

Application of Bayesian Approach Survival Analysis of Under-five Pneumonia Patients in Tercha General Hospital, South West Ethiopia

Lema Abate^{1,*}, Megersa Tadesse²

¹Department of Statistics, College of Natural and Computational Science, Mizan-Tepi University, Tepi, Ethiopia

²Department of Statistics, College of Natural Science, Jimma University, Jimma, Ethiopia

Email address:

lemamtu2007@gmail.com (L. Abate), magetade@gmail.com (M. Tadesse)

*Corresponding author

To cite this article:

Lema Abate, Megersa Tadesse. Application of Bayesian Approach Survival Analysis of Under-five Pneumonia Patients in Tercha General Hospital, South West Ethiopia. *International Journal on Data Science and Technology*. Vol. 6, No. 1, 2020, pp. 44-52.

doi: 10.11648/j.ijdst.20200601.16

Received: December 9, 2019; **Accepted:** December 25, 2019; **Published:** January 16, 2020

Abstract: Pneumonia is among the major killer diseases in under-five children in the world. In developing countries 3 million children die each year due to pneumonia. Ethiopia is one of the 15 pneumonia high burden countries. The aim of this study was to examine the risk factors of the survival time of under-five pneumonia patients using Bayesian approach analysis. Total of 281 under-five pneumonia patients included in this study. The parametric survival models such as Weibull, Lognormal and Log-logistic baseline distributions were used to fit the datasets by introducing prior distributions. The DIC value was used to compare the baseline distributions, and based on the DIC value the Weibull baseline distribution was selected as good model to fit under-five pneumonia dataset well. The results obtained from the Weibull survival model showed that patients from urban residence and patients who were admitted during patient nurse ratio (PNR) was small; were prolong timing death of under-five pneumonia patients, while patients who admitted during Spring and summer season, patients who suffer comorbidity and severe acute malnutrition (SAM) were shorten timing of death of under-five pneumonia patients. Factors such as sex, residence, Season of Diagnosis, Comorbidity, Severe Acute Malnutrition (SAM), Patient refer status and Patient to Nurse Ratio (PNR) were associated with the survival time of under-five pneumonia in this study. The concerned body should give attention for the factors identified in these study to prevent the mortality of under-five children due to pneumonia.

Keywords: Pneumonia, Under-Five, Parametric Models, Risk Factors, Bayesian Approach, WinBUGs

1. Introduction

Pneumonia is the major killer of under-five children's than any other diseases known to affect children, more than the death shares of Acquired Immune Deficiency Syndrome (AIDS), Malaria and Measles combined [1]. More than 50% of all new pneumonia cases of the under-five childhood are concentrated in the poorest world's regions, Sub-Saharan Africa and South Asia. In terms of mortality, about 90% of all under-five Pneumonia deaths burden is reported to occur in these two regions [2]. According to World Bank report the risk of pneumonia in children in developing countries is 3 to 6 times higher than other children. Not only outbreak of pneumonia, but also the mortality rate of this disease is

higher in developing countries [3]. The contribution of pneumonia to the deaths of older children was estimated to reach 14.1% with approximately four percent of childhood-pneumonia related death occurred in the first 28 days of life globally [4]. According to pneumonia and diarrhea progress report of 2015, Ethiopia is among 15 top under five pneumonia high burden countries. And also the study conducted in Gilgel Gibe Field Research Center reported that Neonatal and infant mortality rates were respectively 38 and 76.4 per 1000 live births. The two most common causes of death during neonatal period were prematurity (26.4%) and pneumonia (22.6%). Whereas the top causes of death in post-neonatal period were pneumonia (42%), malaria (37%) and acute diarrheal diseases (30%) [5].

Few studies in this area was tried to use statistical models like Binary logistic regression model and multilevel logistic regression models to examine the determining factors of under-five pneumonia patients and several studies were used descriptive analysis. But those studies are not enable to study the survival time of patients hospitalized with pneumonia. Due to this, the researcher was intended to use the survival analysis models. Survival analysis is a statistical method for data analysis where the outcome variable of interest is the time to the occurrence of an event and there are many standard parametric models such as Weibull, Lognormal and log-logistic models [6]. Lately Bayesian approach have been used in many research studies, especially in the field of medicine, as an alternative to classical or frequentist statistical methods. One of the reasons is that classical methods base their maximum likelihood estimations on asymptotic considerations that are usually only valid for a considerable data size [7]. The study conducted at Beirut Lebanon was reported that the Bayesian approach may have advantages over the frequentist one, particularly in case of a low power of the frequentist analysis [8]. According to Gelfand, A. and Mallick B. K. [9] study Bayesian method is the best method to obtain the appropriate estimates of the model.

The Bayesian methods are applied to relatively small data sets where the validity of the asymptotic assumptions is doubtful [7]. In the Bayesian approach no assumption is made as to the shape of the percentile distribution, rather the data themselves specify the distribution and the Bayesian approach has the possibility of improving the precision of the results by introducing external information in terms of the priori distribution [7]. The choice of prior distribution was determined as the experiences gained in the previous studies, the researcher's expert of knowledge. If prior information comes from the previous research, Bayesian estimation should lead to more precise results than classical methods [10]. The prior distributions which contain exact information and have an impact on the posterior distribution are called informative priors. In Bayesian method, if the researcher intends to obtain correct results, the researcher should use prior distributions which have a minimal impact on the

posterior distribution. Such distributions are called non-informative prior distributions [11]. In this study a small size sample limited to under-five pneumonia was used in order to investigate the survival time of under-five pneumonia patients. As the classical approach central limit theory and requires a big sample, for a small sample it is essential to use Bayesian approach, even if the researcher do not have any prior information, because the researcher can choose non-informative prior distributions [11].

In this study Bayesian survival Analysis are used to identify important risk factors of under-five pneumonia patients. The response variable in this study is the admission time until an event (death) due to pneumonia. Log rank test is used to compare the survival experience of different category of patients and Parametric Accelerating failure time (AFT) models using Bayesian approach analysis is used to identify predictors of mortality of under-five pneumonia patients. There has been a limited literature on the use of Bayesian survival analysis of under-five pneumonia disease. Therefore the main aim of this study was to investigate survival time of under-five pneumonia patients in Tercha General Hospital Using Bayesian approach survival analysis.

2. Method and Materials

2.1. Study Area and Study Design

The study was conducted in Dawro Zone Tercha General Hospital around 491 km away from Addis Ababa capital city of Ethiopia. The study was retrospective study that reviews or visits all under-five aged children cards and pediatric charts hospitalized with Pneumonia in Tercha General Hospital. The data analysis was taken place using latest version R 3.5.0 and WinBUGS software.

The response variable in this study was the survival time (time to death) of under-five pneumonia patients measured in days from the date of admission until the date of patients' death or discharge from hospital. The status variable is coded as 1 if an event (death) occur and 0 for censored in the given time interval. Explanatory variables considered in this study were listed below in Table 1 with their corresponding codes.

Table 1. Description of variables in the study.

	Variables	Coding for Categorical variables		Description
1	Sex	Female= 0	male = 1	Sex of children
2	Age	1-11 = 0	36-47 =3	Age of children
		12-23 = 1	48-59 =4	
3	Residence	24-35 = 2		Residence of children
		Rural = 0	Urban =1	
4	Season	Autumn =0	Winter =1	Season of Diagnosis
		Spring = 2	Summer =3	
5	Co-morbidity	No = 0	Yes =1	Co-morbidity (CAP complicated)
6	SAM	No = 0	Yes =1	Sever Acute Malnutrition
7	Treatment types	Penicillin =0	Ceftriaxone =1	Treatment types taken at time of Diagnosis
		Ampicillin = 2	Combined =3	
8	Patients refer status	No = 0	Yes =1	Patient refer status from other health center
9	Weight	Weight of the children		
10	Bed Occupancy Rate (BOR)	The percentage of official beds occupied by hospital Inpatients for a given period of time; $BOR = \frac{(\text{Total length of hospital stay in a given month})}{30 \times \text{Number of beds in that month}}$		

	Variables	Coding for Categorical variables	Description
11	Patient: Physician Ratio (PPhR)	Ratio of patient to physician counted in a given month; $PPhR = \frac{\text{Number of patients admitted in a given month}}{\text{Number of physician on service in a given month}}$	
12	Patient: Nurse Ratio (PNR)	Ratio of patient to nurse counted in a month $PNR = \frac{\text{Number of patients admitted in a given month}}{\text{Number of nurse on service in a given month}}$	

* Length of Hospital stay is the number of calendar days from the days of patient admission to the day of discharged/died.

2.2. Survival Data Analysis

Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs. By time, mean years, months, weeks, or days from the beginning of follow-up of an individual until an event occurs [6].

2.3. Non-parametric Survival Methods

Nonparametric analyses are more widely used in situations where there is doubt about the exact form of distribution. The nonparametric methods used in these study is Log-rank test method. Assessing whether or not there is a real difference between groups can only be done, with any degree of confidence, by utilizing statistical tests. Log-rank test is the one commonly used nonparametric tests for comparison of two or more survival distributions [12]. The log rank test statistic for comparing two groups is given by:

$$Q = \frac{[\sum_{j=0}^r w_j(d_{1j} - \hat{e}_{1j})]^2}{\sum_{j=0}^r w_j^2 \hat{v}_{1j}} \quad (1)$$

Where: r is the number of rank-ordered failure times (event times), w_j is the weight for censor adjustment at time $t_{(j)}$, $\hat{e}_{1j} = \frac{n_{1j}d_j}{n_j}$, $\hat{v}_{1j} = \frac{n_{1j}n_{2j}d_j(n_j - d_j)}{n_j^2(n_j - 1)}$, d_{1j} is the observed number of failure (event occur) at time $t_{(j)}$ in group 1, n_{1j} is the number of individuals at risk of event occur in the first group just before time $t_{(j)}$, n_{2j} is the number of individuals at risk in the second group just before time $t_{(j)}$, d_j is the total number of events occurred at $t_{(j)}$, n_j is the total number of individuals at risk before time $t_{(j)}$. Q follows a chi-square distribution with $k-1$ degree of freedom.

2.4. Bayesian Method

The Bayesian approach analysis considers the parameters of the model as random variables and requires that prior distributions specified for them and data are considered as fixed. The key ingredients to a Bayesian analysis are the likelihood function, which reflects information about the parameters contained in the data, and the prior distribution, which quantifies what, is known about the parameters before observing data. The prior distribution and likelihood can be easily combined to form the posterior distribution, which represents total knowledge about the parameters after the data have been observed [13].

2.4.1. Prior Distribution

The prior distribution is a probability distribution that represents the prior information associated with the

parameter of interest. It is a key aspect of a Bayesian analysis. In this study the researcher used non-informative multivariate normal prior distribution for coefficients with mean zero and variance 1000 and inverse-gamma prior distribution with scale parameter $a=0.01$ and shape parameter $b=0.01$ [14]. A typical joint prior specification can be expressed as a product of a multivariate normal (for parameter β/σ^2) and an inverse gamma prior (for σ^2), that is

$$\beta/\sigma^2 \sim N_p(\mu_0, \sigma^2 V_0), \sigma^2 \sim IG(a, b) \quad (2)$$

2.4.2. Likelihood Function

A likelihood functions is a function that gives the probability of observing the sample data given the current parameters. Suppose we observe n independent vectors of (t_i, δ_i, x_i) , where t_i is time to the event (death of under-five pneumonia patients) and δ_i is indicator variable telling us whether t_i is uncensored or censored.

$$\delta_i = \begin{cases} 0 & \text{censoring observation} \\ 1 & \text{event or dead (failure)} \end{cases}$$

The likelihood function of the set of unknown parameters θ in right censoring is written as:

$$L(\theta) = \prod_{i=1}^n [f(t_i/x_i)^{I(\delta_i=0)} \times S(t_i/x_i, \theta)^{I(\delta_i=1)}] \quad (3)$$

Log-likelihood would be as follows:

$$l(\theta) = \log \left\{ \prod_{i=1}^n [f(t_i/x_i, \theta)^{I(\delta_i=0)} \times S(t_i/x_i, \theta)^{I(\delta_i=1)}] \right\} \quad (4)$$

$$l(\theta) = \sum_{i=1}^n [\log(f(t_i/x_i, \theta)^{I(\delta_i=0)}) + \log(S(t_i/x_i, \theta)^{I(\delta_i=1)})] \quad (5)$$

Where: $f(t_i/x_i)$ and $S(t_i/x_i)$ are the density and survival distributions [29].

2.4.3. Posterior Distribution

The posterior distribution is obtained by multiplying the prior distribution over all parameters, θ by the full likelihood function $L(X/\theta)$. All Bayesian inferential conclusions are based on the posterior distribution of the model generated. The inference is performed by sampling from posterior distribution until the convergence to the posterior distribution is achieved [15]. The researcher assumed that θ is a random variable and has a prior distribution denoted by $\pi(\theta)$. Inference concerning θ is then based on the posterior distribution, which is obtained by Bayes' theorem. Then posterior distribution of θ is given by:

$$\pi(\theta/X) = \frac{L(X/\theta) \pi(\theta)}{\int L(X/\theta) \pi(\theta) d\theta} \quad (6)$$

Combining the likelihood function with the prior

distribution on (β, σ^2) and the full conditional distributions for unknown parameters, the posterior distribution can be written as:

$$\Pi(\sigma^2/\beta, t, x) \propto \prod_{i=1}^n \left[f\left(\frac{t_i}{x_i}, \theta\right)^{I(\delta_i=0)} \times S\left(\frac{t_i}{x_i}, \theta\right)^{I(\delta_i=1)} \right] \times \Pi(\beta/\sigma^2) \times \Pi(\sigma^2) \quad (7)$$

The posterior distribution for the model specification above does not have closed form solution for the parameters. For those models, MCMC Gibbs sampler is implemented using WinBUGS software. The baseline hazard distributions of parametric survival models such as Weibull, lognormal and log-logistic in Table 2 below are used to fit Bayesian Approach survival models by introducing prior for each parameters.

Table 2. Parametric distributions for the baseline hazards.

AFT models	f(t)	S(t)	h(t)
Weibull	$\lambda \rho t^{\rho-1} \exp(-\lambda t^\rho)$	$\exp(-\lambda t^\rho)$	$\lambda \rho t^{\rho-1}$
Lognormal	$\frac{1}{t\sigma\sqrt{2\pi}} \exp\left(-\frac{[\log(t)-\mu]^2}{2\sigma^2}\right)$	$1 - \Phi\left(\frac{\log(t)-\mu}{\sigma}\right)$	$\frac{\Phi\left(\frac{\log(t)}{\sigma}\right)}{\left(\frac{\log(t)}{\sigma}\right)\sigma t}$
Log-logistic	$\frac{\rho t^{\rho-1} \lambda^\rho}{(1+(\lambda t)^\rho)^2}$	$\frac{1}{1+(\lambda t)^\rho}$	$\frac{\rho t^{\rho-1} \lambda^\rho}{1+(\lambda t)^\rho}$

2.4.4. MCMC Estimation Methods

The MCMC iteration is used to do the integration numerically rather than analytically by sampling from the posterior distribution of interest even when the form of that posterior has no known algebraic form [16]. Gibbs sampler is an algorithm that sequentially generates samples from a joint distribution of two or more random variables [17]. Gibbs Sampler Algorithm is written as follows:

1. Choose an arbitrary initial value of $\theta^{(0)} = \{\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_k^{(0)}\}$
2. For $I = 0, 1, 2, \dots, N-1$, generate each component of θ as follows:
 - a. Draw $\theta_1^{(i+1)}$ from $\pi(\theta_1 | \theta_2^i, \theta_3^i, \dots, \theta_k^i, Y, X)$
 - b. Draw $\theta_2^{(i+1)}$ from $\pi(\theta_2 | \theta_1^{i+1}, \theta_3^i, \dots, \theta_k^i, Y, X)$
 - c. Draw $\theta_3^{(i+1)}$ from $\pi(\theta_3 | \theta_1^{i+1}, \theta_2^{i+1}, \theta_4^i, \dots, \theta_k^i, Y, X)$
 - d. ...
 - e. Draw $\theta_k^{(i+1)}$ from $\pi(\theta_k | \theta_1^{i+1}, \theta_2^{i+1}, \dots, \theta_{k-1}^{i+1}, Y, X)$
3. Repeat step 2 until convergence
4. Return $\theta^{(b+1)} = (\theta_1^{(b+1)}, \theta_2^{(b+1)}, \dots, \theta_k^{(b+1)}), \theta_1^{(b+2)}, \theta_2^{(b+2)}, \dots, \theta_k^{(N)}$

As a rule of thumb, the iteration should be run until the Monte Carlo error for each parameter of interest is less than about 5% of the sample standard deviation [18].

2.4.5. Model Selection and Comparison

The fitted Bayesian approach models are selected based on Deviance information criteria (DIC)

$$pD = E[-2\log\{f(y/\theta)\}] + 2\log\{f(y/\hat{\theta})\} \quad (8)$$

$$DIC = -2\log\{f(y/\hat{\theta})\} + pD \quad (9)$$

Where: pD is effective number of parameters in the model, $\hat{\theta}$ is maximum likelihood estimate.

DIC is used to compare Bayesian approach survival models. The preferable model is the one with the lowest value of the DIC [16].

2.4.6. Model Diagnostics

Once a model has been developed, the researcher would like to know how effective the model is in describing the outcome. This is referred to as goodness of fit. The most common ways of checking goodness of fit in Bayesian approach analysis are diagnosis for convergence and mixing. Therefore the researcher used time series plot, Autocorrelation plot, Gelman-Rubin statistic plot and Kernel density plots to check goodness of fit.

3. Results

A total of 281 under-five aged children who are registered due to pneumonia in Tercha General hospital from September, 2016 up to August, 2017 are included in this study out of those patients 126 (44.84%) are females and 155 (55.16%) are males. Patients whose residence was rural is 194 (69.04%) and whose residence was urban is 87 (30.96%). The case of under-five pneumonia based on the season of diagnosis are in Autumn 78 (27.76%), in Winter 33 (11.74%), in Spring 89 (31.67%) and in Summer 81 (28.83%) out of 281 patients, this implies that the occurrences of the pneumonia in the hospital are higher in spring and in summer season. According to age group in the table 3 below the percentage of under-five pneumonia were 49.82%, 22.06%, 11.39%, 10.32% and 6.41% in age group of 1-11, 12-23, 24-35, 36-47 and 48-59 months respectively. The patients with comorbidity is 39.50% and without comorbidity is 60.50%. This implies that even if there is no comorbidity the pneumonia in under-five is the sever disease. All the descriptive results are presented in the table 3 below.

In this study the log-rank test is used to compare the survival time of two or more groups of under-five pneumonia patients. According to log-rank test in table 3 below there is a significant difference of survival time between female and male patients ($p=0.023$) at 5% level of significance and there is no significant difference in each age groups. And also the log-rank test revealed that residence, comorbidity, SAM, Treatment types and patient refer status has statistical significant different survival time for each categories of covariates as shown in table 3 below.

Table 3. Descriptive Summary and log rank test of under-five pneumonia patients in TGH.

Variable	Category (codes)	frequency	Percent (%)	Log Rank Test		
				Chi-square	df	P-value
Sex	Female (0)	126	44.84	5.18	1	0.023
	Male (1)	155	55.16			
Age	1-11 (0)	140	49.82	0.996	4	0.710
	12-23 (1)	62	22.06			
	24-35 (2)	32	11.39			
	36-47 (3)	29	10.32			
	48-59 (4)	18	6.41			
Residence	Rural (0)	194	69.04	10.91	1	0.001
	Urban (1)	87	30.96			
Season of Diagnosis	Autumn (0)	78	27.76	3.983	3	0.263
	Winter (1)	33	11.74			
	Spring (2)	89	31.67			
Co-morbidity	Summer (3)	81	28.83	14.064	1	0.000
	No (0)	170	60.50			
Sever Acute Malnutrition(SAM)	Yes (1)	111	39.50	12.784	1	0.000
	No (0)	225	80.07			
Treatment types	Penicillin (0)	53	18.86	10.158	3	0.017
	Ceftriaxone (1)	45	16.01			
	Ampicillin (2)	92	32.74			
	Combined (3)	91	32.38			
Patient refer status	No (0)	212	75.44	13.361	1	0.000
	Yes (1)	69	24.56			

df: is degree of freedom, significance level is 0.05, p-value <0.05 is used as significant in this study.

3.1. Bayesian Survival Data Analysis

Bayesian approach analysis was used to make inference about the parameters based on the Gibbs sampler algorithm and it is implemented using 40,000 iterations in three different chains, then after 15,000 terms were discarded due to burn-in state in order to avoid autocorrelation and 75,000 sample obtained for full posterior distribution. The non-informative normal prior distribution with mean zero and variance 1000 and Inverse gamma distribution for sigma with scale=0.01, shape=0.01 parameters [14] were used in this study.

Bayesian approach Accelerated Failure Time Model Comparison.

Table 4. Bayesian AFT model comparison.

Model	Dbar	Dhat	PD	DIC
Weibull	1400.830	1380.550	20.283	1421.110
Log-normal	1430.900	1410.530	20.368	1451.270
Log-logistic	1462.480	1441.670	20.816	1483.300

The distribution with smaller DIC value is a good distribution that fit the data well, due to these Bayesian approach Weibull Accelerating failure time model is selected [16] as the preferable model to analyze the under-five pneumonia dataset based on the DIC value presented in the table 4 above.

Table 5. Posterior summary for Bayesian Weibull AFT model parameter Estimates.

nodes	Variables	mean	Sd	MC-error	median	95% CrI	start	sample
Beta [1]	constant	3.008	0.2859	0.00302	3.078	(1.494, 4.554)*	15001	75000
Sex	Female	Ref.						
Beta [2]	male	-0.125	0.1278	0.00113	-0.1248	(-0.267, -0.0345)*	15001	75000
Residence	Rural	Ref.						
Beta [8]	urban	0.165	0.1397	9.08E-4	0.1647	(0.1025, 0.1712)*	15001	75000
Season D	Autumn	Ref.						
Beta [9]	Winter	-0.148	0.2373	0.00295	-0.1488	(-0.3183, 0.6090)	15001	75000
Beta [10]	Spring	-0.023	0.1142	0.0036	-0.0245	(-0.319, -0.0797)*	15001	75000
Beta [11]	Summer	-0.140	0.2282	0.0097	-0.138	(-0.411, -0.1042)*	15001	75000
Comorbidity	No	Ref.						
Beta [12]	yes	-0.119	0.1438	0.00190	-0.1181	(-1.1, -0.5357)*	15001	75000
SAM	No	Ref.						
Beta [13]	yes	-0.286	0.1362	0.00523	-0.2879	(-0.2482, -0.105)*	15001	75000
prefer status	No	Ref.						
Beta [17]	yes	0.372	0.1454	0.00126	0.371	(0.0916, 0.4623)*	15001	75000
Beta [20]	PNR	0.106	0.1157	0.0052	0.1056	(0.0865, 0.1552)*	15001	75000
Sigma (σ)		3.002	0.1427	0.00391	3.0	(2.721, 3.288)*	15001	75000

Sd=standard deviation, MC error=Mont Carlo error, 95%CrI = 95% credible Intervals.

All parameter estimates in the table 5 above has Monte Carlo error (MC-error) value that is less than 5% of standard deviation which indicates convergence of the parameters. Due to this reason and convergence plots, the researcher uses this posterior summary as final results. The final model results were interpreted using acceleration factor, 95% credible interval of Bayesian approach accelerated failure time estimated values.

When the effect of other factors keep fixed, the estimated acceleration factor for male patient was estimated to be $e^{\beta} = e^{-0.125} = 0.882$ with [95% CrI: -0.2673, -0.0345]. The credible interval for the Bayesian acceleration failure time didn't included zero or on other hand researcher can say that the credible interval for the Bayesian acceleration factor did not include one by exponentiation of the Bayesian acceleration failure time credible interval that is [95% CrI: $e^{-0.2673}, e^{-0.0345}$: 0.765, 0.9661]. This indicates that male patients have less survival time than female patients or in the other way female patients survived 11.8% longer time than male patients. The acceleration factors for patients whose residence was urban were estimated to be 1.179 with [95% CrI: 0.1025, 0.1712]. This indicates that patients whose residence was urban had prolonged survival time than patients from rural residence at 5% level of significance.

Patients diagnosed at spring season and summer season were 0.977 and 0.869 with [95% CrI: -0.3194, -0.0797 and -0.4108, -0.1042] respectively. This implies that patients who were diagnosed at spring and summer season had less survival time than patients who was diagnosed at autumn season. The acceleration factor for patients who were suffered other extra disease or comorbidity was estimated to be 0.888 with [95% CrI: -1.1, -0.5357]. This implies that patients who were not suffered other extra disease had longer survival time than patients who were suffered other extra disease or comorbidity. The acceleration factor for patients who were suffered severe acute malnutrition (SAM) was estimated to be 0.751 with [95% CrI: -0.2482, -0.105]. This indicates that patients who were not suffered severe acute malnutrition (SAM) had longer survival time than patients who were suffered severe acute malnutrition (SAM).

Acceleration factor for patient nurse ratio was estimated to be 1.112 with [95% CrI: 0.0865, 0.1552]. This indicates that patient nurse ratio had significant effect on the survival time of patients.

3.2. Assessment of Convergence

a) Time Series Plots:

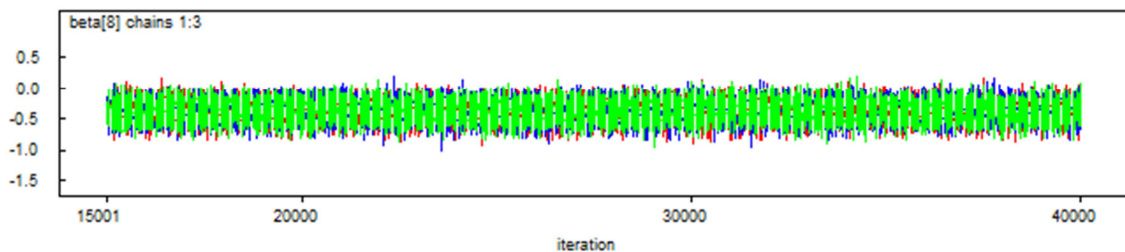


Figure 1. Convergence checking for Residence.

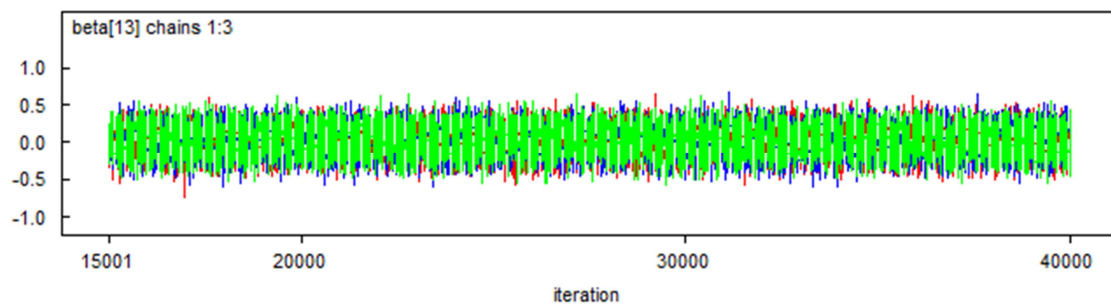


Figure 2. Convergence checking for Sever acute malnutrition (SAM).

b) Autocorrelation plot

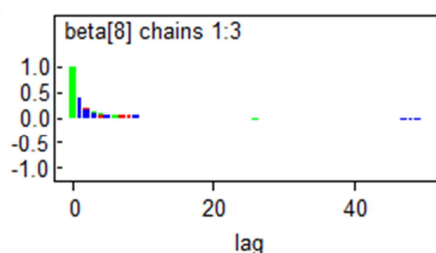


Figure 3. Convergence checking for Residence.

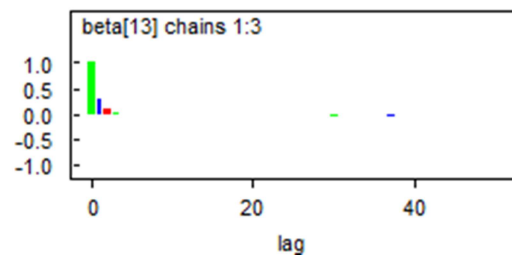


Figure 4. Convergence checking for SAM.

c) Gelman-Rubin Statistics

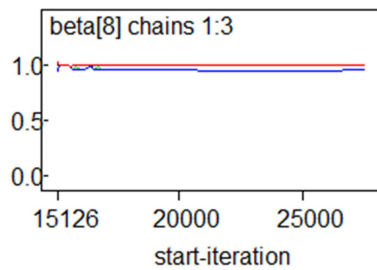


Figure 5. Convergence checking for Residence.

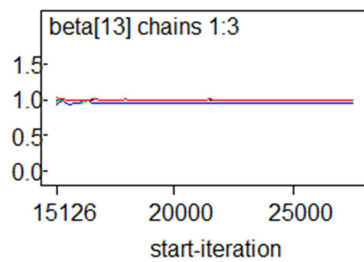


Figure 6. Convergence checking for SAM.

d) Kernel Density Plot

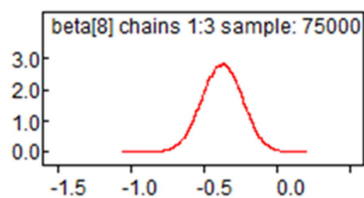


Figure 7. Convergence checking for Residence.

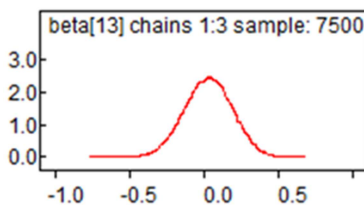


Figure 8. Convergence checking for SAM.

Four different types of plots used in this study to check the convergence of the parameters and indicated that the parameters are converged. To assess the accuracy of Bayesian survival analysis, the researcher used Monte Carlo error of each parameter and checked that MC error value is less than 5% of its posterior standard deviation, then the posterior density is estimated with accuracy. In this study, MC error for each significant variable is less than 5% of its standard deviation. This indicates that convergence and accuracy of posterior estimates are attained and the model is appropriate to estimate posterior statistics.

3.3. Discussion

In this study the Bayesian approach parametric survival models were used to identify the risk factors of survival time of under-five pneumonia patients in Tercha General Hospital. The parametric survival models used in this study were Weibull, Lognormal and Log-logistic distributions. For this

study the source of data was a single center study with 281 under-five pneumonia patients. The Bayesian approach parametric survival analysis is started by MCMC iteration until the convergence of each parameters. The MCMC iterations were generated by setting the initial values and burn-in state without any criteria, since there is no established method for determining an appropriate number of iterations and burn-in size. Rather, the researcher use a trial-and-error process in which the ultimate goal is to obtain stable parameter estimates that minimize simulation error. This statement confirms with study conducted in USA [14]. The MCMC simulation helped to increase the accuracy of the results by narrowing the credible interval and minimizing the standard error, but did not change the direction of the results this agrees with studies [19, 20]. In this study 40,000 samples was generated by using Gibbs sampler algorithm of MCMC iteration method. From those samples 15,000 was used for burn-in state and 25,000 samples with three chains were used for posterior inference using Win BUGS software for iteration and the convergence of the parameters were checked. After the convergence of the parameters three baseline distributions were compared using DIC value and the distribution with smaller DIC value is preferable as stated by Spiegelhalter D. [16]. In this study the Weibull distribution model was selected as a good model to fit the under-five pneumonia dataset in the Bayesian approach survival analysis due to having smaller DIC value.

The results of Bayesian Weibull accelerated failure time model in this study was revealed that female patients had prolonged survival time than male patients and this study agrees with the study conducted in Pakistan by [21] and in JUSH [22]. Patients whose residence was urban had prolonged survival time than patients in rural residence and is in line with study conducted in JUSH [22]. The result of this study revealed that Patients who were admitted in summer and spring season have shorter survival time and had high risk of dying from Pneumonia as compared with autumn and winter seasons; this result agrees with study conducted in Hawassa city [23] and in southern Israel Hospital [24]. And also it is in line with report of CHERG as stated that Altitude, annual rainfall and nature of the seasons and average monthly temperatures are the factors of under-five pneumonia [25].

Patients who were suffered comorbidity or other diseases with pneumonia had shorter survival time than patients without comorbidity, and also under-five pneumonia patients suffered with Severe acute malnutrition (SAM) had shorter survival time than that of patients without Severe acute malnutrition (SAM) and this study agree with the studies conducted in Pakistan [26], in Malawi [27], in JUSH by [22], in southern Israel Hospital [24] and with Child Health Epidemiology Reference Group (CHERG) report [25]. Patients who admitted during patients to nurse ratio (PNR) was high had high risk of dying from pneumonia than patients admitted during PNR low in this study; which is supported by study conducted in Europe that is the higher level of nurse staffing was associated with a decrease in the

risk of in hospital mortality [28] and in Hawassa city [23].

4. Conclusion

The risk factors of survival time of under-five pneumonia patients in the hospital were identified using the Bayesian approach parametric survival models. The Weibull, lognormal and Log-logistic baseline distributions were used in this study. From those baseline distributions Weibull distribution was selected as the good model based on DIC value and Weibull baseline distribution was used as final model to fit under-five pneumonia dataset. Therefore; the results of Bayesian approach Weibull AFT model analysis showed that sex, residence, season of diagnosis, comorbidity, severe acute malnutrition (SAM), patients refer status and patient nurse ratio (PNR) were found to be statistical significant predictors for survival time of under-five pneumonia patients. According to the results obtained from this studies all concerned body should work on awareness by giving health promotions on appropriate and effective treatment in home and early diagnosis to the community to reduce under-five mortality due to pneumonia. The researchers who are interested to investigate on the same area are recommended to introduce frailty modeling to account the correlation which comes from the cluster using Bayesian approach.

Acknowledgements

We would like to acknowledge Jimma University and Mizan-tepi University for their sponsorship and fund for this study. We also thank the Dawro zone department of Health staffs in Tercha General Hospital to undertake this study with their cooperation and permission in using the data with special thanks for Mr. Abashaw Ayele, Dr. Yonatan Fikadu and Dr. Tamirat Tekle for their willingness to help us.

References

- [1] WHO.(2016). Pneumonia [online] Availableat: www.who.int/mediacentre/factsheets/fs331/en/.
- [2] WHO. (2012). Improving health and saving lives by ensuring access to priority medicines.
- [3] Jamison DT, B. J. (2006). Disease control priorities in developing countries. World Bank Publications.
- [4] Li Liu, H. L. (2012). Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet*, 379: 2151–61.
- [5] Amare D. et al. (2007). Determinants of under-five mortality in Gilgel Gibe Field Research Center, Southwest Ethiopia., *Ethiop. J. Health Dev.*, 21 (2): 117-124].
- [6] Klein M. (2005). Survival analysis: techniques for censored and truncated data.
- [7] M. L. Calle et al. (2006). Bayesian survival analysis modeling applied to sensory shelf life of foods. *Food Quality and Preference* 17, 307–312.
- [8] Pascale S., M. W. (2014). Bayesian and Frequentist Comparison for Epidemiologists: The Open Epidemiology Journal, 7, 17-26, Beirut Lebanon.
- [9] Gelfand, A. E., & Mallick, B. K. (2005). Bayesian analysis of proportional hazards models built from monotone functions. *Biometrics*, 51 (3), 843–852.
- [10] Wioletta G. (2013). The Significance Of Prior Information In Bayesian Parametric Survival Models. Institute of Statistics and Demography, Warsaw School of Economics., 285.
- [11] Gelman A. et.al, C. J. (2000). Bayesian data analysis. Chapman.
- [12] Mantel, N. a. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 22, 719-748.
- [13] Christensen R. (2011). Bayesian Ideas and Data Analysis, for Scientists and Statisticians.
- [14] Ghassan H., R. M. (2013). Markov Chain Monte Carlo: an introduction for Epidemiologists,. *International Journal of Epidemiology* 42: 627–634 doi: 10.1093/ije/dyt043.
- [15] Dezfuli H. et al. (2009). Bayesian inference for probabilistic risk and reliability analysis.
- [16] Spiegelhalter D., A. (2004). Bayesian approaches to clinical trials and health-care evaluation.
- [17] Ibrahim J. et al. (2001). Bayesian Survival Analysis. Spring.
- [18] Muluneh S. et al. (2011). Factors Influencing the Intention Not to Use Contraceptives among Sexually Active Women in Ethiopia. *Journals of statistical Society*, vol 20.
- [19] Hakim E., U. M. (2009). Squamous cell carcinoma and keratoa canthoma of the lower lip associated with “Goza” and “Shisha” smoking. *Int J Dermatol*, 108-10.
- [20] Geir Storvik, H. L. (2014). Bayesian methods. *Statistical Machine Learning*.
- [21] Christa L, F. W. (2013). Childhood Pneumonia and Diarrhoea: Global burden of childhood pneumonia and diarrhoea. *Lancet*, 381: 1405–16.
- [22] Firaol B., M. S. (2017). Factors associated with outcomes of severe pneumonia in children aged 2 months to 59 months at JUSH.
- [23] Tariku T. (2017). Modelling Under-Five Mortality among Hospitalized Pneumonia Patients in Hawassa City: A Cross-Classified Multilevel Analysis. *Ann. Data. Sci.* doi: DOI 10.1007/s40745-017-0121-4.
- [24] Lieberman D and Porath A. (2005). Seasonal variation in community-acquired pneumonia in Southern Israel. *Eur Respir J*, 9 (12): 2630–2634.
- [25] Fischer W., (2013). Global burden of childhood pneumonia and diarrhea. *Lancet*, 381, 1405-1416.
- [26] Duke T, P. H. (2002). Chloramphenicol versus benzylpenicillin and gentamicin for the treatment of severe pneumonia in children in Papua New Guinea: *Lancet*, 359: 474–80.

- [27] Ellubey R. (2004). Pneumonia Case Fatality Rate In Children Under-Five.
- [28] Andrea D. (2017). The effect of nurse-to-patient ratios on nurse-sensitive patient outcomes in acute specialist units.
- [29] Ganjali M. and Baghfalaki T. (2012). Bayesian Analysis of Unemployment Duration Data in the Presence of Right and Interval Censoring. *Journal of Reliability and Statistical Studies*, 5 (1), 17-32.