Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/110319

Bayesian Survival Analysis of Acute Encephalitis Syndrome with Censoring Mechanism using Brms Package

*Shazia Farhin** *and Athar Ali Khan*

Dipartiment of Statistics and Operation Reearch, Aligarh Muslim University, Aligarh, Uttar Pradesh, 202002, India

Received: 2 May 2020, Revised: 28 Oct. 2020, Accepted: 3 Nov. 2020
Published online: 1 Sep. 2022

Abstract: Acute encephalitis syndrome(AES) most commonly affects children and young adults and can lead to considerable morbidity and mortality. In June 2019, the outbreak of acute encephalitis syndrome occurred in Muzaffarpur district and their neighbouring district of Bihar. This paper presents the Bayesian survival analysis of AES data of the Muzaffarpur district. AES data extracted from the SKMCH and KM hospital of Muzaffarpur. The Weibull, Log-normal, and Exponential, these survival models have been used for fitting of AES data with the help of `brms` packages of `R` and compared these models with the Leave one out cross-validation. `brms` package uses the Hamiltonian Monte Carlo(HMC) sampler and its extension, no-U-turn sampler (NUTS) algorithm of MCMC, for the simulation study. In addition, the Logistic regression model is used to predict the risk of death on the basis of observed characteristics or covariates.

Keywords: Acute Encephalitis syndrome data, Bayesian Inference, Survival models, brms package, Leave one out cross validation.

1 Introduction

Acute Encephalitis Syndrome(AES) characterized as acute-onset of fever and a change in mental status(mental confusion, disorientation, delirium or coma) or new-onset of seizures in a person of any stage at any time of the year. Early symptoms include headaches and vomiting along with sudden hypoglycaemia(drops in blood sugar level) but may lead to coma, brain dysfunctions, and inflammation of the heart and lungs [1,2]. Severe hypoglycaemia can cause death [3]. AES most commonly affects children and young adults and can lead to considerable morbidity and mortality. AES can be caused by different microorganisms, including viruses, bacteria, fungi, parasites, and spirochaetes as well as chemicals and toxins [4]. In India, AES was chiefly associated with the japanese encephalitis virus(JEV) [1] and although with chandipura virus(CHPV), Nipah Virus and entro-viruses. For AES, high temperature, humidity, malnutrition and poor hygiene are known as a more serious factor.

This paper focuses on the Bayesian survival analysis of AES data of Muzaffarpur district of Bihar. Since 1st June 2019, the temperature in Muzaffarpur has remained above 40°C, and Poverty & Malnutrition are widespread among children in the region. Malnourished children lack a buffer stock of sugar as glycogen in the liver, which puts them at a higher risk of hypoglycaemia. Muzaffarpur is the largest producer of lychee fruits. It was said that the outbreak of AES occurs due to lychee fruits, because of a heavy diet of unripe lychee fruits without having a full meal in a day later may be put malnourished children at risk of hypoglycaemia [3]. In June 2019, an outbreak occurred in 222 blocks of Muzaffarpur district and neighbouring district of Bihar. We have extracted a real AES data from Sri Krishna Medical College and Hospital(SKMCH), the largest state-operated hospital in Bihar and Kejriwal Matrisadan(KM), a trust run hospital. This data set contains 173 observations of children of age less than 12 years, they are those children who were suffered from AES. In 173 observations, 21 observations are from SKMCH and 152 observations are from KM hospital.

The main objective of this paper is to fit the parametric survival models like Weibull, Log-normal, and Exponential to the

* Corresponding author e-mail: shaziafarhin18@gmail.com

AES data in the Bayesian environment with the `brms` package [5], and compared these models with each other using leave one out cross-validation(LOO-CV) criteria [6,7]. `brms` implements the Bayesian regression model in `R` [8] using probabilistic programming language `Stan` [9]. `Stan` implements Hamiltonian Monte Carlo [10] sampler and its extension no-U-turn sampler [11]. Whether the prior are conjugate or not, these algorithms of MCMC converge much more quickly to the target distribution as compared to the Metropolis Hastings algorithm[12,13] and Gibbs sampler[14, 15]. Along with these things, we have made the comparison of non-parametric estimation of survival function using the Kaplan-Meier estimator to the parametric survival models with the help of graphs. Besides this, we have also fitted the logistic regression model to the AES data with the Bayesian approach to predict the risk of death based on the observed characteristics of the patient.

2 Description of AES data

The AES data is nothing but right censored data, in which 77% observations are censored. Discharge, leave against medical advice(LAMA), and refer children are considered to be censored. The data frame contains 173 rows and seven columns, and the columns are the Age, Gender, District, Hospital, Symptom, Time, and Indicator. In which, approx 57% are girl child patients, 43% are boy child patients, most of the children are of age less than or equal seven years (approx 87%), and the rest of children are above seven years but below 12 years. The majority of children belong to the Muzaffarpur district of Bihar(approx 84%), and the rest of the children belongs to other district e.g., Sitamadhi, Vaishali, Champaran, and Darbhanga. The symptom hypoglycaemia is a major cause of outbreak of AES, about 97% of children had symptom of hypoglycaemia and rest of AES suffered children had symptom of hyper-pyrexia.

```
Description of data as given below;
Age       : Age of children at time of disease occur in years.
Gender    : Sex of children; 1=Female, 0=Male.
District  : Name of district; 1=Muzaffarpur, 0=Others, others could be Vaishali, Sitamadhi, Champaran, Darbhanga.
Hospital  : Name of Hospital; 1=KM Hospital, 0=SKMCH.
Symptom   : Symptom found in the children; 1=Hypoglycaemia, 0=Hyper pyrexia.
Time      : Time from Hospitalise to the event in days
Indicator: 1=Death ,0= alive(Discharge, LAMA, Refer)
```

3 Kaplan-Meier Estimate of Survival Function

Kaplan Meir estimator is an estimator, which estimates the survival function of the lifetime data, and it measures the fraction of subjects living for a certain amount of time after treatment. It is also known as a non-parametric maximum likelihood estimate of S(t). Let S(t) be a probability of surviving an individual after having lifetime t.

$$\hat{S}(t) = \prod_{t_i<t} \frac{n_i - d_i}{n_i} \tag{1}$$

where, n_i is number of individual at risk to each corresponding t_i and d_i is the number of death at t_i.

Table 1: Kaplan-Meier estimates of survival function of AES data

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	173	16	0.908	0.0220	0.8654	0.952
2	123	8	0.848	0.0288	0.7938	0.907
3	78	3	0.816	0.0333	0.7531	0.884
4	43	5	0.721	0.0496	0.6301	0.825
5	21	4	0.584	0.0737	0.4557	0.747
7	6	1	0.486	0.1080	0.3148	0.751
8	4	1	0.365	0.1328	0.1787	0.745
9	2	1	0.182	0.1451	0.0384	0.867
11	1	1	0.000	NaN	NA	NA

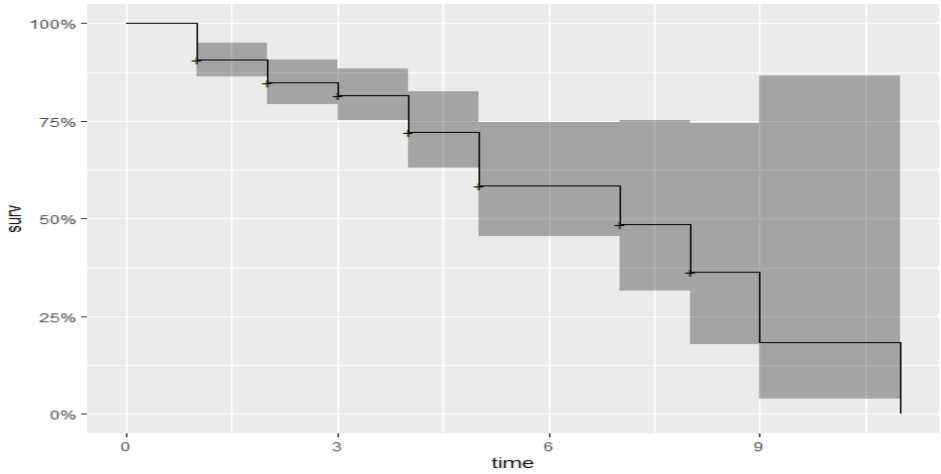

Fig. 1: Graph of the Kaplan-Meier estimates of survival function in table 1

From Figure 1, we can see that the probability of survival of children decreases as the time increases.

4 Bayesian Analysis using Weibull model

If T as a random variable follow Weibull distribution with parameter α and λ, then the pdf of Weibull can be written as given below;

$$f(t) = \frac{\alpha}{\lambda} \times \left(\frac{t}{\lambda}\right)^{\alpha-1} \times e^{-(\frac{t}{\lambda})^{\alpha}} \qquad t \geq 0 \tag{2}$$

cdf,

$$F(t) = 1 - e^{-(\frac{t}{\lambda})^{\alpha}} \qquad t \geq 0 \tag{3}$$

Survival function,

$$S(x) = e^{-(\frac{t}{\lambda})^{\alpha}} \qquad t \geq 0 \tag{4}$$

hazard function

$$h(x) = \frac{\alpha}{\lambda} \times \left(\frac{t}{\lambda}\right)^{\alpha-1} \qquad t \geq 0 \tag{5}$$

The joint likelihood function for right censored data can be written as,

$$L = \prod_{i=0}^{n} Pr(t_i, \delta_i) = \prod_{i=0}^{n} [f(t_i)]^{\delta_i} [S(t_i)]^{1-\delta_i}$$

here δ_i is indicator variable

$$\delta_i = \begin{cases} 0, censored \\ 1, observed \end{cases}$$

To build the Weibull regression model, we have introduced covariates through the log link function i.e.

$$E(t) = exp(X\beta)$$

$$\lambda\Gamma(1+1/\alpha) = exp(X\beta)$$

$$\lambda = \frac{exp(X\beta)}{\Gamma(1+1/\alpha)}$$

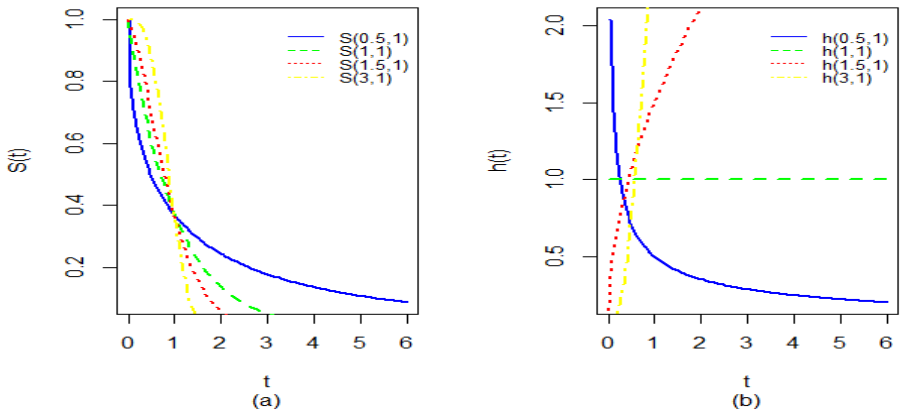

Fig. 2: (a)Survival curve and (b) Hazard curve of weibull distribution with differenet shape, scale=1

Likelihood function becomes as given below,

$$L=\prod_{i=0}^{n}\left[\frac{\alpha}{exp(x_i\beta)/\Gamma(1+1/\alpha)}\times\left(\frac{t}{exp(x_i\beta)/\Gamma(1+1/\alpha)}\right)^{\alpha-1}\times e^{-\left(\frac{t}{exp(x_i\beta)/\Gamma(1+1/\alpha)}\right)^{\alpha}}\right]^{\delta_i}$$

$$\times\left[e^{-\left(\frac{t}{exp(x_i\beta)/\Gamma(1+1/\alpha)}\right)^{\alpha}}\right]^{1-\delta_i} \tag{6}$$

To build the posterior distribution, first we need to specify the prior distribution for the regression coefficient and shape parameter of the model. In this paper, we have chosen Gaussian prior with mean 0, and a standard deviation of 10 as a weakly informative prior for β is a nearly flat prior but not completely. Here the question arises, why do we choose Gaussian prior with mean zero, because this prior places as much probability below zero as it does above zero, which gives more conservative estimates as compared to perfectly flat prior [16].

$$\beta_j \sim N(0,10)$$

For shape parameter α, we have chosen half Cauchy prior because [17] recommended the half-Cauchy prior as a better choice for a variance parameter. The half-Cauchy distribution with scale 25 is also a nearly flat prior but not completely. The prior distribution that are not completely flat provides enough information for the numerical approximation algorithm to continue to explore the target density, the posterior distribution[18]

$$\alpha \sim HC(0,25)$$

Using Bayes theorem, joint posterior distribution can be written as given below;

$$P(\alpha,\beta|X,t) \propto L(t|\alpha,\beta,X)\times P(\alpha)\times P(\beta) \tag{7}$$

$$\propto \prod_{i=0}^{n}\left[\frac{\alpha}{exp(x_i\beta)/\Gamma(1+1/\alpha)}\times\left(\frac{t}{exp(x_i\beta)/\Gamma(1+1/\alpha)}\right)^{\alpha-1}\times e^{-\left(\frac{t}{exp(x_i\beta)/\Gamma(1+1/\alpha)}\right)^{\alpha}}\right]^{\delta_i}$$

$$\times\left[e^{-\left(\frac{t}{exp(x_i\beta)/\Gamma(1+1/\alpha)}\right)^{\alpha}}\right]^{1-\delta_i}\times\frac{2\times 25}{\pi(\alpha^2+25^2)}\times\prod_{j=0}^{J}\frac{1}{10\sqrt{2\pi}}exp\left(-\frac{1}{2\times 100}\beta_j^2\right) \tag{8}$$

4.1 Fitting of Weibull model with brms

We have given the name of data is `Aes`. To access data from csv file, there is function in `R` that is `read.csv()`.

```
Aes=read.csv("Aes.csv")
names(Aes)
[1] "Time"      "Age"       "Gender"    "District"    "Symptom"
[6] "Indicator" "Hospital"

head(Aes)
Time Age   Gender  District Symptom Indicator Hospital
 2   4.0      1         1       1         0        1
 1   2.0      0         1       1         0        1
 1   3.0      0         1       1         1        1
 1   2.6      1         1       1         1        1
 2   1.5      0         1       1         0        1
 3   1.4      1         1       1         0        1
```

To fit the model with `brms`, first we need to require it. In `brms`, 0 stands for observed and 1 stands for censored. So, there is a need to add one more variable to the given data that is censored which is opposite of Indicator variable.

```
require(brms)
Aes$censored=as.numerci(Aes$Indicator==0)
AesWl=brm(Time|cens(censored)~Age+Gender+District+Hospital+Symptom,
      prior=c(set_prior("normal(0,10)",class="b"),set_prior("cauchy(0,25)",
      class="shape")),family = weibull(),chains=2,iter=5000,,data=Aes)
```

4.2 Output Summary

Using `print()` function we can get result in tabular form

```
print(AesWl)
```

Table 2: Summary of simulated results from `brm` function contains the posterior estimates ,Monte Carlo error, Credible Interval, Rhat and Effective sample size.

Parameter	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	9.46	5.94	1.65	24.11	1.00	1931	1941
Age	0.03	0.05	-0.06	0.13	1.00	3693	3429
Gender	-0.24	0.28	-0.80	0.30	1.00	3315	2357
District	-0.11	0.51	-1.23	0.77	1.00	3862	3002
Hospital	0.97	0.32	0.41	1.67	1.00	2351	2732
Symptom	-7.91	5.93	-22.64	-0.11	1.00	1906	1937
shape	1.35	0.18	1.02	1.71	1.00	2191	2819

After fitting of Weibull model to the AES data with `brm` function, we get the result in tabular form, which is given in Table 2. From Table 2, it is an evidence that the intercept of model and covariates Hospital, and Symptom are statistically significant at 95% credible interval and the rest of the covariates like Age, Gender, and District are not statistically significant at 95% credible interval. The posterior estimate of the coefficient of the covariate Hospital is 0.97 ± 0.32, here the positive value of estimate indicates the surviving probability of the children who were admitted to the KM hospital is better than the children who were admitted to the SKMCH, or we can say that the children, who were admitted to the SKMCH have higher risk of death. The posterior estimate of the covariate Symptom is -7.91 ± 5.93, here the negative value of estimate indicates the survival probability of the children, who had hypoglycaemia symptoms is lower than the children, who had hyper-pyrexia symptom. The effective sample size is enough to get the conversion, and the Rhat for all the parameters is one that means the Marcov chain well converge to the target distribution(posterior distribution), or we can say that the conversion of a given model is good.

4.3 Graphical Analysis

For plotting the Caterpillar plot in `brms`, there is a function `stanplot()`
`stanplot(AesWl)`

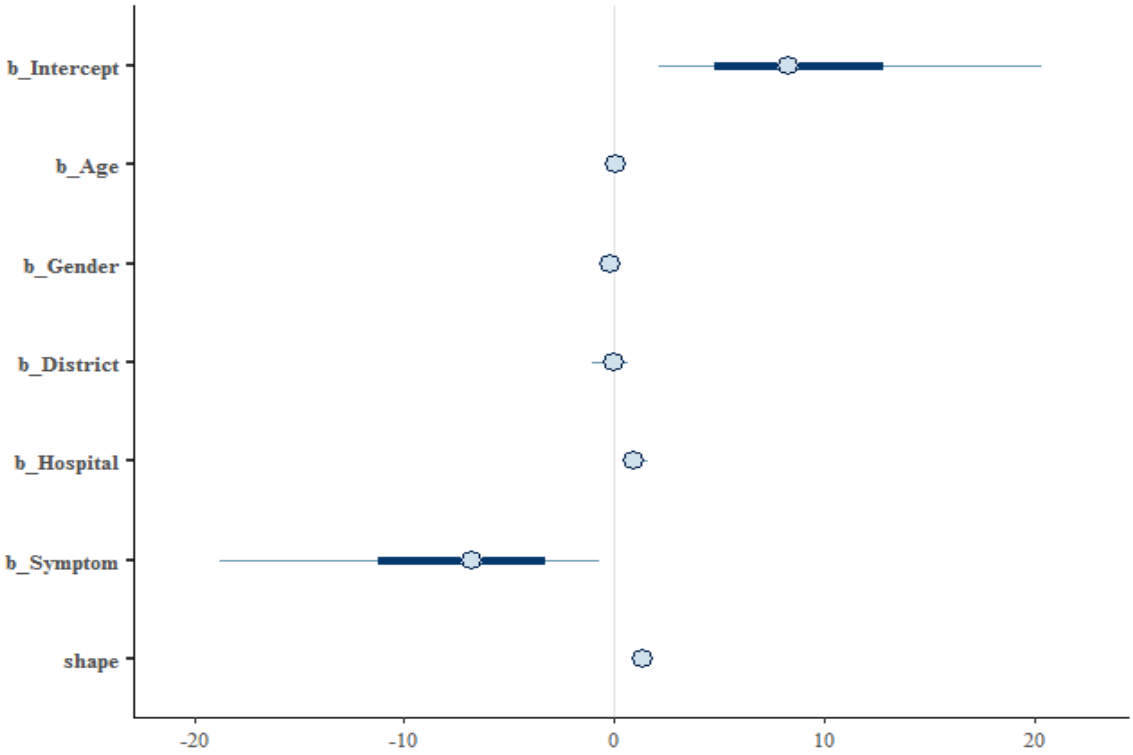

Fig. 3: Caterpillar plot of the Bayesian Weibull regression model

From the above Figure 3, it is evident that the intercept of model, and covariates Hospital & Symptom are statistically significant because they are not crossing the vertical line passes through the zero.

For plotting the trace plot and density plot, we have used the command, as given below.
`plot(AesWl,pars = c("b_Intercept","b_Age","b_Gender","b_District"))`

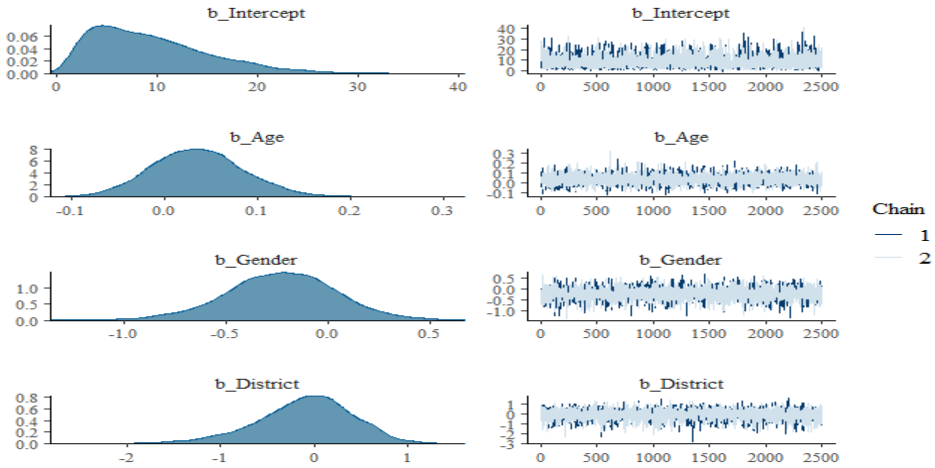

Fig. 4: Density and Trace plot of the parameters the model

```
plot(AesWl,pars = c("b_Symptom","b_Hospital","shape"))
```

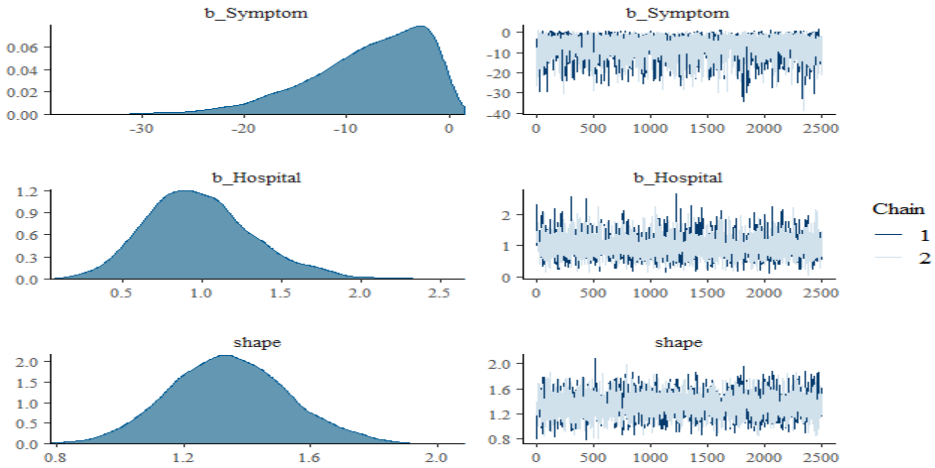

Fig. 5: Density and Trace plot of the parameters of the model

From the above Figure 4 and Figure 5, we can see that the mixing of two chains is good. They overlaid, which means the Marcov chain converge to the target distribution in very well manner.

Marginal Effects Plots

In `brms`, there is provision to plot marginal effects of the model and the command for plotting the marginal effects is

```
marginal_effects(AesWl)
```

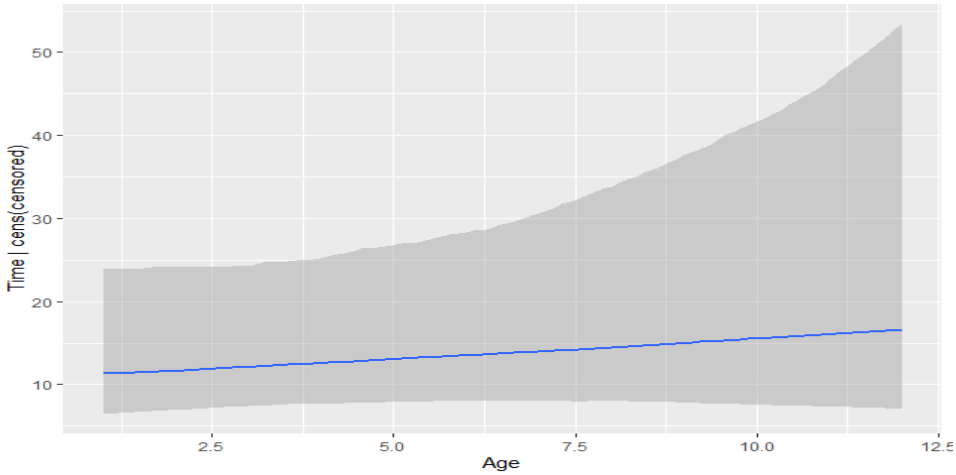

Fig. 6: Marginal effects plot of Age Vs Time

It is evident from Figure 6 that there is a positive relationship between age and survival time. As age increases, the survival time increases, which means that the children as much as younger than others have lower survival time.

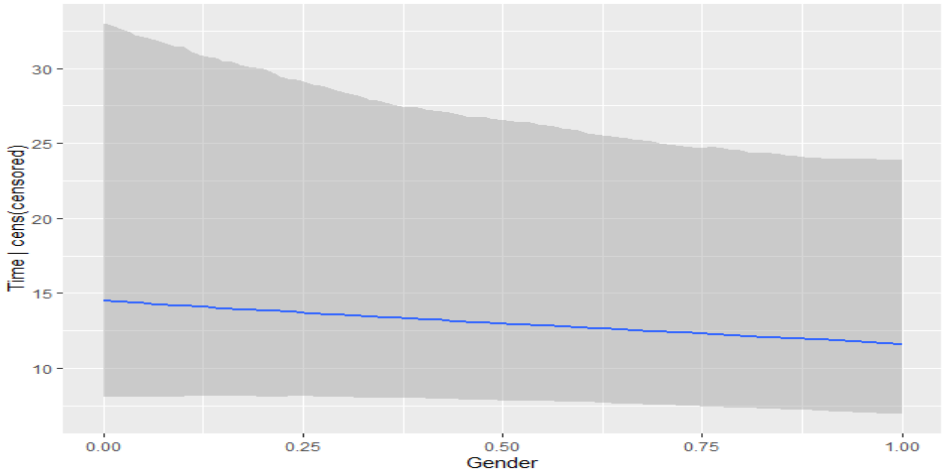

Fig. 7: Marginal effects plot of Gender Vs Time

It is seen from the Figure 7 that female child patients have lower survival time than male child patients.

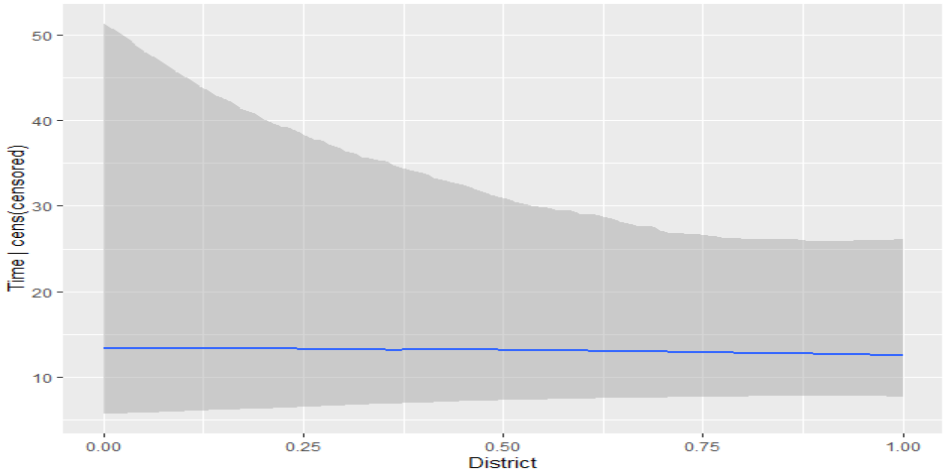

Fig. 8: Marginal effects plot of District Vs Time

It is evident from Figure 8 that the children are from Muzaffarpur district have lower survival time than the children are from other districts(Champaran, Sitamadhi, Vaishali, and Darbhanga).

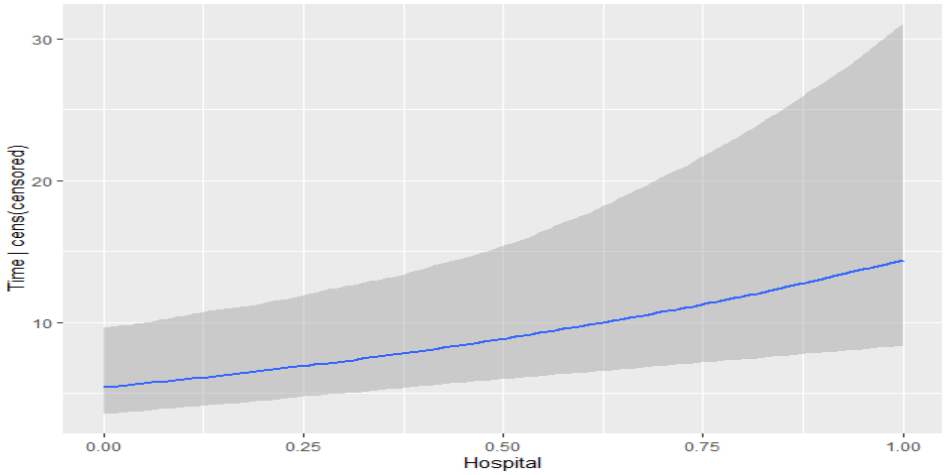

Fig. 9: Marginal effects plot of Hospital vs Time

It is seen from Figure 9 that the children who were admitted to SKMCH hospital have lower survival time as compared to children who were admitted to KM hospital. There is a significant difference in effects of two hospitals.

Table 3: LOO-CV values of the three models

Model	LOO-CV
Weibull	256.1
Log-normal	258.0
Exponential	260.7

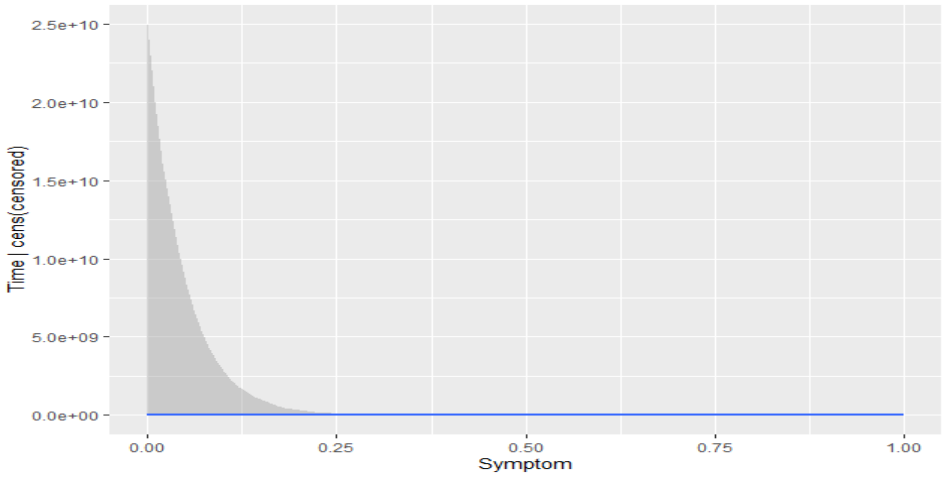

Fig. 10: Marginal effect plot of Symptom Vs Time

It is seen from Figure 10, the children with symptom hypoglycaemia have lower survival time than children with symptom hyper-pyrexia.

4.4 Model comparison

To compare the three models, Weibull, Log-normal, and Exponential, we have used leave one out cross-validation(LOO-CV) criteria, and we have shown the LOO-CV values of three models in the Table 3. However, we have not given the whole analysis of Log-normal and Exponential model, just to save the space and also because of the same thing has been done with Log-normal and Exponential model as we have done with Weibull model.

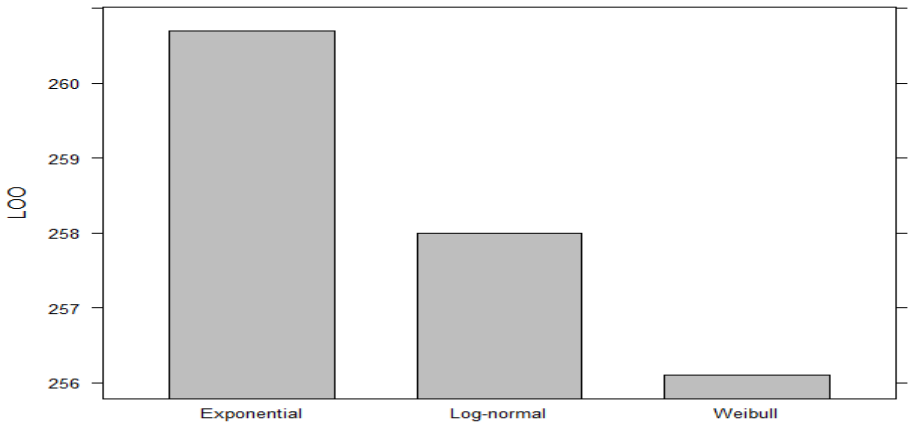

Fig. 11: Barplot of LOO-CV of the three models

It is seen from the Table 3 and Figure 11 that the Weibull model has lowest LOO-CV value as compared to others, which means that the Weibull model is the best model than the others and it would provide a better fit to the AES data.

We have also made the comparison between parametric survival models and non-parametric estimator of the survival function with the help of graph as shown in Figure 12. See Figure 12, the Kaplan-Meier estimator and Weibull model have almost the same behaviour but the distribution of Log-normal model and Exponential model does not seem similar to the Kaplan Meier estimates of the survival function.

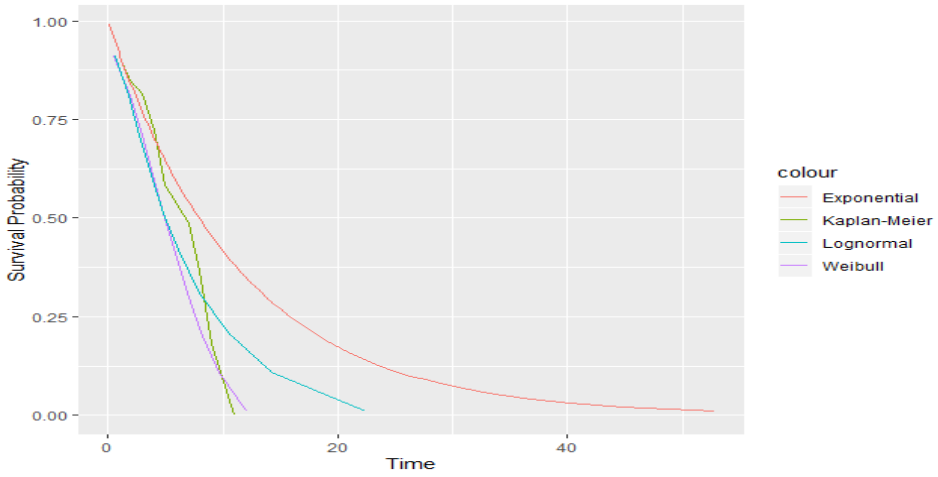

Fig. 12: Graph for the parametric and non parametric model

5 Bayesian Analysis using logistic regression model

If $(y_1, y_2, ..., y_n)$ represents the data for individual units who experienced the death or not, then y_i for (i=1,2,...,n) will follow Bernoulli distribution with values, 0 or 1 . Alternatively, $(y_1, y_2, ..., y_n)$ represents the count of event under study over a specified number of trials for n groups, then y_i follow Binomial distribution with values $1, 2, 3, .., n_i$, where n_i represents the number of individual units for the i^{th} group.
Likelihood function for i^{th} individual with the covariates x_i is given below,

$$likelihood_i = \pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i} \tag{9}$$

where,

$$\pi(x_i) = P(y_i = 1|x_i)$$

As we know that the logistic regression is given by,

$$\pi(x) = \frac{e^{\beta_0+\beta_1X_1+\beta_2X_2+...+\beta_pX_p}}{1+e^{\beta_0+\beta_1X_1+\beta_2X_2+...+\beta_pX_p}}$$

Now, the likelihood function for the i^{th} individual becomes as,

$$likelihood_i = \left(\frac{e^{\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+...+\beta_pX_{ip}}}{1+e^{\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+...+\beta_pX_{ip}}}\right)^{y_i}\left(1-\frac{e^{\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+...+\beta_pX_{ip}}}{1+e^{\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+...+\beta_pX_{ip}}}\right)^{1-y_i} \tag{10}$$

Joint likelihood function over the data set given as,

$$\prod_{i=0}^{n}\left[\left(\frac{e^{\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+...+\beta_pX_{ip}}}{1+e^{\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+...+\beta_pX_{ip}}}\right)^{y_i}\left(1-\frac{e^{\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+...+\beta_pX_{ip}}}{1+e^{\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+...+\beta_pX_{ip}}}\right)^{1-y_i}\right] \tag{11}$$

prior, we have chosen Normal distribution with mean=0 and sd=10 as a weak informative prior for regression coefficients β.

$$\beta_j \sim N(0,10)$$

Using Bayes theorem, the posterior distribution can be constructed as,

$$Posterior = \prod_{i=0}^{n}\left[\left(\frac{e^{\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+...+\beta_p X_{ip}}}{1+e^{\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+...+\beta_p X_{ip}}}\right)^{y_i}\left(1-\frac{e^{\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+...+\beta_p X_{ip}}}{1+e^{\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+...+\beta_p X_{ip}}}\right)^{1-y_i}\right]$$

$$\times\prod_{j=0}^{p}\frac{1}{10\sqrt{2\pi}}exp\left(-\frac{1}{2\times 100}\beta_j^2\right) \tag{12}$$

5.1 Fitting of logistic regression model with brms

```
Aes$Death=Aes$Indicator
AesLt=brm(Death~Age+Gender+District+Symptom+Hospital,data=Aes, chains=2,
iter=5000,prior=c(set_prior("normal(0,10)",class="b")), family = bernoulli())
```

5.2 Output Summary

```
print(AesLt)
```

Table 4: Summary of simulated results from `brm` function contains the posterior estimate, Monte Carlo error, credible interval, Rhat, and effective sample size.

Parameter	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	3.42	7.95	-13.29	18.40	1.00	2025	1869
Age	-0.37	0.17	-0.75	-0.06	1.00	4979	3269
Gender	0.64	0.56	-0.41	1.77	1.00	4926	3270
District	0.43	0.75	-0.89	2.02	1.00	4968	2840
Symptom	7.92	6.11	-0.46	21.97	1.00	3247	2585
Hospital	-12.64	4.99	-25.12	-5.89	1.00	1842	1643

It can be seen from Table 4 that the covariates Age, and Hospital are statistically significant at 95% credible interval, and the remaining covariates like Gender, District, and Symptom are not statistically significant at 95% credible interval. The posterior estimate of the regression coefficient of Age is -0.37$\pm$0.17, the negative sign of estimates indicates that the younger children are at higher risk of death or we can say that as age increases, the risk of death decreases. The posterior estimate of the regression coefficient of Hospital is -12.64$\pm$4.99, which tells us that the children who were admitted to SKMCH hospital are at higher risk of death than the children who were admitted to KM hospital. As it can be seen from Table 4 that the covariate Symptom is not statistically significant, but it has an influence on the risk of death, which we will see in the marginal effects plot. The effective sample size is enough for the conversion of the model, and Rhat for all the parameters of the model is near to one, which means the Marcov chain converge to the target distribution(posterior distribution) in a very well manner.

5.3 Graphical Analysis

```
stanplot(AesLt)
```

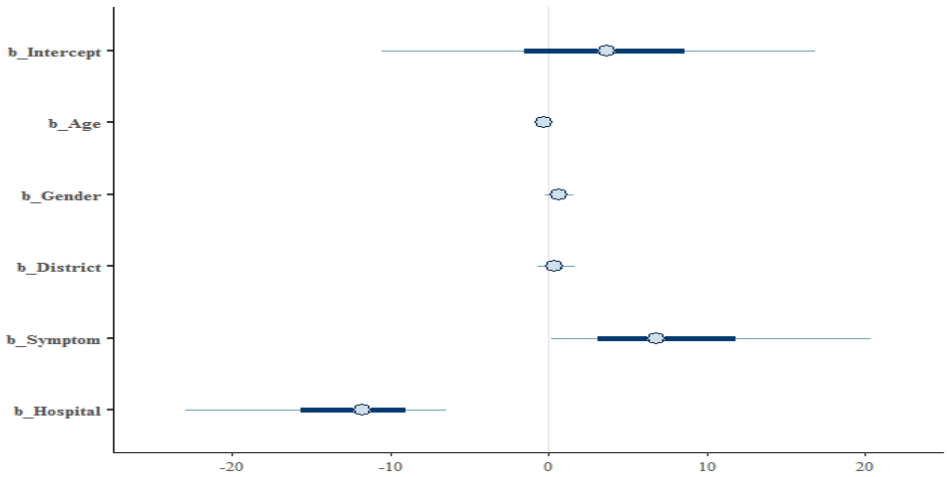

Fig. 13: Caterpillar plot of the Bayesian logistic regression model

From the Figure 13, it is evident that covariates Age and Hospital are statistically significant.

```
plot(AesLt,pars = c("b_Intercept","b_Age","b_Gender","b_Dist"))
```

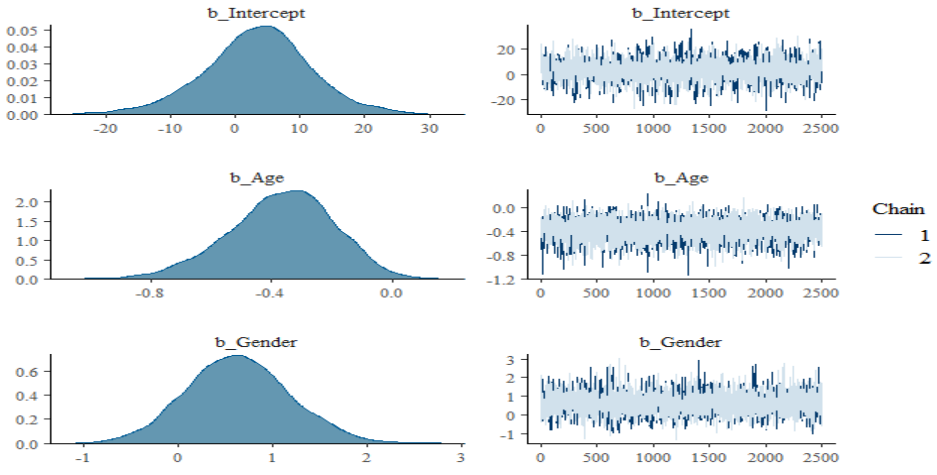

Fig. 14: Density and Trace plot for the parameters of the model

```
plot(AesLt,pars = c("b_Symptom","b_Hospital","shape"))
```

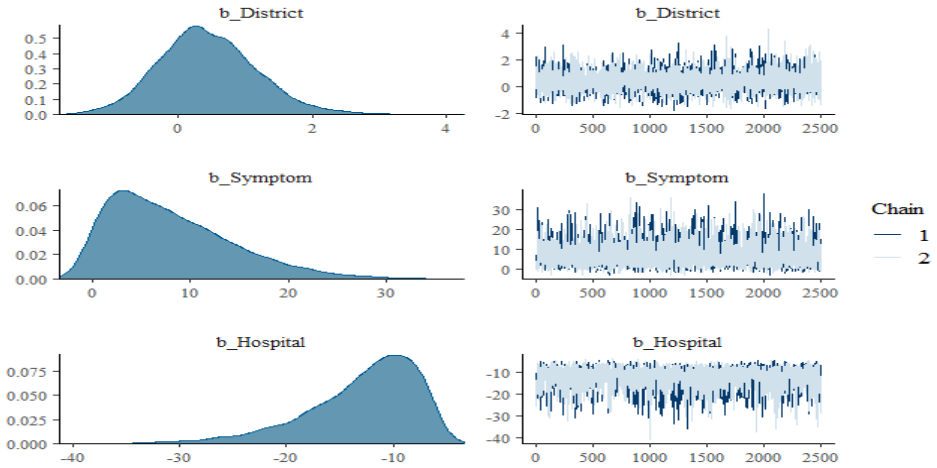

Fig. 15: Density and Trace plot for the parameters of the model

It can be seen from Figure 14 & 15 that the mixing of two chains age good, and the Marcov chain converge to the target distribution in a very well manner.

Marginal Effects Plots

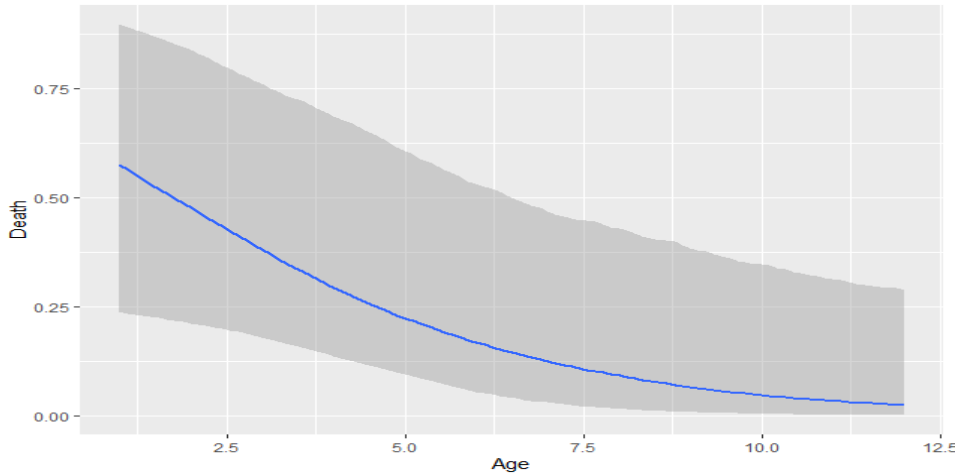

Fig. 16: Marginal effect plot of covariate Age

From Figure16, we can see that the probability of children's death decreases as age of children increases that means the children, who are as much older than the other children are at lower risk of death. See Figure16, the children of age less than two years are at the risk of death with high probability, and the children are of age greater than ten years have lower probability of death near to zero.

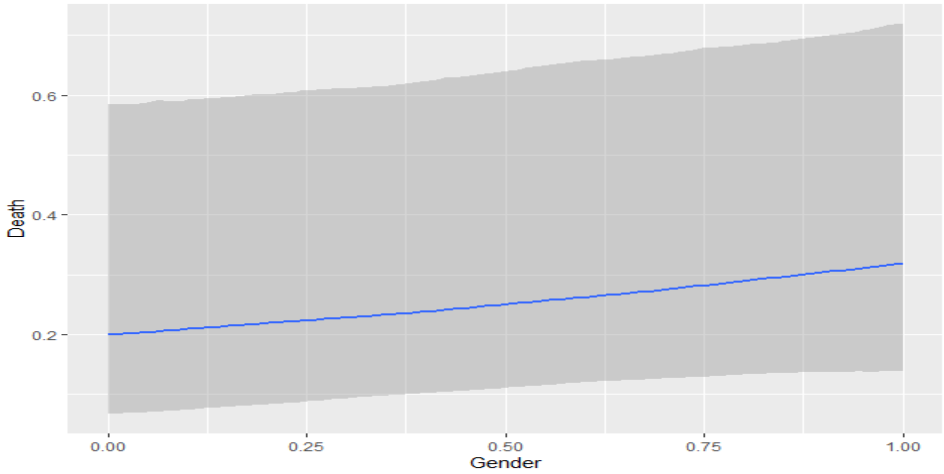

Fig. 17: Marginal effect plot of the covariate Gender

From Figure 17, we can see that the female children have high probability of death than the male children, but this covariate has not much effect on the risk of death.

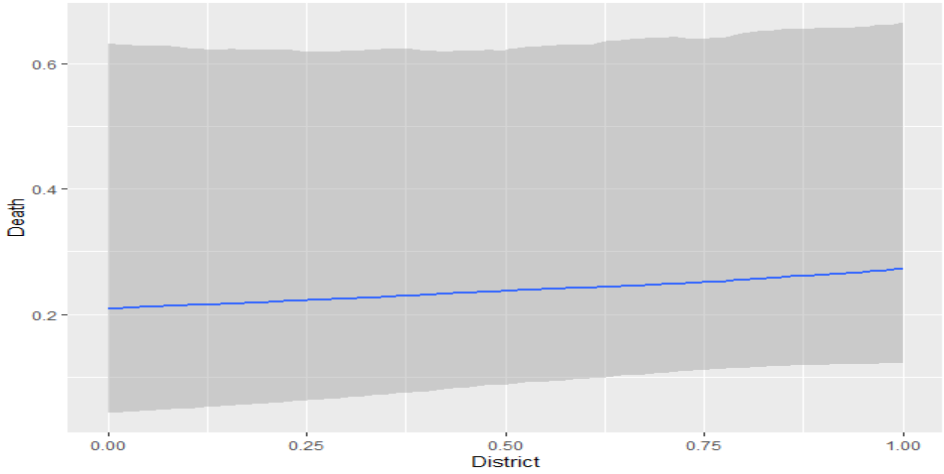

Fig. 18: Marginal effect plot of the covariate District

See Figure 18, the probability of death of the Muzaffarpur district's children and others district's are slightly different, and but the difference is not significant as we have seen Table 4 . It means Districts have not much influence on the risk of death.

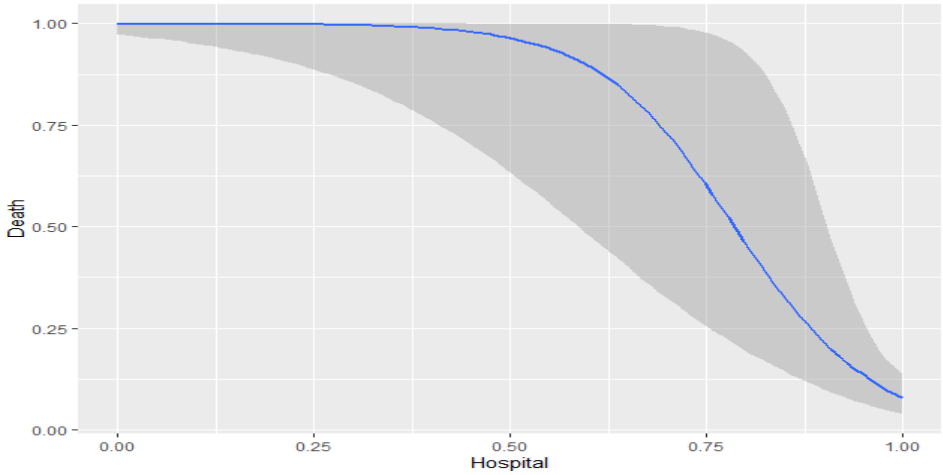

Fig. 19: Marginal effect plot of the covariate Hospital

It is clearly seen from Figure 19 that the children who were admitted to SKMCH hospital are at higher risk of death with a high probability than the children who were admitted to KM hospital.

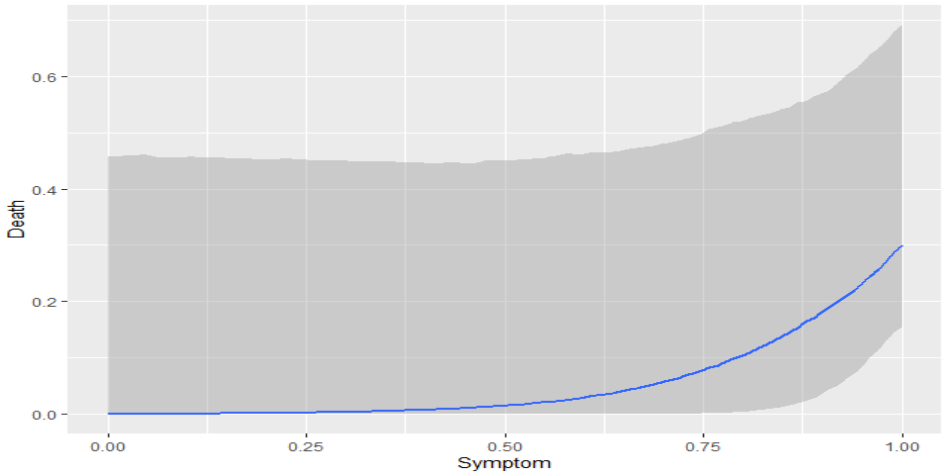

Fig. 20: Marginal effect plot of the covariate symptom

It is seen from Figure 20 that the children who had Hypoglycaemia symptoms are at higher risk of death than the children who had Hyper-pyrexia symptom or the children with hyper-pyrexia symptom have experienced risk of death with almost zero probability.

5.4 Posterior Predictive check

Posterior predictive check is the post-analysis of the Bayesian model after the fitting, and it checks whether the replicated data sets have the same behavior as observed data set or not. The posterior predictive distribution for the replicated observation y^{rep} can be written as,

$$p(y^{rep}|y) = \int p(y^{rep}|\theta)p(\theta|y)d\theta$$

Where θ is a set of all parameters in the model.

Figure 21 is a density plot of the logistic regression model in which light blue lines show the distribution of replicated data set y^{rep}, and the dark blue line shows the distribution of observed data set y. See Figure 21, the y^{rep} and y are overlapping to each other and have the same trend, which means the replicated data set and observed data set have the same behavior and performance of logistic regression model for predicting the risk of death is good.

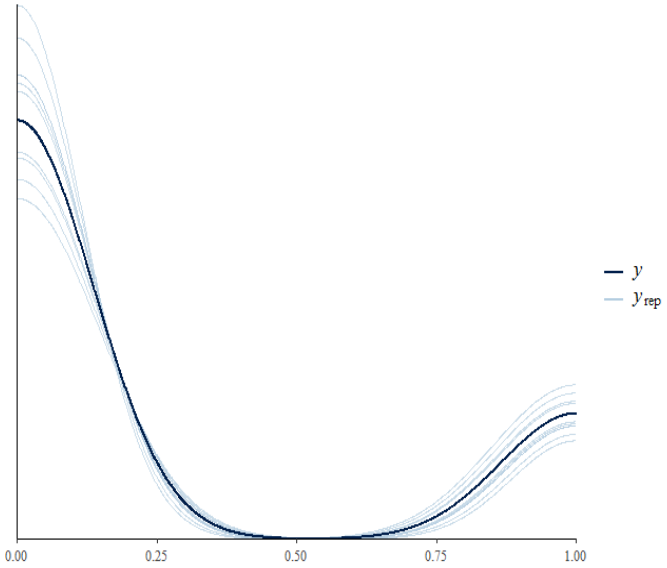

Fig. 21: Posterior predictive check of the logistic regression model

6 Discussion and Conclusions

In this paper, we have done an analysis on the outbreak of acute encephalitis syndrome, which occurred in Muzaffarpur district and their neighboring district of Bihar in June 2019. We have fitted Weibull, Log-normal, and Exponential model to the AES data with the Bayesian approach. We have found that Weibull is the best model for the fitting of AES data as compared to the Log-normal and Exponential model, and the survival time of children follow the Weibull distribution more than the others. In the case of non-parametric and parametric model comparison, the Weibull model behaves similar to the Kaplan-Meier estimator of the survival function for AES data. For predicting the risk of death of children due to AES on the basis of observed characteristic(Age, Gender, District, Hospital, and Symptom), we have fitted logistic regression model to the given data, and we have found that the performance of logistic regression model is good with the help of posterior predictive check. From the whole study, we have found that the children, who are as much as younger than others, have lower survival time and symptom Hypoglycaemia reduces the surviving time of patients with a higher risk of death. The children, who were admitted to KM hospital have better survival than the children who were admitted to SKMCH. It is mean that the probability of survival time, of the AES suffered children are affected by the covariates Age, Hospital, and Symptom, or we can say that these covariates have significant effects on the risk of death.

Acknowledgments

The authors acknowledge with thanks to both the referees and the Editor-in-Chief for their valuable suggestions and comments which led to the overall improvement in the manuscript.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this article.

References

[1] R. Joshi, S. P. Kalantri, A. Reingold, and J. M. Colford Jr, Changing landscape of acute encephalitis syndrome in India: a systematic review, National Medical Journal of India, **25**, 212-220 (2012).
[2] S. Ghosh, and A. Basu, Acute encephalitis syndrome in India: the changing scenario, Annals of neurosciences, **23**, 131 (2016).
[3] A. Shrivastava, A. Kumar, J. D. Thomas, K. F. Laserson, G. Bhushan, M. D. Carter, others, Association of acute toxic encephalopathy with litchi consumption in an outbreak in muzaffarpur, india, 2014: a case-control study, The Lancet Global Health, **5**, 458-466 (2017).
[4] A. Chaudhuri, and P. G. E. Kennedy, Diagnosis and treatment of viral encephalitis, Postgraduate medical journal, **78**, 575-583 (2002).
[5] P. C.Bürkner, brms: Bayesian regression models using Stan, R package version 0.10. 0 (2016).
[6] A. Vehtari, A. Gelman, and J. Gabry, Practical bayesian model evaluation using leave-one-out cross-validation and waic, Statistics and computing, **27**, 1413-143 (2017).
[7] A. Vehtari, A. Gelman, and J. Gabry, Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC, Statistics and computing, **27**, 1413-1432 (2017).
[8] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, (2019).
[9] S. D. Team, Stan Modeling Language Users Guide and Reference Manual Version 2.17. 0., (2017).
[10] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Hybrid monte carlo, Physics letters B, **195**, 216-222 (1987).
[11] M. D. Hoffman, and A. Gelman, The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo, Journal of Machine Learning Research, **15**, 1593-1623 (2014).
[12] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, The journal of chemical physics, **21**, 1087-1092 (1953).
[13] W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika, **57**, 97-109 (1970).
[14] S. Geman, and D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Transactions on pattern analysis and machine intelligence, 721-741 (1984).
[15] A. E. Gelfand, and A. F. Smith, Sampling-based approaches to calculating marginal densities, Journal of the American statistical association, **85**, 398-409 (1990).
[16] R. McElreath, Statistical rethinking: A bayesian course with examples in r and stan, Chapman and Hall/CRC, 94-95 (2018).
[17] A. Gelman, Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper), Bayesian analysis, **1**, 515-534 (2006).
[18] N. Khan, and A. A. Khan, Bayesian Analysis of Topp-Leone Generalized Exponential Distribution, Austrian Journal of Statistics, **47**, 1-15 (2018).