HAZARD PROPORTIONAL REGRESSION STUDY TO DETERMINE STROKE RISK FACTORS USING BRESLOW METHOD

Sudarno, Eri Setiani

Department of Statistics, Faculty of Science and Mathematics, Diponegoro University

e-mail: dsghani@gmail.com

Abstract: Cox proportional hazard regression is a regression model that is often used in survival analysis. Survival analysis is phrase used to describe analysis of data in the form of times from a well-defined time origin until occurrence of some particular be death. In analysis survival sometimes ties are found, namely there are two or more individual that have together event. The objectives of this research are applied Cox proportional hazard regression on ties event using Breslow method and determine factors that affect survival of stroke patients in Tugurejo Hospital Semarang. The response variable is length of stay at hospital, and the predictors are gender, age, type of stroke, history of hypertension, systolic blood pressure, diastolic blood pressure, blood sugar levels, and body mass index. The factors cause stroke disease by significant are type of stroke, history of hypertension, systolic blood pressure, diastolic blood pressure, and blood sugar level. By the survivorship function that the patients have been looked after at hospital greater than 20 days, they have probability of healthy be little even go to death. A person in order to be healthy must notice and prevent some factors cause disease.

1. INTRODUCTION

Survival analysis is a method to describe data analysis in the form of time, that is, from the early detection until the outbreak of certain event (Collet, 2015). According to Kleinbaum and Klein (2005), the intended time is time in units of years, months, weeks, or days from the beginning of the individual's observation until the occurrence of desired event. Conversely, the related event or incident is death, recurrence of disease after treatment, recovery or experience determined by the researcher for the benefit of the individual under study. Survival analysis applies the concept of censorship. Censorship occurs when events have information about the life time of individuals, but do not tell for certain the time of death. Thus, the data is said to be censored (Lee and Wang, 2013).

To analyze survival data, we can use the survival regression method to find the relationship between the independent variable (predictor) and life time as the dependent

variable (response). This study used the semiparametric regression model of Cox proportional hazard regression. This is because the Cox proportional hazard regression model does not require assumptions about the distribution of survival time (Lee and Wang, 2003). In survival analysis, common events are often called ties, which are defined as a situation when two or more individuals experience an event at the same time. Such joint events can cause problems in the estimation of parameters related to the determination of members of the set of risks. Because the estimation results of the parameters can be biased, it can be solved using the weight similarity method (Xin, *et al.*, 2017).

Several previous studies on the Cox proportional hazard regression analysis include the one conducted by Sudana, *et al.* (2013) pinpointing that the factors influencing the length of time to look for work is residence, sex and marital status. Marisa, *et al.* (2016) explained that the length of the study was influenced by the performance index in semester 2, 3 and 4 and the independent admission. Whereas Dewi, *et al.* (2016) revealed that the patient's diet of diabetes mellitus was influenced by gender, blood sugar level status, and concomitant diseases. Also Faruk (2017) articulated that the birth interval of the first child is influenced by age, parental education and economic level.

This study used Cox proportional hazard regression with the Breslow approach to look for a significant effect of the predictors consisting of patient status, sex, age, type of stroke, hypertension record, systolic blood pressure, diastolic blood pressure, blood sugar level at a time, and index body mass response to the variable response time to survival. In addition, the survivorship function was used to predict the relationship between time of illness and the chance of recovery. The Breslow method assumed that the size of the set of risks in a joint event is the same (Xin, *et al.*, 2017). The sample of patients in this study were stroke patients at the Regional General Hospital, RSUD Tugurejo, Semarang City. This study is expected to provide a model that can provide information of significant factors on the causes of stroke. Knowing the risk factors of stroke, a person can prevent their occurrence, receive good medical treatment, stay healthier and have a long life.

2.   LITERATURE REVIEW
2.1. Cox Proportional Hazard Regression and Survivorship Function

There are three conditions to determine survival time, namely the clear definition of origin of time, the determination of time scale, and the clear meaning of failure such as a recurrence or recovery of a disease (Cox and Oakes, 1998). According to Collet (2015), the value of the independent variable in the proportional hazard model is represented by vector $\mathbf{x}$, where $\mathbf{x} = (x_1, x_2, ..., x_p)'$. The hazard function $h_i(t)$ for i-th individuals is written with

$$h_i(t \mid \mathbf{x}_i) = \psi(\mathbf{x}_i) h_0(t) \tag{1}$$

where $\psi(\mathbf{x}_i)$ is the function of the vector value of the independent variable for the i-th individual:

$$\psi(\mathbf{x}_i) = \exp(b_1 x_{1i} + b_2 x_{2i} + \cdots + b_p x_{pi}) \tag{2}$$

In brief, the proportional hazard model can be written as

$$h_i(t \mid \mathbf{x}_i) = \exp(\mathbf{b}'\mathbf{x}_i) h_0(t) \tag{3}$$

where $h_i(t \mid \mathbf{x}_i)$: i-th individual failure function

$h_0(t)$: basic failure function

b'   : coefficient vector of the independent variable $(b_1, b_2, \ldots, b_p)$, measuring 1 x p.

$\mathbf{x}_i$: independent variable value vector $(x_{i1}, x_{i2}, \ldots, x_{ip})'$, measuring p x 1 with $i = 1, 2, \ldots, n$.

By considering the nature and proportional conditions or ratios, equation (3) is equivalent to

$$S(t \mid \mathbf{x}) = \left[S_0(t) \right]^{\exp(\mathbf{b'x})} \tag{4}$$

known as the survivorship function, where $S_0(t)$ is a basic survey function that is time-dependent t. In addition, in natural logarithms equation (3) is expressed as

$$\log \frac{h_i(t \mid \mathbf{x}_i)}{h_0(t)} = \mathbf{b'x}_i \tag{5}$$

which is called a function of the hazard ratio or relative risk. Relative risk is a function of multiple linear regression without constants or regression functions through a central point. The risk function is independent of time t (Lee *et al.*, 2013).

2.2. Assumption of Cox Proportional Hazard Regression

According to Lee and Wang (2013) the Cox proportional hazard regression model does not require information about the distribution that underlies the survival time. The proportional hazard function in the model can take any form of distribution, but it should fulfil an assumption, namely the proportional hazard assumption. The proportional hazard assumption states that the independent variable (predictor) must be independent or independent of time. Testing of the proportional hazard assumption through the Goodness of Fit approach is conducted through residual Schoenfeld. According to Kleinbaum and Klein (2005), the steps for testing the proportional hazard assumption are as follows:
1. Constructing a proportional Cox hazard and residual Schoenfeld model for each individual on each independent variable.
2. Determining variables that state rank of the survival time.
3. Testing the correlation between residual Schoenfeld variables with survival time in step 2. According to Nurfain and Purnami (2017) statistical testing for the i-individual independent variable uses the following hypothesis.
   Hypothesis:

   H_0 : there is no correlation between rank variable and residual Schoenfeld values of each variable (proportional hazard assumption fulfilled).
   H_1 : there is a correlation between rank variables with the Schoenfeld residual value of each variable (proportional hazard assumption not met).

Significance Level: α

Test Statistics:

$$r_h = \frac{\sum_{i=1}^{n} \left(R_{ji} - \bar{R}_{ji} \right) \left(RT_i - \overline{RT_i} \right)}{\sqrt{\sum_{i=1}^{n} \left(R_{ji} - \bar{R}_{ji} \right)^2} \sqrt{\sum_{i=1}^{n} \left(RT_i - \overline{RT_i} \right)^2}} \tag{6}$$

where

r_h : hazard correlation coefficient

R_{jt} : Schoenfeld residual for each j-th independent variable where $j = 1, 2, \ldots, p$ for i-th individual where $i = 1, 2, \ldots, n$.

RT_i : i-th individual survival time.

Rejection Criteria:

H_0 is rejected if $|r_h| > r_{n-2,\frac{\alpha}{2}}$ or the value of $-p < \alpha$.

2.3. Parameter Estimation and Testing

Survival analysis sometimes contains simultaneous events called *ties*. Collet (2015), defined *ties* as conditions when there are two or more individuals experiencing an event of failure at the same time or having the same survival time value. Such simultaneous events can cause problems in the estimation of parameters related to the determination of members of the set of risks. To approach simultaneous events, we used the Breslow method. The Breslow method assumes that the size of the set of risks in *ties* is the same. Following is the likelihood function for the Breslow method:

$$L(\mathbf{b}) = \prod_{i=1}^{n} \frac{\exp(\mathbf{b}'\mathbf{s}_i)}{\left(\displaystyle\sum_{l \in R(t_i)} \exp(\mathbf{b}'\mathbf{x}_i) \right)^{d_i}} \tag{7}$$

Where t_i : time of the i-individual event

$\mathbf{s}_i$: variable vectors of all individuals who died at the time of t_i

$R(t_i)$: risk set of individuals who survive at the time of t_i

$\mathbf{x}_i$: variable vectors of individuals who still survive and are elements of $R(t_i)$

d_i : the number of simultaneous events at t_i

According to Hosmer *et al.* (2008) testing the significance of parameters can use the likelihood ratio test and the Wald test. This test aims to determine the effect of predictor variables in the model. The likelihood ratio test is used for simultaneous tests, while the Wald test is used for partial tests of parameters.

a. Likelihood Ratio Test

Hypothesis:

H_0 : Together the predictors do not affect the model
H_1 : at least one predictor affects the model

Significance Level: α

Test Statistics:

$$G = -2[\ln L(0) - \ln L(b)]$$

where $\ln L(0)$: log partial likelihood of a model without independent variables.

$\ln L(b)$: log partial likelihood of the model consisting of p independent variables.

Rejection Criteria

H_0 is rejected if $G \geq \chi^2_{p,\alpha}$ or the value of $-p < \alpha$

b. Wald Test

Hypothesis:

H_0: j-predictor has no effect on survival time

H_1: j-predictor influences survival time, $j = 1, 2, \ldots, p$

Significance Level: α

Test Statistics:

$$Z^2 = \left[\frac{b_j}{se(b_j)} \right]^2$$

where $se(b_j) = \sqrt{\mathrm{var}(b_j)}$

Rejection Criteria

H_0 is rejected when $Z^2 > \chi^2_{1;\alpha}$ or the value of $-p < \alpha$

2.4. Failure Ratio

According to Kleinbaum and Klein (2005), the failure ratio is failure for one group of individuals divided by failure for another group of individuals. The two groups of compared individuals were distinguished by the value of the independent variable. The failure ratio is stated in the following formula:

$$HR = \exp\left(\sum_{j=1}^{p} x_{ij}^{*} - x_{ij} \right) \tag{8}$$

where x_{ij} : variable value of x_j, of i-th individual, where $i = 1, 2, \ldots, n$

x_{ij}^{*} : variable value of x_j^{*}, the-i-th individual.

According to Allison (2010) the ratio of failure to independent variables in the Cox proportional hazard model which is categorical with dummy variables of 1 and 0 can be interpreted as a comparison of the failure predictors for individuals who are 1 to the failure predictors of individuals who are 0. While the comparison result of independent quantitative or continuous variable will be more meaningful if the failure ratio is expressed as a percentage.

2.5. Stroke

Stroke is a type of non-communicable disease. In medical language stroke is called celebro-vascular accident (CVA). Stroke is defined as a permanent nerve disorder due to disruption of blood circulation to the brain that occurs around 24 hours or more (Lingga, 2013). Based on its trigger, the stroke is divided into two types, namely ischemic stroke and hemorrhagic stroke. Ischemic stroke is a type of stroke that is attributable to blocked blood vessels of the brain by plaque (material consisting of protein, calcium, and fat) which causes the blocked flow of oxygen through the arteries. The hemorrhagic stroke is a type of stroke caused by brain hemorrhage due to rupture of blood vessels in the brain.

Stroke risk factors are divided into two, namely uncontrolled factors (fixed factors) and controlled factors (non-fixed factors). Uncontrollable factors in the form of genetic factors (race) are age, gender, and history of illness experienced by parents or siblings. While controlled factors are in the form of smoking, alcoholic consumption and sedentary lifestyle (Ministry of Health, Republic of Indonesia, 2014).

3. RESEARCH METHODS

3.1. Type and Source of Data

This study used secondary data obtained from the medical records of stroke inpatients at the Tugurejo Regency General Hospital, Semarang City from January 2018 to December 2018. Data processing was conducted using SAS software in Allison (2010) and R in Fox & Wisberg (2018).

3.2. Research Variables

The variables used were length of survival, status, sex, age, type of stroke, hypertension record, systolic blood pressure, diastolic blood pressure, blood sugar levels and body mass index. The study observed 10 variables. The response variable is survival time while the covariate is 9 other variables. Full variable descriptions are listed in the following Table 1.

Table 1 Research Variable

No	Variable	Description
1	Survival Time (T)	Length of Treatment of stroke patients (days)
2	Patient Status (Cens)	0 = survived patients during treatment, 1 = dead patient
3	Sex (X_1)	0 = female, 1 = male
4	Age (X_2)	age (tahun)
5	Type of Stroke (X_3)	0 = ischemic stroke, 1 = hemorrhagic stroke
6	Hypertension Record (X_4)	0 = non hypertension, 1 = hypertension
7	Systolic Blood Pressure (X_5)	mmHg
8	Diastolic Blood Pressure (X_6)	mmHg
9	Blood Sugar Level (X_7)	mg/dl
10	Body Mass Index (X_8)	kg/m^2

3.3. Data Analysis

The analysis covers the following steps:
1. Conducting descriptive analysis,
2. Standardizing the continuous free variables using Z-Score normalization.
3. Cox proportional hazard regression modeling using the Breslow method, by way of
   a. Making an initial Cox proportional hazard regression model.
   b. Testing the Cox proportional hazard assumption using the Goodness of Fit test.
   c. Conducting parameter significance tests consisting of simultaneous tests and partial tests.
   d. Forming the final Cox proportional hazard regression model and the best survivorship function.
4. Interpreting the best model of Cox proportional hazard regression that has been formed.

4. RESULTS AND DISCUSSION

The research process and discussion are presented in the following description in the form of tables and explanation.

4.1. Descriptive Analysis

Descriptive analysis was used to show the characteristics of stroke patients at Tugurejo Regency Hospital, Semarang City. This study involved 288 patients as sample size with 51 samples or 18% of the sample was observable or uncensored data and the remaining 237 samples or 82% were censored data. Descriptive analysis for continuous independent variables is presented in Table 2.

Table 2 Descriptive Statistics for Continuous Independent Variables

Variables	Average	Minimum	Maximum
Age	59	25	89
Systolic Blood Pressure	170	70	257
Diastolic Blood Pressure	97	14	174
Blood Sugar Level	160	61	957
Body Mass Index	24.21	14.10	39

Descriptive analysis for the categorical independent variables are presented in Table 3.

Table 3 Descriptive Statistics for Categorical Independent Variables

Variables	Category	Status		Total
		Uncensored	Censored	
Sex	Female	20	90	110
	Male	31	147	178
Type of Stroke	Ischemic	22	197	219
	Hemorrhagic	29	40	69
Hypertension Record	No	14	33	47
	Yes	37	204	241

4.2. Cox Proportional Hazard Regression Modeling Using the Breslow Method

Initial Modeling

Cox proportional hazard regression models are used to determine the effect of predictors on response. This study used the independent variables of gender (X_1), age (X_2), type of stroke (X_3), hypertension record (X_4), systolic blood pressure (X_5), diastolic blood pressure (X_6), blood sugar levels (X_7) and body mass index (X_8). The computational process of stroke patients are summarized in Table 4.

If a significant level is taken $\alpha = 10\%$, it can be seen that the non-influential variables are the variables X_1, X_2, and X_8 while the influential variables are the variables X_3, X_4, X_5, X_6, and X_7 on the survival time of patients with stroke. Thus, the initial/complete model of Cox proportional hazard regression is

$$h(t \mid \mathbf{X}) = h_0(t)\exp(-0.135151X_1 + 0.056161X_2 + 1.922832X_3 - 1.379698X_4 + 0.020028X_5 - 0.021741X_6 + 0.001804X_7 + 0.295581X_8 \tag{9}$$

Table 4 Regression Coefficient and Significant Variable

Variable	Regression Coefficient	Standard Error	z	Pr(>|z|)	Confidence Interval 95%	
					Lower	Upper
X_1	-0.135151	0.325935	-0.415	0.67839	0.46118	1.6548
X_2	0.056161	0.301400	0.186	0.85218	0.58592	1.9096
X_3	1.922832	0.333843	5.760	8.43e-09	3.55556	13.1596
X_4	-1.379698	0.519906	-2.654	0.00796	0.09084	0.6972
X_5	0.020028	0.008333	2.403	0.01624	1.00370	1.0370
X_6	-0.021741	0.011751	-1.850	0.06429	0.95622	1.0013
X_7	0.001804	0.001066	1.692	0.09067	0.99971	1.0039
X_8	0.295581	0.327960	0.901	0.36744	0.70666	2.5558

Testing Hazard Proportional Assumptions

After knowing the initial model, it was followed by testing the proportional hazard assumption. The assumption testing aims to determine whether the predictors affect the survival time of stroke patients. The assumption testing was conducted using residual Schoenfeld.

Hypothesis:

$H_0 : \rho = 0$ (assumption of proportional hazard is met)
$H_1 : \rho \neq 0$ (assumption of proportional hazard is not met)

Given the significance level is $\alpha = 0,1$ the obtained test statistic is in the form of a correlation test between the residual Schoenfeld value and the survival time presented in Table 5.

Tabel 5 Hazard Correlation Coefficient and Hypothesis Decision

Variable	r_h	Chisq	*p*-value	Decision
X_1	0.13673	1.04278	0.307	H_0 is accepted
X_2	0.01486	0.00984	0.921	H_0 is accepted
X_3	0.02611	0.03939	0.843	H_0 is accepted
X_4	0.02801	0.04166	0.838	H_0 is accepted
X_5	-0.00811	0.00398	0.950	H_0 is accepted
X_6	0.08893	0.42538	0.514	H_0 is accepted
X_7	-0.07748	0.12895	0.720	H_0 is accepted
X_8	-0.15832	1.21781	0.270	H_0 is accepted

The Product Moment correlation coefficient values for the two-way test, with the sample size of 288 lead to the $r_{286;0.05} = 0.195$. Thus, it can be concluded that there is no correlation between the Schoenfeld residual value with survival time. This means that all variables thought to influence the survival time of stroke patients are to meet the proportional hazard assumption.

Parameter Testing

After the initial Cox proportional hazard model and the proportional hazard assumption are met, the parameters of the model are tested simultaneously and partially using the Likelihood Ratio test and the Wald test.

a. Testing the Likelihood Ratio

Hypothesis:

H_0 : $\beta_1 = \beta_2 = \cdots = \beta_8 = 0$

      (together the predictors have no effect on response)

H_1 : at least one predictor $\beta_j \neq 0$, where $j = 1, 2, \ldots, 8$

      (at least one predictor affects the response)

Suppose the significance level is $\alpha = 0,1$ the test statistics is

$G = -2\,[\ln L(0) - \ln L(\hat{\beta})] = = -2\,[-247.5963 - (-222.8770)] = 49.4386.$

Because the table statistics is $\chi^2_{8;0.1} = 13.326$ and the value of G is $= 49.4386$, H_0 is rejected. Thus, it can be concluded that there is at least one predictor variable from the model that significantly influences the survival time of stroke patients.

b. Wald Testing

To find out a predictor variable that significantly influences the survival time of stroke patients, a partial test was performed using the Wald test.

Hypothesis:

H_0 : $\beta_j = 0$ (the j-th predictor has no effect on survival time)

H_1 : $\beta_j \neq 0$, where $j = 1, 2, \ldots, 8$ (the j-th predictor affects the survival time)

Given the significance level of $\alpha = 0.1$ the test statistical equivalents are tabulated in Table 6.

Tabel 6 Wald Test Result

Variable	Regression Coefficient	Standard Error	p-value	Decision
X_1	-0.135151	0.325935	0.67839	H_0 is accepted
X_2	0.056161	0.301400	0.85218	H_0 is accepted
X_3	1.922832	0.333843	8.43e-09	H_0 is rejected
X_4	-1.379698	0.519906	0.00796	H_0 is rejected
X_5	0.020028	0.008333	0.01624	H_0 is rejected
X_6	-0.021741	0.011751	0.06429	H_0 is rejected
X_7	0.001804	0.001066	0.09067	H_0 is rejected
X_8	0.295581	0.327960	0.36744	H_0 is accepted

The value of $\chi^2_{(1;0.05)} = 3.841$, then based on Table 6, it can be decided that the Cox proportional hazard model with the Breslow method with the variables of type of stroke type (X_3), hypertension record (X_4), systolic blood pressure (X_5), and diastolic blood pressure (X_6) and blood sugar levels (X_7) partially is significant towards the model.

Given the insignificant effects of variables to the model, namely the variables of sex (X_1), age (X_2) and body mass index (X_8), the insignificant variables were excluded from the model using backward selection. The exclusion of insignificant variables was conducted one by one based on the order of the largest value of the p-value. There were three excluded independent variables. The backward selection procedure would stop if all the predictors in the model were significant. Then, the Likelihood Ratio test and the

Wald test were performed for each model. The backward selection resulted in regression models of all significant independent variables. The discussion is presented below.

4.3. Cox Proportional Hazard Regression Model and the Best Survivorship Function

Testing the Best Regression Model and Function

The backward selection process led the results tabulated in Table 7, which is the best builder of the Cox proportional hazard regression model.

Tabel 7 Regression Coefficient and The Best Significant Variable

Variable	Regression Coefficient	Standard Error	z	Pr(>|z|)	Confidence Interval 95%	
					Lower	Upper
X_3	1.930458	0.332333	5.809	6.29e-09	3.59339	13.2212
X_4	-1.306462	0.508903	-2.567	0.0103	0.09987	0.7342
X_5	0.019780	0.007953	2.487	0.0129	1.00420	1.0360
X_6	-0.022032	0.010907	-2.020	0.0434	0.95752	0.9993
X_7	0.001928	0.001015	1.898	0.0577	0.99994	1.0039

The next step is to test the parameters simultaneously using the likelihood ratio test and partially using the Wald test.

a. Testing the Likelihood Ratio

Hypothesis:

H_0 : $\beta_3 = \cdots = \beta_6 = \beta_7 = 0$
(together, the predictors have no impact on response)

H_1 : at least one predictor $\beta_j \neq 0$, where $j = 3, 4, \ldots, 7$
(at least one predictor affects response)

Given the significance level of $\alpha = 0,1$ the computational process leads to the testing statistics of

$$G = -2 [\ln L(0) - \ln L(\hat{\beta})] = -2 [-247.5963 - (-223.3450)] = 48.5026$$

Thus, the chi-square table obtained a value of $\chi^2_{5;0.1} = 9.236$, and the value of G = 48.5026, so that H_0 is rejected. This means that there is at least one predictor of the model that significantly influences the survival time of stroke patients.

b. Wald Testing

Hypothesis:

H_0 : j-predictor has no effect on survival time
H_1 : j-predictor has no effect on survival time , $j = 3, 4, \ldots, 7$

Given the significance level of $\alpha = 10\%$ the test statistics lead to the result presented in the following Table 8.

Table 8 The Best Wald Model Parameter Test Results

Variables	Regression Coefficient	Standard Deviation	p-value	Decision
X_3	1.930458	0.332333	6.29e-09	H_0 is rejected
X_4	-1.064620	0.508903	0.0103	H_0 is rejected
X_5	0.019780	0.007953	0.0129	H_0 is rejected
X_6	-0.022032	0.010907	0.0434	H_0 is rejected
X_7	0.001928	0.001015	0.0577	H_0 is rejected

Because all p-values are smaller than the significance level, H_0 is rejected. This means that variable type of stroke (X_3), history of hypertension (X_4), systolic blood pressure (X_5), and diastolic blood pressure (X_6) and blood sugar levels when (X_7) are partially significant on the model. Because all variables in the model are significant, the best Cox proportional hazard regression model is resulted from the Breslow method, as given below:

$$h(t \mid \mathbf{X}) = h_0(t) \exp \left(\begin{array}{l} 1.930458 X_3 - 1.306462 X_4 + 0.019780 X_5 \\ -0.022032 X_6 + 0.001928 X_7 \end{array} \right) \quad (10)$$

The above Cox proportional hazard regression leads to the best survivorship function, as given below

$$S(t \mid \mathbf{X}) = \left[S_0(t) \right]^{\exp(\mathbf{b'x})} \quad (11)$$

where

$$\exp(\mathbf{b'x}) = \exp \left(\begin{array}{l} 1.930458 X_3 - 1.306462 X_4 + 0.019780 X_5 \\ -0.022032 X_6 + 0.001928 X_7 \end{array} \right) \quad (12)$$

If the survivorship function is applied to the data of this study, we get a picture of the relationship between survival time and the chance of survival in Figure 1.
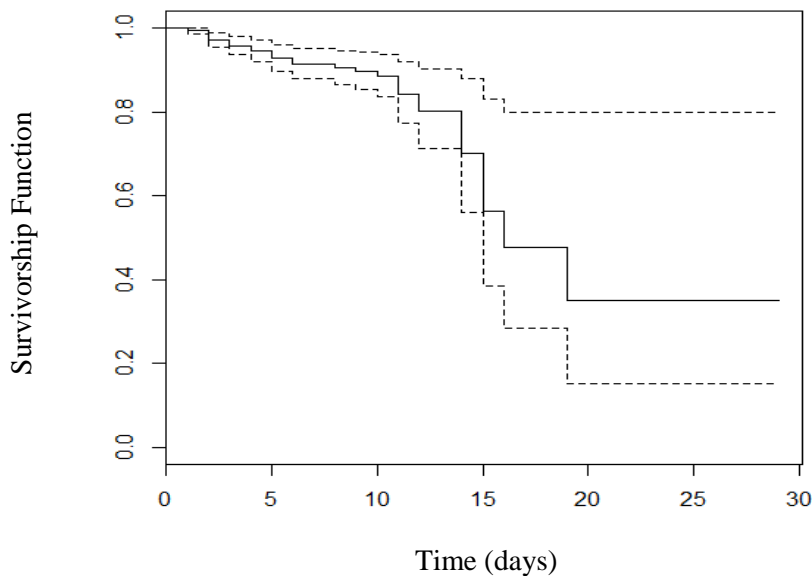

Figure 1 Survivorship Function

The function of survivorship indicates the classification of healthy opportunities for the duration of illness. The longer a person is sick, the less his chance to survive. During more than 20 days of hospitalization, it was seen that the chance to recover were low and even leading to death.

Table 9 Recovery Opportunities based on the Survivorship Function

Time (days)	Recovery Function	Description
5	0.98	High
10	0.87	High
15	0.55	Medium
20	0.35	Low
25	0.35	Low

The following section discusses the properties of the relative risk function $y = \log \dfrac{h(t \mid \mathbf{x})}{h_0(t)} = \log x$. The properties of the relative risk function are presented in Table 10.

Table 10 Characteristics of Risk Functions

X	Y	Risks
$0 < x < 1$	Less than zero	Negative Risk
1	Equals zero	No risk
$1 < x$	More than zero	Positive Risk

4.4. Model Interpretation

The best Cox proportional hazard regression model that has been obtained is

$$h(t \mid \mathbf{X}) = h_0(t) \exp\left(1.930458X_3 - 1.306462X_4 + 0.019780X_5 - 0.022032X_6 + 0.001928X_7\right)$$

In this case, it contains the value of $\exp(b_j), j = 3, 4, 5, 6, 7$ that indicates the value of the failure ratio of each predictor. The failure ratio is calculated only for variables that affect the survival time of stroke patients. Based on the best model, the failure ratio calculated from the model is presented in Table 11.

The followings are interpretations of the obtained model.
1. Variables affecting the risk factors of stroke for a person are stroke type, hypertension record, systolic blood pressure and diastolic blood pressure, and blood sugar levels.
2. The risk of someone to have an ischemic stroke is 5.893 times higher than someone having a hemorrhagic stroke. Patients affected by hemorrhagic stroke have a 6.893 times chance of survival than patients affected by ischemic stroke.
3. The risk of stroke patients who have no record of hypertension failure to survive is lower 0.729 times than that of stroke patients who have a record of hypertension.
4. An increase in someone's systolic blood pressure results in risk of stroke that is 0.02 times higher than someone with normal systolic blood pressure.
5. A decrease in someone's diastolic blood pressure results in the risk of stroke that is 2.2% smaller than someone with normal diastolic blood pressure.
6. An increase in blood sugar levels in a person results in a risk factor of stroke that is 0.2% higher than someone with a normal blood sugar level.

5. CONCLUSION

Based on the study of Cox proportional hazard regression using the Breslow method on the survival data of stroke patients at the Tugurejo Regency Hospital in Semarang it was concluded that Risk Factors of stroke are type of stroke, hypertension record, systolic blood pressure, and diastolic blood pressure. The best Cox proportional hazard regression model is

$$h(t \mid \mathbf{X}) = h_0(t) \exp \left(\begin{array}{l} 1.930458 X_3 - 1.306462 X_4 + 0.19780 X_5 \\ -0.022032 X_6 + 0.001928 X_7 \end{array} \right)$$

The best survivorship function is $S(t \mid \mathbf{X}) = \left[S_0(t) \right]^{\exp(\mathbf{b'x})}$, where

$$\exp(\mathbf{b'x}) = \exp \left(\begin{array}{l} 1.930458 X_3 - 1.306462 X_4 + 0.019780 X_5 \\ -0.022032 X_6 + 0.001928 X_7 \end{array} \right)$$

A person's risk factor for hemorrhagic strokes is lower than that of an ischemic stroke, and his survival is shorter. Someone who has a record of hypertension has a higher risk of stroke than those without. An increase in someone's systolic blood pressure results in a higher failure of life. On this account, a decrease in someone's diastolic blood pressure, the higher the survival of stroke patients.

The longer the period of sickness, the lower the chance of survival. During more than 20 days of hospitalization, it appears that the chances of recovery are low and even leading to death.

REFERENCES

Allison, P. D. 2010. *Survival Analysis Using SAS: A Practical Guide,* Second Edition. North Carolina: SAS Institute Inc.

Bain, L. J and Engelhardt, M. 1992. *Introduction to Probability and Mathematical Statistics*. California: Duxbury press.

Bhandare, S. K. and Jain, Y. K. 2014. Min Max Normalization Based Data Pertubation Method for Privacy Protection. *International Journal of Computer & Communication Technology*, Vol. 3, No. 4, pp. 45–50.

Collett, D. 2015. *Modelling Survival Data in Medical Research,* Third Edition. CRC Press.

Cox, D. R. and Oaks, D. 1998. *Analysis of Survival Data.* United State of America: Chapman and Hall, Inc.

Dewi, I. A. P. R., Suciptawati, N. L. P, and Tastrawati, N. K. T. 2016. Aplikasi Regresi Cox Proportional Hazard pada Sitasan Pasien Diabetes Melitus, *E-Jurnal Matematika,* Vol. 7, No. 3, pp. 278–285.

Faruk, A. 2017. Aplikasi Regresi Cox pada Selang Kelahiran Anak Pertama di Provinsi Sumatera Selatan. *Jurnal Matematika,* Vol. 7, No. 1, pp. 19–29.

Fox, J. and Weisberg, S. 2018. *Cox Proportional Hazard Regression for Survival Data in R,* Third Edition. CA: Thousand Oaks.

Hosmer, D. W., Lemeshow, S., and May, S. 2008. *Applied Survival Analysis: Regression Modelling of Time to Event Data,* Second Edition. New Jersey: John Wiley & Sons, Inc.

Jong, P. D. and Heller, G. Z. 2008. *Generalized Linear Models for Insurance Data,* United Kingdom: Cambridge University Press.

Kementerian Kesehatan RI. 2014. *Info Datin Pusat dan Informasi Kementerian Kesehatan RI, Situasi Kesehatan Jantung.* Jakarta.

Kleinbaum, D. G. and Klein, M. 2005. *Survival Analysis A Self-Learning Text,* Second Edition. London: Springer.

Lee, E. T. and Wang, J. W. 2013. *Statsitical Methods for Survival Data Analysis*, Fourth Edition, Canada: JohnWiley & Sons, Inc.

Lingga, L. 2013. *All About Stroke Hidup Sebelum dan Pasca Stroke*. Jakarta: PT Elex Media Komputindo.

Marisa, Yozza, H. and Mayastri. 2016. Model Regresi Cox *Proportional Hazard* pada Laju Tamat Mahasiswa Jurusan Matematika Universitas Andalas, *Jurnal Matematika UNAND,* Vol. VI, No. 1, pp. 33–46.

Nurfain and Purnami S. W. 2017. Analisis Regresi Cox Extended pada Pasien Kusta di Kecamatan Brondong Kabupaten Lamongan. *Jurnal Sains dan Seni ITS*, Vol. 6, No. 1, pp. 94–100.

Sudana, I. G. A., Suciptawati, N. L. P., and Harini, L. P. I. 2013. Penerapan Regresi Cox Proportional Hazard untuk Menduga faktor-faktor yang Mempengaruhi Lama Mencari Kerja. *E-Jurnal Matematika,* Vol. 2, No. 3, pp. 7–10.

Xin, X., Harrocks, J., and Darlington, G. A. 2017. Weighting Methods for Ties between Event Times and Covariate Change Times, *Journal Communication in Statistics-Simulation and Computation,* Vol. 47, Issue 1.