

# Bayesian Survival Dagum 3 Parameter Link Function Models in the Suppression of Dengue Fever in Bojonegoro

Nur Mahmudah and Fetrika Anggraeni

**Abstract**—Dengue Hemorrhagic Fever (DHF) is currently attacking residential areas unprepared to prevent the spread of the *Aedes Aegypti* mosquito. This has resulted in many residents suffering from DHF and eventually decreases economic productivity in the particular area or country. Furthermore, if the infected people's ability to recover from this disease tends to be slow, the economic strait will continue to weaken and the death risk will rise. In addition to health quality and economy, other factors such as knowledge and awareness of the danger of DHF also influence how fast the recovery rate of the infected people in a particular area, especially in Bojonegoro. Taking into consideration these factors, a mathematical modeling can be carried out to estimate the duration of survival rate comprehensively. A survival model is a mathematical model to estimate the duration of a certain population's resistance to an event. This study aims to find out what factors affect the recovery rate of DHF, such as length of hospitalization, sex, age, education, occupation, marital status, hematocrit levels, thrombocyte count, and hemoglobin count. The model used is the Survival Dagum 3 Parameter Link Function which parameters were estimated using the Bayesian MCMC-Gibbs Sampling method. The best survival model found was Dagum 3 parameter with normal distribution random effects. The factors that influence DHF were Sex (X1), Age (X2), Education (X3), Occupation (X4), Hematocrit Level (X5), Thrombocyte Count (X6), and Marital Status (X8).

**Index Terms**—Bayesian; Dagum 3 Parameter; Dengue Hemorrhagic Fever; MCMC-Gibbs Sampling; Survival.

## I. INTRODUCTION

**D**ENGUE hemorrhagic fever (hereafter abbreviated as DHF) is a serious health problem that can lead to hospitalization and significant rate of mortality in tropical Asian countries, including Indonesia, which globally, the number of cases increases significantly every year [1]. DHF is a disease caused by the dengue virus transmitted by *Aedes Aegypti* mosquito [2]. The mosquito spreads the virus from a sufferer to healthy people through bites [3]. Several factors influence the incidence of DHF, including host susceptibility, environment, and the virus itself [4]. The factors that influence the recovery rate of DHF patients are demographic factors including age, sex, and education level [5]. Those factors are interesting to be observed further [6].

Manuscript received September 11, 2020; revised April 26, 2021. This work was supported in part by the Penelitian Dosen Pemula (PDP), Research from the Ministry of Research, Technology, and Higher Education of the Republic of Indonesia (KEMENRISTEKDIKTI-BRIN) 2020 under Grant 96/SK/LPPM-UNUGIRI/III/2020.

Nur Mahmudah is a Lecturer in the Department of Physics, Universitas Nahdlatul Ulama Sunan Giri Bojonegoro, Jalan Jendral Ahmad Yani No.10, Bojonegoro, Jawa Timur, Indonesia (Corresponding author mail: mahmudahnur437@gmail.com).

Fetrika Anggraeni is a Lecturer in the Department of Physics, Universitas Nahdlatul Ulama Sunan Giri Bojonegoro, Jalan Jendral Ahmad Yani No.10, Bojonegoro, Jawa Timur, Indonesia (fetrikaanggraini@gmail.com).

Survival analysis is a set of statistical procedures for analyzing data in which the response variable results from a period of time until an event occurs. It aims to predict the probability of survival, recurrence, death, and other events over a period of time [7]. Survival model is used to explicate that the hazard of an event is influenced by a number of covariates mentioned in supporting theory [8]. Hazard rate is the instantaneous risk of a unit experiment that survives an event at a particular point of time, that is, it does not experience the event in interest until the period is over [9]. Baseline hazard is the risk of an event to occur without taking covariate effect into account, such as time dependency of an event [10]. Survival analysis has three functions, namely the survival function, the hazard function and the probability density function [11].

A survival model is a mathematical model often applied in various fields of research, especially in the health sector. Survival models can also be used to identify risk factors of events and to address situations in which risk factors change over time [12]. Based on this information, the researchers aimed to find out the factors that influence the occurrence of an event. With the risk factor of incidence versus time, the survival model of a tool will be more adequate [13]. Survival analysis has been widely applied in various health or medical fields and is known by various terms in other fields such as event history analysis in sociology, and reliability analysis and duration transition analysis in economics [14].

Research conducted by [15] applied the spatial survival model to political science by modelling the time period until parliament membership structure in the United States government was announced by NAFTA. The application of survival model in political science does not define death as an actual death, but instead, it refers to the survival time of a unit before experiencing a political event. A similar study was carried out by [16] who applied spatial survival model to HIV/AIDS incidence in East Java Province by modelling the time until the patients were pronounced dead, or got referrals to stop Anti Retro Viral Treatment (ART), measured by a 3-parameter Lognormal distribution with spatial effects. In health sciences, survival model is used to observe actual death cases. Similar research conducted by [5] examined a spatial survival model for DHF cases in Makassar by modelling the length of hospitalization until the patients recovered and were discharged and by identifying any censored or failure data. According to a study conducted by [6] using a Bayesian Mixture Survival analysis with Weibull Mixture model distribution on DHF cases, it was found that the duration of hospitalization or survival time of the patients before recovery had mixed distribution.

Previous studies regarding survival model proposed that a good estimate will be produced if the survival data are assumed to have certain distributions such as the 3-Parameter Weibull and the 3-Parameter Log Normal. Unfortunately, not all survival data distributions can be signified clearly. Thus, this study examines the survival Dagum 3 parameter model using Bayesian MCMC approach to estimate the best model and the parameters of Dagum 3 parameter link function and the factors influencing DHF patients' recovery rate. Its analysis result can be presented in DHF management dissemination in order to optimally flatten the case number in Bojonegoro Regency, to raise people's awareness of the disease, and to inform them about the factors influencing the patients' recovery rate. The results can also be the basis for Bojonegoro Health Department to create policies and strategies in accelerating DHF recovery rate.

## II. BAYESIAN MCMC

Bayesian model is based on posterior model which combines historical data as prior information and observation data which derives likelihood function [17]. The estimator of Bayesian approach is the *meanor* mode of the posterior distribution. If a  $\theta$  parameter is considered as a variable, the information before an observation is called prior distribution [18]. When collected, some observation data will display likelihood called the likelihood data [19]. The distribution of posterior data is constructed from the combination of prior information used as prior distribution and sample information represented by the likelihood function [20]. The equation of posterior distribution is as follows [21]:

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)} \propto f(x|\theta)f(\theta) \quad (1)$$

with:

$f(\theta|x)$  = Posterior distribution  
 $f(x|\theta)$  = Likelihood function  
 $f(\theta)$  = Prior distribution  
 $x$  = Data  
 $\theta$  = Parameter

Bayesian method is an alternative method for estimating model parameters [22]. The availability of a program package for Bayesian analysis makes this method more effective and flexible in complex stochastic modeling analysis [15]. As a result, some limitations in classical modeling can be overcome, such as complex models, assumptions that are not in accordance with the reality, and simplifications that can be avoided [23]. One of the Bayesian methods is the Markov Chain Monte Carlo (MCMC), which according to [21], is a numerical approach to obtain a posterior distribution, from a very complicated Bayesian simulation method which is a combination of Monte Carlo and Markov Chain properties to obtain sample data based on a specific sampling scenario [24]. Markov chain is a stochastic process  $\{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(K)}\}$ , therefore, it can be expressed in the following equation [25]:

$$f(\theta^{(K+1)}|\theta^{(K)}, \dots, \theta^{(1)}) = f(\theta^{(K+1)}|\theta^{(K)}) \quad (2)$$

In generating samples from  $p(\theta|x)$ , firstly, the Markov Chain must be arranged on the condition that

$f(\theta^{(K+1)}|\theta^{(K)})$  must be easily generated and the stationary distribution of the Markov Chain is a posterior distribution  $p(\theta|x)$  through the following steps [26]:

- 1) Determine the initial value  $\theta^{(0)}$ .
- 2) Generating samples with as many iterations as K.
- 3) Observe the convergence of the sample data (if the convergence conditions have not been reached then more samples are carried out, continue with step 2 until they are convergent).
- 4) Carry out the burn-in process by removing as much as the first sample B (the burn-in period is the initial iteration period for parameter estimation in the MCMC process to remove B as many as the first iteration in order to eliminate the effect of using the initial value. The burn-in period will end until equilibrium conditions are reached).
- 5) Use  $\{\theta^{(B+1)}, \theta^{(B+2)}, \dots, \theta^{(K)}\}$  as a sample for posterior analysis.
- 6) Create a posterior distribution plot.
- 7) Summarize the posterior distribution such as mean, median, standard deviation and standard error [21].

## III. BAYESIAN 3-PARAMETER DAGUM SURVIVAL LINK FUNCTION

The survival model is semi parametric because it does not require information about the distribution of the underlying survival time and the baseline hazard function does not have to be determined to estimate the parameter [27]. Baseline hazard is a function that is not specific [28]. According to [13], hazard function estimates the probability of the object experiencing an event at time  $t$ . The survival model can be written in the following equation [29]:

$$h(t_{ij}, x_{ij}) = h_0(t_{ij}) \exp(\beta_1 x_{1ij} + \beta_2 x_{2ij} + \dots + \beta_p x_{pij}) \quad (3)$$

In this study, the distribution of length of care (survival time) for DHF patients followed the three-parameter dagum distribution  $(\alpha, \beta, k)$  [30]. The distribution of dagum has a function of opportunity density as follows [16]:

$$f(t; k, \alpha, \beta) = \frac{\alpha k \left(\frac{t}{\beta}\right)^{\alpha k - 1}}{\beta \left(1 + \left(\frac{t}{\beta}\right)^{\alpha}\right)^{k+1}} \quad (4)$$

where  $t > 0$ ,  $k > 0$ ,  $\alpha > 0$ ,  $\beta > 0$  and  $k, \alpha$  the shape parameter and the scale parameter  $\beta$ .  $t$  is the response variable which has a 3-parameter dagum distribution. While the cumulative distribution function is as follows [31]:

$$\begin{aligned}
 F(t) &= p(T \leq t) \\
 &= \int_0^t \frac{\alpha k \left(\frac{t}{\beta}\right)^{\alpha k - 1}}{\beta \left(1 + \left(\frac{t}{\beta}\right)^{\alpha}\right)^{k+1}} dt \\
 &= \left(1 + \left(\frac{t}{\beta}\right)^{\alpha}\right)^{-k}
 \end{aligned} \quad (5)$$

Based on the survival function in equation (5), the survival function of the 3 parameter dagum distribution can be determined as follows [16]:

$$\begin{aligned}
 S(t) &= 1 - F(t) \\
 &= 1 - \left(1 + \left(\frac{t}{\beta}\right)^{\alpha}\right)^{-k}
 \end{aligned} \quad (6)$$

Then the Hazard function based on equation (6) is as follows [32]:

$$h(t) = \frac{f(t)}{S(t)} = \frac{\frac{\alpha k \left(\frac{t}{\beta}\right)^{\alpha k-1}}{\beta \left(1 + \left(\frac{t}{\beta}\right)^{\alpha}\right)^{k+1}}}{1 - \left(1 + \left(\frac{t}{\beta}\right)^{-\alpha}\right)^{-k}} = \frac{\frac{\alpha k \left(\frac{t}{\beta}\right)^{\alpha k-1}}{\beta \left(1 + \left(\frac{t}{\beta}\right)^{\alpha}\right)^{-k-1}}}{1 - \left(1 + \left(\frac{t}{\beta}\right)^{-\alpha}\right)^{-k}} \quad (7)$$

The Survival regression equation in equation (7) can form the 3 parameter dagum distribution model as follows [33]:

$$h(t, X) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p) = \frac{\alpha k \left(\frac{t}{\beta}\right)^{\alpha k-1} \left(1 + \left(\frac{t}{\beta}\right)^{\alpha}\right)^{-k-1}}{\beta \left(1 + \left(\frac{t}{\beta}\right)^{-\alpha}\right)^{-k}} \quad (8)$$

where [16]:

$y$  dagum( $\alpha, \beta, k$ )

$\mu = \beta^T x_{ij} + \varepsilon_i, \quad \varepsilon_i | \varepsilon_{-i} \sim \text{Normal}(a, b), \beta \sim \text{Normal}(s, r)$

The posterior marginal distribution for each of the  $k$ , parameters and is done by integrating out the relevant parameters and can be explained as follows [16]:

$$\begin{aligned} p(\alpha | k, \beta_{1+i}) &\cong \int \int \dots \int_{\beta_1 \dots \beta_{1+p}} I(t | k, \beta_1, \dots, \beta_p) \\ &\quad p(k) p(\beta_1) \dots p(\beta_p) d\beta_1 \dots d\beta_p \\ p(k | \alpha, \beta_{1+i}) &\cong \int \int \dots \int_{\alpha \beta_1 \dots \beta_{1+p}} I(t | \alpha, \beta_1, \dots, \beta_p) \\ &\quad p(\alpha) p(\beta_1) \dots p(\beta_p) d\alpha d\beta_1 \dots d\beta_p \\ p(\beta_0 | \alpha, k, \beta_{1+i}) &\cong \int \int \dots \int_{\alpha k \beta_p} I(t | k, \alpha, \beta_1, \dots, \beta_p) \\ &\quad p(k) p(\alpha) p(\beta_1) \dots p(\beta_p) dk d\alpha d\beta_1 \dots d\beta_p \\ p(\beta_1 | \alpha, k, \beta_{1+i} \neq 1) &\cong \int \int \dots \int_{\alpha k \beta_p} I(t | \alpha, k, \beta_0, \dots, \beta_p) \\ &\quad p(\alpha) p(k) p(\beta_0) p(\beta_2) \dots \\ &\quad p(\beta_{1+p}) d\alpha dk d\beta_0 d\beta_2 \dots d\beta_p \dots \\ &\quad I(t | \alpha, k, \beta, \dots, \beta_{1+i}) \\ &\quad p(\alpha) p(k) p(\beta_0) p(\beta_2) \dots \\ &\quad p(\beta_{1+i}) d\alpha dk d\beta_0 d\beta_2 \dots d\beta_{1+i} \end{aligned} \quad (9)$$

#### IV. METHOD

This research used secondary data of DHF hospitalization record of patients' condition in RSUD Dr. R. Sosodoro Djatikoesoemo, Bojonegoro. The data taken were the length of hospitalization until the patients discharged, called Failure event, and the recording period is from May 1st 2019 to March 30th 2020. The variables taken were Hospitalization Duration ( $Y$ ), Sex ( $X_1$ ), Age ( $X_2$ ), Education ( $X_3$ ), Occupation ( $X_4$ ), Hematocrit Level ( $X_5$ ), Thrombocyte Count ( $X_6$ ), Hemoglobin Count ( $X_7$ ), and Marriage Status ( $X_8$ ). The following table presents the information of response variables and predictors.

The following are the steps to finish the analysis of Bayesian survival Dagum 3 parameter distribution.

- 1) Create the model of Bayesian survival Dagum 3 parameter link function by:
  - Determining the prior and joint distributions.
  - Estimating the survival model parameter using MCMC and Gibbs sampling.
  - Obtaining the model of survival Dagum 3 parameter link function.
- 2) Collect the data of DHF patients at RS Sosodoro Djatikoesoemo.

- 3) Identify the events, and the censored and uncensored data represented below:
  - $\delta$ : 0 is censored data which include patients who experienced failure, such as death, forced discharge, or transfer to other hospital.
  - $\delta$ : 1 is uncensored data which include patients who did not experience failure, such discharge upon better condition or recovery.
- 4) Install "add-ins" of Dagum 3 parameter distribution in WinBUGS as parameter generator through the following steps:
  - Install WinBUGS 1.4
  - Install Blackbox Component Builder
  - Prepare a file containing the connection of new combined distribution to WinBUGS
  - Prepare a template UnivariateTemplate.odc to add new distributions
  - Organize the input needed in UnivariateTemplate.odc to add Dagum distribution which consists of the pdf file, the log-likelihood function, and the CDF of Dagum distribution.
  - Create program coding based on the input in (e) and put it in the corresponding procedure.
  - Compile the program.
  - Validate the program.
- 5) Specify the survival model using open source WinBUGS package. Markov Chain Monte Carlo (MCMC) simulation and Gibb Sampling can be used to determine the survival model and parameter of Dagum distribution by following the steps below:
  - Determine the Likelihood function
  - Signify the prior distribution of Dagum parameters based on the information provided by the data
  - Specify the parameter initialization ( $\alpha, \beta$  and  $k$ ) using 1-Step MCMC
  - Calculate the values of hazard and survival functions in Dagum distribution based on the obtained posterior summaries.
- 6) Determine the mean and variance of the survival Dagum 3 parameter model by estimating its parameters ( $\alpha, \beta, k$ ) through MCMC simulation and gibbs sampling as described in the following steps:
  - Specify the likelihood function.
  - Determine the prior distribution of each parameter based on data information.
  - Determine the initial value of each parameter model using 2-steps MCMC.
- 7) Build T sample  $\theta^1, \theta^2, \dots, \theta^T$  from posterior distribution  $p(\theta | x)$  by updating T n times with enough thin to complete *Markov* Chain process.
- 8) Convergence of algorithm is known as the condition when the algorithm reaches stationary condition in Dagum 3 parameter posterior distribution.
- 9) Obtain the posterior distribution summaries (mean, median, standard deviation, MC error, and confidence interval 95%).
- 10) Choose the best model.
- 11) Create and interpret the survival model of Dagum 3 parameter distribution.

TABLE I  
RESEARCH VARIABLES

No.	Variable	Informasi
1	Time (t)	0=censored 1=uncensored
2	Hospitalization Duration (Y)	Interval
3	Sex ( $X_1$ )	1= male 0= female
4	Age ( $X_2$ )	0 = < 25 years old 1 = 25-50 years old 2 = > 50 years old
5	Education ( $X_3$ )	0 = No formal education 1 = Elementary school or the equivalence 2 = Junior high School 3 = Senior high School 4 = University
6	Occupation ( $X_4$ )	0 = Student 1 = Unemployed 2 = Employed
7	Hematocrit Level ( $X_5$ )	0 = Hematocrit level < 42 1 = Hematocrit level > 42
8	Thrombocyte Count ( $X_6$ )	0 = Thrombocyte count < 150.000 1 = Thrombocyte count > 150.000
9	Hemoglobin Count ( $X_7$ )	0 = Hemoglobin count < 15 1 = Hemoglobin count > 15
10	Marriage Status ( $X_8$ )	0 = Married 1 = Unmarried

## V. RESULT AND DISCUSSION

This study conducted a survival analysis of Dagum 3 parameter function on the factors that affect DHF patients at RSUD Dr. R. Sosodoro Djatikoesoemo, Bojonegoro. TABLE 1 shows the estimated survival time distribution (t) in the length Dengue Fever inpatients by implementing the Kolmogorov Smirnov test through the Easy-Fit Program with the following hypotheses:  $H_0$ : Selection of survival time is in accordance with the dagum estimated distribution of 3 Parameters.  $H_1$ : Selection of survival time is not in accordance with the dagum estimated distribution of 3 Parameters.

TABLE II  
SURVIVAL TIME DISTRIBUTION TEST

Distribution	Statistics Test	p_value	Rank	Decision
Dagum	0.14723	0.15324	4	Accept $H_0$

Based on the testing results of the data distribution of survival time for dengue hemorrhagic fever patients shows that the appropriate predictive distribution is the 3-parameter chin distribution with the p\_value Kolmogorov Smirnov greater than the critical value  $\alpha = 0.05$ . Dagum distribution of 3 parameters is a positive distribution. It shows that the patient will recover after hospitalization in Dr. R. Sosodoro Djatikoesoemo, Bojonegoro. TABLE 2 below presents the patients' survival and hazard functions based on the estimation of Bayesian Dagum 3 parameter distribution versus survival time data:

The result showed that the longer is the hospitalization duration, the higher is the recovery rate and the lower is the survival chance until t time. It means that patients who were hospitalized longer had bigger chance for recovery. For example, the survival chance on day 4 was 0.54. It means that the number of patients who had not recovered on day 4 was 54%. Meanwhile, based on the hazard function, on day 4, the patients' recovery rate was 0.55. It means that the

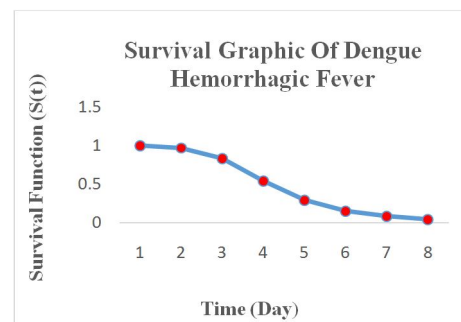


Fig. 1. Survival Function of DHF Patients

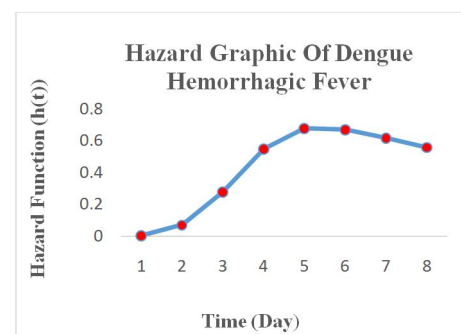


Fig. 2. Hazard Function of DHF Patients

number of patients who recovered on day 4 was 55%. Fig. 1 illustrates the survival function over time. It can be seen that the patients' probability to survive on day 8 nearly reached 0, meaning that all patients had recovered.

Fig. 2 shows that the hazard value or the recovery rate of DHF patients increased. It means that longer hospitalization increased the recovery rate, thus the probability to recover was bigger. However, on day 7, the hospitalization duration decreased, meaning that there was a patient on day 7 who had not recovered. Next, in order to identify whether the normal

TABLE III  
HAZARD AND SURVIVAL FUNCTIONS IN DAGUM 3  
PARAMETER DISTRIBUTION

Day	S(t)	h(t)
1	1.00	0.003
2	0.97	0.07
3	0.83	0.28
4	0.54	0.55
5	0.29	0.68
6	0.15	0.67
7	0.08	0.62
8	0.04	0.56

frailty model was the best one on Dagum 3 parameter, it was then compared to a non-frailty one. Based on the analysis result of TABLE 3 using DIC goodness of fit model, the survival model with random effect showed the lowest DIC value compared to that without random effect. Therefore, it can be assumed that the model with normal random effect is a better model for DHF in Bojonegoro than non-frailty survival model. It can be assumed that there was heterogeneity that could not be explained or represented by the factors in non-frailty survival model.

TABLE IV  
DIC IN SURVIVAL MODEL

Model	DIC Dagum 3 Parameter
Non-Fraily	426.563
Normal Frailty	426.391

The benefit of using frailty model was also shown by the 95% confidence interval on the significant factors that influence DHF patients' survival time as compared to non-frailty model. The efficiency of 95% confidence interval was because the model could specify the covariance structure and the random effect very well. Random effect survival model could reduce bias, thus explaining the heterogeneity effect of the model. Below is the result of Dagum 3 Parameter survival model with normal random effect:

TABLE 4 presents the significant factors of DHF patients recovery rate if the values between 2,5% to 97,5% did not contain 0. It shows that not all factors were significant for the recovery rate. The variable column displays the factors that seemed to predict the recovery rate, mean column displays model parameter values, while the other four columns show the estimated values of confidence interval 97,5%. The factors predicting DHF recovery rate were Sex ( $X_1$ ), Age ( $X_2$ ), Education ( $X_3$ ), Occupation ( $X_4$ ), Hematocrit Level ( $X_5$ ), Thrombocyte Count ( $X_6$ ), and Marriage Status ( $X_8$ ). Furthermore, parameter  $\alpha$  and  $k$  were also significant to the recovery rate. It shows that DHF occurrences cause the increase and decrease of hazard function value or the recovery rate versus time. Below is the Cox Proportional Hazard Regression model [16]:

$$h(t, X) = h_0(t) \exp(5.047 + 0.048X_{1.0} - 0.084X_{1.1} + \dots - 0.843X_{8.1})$$

Next, odd ratio was used to determine the risk level/bias of a certain factor. The odd ratio value shows that the survival rate from death on individuals with bigger hazard factor was  $\exp(\beta)$  times higher than those with lower hazard factor. In the posterior summaries of survival Dagum 3 parameter model link function presented in TABLE 4, it can be seen that Sex ( $X_1$ ) with ( $\hat{\beta} = 0.048$ ) was significant to DHF recovery

TABLE V  
POSTERIOR SUMMARIES OF NORMAL FRAILITY SURVIVAL  
MODEL

Variable	Parameter	Mean	2,50%	Median	97,50%
Random Effect	W	0.046	-3.286	0.011	3.665
Alpha	A	5.122	3.655	5.013	7.162
$X_{1.0}$	$b_1$	0.048	-1.058	0.114	1.034
$X_{1.1}$	$b_2$	-0.084	-1.193	-0.017	0.907
$X_{2.0}$	$b_3$	-1.297	-1.970	-1.295	-0.415
$X_{2.1}$	$b_4$	-1.213	-1.837	-1.203	-0.416
$X_{2.2}$	$b_5$	-0.991	-1.650	-0.985	-0.188
$X_{3.0}$	$b_6$	2.156	1.143	2.124	3.419
$X_{3.1}$	$b_7$	0.952	0.185	1.049	1.614
$X_{3.2}$	$b_8$	0.947	0.167	1.030	1.626
$X_{3.3}$	$b_9$	0.765	-0.002	0.868	1.440
$X_{4.0}$	$b_{10}$	-0.229	-1.438	-0.054	0.868
$X_{4.1}$	$b_{11}$	-0.409	-1.686	-0.303	0.688
$X_{4.2}$	$b_{12}$	-0.361	-1.584	-0.211	0.693
$X_{5.0}$	$b_{13}$	-0.647	-1.371	-0.701	-0.012
$X_{5.1}$	$b_{14}$	-0.567	-1.251	-0.641	0.085
$X_{6.0}$	$b_{15}$	-0.652	-1.576	-0.534	-0.002
$X_{6.1}$	$b_{16}$	-0.688	-1.670	-0.564	-0.032
$X_{7.0}$	$b_{17}$	-0.813	-1.705	-0.909	0.239
$X_{7.1}$	$b_{18}$	-0.751	-1.703	-0.828	0.308
$X_{8.0}$	$b_{19}$	-0.891	-1.617	-0.848	-0.311
$X_{8.1}$	$b_{20}$	-0.843	-1.586	-0.813	-0.229
Constanta	$b_0$	5.047	4.095	4.952	6.104
K	k	0.892	0.407	0.809	1.792
Lambda	$\lambda$	1.558	0.737	1.338	3.716
Rho	$\rho$	0.661	0.073	0.559	1.844

rate at  $\exp(-0.1067) = 0.13$ . It shows that female patients were 0.13 times slower to recover than the male ones. Thus, overall, the death rate of female patients is higher than the male ones because their bodies are more prone to *dengue* virus. Similar interpretation applies to all variables.

The survival analysis result Dagum 3 parameter Dagum link function using Bayesian MCMC method shows that the factors predicting DHF were Sex ( $X_1$ ), Age ( $X_2$ ), Education ( $X_3$ ), Occupation ( $X_4$ ), Hematocrit Level ( $X_5$ ), Thrombocyte Count ( $X_6$ ), and Marriage Status ( $X_8$ ). This finding differs from that of [6] who found three significant factors using Bayesian mixture survival with Weibull distribution, namely Sex, Hematocrit Level, and Thrombocyte Count. Study conducted by [5] using Weibull distribution and Cox semi-parametric found Age, Leukocytes, and Thrombocyte variables to be significant. Another DHF study carried out by [34] in Pakistan using Weibull distribution survival analysis found that Sex and Age were significant. Similar study by [35] concluded that Age and Thrombocyte Count were significant to recovery rate. Thus, it can be concluded that the epidemiologically huge number of DHF cases needs to be overcome by implementing health program to the society through survey system and routine assessment of recovery rate. The result also emphasized the importance of using data distribution in survival analysis. The survival Dagum 3 parameter link function with normal random effect was able to provide information important to be presented in DHF management dissemination in Bojonegoro.

Figure 3 shows that Cox proportional hazard regression dagum three-parameter link function with frailty normal value value or recovery rate for DHF patients is based continuously on Sex ( $X_1$ ), Age ( $X_2$ ), Education ( $X_3$ ), Occupation ( $X_4$ ), Hematocrit Level ( $X_5$ ), Thrombocyte Count ( $X_6$ ), and Marital Status ( $X_8$ ). This shows the recovery rate in Bojonegoro regency same. The recovery rate is affected by dengue fever based on a person's lifestyle factors such

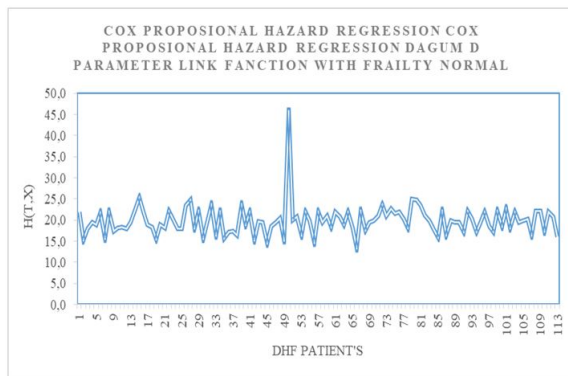


Fig. 3. Cox Propotional Hazard Regression dagum 3 parameter link function with frailty Normal

as maintaining cleanliness and health[42]. One thing that distinguishes it is the width of the interval between the patient's recovery rate because the average random effect parameter has a significant effect on the cure rate [43]. So it can be said that the case of dengue fever in Bojonegoro does depend on the variety component [42] It means that the difference in the value of the variance from the usual random effect results in a different confidence interval for the cure rate for dengue hemorrhagic fever. Cox proportional hazard regression of the three-link function parameters with frailty normal distribution can be used as a basic model for the Bojonegoro regency Health Department's consideration in taking policies to formulate strategic steps to accelerate recovery rate for dengue hemorrhagic fever.

## VI. CONCLUSION

Based on the analysis result, it can be concluded that time distribution of survival Dagum 3 parameter link function model could show the significant factors affecting DHF cases in Bojonegoro. Those factors are Sex ( $X_1$ ), Age ( $X_2$ ), Education ( $X_3$ ), Occupation ( $X_4$ ), Hematocrit Level ( $X_5$ ), Thrombocyte Count ( $X_6$ ), and Marriage Status ( $X_8$ ). This distribution can be used to improve previous researches. Survival Dagum 3 parameter model with MCMC computation method using winBUGS was not only successful in estimating the parameters accurately but also easier for researchers who are able to explain complex models, when the assumptions do not conform with the reality, and to predict the survival time.

## REFERENCES

- [1] S. A. Thamrin, and I. Taufik, "Spatial Random Effects Survival Models to Assess Geographical Inequalities in Dengue Fever Using Bayesian Approach: A Case Study." *In Journal of Physics: Conference Series*, vol. 979, pp. 1-9, 2018.
- [2] H. M. Khormi, and L. Kumar, and R. A. Elzahrany, "Modeling Spatio-Temporal Risk Changes in the Incidence of Dengue Fever in Saudi Arabia: A Geographical Information System Case Study." *Geospatial Health*, vol. 6, no. 1, pp. 77-84, 2011.
- [3] P. Siriyasatien, and A. Phumee1, and P. Ongruk, and K. Jampachaisri4, and K. Kesorn3, "Analysis of Significant Factors for Dengue Fever Incidence Prediction." *BMC Bioinformatics*, vol. 17, no. 1, pp. 1-9, 2016.
- [4] A. Aswi, and S. M. Cramb, and P. Moraga, and K. Mengersen, "Bayesian Spatial and Spatio-Temporal Approaches to Modelling Dengue Fever: A Systematic Review." *Epidemiology and Infection*, pp. 1-14, 2018.
- [5] A. Aswi, and S. Cramb, and E. Duncan, and W. Hu, and G. White, and K. Mengersen, "Bayesian Spatial Survival Models for hospitalization of Dengue: A Case Study of Wahidin Hospital in Makassar, Indonesia." *International Journal of Environmental Research and Public Health*, vol. 3, no.1, pp. 1-12, 2020.
- [6] S. Amalia, and N. Iriawan, and D. D. Prastyo, "Survival Analysis and Factor Influencing The Recovery Of Dengue Hemorrhagic Fever Patient By Using Bayesian Mixture Survival." *Proceedings of the Third International Conference on Mathematics and Natural Sciences(ICMNS)*, pp. 91-97, 2010.
- [7] S. Selvin, "Survival Analysis for Epidemiologic and Medical Research." New York: Cambridge University Press, 2008.
- [8] S. Guo, "Survival analysis." Oxford University Press, 2010.
- [9] P. Makkar, and P. K. Srivastava, and R. S. Singh, and S. K. Upadhyay, "Bayesian Survival Analysis of Head and Neck Cancer Data Using Lognormal Model." *Taylor and Francis*, vol. 43, no. 2, pp. 392-407, 2014.
- [10] P. J. Smith, "Analysis of Failure and Survival Data." CRC PRESS: New York, 2017.
- [11] K. Bogaerts, and A. Komarek, and E. Lesaffre, "Survival Analysis with Interval-Censored Data: A Practical Approach with Examples in R, SAS, and BUGS." New York: Chapman and Hall/CRC, 2018.
- [12] Y. Khan, and A. A. Khan, "Bayesian Analysis of Weibull and Lognormal Survival Models with Censoring Mechanism." *International Journal of Applied Mathematics*, vol. 26, no. 6, pp. 671-683, 2013.
- [13] D. G. Kleinbaum, and M. Klein, "Survival Analysis: A self-Learning Text." 3rd Springer. New York, 2012.
- [14] M. R. Karim, and M. A. Islam, "Reliability and Survival Analysis." Singapore: Springer, 2019.
- [15] D. Darmofal, "Bayesian Spatial Survival Models for Political Event Processes." *Department of Political, Science University of South Carolina*, 350 Gambrell Hal. Columbia, 2008.
- [16] N. Mahmudah, and N. Iriawan, and S. W. Purnami, "Bayesian Spatial Survival Models For HIV/AIDS Event Processes In East Java." *Indian Journal of Public Health Research and Development*, vol. 9, no. 11, pp. 1586-1591, 2018.
- [17] M. A. Erango, "Bayesian Joint Modeling of Longitudinal and Survival Time Measurement of Hypertension Patients." *Risk Management and Healthcare Policy*, vol. 13, pp. 73-81, 2020.
- [18] J. Kruschke, "Doing Bayesian Data Analysis." USA: Elsevier Science, Academic Press, 2014.
- [19] K. M. Banner, and K. M. Irvine, and T. J. Rodhouse, "The Use of Bayesian Priors In Ecology : The Good, The Bad and not great." *Methods In Ecology And Evolution*, vol. 11, no. 8, pp. 882-889, 2020.
- [20] P. Congdon, "Applied Bayesian Modelling." USA: John Wiley and Sons, 2003.
- [21] C. Li, and H. Hao, "Degradation Data Analysis using Wiener Process and MCMC Approach." *Engineering Letters*, V25, N3, pp. 234-238, 2017.
- [22] T. Murniati, and N. Iriawan, and D. D. Prastyo, "Bayesian Mixture Poisson Regression for Modeling Spatial Point Pattern of Primary Health Centers in Surabaya." *MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics*, vol. 36, no. 1, pp. 51-67, 2020.
- [23] J. G. Ibrahim, and M. H. Chen, and D. Sinha, "Bayesian Survival Analysis." Singapore: Springer, 2010.
- [24] N. Iriawan, and S. Astutik, and D. D. Prastyo, "Markov Chain Monte Carlo - Based Approaches for Modeling the Spatial Survival with Conditional Autoregressive (CAR) Frailty." *International Journal of Computer Science and Network Security*, vol. 10 no. 12, pp. 1-12, 2010.
- [25] B. Yu, "A Bayesian MCMC Approach to Survival Analysis with Doubly-Censored Data." *Computational Statistics and Data Analysis*, vol. 54, no. 8, pp. 1921-1929, 2010.
- [26] S. M. Lynch, "Introduction to Applied Bayesian Statistics and Estimation for Social Scientists." New York: Springer, 2017.
- [27] W. M. Bolstad, and J. M. Curran, "Introduction to Bayesian Statistics." Canada: John Wiley and Sons, 2017.
- [28] Q. Feng, and S. Sha, and L. Dai, "Bayesian Survival Analysis Model for Girth Weld Failure Prediction." *Applied Sciences*, vol. 9, no. 1, pp. 1-11, 2019.
- [29] P. C. Austin, "A Tutorial on Multilevel Survival Analysis: Methods, Models and Applications." *International Statistical Review*, vol. 2, no. 85, pp. 185-203, 2017.
- [30] D. F. Moore, "Applied Survival Analysis Using R." New York: Springer International Publishing, 2016.
- [31] A. Wienke, "Frailty Models in Survival Analysis." *Chapman and Hall CRC Biostatistics Series*, New York: Chapman and Hall/CRC, 2010.
- [32] Z. Mai, "Empirical Likelihood Method in Survival Analysis." New York: CRC Press, 2016.
- [33] E. T. Lee, and J. Wang, "Statistical Methods for Survival Data Analysis." Third Edition (Wiley Series in Probability and Statistics), New York: Wiley-Interscience, 2003.



- [34] K. A. Chaudhry, and F. Jamil, and M. Razzaq, and B. F. Jilani, "Survival Analysis of Dengue Patients of Pakistan." *International Journal of Mosquito Research*, vol. 2, no. 6, pp. 5, 2018.
- [35] L. Handayani, and M. Fatekurohman, and D. Anggraeni, "Survival Analysis in Patients with Dengue Hemorrhagic Fever (DHF) Using Cox Proportional Hazard Regression." *International Journal of Advanced Engineering Research and Science (IJAERS)*, vol. 4, no. 7, pp. 138-145, 2017.
- [36] S. Banerjee, and M. M. Wall, and B. P. Carlin, "Frailty Modeling for Spatially Correlated Survival Data, with Application to Infant Mortality in Minnesota." *Biostatistics*, vol. 4, no. 1, pp. 123-142, 2003.
- [37] N. Mahmudah, and H. Pramodyo, "Pemodelan Spasial Survival Weibull-3 Parameter dengan Frailty Berdistribusi Conditional Autoregressive (CAR)." *Natural B*, vol. 3, no. 1, pp. 93-102, 2015.
- [38] P. D. Allison, "Survival Analysis Using SAS origin A Practical Guide Second Edition." USA: SAS Institute Inc, Cary, 2010.
- [39] Zang, "Survival Analysis." California: Wadsworth, 2008.
- [40] S. Momenyan, and A. Kavousi, and T. Baghfalaki, and Poorolajal, J. "Bayesian Modeling of Clustered Competing Risks Survival Times with Spatial Random Effects." *Epidemiology Biostatistics and Public Health*, vol. 17, no. 2, pp. 1-10, 2020.
- [41] P. H. O. Shireesha, and P. T. Sri, and G. Jacob, and Nagarajan, and A. Pravin, "Dengue Prediction Using Machine Learning Techniques." *International Conference on Emerging Trends and Advances in Electrical Engineering and Renewable Energy*, vol. 3, no. 1, pp. 509-517, 2021.
- [42] L. Biard, and A. Bergero, and V. Lévy and S. Chevret, "Bayesian Survival Analysis for Early Detection of Treatment Effects in Phase 3 Clinical Trials." *Contemporary Clinical Trials Communications*, vol. 21, art. id. 100709, 2021.
- [43] X. W. Liu and D. G. Lu, "Survival Analysis of Fatigue Data: Application of Generalized Linear Models and Hierarchical Bayesian Model." *International Journal of Fatigue*, vol. 117, no. 2, pp. 39-46, 2018.



**Nur Mahmudah** received the B.S. degree in Statistics, from Brawijaya University, Malang, Indonesia, in 2010 and the M.S. degree in Statistics from ITS, Surabaya, Indonesia, in 2018, respectively. At present, she is currently a lecturer in the Department of Statistics, Universitas Nahdlatul Ulama Sunan Giri, Bojonegoro, Indonesia. Her research interests include modeling, computational statistics, Bayesian and Stochastic modeling.



**Fetrika Anggraeni** received the B.S. and M.S. degrees from Universitas Negeri Yogyakarta, Yogyakarta, Indonesia, in 2012 and 2015, respectively. At present, she is currently a lecturer in the Department of Statistics, Universitas Nahdlatul Ulama Sunan Giri, Bojonegoro, Indonesia. Her research interests include modeling, economics and statistics.