. Section 2 describes the competing risks model assumptions. In Section 3, we derive the likelihood function and obtain the maximum likelihood estimates and discuss the two-sided confidence intervals for the model parameters. The Bayesian procedure is discussed in Section 4. The risk due a specific cause in the presence of all other causes, for the underlying model, is discussed in Section 5. A complete analysis of a real data set is provided in Section 6. Finally, Section 7 concludes the paper.

## 2. Model Assumptions

Assume that there is a set of  $N$  identical and independent objects (or systems) on life test. Each object is assigned to one of  $k, k \geq 2$ , independent causes of failure. Every object is tested until it fails or reaches a censored time. In the failure case, the object fails due to only one cause. The whole test will be terminated after either all objects fail or it reach the censored times or a combination of the two. When an object fails, there will be two observable quantities, the object's lifetime, say  $T_i$ , and the cause of failure, say  $\delta$ , where  $\delta \in \{1, 2, \dots, k\}$ . While in the censored case, we observe only the censoring time. To simplify the notations, we set  $\delta = 0$  for the censoring case. Furthermore, we need the following assumptions throughout the paper:

- 1)  $T_i$  is the lifetime of object which has a cumulative distribution function  $F(\cdot)$ , survival function  $\bar{F}(\cdot)$ , and probability density function  $f(\cdot)$ ,  $i = 1, 2, \dots, N$ .
- 2)  $T_{ji}$  is the random time at which cause  $j$  ( $j = 1, 2, \dots, k$ ) may hit object  $i$ ,

$i = 1, 2, \dots, N$ . Since the  $N$  objects on the life test are identical, therefore  $T_{ji}$ , for  $i = 1, 2, \dots, N$ , are identically distributed random variables. Furthermore, we assume that  $T_{ji}$ , for  $i = 1, 2, \dots, N$ , has cumulative distribution function  $F_j(\cdot)$  (which is called also as sub-distribution function of cause  $j$ ), survival function (or sub-survival function)  $\bar{F}_j(\cdot)$ , probability density function  $f_j(\cdot)$ , and hazard rate function  $h_j(\cdot)$ , for  $j = 1, 2, \dots, k$ .

Obviously,  $T_i = \min_{1 \leq j \leq k} \{T_{ji}\}$ . That is, when  $\delta_i = j$ , where  $j \in \{1, 2, \dots, k\}$ ,  $T_i = T_{ji}$ , therefore,  $F(t) = P(T_i \leq t) = P(T_{ji} \leq t) = F_j(t)$ . While when  $\delta_i = 0$ ,  $\bar{F}(t) = P(T_i > t) = P(\min_{1 \leq j \leq k} \{T_{ji}\} > t) = \prod_{j=1}^k P(T_{ij} > t) = \prod_{j=1}^k \bar{F}_j(t)$ .

3) More specifically, we assume that  $T_{ji}$  follow Weibull distributions with unknown parameters  $\alpha_j$  and  $\beta_j$ , say  $W(\alpha_j, \beta_j)$ , for  $i = 1, 2, \dots, N$  and  $j = 1, 2, \dots, k$ . That is,  $T_{ji}$  has the cumulative distribution function

$$F_j(t) = 1 - \exp(-\alpha_j t^{\beta_j}), \quad \alpha_j, \beta_j > 0, t \geq 0, \quad (1)$$

the probability density function is

$$f_j(t) = \alpha_j \beta_j t^{\beta_j - 1} \exp(-\alpha_j t^{\beta_j}), \quad (2)$$

and the survival rate function is

$$\bar{F}_j(t) = \Pr(T > t) = \exp(-\alpha_j t^{\beta_j}), \quad (3)$$

and the hazard rate function is

$$h_j(t) = \alpha_j \beta_j t^{\beta_j - 1}, \quad (4)$$

where  $\beta_j$  is the shape parameter and  $\alpha_j$  is the scale parameter.

### 3. The Likelihood Function and Maximum Likelihood Estimates

In this section we will derive the likelihood function of the model parameters using the available data described above. The available data can be expressed as  $(T_1, \delta_1), (T_2, \delta_2), \dots, (T_N, \delta_N)$ , where

$$T_i = \begin{cases} T_{ji} & \text{if } \delta_i = j \in \{1, 2, \dots, k\}, \\ T_i & \text{if } \delta_i = 0 \text{ (here } T_i \text{ is a censoring time)}. \end{cases}$$

#### 3.1. The Likelihood Function

Using data described above, the likelihood function is

$$L(\theta | \cdot) = \prod_{i=1}^N \left[ \prod_{j=1}^k [f_j(t_i) \bar{F}_j(t_i)]^{I(\delta_i=j)} [\bar{F}(t_i)]^{I(\delta_i=0)} \right], \quad (5)$$

where  $\theta$  is the vector of  $2k$  unknown parameters,

$\theta = (\alpha_1, \alpha_2, \dots, \alpha_k, \beta_1, \beta_2, \dots, \beta_k)$ . Based on the model assumptions and using the well-known relations between the reliability measures “the survival, hazard rate and the probability density functions”, we can write the likelihood function as

$$L(\theta | \cdot) = \left[ \prod_{i=1}^N \prod_{j=1}^k [h_j(t_i)]^{I(\delta_i=j)} \right] \left[ \prod_{i=1}^N \prod_{j=1}^k \bar{F}_j(t_i) \right], \tag{6}$$

by taking the natural logarithm for both sides of Equation (6), we get the log-likelihood function in the general case as

$$\mathcal{L} = \sum_{i=1}^N \sum_{j=1}^k [I(\delta_i = j)(\ln h_j(t_i)) + \ln \bar{F}_j(t_i)], \tag{7}$$

where  $I(A)$  is an indicator function that is defined as  $I(A) = 1$  if  $A$  is true and 0, otherwise.

Substituting from (3) and (4) into (6) and (7), we get the likelihood function and the log-likelihood function for the competing risks model, when the causes follow Weibull distribution with unknown parameters, as

$$L(\theta | \cdot) = \left\{ \prod_{i=1}^N \prod_{j=1}^k [\alpha_j \beta_j t_i^{\beta_j - 1}]^{I(\delta_i=j)} \right\} e^{-\sum_{j=1}^k \alpha_j \left( \sum_{i=1}^N t_i^{\beta_j} \right)}, \tag{8}$$

and

$$\mathcal{L} = \sum_{i=1}^N \sum_{j=1}^k [I(\delta_i = j) [\ln \alpha_j + \ln \beta_j + (\beta_j - 1) \ln t_i] - \alpha_j t_i^{\beta_j}]. \tag{9}$$

As well known, the maximum likelihood point estimate of  $\theta$  is the set of values of its elements that maximize the likelihood (or log-likelihood) function.

The first partial derivatives of  $\mathcal{L}$  with respect to  $\alpha_l$  and  $\beta_l, l=1,2,\dots,k$  are

$$\frac{\partial \mathcal{L}}{\partial \alpha_l} = \sum_{i=1}^N \sum_{j=1}^k \delta_{jl} \left\{ I(\delta_i = j) \frac{1}{\alpha_j} - t_i^{\beta_j} \right\}, \tag{10}$$

$$\frac{\partial \mathcal{L}}{\partial \beta_l} = \sum_{i=1}^N \sum_{j=1}^k \delta_{jl} \left\{ I(\delta_i = j) \left[ \frac{1}{\beta_j} + \ln t_i \right] - \alpha_j t_i^{\beta_j} \ln t_i \right\}, \tag{11}$$

where  $\delta_{ij}, i, j=1,2,\dots,k$  is Kronecker delta. To get the maximum likelihood point estimates for the parameters, we set  $\frac{\partial \mathcal{L}}{\partial \alpha_l} = 0$  and  $\frac{\partial \mathcal{L}}{\partial \beta_l} = 0, l=1,2,\dots,k$ ,

then solve the system of  $2k$  equations obtained with respect to  $\alpha_j$  and  $\beta_j, j=1,2,\dots,k$ . The obtained system has no analytical solution, therefore we should use a numerical technique to get the maximum likelihood estimates of the parameters.

In order to obtain the information matrix, we need the second partial derivatives of  $\mathcal{L}$  with respect to  $\alpha_l$  and  $\beta_l, l=1,2,\dots,k$ , which are given below

$$\frac{\partial^2 \mathcal{L}}{\partial \alpha_l \partial \alpha_m} = -\sum_{i=1}^N \sum_{j=1}^k [I(\delta_i = j) \frac{\delta_{jl} \delta_{jm}}{\alpha_j^2}], \tag{12}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \alpha_l \partial \beta_m} = -\sum_{i=1}^N \sum_{j=1}^k \delta_{jl} \delta_{jm} [t_i^{\beta_j} \ln t_i], \tag{13}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \beta_l \partial \beta_m} = - \sum_{i=1}^N \sum_{j=1}^k \delta_{jl} \delta_{jm} \left[ \frac{I(\delta_i = j)}{\beta_j^2} + \alpha_j t_i^{\beta_j} (\ln t_i)^2 \right], \quad (14)$$

where  $l, m = 1, 2, \dots, k$ .

### 3.2. Confidence Intervals

The maximum likelihood estimators for the parameters cannot be obtained in analytic forms. Therefore, their actual distributions cannot be derived. However, we can use the asymptotic distribution of the maximum likelihood estimator to derive confidence intervals for  $\theta = (\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_k)$ . It is well known that

$$\hat{\theta} \rightarrow N_{2k}(\theta, I^{-1}(\hat{\theta})). \quad (15)$$

where  $N_{2k}$  denotes 2  $k$ -multidimension normal distribution and  $I^{-1}(\hat{\theta})$  is the variance-covariance matrix that can be obtained as the inverse of the information matrix of  $\hat{\theta}$ . The elements of the information matrix are the second partial derivatives of  $\mathcal{L}$  evaluated at the maximum likelihood point estimate of the unknown parameters. That is  $I(\theta) = (I_{ij}(\theta))$ ,  $i, j = 1, 2, \dots, k$ , where

$$I_{ij}(\theta) = -E \left[ \frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta_i \partial \theta_j} \right], \quad i, j = 1, 2, \dots, k, \quad (16)$$

and  $\theta = (\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_k)$ .

### 4. Bayesian Procedure

In this section, we use Bayes approach to estimate the unknown model parameters  $\alpha_j$  and  $\beta_j$ ,  $j = 1, 2, \dots, k$ . We assume that all parameters are independent and follow gamma distributions with different but known prior hyperparameters. That is, we assume that  $\alpha_j$  follows Gamma prior distribution with shape parameter  $a_{j1}$  and scale parameter  $a_{j2}$ , and  $\beta_j$  follows Gamma prior distribution with shape parameter  $b_{j1}$  and scale parameter  $b_{j2}$ , for  $j = 1, 2, \dots, k$ . Therefore, the joint prior density of  $\theta = (\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_k)$ , up to a constant, is

$$g(\theta) \propto \prod_{j=1}^k \alpha_j^{a_{j1}-1} e^{-a_{j2}\alpha_j} \beta_j^{b_{j1}-1} e^{-b_{j2}\beta_j}, \quad \alpha_j, \beta_j > 0. \quad (17)$$

Combine the likelihood function (8) and the joint prior density (17), and using the Bayes' theorem, we get the joint posterior density function of  $\theta$ , up to a constant, as

$$g(\theta | \cdot) \propto \prod_{j=1}^k \alpha_j^{a_{j1}+n_j-1} \beta_j^{b_{j1}+n_j-1} e^{-\alpha_j \left( a_{j2} + \sum_{i=1}^N I_i^{\beta_j} \right)} e^{-\left( b_{j2} - \sum_{i=1}^N I(\delta_i=j) \log t_i \right) \beta_j}, \quad (18)$$

where  $n_j = \sum_{i=1}^N I(\delta_i = j)$ ,  $j = 1, 2, \dots, k$ .

Under quadratic loss function, the Bayesian estimate of any function of the vector of unknown parameters  $\theta$ , say  $v(\theta)$ , is the posterior mean of that function. That is,

$$\hat{v} = E_{\theta} (v(\theta)) = \int_0^{\infty} v(\theta) g(\theta | \cdot) d\theta. \quad (19)$$

The integral in Equation (19) as well as the normalized constant included in (18) have no analytical solutions. Therefore, numerical approaches should be used to do Bayesian analysis of the underlying model. Among several approaches, we will use Markov Chain Monte Carlo (MCMC) simulation technique to do the analysis. MCMC algorithm can be used to get random draws from the posterior distribution with density given in (18) without calculating the normalized constant. Then we can use the random draws to do any analysis we wish about the model parameters and model characteristics.

### Markov Chain Monte Carlo Method

One of the most successful methods in modern Bayesian statistics is the Markov Chain Monte Carlo (MCMC) technique. MCMC method is an algorithm to summarize the posterior distribution without calculating the normalized constant. MCMC techniques have been extensively used to become among the main computational tools in the Bayesian statistical inference [19]. The Metropolis-Hasting sampler is a modified version of the MCMC method. One of the main ideas in MCMC is to find a suitable distribution function, called as “proposal” that satisfies two conditions: 1) easy to simulate from, and 2) it mimics the posterior distribution function of interest. Once we determine such proposal, we get random draws from it, we apply the acceptance-rejection rule to get random draws from our target posterior distribution.

The following describes the steps of Metropolis-Hasting algorithm to simulate random draws from the posterior distribution  $g(\theta | \cdot)$ :

- 1) Set starting point of the chain, say  $\theta^{(0)}$ .
- 2) Set a size of trails we get for the random draws, say  $M$ .
- 3) For  $i = 1, \dots, M$  repeat the following steps:
  - a) Set  $\theta = \theta^{(i-1)}$ .
  - b) Generate a candidate  $\theta^*$  from a proposal distribution  $p(\theta^* | \theta)$ .
  - c) Calculate the acceptance probability  $\hat{h}$  as  $\hat{h} = \min\{1, R\}$ , where

$$R = \frac{g(\theta^* | \cdot) p(\theta | \theta^*)}{g(\theta | \cdot) p(\theta^* | \theta)}.$$

- d) Set  $\theta^{(i)} = \theta^*$  with probability  $\hat{h}$  or otherwise set  $\theta^{(i)} = \theta$ .

Under some regularity conditions on the proposal density  $p(\theta^* | \theta)$ , the sequence of the simulated draws  $\{\theta^{(i)}\}_{i=1}^M$  will converge to random draws that follow the posterior density  $g(\theta | \cdot)$ .

### 5. The Risks

One of the important characteristics in competing risks models is the relative risk of one of the competing risks in the presence of all other risks. In this sec-

tion, we discuss how to estimate the failure probability distribution for each cause of failure in the existence of all others and the risk due to a specific cause of failure. The probability of failure due to cause  $j$  at time  $t$  in the presence of all other causes is, see [17] [20],

$$F_{C_j}(t) = \int_0^t h_j(u) \prod_{l=1}^k \bar{F}_l(u) du, \quad j = 1, 2, \dots, k. \quad (20)$$

The risk due to cause  $j = 1, 2, \dots, k$ , say  $\pi_j$ , is

$$\pi_j = \lim_{t \rightarrow \infty} F_{C_j}(t) = \int_0^{\infty} h_j(t) \prod_{l=1}^k \bar{F}_l(t) dt. \quad (21)$$

For the Weibull competing risks model discussed here,  $\pi_j$  can be obtained by solving the following integral

$$\pi_j = \int_0^{\infty} \alpha_j \beta_j t^{\beta_j - 1} e^{-\sum_{l=1}^k \alpha_l t^{\beta_l}} dt, \quad j = 1, 2, \dots, k, \quad (22)$$

The above integral has no analytic solution. As special case, when all shape parameters for causes are equal,  $\beta_1 = \beta_2 = \dots = \beta_k$ , we get

$$\pi_j = \frac{\alpha_j}{\sum_{l=1}^k \alpha_l}, \quad j = 1, 2, \dots, k. \quad (23)$$

Using the invariant property, the maximum likelihood estimate of  $\pi_j$  can be obtained by numerical integration when the integrand function in (22) evaluated at the maximum likelihood estimates of the parameters. For Bayesian analysis, we will use the random draws that obtained from the joint distribution along with the integral above to get random draws from the posterior distribution of  $\pi_j$ , then use it to do all Bayes analysis we wish on  $\pi_j$ ,  $j = 1, 2, \dots, k$ .

## 6. Application

We study in this section an application of a real-life data set. This data set gives the times (in years) from HIV infection to AIDS, SI switch and death in 329 men who have sex with men (MSM). Data are from the period until combination anti-retroviral therapy became available (1996). For more background information on this data, see [21] [22]. It was used as example for the competing risks analyses in [23] [24]. These data can be considered as competing risks data with two risks. There are some individuals in the study left with no infection switch or death. Those are considered as censored observations.

Here, we use the model discussed in this paper to analyze this data set. **Table 1** shows the maximum likelihood and Bayes point estimates of the four model parameters. We used R language to do all calculations.

As shown in **Table 1**, the results from both methods are completely different. Since we do not know the actual values of the unknown parameters, then using those point estimates will not be sufficient to judge which method provides better estimates. One of the comparison tools is the interval estimation or the corresponding variance to each point estimate. This why we calculated them as

**Table 1.** Estimates the parameters  $\alpha_1, \beta_1, \alpha_2$  and  $\beta_2$  by maximum likelihood and bayesian methods.

Methods	Parameters			
	$\alpha_1$	$\beta_1$	$\alpha_2$	$\beta_2$
Maximum likelihood estimates	0.0240	1.4362	0.0488	1.1515
Bayes pestimates	0.0065	1.9500	0.0206	1.4010

given below.

Below is the inverse of the Fisher information matrix that approximates the variance-covariance matrix for the maximum likelihood estimates of the vector of unknown parameters  $\theta$ :

$$I^{-1} = \begin{pmatrix} 96.0940576 & -0.44433333 & 0.00000000 & 0.00000000 \\ -0.44433333 & 0.002359471 & 0.00000000 & 0.00000000 \\ 0.00000000 & 0.00000000 & 13.3737256 & -0.183498758 \\ 0.00000000 & 0.00000000 & -0.1834988 & 0.003166101 \end{pmatrix}$$

As we can see, the variances of  $\hat{\alpha}_1$  and  $\hat{\alpha}_2$  are very large comparing to their values. This would lead to negative lower bound of the asymptotic confidence intervals of  $\alpha_1$  and  $\alpha_2$ . Since  $\alpha_1$  and  $\alpha_2$  cannot be negative, we truncate the lower limits at zero. This is one of the disadvantages of the maximum likelihood method.

**Table 2** shows asymptotic 95% confidence intervals of the four unknown parameters. **Table 3** shows the estimated risks using both two methods. The 95% credible intervals of the four unknown parameters are shown in **Table 4**, while **Table 5** gives the measures of central tendency of the four parameters using their marginal posterior distributions. As **Table 4** and **Table 5** show, not only the lower limits of the credible intervals of  $\alpha_1$  and  $\alpha_2$  are larger than zero, but the minimum values of  $\alpha_1$  and  $\alpha_2$  are larger than zero. This is one of the advantages of the Bayes method.

In Bayes case, we set all hyperparameters equal and equal to 0.001 that reflects non-informative prior. To apply MCMC, we used the proposal as a multivariate t distribution with mode equal to the vector of the maximum likelihood estimates, variance-covariance matrix equals to the inverse Fisher matrix and four degrees of freedom. Also, the number of draws, M, was 10000. As a diagnostic test for the MCMC, we plotted the autocorrelations and the traces for each parameter as shown in **Figure 1** and **Figure 2**. The trace plots show good mix of the sampled draws and the autocorrelation plots show that the Lag decreases rapidly, which indicates that the draws become approximately independent over time and the draws come from the actual posterior distribution. We used the random draws from the joint posterior distribution of  $\theta$  to get random draws for the risks  $\pi_j, j=1,2$  from their marginal posterior distributions without obtaining those actual marginal distributions. We used those draws to calculate Bayes estimates of  $\pi_j$  and sketch their posterior density functions. **Figure 3**

**Table 2.** The asymptotic confidence intervals using Maximum likelihood method.

Parameters	Confidence intervals	
	lower bound	upper bound
$\alpha_1$	0.00000	19.23708
$\beta_1$	1.340984	1.531392
$\alpha_2$	0.000000	7.216447
$\beta_2$	1.041220	1.261787

**Table 3.** Estimates the two risks.

Method	Risk	
	$\hat{\pi}_1$	$\hat{\pi}_2$
MLE	0.3296700	0.6703300
BE	0.5511003	0.4488998

**Table 4.** 95% credible intervals for the four parameters.

parameters	Probability intervals	
	lower bound	upper bound
$\alpha_1$	0.003143687	0.01227029
$\beta_1$	1.679310071	2.22135923
$\alpha_2$	0.012470806	0.03273711
$\beta_2$	1.186797437	1.61211860

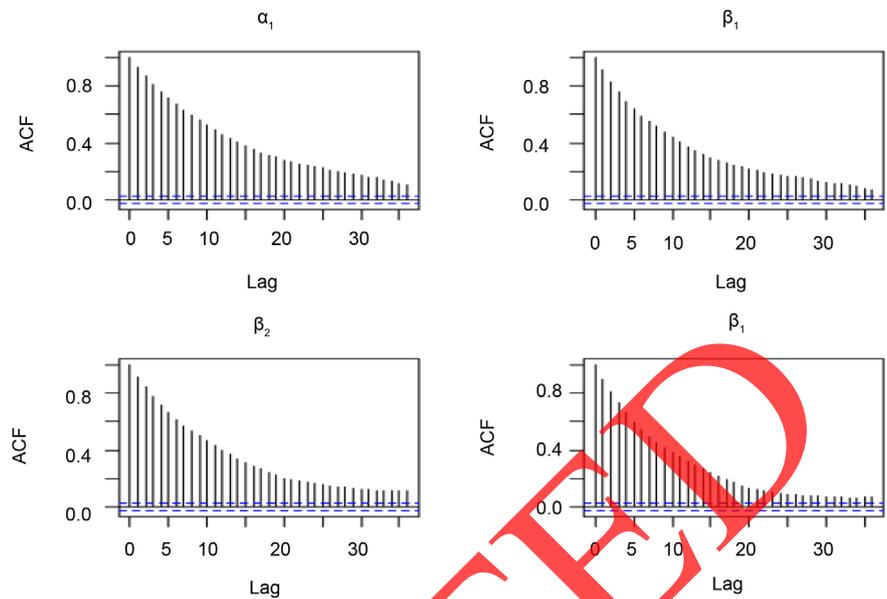
**Table 5.** The measures of central tendency of the parameters  $\alpha_1, \beta_1, \alpha_2$  and  $\beta_2$ .

Measures	$\alpha_1$	$\beta_1$	$\alpha_2$	$\beta_2$
Min.	0.001972	1.522	0.008279	1.038
First Quartile	0.004857	1.852	0.016806	1.327
Median	0.006273	1.947	0.020028	1.398
Mean	0.006509	1.953	0.020612	1.401
Third Quartile	0.007707	2.054	0.023866	1.478
Max.	0.016232	2.466	0.048149	1.816

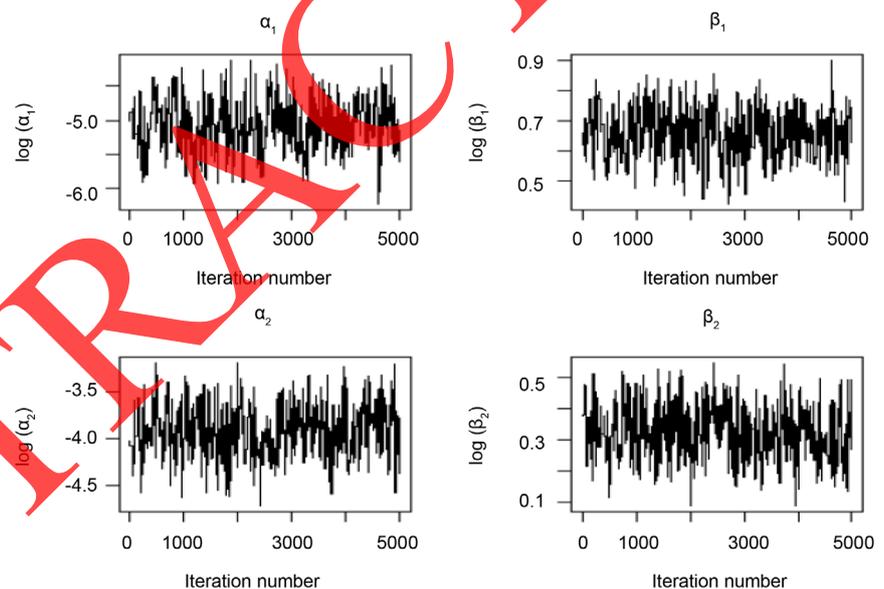
shows the trace plots of the draws of  $\pi_j$  and the corresponding marginal posterior density functions. Furthermore, we used the random draws of  $\theta$  to get random draws for the sub-survivors and the overall survivor and used them to calculate the Bayes estimators along with 95% credible intervals for those functions at different time as shown in **Figure 4**.

## 7. Conclusions

In this paper, we attempted to examine competing risks models with  $k, k \geq 2$ ,



**Figure 1.** The autocorrelation plot of the simulated draws using MCMC after discarding the first 50% of the draws.



**Figure 2.** The trace plots of the simulated draws after discarding the early 50% of the draws.

independent causes of failure in the presence of censoring data. We derived the likelihood equation of the model in general and used it to derive the likelihood function when the risks follow Weibull distributions with unknown shape and scale parameters.

We discussed how to get the maximum likelihood estimates and Bayes estimates of the model parameters as well as for some important reliability measures such as the individual risks, sub-survivors and overall survivor. In Bayes analysis, we used gamma prior distribution for all unknown parameters with known

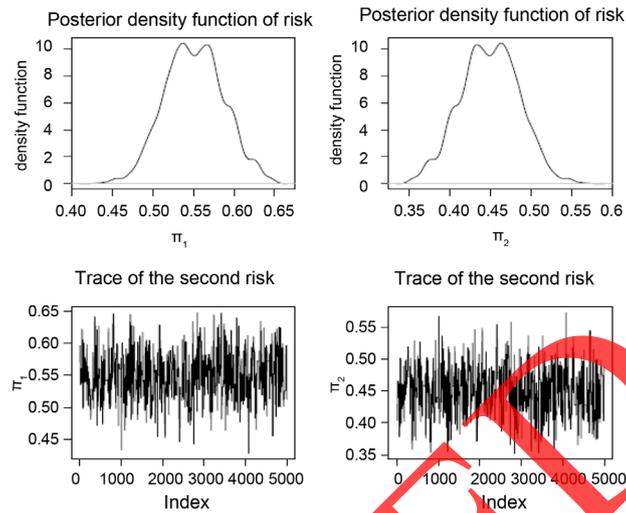


Figure 3. Trace plots and posterior density functions of  $\pi_j, (j = 1, 2)$ .

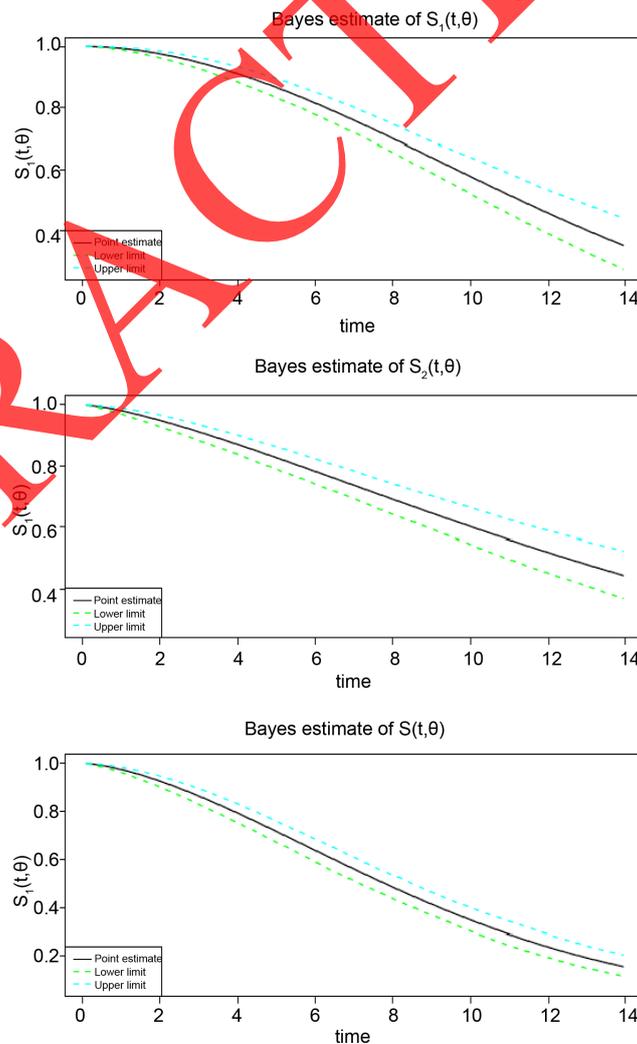


Figure 4. Bayes point and interval estimates of the sub-survivor functions and the overall survivor function.

hyperparameters and used MCMC to get random draws from the joint posterior distribution function. Also, we discussed the asymptotic confidence intervals and credible intervals of all unknown parameters. The model is applied to a real data set and detailed discussion was provided.

### Acknowledgements

The authors are grateful to the anonymous referees for a careful checking of the details and for helpful comments that improved this paper.

### References

- [1] De Wreede, L.C., Fiocco, M. and Putter, H. (2010) The Mstate Package for Estimation and Prediction in Non- and Semiparametric Multi-State and Competing Risks Models. *Computer Methods and Programs in Biomedicine*, **99**, 261-274. <https://doi.org/10.1016/j.cmpb.2010.01.001>
- [2] Brooks, S., Gelman, A., Jones, G.L. and Meng, X.L. (2011) Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Chapman & Hall/CRC, Boca Raton.
- [3] Schmoor, C., Schumacher, M., Finke, J. and Beyersmann, J. (2013) Competing Risks and Multistate Models. *Clinical Cancer Research*, **19**, 12-21. <https://doi.org/10.1158/1078-0432.CCR-12-1619>
- [4] Geskus, R.B. (2015) Data Analysis with Competing Risks and Intermediate States. CRC Press.
- [5] Hinchliffe, S.R. (2013) Advancing and Appraising Competing Risks Methodology for Better Communication of Survival Statistics. University of Leicester, Leicester.
- [6] Lawless, J.F. (2011) Statistical Models and Methods for Lifetime Data. John Wiley & Sons.
- [7] Haller, B. (2014) The Analysis of Competing Risks Data with a Focus on Estimation of Cause-specific and Subdistribution Hazard Ratios from a Mixture Model, Imu.
- [8] Cox, D.R. (1959) The Analysis of Exponentially Distributed Lifetimes with Two Types of Failure. *Journal of the Royal Statistical Society. Series B (Methodological)*, 411-421.
- [9] Berkson, J. and Elveback, L. (1960) Competing Exponential Risks, with Particular Reference to the Study of Smoking and Lung Cancer. *Journal of the American Statistical Association*, **55**, 415-428. <https://doi.org/10.1080/01621459.1960.10482072>
- [10] David, H.A. and Moeschberger, M.L. (1978) The Theory of Competing Risks. Griffin, London.
- [11] Kundu, D. and Sarhan, A.M. (2006) Analysis of Incomplete Data in Presence of Competing Risks among Several Groups. *IEEE Transactions on Reliability*, **55**, 262-269. <https://doi.org/10.1109/TR.2006.874919>
- [12] Ahmad, S.P. and Ahmad, K. (2013) Bayesian Analysis of Weibull Distribution using R Software. *Australian Journal of Basic and Applied Sciences*, **7**, 156-164.
- [13] Kaplan, E.L. and Meier, P. (1958) Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, **53**, 457-481.
- [14] Park, C., *et al.* (2005) Parameter Estimation of Incomplete Data in Competing Risks using the EM Algorithm. *IEEE Transactions on Reliability*, **54**, 282-290.
- [15] Sarhan, A.M. (2007) Analysis of Incomplete, Censored Data in Competing Risks Models with Generalized Exponential Distributions. *IEEE Transactions on Reliability*

ity, **56**, 132-138. <https://doi.org/10.1109/TR.2006.890899>

- [16] Hossain, A. and Zimmer, W. (2003) Comparison of Estimation Methods for Weibull Parameters: Complete and Censored Samples. *Journal of Statistical Computation and Simulation*, **73**, 145-153. <https://doi.org/10.1080/00949650215730>
- [17] Sarhan, A.M., Hamilton, D.C. and Smith, B. (2010) Statistical Analysis of Competing Risks Models. *Reliability Engineering & System Safety*, **95**, 953-962.
- [18] Kundu, D. and Basu, S. (2000) Analysis of Incomplete Data in Presence of Competing Risks. *Journal of Statistical Planning and Inference*, **87**, 221-239.
- [19] Albert, J. (2009) Bayesian Computation with R. Springer Science & Business Media. <https://doi.org/10.1007/978-0-387-92298-0>
- [20] Bocchetti, D., Giorgio, M., Guida, M. and Pulcini, G. (2009) A Competing Risk Model for the Reliability of Cylinder Liners in Marine Diesel Engines. *Reliability Engineering & System Safety*, **94**, 1299-1307.
- [21] Dukers, N.H.T.M., Renwick, N., Prins, M., Geskus, R.B., Schulz, T.F., Weverling, G.-J., Coutinho, R.A. and Goudsmit, J. (2000) Risk Factors for Human Herpesvirus 8 Seropositivity and Seroconversion in a Cohort of Homosexual Men. *American Journal of Epidemiology*, **151**, 213-224.
- [22] Xiridou, M., Geskus, R., de Wit, J., Coutinho, R. and Kretzschmar, M. (2003) The Contribution of Steady and Casual Partnerships to the Incidence of HIV Infection among Homosexual Men in Amsterdam. *Aids*, **17**, 1029-1038. <https://doi.org/10.1097/00002030-200305020-00012>
- [23] Putter, H., Fiocco, M. and Geskus, R.B. (2007) Tutorial in Biostatistics: Competing Risks and Multi-State Models. *Statistics in Medicine*, **26**, 2389-2430. <https://doi.org/10.1002/sim.2712>
- [24] Geskus, R.B., Gonzalez, C., Torres, M., Del Romero, J., Viciano, P., Masia, M., Blanco, J.R., Iribarren, M., De Sanjose, S., Hernandez-Novoa, B., et al. (2016) Incidence and Clearance of Anal High-Risk Human Papillomavirus in HIV-Positive Men Who Have Sex with Men: Estimates and Risk Factors. *AIDS (London, England)*, **30**, 37.