RedWhitepress https://series.gci.or.id

Global Conferences Series:
Sciences and Technology (GCSST), Volume 2, 2019
The 1st International Conference on Education, Sciences and Technology
DOI: https://doi.org/10.32698//tech1315171

Survival analysis in dengue hemorrhagic fever using Cox proportional hazard model

Suwardi Annas, Nurfadhila Fahmi Utami, Muh. Nusrang

Universitas Negeri Makassar, Makassar, Indonesia

Abstract. Survival analysis is a statistical procedure used to analyze the distribution of time to event data. Dengue hemorrhagic fever data has characteristics that suitable for survival analysis. This study presents a survival analysis to identify the correlation between the times of event of dengue hemorrhagic fever with the measured independent variables by using Cox proportional hazard model. The hazard ratio for platelets obtained that the recovery rate of the patients of dengue hemorrhagic fever with below normal platelet is 2.625 times to the normal platelet count. The result indicated that patients with below normal platelet counts would need a long time to recovery compared than patients with normal platelet counts.

Keywords: Survival analysis, Cox proportional hazard, Dengue hemorrhagic fever

1. Introduction

Survival analysis is a set of statistical procedures that are used to analyze data where the variables considered are the time of an event. Time can be expressed in years, months, weeks, or days from the beginning of observations to an individual until an event occurs (Collet, 2003). The purpose of survival analysis is to determine the relationship between the time of occurrence and the independent variables measured at the time. In addition, it is also used to identify the factors that most influence to the occurance of an event.

Survival analysis is widely used in the field of health research. In this study, survival analysis will be carried out on data of dengue hemorrhagic fever (DHF) patients at the Regional General Hospital (RSUD) in Makassar, Indonesia. DHF is a disease that has a very rapid occurances in the rainy season. In this condition it easily spreads and can cause death. DHF patients especially in the city of Makassar every year has always increased in sufferers and has a high risk of death. Data shows that the number of people with dengue disease in South Sulawesi has continued to increase since of the year 2015.

Several previous studies have been carried out for modeling the spread of DHF. Li , et al (2017) in modeling and projection of dengue fever case in Guangzhou based on variation of weather factors. This study suggested that seasional disease control and mitigation of greenhouse gas emissions could help reduce the incidence of dengue fever. Other study by Anggraeni, et al (2017) have predicted the

number of dengue fever incidents in Malang Indonesia by using modified regrsion approach. They were conclueded that the variability variables can be well explained by the independent variable if both dependent and independent variables have relatively similar variances.

Furthermore, Ernawatiningsih (2012) conducted a survival analysis of patients of DHF with two influencing factors, namely age and platelets, at the Surabaya Hajj Hospital. Fa'rifah and Purhadi (2012) than conducted a survival analysis to determine the factors that influence to the recovery rate of patients of DHF. This study concluded that there were two influential factors, namely age and platelets. Both studies used survival analysis with the Cox regression model approach. Another method that can be used in survival analysis to identify the effect relationship between variables is Cox proportional hazard model.

The proportinal hazard model developed by Cox (1975) is used in survival analysis. This model was initially used in estimating the reliability of mobile handsets (Tiwari and Roy, 2013) and recently by A'Hern (2017) in cancer trials. In this study, Cox proportional hazard will be used to determine the factors that influence the recovery rate of DHF patients. The advantage of Cox proportional hazard model is that it does not depend on the assumption of the distribution of the time it occurred. In addition, this model is a suitable model chosen when in doubt to determine its parametric model (Kleinbaum and Klein 2005).

In this study we developed the fit of Cox proportional hazards for modeling the data of DHF patients. Then, based on this model, we identify the variables that most influence to the rate of recovery of patients with DHF. In addition we also determine how long time it will take to cure DHF patients.

2. Characteristics of Data

DHF is a disease caused by the dengue virus that is transmitted through the bites of mosquitoes Aedes aegypti and Aedes albopictus. There are three factors that cause in the transmission of dengue infection, namely humans, viruses, and intermediary factors. The dengue virus is transmitted to humans through the bite of the Aedes aegypti mosquito. In humans, transmission can only occur when the body is in a state of viremia which is between 3-5 days. To get a higher accuracy of diagnosis, laboratory tests are generally carried out, such as counting the number of antibodies against dengue virus, and complete blood counts such as hemoglobin, leukocytes, hematocrit, and platelets (Hadinegoro & Satari, 1999).

The data used in this study are data of DHF patients obtained from medical records in Makassar City Hospital in 2015. The reason for using this data is because the number of DHF patients in South Sulawesi, especially in the city of Makassar since 2015 continues to increase. For this reason, efforts need to be done to overcome the increase in sufferers of DHF. One effort was made by looking at the recovery rate of DHF patients. Whereas the independent variables identified influence the recovery rate of patient survival time (Y) of DHF, namely; patient age (X1), gender (X2), hemoglobin number (X3), leukocyte count (X4), hematocrit percentage (X5), platelet count (X6), and body temperature (X7). In observing the recovery rate, we performed a survival analysis by using Cox proportional hazard models.

This study involved 105 samples of DHF patients in Makassar City General Hospital during the period January to December 2015. The average cure rate in 105 DHF patients was around 5 days. The average age of patients is 17 years old and more are male. The number of platelet content of patients is below normal with an average leukocyte level of 5.30/µl and an average hematocrit level of 14 g/dl. The average hematocrit level of patients is 39% and the patient's body temperature averages around 38oC.

3. Method

In survival analysis there are two models, namely parametric model and semiparametric model. The Cox proportional hazard regression model is a semiparametric distribution model because it does not require information about the distribution underlying survival time and regression parameters can be estimated from the model. This semiparametric model, although the functional form $h_0(t)$ is unknown, but it can still provide information in the form of a hazard ratio that does not depend on $h_0(t)$. Hazard ratio is defined as the ratio of the hazard rate of one individual to the hazard rate of another individual.

One of the objectives of the Cox proportional hazard model is to model the relationship between survival time and variables that are thought to influence survival time. Therefore, it is necessary to do proportional hazard assumption on the data to form Cox proportional hazard model. In this study, Testing assumptions used Global test with a significance level of $\alpha = 0.05$. Furthermore, the step of data analysis is started with perform Cox proportional hazard modeling as written below.

$$h_i(t) = h_0(t)\exp\left(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p\right) = h_0(t) e^{\sum_{j=1}^{p} \beta_j x_j} \quad (1)$$

Where:
$h_i(t)$ = individual failure function of i
$h_0(t)$ = basic failure function (Hazard function)
x_j = value of j-variable, with $j = 1,2, \ldots, p$
β_j = regression coefficient of j, with $j = 1,2. \ldots, p$

3.1 Parameter Estimation

The parameter estimator of the independent variables of $X_1, X_2, X_3, \ldots, Xp$ are given by $\beta_1, \beta_2, \beta_3, \ldots, \beta p$. The coefficient β in the proportional hazard model can be estimated using the Maximum Likelihood Estimator (MLE) method. Parameter estimation is carried out to determine the effect each independent variable to the dependent variable which has the potential to be a risk factor. To make it easier to get parameter estimating values, Newton-Raphson iteration method can be used. The likelihood function is stated as follows:

$$L(\beta) = \prod_{j=1}^{r} \frac{\exp(\beta x_{(i)})}{\sum_{i \in R(t_i)} \exp\ (\beta x_{(i)})} \quad (2)$$

Where:
$x_{(i)}$ = variable vectors of individuals that fail when i with t_i.
$R(t_i)$ = all individuals who have a risk of failure at the time of i.

3.2 Selecting of the best model

Testing for the significance of parameters include simultaneous test and partial test (Hosmer, et al., 2008). In order to test the hypothesis of one or several regression β_i is zero, it can use a simultaneous test with the partial likelihood ratio. This test statistic follows a chi-square distribution with a degree of freedom p. The hypothesis used is:

H_0: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$
H_1: at least one $\beta_k \neq 0$, with $k = 1, \ldots, 7$

Rejection of H_0 if $G \geq X^2_{(\alpha;\ db\ =\ p)}$ or p-value $\leq \alpha$ which means that there is at least one independent variable that affects survival time.

The partial test is then used to know the partial effect each independent variables to dependent variable. If there are independent variables not significant, then it is necessary to reduce the independent variable. The hypothesis used is:

$H_0: \beta_k = 0$, with $k = 1,2,\ldots,p$
$H_1: \beta_k \neq 0$, with $k = 1,2,\ldots,p$

Rejection of H_0 if $W > X^2_{(1;df=p)}$ or p-value < α which means that independent variables affect survival time.

Furthermore, according to Tustianto and Soehono (2012) that the selection of the best models using Akaike Information Criterion (AIC). The best model has the smallest AIC value of

$$AIC = -2\log\hat{L} + 2P \quad (3)$$

Where:
$\log\hat{L}$ = the maximum value of the likelihood function Cox proportional hazard model
P = the number of independent variables

The selection of the best Cox proportional hazard model is done by selecting the variables which can be done in three ways, namely forward selection, backward selection and stepwise procedures. Forward selection or advanced selection by adding variables one by one in each step. Backward selection is the process of eliminating variables that enter the model, starting with removing or deleting one by one according to the significance criteria. The stepwise selection procedure is a combination of two processes, namely forward and backward selection.

3.3 Hazard Ratio

The patient's recovery rate can be seen from the hazard ratio or odds ratio Dahlan, (2013) said that the value of the hazard ratio is a measure used to determine the level of risk (tendency). It can be seen from the comparison between individuals with the conditions of independent variables in the success category with the failure category.

For example X is an independent variable with two categories, namely 0 and 1. The relationship between variable X with $h_0(t)$ is expressed by $h_0\,(t|x) = h_0(t)e^{\widehat{\beta x}}$, then:

Individual with x = 1, hazard function:

$$h_0\,(t|x=1) = h_0(t)e^{\widehat{\beta.1}} = h_0(t)e^{\hat{\beta}} \quad (4)$$

Individual with x = 0, hazard function:

$$h_0\,(t|x=0) = h_0(t)e^{\widehat{\beta.0}} = h_0(t) \quad (5)$$

Hazard ratio for individuals with x = 0 compared with x = 1 is:

$$Hazard\ Ratio = \frac{h_0\,(t|x=0)}{h_0\,(t|x=1)} = \frac{h_0(t)}{h_0(t)e^{\hat{\beta}}} = e^{-\widehat{\beta}} \quad (6)$$

If the value of independent variable with a hazard rati less than 1, then the increase in the value of the independent variable is related to a decrease in the risk of death and a longer survival time. When a hazard ratio more than 1, then the increase in the value of the independent variable is associated with the increase of risk of death and shorter survival time.

4. Results and discussion

Fulfillment of assumptions in survival analysis is very important to ensure that the survival function used meets Cox proportional hazard assumptions. This study uses assumption testing with a Global test or Goodness of Fit test. The result of assumption testing for each variable indicate that each

variable has a *p*-value > α = 0.05 so that the proportional hazard assumption is fulfilled for all independent variables. Furthermore, estimating the parameters of the Cox proportional hazard model using the Maximum Likelihood Estimator (MLE) method. Simultaneous parameter testing of all variables indicates that the Likelihood test value is 17.27 with a significance level of 0.015 < α = 0.05. These results mean that the overall model can contribute to the recovery rate of DHF patients. This also indicates that there is at least one significant independent variable so that it can be continued with partial parameter testing.

The results of partial parameter testing in Table 1 show that variables that have *p*-value < α = 0.05 are only platelet variables. It can be concluded that the platelet variable is a variable that has a significant effect on the model while the other variables have no significant effect. This means that platelets can contribute to the recovery rate of dengue hemorrhagic fever patients. These results are relevant with the results of survival analysis with cox regression conducted by Fa'rifah and Purhadi (2012) who concluded that besides platelet, the age of patient is as a significant factor influence the recovery rate of DHF.

Table 1 Partial test for all independet variables

Variables	$\widehat{\boldsymbol{\beta}}$	***p-value***	**Sig. criteria**
Age (X_1)	0.0061	0.572	not significant
Gender (X_2)	-0.1520	0.470	not significant
Hemaglobin (X_3)	-0.2515	0.098	not significant
Leucocyte (X_4)	0.0429	0.206	not significant
Hematocrit (X_5)	0.0949	0.110	not significant
Platelet (X_6)	-0.8054	0.030	significant
Body temperature (X_7)	-0.1754	0.066	not significant

The selection of the best model is then carried out by the backward selection method. It begins by removing or deleting one by one the variables according to the significance criteria. In order to choose the best model, the smallest AIC value of the model will be consedered. The AIC value in Table 2 show that the selection of the best model in step 1 includes all variables into the model, then for step 2 and so on, and reduce the variables one by one. The backward selection process stops at step 6, where the model formed is a model leaving only two independent variables, namely platelet variable (X_6) and body temperature (X_7) with the lowest AIC value of 747.06.

Table 2 AIC values for best model selection

Model	**Variable**	**AIC**
1	All free variable are entered	752.08
2	Without age variable (X_1)	750.38
3	Without age variable (X_1) and gender (X_2)	748.89
4	Without age variable (X_1), gender (X_2), and leucocyte (X_4)	748.42
5	Without age variable (X_1), gender (X_2), leucocyte (X_4), and hematocrit (X_5)	748.06
6	Without age variable (X_1), gender (X_2), leucocyte (X_4), hematocrit (X_5), and hemoglobin (X_3)	**747.06**

Based on the backward selection process, the model formed is a model without age (X_1), gender (X_2), hemoglobin (X_3), leukocytes (X_4) and hematocrit (X_5). After estimating the parameters of the remaining dependent variables, namely platelet (X_6) and body temperature (X_7), the best Cox proportional hazard regression was obtained with the following model.

$$h(t) = h_0(t)\exp(-0.96418\,X_6 - 0.17331\,X_7) \qquad (7)$$

To find out the recovery rate of patients with dengue disease can be done by looking for the hazard ratio value of the variables included in the best model. The value of the hazard ratio is a measure used to determine the level of risk (failure) that can be seen from the comparison between individuals. For platelet variables in the types of normal and below normal categories which have been categorized as 1 and 0, the hazard ratio values as presented in Table 3.

Table 3 Hazard ratio value for platelet variation and body temperature

Variable	$\widehat{\boldsymbol{\beta}}$	***Hazard Ratio*** $(e^{-\widehat{\beta}})$
Platelet (X_6)	-0.9642	2.625
Body temperature (X_7)	-0.1733	1.189

According parameter estimates in Table 3, the hazard ratio for platelet variables is 2.625. This value means that the recovery rate of patients in dengue hemorrhagic fever with platelet counts below normal is 2.625 times the normal platelet count. Whereas for body temperature variables, the hazard ratio value is 1.189 or more than 1 which means that the higher the temperature of a patient's body, that cause the patient's recovery rate will be longer. These results support the study conducted by Ernawatiningsih (2012) who concluded that there was a significant relationship between platelet counts and the recovery rate of DHF patients.

5. Conclusion

The study presented in this paper was using Cox proportional hazard analysis in DHF. The parameter values can be well estimated by Maximum Likelihood Estimator method. Backward selection method has than selected ca the best model of Cox proportional hazard by involving two independent variables that influence the recovery rate of dengue fever patients, namely platelets and body temperature. However, based on the parameter testing indicate that only platelet variable has a significant effect on the recovery rate of dengue hemorrhagic fever patients. In the future study, it will be compared Cox proportional hazard analysis with others semiparametric statistical approach that can be modeling the DHF data.

6. References

A'Hern, R. P. (2018). Cancer Biology and Survival Analysis in Cancer Trials: Restricted Mean Survival Time Analysis versus Hazard Ratios, Clinical Oncology, Elsevier, 30 (75-80).

Anggraeni, W., Nurmasari, R., Riksakomara, E., Samopa, F., Wibowo, R. P., Lulus, C. T., Pujiadi (2017). Modified regression approach for predicting number of dengue fever incidents in Malang Indonesia. Procedia Computer Science, Elsevier, 124 (142-150).

Collectt, D. (2003). *Modelling Survival Data In Medical Research "Second Edition"*. Chapman & Hall: New York.

Cox, D. R. (1975). Partial Likelihoohd. Biometrika, 62 (269-276).

Dahlan, M. S. (2013). Survival Analysis "The Basics of Theory and Application of the Stata Program". Sagung Seto: Jakarta. (In Indonesian)

Ernawatiningsih, N. P. L. (2012). Survival Analysis With Cox Regression Model. Mathematical Journal, 2 (2), 1693-1394. (In Indonesian)

Fa'rifah, R. Y., Purhadi, P. (2012). Survival Analysis of Factors Affecting the Healing Rate of Patients with Dengue Hemorrhagic Fever (DHF) in Surabaya Hajj Hospital with Cox Regression. ITS Science and Art Journal, 1 (1), D271-D276. (In Indonesian)

Hadinegoro, S. R. H., Satari, H. I. (2002). Dengue Hemorrhagic Fever (Training for trainers of Pediatricians & Internists in the Management of DHF Cases). FK UI: Jakarta. (In Indonesian)

Hosmer, D. W., Lemeshow, S., Mya, S. (2008). Applied Survival Analysis: *Regression Modelling of Time to Event Data*. New Jersey: John Wiley.

Kleinbaum, D. G., Klein, M. (2005). *Survival analysis: a self-learning text*. Springer-Verlag: New York. April 4, 2016.

Li, C., Wang, W., Wu, X., Liu, J., Ji, D., Du, J. (2017) Modeling and projection of dengue fever case in Guangzhou based on variation of weather factors. Science of the Total Environment, Elsevier, 605-606 (867-873).

Tiwari, A., Roy, D. (2013) Estimation of Reliability of Mobile Handsets Using Cox-proportional Hazards Model. Microelectronics Reliability, Elsevier, 53: 481-487.

Tustianto, K., Soehono, L. A. (2012). Cox regression modeling proportional hazard old factors of Malang City IMB process. Mathematical Journal. October 3, 2016. (In Indonesian)

Acknowledgment

The authors are grateful to all the DHF patients and Makassar City Hospital for supporting the DHF data.