Hacettepe Journal of Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 52 (5) (2023), 1263 – 1281
DOI : 10.15672/hujms.1066111

Research Article

Analysis and modelling of competing risks survival data using modified Weibull additive hazards regression approach

Habbiburr Rehman[1], Navin Chandra[*2], Ali H. Abuzaid[3]

[1] *Department of Medicine, Biomedical Genetics, Boston University Chobanian & Avedisian School of Medicine, Boston, USA*
[2] *Department of Statistics, Ramanujan School of Mathematical Sciences, Pondicherry University, Puducherry, India*
[3] *Department of Mathematics, Faculty of Science, Al Azhar University-Gaza, Gaza, Palestine*

Abstract

The cause-specific hazard function plays an important role in developing the regression models for competing risks survival data. Proportional hazards and additive hazards are the commonly used regression approaches in survival analysis. Mostly, in literature, the proportional hazards model was used for parametric regression modelling of survival data. In this article, we introduce a parametric additive hazards regression model for survival analysis with competing risks. For employing a parametric model we consider the modified Weibull distribution as a baseline model which is capable to model survival data with non-monotonic behaviour of hazard rate. The estimation process is carried out via maximum likelihood and Bayesian approaches. In addition to Bayesian methods, a class of non-informative types of prior is introduced with squared error (symmetric) and linear-exponential (asymmetric) loss functions. The relative performance of the different estimators is assessed using Monte Carlo simulation. Finally, using the proposed methodology, a real data analysis is performed.

Mathematics Subject Classification (2020). 62N02, 62F15, 62P10

Keywords. Additive hazard, competing risks, modified Weibull distribution, Bayes estimation, non-informative priors, MCMC

1. Introduction

In survival studies, it is often interesting to observe individuals lifetime with $p \geq 2$ mutually exclusive types of events or competing risks [9, 13]. In this situation, occurrence of one type of failure alter the chance of the occurrence of other types of failure. For example, primary biliary cirrhosis (PBC) is a chronic liver disease in which individual may receive the transplant and experience the death in waiting queue. In breast cancer clinical trial, investigators may be interested to observe events such as local relapse, auxiliary relapse, remote relapse, second malignancy of any kind, and death. The frequently used

*Corresponding Author.
Email addresses: rehmanh79@gmail.com (H. Rehman), nc.stat@gmail.com (N. Chandra), alizaid33@yahoo.com (A.H. Abuzaid)
Received: 15.11.2020; Accepted: 03.03.2023

competing risks modelling methods depend on the observed value of the bivariate random vector (T, C), where T denotes the lifetime (possible censored) and $C = j, j \in 1, 2, \ldots, p$ is the set of possible causes of failure. In this framework, the basic identifiable quantities are the cause-specific hazard (CSH) function and the cumulative incidence function (CIF). For comprehensive review and discussion on competing risks one may refer to [13, 17, 25].

The survival time is frequently arises with auxiliary information in the form of the covariates such as treatment, sex, age, and so on. In this scenario, regression models are useful tools for understanding and exploiting the relationship between survival time and covariates. Generally, the regression models are developed through CSH function for competing risks survival analysis. In practice, the widely used analysis of survival data with competing risks is non-parametric and semi-parametric. However, if the model is correctly specified then parametric models are more efficient than semi-parametric models [17].

In literature, there is a considerable amount of work on parametric modelling of competing risks using the CSH function. Jeong and Fine [11] considered the Weibull CSH function and the direct improper Gompertz distribution for parametric modelling of the CIF. Anjana and Sankaran [2] proposed the parametric reverse cause-specific proportional hazards (PH) regression model by assuming inverse Weibull model under left censoring. Lee [18] provided the parametric quantile inference for CSH function with adjustment of covariates. Rehman et al. [29] presented the survival analysis with competing risks under parametric PH model with Bayesian approach.

Parametric regression analysis of competing risks survival data in the above mentioned literature is mainly based on the Cox's PH model [4]. In the PH model, the effects of the covariates act multiplicatively on some baseline hazard rate. Aalen [1] introduced an important alternative to the PH model that is the additive hazards (AH) regression model and later studied by [20, 21]. In the AH model, the hazard rate with the associated covariates is defined as the sum of the baseline hazard rate and regression function of the covariates. In a two sample set-up, the PH model concern the risks ratio, whereas AH model addresses the risks difference. Shen and Cheng [32] proposed the confidence bands for CIF under AH model. Sun et al. [35] considered the AH model for competing risks analysis of the case-cohort design. Zhang et al. [38] proposed the regression analysis of competing risks data via semi-parametric AH model. Li et al. [19] analyzed an additive sub-distribution hazard model for competing risks data.

Parametric AH regression model may developed by assuming some known distributional form for baseline hazard function [31]. As much as we know, survival analysis with competing risks based on parametric AH regression model has not received any attention. Therefore, the objective of the present work is to employ parametric AH regression model for competing risks survival data. In this article, we study the modified Weibull distribution (MWD) with one scale and two shape parameters which is capable to capture various shapes of the hazard rate like bathtub failure [15]. The aim of this attempt is to consider both classical and Bayesian methods of estimation. Recently, Rehman et al. [28] proposed the Bayesian estimation based on the class of informative priors for the modified Weibull AH regression model under competing risks. Therefore, in this article, we consider Bayesian estimation based on a class of non-informative types of prior namely, uniform, Jeffreys and half-t for baseline parameters and uniform prior for regression parameters. The squared error and linear-exponential (LINEX) loss functions, which are symmetric and asymmetric loss functions, respectively, are used to derive the Bayes estimates.

The rest of the paper is organised as follows: we introduce a parametric cause-specific AH regression model in Section 2. In Section 3, we estimate the model parameters by using maximum likelihood method. In Section 4, the Bayesian estimation is considered under non-informative priors with two loss functions. A Monte Carlo simulation study is carried out to examine the finite sample behaviour of the estimators in Section 5. In

Section 6, the applicability of the proposed model is demonstrated with real data. Finally, concluding remarks are given in Section 7.

2. Model specification

According to [27], the CSH function of the survival time T, simply gives the instantaneous failure rate from a particular cause j among the individuals who are currently event free in the presence of remaining competing causes (other than j). Mathematically, the CSH function at t due to cause j conditional on $m \times 1$ vector of covariates $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)^\top$ is given by

$$h_j(t|\boldsymbol{x}) = \lim_{\Delta t \to 0} \frac{Pr(t \leq T < t + \Delta t, C = j \mid T \geqslant t, \boldsymbol{x})}{\Delta t}, \quad j = 1, 2, \ldots, p. \tag{2.1}$$

In the context of competing risks analysis, calculating the probability of a certain type of failure, say j at time t in the presence of competing risks may be useful. Such probability is known as CIF and defined by

$$F_j(t|\boldsymbol{x}) = \Pr(T \leq t, C = j \mid \boldsymbol{x}), \quad j = 1, 2, \ldots, p. \tag{2.2}$$

Since the causes of failure are mutually exclusive. So, the overall hazard $h(t)$ and cumulative distribution function $F(t)$ of T are given below

$$h(t|\boldsymbol{x}) = \sum_{j=1}^{p} h_j(t|\boldsymbol{x}), \quad F(t|\boldsymbol{x}) = \sum_{j=1}^{p} F_j(t|\boldsymbol{x}).$$

Therefore, the CIF can be evaluated in terms of the CSH function as follows

$$F_j(t|\boldsymbol{x}) = \int_0^t h_j(u|\boldsymbol{x}) \exp\left(-\sum_{j=1}^{p} H_j(u|\boldsymbol{x})\right) du, \tag{2.3}$$

where $H_j(t|\boldsymbol{x})$ is the cumulative CSH function. The overall survival function in terms of CSH function is defined as

$$S(t|\boldsymbol{x}) = \exp\left(-\sum_{j=1}^{p} H_j(t|\boldsymbol{x})\right). \tag{2.4}$$

For the detail interpretation and use of the equations (2.1)-(2.4) one may refer to [26].

In this study, to develop a regression model for competing risks survival data we consider the AH regression model given by [21]. In this model, the constant effect of covariates on the baseline hazard function is additive in nature. This model for CSH rate turns out in the following form

$$h_j(t|\boldsymbol{x}) = h_{0j}(t) + \boldsymbol{\beta}_j^\top \boldsymbol{x}, \quad j = 1, 2, \ldots, p, \tag{2.5}$$

where $h_j(t|\boldsymbol{x})$ and $h_{0j}(t)$ denotes the CSH functions for given covariates $\boldsymbol{x}$ and at baseline, respectively and $\boldsymbol{\beta}_j = (\beta_{j1}, \beta_{j2}, \ldots, \beta_{jm})^\top$ is the $m \times 1$ vector of cause-specific regression parameters. In the present work, we study the MWD with one scale parameter a and two shape parameters α and λ for lifetime variate T with cumulative distribution function, and hazard function are given as

$$F(t) = 1 - \exp(-at^\alpha e^{\lambda t}), \quad t \geq 0, a > 0, \alpha \geq 0, \lambda > 0, \tag{2.6}$$

$$h(t) = a(\alpha + \lambda t)t^{\alpha-1}e^{\lambda t}, \quad t \geq 0, a > 0, \alpha \geq 0, \lambda > 0. \tag{2.7}$$

The commonly used parametric models for survival analysis are exponential, Weibull and gamma distributions etc., which can only capture the monotonic behaviour of the hazard rate, for example, increasing, decreasing and constant hazard rates. However, MWD can capture non-monotonic behaviour of the hazard rate and has exponential, Weibull, and Type I extreme value distributions as special cases [15]. In equation (2.6), if we assume that $\lambda = 0$ and $\lambda = 0, \alpha = 1$ then the distribution function $F(t)$ reduces to

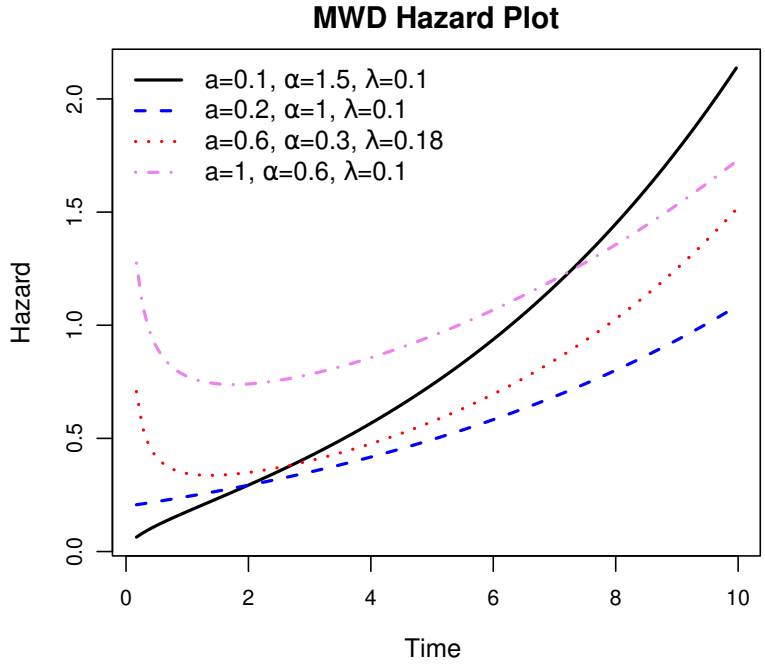

Figure 1. Hazard plot of MWD for different parameter values.

Weibull and exponential distribution, respectively. Therefore, MWD covers a wider range of real-life applications compared to exponential, Weibull and Gamma distributions. So, motivated by this, we assume MWD as a baseline parametric model. Lai et al. [15] developed the MWD and discussed some of its theoretical properties, for example, the behaviour of the hazard rate see Figure 1. Ng [24] estimated the parameters of the MWD for progressively type -II censored samples. Further, some Bayesian estimations of MWD parameters are considered by [12] and [37]. The MWD is assumed here as a baseline model of the cause-specific AH analysis in (2.5) due to its flexibility to accommodate various shapes of the hazard function.

Accordingly, the CSH function, cumulative CSH function, and overall survival function are obtained as

$$h_j(t;\boldsymbol{\Theta}_j,\boldsymbol{x}) = a_j(\alpha_j+\lambda_j t)t^{\alpha_j-1}e^{\lambda_j t} + \boldsymbol{\beta}_j^{\top}\boldsymbol{x}, \tag{2.8}$$

$$H_j(t;\boldsymbol{\Theta}_j,\boldsymbol{x}) = a_j t^{\alpha_j}e^{\lambda_j t} + \boldsymbol{\beta}_j^{\top}\boldsymbol{x}t, \tag{2.9}$$

and

$$S(t;\boldsymbol{\Theta},\boldsymbol{x}) = \exp\left\{-\left(\sum_{j=1}^{p} a_j t^{\alpha_j}e^{\lambda_j t} + \boldsymbol{\beta}_j^{\top}\boldsymbol{x}t\right)\right\}, \tag{2.10}$$

where $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \ldots, \boldsymbol{\Theta}_p)$, and $\boldsymbol{\Theta}_j = (a_j, \alpha_j, \lambda_j, \boldsymbol{\beta}_j)$ vector of cause-specific parameters. The main attention of this article is to estimate the unknown parameters and cumulative CSH function as the quantity of interest.

3. Maximum likelihood estimation

In the competing risks framework, suppose that T is the observed lifetime which is defined by $T = \min(T^*, D)$, where T^* is the failure time and D is the censoring time. For the given covariate $\boldsymbol{x}$, T^* and D are assumed to be independent. Further, we assume that

for each observed failure time, the associated cause of failure is also observed. Therefore, the censoring indicator is defined as $\delta_{ij} = I(T_i = T_i^*, C_i = j)$. Let $(t_i, \delta_{ij}, \boldsymbol{x}_i)$, $i = 1, 2, \ldots, n$ be the $n \in \mathbb{N}$ independently and identically distributed samples of $(T, \delta, \boldsymbol{x})$. Now we can write the likelihood function for the observed data as

$$L(\boldsymbol{\Theta}) = \prod_{i=1}^{n} \left(\prod_{j=1}^{p} h_j(t_i; \boldsymbol{\Theta}_j, \boldsymbol{x}_i)^{\delta_{ij}} S(t_i; \boldsymbol{\Theta}, \boldsymbol{x}_i) \right). \tag{3.1}$$

The fully parameterized likelihood function based on (2.8) and (2.10) is given by

$$L(\boldsymbol{\Theta}) = \prod_{i=1}^{n} \left[\prod_{j=1}^{p} \left(a_j(\alpha_j + \lambda_j t_i) t_i^{\alpha_j - 1} e^{\lambda_j t_i} + \boldsymbol{\beta}_j^\top \boldsymbol{x}_i \right)^{\delta_{ij}} \exp\left\{ - \left(\sum_{j=1}^{p} a_j t_i^{\alpha_j} e^{\lambda_j t_i} + \boldsymbol{\beta}_j^\top \boldsymbol{x}_i t_i \right) \right\} \right]. \tag{3.2}$$

The log likelihood function $\ell(\boldsymbol{\Theta}) = \log L(\boldsymbol{\Theta})$ is given as

$$\ell(\boldsymbol{\Theta}) = \sum_{j=1}^{p} \sum_{i=1}^{n_j} \log \left(a_j(\alpha_j + \lambda_j t_i) t_i^{\alpha_j - 1} e^{\lambda_j t_i} + \boldsymbol{\beta}_j^\top \boldsymbol{x}_i \right) - \sum_{i=1}^{n} \left(\sum_{j=1}^{p} a_j t_i^{\alpha_j} e^{\lambda_j t_i} + \boldsymbol{\beta}_j^\top \boldsymbol{x}_i t_i \right). \tag{3.3}$$

In equation (3.3), n_j denotes the number of failure of type j. To obtain the estimates of the unknown parameters a_j, α_j, λ_j and $\boldsymbol{\beta}_j$ we maximize the (3.3) by equating the partial derivatives of each parameter to zero. The score equations are obtained as

$$\frac{\partial \ell(\boldsymbol{\Theta})}{\partial a_j} = \sum_{i=1}^{n_j} \frac{(\alpha_j + \lambda_j t_i) t_i^{\alpha_j - 1} e^{\lambda_j t_i}}{a_j(\alpha_j + \lambda_j t_i) t_i^{\alpha_j - 1} e^{\lambda_j t_i} + \boldsymbol{\beta}_j^\top \boldsymbol{x}_i} - \sum_{i=1}^{n} t_i^{\alpha_j} e^{\lambda_j t_i} = 0, \tag{3.4}$$

$$\frac{\partial \ell(\boldsymbol{\Theta})}{\partial \alpha_j} = \sum_{i=1}^{n_j} \frac{a_j t_i^{\alpha_j - 1} e^{\lambda_j t_i} + a_j \alpha_j t_i^{\alpha_j - 1} \log t_i e^{\lambda_j t_i} + a_j \lambda_j t_i^{\alpha_j} \log t_i e^{\lambda_j t_i}}{a_j(\alpha_j + \lambda_j t_i) t_i^{\alpha_j - 1} e^{\lambda_j t_i} + \boldsymbol{\beta}_j^\top \boldsymbol{x}_i} - \sum_{i=1}^{n} a_j t_i^{\alpha_j} \log t_i e^{\lambda_j t_i} = 0, \tag{3.5}$$

$$\frac{\partial \ell(\boldsymbol{\Theta})}{\partial \lambda_j} = \sum_{i=1}^{n_j} \frac{a_j \alpha_j t_i^{\alpha_j} e^{\lambda_j t_i} + a_j t_i^{\alpha_j} e^{\lambda_j t_i} + a_j \lambda_j t_i^{\alpha_j + 1} e^{\lambda_j t_i}}{a_j(\alpha_j + \lambda_j t_i) t_i^{\alpha_j - 1} e^{\lambda_j t_i} + \boldsymbol{\beta}_j^\top \boldsymbol{x}_i} - \sum_{i=1}^{n} a_j t_i^{\alpha_j + 1} e^{\lambda_j t_i} = 0, \tag{3.6}$$

$$\frac{\partial \ell(\boldsymbol{\Theta})}{\partial \boldsymbol{\beta}_j} = \sum_{i=1}^{n_j} \frac{\boldsymbol{x}_i}{a_j(\alpha_j + \lambda_j t_i) t_i^{\alpha_j - 1} e^{\lambda_j t_i} + \boldsymbol{\beta}_j^\top \boldsymbol{x}_i} - \sum_{i=1}^{n} \boldsymbol{x}_i t_i = 0. \tag{3.7}$$

The score equations (3.4)-(3.7) are not in explicit form and cannot be solved analytically. Therefore, we use numerical methods to estimate the parameters.

In the literature, several techniques are available for estimating the parameters by solving score equations or directly maximising the log-likelihood function. The Newton-Raphson technique is the most commonly used method for estimating the unknown parameters because the derivatives of the scoring equations are simple to calculate. The initial values are important in the numerical iterative procedure because of the logarithm function. In the literature, we do not have any theoretical method to choose the initial values for the complex models. To avoid the effect of initial values, we arbitrarily tried different sets of initial values in the iterative procedure. We chose those values which have the maximum value of the likelihood function. We use the simplex method proposed by [23] to estimate the parameters through `optim` function in `R software`. The simplex method is a straightforward method for estimating the parameters by maximising the likelihood function without having to optimise the function's derivatives. The results of the `optim` function shows the appropriate convergence of the algorithm, and model shows its identifiabilty in terms of time taking during the simulation study. Once the parameter estimates are obtained, the function of the parameters can be estimated using

the invariance property of the maximum likelihood estimator (MLE). Therefore, the MLE of cumulative CSH $H_j(t;\boldsymbol{\Theta}_j,\boldsymbol{x})$ is given by

$$\hat{H}_j(t;\hat{\boldsymbol{\Theta}}_j,\boldsymbol{x}) = \hat{a}_j t^{\hat{\alpha}_j} e^{\hat{\lambda}_j t} + \hat{\boldsymbol{\beta}}_j^{\top}\boldsymbol{x}t.$$

4. Bayes estimation

Frequentist statistical techniques do not incorporate the prior knowledge into data analysis. Bayesian inference is attractive in that because it incorporates prior or previous information with the observed data. Therefore, in this section, we introduced the Bayesian analysis of parametric cause-specific AH regression model. Prior assumption is based entirely on previous experiences, mathematical convenience and expert judgments, which can be informative, non-informative, or weakly informative. If the previous data is large enough, informative priors can be used. A non-informative prior can be used when only limited or vague knowledge (a priori) about the parameters is available. In this article, non-informative types of priors such as the uniform, Jeffreys, and half-t distributions for baseline parameters are described. A uniform non-informative prior is assumed for regression parameters. Also, it is assumed that all the chosen priors are independent.

4.1. Uniform prior

Suppose that the prior distributions of the random variables a_j, α_j and $\lambda_j, j = 1, 2, \ldots, p$ are the uniform distributions of the following form

$$\begin{aligned}
\pi_{1j}(a_j) &\propto \frac{1}{M_{a_j}}, \qquad 0 < a_j < M_{a_j},\\
\pi_{1j}(\alpha_j) &\propto \frac{1}{M_{\alpha_j}}, \qquad 0 < \alpha_j < M_{\alpha_j},\\
\pi_{1j}(\lambda_j) &\propto \frac{1}{M_{\lambda_j}}, \qquad 0 < \lambda_j < M_{\lambda_j}.
\end{aligned} \tag{4.1}$$

For regression parameter $\beta_{jl}, l = 1, 2, \ldots, m, j = 1, 2, \ldots, p$ we assumed a uniform distribution on (c_{jl}, d_{jl}) where $-\infty < c_{jl} < d_{jl} < \infty$ as a non-informative prior. Hence, the prior for $\boldsymbol{\beta}_j$ is given by

$$\pi_{1j}(\boldsymbol{\beta}_j) \propto \prod_{l=1}^{m} \frac{1}{(d_{jl} - c_{jl})}, \qquad -\infty < c_{jl} < d_{jl} < \infty. \tag{4.2}$$

Thus the joint prior distribution for a_j, α_j, λ_j and $\boldsymbol{\beta}_j, j = 1, 2, \ldots, p$ is given by

$$\pi_1(\boldsymbol{\Theta}) \propto \prod_{j=1}^{p} \frac{1}{M_{a_j} M_{\alpha_j} M_{\lambda_j}} \prod_{l=1}^{m} \frac{1}{(d_{jl} - c_{jl})}. \tag{4.3}$$

4.2. Jeffreys prior

A commonly used non-informative prior in Bayesian analysis is Jeffreys prior. Jeffreys prior exhibits nice features that make it an attractive non-informative prior. According to Jeffrey's rule [34] we choose the priors for baseline parameters i.e. if the domain of the parameters is on positive real line then log of the parameters is uniformly distributed. Hence, the priors for baseline parameters are defined accordingly to Jeffrey's rule and for

regression parameter uniform prior is considered as follows

$$\begin{aligned}
\pi_{2j}(a_j) &\propto \frac{1}{a_j}, \qquad 0 < a_j < \infty, \\
\pi_{2j}(\alpha_j) &\propto \frac{1}{\alpha_j}, \qquad 0 < \alpha_j < \infty, \\
\pi_{2j}(\lambda_j) &\propto \frac{1}{\lambda_j}, \qquad 0 < \lambda_j < \infty, \\
\pi_{2j}(\boldsymbol{\beta}_j) &\propto \prod_{l=1}^{m} \frac{1}{(d_{jl} - c_{jl})}, \qquad -\infty < c_{jl} < d_{jl} < \infty.
\end{aligned} \tag{4.4}$$

Therefore, the joint prior distribution of a_j, α_j, λ_j and $\boldsymbol{\beta}_j, j = 1, 2, \ldots, p$ is equivalent to

$$\pi_2(\boldsymbol{\Theta}) \propto \prod_{j=1}^{p} \frac{1}{a_j \alpha_j \lambda_j} \prod_{l=1}^{m} \frac{1}{(d_{jl} - c_{jl})}. \tag{4.5}$$

4.3. Half-t prior

Gelman [5] recommended half-t as a default non-informative prior with large and finite value of the variance (scale parameter) of the t distribution. This prior is more suitable in the situation when more information is required compare to uniform/Jeffreys prior because is not completely flat but nearly flat. When the prior distribution has sufficient information, the numerical approximation algorithm can be easily explore the target density, or posterior distribution. We use an independent and identical half-t prior for baseline parameters and a uniform prior for regression parameters as follows

$$\begin{aligned}
\pi_{3j}(a_j) &\propto \left(1 + \frac{1}{\nu}\left(\frac{a_j}{\sigma}\right)^2\right)^{-\left(\frac{\nu+1}{2}\right)}, \qquad 0 < a_j < \infty, \\
\pi_{3j}(\alpha_j) &\propto \left(1 + \frac{1}{\nu}\left(\frac{\alpha_j}{\sigma}\right)^2\right)^{-\left(\frac{\nu+1}{2}\right)}, \qquad 0 < \alpha_j < \infty, \\
\pi_{3j}(\lambda_j) &\propto \left(1 + \frac{1}{\nu}\left(\frac{\lambda_j}{\sigma}\right)^2\right)^{-\left(\frac{\nu+1}{2}\right)}, \qquad 0 < \lambda_j < \infty, \\
\pi_{3j}(\boldsymbol{\beta}_j) &\propto \prod_{l=1}^{m} \frac{1}{(d_{jl} - c_{jl})}, \qquad -\infty < c_{jl} < d_{jl} < \infty.
\end{aligned} \tag{4.6}$$

Hence, the joint prior distribution of a_j, α_j, λ_j and $\boldsymbol{\beta}_j, j = 1, 2, \ldots, p$ is given by

$$\pi_3(\boldsymbol{\Theta}) \propto \prod_{j=1}^{p} \left(\left(1 + \frac{1}{\nu}\left(\frac{a_j}{\sigma}\right)^2\right)\left(1 + \frac{1}{\nu}\left(\frac{\alpha_j}{\sigma}\right)^2\right)\left(1 + \frac{1}{\nu}\left(\frac{\lambda_j}{\sigma}\right)^2\right)\right)^{-\left(\frac{\nu+1}{2}\right)} \prod_{l=1}^{m} \frac{1}{(d_{jl} - c_{jl})}. \tag{4.7}$$

The parameters ν denotes the degree of freedom and $\sigma > 0$ is the scale parameter of the half-t distribution. From Figure 2, it is clear that at $\sigma = 25$ and $\nu = 4$ half-t becomes approximate to uniform.

4.4. Posterior analysis

The posterior probability distribution is obtained by combining past information with the observed sample using likelihood and prior distribution. Therefore, the joint posterior density of the random variables a_j, α_j, λ_j and $\boldsymbol{\beta}_j, j = 1, 2, \ldots, p$ given the data can be written as

$$p(\boldsymbol{\Theta}|\text{data}) = \frac{L(\text{data}|\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})}{\int\int\cdots\int L(\text{data}|\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})d\boldsymbol{\Theta}}, \tag{4.8}$$

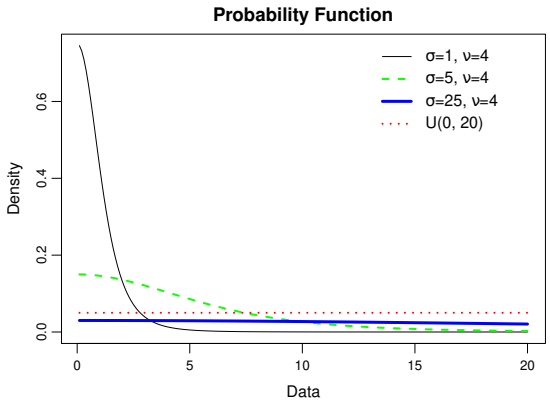

Figure 2. Half-t density plot.

where $p(\boldsymbol{\Theta}|\text{data})$ is the joint posterior density, $L(\text{data}|\boldsymbol{\Theta})$ is the likelihood function for the given observed data as in (3.2) and $\pi(\boldsymbol{\Theta})$ is the joint prior density. Under the joint priors $\pi_1(\boldsymbol{\Theta}), \pi_2(\boldsymbol{\Theta})$ and $\pi_3(\boldsymbol{\Theta})$ given in (4.3), (4.5) and (4.7), respectively, the joint posterior densities are obtained as

$$\begin{aligned}\pi_1(\boldsymbol{\Theta}|\text{data}) = K_1 \prod_{j=1}^{p} \left[\prod_{i=1}^{n_j} \left(a_j(\alpha_j + \lambda_j t_i) t_i^{\alpha_j - 1} e^{\lambda_j t_i} + \boldsymbol{\beta}_j^\top \boldsymbol{x}_i \right) \right] \\ \times \exp\left\{ - \sum_{i=1}^{n} \sum_{j=1}^{p} \left(a_j t_i^{\alpha_j} e^{\lambda_j t_i} + \boldsymbol{\beta}_j^\top \boldsymbol{x}_i t_i \right) \right\},\end{aligned} \tag{4.9}$$

$$\begin{aligned}\pi_2(\boldsymbol{\Theta}|\text{data}) = K_2 \prod_{j=1}^{p} \left[\frac{1}{a_j \alpha_j \lambda_j} \prod_{i=1}^{n_j} \left(a_j(\alpha_j + \lambda_j t_i) t_i^{\alpha_j - 1} e^{\lambda_j t_i} + \boldsymbol{\beta}_j^\top \boldsymbol{x}_i \right) \right] \\ \times \exp\left\{ - \sum_{i=1}^{n} \sum_{j=1}^{p} \left(a_j t_i^{\alpha_j} e^{\lambda_j t_i} + \boldsymbol{\beta}_j^\top \boldsymbol{x}_i t_i \right) \right\},\end{aligned} \tag{4.10}$$

$$\begin{aligned}\pi_3(\boldsymbol{\Theta}|\text{data}) = K_3 \prod_{j=1}^{p} \left[\prod_{i=1}^{n_j} \left(a_j(\alpha_j + \lambda_j t_i) t_i^{\alpha_j - 1} e^{\lambda_j t_i} + \boldsymbol{\beta}_j^\top \boldsymbol{x}_i \right) \right. \\ \left. \times \left(\left(1 + \frac{1}{\nu} \left(\frac{a_j}{\sigma} \right)^2 \right) \left(1 + \frac{1}{\nu} \left(\frac{\alpha_j}{\sigma} \right)^2 \right) \left(1 + \frac{1}{\nu} \left(\frac{\lambda_j}{\sigma} \right)^2 \right) \right)^{-\left(\frac{\nu+1}{2} \right)} \right] \\ \times \exp\left\{ - \sum_{i=1}^{n} \sum_{j=1}^{p} \left(a_j t_i^{\alpha_j} e^{\lambda_j t_i} + \boldsymbol{\beta}_j^\top \boldsymbol{x}_i t_i \right) \right\},\end{aligned} \tag{4.11}$$

where K_1, K_2 and K_3 are the normalizing constant or they are the denominator part in the right hand side of equation (4.8) according to each joint posterior distribution.

It is not possible to compute the integral in the denominator of (4.8) analytically under each considered prior due to the complex form of likelihood function. Therefore, we cannot obtain the posterior density in closed form. Hence, in such situation Markov Chain Mote Carlo (MCMC) method [30] can be used to approximate the integrals. Popularly used MCMC algorithms are Gibbs sampling algorithm [6] and Metropolis-Hastings (M-H) algorithm [10]. For the implementation of the Gibbs sampling algorithm the full

conditional distribution of each parameter is required. So, in this situation the M-H algorithm is preferable.

The first step in the M-H algorithm is to generate a candidate point, denoted here by $\boldsymbol{\Theta}^*$, from a proposal density $q(\boldsymbol{\Theta}^*|\boldsymbol{\Theta})$ also known as candidate density [33]. Let u be a random variable which is generated from a uniform distribution $\mathcal{U}(0,1)$. Then a general version of the M-H algorithm for sampling from the posterior distribution $\pi(\boldsymbol{\Theta}|\text{data})$ can be described as follows:

(1) Choose the initial values $\boldsymbol{\Theta}^{(0)}$ and set $i = 0$.
(2) For $i = 1, 2, \ldots, M$ repeat the following steps, where M is typically of the order of many thousands.
   (a) Set $\boldsymbol{\Theta} = \boldsymbol{\Theta}^{(i-1)}$.
   (b) Generate a candidate point $\boldsymbol{\Theta}^*$ from a proposal density $q(\boldsymbol{\Theta}^*|\boldsymbol{\Theta})$.
   (c) Generate u from $\mathcal{U}(0,1)$.
   (d) Calculate the acceptance probability

$$\alpha(\boldsymbol{\Theta}, \boldsymbol{\Theta}^*) = \min\left\{1, \frac{\pi(\boldsymbol{\Theta}^*|\text{data})q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^*)}{\pi(\boldsymbol{\Theta}|\text{data})q(\boldsymbol{\Theta}^*|\boldsymbol{\Theta})}\right\}.$$

   (e) Set

$$\boldsymbol{\Theta}^{(i+1)} = \begin{cases} \boldsymbol{\Theta}^* & \text{if } u \le \alpha(\boldsymbol{\Theta}^{(i)}, \boldsymbol{\Theta}^*) \\ \boldsymbol{\Theta}^{(i)} & \text{otherwise.} \end{cases}$$

The performance of the M-H algorithm depends on the choice of a proposal density $q(\cdot)$. Nevertheless, in practice, the choice of the proposal is essential since poor choices will considerably delay convergence towards the equilibrium distribution. Gibbs sampling is a particular case of the M-H algorithm using a specified form of a proposal density. Special, cases of M-H algorithm are random-walk Metropolis algorithm, independence sampler, or independent Metropolis algorithm [7, 22, 30]. As for applying these algorithms, we use `BUGS` software via `OpenBUGS` [22] interface, a free software and it is the most popular and has good documentation. For understanding the procedure of Bayesian analysis through `OpenBUGS` and its implementation in `R software` one may refer to (https://CRAN.R-project.org/package=R2OpenBUGS), (https://www.mrc-bsu.cam.ac.uk/software/bugs/openbugs/), and [22].

4.5. Loss function

The choice of loss function plays an important role in Bayesian computation. In this article, we consider two different types of loss functions, namely, squared error (symmetric) and LINEX (asymmetric) loss functions to compare Bayes estimates. Squared error loss function (SELF) for a parameter $\boldsymbol{\Theta}$ is defined as

$$L_1(\boldsymbol{\Theta}, \hat{\boldsymbol{\Theta}}) = (\boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}})^2.$$

Then the Bayes estimate for parameter $\boldsymbol{\Theta}$ under SELF can be obtained as the posterior means and calculated by

$$\hat{\boldsymbol{\Theta}}^{self} = \frac{1}{M - M^*} \sum_{l=M^*+1}^{M} [\boldsymbol{\Theta}]_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^{(l)}},$$

where $\boldsymbol{\Theta}^{(l)}, l = 1, 2, \ldots, M$ are the MCMC random samples generated from the posterior distribution of $\boldsymbol{\Theta}$ and M^* is the number of iteration used in burn-in period.

However, we also consider LINEX loss function (LLF) as an asymmetric loss function, which is given by

$$L_2(\boldsymbol{\Theta}, \hat{\boldsymbol{\Theta}}) = e^{\rho(\hat{\boldsymbol{\Theta}}-\boldsymbol{\Theta})} - \rho(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}) - 1, \quad \rho \neq 0.$$

Under LLF the Bayes estimates of parameter $\boldsymbol{\Theta}$ can be obtained as follows

$$\hat{\boldsymbol{\Theta}}^{llf} = -\frac{1}{\rho}\log\left(\frac{1}{M-M^*}\sum_{l=M^*+1}^{M} e^{-\rho[\boldsymbol{\Theta}]_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^{(l)}}}\right),$$

where ρ is the hyper parameter of the LLF and magnitude of ρ reflects the degree of asymmetry. For $\rho > 0$, the LLF is quite asymmetric about 0 with overestimation being more serious than underestimation. Estimates under LLF are roughly equal to estimates obtained under SELF if ρ is close to zero. LLF is more applicable in lifetime modeling because underestimating the survival and failure rate functions is typically much more harmful than overestimating [8].

4.6. Asymptotic confidence interval

In this subsection, we obtained the interval estimates of the unknown parameters based on the asymptotic property of MLE. Since the MLEs of the unknown parameters are not in closed form, therefore, it is not possible to obtained exact distribution of MLEs. Using the asymptotic property of MLE, the sampling distribution of $\hat{\boldsymbol{\Theta}}$ can be approximated by a $(2p + (p \times m))$ variate normal distribution with mean $\boldsymbol{\Theta}$ and variance covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\Theta})$, which is nothing but the inverse of the Fisher information matrix $\mathbf{I}(\boldsymbol{\Theta})$ and given by

$$\mathbf{I}(\boldsymbol{\Theta}) = \mathrm{E}\left[-\frac{\partial^2 \ell(\boldsymbol{\Theta})}{\partial(\boldsymbol{\Theta})\partial(\boldsymbol{\Theta})^\top}\right]_{\boldsymbol{\Theta}=\hat{\boldsymbol{\Theta}}}.$$

Since the exact mathematical expressions for the above expectations are difficult to obtain. Therefore, the observed Fisher information matrix $\mathbf{I}_O(\boldsymbol{\Theta})$ can be used to approximate Fisher information matrix $\mathbf{I}(\boldsymbol{\Theta})$, which is obtained by dropping the expectation operator E in $\mathbf{I}(\boldsymbol{\Theta})$. The variance of MLEs of the unknown parameters, i.e. $\mathrm{var}(\hat{\boldsymbol{\Theta}})$ are the diagonal elements of the asymptotic variance covariance matrix $\boldsymbol{\Sigma}(\hat{\boldsymbol{\Theta}})$. Thus, for a given confidence level $\gamma, 0 < \gamma < 1$, a two-sided $100(1-\gamma)\%$ asymptotic confidence interval (ACI) for $\hat{\boldsymbol{\Theta}}$ can be constructed as follows

$$\left[\hat{\boldsymbol{\Theta}} - z_{\gamma/2}\sqrt{\mathrm{var}(\hat{\boldsymbol{\Theta}})}, \hat{\boldsymbol{\Theta}} + z_{\gamma/2}\sqrt{\mathrm{var}(\hat{\boldsymbol{\Theta}})}\right],$$

where, $z_{\gamma/2}$ is the upper $\gamma/2$ quantile of the standard normal distribution. Further, we also computed two-sided $100(1-\gamma)\%$ confidence interval for the estimates of cumulative CSH $\hat{H}_j(t;\hat{\boldsymbol{\Theta}}_j,\boldsymbol{x})$, which is given by

$$\left[\hat{H}_j(t;\hat{\boldsymbol{\Theta}}_j,\boldsymbol{x}) - z_{\gamma/2}\sqrt{\mathrm{var}(\hat{H}_j(t;\hat{\boldsymbol{\Theta}}_j,\boldsymbol{x}))}, \hat{H}_j(t;\hat{\boldsymbol{\Theta}}_j,\boldsymbol{x}) + z_{\gamma/2}\sqrt{\mathrm{var}(\hat{H}_j(t;\hat{\boldsymbol{\Theta}}_j,\boldsymbol{x}))}\right],$$

where variance of cumulative CSH $\mathrm{var}(\hat{H}_j(t;\hat{\boldsymbol{\Theta}}_j,\boldsymbol{x}))$ is obtained by using the delta method as follows

$$\mathrm{var}(\hat{H}_j(t;\hat{\boldsymbol{\Theta}}_j,\boldsymbol{x})) = \left(\frac{\partial H_j(t;\boldsymbol{\Theta}_j,\boldsymbol{x})}{\partial\boldsymbol{\Theta}}\right)\Bigg|_{\boldsymbol{\Theta}=\hat{\boldsymbol{\Theta}}} \boldsymbol{\Sigma}(\hat{\boldsymbol{\Theta}}) \left(\frac{\partial H_j(t;\boldsymbol{\Theta}_j,\boldsymbol{x})}{\partial\boldsymbol{\Theta}}\right)^\top\Bigg|_{\boldsymbol{\Theta}=\hat{\boldsymbol{\Theta}}}.$$

4.7. Bayes credible interval

In Bayesian approach, for a γ level of significance, the $(1-\gamma)$ interval estimate of a parameter $\boldsymbol{\Theta}$ is a credible interval based on given data, that covers the parameter with $(1-\gamma)$ level of confidence. The $100(1-\gamma)\%$ Bayes credible interval (BCI) $[\boldsymbol{\Theta}_L, \boldsymbol{\Theta}_U]$ for $\boldsymbol{\Theta}$ is obtained by setting $\boldsymbol{\Theta}_L$ equal to the $\gamma/2\%$ quantile and $\boldsymbol{\Theta}_U$ equal to $(1-\gamma/2)\%$ quantile of $\boldsymbol{\Theta}_l, l = 1, 2, \ldots, M - M^*$. Similarly, same procedure also adopt for obtaining Bayes credible interval for $H_j(t;\boldsymbol{\Theta}_j,\boldsymbol{x})$.

Table 1. AE, MSE, AVL and CP values for MLE and Bayes estimates for sample size $n = 75$

Method		a_1	α_1	β_{11}	H_1	a_2	α_2	β_{21}	H_2
True value		0.5	0.8	0.5	0.4269	0.7	0.6	0.4	0.5855
MLE	AE	0.5241	0.8592	0.5385	0.4397	0.7260	0.6437	0.4166	0.5907
	MSE	0.0301	0.0502	0.1468	0.0107	0.0414	0.0132	0.1642	0.0114
ACI	AVL	0.6204	0.7305	1.4002	0.5473	0.7137	0.4347	1.5292	0.5577
	CP	0.9340	0.9460	0.9420	0.9920	0.9280	0.9600	0.9500	0.9880
U-self	AE	0.5331	0.9033	0.6614	0.4671	0.7148	0.6626	0.5988	0.6230
	MSE	0.0244	0.0524	0.1132	0.0119	0.0276	0.0163	0.1201	0.0131
U-llf1	AE	0.5168	0.8755	0.5893	0.4609	0.6952	0.6524	0.5234	0.6148
	MSE	0.0217	0.0427	0.0811	0.0111	0.0254	0.0143	0.0797	0.0122
U-llf2	AE	0.5507	0.9334	0.7426	0.4737	0.7356	0.6734	0.6864	0.6314
	MSE	0.0280	0.0649	0.1594	0.0129	0.0308	0.0187	0.1791	0.0143
U-BCI	AVL	0.5717	0.7397	1.1951	0.3561	0.6305	0.4535	1.2120	0.4064
	CP	0.9620	0.9320	0.9480	0.9080	0.9440	0.9560	0.9620	0.9360
J-self	AE	0.4843	0.8580	0.6946	0.4561	0.6689	0.6341	0.6307	0.6088
	MSE	0.0220	0.0503	0.1263	0.0108	0.0270	0.0127	0.1375	0.0118
J-llf1	AE	0.4693	0.8281	0.6219	0.4500	0.6504	0.6240	0.5538	0.6009
	MSE	0.0210	0.0397	0.0902	0.0101	0.0266	0.0114	0.0922	0.0111
J-llf2	AE	0.5005	0.8922	0.7761	0.4625	0.6887	0.6446	0.7190	0.6170
	MSE	0.0237	0.0673	0.1770	0.0115	0.0284	0.0144	0.2014	0.0127
J-BCI	AVL	0.5465	0.7606	1.2068	0.3518	0.6117	0.4489	1.2285	0.4015
	CP	0.9380	0.9300	0.9420	0.9140	0.9280	0.9700	0.9480	0.9440
HT-self	AE	0.5324	0.9108	0.6648	0.4671	0.7163	0.6645	0.6004	0.6238
	MSE	0.0245	0.0628	0.1143	0.0120	0.0276	0.0167	0.1202	0.0132
HT-llf1	AE	0.5161	0.8805	0.5930	0.4608	0.6964	0.6541	0.5246	0.6156
	MSE	0.0218	0.0485	0.0819	0.0112	0.0253	0.0146	0.0794	0.0122
HT-llf2	AE	0.5501	0.9453	0.7454	0.4736	0.7377	0.6755	0.6888	0.6323
	MSE	0.0282	0.0840	0.1608	0.0129	0.0312	0.0193	0.1798	0.0144
HT-BCI	AVL	0.5726	0.7654	1.1927	0.3559	0.6354	0.4557	1.2175	0.4081
	CP	0.9540	0.9260	0.9460	0.8960	0.9540	0.9480	0.9600	0.9420

5. Numerical illustration

We conducted a Monte Carlo simulation study to observe the finite sample behaviour of the proposed estimators of the unknown parameters and cumulative CSH functions. For simplicity, we considered two causes of failure i.e. $j = 1, 2$ and one covariate, say x, which is generated using a Bernoulli random number for each sample with equal probability of success and failure. The survival time T is generated through inverse transformation following the steps given in [3]. For each simulated survival time, the causes of failure are generated from Binomial distribution with probability of success $\frac{h_1(t;\boldsymbol{\Theta}_j,\boldsymbol{x})}{h_1(t;\boldsymbol{\Theta}_j,\boldsymbol{x})+h_2(t;\boldsymbol{\Theta}_j,\boldsymbol{x})}$ for cause 1 and failure outcome is considered as cause 2.

In this simulation study, the data set are generated for various sample sizes such as $n = 75, 150$ and 300. Without loss of generality we have arbitrary taken the true value of the parameters as $a_1 = 0.5, \alpha_1 = 0.8, \lambda_1 = 0.1, \beta_{11} = 0.5, a_2 = 0.7, \alpha_2 = 0.6, \lambda_2 = 0.1, \beta_{21} = 0.4$. We assume that λ_j to be known for mathematical simplicity. The censored time D is generated from $U(0, d)$, where d is chosen in such a way that on an average 20% observations are right censored. Under this setting we had approximately 34% and 46% failures from cause 1 and cause 2, respectively. For each sample size we have calculated the

Table 2. AE, MSE, AVL and CP values for MLE and Bayes estimates for sample size $n = 150$

Method		a_1	α_1	β_{11}	H_1	a_2	α_2	β_{21}	H_2
True value		0.5	0.8	0.5	0.4269	0.7	0.6	0.4	0.5855
MLE	AE	0.5151	0.8237	0.5030	0.4317	0.7223	0.6315	0.3897	0.5867
	MSE	0.0147	0.0170	0.0666	0.0040	0.0174	0.0064	0.0754	0.0053
ACI	AVL	0.4318	0.4784	0.9658	0.3661	0.4971	0.2963	1.0517	0.3847
	CP	0.9220	0.9480	0.9400	0.9960	0.9400	0.9540	0.9500	0.9840
U-self	AE	0.5206	0.8484	0.5675	0.4463	0.7171	0.6393	0.4884	0.6050
	MSE	0.0129	0.0196	0.0581	0.0044	0.0129	0.0070	0.0576	0.0060
U-llf1	AE	0.5122	0.8362	0.5273	0.4433	0.7069	0.6348	0.4461	0.6010
	MSE	0.0121	0.0175	0.0491	0.0042	0.0122	0.0065	0.0449	0.0057
U-llf2	AE	0.5294	0.8611	0.6113	0.4495	0.7277	0.6438	0.5357	0.6091
	MSE	0.0138	0.0222	0.0714	0.0046	0.0138	0.0075	0.0759	0.0063
U-BCI	AVL	0.4121	0.4937	0.9019	0.2499	0.4571	0.3004	0.9122	0.2860
	CP	0.9300	0.9360	0.9480	0.9480	0.9460	0.9460	0.9660	0.9300
J-self	AE	0.4957	0.8250	0.5866	0.4409	0.6928	0.6260	0.5087	0.5979
	MSE	0.0121	0.0176	0.0616	0.0041	0.0124	0.0061	0.0631	0.0057
J-llf1	AE	0.4877	0.8129	0.5464	0.4379	0.6829	0.6216	0.4656	0.5940
	MSE	0.0118	0.0161	0.0513	0.0039	0.0122	0.0057	0.0487	0.0055
J-llf2	AE	0.5040	0.8378	0.6302	0.4440	0.7031	0.6305	0.5566	0.6019
	MSE	0.0126	0.0196	0.0763	0.0043	0.0128	0.0065	0.0835	0.0059
J-BCI	AVL	0.4020	0.4935	0.9040	0.2485	0.4489	0.2995	0.9217	0.2841
	CP	0.9180	0.9440	0.9420	0.9620	0.9400	0.9520	0.9620	0.9340
HT-self	AE	0.5192	0.8500	0.5714	0.4463	0.7177	0.6402	0.4885	0.6052
	MSE	0.0127	0.0198	0.0587	0.0044	0.0129	0.0071	0.0575	0.0060
HT-llf1	AE	0.5109	0.8376	0.5313	0.4432	0.7074	0.6357	0.4462	0.6012
	MSE	0.0120	0.0176	0.0494	0.0042	0.0122	0.0066	0.0448	0.0057
HT-llf2	AE	0.5279	0.8629	0.6147	0.4494	0.7284	0.6448	0.5358	0.6093
	MSE	0.0137	0.0225	0.0723	0.0046	0.0139	0.0076	0.0759	0.0063
HT-BCI	AVL	0.4101	0.4972	0.8996	0.2486	0.4585	0.3020	0.9136	0.2861
	CP	0.9200	0.9420	0.9440	0.9460	0.9440	0.9400	0.9680	0.9380

average estimate (AE) and mean square error (MSE) for point estimates and average length (AVL) and coverage probability (CP) for interval estimates of a_j, α_j, β_j and $H_j(t; \boldsymbol{\Theta}_j, \boldsymbol{x})$ over 500 replications. The estimates of $H_j(t; \boldsymbol{\Theta}_j, \boldsymbol{x})$ for $j = 1, 2$ are obtained at $t = 0.5$ with covariates values $x = 0.5$ and denoted as H_1 and H_2 with true values $H_1 = 0.4269$ and $H_2 = 0.5855$.

Next, as we mentioned in Sub-Section 4.4 that the conditional posterior densities of the unknown parameters are not turn out in known distributional form. So, we employed the MCMC procedure for generating the random samples from conditional posteriors. For this purpose we used the `BUGS software` via `R2OpenBUGS` package in `R software` [22]. The non-informative priors often lead to a class of improper priors [34]. The use of a proper prior ensures that posterior will be a proper density, but using an improper prior does not guarantee that posterior will be a proper density. Thus, with the use of improper prior, it is necessary to ensure that the resulting posterior density will be proper. Although we are using the BUGS software for Bayesian analysis, which requires a valid probability distribution to define the prior as a result, we will get a proper posterior density. In BUGS, the Jeffreys prior is defined using an appropriate distribution approximation such as the gamma distribution [22].

Table 3. AE, MSE, AVL and CP values for MLE and Bayes estimates for sample size $n = 300$

Method		a_1	α_1	β_{11}	H_1	a_2	α_2	β_{21}	H_2
True value		0.5	0.8	0.5	0.4269	0.7	0.6	0.4	0.5855
MLE	AE	0.5129	0.8158	0.4970	0.4305	0.7208	0.6212	0.3750	0.5859
	MSE	0.0067	0.0070	0.0301	0.0020	0.0101	0.0031	0.0400	0.0028
ACI	AVL	0.3038	0.3307	0.6751	0.2554	0.3494	0.2050	0.7339	0.2679
	CP	0.9460	0.9620	0.9560	0.9920	0.9200	0.9440	0.9320	0.9820
U-self	AE	0.5194	0.8280	0.5208	0.4378	0.7207	0.6254	0.4195	0.5952
	MSE	0.0067	0.0075	0.0289	0.0021	0.0084	0.0032	0.0295	0.0030
U-llf1	AE	0.5150	0.8224	0.4994	0.4363	0.7152	0.6234	0.3966	0.5932
	MSE	0.0064	0.0070	0.0273	0.0021	0.0081	0.0031	0.0270	0.0029
U-llf2	AE	0.5239	0.8336	0.5430	0.4393	0.7262	0.6275	0.4435	0.5972
	MSE	0.0070	0.0080	0.0316	0.0022	0.0088	0.0033	0.0332	0.0030
U-BCI	AVL	0.2994	0.3359	0.6621	0.1744	0.3328	0.2052	0.6727	0.1998
	CP	0.9480	0.9560	0.9560	0.9460	0.9240	0.9300	0.9520	0.9460
J-self	AE	0.5062	0.8162	0.5318	0.4351	0.7083	0.6191	0.4313	0.5917
	MSE	0.0062	0.0068	0.0294	0.0021	0.0080	0.0029	0.0310	0.0029
J-llf1	AE	0.5019	0.8106	0.5105	0.4336	0.7029	0.6170	0.4081	0.5898
	MSE	0.0060	0.0065	0.0274	0.0020	0.0078	0.0028	0.0279	0.0029
J-llf2	AE	0.5106	0.8218	0.5540	0.4366	0.7137	0.6212	0.4556	0.5937
	MSE	0.0064	0.0072	0.0325	0.0021	0.0083	0.0030	0.0353	0.0030
J-BCI	AVL	0.2958	0.3351	0.6613	0.1740	0.3302	0.2056	0.6784	0.1995
	CP	0.9500	0.9680	0.9560	0.9540	0.9360	0.9500	0.9480	0.9520
HT-self	AE	0.5180	0.8285	0.5245	0.4378	0.7211	0.6261	0.4184	0.5950
	MSE	0.0066	0.0075	0.0291	0.0021	0.0084	0.0032	0.0293	0.0030
HT-llf1	AE	0.5136	0.8229	0.5031	0.4364	0.7156	0.6240	0.3955	0.5930
	MSE	0.0063	0.0070	0.0273	0.0021	0.0081	0.0031	0.0268	0.0029
HT-llf2	AE	0.5225	0.8342	0.5466	0.4393	0.7267	0.6282	0.4426	0.5970
	MSE	0.0069	0.0081	0.0319	0.0022	0.0089	0.0034	0.0330	0.0030
HT-BCI	AVL	0.2984	0.3364	0.6609	0.1736	0.3342	0.2069	0.6744	0.1999
	CP	0.9480	0.9620	0.9560	0.9500	0.9260	0.9360	0.9500	0.9520

We generated $M = 10000$ Markov chains for each parameter, and the first $M^* = 4000$ samples are used in the burn-in period to reduce the effect of initial values. Furthermore, for minimizing the effect of the autocorrelation, every second equally spaced outcome is considered i.e., thin=2. By visualizing the convergence diagnostics plots, it is realized that chains converge nicely. Therefore, the last 6000 MCMC samples are used to obtain the Bayes estimates of $a_1, \alpha_1, \beta_{11}, H_1, a_2, \alpha_2, \beta_{21}$ and H_2 under both the loss functions. The Bayes estimates for a uniform prior are denoted as U-self: Bayes estimates under SELF, U-llf1: Bayes estimates under LLF at $\rho = 1.5$, U-llf2: Bayes estimates under LLF at $\rho = -1.5$. Similarly, for Jeffreys and half-t are denoted as J-self, J-llf1, J-llf2 and HT-self, HT-llf1, HT-llf2, respectively. The BCI for uniform, Jeffreys and half-t priors are denoted as U-BCI, J-BCI and HT-BCI, respectively. The numerical results of the simulation study are presented in Tables 1-3.

From Tables 1-3 it is clear that as the sample size increases, MSEs decrease for MLE and Bayes estimates which verifies the consistency property of all the estimators. It is also noticed that the AVLs for ACI and BCIs are decreasing and CPs maintain the nominal level (95%). From Table 1, we observed that for sample size 75, the Bayes estimates based on Jeffreys prior are better than the MLE and Bayes estimate based on uniform and half-t

priors under both the loss functions except for β_{11} and β_{21}. For large sample sizes 150 and 300 in Tables 2-3, MLE gets better except for some cases. MLE of the cumulative CSH functions are dominating over the Bayes estimates for sample sizes 75, 150 and 300. However, half-t prior work well under LINEX loss function at $\rho = 1.5$ compare to MLE. For large sample size it is observed that the magnitude of the MSE is negligible for the MLE and Bayes estimates. Among the priors it is observed that Jeffreys prior is performing well, however, in some cases uniform and half-t priors shows their applicability. It is also noticed that the performance of LINEX loss function at $\rho = 1.5$ relatively good compared to SELF and LINEX $\rho = -1.5$ corresponding to each prior. From Tables 1-3 it is noticed that the interval estimates, BCI under uniform, Jeffreys and half-t prior are dominating over ACI except for α_1 and α_2 in terms of AVL.

6. Application

In this section, we used real data from a Mayo Clinic trial in PBC of the liver conducted between 1974 and 1984 to demonstrate the applicability of the proposed model. This data set is available in *survival* package of `R software`. During this ten-year period, 312 patients were randomly assigned to receive D-penicillamine or placebo treatment from a total of 424 patients. Furthermore, the remaining 112 patients did not take part in the clinical trial but agreed to have their basic measurements taken and to be observed for survival. Six of those patients were lost to follow-up shortly after diagnosis, so these patients were removed from the study.

In the end of the study, 161 patients died, another 25 patients were received liver transplant and 232 patients were lost to follow-up. Therefore, the competing risks model becomes more reasonable for two competing outcome variables liver transplant and death. The survival time is measured in days for all individuals. Although there are several covariates in the original data such as treatment, sex, age, etc. For the analysis purpose treatment is considered as a covariate. For more information on PBC data one could refer to [36] and application of competing risk on PBC data is available in [16]. Recently, Bayes estimates based on informative priors for PBC data using a modified Weibull AH regression model with competing risks are obtained in [28].

In order to compute survival time in terms of years, we divided it by 365, which has a median survival time of 4.74 years. We also assume that 106 patients who did not participate in the trial they received the D-penicillamine treatment. Further, we apply the proposed estimation methods to obtain the estimates of unknown parameters and cumulative CSH functions. The results of the estimates of unknown parameters are presented in Table 4. Figure 3 depicts diagnostic plots of the 6000 posterior samples of $a_1, \alpha_1, \beta_{11}, a_2, \alpha_2$, and β_{21}

Table 4. Estimates of the unknown parameters

Method	a_1	α_1	λ_1	β_{11}	a_2	α_2	λ_2	β_{21}
MLE	0.00159	1.36052	0.12696	0.00481	0.06098	0.86681	0.05123	0.00588
U-self	0.00713	0.78394	0.10336	0.00454	0.06215	0.88388	0.04189	0.01178
U-llf1	0.00712	0.7546	0.10083	0.00453	0.06207	0.87867	0.04153	0.01173
U-llf2	0.00713	0.80735	0.10601	0.00454	0.06223	0.88876	0.04225	0.01183
J-self	0.00447	1.29811	0.04431	0.00469	0.05853	1.06259	0.00677	0.0113
J-llf1	0.00446	0.96344	0.03928	0.00468	0.05843	1.05539	0.00663	0.01126
J-llf2	0.00448	1.48471	0.05018	0.00469	0.05864	1.06953	0.00691	0.01135
HT-self	0.00446	1.27911	0.05193	0.00414	0.0586	0.94879	0.03331	0.01184
HT-llf1	0.00446	1.18592	0.05029	0.00413	0.05851	0.93932	0.03295	0.01179
HT-llf2	0.00447	1.36911	0.0537	0.00414	0.05869	0.95804	0.03368	0.01189

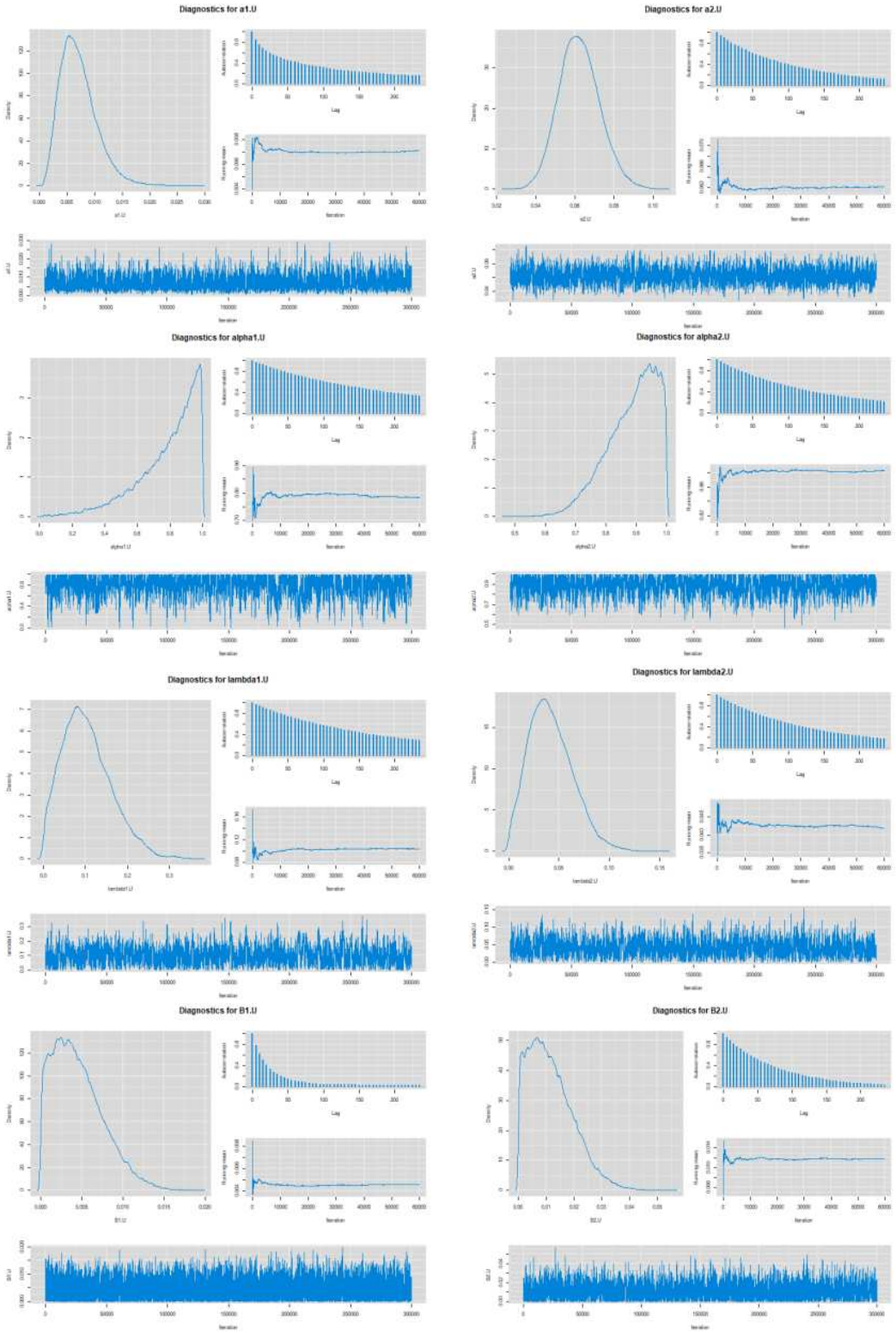

Figure 3. MCMC diagnostic plots of the Bayes estimates under uniform prior, where B1.U and B2.U are cause-specific regression parameters.

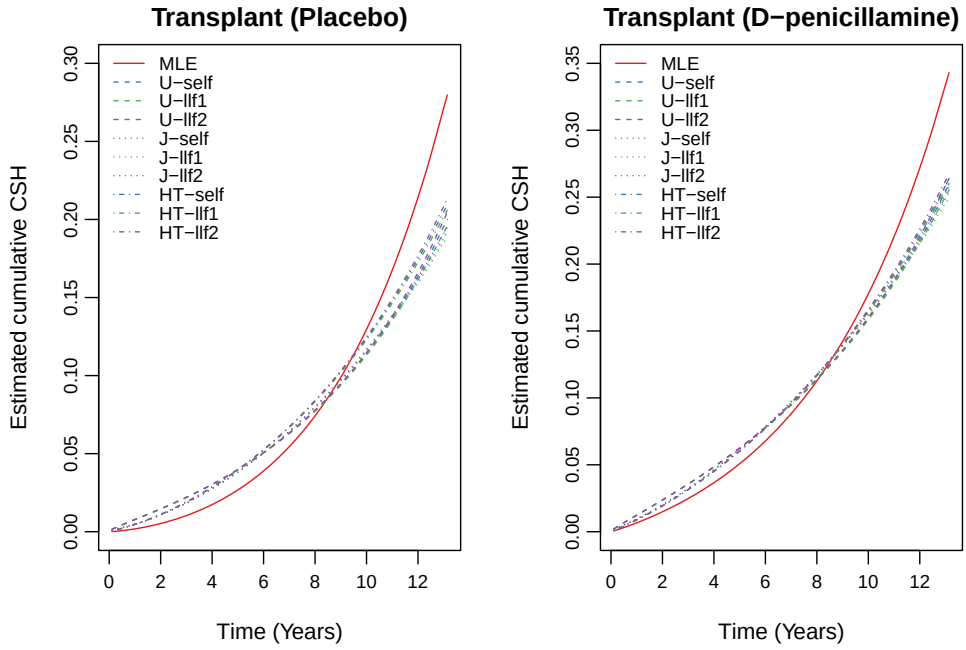

Figure 4. Cumulative CSH plot for transplant.

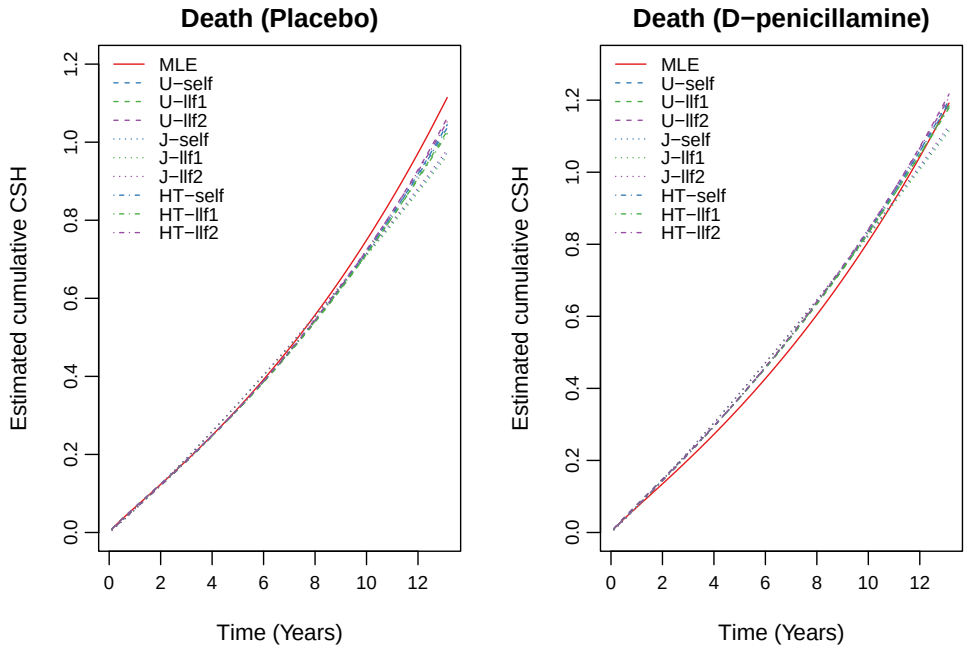

Figure 5. Cumulative CSH plot for death.

under uniform prior generated from MCMC algorithm using `BUGS software`. The model parameters appear to be convergent based on the trace and density plots, as seen in Figure 3 where the auto-correlation for each parameter is decreasing with increasing values of lag. Although similar results for posterior samples under Jeffreys and half-t priors can also be obtained.

We estimates the cumulative CSH functions using (2.9) based on the proposed estimators which are presented in Figures 4 and 5. These plots indicates that cumulative CSH rate for transplant patients is small compared to the same for the patients who experienced the death. Figure 4 shows that value of the cumulative CSH function due to transplant is small for the patients who received placebo treatment. Similarly, the same is observed for cumulative CSH rate due to death see Figure 5. Moreover, the likelihood test procedure is used to test the significance of the covariate effect on transplant and death separately. The hypotheses of interest are $H_0 : \beta_{11} = 0$ against $H_1 : \beta_{11} \neq 0$ and $H_0 : \beta_{21} = 0$ against $H_1 : \beta_{21} \neq 0$. We calculated the likelihood ratio test statistics and corresponding $p - values$ are 0.016 and 0.008, respectively. This indicates that treatment has a significant effect on transplant and death.

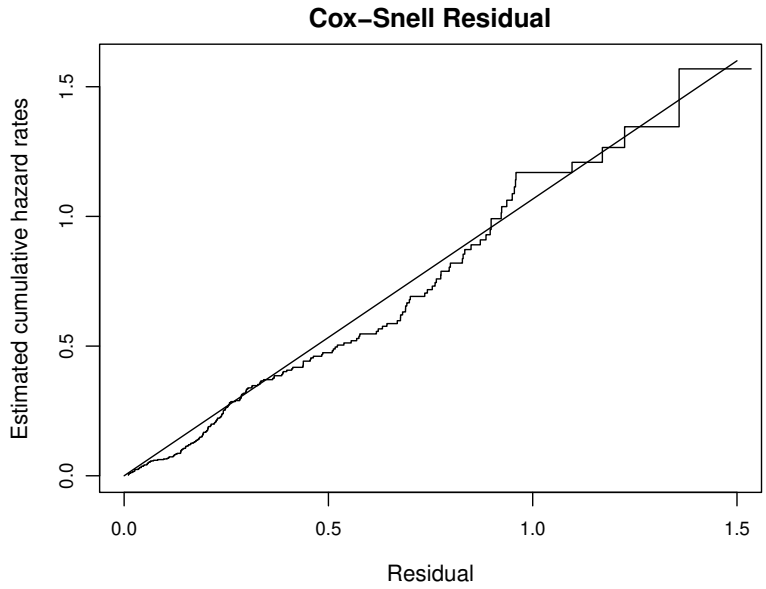

Figure 6. Plot of Cox-Snell residual versus its estimates of cumulative hazard rate.

To test the goodness of fit of the model (2.5) to the PBC data in competing risks framework we use the Cox-Snell residual plot [14]. The Cox-Snell residual is defined as

$$r_i = \hat{H}(t_i|x_i), \qquad i = 1, 2, \ldots, n, \tag{6.1}$$

where $\hat{H}(t|x)$ is the estimator of cumulative CSH rate $H(t|x) = \sum_{j=1}^{2} H_j(t|x)$ and $j = 1, 2$ for transplant and death, respectively. If the model holds, then these residuals should be a sample from unit exponential distribution. Therefore, the hazard plot of residuals versus the Nelson-Aalen estimator of the cumulative hazard of the residuals will be a straight line with slope one. From Figure 6 it is clearly observed that fit is reasonably good.

7. Conclusion

In this article, we considered the parametric cause-specific AH regression analysis through MWD for competing risks survival data. AH model is a good alternative of the Cox PH model and it is useful when excess risk is of concern. The estimation of the unknown parameters and cumulative CSH function have been dealt through MLE and Bayes estimates. In addition to Bayes estimation, we proposed three types of non-informative priors for baseline parameters and uniform priors are considered for regression parameters. Monte Carlo simulations are used to evaluate the relative performance of the proposed estimators using various sample sizes. The results of simulations showed that the proposed estimators perform admirably. We demonstrated the model utility with the PBC data. This data fits well to the model and the covariate have significant effect on transplant and death. The proposed work can be extent for different censoring schemes such as interval, current status and middle censoring schemes. Furthermore, the situation of masking in competing risks analysis is widespread, therefore, the analysis of masked competing risks data using proposed model seems an interesting attempt.

References

[1] O.O. Aalen, *A linear regression model for the analysis of life times*, Stat. Med. **8** (8), 907-925, 1989.

[2] S. Anjana and P. Sankaran, *Parametric analysis of lifetime data with multiple causes of failure using cause-specific reversed hazard rates*, Calcutta Stat. Assoc. Bull. **67** (3-4), 129-142, 2015.

[3] J. Beyersmann, A. Allignol and M. Schumacher, *Competing Risks and Multistate Models with R*, Springer Science & Business Media, New York, 2012.

[4] D.R. Cox, *Regression models and life-tables*, J. R. Stat. Soc. Ser. B (Methodol.) **34**(2), 187-220, 1972.

[5] A. Gelman, *Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)*, Bayesian Anal. **1** (3) 515-534, 2006.

[6] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-**6** (6), 721-741, 1984.

[7] W.R. Gilks, R. Sylvia and S. David, *Markov Chain Monte Carlo in Practice*, CRC press, Boca Raton, 1996

[8] C.B. Guure and N.A. Ibrahim, *Bayesian analysis of the survival function and failure rate of Weibull distribution with censored data*, Math. Probl. Eng. 2012.

[9] B. Haller, G. Schmidt and K. Ulm, *Applying competing risks regression models: an overview*, Lifetime Data Anal. **19** (1), 33-58, 2013.

[10] W.K. Hastings, *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika **57** (1), 97-109, 1970.

[11] J.H. Jeong and J. Fine, *Direct parametric inference for the cumulative incidence function*, J. R. Stat. Soc. Ser. C (Applied Stat.) **55** (2), 187-200, 2006.

[12] H. Jiang, M. Xie and L. Tang, *Markov chain monte Carlo methods for parameter estimation of the modified Weibull distribution* , J. Appl. Stat. **35** (6), 647-658, 2008.

[13] J.D. Kalbeisch and R.L. Prentice, *The Statistical Analysis of Failure Time Data*, volume 360, John Wiley & Sons, New Jersey, 2002.

[14] J.P. Klein and M.L. Moeschberger, *Survival Analysis: Techniques for Censored and Truncated Data*, Springer-Verlag, New York, 2003.

[15] C. Lai, M. Xie and D. Murthy, *A modified Weibull distribution*, IEEE Trans. Reliab. **52** (1), 33-37, 2003.

[16] X. Lai, K.K. Yau and L. Liu, *Competing risk model with bivariate random eects for clustered survival data*, Comput. Stat. Data Anal. **112**, 215-223, 2017.

[17] J.F. Lawless, *Statistical Models and Methods for Lifetime Data*, volume 362, John Wiley & Sons, New Jersey, 2003.
[18] M. Lee, *Parametric inference for quantile event times with adjustment for covariates on competing risks data*, J. Appl. Stat. **46** (12), 2128-2144, 2019.
[19] W. Li, X. Xue and Y. Long, *Long. An additive subdistribution hazard model for competing risks data*, Commun. Stat. - Theory Methods **46** (23), 11667-11687, 2017.
[20] D. Lin and Z. Ying, *Semiparametric analysis of general additive-multiplicative hazard models for counting processes*, Ann. Stat. **23** (5), 1712-1734, 1995.
[21] D. Lin and Z. Ying, *Semiparametric analysis of the additive risk model*, Biometrika **81** (1), 61-71, 1994.
[22] D. Lunn, C. Jackson, N. Best, D. Spiegelhalter and A. Thomas, *The BUGS book: A Practical Introduction to Bayesian Analysis*, Chapman and Hall/CRC, Boca Raton, 2012.
[23] J.A. Nelder and R. Mead, *A simplex method for function minimization*, Comput. J. **7** (4), 308-313, 1965.
[24] H.K.T. Ng, *Parameter estimation for a modified Weibull distribution, for progressively type-II censored samples*, IEEE Trans. Reliab. **54** (3), 374-380, 2005.
[25] M. Pintilie, *Competing Risks: A Practical Perspective*, volume 58, John Wiley & Sons, England, 2006.
[26] N. Porta Bleda, G. Gómez Melis and M.L. Calle Rosingana, *The role of survival functions in competing risks*, Technical report, Universitat Politùcnica de Catalunya, 2008.
[27] R.L. Prentice, J.D. Kalbeisch, A.V. Peterson Jr, N. Flournoy, V.T. Farewell and N.E. Breslow, *The analysis of failure times in the presence of competing risks*, Biometrics **34** (4), 541–554, 1978.
[28] H. Rehman, N. Chandra, T. Emura and M. Pandey, *Estimation of the modified Weibull additive hazards regression model under competing risks*, Symmetry, **15** (485), 2023, https://doi.org/10.3390/sym15020485.
[29] H. Rehman, N. Chandra, F.S. Hosseini-Baharanchi, A.R. Baghestani and M.A. Pourhoseingholi, *Cause-specific hazard regression estimation for modified Weibull distribution under a class of non-informative priors*, J. Appl. Stat. **49** (7), 1784–1801, 2022.
[30] C.P. Robert, G. Casella and G. Casella, *Introducing Monte Carlo methods with R*, volume 18, Springer Science & Business Media, New York, 2010.
[31] P. Sankaran and S. Prasad, *Additive risks regression model for middle censored exponentiated-exponential lifetime data*, Commun. Stat. Simul. Comput. **47** (7), 1963-1974, 2018.
[32] Y. Shen and S. Cheng, *Confidence bands for cumulative incidence curves under the additive risk model*, Biometrics 55(4), 1093-1100, 1999.
[33] C. Siddhartha and G. Edward, *Understanding the metropolis-hastings algorithm*, Stat. Methods Appt. **49** (4), 327-335, 1995.
[34] S. Sinha, *Bayesian Estimation*, New Age International (P) Limited Publisher, New Delhi, 1998.
[35] J. Sun, L. Sun and N. Flournoy, *Additive hazards model for competing risks analysis of the case-cohort design*, Commun. Stat. - Theory Methods **33** (2), 351-366, 2004.
[36] T.M. Therneau and P.M. Grambsch, *Modeling Survival Data: Extending the Cox Model*, Springer Science & Business Media, New York, 2000.
[37] S. Upadhyay and A. Gupta, *A Bayes analysis of modified Weibull distribution via Markov chain Monte Carlo simulation*, J. Stat. Comput. Simul. **80** (3), 241-254, 2010.
[38] X. Zhang, H. Akcin and H.J. Lim, *Regression analysis of competing risks data via semi-parametric additive hazard model*, Stat. Methods Appt. **20** (3), 357-381, 2011.