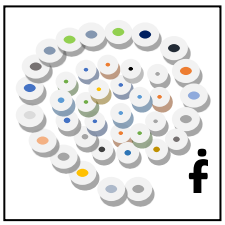

International Journal of Mathematical Analysis and Modelling

(Formerly Journal of the Nigerian Society for Mathematical Biology)

Volume 5, Issue 2 (Sept.), 2022

ISSN (Print): 2682 - 5694
ISSN (Online): 2682 - 5708

International Journal of Mathematical Analysis and Modelling
Volume 5, Issue 2, September 2022, Pages 171 – 190

A Bayesian analysis and estimation of random censored data: application to pharmacovigilance data on Malaria drugs

Y.S. Jackson[*†], S.C. Nwaosu[‡], N.P Dibal[*], and Z. Wudiri[§]

Abstract

In this study, we propose and derive a new Hierarchical Bayesian Weibull Regression model on random censored data using the Jeffrey's prior. We derive the posterior density, marginal and obtain their kernels. The model parameters were estimated using Integrated Nested Laplace Approximation (INLA) and assess the model performance using Kulback Leibler Divergence and Deviance Information Criterion. The data is obtained from a pharmacovigilance study conducted by NAFDAC in collaboration with World Health Organization in Federal republic of Nigeria across six geopolitical zones. The shape parameter θ , were selected in the closed interval between zero and one with their respective priors distributions $\psi^{J}(\theta)$. It is evident that choice of prior $\psi^{J}(\theta) = 0.9$ results in the posterior for θ to be 2.52 for model I. In model II, we use the prior $\psi^{J}(\theta) = 0.001$ and obtain the posterior estimate of 2.514. We obtain the parameter estimate for model III using classical Integrated Nested Laplace Approximation with shape parameter 2.673. For choice of prior $\psi^{J}(\theta) = 80$, the posterior for θ is 3.007 in model IV. When we tune the prior to higher density region the posterior distribution for θ increases. For model I, II and III, the density is Rayleigh, but model IV tends to Normal for $\theta = 3.007$. It implies that higher prior change the shape of the posterior distribution. The predictors on recovery are not sensitive to low priors. The DIC and KLD show that the posterior converges to the real density with KLD of zero for all the models but model IV has the minimum DIC, which will perform better.

Keywords: Jeffrey's prior; hierarchical Bayes model; integrated nested Laplace approximation; latent Gaussian Models; random censoring

[*] Department of Mathematical Sciences, University of Maiduguri, Maiduguri, Nigeria
[†] Corresponding Author. E-mail: samailajyaga@unimaid.edu.ng
[‡] Department of Math/Statistics/Computer Science, University of Agriculture, Makurdi, Nigeria
[§] Department of Community Medicine, College of Medical Sciences, University of Maiduguri, Maiduguri, Nigeria

1 Introduction

Latent Gaussian models are class of hierarchical models with some good number of hyperparameters. The response variable in the model is assume to be non-Gaussian and the latent field have Gaussian density with non-Gaussian hyperparameters. There are quite good approximate methods for performing Bayesian analysis for hierarchical structured models such as the MCMC method, Lindley's approximation, Tierney-Kadane, Lindely-Smith, some can be approximated to Gaussian distribution using Laplace approximation method. In this paper, we proposed a new hierarchical Bayesian Weibull regression model with using Jeffreys prior on random censored data. In Bayesian, statistics there is more concern on the elicitation of prior distribution..

In this study, we use Jeffreys prior for the shape parameter of Weibull density and normal priors for regression parameters and frailty parameter was assign a gamma prior distribution. A Weibull regression model with accelerated failure time was discussed [1]. Accelerated failure time regression model is a linear function of covariates and extreme value distribution. In [2] Weibull regression model was used on time to first and second recurrence of kidney infection with covariates for survival function and inclusion of additive random effect using Bayesian method. They analyze time to first and second recurrence of infection for kidney patients on dialysis. Using the Cox model with multiplicative frailty parameter for each individual. The survival distribution used is a truncated Weibull due to the censored observations. They used the non-informative prior for the regression co-efficient and the precision of random effect.

Bayesian method was used to analyzed data on photo carcinogenicity in four groups of mice each containing 20 mice with [3], they recorded a survival time of death event and the survival distribution assumed a Weibull with failure time and covariate vector with base line hazard function due to censored observations and implemented a truncated Weibull regression model with lower bound corresponding to the censoring time, and regression co-efficient β, where assumed a priori distribution to follow independent normal distribution with zero mean and vague precision of 0.0001. The shape parameter is distributed as $\text{Gamma}(1, 0.0001)$ which coincides with the work of [2]. Bayesian analysis in the presence of covariates for multivariate survival data was developed by [4] and used different frailties (latent variables) to capture correlations among the survival times for the same individual. They use the Weibull and generalized gamma distribution considering right censored lifetime data and the MCMC method for Bayesian analysis and use gamma prior for the scale parameter and gamma prior for the latent parameter resulting to a hierarchical Bayesian model with two levels due to the hyper-parameters in the model.

In determining the best estimator of the two-parameter Weibull distribution, [5] adopted the frequentist maximum likelihood estimator, least squares regression estimator and the Bayesian estimator by using two loss functions, which are, the quadratic loss function, and the linear exponential loss function. The Bayes estimates were obtained using the Lindleys approximation, the comparison was made through simulation to determine the performance of these methods. From the result, they found the Bayesian estimator with linear exponential loss function (LINEX), is found to be the best compared to the conventional MLE and Least squares method. [5] used the similar approach on censored data using the Lindleys approximation for the Bayesian approach with a standard non-informative prior and proposed generalization of the non-informative priors. The Bayes estimator using LINEX loss function with the proposed prior gives better result. They study also coincides with the work of [6] but

the difference is that they use Jeffreys prior and the extension of Jeffreys prior for the Bayesian paradigm on censored data. The parameters, was estimated using the Lindleys approximation.

The Bayesian with extension of Jeffreys prior gives better result compared to the MLE. A two-parameter Weibull density considering the properties of Bayes estimator with different loss functions was estimated in [7], the simulation scheme was done using non-informative priors and informative prior such as gamma prior, exponential-gamma prior, gamma-gamma prior and finally determine the prior and posterior predictive distributions. The parameters were estimated using the quadratic loss, weighted loss, square error loss, and square logarithmic loss functions. Numerical integration was tool for parameter estimation. The posterior risk of joint exponential gamma prior along with quadratic loss are most appropriate for the shape parameter while the joint lognormal-gamma prior with quadratic loss function are suitable for the scale parameter [8] use similar method but used the modified Weibull density. [6] Compared the Bayesian method on interval-censored data for Weibull distribution. The study considered the MLE and Bayes estimates. The Bayesian estimation method cannot solve for the parameters analytically instead, the MCMC method. The full conditional distribution for the scale and shape parameters are obtain using Metropolis Hastings Algorithm (M-H Algorithm), and Lindley's approximation to determine the shape parameter. The result shows that the Bayesian using the M-H Algorithm for estimating the shape and scale parameter based on interval-censored data is the best.

The work in [9] Performed approximate Bayesian inference in a sub-class of structured additive regression called the latent Gaussian models. They use integrated nested Laplace approximation on structured model and assume the response variable belong to the exponential family. The mean is linked to a structured additive linear predictor through a given log-link function and account for effects of various covariates in an additive way. They also provide accurate and fast deterministic approximations to all posterior marginal for the latent field and posterior marginal for the hyper-parameters. Model comparison, was obtained using the marginal likelihood and Deviance Information Criterion (DIC) also [10] investigates the use of INLA to solve Bayesian inferential problem in Bayesian survival analysis, the study uses the exponential and Weibull distributed lifetimes with and without censoring and frailty. Cox model, was fit with piece wise linear baseline hazard and express the model as latent Gaussian model so that INLA proposed by [11] can be a good tool for computation of posterior marginal. The model was tested using time to infection of kidney dialysis patients.

The work in [12] developed a Bayesian Semi-parametric model for occurrence of Asthma attacks in respiratory trials; they decompose the intensity function into product of various terms, first is the baseline intensity and second is the covariate and lastly the frailty term to take care of heterogeneity among different individuals with ability to have quick response to the event. The computations, was done using the INLA and the result turns out that this method has the highest speed and accuracy compared to Monte Carlo based inference.

2 Methods: Bayes theorem for Parametric Inference

Consider a general problem in which we have data y and we require inference about a parameter θ. In Bayesian analysis, θ is unknown and viewed as a random quantity. Thus, it possesses a density function $\psi(\theta)$. Using Bayes theorem, the marginal distribution for θ given the data y is

$$\psi(\theta|y) = \frac{\psi(y|\theta)\psi(\theta)}{\psi(y)} \tag{1}$$

where $\psi(\theta|y)$ is the posterior density, $\psi(\theta)$ is the prior density, $\psi(y|\theta)$ is the likelihood, $\psi(y)$ is the marginal for the data (Congdon, 2003).

3 Non-Informative Prior Distributions

Jeffrey's prior is a non-informative prior that is invariant under reparameterization [16]. The invariance property makes the prior distribution to be more formal than the application of uniform prior distribution. There are so many formal rule for selecting prior but we restrict to the Jeffreys prior. [13], explicitly define Jeffrey's prior in terms of lebesgue measure on the parameter space. Consider a density function $\psi_1(t|\theta)$ with respect to σ-finite dominating measure $\mu(dt)$ on a sample space $(\mathcal{X}, \mathcal{B})$. Θ is a non-empty open subset of k-dimensional parameter space $\mathbb{R}^k$.The function $\ln L(\psi_1(t|\theta))$ is a well-defined differentiable function of θ. We construct a well-defined score function $S_1(t|\theta)$ and its covariance exist. The score function, is given by the expression

$$S_1(\theta) = \left\{ \frac{\partial^2}{\partial\theta_i\theta_j} \ln L(f(t|\theta)), i, j = 1,2, \dots, n \right\} \tag{2}$$

The Fisher Information is derive from the covariance of the score function.
The Jeffreys prior with respect to the parameter θ is

$$\pi^J(\theta) \propto (\det(I_1(\theta)))^{1/2} .$$

4 Hierarchical Bayesian model

Bayesian statistics assumes that the parameters of a given distribution are random hereby having a probability density. Hierarchical Bayes model is a complex Bayesian model that consist of a sampling distribution (Data model) conditional on the parameter θ and the parameter has a distribution $\psi(\theta|\gamma)$ conditional on hyper-parameter γ. Finally if γ has a distribution making the model in different layers or levels. Such analogue in Bayesian modeling is termed as the hierarchical Bayes model. Example of a Hierarchical Bayesian model equation is

$$\psi(y, \theta, \gamma) = \psi(y|\theta) f(\theta|\gamma)\psi(\gamma) . \tag{3}$$

It is full joint distribution of all quantities. The inference about θ and γ is simply obtained through their joint posterior distribution

$$\psi(\theta,\gamma|y) = \frac{\psi(\theta,\gamma,y)}{\psi(y)} = \frac{\psi(y|\theta)\psi(\theta|\gamma)\psi(\gamma)}{\iint \psi(y|\theta)\psi(\theta|\gamma)\psi(\gamma)d\theta d\gamma}. \quad (4)$$

And inference about γ is given by its marginal posterior

$$\psi(\gamma|y) = \int \psi(\theta,\gamma|y)d\theta \propto \psi(\gamma)\int \psi(y|\theta)\psi(\theta|\gamma)d\theta. \quad (5)$$

We notice that γ is un-identifiable. The likelihood involves only θ, so prior belief about γ is influenced by the data y only through θ.

5 Model Formulation

Let T be a recovery time with density function ψ and non-recovery function S_r. Define a random censoring time C with distribution function G having probability density function g and non-recovery function S_g. Each unit has a recovery time T_i and a censoring time C_i.We define n independent and identically distributed random pair (Y_i, φ_i), where

$$Y_i = Min(T_i, C_i) \quad (6)$$

$$\varphi = \begin{cases} 1, for\ T_i \le C_i \\ 0, for\ C_i < T_i \end{cases} \quad (7)$$

The times T_i and C_i are assumed to be independent [14]. Let ϑ and S_v be distribution and non-recovery function. By independence assumption, it easily follows that the non-recovery function is

$$S_v(y) = [S_r(y)][S_g(y)]. \quad (8)$$

The distribution function of Y is

$$v(y) = 1 - S_r(y)S_g(y). \quad (9)$$

The likelihood function for the n independent and identically distributed random pairs (Y_i, φ_i) is

$$L = \left[\prod_{i=1}^{n}[S_g(y_i)]^{\varphi_i}[g(y_i)]^{1-\varphi_i}\right]\left[\prod_{i=1}^{n}[\psi(y_i)]^{\varphi_i}[S_r(y_i)]^{1-\varphi_i}\right] \quad (10)$$

If C, is not parameterized, then the first factor plays role in the maximization process, hence the likelihood function can be taken to be

$$L = \prod_{i=1}^{n}[(\psi(y_i)]^{\varphi_i}[S_r(y_i)]^{1-\varphi_i}. \quad (11)$$

It has the same form as the likelihood derived for right and left censoring [10]. The joint pdf for the pair,(Y, φ) is given by

$$\mathrm{P}(\mathrm{Y}, \varphi) = [(g(y))(S_r(y))]^{1-\varphi}[(\psi(y))(S_g(y))]^{\varphi}. \quad (12)$$

6 Weibull Regression Model

The time T_i for recovery, is distributed as Weibull with density function

$$\psi(t|\theta, \lambda) = \theta\, \lambda t^{\theta-1}\exp\left(-\lambda t^{\theta}\right), \quad (13)$$

where $\mathrm{t} > 0$, the shape parameter $\theta > 0$, and $\lambda > 0$, is the scale parameter. For Weibull density, we have a vector of covariate that affect t. Let the covariate be Z, where Z' is a $(p \times 1)$ vector

$$\lambda_i = \exp\left(\beta_1 Z_1 + \beta_2 Z_2 + \cdots + \beta_p Z_p\right) = \exp\left(\mathrm{z_i^T}\beta\right). \quad (14)$$

Density function for modeling event time is the product of the hazard function and the non-recovery function. The parameterization require modeling function for the mean of the distribution through λ. Covariate is link through λ and the parameter θ provides the shape of the distribution. Considering the change of parameter, we define the density for the recovery time with covariate Z as

$$\psi(t_i|\theta, z, \beta) = \theta \exp(Z_i^T \beta)\, t^{\theta-1} \exp\left(-\exp(Z_i^T \beta)\, t^{\theta}\right), \quad (15)$$

where the vector β is the co-efficient of regression parameters for covariates Z, t is the recovery time, and θ is the shape parameter. The hazard function, is given by

$$h_r(t) = \frac{\theta \exp\,(Z_i^T \beta)\; t^{\theta-1} \exp\,(-\exp(Z_i^T \beta)\, t^{\theta})}{\exp\,(\exp(Z_i^T \beta)\, t^{\theta})}. \quad (16)$$

The non-recovery function for each person at time t is express as

$$S_r\,(t|z_i, \theta) = \exp\left(-\exp(z_i^T \beta)\, t^{\theta}\right), \quad (17)$$

after substitution, we obtain the non-recovery function as we link equation (15), (17) into equation (12) the joint distribution becomes

$$\psi(t|z, \beta, \theta) = \theta^{\varphi} \exp\,\varphi(z_i^T \beta) t^{\varphi(\theta-1)} \exp\left(-\exp(z_i^T \beta)\, t^{\theta}\right). \quad (18)$$

7 Constructing Jeffreys Prior $\boldsymbol{\pi^J(\theta)}$ for Bayesian HW-R Model

We have two methods for constructing prior for the shape parameter of Bayesian HW-R model, in first method, we modified the work of [6]) and in the second method, we establish another method for constructing Jeffrey's prior for the shape parameter of a Weibull distribution.

Method I

Jeffrey's prior as derived in [6] shows the joint distribution of both the scale and shape parameter. We will modify the prior, adopting some assumptions. Jeffreys prior for the shape parameter θ is obtained by determining the Marginal density from the joint Jeffreys prior distribution for scale and shape parameter (θ, λ), secondly we assume Independence between the shape and scale parameter (θ, λ). The joint prior for two-parameter Weibull distribution is

$$\psi^J(\theta,\lambda) \propto \sqrt{detI(\theta,\lambda)} \tag{19}$$

$$I(\theta,\lambda) = -E_{(\theta,\lambda)}\begin{bmatrix} \frac{\partial^2 \ln L(\psi(t|\theta,\lambda)}{\partial\theta^2} & \frac{\partial^2 \ln L(\psi(t|\theta,\lambda)}{\partial\theta\partial\lambda} \\ \frac{\partial^2 \ln L(\psi(t|\theta,\lambda)}{\partial\theta\partial\lambda} & \frac{\partial^2 \ln L(\psi(t|\theta,\lambda)}{\partial\lambda^2} \end{bmatrix}, \tag{20}$$

where $\mathrm{I}(\theta,\lambda)$ is the Fisher Information matrix (covariance of the score function), we derive Jeffreys prior as follows. Since we are considering the model with correction for censoring, we will determine the Jeffreys prior from the general representation with likelihood function

$$L\big(\psi(t|\lambda,\theta)\big) = \prod_{i=1}^{n}\Big[\big(\theta\lambda t^{\theta-1}\exp\left(-\lambda t^{\theta}\right)\big)^{\varphi_i}\big(\exp\left(-\lambda t^{\theta}\right)\big)^{1-\varphi_i}. \tag{21}$$

Before we determine the prior, we need to form Fisher's Information matrix. We derive Fisher's Information matrix from equation (21) by taking the second order partial derivative of log-likelihood function for each parameter forming the matrix

$$I(\theta,\lambda) = \begin{bmatrix} \frac{1}{\theta^2} & \frac{In\theta + \Gamma'(2)}{\theta\lambda} \\ \frac{In\theta + \Gamma'(2)}{\theta\lambda} & \frac{\sum \varphi_i + (In\theta)^2 + \Gamma''(2) + 2In\theta\Gamma'(2)}{\lambda^2} \end{bmatrix}$$

$$= \frac{\sum \varphi_i + (In\theta)^2 + \Gamma''(2) + 2In\theta\Gamma'(2)}{(\lambda\theta)^2} - \left[\frac{In\theta + \Gamma'(2)}{\theta\lambda}\right]^2$$

$$= \frac{\sum \varphi_i + \Gamma''(2) - (\Gamma'(2))^2}{(\theta\lambda)^2} = \frac{\sum \varphi_i + r}{(\theta\lambda)^2}, \tag{22}$$

where

$\Gamma'(2) = \int_0^\infty yIn(y)\exp(-y)\,dy$, $\Gamma''(2) = \int_0^\infty yIn^2(y)\exp(-y)\,dy$

$r = \Gamma''(2) - (\Gamma'(2))^2$.

Since we are only interested in the value of the prior for θ we ignore the constant term and use the kernel function for the prior distribution

$$\psi^J(\theta,\lambda) = \sqrt{\frac{\sum \varphi_i + r}{(\theta\lambda)^2}} \propto \frac{1}{\theta\lambda}. \tag{23}$$

We will derive the density for the shape parameter θ by looking for its marginal distribution. Using Bayes theorem for parametric inference

$$\psi^J(\theta|\lambda) = \frac{\psi^J(\theta,\lambda)}{\psi(\lambda)}, \tag{24}$$

where $\psi^J(\theta,\lambda)$ is the joint prior distribution in equation (23), and $\psi(\lambda)$ is the marginal distribution for scale parameter, $1/\lambda$, since the distribution of the parameters are conditionally independent. The distribution of $\psi^J(\theta)$ given by

$$\psi^J(\theta) \propto \frac{1}{\theta} \tag{25}$$

$$\psi^J(\theta) = \frac{k}{\theta}. \tag{26}$$

Finally, Jeffreys prior for shape parameter of a Weibull density is Bernoulli distribution with parameter zero and one 'B $(0,1)$'.

Method II

On expanding equation (42), the log-likelihood function, is given by

$$\ln L(\psi(t|\theta,\lambda,\varphi) = \left(\sum_{i=1}^{m}\varphi_i\right)\log\theta + \left(\sum_{i=1}^{m}\varphi_i\right)\log\lambda - \lambda\sum_{i=1}^{m}t_i\varphi_i - \lambda\sum_{i=1}^{m}t_i(1-\varphi_i) + (\theta-1)\varphi_i\sum_{i=1}^{m}\log t_{i.} \tag{27}$$

The Fisher Information for this likelihood, is obtained by taking the expectation of the second order partial derivative of the log-likelihood function with respect to each parameter, which is shown as

$$I(\theta) = -E\left[\frac{\partial^2 \ln L(f(t|\theta,\lambda))}{\partial\theta^2}\right] = -E\left[-\frac{\sum_{i=1}^{m}\varphi_i}{\theta^2}\right] = \left(\frac{\sum_{i=1}^{m}E(\varphi_i)}{\theta^2}\right) \propto \frac{1}{\theta^2}. \tag{28}$$

Before we continue with the derivation, we need to establish a theorem that will enable us to get the expected value of the indicator variable φ_i

Theorem 1

Let $\{y_r\}_{r=1}^{m}$ be a finite sequence of sampled patients having attribute d with survival times $\{t_r\}_{r=1}^{m}$, and random censoring times $\{c_r\}_{r=1}^{n}$ with censoring indicator$\{\varphi_r\}_{r=1}^{m}$. We define a probability space $D = \{\Omega, \varphi, P(\varphi)\}$, with $P(\varphi)$, a probability measure satisfying $0 \le P(\varphi) \le 1$, and $\sum P(\varphi) = 1$. For m number of patients there will be n censored patients and $(m-n)$ uncensored patients with the following characterization.

$$E(\varphi) = (m-n)P$$
$$Var(\varphi) = (m-n)P(1-P).$$

Proof:

Let $P(\varphi_y)$ be the probability that y^{th} individual is censored and $\left(1-P(\varphi_y)\right)$ be the probability of being uncensored. φ_y , is distributed as Bernoulli with the parameter P with probability mass function

$$P(\varphi_y = r) = P^{\varphi_y}(1-P)^{1-\varphi_y} \text{ for } r, \varphi = 0,1 \tag{29}$$

The first characterization is the mean which is derive as follows

$$E(\varphi_y) = \sum_{y=1}^{m}\varphi_y P(\varphi_y = r). \tag{30}$$

Since there are n number of censored observations and $(m-n)$ number of uncensored observation the indicator variable assumes values

$$\varphi = \begin{cases} 1, & \text{if uncensored} \\ 0, & \text{if censored} \end{cases}$$

Equation (30) becomes

$$E(\varphi) = \sum_{y=1}^{m} \varphi_y P(\varphi_y = r) = (m-n)P, \tag{31}$$

where

$$y,\ x = 1,2,\dots,m \text{ , for } \quad P_y = P_x.$$

The second prove is obtain directly using the variance of a Bernoulli distribution

$$V(\varphi) = \sum_{i=1}^{m} \varphi_y^2 P(\varphi_y) - \left(\sum_{i=1}^{m} \varphi_y P(\varphi_y) \right)^2 = (m-n)P(1-P). \tag{32}$$

Applying Theorem 1 to equation (28) we get

$$V(\varphi) = \frac{\sum_{i=1}^{m} E(\varphi_i)}{\theta^2} = \frac{\sum_{i=1}^{m} (m-n)P_y}{\theta^2} \propto \frac{k}{\theta^2}\ , 0 < k < h. \tag{33}$$

$$h \geq m+n$$

The Fisher information $I(\theta)$ is finally, $I(\theta) \propto k/\theta^2$ and Jeffreys prior becomes

$$\psi^J(\theta) \propto \left[\frac{k}{\theta^2} \right]^{1/2} = \frac{k^{1/2}}{\sqrt{\theta^2}} \propto \frac{1.}{\theta} \tag{34}$$

This result is the same with the modified version of Alomari et al (2012)

8 Analysis of Bayesian HW-R Model using Jeffrey's prior $\boldsymbol{\pi^J(\theta)}$

In Bayesian analysis, the basic concept is the determination of posterior density using the Bayes method of parametric inference. The posterior density for Bayesian HW-R model using Bayes theorem is given by

$$\psi(\theta, \beta, \phi | t) \propto L(\psi(t|\theta, \beta, \eta)\ \psi(\beta)\psi^J(\theta)\psi(\phi) \tag{35}$$

$$= \frac{L(\psi(t|\theta, \beta, \eta)\ \psi(\beta|\mu, \Phi)\psi^J(\theta) f(\phi)}{\iint L(\psi(t|\theta, \beta, \eta)\ \psi(\beta|\mu, \Phi)\psi^J(\theta)\psi(\phi) d\theta d\beta}. \tag{36}$$

With prior specification for hyperparameters. The first one is the gamma prior for parameter ϕ with density function

$$\psi(\phi) = \frac{b^a}{\Gamma(a)} exp(-b\phi). \tag{37}$$

Secondly is the normal or Gaussian prior distribution for the parameter of predictor variables β.

$$\psi(\beta) = \frac{\exp - \left(\frac{1}{2}(\beta - \hat{\beta})^T \Phi^{-1}(\beta - \hat{\beta})\right)}{(\sqrt{2\pi})|\Phi|^{-1/2}}. \tag{38}$$

Finally, the prior distribution for the shape parameter of Weibull distribution, which is derived using Jeffreys rule. It is given, as

$$\psi^J(\theta) = \frac{k^{1/2}}{\theta}, \qquad 0 \le k \le h. \tag{39}$$

The posterior density and marginal in equation (35) after substitution and Simplification becomes

$$\psi(\theta, \beta, \phi|t) \propto \theta^{r-1} \exp\left(\sum_{i=1}^{n}\left(\varphi_i + z_i^T\beta + \varphi_i(\theta-1)Ln(t_i)\right) - t_i^{\theta}\exp(z_i^T\beta) - \frac{1}{2}(\beta - \hat{\beta})^T\Phi^{-1}(\beta - \hat{\beta}) - a\phi\right). \tag{40}$$

8.1 Kernel for posterior and marginal densities

Kernel for Posterior density

$$\psi(\theta, \beta, \phi|t) \propto \theta^{r-1} exp\left(\sum_{i=1}^{n}\left(z_i^T\beta + \varphi_i(\theta-1)Ln(t_i)\right) - t_i^{\theta}\exp(z_i^T\beta) - \left(\frac{1}{2}(\beta - \hat{\beta})^T\Phi^{-1}(\beta - \hat{\beta})\right) - a\phi\right). \tag{41}$$

Posterior marginal for θ in Kernel form

$$\psi(\theta|\beta, \phi, t) \propto \theta^{r-1} exp\left[(\theta-1)\sum_{i=1}^{m}\varphi_i \ln t_i - t_i^{\theta} e^{(z_i^T\beta)}\right]. \tag{42}$$

Posterior Marginal for ϕ in Kernel form

$$\psi(\phi|\beta, \theta, t) \propto exp(-a\phi) \tag{43}$$

Posterior Marginal for β in Kernel form

$$\psi(\beta|\theta, \phi, t, z) \propto exp\left\{-\tfrac{1}{2}(\beta - \hat{\beta})^T\Phi^{-1}(\beta - \hat{\beta}) + \sum z_i^T\beta\right\}. \tag{44}$$

9 Parameter Estimation

Parameter estimation is one of the most difficult aspect Bayesian analyses more especially complex models that the prior distribution is not conjugate and class of hierarchical models which sometimes MCMC method and Numerical Integration cannot give accurate estimates due to the hierarchical structure. Lindley (1980) developed a method for approximating posterior distribution for integral of the form

$$\mathrm{I} = \frac{\int U(\lambda, \theta)\exp[\ln L(\psi(\lambda, \theta)) + \rho(\lambda, \theta)]\, d\theta}{\int \exp[\ln L(\lambda, \theta) + \rho(\lambda, \theta)]\, d\theta}, \tag{45}$$

where $U(\lambda, \theta)$, is the function of λ and θ only, $\ln L(\psi(\lambda, \theta))$ is the log likelihood and $\rho(\lambda, \theta) = \ln L[\psi(\lambda, \theta)]$. Lindleys method evaluates this form of integral.

9.1 Laplace Approximation

The Laplace method approximate the log-likelihood of the posterior density as a quadratic form with mode $(\hat{\theta})$. The method approximates equation (45) using the Taylor's series expansion on $\psi^*(\theta)$.The expansion yielded

$$\ln[L(\psi^*(\theta))] \approx b - \frac{1}{2}(\theta - \hat{\theta})^T H_1(\theta - \hat{\theta}) \tag{46}$$

$L(\psi^*(\theta))$, is the likelihood function for density $\psi^*(\theta)$ evaluated at θ, b is a constant and H_1 is defined as

$$H_1 = -\frac{\partial^2 \ln[L(\psi^*(\theta))]}{\partial \theta^2} = -\nabla^2{}_1 \ln[L(\psi^*(\theta))] \;\; at\ \theta = \hat{\theta}.$$

The posterior density $f^*(\theta)$, is approximated by an un-normalized Gaussian density

$$\psi(\theta) \equiv \psi^*(\theta) exp\left(-\frac{\nabla^2}{2}\ln[L(\psi^*(\theta))](\theta - \hat{\theta})^T(\theta - \hat{\theta})\right). \tag{47}$$

The normalizing constant for Gaussian density is approximated using normalizing constant of this Gaussian:

$$C \equiv \psi^*(\theta)\sqrt{\frac{2\pi}{\nabla^2 \ln[L(f(\theta))]}}. \tag{48}$$

9.2 Quasi-Newton Optimization

Quasi-Newton method is one of the ideas in solving non-linear optimization problem. It is like the steepest descent that require only the gradient of the objective function that is good enough to produce super-linear convergence. It can solve unconstraint, constraint and large scale optimization problems. The Quasi-Newton method is performed using the following steps. Consider a non-linear objective function

$$\psi(\theta_\sigma, \omega) = a{\theta_\sigma}^r + b{\theta_\sigma}^{r-1} + \cdots + q + u\omega^t + v\omega^{t-1} + \cdots + z \tag{49}$$

where $a, b, \ldots \in \mathbb{R}$, and θ_σ, ω are variables

Given a starting point $\theta_{(0)} \in \text{domain } \psi \text{ for } \theta = (\theta_\sigma, \omega)$
Compute the Quasi-Newton direction $\Delta\theta = -H_{k-1}^{-1}\psi'(\theta_{k-1})$
Determine the step size t_k
Compute $\theta_k = \theta_{k-1} + t_k\psi'(\theta_k)$
Compute H_k

Broyden, Fletcher, Goldferb, and Shanno (BFGS) for updating Hessian ($\mathbf{H_k}$)

BFGS is the popular Quasi-Newton method; it is use for updating the Hessian H_k. Following the model of the objective function at the current iterate θ_k, we use the finite difference approximation method

$$\psi(\theta_k + \Delta\theta) = \psi(\theta_k) + \Delta\theta_k \psi'(\theta_k) + \frac{\Delta\theta_k{}^T}{2!} H_k \Delta\theta_k + \cdots \tag{50}$$

H_k, is a $n \times n$ symmetric positive definite matrix and is revised or updated at every iteration. The function value and gradient of this model at $\Delta\theta_k = 0$ match ψ_k and $\Delta\psi_k$, the minimizer $\Delta\theta_k$ of this convex quadratic model is written as $\Delta\theta = -H_{k-1}^{-1}\psi'(\theta_{k-1})$, will be use as search direction and new iterate is

$$\theta_{k+1} = \theta_k + t_k \psi'(\theta_k), \tag{51}$$

where t_k is the step length, satisfying the Wolfe condition.
The approximate Hessian H_k is used in place of the true Hessian and is updated using

$$H_k = H_{k-1} + \frac{vv^T}{v^T s} - \frac{H_{k-1} s s^T H_{k-1}}{s^T H_{k-1} s}, \tag{52}$$

where

$$s = \theta_k - \theta_{k-1}, \qquad v = \Delta\psi(\theta_k) - \psi\Delta(\theta_{k-1})$$

The inverse update is obtain for $v^T s > 0$ and strictly convex function ψ is given by

$$H_k^{-1} = \left(I - \frac{sv^T}{v^T s}\right) H_{k-1}^{-1} \left(I - \frac{v^T s}{v^T s}\right) + \frac{ss^T}{v^T s}. \tag{53}$$

The BFGS update satisfies the secant condition as $H_k s > v$ i.e. $H_k(\theta_k - \theta_{k-1}) = \Delta\psi(\theta_k) - \psi\Delta(\theta_{k-1})$, and θ_{k+1} is obtain as

$$\theta_{k+1} = \theta_k - \frac{\psi'(\theta_k)}{H_k} \tag{54}$$

$$H_k = \frac{\psi'(\theta_k) - \psi'(\theta_{k-1})}{\theta_k - \theta_{k-1}}. \tag{55}$$

10 Integrated Nested Laplace Approximation (INLA) Method

Numerical approach in Bayesian analysis for Latent Gaussian models is discussed in [11] and [15]. This approach makes the Laplace approximations very accurate when applied to Latent Gaussian Model (LGM).

Numerical computation of the Latent Gaussian Model
The posterior density $\psi(x, w|t)$ for Gaussian Latent Field x and hyperparameters w conditional on t is given by the expression

$$\psi(x, w|t) \propto \psi(w)\psi(x|w) \prod_{i=1}^{n} \psi(t_i|x_i), \tag{56}$$

$\psi(x_i|t)$: The posterior marginal for Latent Gaussian Field for all i
$\psi(w_i|t)$: Posterior for the marginal, for hyperparameters, which is not Gaussian.

Numerical computation of posterior marginal
The posterior marginal for the hyper-parameters, conditionally on time t is express as

$$\psi(w|t) \propto \frac{\psi(x,t,w)}{\psi(x|t,w)}. \tag{57}$$

We locate the mode in log scale and the posterior marginal will be interpolated, and we use the Gaussian Approximation to the conditional for the marginal for hyper- parameters

$$\psi(w|t) \propto \frac{\psi(x,w|t)}{\psi_G(x|w,t)} \quad at\ x = x\text{mode}(w). \tag{58}$$

For the Latent field, the marginal for $x_i|t$ we do similarly

$$\psi(x_i|t) \approx N_c \frac{\psi(x,t|w)}{\psi_G(x_{-i}|x_i,t,w)} \quad at\ x_{-i} = x\text{mode}(w,x_i). \tag{59}$$

Taking $\psi_G(x_{-i}|x_i,t,w)$ as the Gaussian Approximation which gives us Laplace Approximation

$$\psi(x|t,w) \propto exp\left(-\frac{1}{2}x^T\Phi x + \sum_{i=1}^{n} \ln L(\psi(t_i|x_i))\right)$$

$$\psi_G(x|w,t) \approx exp\left(-\frac{1}{2}(x-\mu)^T\left(\Phi + diag\left(\Phi_{(\theta,\phi)}\right)\right)(x-\mu)\right). \tag{60}$$

Implementing INLA for Bayesian HW-R model

Consider the linear predictor of a structured additive model

$$\eta = \beta_0 + \beta_1 Z_1 + \cdots \beta_n Z_n + \phi\sigma + \epsilon, \tag{61}$$

where Z's are the covariates

β_0 = Intercept for the linear predictor model
$\beta's$ = linear effects of covariates Z (fixed)
σ = The variable accounting for frailty term in the linear predictor
ϕ = is the co-efficient for frailty term in the model.
ϵ = is an unstructured term accounting for unexplained variation in the model

Computing Posterior marginal for $\psi(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{\beta}, \boldsymbol{t},)$ using Quasi-Newton method

In computing the posterior marginal for the kernel $\psi(\theta, \phi|\beta, t,)$, we first, locate the mode for $\ln\psi(\theta, \phi|\beta, t,)$, by optimizing with respect to θ using Quasi-Newton method which build up to an approximation to the second derivative of $\ln\psi(\theta, \phi|\beta, t,)$ using difference between successive gradient vectors and the gradient is obtain using finite difference.

The kernel is

$$\psi(\theta, \phi|\beta, t) \propto \theta^{r-1} exp\left\{(\theta-1)\sum_{k=1}^{m} \varphi_k \ln t_i - t_k^{\theta} e^{\eta} - a\phi\right\} \tag{62}$$

$$\ln\psi(\theta, \phi|\beta, t,) = (r-1)\ln\theta + \sum_{k=1}^{m} \varphi_k(\theta-1)\ln t_k - t_k^{\theta} e^{\eta} - a\phi.$$

We locate the mode for θ_k and ϕ_k by using the Quasi-Newton method. The quadratic model for the objective function, is given by the second order approximation

$$\mathcal{M}_k(\Delta\hat{\theta}_k) = \psi(\hat{\theta}_k) + \Delta\hat{\theta}_k\psi'(\hat{\theta}_k) + \frac{\Delta\hat{\theta}_k^{\,T}}{2!}H_k\Delta\hat{\theta}_k + \cdots \tag{63}$$

where $\hat{\theta} = (\hat{\theta}, \hat{\phi})$, $\Delta\hat{\theta} = (\hat{\theta} - \hat{\theta}_k)$ and $\mathcal{M}_k(\Delta\hat{\theta}_k)$ is the objective function.

Using finite difference, the Hessian is obtain using the formula

$$H_k(\theta) = \frac{\left(\frac{\partial \ln \psi(\hat{\theta}_k, \phi|\beta, t,)}{\partial\hat{\theta}_k} - \frac{\partial \ln \psi(\hat{\theta}_{k-1}, \phi|\beta, t,)}{\partial\hat{\theta}_{k-1}}\right)}{\theta_k - \theta_{k-1}}. \tag{64}$$

For $\hat{\theta}_k$

$$\frac{\partial \ln \psi(\hat{\theta}_k, \phi|\beta, t,)}{\partial\hat{\theta}_k} = \frac{r-1}{\hat{\theta}_k} + \sum_{k=1}^{m} \varphi_k \ln t_k - t_k^{\hat{\theta}_k} \ln t_k \, e^{\eta}. \tag{65}$$

For $\hat{\theta}_{k-1}$

$$\frac{\partial \ln \psi(\hat{\theta}_{k-1}, \phi|\beta, t,)}{\partial\hat{\theta}_{k-1}} = \frac{r-1}{\hat{\theta}_{k-1}} + \sum_{k=1}^{m} \varphi_k \ln t_k - t_k^{\hat{\theta}_{k-1}} \ln t_k \, e^{\eta}. \tag{66}$$

Using this we compute the Hessian for the objective function to be

$$H_k = \frac{\left\{r - 1\left(\frac{1}{\hat{\theta}_k} - \frac{1}{\hat{\theta}_{k-1}}\right) - e^{\eta}\left(t_k^{\hat{\theta}_k} \ln t_k + t_k^{\hat{\theta}_{k-1}} \ln t_k\right)\right\}}{\hat{\theta}_k - \hat{\theta}_{k-1}}. \tag{67}$$

The value of $\hat{\theta}_k$ can be iterated to get new value for $\hat{\theta}_{k+1}$ using

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \left\{\left(\frac{\partial \ln \psi(\hat{\theta}_k, \phi|\beta, t,)}{\partial\hat{\theta}_k}\right)/H_k\right\}, \tag{68}$$

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \frac{\frac{r-1}{\hat{\theta}_k} + \sum_{k=1}^{m} \varphi_k \ln t_k - t_k^{\hat{\theta}_k} \ln t_k \, e^{\eta}}{\left\{\frac{r - 1\left(\frac{1}{\hat{\theta}_k} - \frac{1}{\hat{\theta}_{k-1}}\right) - e^{\eta}\left(t_k^{\hat{\theta}_k} \ln t_k + t_k^{\hat{\theta}_{k-1}} \ln t_k\right)}{\hat{\theta}_k - \hat{\theta}_{k-1}}\right\}}. \tag{69}$$

We routinely get new values of Hessian and compute new values for $\hat{\theta}$ .Finally; we construct an interpolant and compute the marginal using numerical integration from this interpolant. We standardize values by converting $\hat{\theta}$ using

$$\hat{\theta}(z) = \hat{\theta} + A\,\lambda^{\frac{1}{2}}\,\mathrm{Z}, \tag{70}$$

where $\lambda^{\frac{1}{2}}$, A are obtained from Eigen decomposition of inverse Hessian obtained from the finite difference approximation and $\mathrm{Z}\sim\mathrm{N}(\mathbf{0}, \mathbf{I})$ which corrects for scale and rotation.

Computing posterior marginal for latent variables $\boldsymbol{\psi(x_i|t)}$

We use numerical integration to obtain the posterior marginal for x_i

$$\psi(x_k|t) = \sum_{i\epsilon t} \psi\left(x_k|\hat{\theta}_k, t\right) \psi\left(\hat{\theta}_k|t\right)\Delta_k \,, \tag{71}$$

which is the sum over the values of $\hat{\theta}_k$ with area weight Δ_k, we then use laplace approximation to provide the accurate approximation using the set of weighted points $\{\hat{\theta}_k\}$ to be use in the integration.

11 Assessing Model Performance

11.1 Deviance Information Criterion

The models, will be assessed using the deviance Information Criterion (DIC), the DIC is given by the expression

$$D(\theta) = -2Ln\mathcal{L}(\theta) + \mathcal{C}, \tag{72}$$

θ is the vector of unknown parameters of the model, $\mathcal{L}(\theta)$ is the likelihood function of the model and $\mathcal{C}$ is a constant that does not need to be known. The DIC define by [17] is well applied in the Bayesian model comparison, it has the form

$$\text{DIC } = \mathcal{D}\left(\hat{\theta}\right) + 2n_{\mathcal{D}}, \tag{73}$$

where $\mathcal{D}\left(\hat{\theta}\right)$, is the deviance evaluated at the posterior mean and $n_{\mathcal{D}}$ is the effective number of parameter for the model.

11.2 Kulback Leibler Divergence (KLD)

Kullback and Leibler introduce another information measure aside the Bayesian information and Akaike Information criterion to assess the distance between the true and approximated Bayesian posterior distribution. Let $f(\theta|y)$ be the posterior density for the true model and $g(\theta|y)$ be the probability density for the approximated model on the measurable space. The Kullback Leibler Divergence, is defined by the relation

$$I(\psi, g) = E\left[\frac{\psi(\theta|y)}{g(\theta|y}\right] = \int \log\left[\frac{\psi(\theta|y)}{g(\theta|y}\right]\psi(\theta|y)dy\,, \tag{74}$$

$I(\psi, g)$ is non-negative and it reach 0 when $\psi(\theta|y)$ is the same as $g(\theta|y)$ almost surely, and is interpreted as information lost when $g(\theta|y)$ is used to approximate $\psi(\theta|y)$. The smaller values of $I(\psi, g)$ we consider the model g to be closer to the true distribution $\psi(\theta|y)$

12 Results

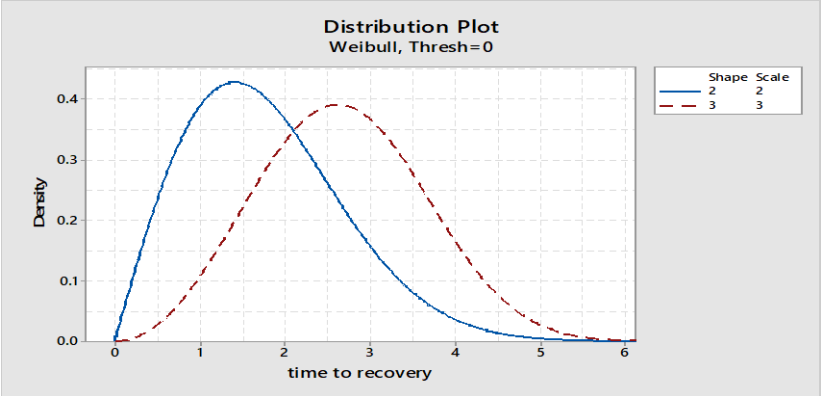

Figure 1a: Density Plot for Weibull distribution with Scale $\lambda = 2$ and Shape Parameter $\theta = 2$, Scale $\lambda = 3$ and shape $\theta = 3$

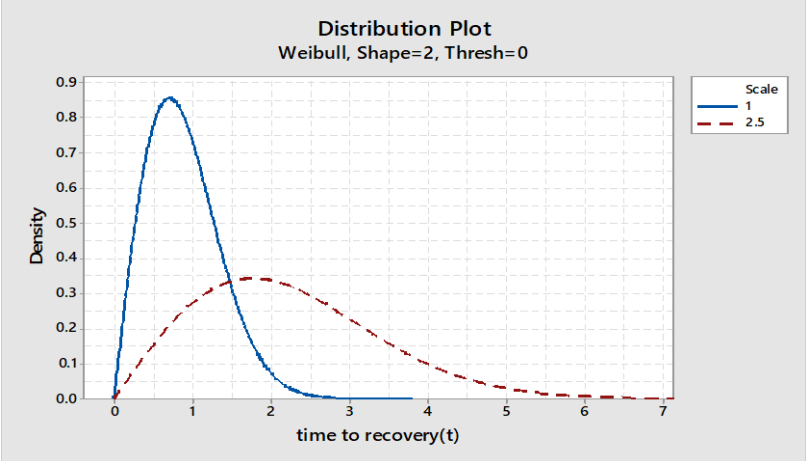

Figure 1b: Density Plot for Weibull distribution with Scale parameter $\lambda = 1,\ \lambda = 2.5$ with shape parameter $\theta = 2$

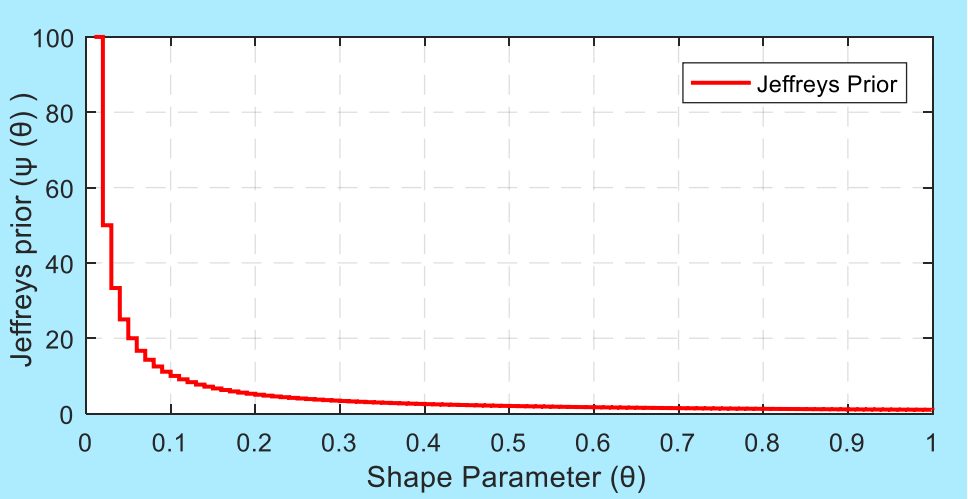

Figure 2: Density plot for Jeffrey's prior for shape parameter θ in the close interval $[\,0, 1]$

Specification for the drug precision ϕ

Variables	Mean	S.E	0.025 Quant	0.5Quant	0.975 Quant	Mode
β_0	-4.6920	0.4306	-5.5479	-4.6881	-3.8584	-4.6754
$\beta_{(Sex)}$	0.0093	0.1247	-0.2350	0.0091	0.2546	0.0087
$\beta_{(Age)}$	-0.0039	0.0063	-0.0165	-0.0039	0.0083	-0.0037
$\beta_{(Wght)}$	0.0081	0.0052	-0.0020	0.0081	0.0182	0.0081
$\theta_{(Shape)}$	2.52	0.1272	2.277	2.518	2.777	2.515

Table 1: INLA estimate for Bayesian HW-R using Jeffreys prior, $\psi^J(\theta) = 0.9$ and no prior

Variables	Mean	S.E	0.025 Quant	0.5Quant	0.975 Quant	Mode
β_0	-4.6815	0.4304	-5.5372	-4.6776	-3.8482	-4.6696
$\beta_{(Sex)}$	0.0095	0.1247	-0.2348	0.0092	0.2548	0.0088
$\beta_{(Age)}$	-0.0039	0.0063	-0.0165	-0.0038	0.0084	-0.0037
$\beta_{(Wght)}$	0.0081	0.0052	-0.0020	0.0081	0.0182	0.0081
$\theta_{(Shape)}$	2.512	0.1272	2.271	2.512	2.771	2.509

Table 2: INLA estimate for Bayesian HW-R using Jeffreys prior, $\psi^J(\theta) = 0.001$ with Gamma (1, 0.001) prior specification for the drug precision ϕ

Variables	Mean	S.E	0.025 Quant	0.5Quant	0.975 Quant	Mode
β_0	-4.9686	0.4339	-5.8309	-4.9647	-4.1285	-4.9567
$\beta_{(Sex)}$	0.0061	0.1247	-0.2382	0.0059	0.2514	0.0054
$\beta_{(Age)}$	-0.0042	0.0063	-0.0168	-0.0042	0.0080	-0.0041
$\beta_{(Wght)}$	0.0085	0.0052	-0.0017	0.0085	0.0186	0.0084
$\theta_{(Shape)}$	2.673	0.1273	2.425	2.671	2.934	2.668

Table 3: INLA estimate for Bayesian HW-R with no prior specification for θ.

Variables	Mean	S.E	0.025 Quant	0.5Quant	0.975 Quant	Mode
β_0	-5.5796	0.4429	-6.4593	-5.5758	-4.7217	-5.5681
$\beta_{(Sex)}$	-0.0029	0.1248	-0.2472	-0.0031	0.2426	-0.0035
$\beta_{(Age)}$	-0.0049	0.0063	-0.0175	-0.0048	0.0074	-0.0047
$\beta_{(Wght)}$	0.0092	0.0052	-0.0010	0.0092	0.0194	0.0092
$\theta_{(Shape)}$	3.007	0.1334	2.752	3.005	3.276	3.001

Table 4: INLA estimate for Bayesian HW-R using Jeffrey's prior Jeffrey's prior $\psi^J(\theta) = 80$ with Gamma (1, 0.001) prior specification for the drug precision ϕ.

Variables	Mean	S.E	0.025 Quant	0.5Quant	0.975 Quant	Mode
I	18643.4	1.842E+04	1282.9	13233.4	67312.8	3490.4
II	976.60	938.94	75.43	705.00	3563.90	211.35
III	18691.46	1.846E+04	1286.54	13258.45	67506.76	3500.54
IV	992.14	957.67	77.717	714.82	3529.05	217.49

Table 5: Hyper-Parameter (Random effect component) for drug precsion ϕ , using Gamma (1, 0.001) Prior for Model I, II, IV and No prior for Model III.

Model	Marginal Log-likelihood	DIC	Rank	KLD
I	-642.38	1122.23	3	0
II	-652.98	1122.72	4	0
III	-596.55	1116.17	2	0
IV	-576.73	1111.35	1	0

KLD: Kullback Leibler Divergence DIC: Deviance Information Criterion

Table 6: Assessing Model performance using the Deviance Information Criterion, Marginal Log- likelihood and Kulback Leibler Divergence

13 Summary and Conclusion

The study proposed a Bayesian HW-R model using Jeffreys prior on random censored data. The model was tested using data for recovery times of patients tested positive of malaria and where administered drugs. The study, also determine the posterior density for the proposed model and established a new method for obtaining Jeffreys prior for random censored data. Also, modify the work of Alomari [6] and obtain the Jeffreys prior for shape parameter since the scale parameter is an exponential function of the predictors, which was thus, incorporated forming a structured additive model for the linear predictor through log link function.

The predictor's influence the recovery time only through the predictor. The model used is termed the Latent Gaussian Model. Consider the result in Table1, $\psi^J(\theta) = 0.9$, the posterior mean with this prior yield an estimate of $\theta = 2.52$, also the result in table 2 for prior $\psi^J(\theta) = 0.001$, has posterior mean of $\theta = 2.514$. The prior in table 1 is higher than the prior in table 2 as such the posterior estimate for θ was influenced by the choice of θ in the θ-coordinate. When the prior is non-informative, the likelihood takes part in maximizing the posterior density. Table 3 has no prior specification for the shape parameter θ , the estimated value is $\theta = 2.673$. It is clear that the prior in table 2 decreases the posterior estimate from 2.673 to 2.52 for $\psi^J(\theta) = 0.9$ and 2.514 for $\psi^J(\theta) = 0.001$, but the distribution is Rayleigh for value of θ within the range of 2, the estimate without prior specification uses only the data at hand that's why the estimate is 2.673. Table 4 is exceptional because when we specified the prior to be high in θ-coordinate, $\psi^J(\theta) = 80$, the posterior for θ increases suddenly to the peak with posterior

estimate $\theta = 3.007$ which approximately changed the density to normal, while the other three follows the Rayleigh density as shown in fig1 and fig1.b.

When the value for $\psi^J(\theta)$ in θ-coordinate increases, the posterior estimate for θ increases and change the shape of the distribution. For prior density $\psi^J(\theta)$ with censoring, the posterior density attains normality with increase in prior for shape parameter θ in θ-Space. The assessment of covariate effect with normal prior shows small ignorable differences in the parameter for table 1 and 2. Table 4 shows a very large difference because of high prior updating. In addition, the random component of the predictor, η of the structured additive class of Latent Gaussian Model assumes gamma (1, 0.001) prior. The posterior for drug precision ϕ in model I and III is not large and model II and IV shows no large difference in estimating the drug precision estimate. Finally, assessing model performance, the Jeffreys prior with increase in $\psi^J(\theta)$ in θ-space increases the posterior estimate for θ. Performance evaluation was done using the deviance Information Criterion (DIC) and Kullback Leibler Divergence (KLD). Using this two approaches model IV is the best with minimum DIC and KLD shows that the posterior density converges to the true density with KLD=0.

References

[1] Haile S. (2015). Weibull accelerated failure time regression function in R. R Documentation, St. Gallon 4pp.
[2] McGilchrist C.A. and Aisbett C.W. (1991). Regression with frailty in survival analysis. Biometrics, Vol. 47:461-466.
[3] Dellaportas P and Smith A.F.M. (1993). Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. Applied Statistics, Vol. 42(3):443– 459.
[4] Santos C.A. and Achcar J.A. (2011). A Bayesian analysis in the presence of covariates for multivariate survival data, an example of application. Revista Columbiana de Estadistica, Vol. 34:111-131
[5] Guure C.B. and Ibrahim N.A. (2014). Bayesian analysis of the survival function and failure rate of Weibull distribution with censored data. Maths Probl Eng, 18pp.
[6] Alomari M.A., Arasanj N.A., and Adam M.B. (2012). Bayesian Survival and hazard estimate for Weibull censored time distribution. Journal of Applied Sciences, Vol. 12:1313-1317.
[7] Aslam M., Mohsin S.A.K., Ahmad I., and Haider, S.S. (2014). Bayesian estimation for parameters of the Weibull distribution. Science Int. (Lahore), Vol. 26(5):1915-1920.
[8] Vasile P., Eugenia P., and Alina C. (2010). Bayes estimators of modified Weibull distribution parameters using Lindley's approximation. WSEAS Transactions on Mathematics, Vol. 9: 539-549.
[9] Rue H. and Martino S. (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random fields models. J. Statist. Planng. Inf., Vol. 137: 3177–3192.
[10] Akerkar R. (2012). Investigating posterior contour probabilities using INLA: A case study on recurrence of bladder tumors, Norway. http://www.math.ntnu.no/preprint/statistics/2012/s4-2012.pdf, 07.12.2016.
[11] Rue H., Martino S., and Chopin N. (2009). Approximate Bayesian inference for latent gaussian models using integrated nested Laplace approximation. University of Science and technology Norway, Journal of Royal Statistical Society Ser B, Vol. 71(2):319-392.

[12] Martino S., Akerkar R., and Rue H. (2010). Approximate Bayesian Inference for survival models. Scandinavian Journal of Statistics, Vol. 28 (3):514-528.
[13] Eaton M.L. and Sudderth W.D. (2000). Invariance of posterior distributions under Reparameterization. Indian Journal of Statistics, Special issue in memory of A, Maitra.
[14] Siannis F., John C., and Guobing L.U. (2005). Sensitivity analysis for informative censoring in parametric survival models, Journal of Biostatistics, Vol. 1(6):77-91.
[15] Rue H. and Held L. (2005). Gaussian Markov Random fields, Theory and Application, monographs on statistics and applied probability, Chapman and Hall London, 104.
[16] Kass R.E. and Wasserman L. (1996). The selection of prior distribution by formal Rules. Journal of the American Statistical Association, Vol. 96: 1343–1370.
[17] Spiegelhalter D.J., Best N.G., Carlin B.P. and van der Linde A. (2002). Bayesian measures of model complexity and fit (with discussion). J. R. Statist. Soc. B, Vol. 64: 583–639.