

# Comparison of Accelerated Failure Time Models: A Bayesian Study on Head and Neck Cancer Data

Md. Ashraf-Ul-Alam\* and Athar Ali Khan

Department of Statistics and Operations Research, Aligarh Muslim University, Aligarh-202 002, India

Received: 3 Feb. 2020, Revised: 30 Mar. 2020, Accepted: 18 Apr. 2020

Published online: 1 Nov. 2021

---

**Abstract:** Comparison of treatments is a frequently used phenomenon in clinical studies. Accelerated failure time (AFT) models that express the relationship between logarithm of survival time and covariates are used for such type of comparison. Three log-location-scale models- Weibull, log-normal and log-logistic are evaluated to compare two treatment procedures of head-and-neck cancer data. Censored data are analyzed under Bayesian framework using Stan language. The models are assessed on the basis of LOOIC and WAIC.

**Keywords:** Accelerated failure time, log-location-scale, Head-and-neck cancer, Stan, LOOIC, WAIC

---

## 1 Introduction

In survival analysis main response variable is the time between a well defined origin and an event. Comparison of treatments is frequently made in clinical studies. Researchers in this arena are interested to know whether a new treatment procedure prolongs the survival process more than that of an existing standard treatment procedure. Accelerated failure time models and proportional hazards models are mostly used for comparing treatments. Proportional hazards (PH) model, proposed by [1] is a popular choice by the researchers for analyzing survival data. In PH models, the main assumption is that the hazard rate of an individual is proportional to the hazard rate of another individual. Logarithm of hazard ratio does not depend on time and, as such, no parametric model is required for survival times. Under proportionality hazards assumption, logarithm of hazard rate is expressed in terms of linear combination of a number of potential covariates and the effects of covariates are measured in terms of hazards.

Accelerated failure time(AFT) model is considered as an alternative to Cox's proportional hazards model that study the effect of the covariates on the time to event [2]. AFT models do not require the assumption of proportionality. [1] mentioned accelerated life tests in his famous article 'Regression models and life-tables'. AFT models establish relationship between logarithm of survival time and covariates and have become popular among the survival analysis researchers because of its easy interpretation in terms of lifetime. Professor Nancy Reid of University of Toronto had a long conversation with Sir D. R. Cox on October 26 and 27, 1993. During the conversation, comparing PH and AFT models, Cox mentioned, "that accelerated lifetime models are in many ways more appealing because of their direct physical interpretation" [3].

In this paper, we discuss the head-and-neck cancer data with Bayesian modelling of Weibull, log-normal and log-logistic distributions as survival models; Stan language is used for analysis, treatments are compared with estimated survival and hazard curves and finally fitted models are assessed based on LOOIC and WAIC.

---

\* Corresponding author e-mail: [mdashraful@gmail.com](mailto:mdashraful@gmail.com)

## 2 Accelerated Failure Time Models

The accelerated failure time models are a general class of models for survival data that assume the covariates to act multiplicatively on survival time and additively on logarithm of survival time. Thus the covariates ‘speed up’ or ‘slow down’ the survival process of a patient and the death will occur faster or slower [4]. Accelerated failure time models are parametric in nature which include many survival distributions. For example, gamma, inverse Gaussian, log-logistic, log-normal, Weibull and many others as well as their extensions or generalized forms are used as accelerated failure time models [5].

The AFT model for survival time assumes that the relationship of logarithm of survival time  $T$  and the covariates is linear [4,6] and can be written as:

$$\log(T) = \beta_0 + \sum_{j=1}^p \beta_j x_j + \sigma \varepsilon = \mathbf{x}'\beta + \sigma \mathbf{Z} \quad (1)$$

where,  $x_j, j = 1, 2, \dots, p$  are the covariates,  $\beta_j, j = 0, 1, 2, \dots, p$  are the regression coefficients,  $\sigma (> 0)$  is a scale parameter, and random error  $Z$  has a specified probability distribution. Exponentiation of both sides of Equation (1) leads to the following model:

$$T = \exp(\beta_0 + \sigma Z) \exp\left(\sum_{j=1}^p \beta_j x_j\right) = T_0 \exp\left(\sum_{j=1}^p \beta_j x_j\right) \quad (2)$$

Models (1) and (2) indicate that the covariates act multiplicatively on survival time and additively on logarithm of time. The term  $\exp(\sum_{j=1}^p \beta_j x_j)$  is known as time ratio or acceleration factor indicating that the role of the covariates is to accelerate or to decelerate time to failure. Thus, the model is referred to as the accelerated failure time (AFT) model. If  $\beta_j > 0$ , consequently,  $\exp(\sum_{j=1}^p \beta_j x_j) > 1$ , the covariate  $x_j$  decelerates the survival process and if  $\beta_j < 0$ , consequently,  $\exp(\sum_{j=1}^p \beta_j x_j) < 1$ , the covariate accelerates it [5,6].

### Location-scale models

An accelerated failure time model of survival time  $T$  is also known as a log-location-scale model as the distribution of  $Y = \log(T)$  is a location-scale model. A random variable  $Y$  is said to have a location-scale distribution if its probability density function (pdf)  $f(y)$ , cumulative distribution function (cdf)  $F(y)$  and survival function  $S(y)$  have the following form [5]:

$$\left. \begin{aligned} f(y|\mu, \sigma) &= \frac{1}{\sigma} g\left(\frac{y-\mu}{\sigma}\right) \\ F(y|\mu, \sigma) &= G\left(\frac{y-\mu}{\sigma}\right) \\ S(y|\mu, \sigma) &= S\left(\frac{y-\mu}{\sigma}\right) \end{aligned} \right\} \quad (3)$$

where  $t > 0$ ,  $-\infty < y < \infty$ ,  $\mu$  ( $-\infty < \mu < \infty$ ) is a location parameter,  $\sigma (> 0)$  is a scale parameter,  $g(z)$  and  $G(z)$  are the pdf and the cdf of standardized location-scale distribution of  $Z = (y - \mu)/\sigma$  respectively.

A random variable  $T$  is said to have a log-location-scale distribution if  $Y = \log(T)$  has a location-scale distribution given by Equation (3) and the pdf, cdf and survival function of  $T$  are given as [5]:

$$\left. \begin{aligned} f(t|\mu, \sigma) &= \frac{1}{\sigma t} g\left(\frac{\log(t) - \mu}{\sigma}\right) \\ F(t|\mu, \sigma) &= G\left(\frac{\log(t) - \mu}{\sigma}\right) \\ S(t|\mu, \sigma) &= S\left(\frac{\log(t) - \mu}{\sigma}\right) \end{aligned} \right\} \quad (4)$$

Suppose that a random variable  $Z$  has a standard form location-scale distribution with survival function  $S_0(z)$ . Then the survival function of  $T$  defined by  $\log(t) = \mathbf{x}'\beta + \sigma \mathbf{z} = \mu + \sigma \mathbf{z}$  can be written as

$$\begin{aligned} S(t) &= Pr(T > t) = Pr\left(Z > \left(\frac{\log(t) - \mu}{\sigma}\right)\right) \\ &= S_0\left(\frac{\log(t) - \mu}{\sigma}\right) \\ &= S_0^*\left(\left(\frac{t}{\exp(\mathbf{x}'\beta)}\right)^{1/\sigma}\right) \end{aligned} \quad (5)$$

where  $S_0^*(t) = S_0(\log(t))$  is the survival function of the standard form of  $\log(t)$ , survival time  $t$  is rescaled by  $\exp(\mathbf{x}'\beta) = e^\mu$  and the effect of rescaling can be thought of as ‘accelerating time’ which is the rationale to consider log-location-scale distributions as accelerated failure time models. The effect of the covariates in an accelerated failure time model is to change the scale but not the location of a baseline distribution of survival time.

The present paper aims to discuss log-location-scale models of  $T$ , Weibull, log-normal and log-logistic corresponding to standard location-scale, extreme value, normal and logistic distributions of  $Z = (\log(T) - \mu)/\sigma$  for the analysis of head-and-neck cancer data to compare the treatments whether they accelerate or decelerate the survival process.

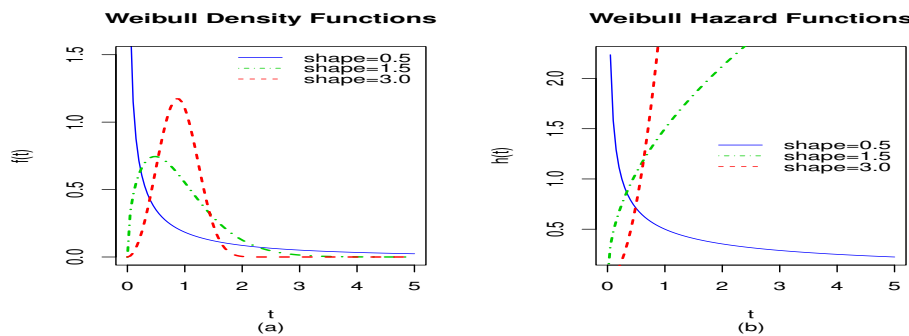
### 2.1 Weibull Survival Model

Weibull distribution is a widely used model in reliability and survival analysis. Its hazard function is monotone increasing and decreasing. Moreover, algebraic expressions for the survival and hazard functions can be obtained explicitly. Because of flexibility and tractability of hazard and survival functions, Weibull model is popular among the researchers.

Suppose survival time  $T$  follows Weibull distribution with shape  $\alpha(> 0)$  and scale parameter  $\lambda(> 0)$  then the pdf  $f(t)$ , survival function  $S(t)$  and hazard function  $h(t)$  of Weibull( $\alpha, \lambda$ ) distribution are given as follows [5]:

$$\left. \begin{aligned} f(t|\alpha, \lambda) &= (\alpha/\lambda)(t/\lambda)^{(\alpha-1)} \exp[-(t/\lambda)^\alpha], \quad t > 0 \\ S(t|\alpha, \lambda) &= \exp[-(t/\lambda)^\alpha] \\ h(t|\alpha, \lambda) &= (\alpha/\lambda)(t/\lambda)^{(\alpha-1)} \end{aligned} \right\} \tag{6}$$

The density and hazard curves of Weibull model for different values of parameters are shown in Figure 1.



**Fig. 1:** Weibull density and hazard functions for different values of shape and scale is unity.

### 2.2 Log-normal Survival Model

Log-normal distribution is a survival model whose hazard increases from zero to a maximum, then decreases to zero as time approaches infinity [5]. The survival and hazard function of log-normal distribution can not be expressed explicitly. Log-normal model does not behave well in the presence of heavy censoring. Accordingly, this model can be applied to describe non-monotonic hazards only when the survival data do not contain many censored observations [7].

Suppose that lifetime  $T$  is such that  $Y = \log(T)$  follows normal distribution with mean  $\mu$ , and variance  $\sigma^2$  then  $T$  follows

log-normal distribution,  $\text{LogNormal}(\mu, \sigma^2)$  with location parameter  $\mu$  ( $-\infty < \mu < \infty$ ), scale parameter  $\sigma$  ( $> 0$ ) having pdf  $f(t)$ , survival function  $S(t)$  and hazard function  $h(t)$  given below:

$$\left. \begin{aligned} f(t|\mu, \sigma) &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma t} \exp\left(-\frac{1}{2} \left(\frac{\log(t) - \mu}{\sigma}\right)^2\right), t > 0 \\ S(t|\mu, \sigma) &= 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right) \\ h(t|\mu, \sigma) &= \frac{f(t|\mu, \sigma)}{S(t|\mu, \sigma)} \end{aligned} \right\} \quad (7)$$

where  $\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp(-u^2/2) du$ .

The density and hazard functions of log-normal model for different values of the parameters are shown in Figure 2.

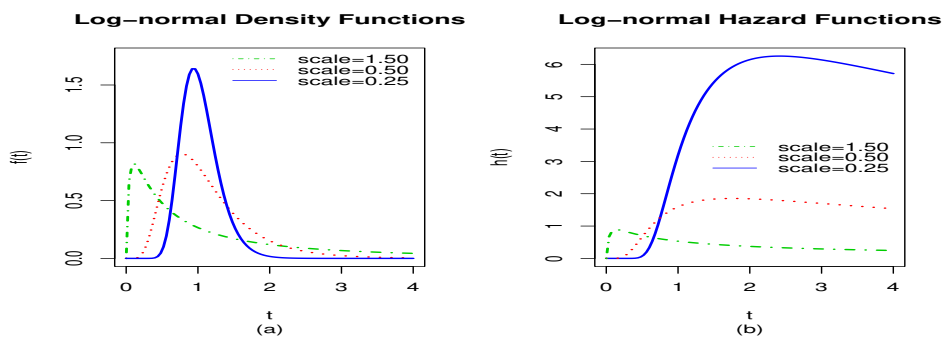


Fig. 2: Log-normal density and hazard functions for different values of scale and location is zero.

### 2.3 Log-logistic Survival Model

Log-logistic distribution is also a frequently used model in reliability and survival analysis. It has monotone decreasing or non-monotone hazard function having a single maximum that increases to the maximum, then decreases thereafter [8, 5]. Its shapes of density and hazard are similar to the shapes of the density and hazard of log-normal distribution. Log-logistic model has explicit algebraic expression for survival and hazard functions which make it more suitable for the analysis of censored survival data than the log-normal model. This is the only lifetime model that belongs to both accelerated failure time model and proportional odds model. [9] and [10] explored the distribution as survival models from classical framework. [11] studied this distribution as a reliability model under Bayesian perspective.

Suppose that survival time  $T$  follows log-logistic distribution with shape parameter  $\alpha$  ( $> 0$ ) and scale parameter  $\lambda$  ( $> 0$ ). Then the pdf  $f(t)$ , survival function  $S(t)$  and hazard function  $h(t)$  of log-logistic distribution,  $\text{LLogist}(\alpha, \lambda)$  [5], are given as follows:

$$\left. \begin{aligned} f(t|\alpha, \lambda) &= \left(\frac{\alpha}{\lambda}\right) \left(\frac{t}{\lambda}\right)^{\alpha-1} \left[1 + \left(\frac{t}{\lambda}\right)^\alpha\right]^{-2}, t > 0 \\ S(t|\alpha, \lambda) &= \left[1 + \left(\frac{t}{\lambda}\right)^\alpha\right]^{-1} \\ h(t|\alpha, \lambda) &= \left(\frac{\alpha}{\lambda}\right) \left(\frac{t}{\lambda}\right)^{\alpha-1} \left[1 + \left(\frac{t}{\lambda}\right)^\alpha\right]^{-1} \end{aligned} \right\} \quad (8)$$

The density and hazard curves of log-logistic model for different values of parameters are shown in Figure 3.

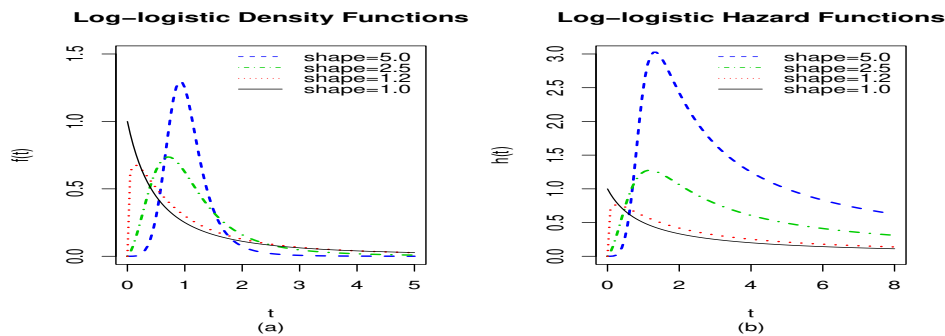


Fig. 3: Log-logistic density and hazard functions for different values of shape and scale is unity.

### 3 Bayesian Analysis of AFT Models

The fundamental assumption of Bayesian statistics is that parameters are random variables having prior distributions  $p(\theta)$ . In Bayesian analysis, we seek the exact distributions of parameters combining prior distribution and data which is called posterior distribution of parameters. Bayesian statistics is based on Bayes theorem. Suppose that the data values  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  are obtained independently from the model  $f(\mathbf{y}|\theta)$ , then the likelihood function is given by

$$L(\theta|y) = f(y_1, y_2, \dots, y_n|\theta) = \prod_{i=1}^n f(y_i|\theta) \tag{9}$$

The posterior distribution  $f(\theta|y)$  is obtained by applying Bayes theorem [19]:

$$f(\theta|y) = \frac{f(\mathbf{y}, \theta)}{h(\mathbf{y})} = \frac{p(\theta)f(\mathbf{y}|\theta)}{\int p(\theta)f(\mathbf{y}|\theta)d\theta} = \frac{p(\theta)L(\theta|y)}{\int p(\theta)L(\theta|y)d\theta} \propto L(\theta|y)p(\theta) \tag{10}$$

where  $h(y) = \int L(\theta|y)p(\theta)d\theta$  is the marginal distribution of  $y$  which is independent of  $\theta$ . That is,

$$\text{Posterior} \propto \text{Likelihood} \times \text{Priors.}$$

#### Likelihood Function for Right Censored Data:

Suppose that  $\mathbf{t} = (t_1, t_2, \dots, t_n)'$  are independent observed survival times- complete or censored, each having a survival model;  $\delta = (\delta_1, \delta_2, \dots, \delta_n)'$  are censoring indicator with  $\delta_i = 1$  indicating the event occurs and  $\delta_i = 0$  indicates censored observation,  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$  is the vector of covariates for the  $i$ th individuals,  $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)'$  is the  $n \times (p + 1)$  design matrix and  $D = (\mathbf{t}, \delta, \mathbf{X})$  denotes the observed data for the model. Then, the likelihood function of the parameters  $\theta = (\sigma, \beta) = (\sigma, \beta_0, \beta_1, \dots, \beta_p)$  for a right censored sample is [4, 5, 12] given as:

$$\begin{aligned} L(\sigma, \beta|D) &= \prod_{i=1}^n f(t_i|\sigma, \beta)^{\delta_i} S(t_i|\sigma, \beta)^{1-\delta_i} \\ &= \prod_i^n \left( \frac{f(t_i|\sigma, \beta)}{S(t_i|\sigma, \beta)} \right)^{\delta_i} S(t_i|\sigma, \beta) \\ &= \prod_{i=1}^n \left( h(t_i|\sigma, \beta) \right)^{\delta_i} S(t_i|\sigma, \beta) \end{aligned} \tag{11}$$

Taking logarithm of both sides of likelihood function, the log likelihood can be written by the following two alternative equations:

$$l(\sigma, \beta | D) = \sum_{i=1}^n \left( \delta_i (\log f(t_i | \sigma, \beta) - \log S(t_i | \sigma, \beta)) + \log S(t_i | \sigma, \beta) \right) \quad (12)$$

$$l(\sigma, \beta | D) = \sum_{i=1}^n \left( \delta_i (\log h(t_i | \sigma, \beta)) + \log S(t_i | \sigma, \beta) \right) \quad (13)$$

### 3.1 Description of Data

[13] reported the survival times of two groups of head and neck cancer patients treated with two treatments in a randomized clinical trial. The study was conducted by the Northern California Oncology Group. One group of patients were treated with radiation therapy alone (RT) and the patients in the other group were treated with radiation plus chemotherapy (RCT). Survival time was measured in days. Table 1 shows the survival times of patients treated with RT and the survival times of patients treated with RCT are given in Table 2. Censored observations are indicated with a plus sign. Efron analyzed the data with classical parametric and nonparametric methods and compared the survival curves under the two treatments. The treatment procedure RCT showed higher survival than RT. [14] discussed the data with Bayesian approach using log-normal model.

**Table 1:** Survival times (in days) of 51 HNC patients treated with RT

7, 34, 42, 63, 64, 74+, 83, 84, 91, 108, 112, 129, 133, 133, 139, 140, 140, 146, 149, 154, 157, 160, 160, 165, 173, 176, 185+, 218, 225, 241, 248, 273, 277, 279+, 297, 319+, 405, 417, 420, 440, 523, 523+, 583, 594, 1101, 1116+, 1146, 1226+, 1349+, 1412+, 1417
--

**Table 2:** Survival times (in days) of 45 HNC patients treated with RCT

37, 84, 92, 94, 110, 112, 119, 127, 130, 133, 140, 146, 155, 159, 169+, 173, 179, 194, 195, 209, 249, 281, 319, 339, 432, 469, 519, 528+, 547+, 613+, 633, 725, 759+, 817, 1092+, 1245+, 1331+, 1557+, 1642+, 1771+, 1776, 1897+, 2023+, 2146+, 2297+
--

### 3.2 Stan in brief

Stan is a probabilistic programming language for Bayesian analysis in the sense that a random variable is a bonafide first-class object. In Stan, variables may be treated as random, and among the random variables, some are observed and some are unknown and need to be estimated or used for posterior predictive inference. It uses No-U-Turn (NUTS) sampler, an adaptive form of Hamiltonian Monte Carlo sampling that is more efficient than other Metropolis-Hastings algorithms, specially for high-dimensional models regardless of whether the priors are conjugate or not [15, 16]. A complete Stan program consists of six code blocks. A sequence of programming statements surrounded by curly braces `{}` form a block. A statement ends with a semi-colon. A comment in Stan is indicated by a double slash `//`. Each block contains a list of instructions for specific tasks. In Stan, statements are

executed imperatively in the order in which they occur in a program . [17] called the language as ‘Stan’ in honor of Stanislaw Ulam, one of the creators of the Monte Carlo methods. The component blocks of a Stan algorithm are described below.

- Data Block: The data code block declares the variables that must be input into the algorithm- the type, dimension and name of every variable has to be declared.
- Transformed Data Block: The transformed data block may be used to define new variables that can be computed based on the data. Any temporary variable used to store a transformation performed on the data without the involvement of parameters should be defined here. The transformed data block starts with a sequence of variable declarations and continues with a sequence of statements defining the variables. For example, standardized versions of data can be defined in a transformed data block.
- Parameters Block: In the parameters block, all the unknown model parameters are declared that are to be sampled by Stan from the posterior density.
- Transformed Parameters Block: The transformed parameters are functions of data and parameters. Any variable declared as a transformed parameter is part of the output produced for samples. Any variable that is defined wholly in terms of data or transformed data should be declared and defined in the transformed data block, defining such quantities in the transformed parameters block is legal, but much less efficient than defining them as transformed data.
- Model Block: The model block contains the model specification. This block is the core of the code structure in which the Bayesian model is defined. The variables defined in the model block are local variables, i.e. other blocks do not know about the variable initialized in this block. After defining the local variables, the model block defines a sampling statement. The sampling statement indicates the priors and the likelihood. The default prior distribution for a parameter is uniform over its support. Stan does not require proper priors.
- Generated Quantities Block: The generated quantities block allows values that depends on parameters and data but do not affect the sampled parameter values. The block is executed only after a sample has been generated. It may be used to calculate posterior expectations, log-likelihood and deviances to generate predictions for new data and to carry out forward simulation for posterior predictive check. Pseudo-random number generators are also available in generated quantities block.

### Assessing convergence of MCMC algorithm and evaluating model fit:

After implementing a Stan program, it is essential to check whether an MCMC algorithm converges to the target posterior distribution, because all inference are made from the simulated samples from the distribution. Convergence of MCMC sampling process to the target posterior distribution is checked quantitatively by *potential scale reduction factor*,  $\hat{R}$  [18], *effective sample size*,  $n_{eff}$  and *Monte Carlo (MC) error*,  $se_{mean}$  and visually by *trace plot* and *autocorrelation plot* [19,20,21].  $\hat{R}$  is defined based on between chain variance and within chain variance and is approximately 1 if convergence is reached. The effective sample size  $n_{eff}$  is a measure of the number of independent samples from the posterior distribution. The larger the effective sample size, the greater the precision of the MCMC estimates. Monte Carlo error,  $se_{mean}$  is a measure of variability of each estimate due to simulation and it is obtained by dividing standard deviation (sd) by the square root of the effective sample size. A low MC error, relative to standard deviation will result in a higher number of independent samples, which is expected. [19] recommended acceptable limit of effective sample size as  $n_{eff} = 100$  and of potential scale reduction factor  $\hat{R} < 1.1$ .

Visual interpretation of convergence is also important. Plotting the values of draws against the iteration, trace plot is obtained. If all the values are within a band showing no discernible periodic

tendencies, then convergence can be assumed. The adjacent samples produced by MCMC algorithms are autocorrelated. If the values of autocorrelation function quickly decrease to 0 with the increase of lag, the distance between successive samples and the MCMC algorithm can be said to be converged. A chain is converged means that it is a stationary chain and adding more samples will not meaningfully change the location and shape of the density of the posterior distribution and so will not change the estimates and other relevant results.

A fitted Bayesian model is accepted as adequate if it predicts the future observations that are consistent with the present data. *Posterior predictive density plot* is used for evaluating model fit.

### 3.3 Weibull Accelerated Failure Time Model

Suppose that a random variable  $Z$  has a standard extreme value distribution with density function  $g(z) = \exp[z - \exp(z)]$  and survival function  $S(z) = \exp[-\exp(z)]$ . Substituting  $z = (\log t - \mathbf{x}'\beta)/\sigma$  from Equation (1) in the extreme value distribution and using Equation (4) and Equation (6), Weibull AFT model,  $T \sim \text{Weibull}(\frac{1}{\sigma}, \exp(\mathbf{x}'\beta))$ , is obtained as follows:

$$\left. \begin{aligned} f(t|\sigma, \beta) &= \left( \frac{\sigma^{-1}}{\exp(\mathbf{x}'\beta)} \right) \left( \frac{t}{\exp(\mathbf{x}'\beta)} \right)^{(1/\sigma-1)} \exp\left( - \left( \frac{t}{\exp(\mathbf{x}'\beta)} \right)^{1/\sigma} \right) \\ S(t|\sigma, \beta) &= \exp\left( - \left( \frac{t}{\exp(\mathbf{x}'\beta)} \right)^{1/\sigma} \right) \\ h(t|\sigma, \beta) &= \left( \frac{\sigma^{-1}}{\exp(\mathbf{x}'\beta)} \right) \left( \frac{t}{\exp(\mathbf{x}'\beta)} \right)^{(1/\sigma-1)} \end{aligned} \right\} \quad (14)$$

where  $h(t) = f(t)/S(t)$ .

#### Bayesian fitting of Weibull AFT model:

The weakly informative prior distribution for scale parameter  $\sigma$  is considered half-Cauchy(0, 25) and for the regression coefficients  $\beta$  as normal (0, 100). That is,  $p(\sigma) = \text{half-Cauchy}(0, 25)$  and  $p(\beta_j) = \text{normal}(0, 100)$  [22]. [23] used half-Cauchy(0, 25) as prior for scale parameter. Thus, the joint posterior distribution of the parameters  $(\sigma, \beta) = (\sigma, \beta_0, \beta_1, \dots, \beta_p)$  of Weibull AFT model can be written by Equation (10) and (11), assuming the parameters are independent, as follows:

$$f(\sigma, \beta|t, X) \propto L(\sigma, \beta|D) \times p(\sigma) \times p(\beta) \quad (15)$$

where, the priors  $p(\sigma) = \frac{2 \times 25}{\pi(\sigma^2 + 25^2)}$ ,  $p(\beta) = \prod_{j=0}^p \frac{1}{\sqrt{2\pi}100} \exp\left(-\frac{1}{2} \frac{\beta_j^2}{100^2}\right)$ .

The likelihood function,  $L(\sigma, \beta|D)$  is obtained substituting  $f(t_i|\sigma, \beta)$  and  $S(t_i|\sigma, \beta)$  from Equation (14) in Equation (11). The marginal distributions of the parameters can not be obtained in closed form and, as such, MCMC algorithm is employed to get the estimates and other relevant results. Stan language is used for simulation and inference and codes are given in the following section.

#### Stan code for fitting Bayesian Weibull AFT model:

The six code blocks for implementing Bayesian Weibull AFT model are showed below. Codes are explained with comments. The model block is mandatory and all other blocks are optional. Half-Cauchy and normal distributions are considered prior distributions for scale parameter sigma and beta coefficients respectively. The Weibull model for response variable is specified here. In generated quantities block, the pointwise log-likelihood `log_lik` and posterior prediction `y_rep`



are calculated and stored for future use. Three functions, i.e. (i) log survival function, (ii) log hazard function and (iii) log-likelihood function are defined at the beginning, before writing the code blocks. Stan codes are written in the editor RStudio [24]. Stan has interface with R [25] by rstan [26].

```
library(rstan)
stancode_waft = "
functions{
// defines the log survival
vector log_S (vector t, real shape, vector scale){
    vector[num_elements(t)] log_S;
for (i in 1:num_elements(t)){
    log_S[i] = weibull_lccdf(t[i]|shape, scale[i]);
}
return log_S;
}
//defines the log hazard
vector log_h (vector t,real shape, vector scale){
    vector[num_elements(t)] log_h ;
    vector[num_elements(t)] ls ;
    ls = log_S(t,shape, scale) ;
    for (i in 1:num_elements(t)){
log_h[i] = weibull_lpdf(t[i]|shape,scale[i])-ls[i];
    }
return log_h;
}
//defines the log likelihood for right censored data
real surv_weibull_lpdf( vector t,vector d,
                        real shape,vector scale){
vector[num_elements(t)] log_lik;
real prob;
log_lik = d .* log_h(t,shape,scale)+log_S(t,shape,scale);
prob = sum(log_lik);
return prob;
}
}
//data block
data{
int N; // number of observations
vector <lower=0> [N] y; // observed times
vector <lower=0,upper=1> [N] event;//censoring (1=obs.,
// 0=cens.)

int M; // number of covariates
matrix[N,M] x;//matrix of covariates (N rows, M columns)
}
//parameters block
parameters{
```

```

vector [M] beta;//coeff.in the linear predictor
real <lower=0> sigma; //scale parameter sigma=1/shape
}
// transformed parameters block
transformed parameters{
vector[N] linpred;
vector[N] mu;
linpred = x*beta;//linear predictor
for (i in 1:N){
mu[i] = exp(linpred[i]);
}
}
// model block
model{
sigma ~ cauchy(0,25); // prior for sigma
beta ~ normal(0,100);//prior for beta coefficients
y ~ surv_weibull(event,1/sigma,mu);//model for data
}
// generated quantities block
generated quantities{
vector[N] y_rep;//posterior predictive value
vector[N] log_lik;//log-likelihood
{ for(n in 1:N){
log_lik[n] = (((weibull_lpdf(y[n]|1/sigma,exp(x[n,]*beta)))-
(weibull_lccdf(y[n]|1/sigma,exp(x[n,]*beta))))*
event[n])+(weibull_lccdf(y[n]|1/sigma,
exp(x[n,]*beta)));}
}
{ for (n in 1:N){
y_rep[n] = weibull_rng( 1/sigma, exp((x[n,]*beta)));}
}
}
"

```

The whole code blocks are saved as `stancode_waft` that is used afterwards.

### Data preparation for Stan (Weibull model):

Stan requires a data list that might include a matrix, vector and values. Hence, data must be prepared to feed into Stan. An object `list` is constructed using R and assigned it to `dat1`.

```

# survival times of head-neck cancer patients in days
y <- c(7, 34, 42, 63, 64, 74, 83, 84, 91, 108, 112, 129, 133, 133, 139, 140,
140, 146, 149, 154, 157, 160, 160, 165, 173, 176, 185, 218, 225, 241,
248, 273, 277, 279, 297, 319, 405, 417, 420, 440, 523, 523, 583, 594,
1101, 1116, 1146, 1226, 1349, 1412, 1417, 37, 84, 92, 94, 110, 112,
119, 127, 130, 133, 140, 146, 155, 159, 169, 173, 179, 194, 195, 209,
249, 281, 319, 339, 432, 469, 519, 528, 547, 613, 633, 725, 759, 817,
1092, 1245, 1331, 1557, 1642, 1771, 1776, 1897, 2023, 2146, 2297)

```

```
# event status, 0 indicates censored and 1 indicates event
event <- c(rep(1,5),0,rep(1,20),0,rep(1,6),0,1,0,rep(1,5),
          0,1,1,1,0,1,0,0,0,rep(1,15),0,rep(1,12),0,0,0,1,
          1,0,1,rep(0,3),1,0,0,1,rep(0,4))
# x1=0, patients treated with RT, x1=1, treated with RCT
x1 <- rep(c(0,1), c(51,45))
x <- cbind(1,x1)
N <- nrow(x)
M <- ncol(x)
dat1 = list(y=y,event=event,x=x,N=N,M=M)
```

### Model fitting

To fit the Weibull AFT model under Bayesian framework and to simulate from the posterior distribution, the function `stan()` from the package `rstan` [26] is called and a `stanfit` object `M1` (say) is created. In Stan, default choices for chain and iteration are 4 and 2000 respectively. We have also fixed 4 chains and 2000 iterations for each chain. To explain, for each of the 4 chains 2000 samples are drawn for each of the parameters. Stan uses half of the iterations as warmup iterations, so post-warmup draws per chain is 1000.

```
M1 <- stan(model_code=stancode_waft,data=dat1,
          iter=2000,chains=4)
```

### Summarizing output of Stanfit Weibull AFT model:

Using `print()` command, summary results are obtained from the fitted object `M1` and are reported in Table 3. Trace plots and autocorrelation plots are made for visual convergence checking. For posterior predictive density plot `bayesplot` package [27] and for model comparison criteria LOOIC and WAIC, `loo` package [28] are used and they are reported in Table 6.

```
print(M1,c("beta","sigma"),digits=3,
      probs= c(0.025,0.50,0.975))

require(bayesplot)
require(loo)
stan_trace(M1, pars=c("beta","sigma"))+
  ggtitle("Trace plot (Weibull AFT model)")
stan_ac(M1, pars=c("beta","sigma"))+grid_lines()+
  ggtitle("Autocorrelation plot (Weibull AFT model)")
#posterior predictive check
# posterior predictive value y_rep
y_rep <- as.matrix(M1,pars="y_rep")
ppc_dens_overlay(y,y_rep[100:130,])+grid_lines()+
  ggtitle("PPD plot (Weibull AFT model)")
# Caterpillar plot for showing credible interval
stan_plot(M1,pars=c("beta","sigma"),ci_level=0.95)+
  grid_lines()+
  ggtitle("Caterpillar plot (Weibull AFT model)")
#calculating LOOIC and WAIC using loo package
log_lik_1 <- extract_log_lik(M1,parameter_name="log_lik",
```

```

merge_chains = TRUE)
loo_1 <- loo(log_lik_1, r_eff=NULL, save_psis=FALSE)
print(loo_1)
waic1 <- waic(log_lik_1)
print(waic1)

```

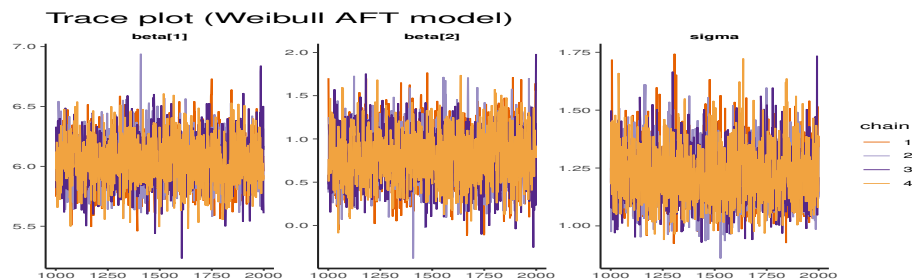
**Table 3:** Summary results of fitted Weibull AFT model

Parameter	mean	se_mean	sd	2.5%	50%	97.5%	n_eff	$\hat{R}$
beta[1]	6.050	0.004	0.189	5.677	6.044	6.430	2309	1.001
beta[2]	0.794	0.006	0.291	0.225	0.794	1.366	2279	1.000
sigma	1.219	0.002	0.115	1.016	1.210	1.469	2623	1.000

### Convergence check and evaluating Weibull AFT model fit:

From the summary results, it is seen that  $\hat{R}$  is 1,  $n_{eff}$  is greater than 100 and  $se_{mean}$  is less relative to the standard deviations for all of the parameters which indicate that convergence of MCMC algorithm has been achieved. Trace plot (Figure 4) shows no periodicity and autocorrelation plot (Figure 5) shows that autocorrelation function drops to near zero quickly with the increase of lag indicating convergence of the MCMC sampling process to the joint posterior distribution. That is, MCMC algorithm performs correctly to explore the target posterior distribution.

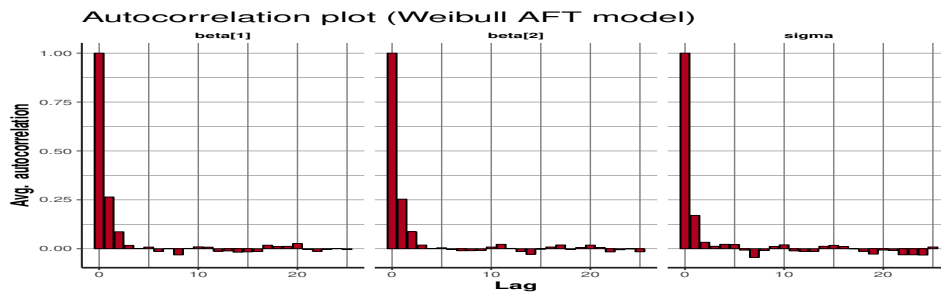
Model fit is assessed visually by posterior predictive density plot (Figure 6) which is made using `bayesplot` package. From posterior predictive density plot (Figure 6), it is observed that Weibull AFT model is well to predict the future observations that are compatible with the current data.



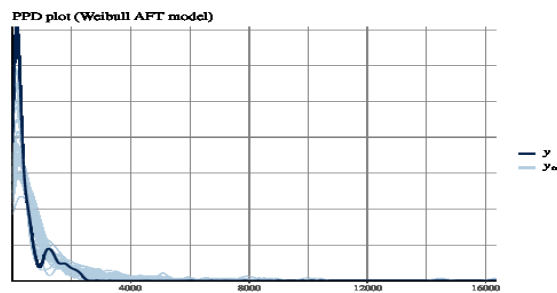
**Fig. 4:** Trace plot of fitted Weibull AFT model are obtained by plotting parameter values along the Y-axis against their corresponding iterations along the X-axis and there is no tendency of periodicity of the plot showing convergence of the algorithm.

### Interpretation of results of Bayesian fitted Weibull AFT model:

Bayesian point estimates of the parameters obtained from posterior distribution, their standard deviations and quantiles are given under the columns mean, sd, 2.5%, 50% and 97.5% quantiles (Table 3). Radiation therapy (RT) is considered the reference category. In regression modelling with Stan, intercept is denoted as  $\beta[1]$ . The coefficient  $\beta[2]$  of treatment variable  $x_1$  ( $x_1$  is zero if a patient is treated with RT and  $x_1$  is one if the patient is treated with RCT) is positive which means that new treatment RCT will delay the event, so, length of lifetime will increase. Estimated value of the coefficient  $\beta[2] = 0.794$  that belongs to the 95% credible interval (0.225, 1.366) which does not include zero value indicating statistical significance. Moreover, it is seen from the caterpillar plot (Figure 7) that 95% credible interval for the coefficient does not include zero value, so the coefficient



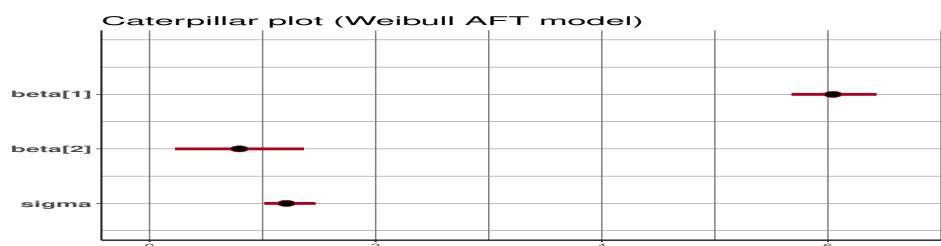
**Fig. 5:** Autocorrelation plot of fitted Weibull AFT model shows that autocorrelation drops to values close to zero at around lags of 4.



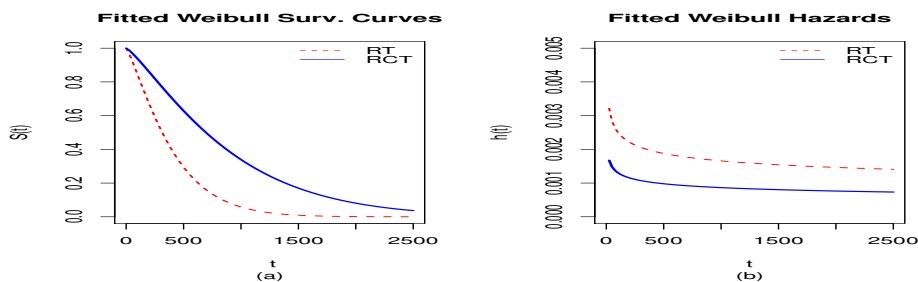
**Fig. 6:** Posterior predictive density (PPD) plot of Weibull AFT model is done by plotting the data  $y$  and then overlaying the density of the predicted values  $y_{rep}$ . The plot shows that the posterior predictive density fits the data well.

of the treatment is statistically significant. The acceleration factor is  $\exp(0.794) = 2.21$  for a patient treated with RCT. The time to death of a patient treated with RCT is therefore delayed by a factor of about 2.21, compared to a patient treated with RT under Weibull AFT model.

Fitted survival curves and hazard curves are drawn in Figure 8 and the curves resemble the numerical results that a patient treated with RCT would survive longer than a patient treated with RT.



**Fig. 7:** Caterpillar plot of Weibull AFT model shows that 95% credible intervals of the parameters do not include zero value so the parameters are statistically significant.



**Fig. 8:** Fitted survival curves and hazard curves of Weibull AFT model. Survival curve is higher and hazard rate is lower for the patients treated with RCT than that of the patients treated with RT.

### 3.4 Log-normal Accelerated Failure Time Model

Suppose that a random variable  $Z$  has a standard normal distribution with density function  $g(z) = N(0, 1)$  and survival function  $S(z) = 1 - \Phi(z)$ . Substituting  $z = (\log t - \mathbf{x}'\beta)/\sigma$  from Equation (1) in the standard normal distribution and using Equation (4) and (7) log-normal AFT model,  $T \sim \text{log-normal}(\mathbf{x}'\beta, \sigma^2)$  is obtained, as follows:

$$\left. \begin{aligned} f(t|\beta, \sigma) &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma t} \exp\left(-\frac{1}{2} \left(\frac{\log(t) - \mathbf{x}'\beta}{\sigma}\right)^2\right) \\ S(t|\beta, \sigma) &= 1 - \Phi\left(\frac{\log t - \mathbf{x}'\beta}{\sigma}\right) \\ h(t|\beta, \sigma) &= \frac{f(t|\beta, \sigma)}{S(t|\beta, \sigma)} \end{aligned} \right\} \quad (16)$$

#### Bayesian fitting of log-normal AFT model

The weakly informative prior distributions are considered  $p(\sigma) = \text{half-Cauchy}(0, 25)$  and  $p(\beta_j) = \text{normal}(0, 100)$  [22]. Thus, the joint posterior distribution of the parameters  $(\sigma, \beta) = (\sigma, \beta_0, \beta_1, \dots, \beta_p)$  of log-normal AFT model can be written by Equation (10) and (11), assuming independence of the parameters, as below:

$$f(\sigma, \beta | t, X) \propto L(\sigma, \beta | D) \times p(\sigma) \times p(\beta)$$

The likelihood function,  $L(\sigma, \beta | D)$  is obtained substituting  $f(t_i | \sigma, \beta)$  and  $S(t_i | \sigma, \beta)$  from Equation (16) in Equation (11). The marginal distributions of the parameters can not be obtained explicitly, so MCMC algorithm is applied to get the estimates and other relevant results. Stan language is used for the analysis and codes are given in the following section.

#### Stan code for fitting Bayesian log-normal AFT model:

Stan codes are written with comments for fitting the log-normal AFT model under Bayesian framework.

```
library(rstan)
stancode_lnaft = "
functions{
// defines the log survival
vector log_S (vector t, vector location, real scale){
```

```

    vector[num_elements(t)] log_S ;
for (i in 1:num_elements(t)){
  log_S[i] = lognormal_lccdf(t[i]|location[i],scale);
}
return log_S;
}
//defines the log hazard
vector log_h (vector t,vector location,real scale){
  vector[num_elements(t)] log_h ;
  vector[num_elements(t)] ls ;
  ls = log_S(t,location, scale) ;
  for (i in 1:num_elements(t)){
log_h[i] = lognormal_lpdf(t[i]|location[i],scale)-
          ls[i];
  }
return log_h;
}
//defines the sampling distribution for right censored data
real surv_lognormal_lpdf( vector t,vector d,
                          vector location,real scale){
vector[num_elements(t)] log_lik;
real prob;
log_lik = d .* log_h(t, location, scale) +
          log_S(t, location, scale);
prob = sum(log_lik);
return prob;
}
}
//data block
data{
int N;          // number of observations
vector <lower=0> [N] y; // observation vector
vector <lower=0,upper=1> [N] event;//censoring(1=obs.,
// 0=cens.)

int M;    // number of covariates
matrix [N,M] x;//matrix of covariates (N rows, M columns)
}
//parameters block
parameters{
vector [M] beta;//coeff. in the linear predictor
real<lower=0> sigma; // scale parameter sigma=1/shape
}
// transformed parameters block
transformed parameters{
vector[N] linpred;
vector[N] mu;
linpred = x*beta;//linear predictor

```

```

for (i in 1:N){
mu[i] = linpred[i];
}
}
// model block
model{
sigma ~ cauchy(0,25); // prior for sigma
beta ~ normal(0,100); //prior for beta coefficients
y ~ surv_lognormal(event,mu,sigma); // model for data
}
//generated quantities block
generated quantities{
vector[N] y_rep; //posterior predictive value
vector[N] log_lik; //log-likelihood
for(n in 1:N)
log_lik[n] = (((lognormal_lpdf(y[n] | (x[n,]*beta), sigma)) -
(lognormal_lccdf(y[n] | (x[n,]*beta), sigma))) *
event[n]) +
(lognormal_lccdf(y[n] | (x[n,]*beta), sigma));
for(n in 1:N)
y_rep[n] = lognormal_rng((x[n,]*beta), sigma);
}
"

```

The whole code blocks are saved as `stancode_lnaft`. The same head and neck cancer data object (`dat1`) prepared for Weibull AFT model is applied here for fitting log-normal AFT model.

### Model fitting and summarizing output of Stanfit log-normal AFT model:

A stanfit object M2 (say) is created with the function `stan()` from the package `rstan`. Summary results are obtained, using `print()` command, from the fitted object M2 and are reported in Table 4. Then, from M2, trace plots and autocorrelation plots are made for visual convergence checking. Caterpillar plot is made for credible interval that indicates significance of the parameters and for model comparison criteria LOOIC and WAIC are reported in Table 6.

```

M2 <- stan(model_code=stancode_lnaft, data=dat1,
           iter=2000, chains=4)
print(M2, c("beta", "sigma"), digits=3,
      probs=c(0.025, 0.50, 0.975))

```

**Table 4:** Summary results of fitted log-normal AFT model

Parameter	mean	se_mean	sd	2.5%	50%	97.5%	n_eff	Rhat
beta[1]	5.547	0.004	0.188	5.188	5.547	5.923	2147	1.002
beta[2]	0.623	0.006	0.283	0.084	0.619	1.185	2237	1.002
sigma	1.336	0.002	0.119	1.125	1.329	1.597	2501	1.002

### Convergence diagnostics and evaluating model fit for log-normal AFT model:

The summary results show that  $\hat{R}$  is 1,  $n_{eff}$  is greater than 100 and  $se_{mean}$  is less relative to the



standard deviations for all of the parameters that indicate convergence of MCMC algorithm. Trace plot (Figure 9) and autocorrelation plot (Figure 10) also show that the MCMC sampling process converged to the joint posterior distribution. Moreover, from posterior predictive density plot (Figure 11), it is observed that log-normal model is well suited to the data.

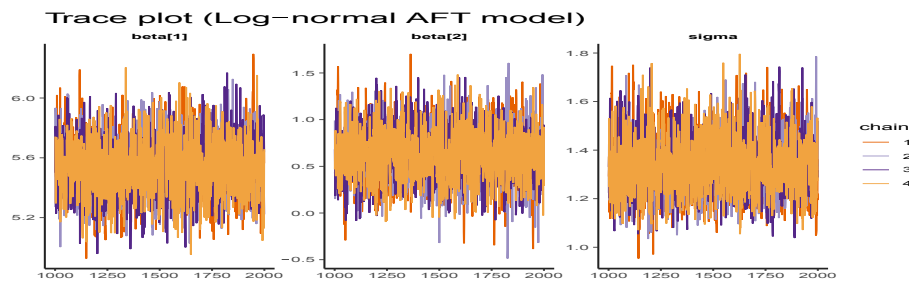


Fig. 9: Trace plot of fitted log-normal AFT model shows no tendency of periodicity indicating convergence of the algorithm.

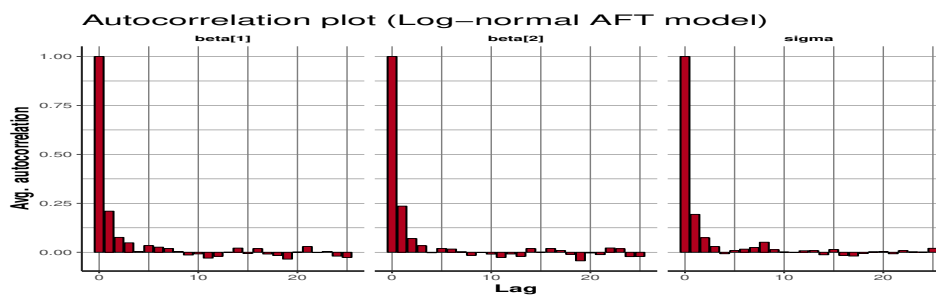


Fig. 10: Autocorrelation plot of fitted log-normal AFT model shows that autocorrelation drops to values close to zero as lag increases.

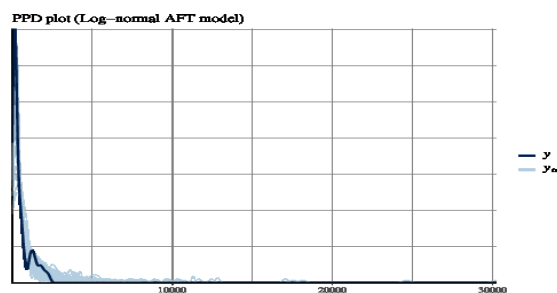


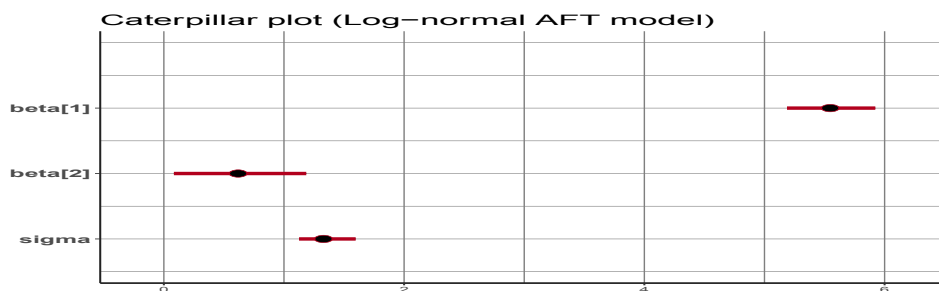
Fig. 11: Posterior predictive density (PPD) plot of log-normal AFT model shows that the PPD fits the data well.

**Interpreting results of Bayesian fitted log-normal AFT model):**

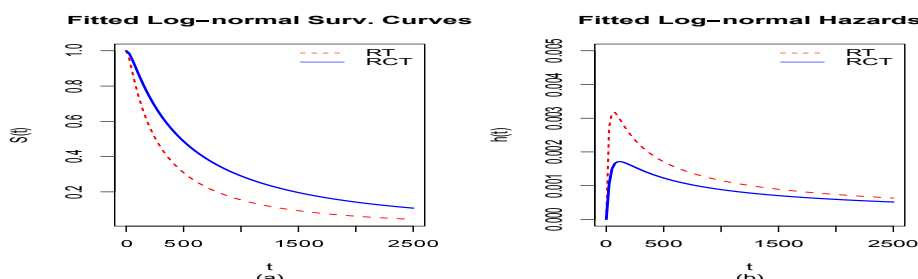
The coefficient  $\beta[2]$  of treatment variable  $x_1$  is positive, suggesting that new treatment RCT will prolong life of patients. Consequently, the failure will be delayed. Bayesian point estimate of the coefficient  $\beta[2]$  is 0.623 which is contained in the 95% credible interval (0.084, 1.185) that does not

include zero. It is evident from the credible interval and caterpillar plot (Figure 12) that the coefficient of the treatment is statistically significant. The acceleration factor is  $\exp(0.623) = 1.86$  that explains that the time to death of a patient treated with RCT is delayed by a factor of about 1.86, compared to a patient treated with RT under log-normal AFT model.

Fitted survival curves and hazard curves are drawn in Figure 13 and the curves mimic the quantitative results that a patient treated with RCT would decelerate death than a patient treated with RT.



**Fig. 12:** Caterpillar plot of log-normal AFT model shows that 95% credible intervals of the parameters do not include zero value, so the parameters are statistically significant.



**Fig. 13:** Fitted survival curves and hazard curves of log-normal AFT model. Survival curve is higher and hazard rate is lower for the patients treated with RCT than that of the patients treated with RT.

### 3.5 Log-logistic Accelerated Failure Time Model

Suppose that a random variable  $Z$  has a standard logistic distribution with density function  $g(z) = \exp(z)[1 - \exp(z)]^{-2}$  and survival function  $S(z) = [1 - \exp(z)]^{-1}$ . Substituting  $z = (\log t - \mathbf{x}'\beta)/\sigma$  from Equation (1) in the standard logistic distribution and using Equation (4) and Equation (8) log-logistic AFT model,  $T \sim \text{LLogist}(\frac{1}{\sigma}, \exp(\mathbf{x}'\beta))$ , is obtained, as follows:

$$\left. \begin{aligned} f(t|\sigma, \beta) &= \left(\frac{\sigma^{-1}}{\exp(\mathbf{x}'\beta)}\right) \left(\frac{t}{\exp(\mathbf{x}'\beta)}\right)^{(1/\sigma-1)} \left[1 + \left(\frac{t}{\exp(\mathbf{x}'\beta)}\right)^{1/\sigma}\right]^{-2} \\ S(t|\sigma, \beta) &= \left[1 + \left(\frac{t}{\exp(\mathbf{x}'\beta)}\right)^{1/\sigma}\right]^{-1} \\ h(t|\sigma, \beta) &= \left(\frac{\sigma^{-1}}{\exp(\mathbf{x}'\beta)}\right) \left(\frac{t}{\exp(\mathbf{x}'\beta)}\right)^{(1/\sigma-1)} \left[1 + \left(\frac{t}{\exp(\mathbf{x}'\beta)}\right)^{1/\sigma}\right]^{-1} \end{aligned} \right\} \quad (17)$$

### Bayesian fitting of Log-logistic AFT model:

The joint posterior distribution of the parameters  $\theta = (\sigma, \beta) = (\sigma, \beta_0, \beta_1, \dots, \beta_p)$  of log-logistic AFT model can be written by Equation (10) and Equation (11), as follows:

$$f(\sigma, \beta | t, X) \propto L(\sigma, \beta | D) \times p(\sigma) \times p(\beta)$$

where, the priors  $p(\sigma) = \frac{2 \times 5}{\pi(\sigma^2 + 25^2)}$ ,  $p(\beta) = \prod_{j=0}^p \frac{1}{\sqrt{2\pi}10^2} \exp\left(-\frac{1}{2} \frac{\beta_j^2}{100^2}\right)$ .

The likelihood function,  $L(\sigma, \beta | D)$  is obtained by substituting  $f(t_i | \sigma, \beta)$  and  $S(t_i | \sigma, \beta)$  from Equation (17) in Equation (11). The joint posterior distribution is obtained using Bayesian software Stan and MCMC algorithm is implemented to find the estimates and other relevant results. Stan codes are given in the following section.

#### Stan code for fitting Bayesian Log-logistic AFT model:

Stan codes are written with comments for fitting the log-logistic accelerated failure time (AFT) model under Bayesian setting.

```

library(rstan)
stancode_llaft = "
functions{
// defines the log survival
vector log_S (vector t, real shape, vector scale){
  vector[num_elements(t)] log_S ;
  for (i in 1:num_elements(t)){
    log_S[i] = -log(1+(t[i]/scale[i])^shape);
  }
  return log_S;
}
//defines the log hazard
vector log_h (vector t, real shape, vector scale){
  vector[num_elements(t)] log_h ;
  vector[num_elements(t)] ls ;
  ls = log_S(t, shape, scale) ;
  for (i in 1:num_elements(t)){
log_h[i] = log(shape)-shape*log(scale[i])+
          (shape-1)*log(t[i])-
          2*log(1+(t[i]/scale[i])^shape)-ls[i];
  }
  return log_h;
}
//defines the log likelihood for right censored data
real surv_lllogist_lpdf( vector t, vector d,
                        real shape, vector scale){
vector[num_elements(t)] log_lik;
real prob;
log_lik = d .* log_h(t, shape, scale)+
          log_S(t, shape, scale);
prob = sum(log_lik);

```

```

return prob;
}
}
//data block
data{
int N;          // number of observations
vector <lower=0> [N] y;//observation vector(times)
vector <lower=0,upper=1> [N] event;//censoring (1=obs.,
//          0=cen.)

int M;    // number of covariates
matrix[N,M] x;//matrix of covariates(N rows, M columns)
}
//parameters block
parameters{
vector [M] beta;//coeff.in the linear predictor
real<lower=0> sigma;//scale parameter sigma=1/shape
}
// transformed parameters block
transformed parameters{
vector[N] linpred;
vector[N] mu;
linpred = x*beta;//linear predictor
for (i in 1:N){
mu[i] = exp(linpred[i]);
}
}
// model block
model{
sigma ~ cauchy(0,25);//prior for sigma
beta ~ normal(0,100);//prior for beta coefficients
y ~ surv_llogist(event,1/sigma,mu);//density for data
}
// generated quantities block
generated quantities{
vector[N] y_rep;//posterior predictive value
vector[N] log_lik;// log-likelihood
{ for(n in 1:N)
log_lik[n] = (((log(1/sigma)-(1/sigma)*(x[n,]*beta)+
((1/sigma)-1)*log(y[n])-
2*log(1+(y[n]/(exp(x[n,]*beta)))^(1/sigma)))-
(-log(1+(y[n]/(exp(x[n,]*beta)))^(1/sigma))))*
event[n])+
(-log(1+(y[n]/(exp(x[n,]*beta)))^(1/sigma)))));
}
{ real u;
u=uniform_rng(0,1);
for (n in 1:N){

```

```

y_rep[n]= (exp(x[n,]*beta)) * (((1-u)^(-1))-1)^sigma);
}
}
"

```

The whole code blocks are saved as `stancode_llaft` that is used afterwards. The head and neck cancer data have already been prepared and the same data object `dat1` is used for fitting log-logistic AFT model under Bayesian paradigm.

**Model fitting and summarizing output of Stanfit log-logistic AFT model:**

Now, the function `stan()` is called from the package `rstan` to fit the log-logistic AFT model; a Bayesian fitted object `M3`(say), which is used for inference, is created. Summary results are reported in Table 5.

```

M3 <- stan(model_code=stancode_llaft,data=dat1,
          iter = 2000, chains = 4)
print(M3,c("beta","sigma"),digits=3,
      probs=c(.025,0.50,.975))

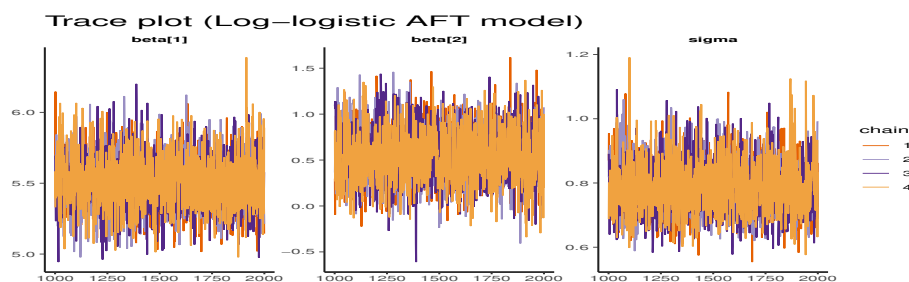
```

**Table 5:** Summary results of fitted log-logistic AFT model

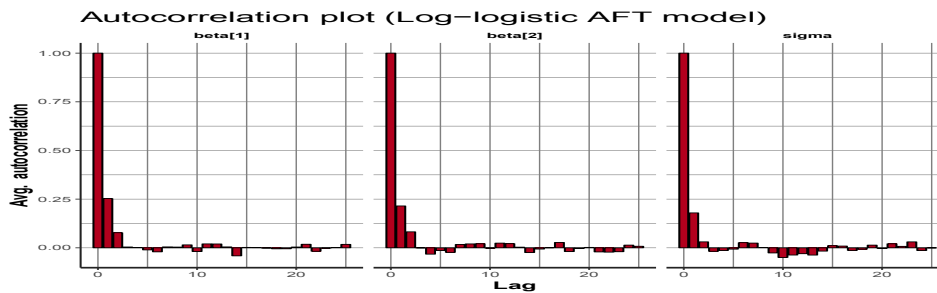
Parameter	mean	se_mean	sd	2.5%	50%	97.5%	n_eff	Rhat
beta[1]	5.507	0.004	0.185	5.154	5.505	5.889	2336	1.002
beta[2]	0.556	0.006	0.283	0.009	0.550	1.111	2480	1.001
sigma	0.785	0.002	0.080	0.646	0.778	0.961	2775	1.001

**Convergence check and model assessment of log-logistic AFT model:**

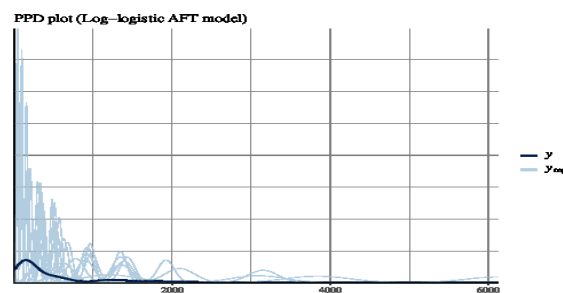
On the basis of summary results,  $\hat{R}$ ,  $n_{eff}$ , and  $se_{mean}$  (Table 5), it can be said that MCMC algorithm has converged to the target joint posterior distribution. Trace plot (Figure 14) and autocorrelation plot (Figure 15) also indicate convergence of the MCMC algorithm. Moreover, from posterior predictive density plot (Figure 16), it is evident that log-logistic model matched the data well.



**Fig. 14:** Trace plot of fitted log-logistic AFT model depicts no tendency of periodicity showing convergence of the algorithm.



**Fig. 15:** Autocorrelation plot of fitted log-logistic AFT model shows that autocorrelation drops to values close to zero as lag increases.

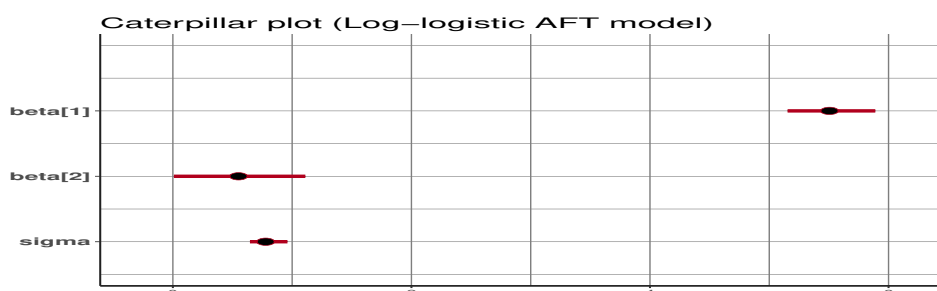


**Fig. 16:** Posterior predictive density (PPD) plot of log-logistic AFT model shows that the PPD fits the data well.

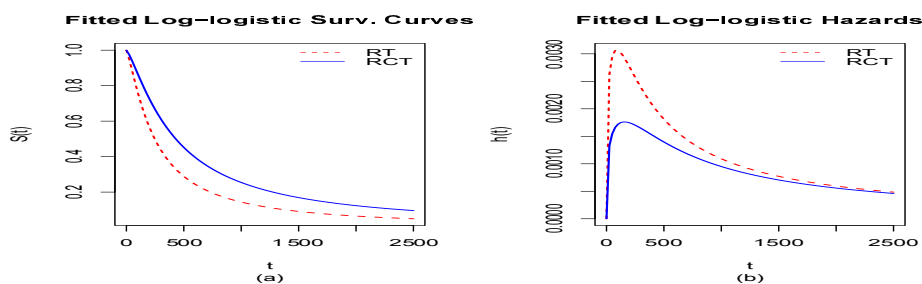
### Interpretation of results of log-logistic AFT model:

The coefficient of treatment variable  $x_1$  ( $\beta[2] = 0.556$ ) is positive which means that new treatment RCT will prolong life of patients, so the failure will be delayed. It is evident from the 95% credible interval (0.009, 1.111) and caterpillar plot (Figure 17) that the coefficient of the treatment is statistically significant. The acceleration factor is  $\exp(0.556) = 1.74$  for a patient treated with RCT. The time to death of a patient treated with RCT is therefore delayed by a factor of about 1.74 compared to a patient treated with RT under log-logistic AFT model.

Fitted survival curves and hazard curves are drawn in Figure 18 and the curves resemble the quantitative results that a patient treated with RCT would slow down death more than a patient treated with RT.



**Fig. 17:** Caterpillar plot of log-logistic AFT model shows that 95% credible intervals of the parameters do not contain zero value so the parameters are statistically significant.



**Fig. 18:** Survival curve is higher and hazard rate is lower for the patients treated with RCT than that of the patients treated with RT.

### 4 Model Comparison

Selecting the best model from among the several competitive models is always crucial in Bayesian statistics and in classical statistics as well. Based on information criteria Leave-one-out cross validation  $LOO$  and Widely Applicable or Watanabe Akaike Information Criterion  $WAIC$  [29, 30, 19] the fitted models are compared. Pointwise log-likelihoods are calculated in the generated quantities block of Stan program and afterwards ‘loo-package’ [28] extracts and uses these quantities to obtain numerical measures  $LOOIC$  ( $LOO$  information criterion) or  $WAIC$  for model comparison. A model with smaller  $LOOIC$  or  $WAIC$  is a better fitted model than the others. On the basis of these measures (Table 6), the log-normal and log-logistic models are almost indistinguishable in fitting the head-and-neck cancer data. However, both the models fit the data better than the Weibull model.

**Table 6:**  $LOOIC$  and  $WAIC$  for model comparison and their standard errors (SE)

Model	$LOOIC$	SE	$WAIC$	SE
Log-logistic	1067.5	47.4	1067.5	47.4
Log-normal	1066.4	48.0	1066.4	48.0
Weibull	1082.8	48.7	1082.8	48.7

### 5 Conclusion

Three accelerated failure time models- Weibull, log-normal and log-logistic are fitted under Bayesian framework to the head and neck cancer data. For all the models, treatment variable was statistically significant. The acceleration factor was greater than one, i.e. survival time is longer for the patients treated with radiation and chemotherapy (RCT) than that of the patients treated with radiation therapy (RT) only. Considering posterior predictive density plots for the models and comparing  $LOOIC$  and  $WAIC$ , it can be concluded that log-normal model fits the data better than log-logistic and Weibull models.

### Conflict of Interest

The authors declare that they have no conflict of interest.

## References

- [1] D.R. Cox, *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187–202 (1972).
- [2] L.-J. Wei, *Statistics in Medicine*, 11(14-15), 1871–1879 (1992).
- [3] N. Reid, *Statistical Science*, 9(3), 439–455 (1994).
- [4] D. Collett, *Modelling Survival Data in Medical Research*, Chapman and Hall/CRC, 2015.
- [5] J.F. Lawless, *Statistical Models and Methods for Lifetime data*, John Wiley & Sons, 2003.
- [6] E.T. Lee, and J. Wang, *Statistical Methods for Survival Data Analysis*, John Wiley & Sons, 2013.
- [7] X. Liu, *Survival Analysis: Models and Applications*, John Wiley & Sons, 2012.
- [8] D.R. Cox and D. Oakes, *Analysis of Survival Data*, Chapman and Hall, 1984.
- [9] S. Bennett, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 32(2), 165–171 (1983).
- [10] J. O'Quigley and L. Struthers, *Computer Programs in Biomedicine*, 15(1), 3–11 (1982).
- [11] M.T. Akhtar and A.A. Khan, *American Journal of Mathematics and Statistics*, 4(3), 162–170 (2014).
- [12] X. Wang, Y.R. Yue and J.J. Faraway, *Bayesian Regression Modeling with INLA*, Chapman and Hall/CRC, 2018.
- [13] B. Efron, *Journal of the American Statistical Association*, 83(402), 414–425 (1988).
- [14] P. Makkar, P.K. Srivastava, R.S. Singh and S.K. Upadhyay, *Communications in Statistics-Theory and Methods*, 43(2), 392–407 (2014).
- [15] B. Carpenter, A. Gelman, M.D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li and A. Riddell, *Journal of Statistical Software*, 76(1), (2017).
- [16] M.D. Hoffman and A. Gelman, *Journal of Machine Learning Research*, 15(1), 1593–1623 (2014).
- [17] Stan Development Team, *Stan Modeling Language Users Guide and Reference Manual*, Version 2.16.0, 2017, <http://mc-stan.org/>.
- [18] A. Gelman and D.B. Rubin, *Statistical Science*, 7(4), 457–472 (1992).
- [19] A. Gelman, H.S. Stern, J.B. Carlin, D.B. Dunson, A. Vehtari and D.B. Rubin, *Bayesian Data Analysis*, Chapman and Hall/CRC, 2013.
- [20] I. Ntzoufras, *Bayesian Modeling using WinBUGS*, Vol. 698, John Wiley & Sons, 2009.
- [21] G. Hamra, R. MacLehose and D. Richardson, *International Journal of Epidemiology*, 42(2), 627–634 (2013).
- [22] A. Gelman, *Bayesian analysis*, *International Society for Bayesian Analysis*, 1(3), 515–534 (2006).
- [23] N. Khan and A.A. Khan, *Austrian Journal of Statistics*, 47(4), 1–15 (2018).
- [24] RStudio Team, *RStudio: Integrated Development Environment for R*, Boston, MA, <http://www.rstudio.com/>, 2015.
- [25] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>, 2017.
- [26] Stan Development Team, *RStan: the R interface to Stan*, R package version 2.17.3, <http://mc-stan.org/>, 2018.
- [27] J. Gabry, T. Mahr, P.-C. Bürkner, M. Modrák and M. Barrett, *bayesplot: Plotting for Bayesian models*, R package version 1.6.0, <http://mc-stan.org/bayesplot>, 2018.
- [28] A. Vehtari, A. Gelman and J. Gabry and Y. Yao, *loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models*, R package version 2.0.0, 2018.
- [29] A. Vehtari, A. Gelman and J. Gabry, *Statistics and Computing*, (27)5, 1413–1432 (2017).
- [30] R. McElreath, *Statistical rethinking: A Bayesian course with examples in R and Stan*, Chapman and Hall/CRC, 2015.
- [31] T. Oetiker, H. Partl, I. Hyna and E. Schlegel, *The Not So Short Introduction to LATEX 2e or LATEX 2e in 157 minutes*, 2016.