Survival Analysis of Un-identical Recurrence Using Conditional I, Conditional II, and Marginal Method

IOP Conf. Series: Materials Science and Engineering **546** (2019) 052058 doi:10.1088/1757-899X/546/5/052058

Survival Analysis of Un-identical Recurrence Using Conditional I, Conditional II, and Marginal Method

Dita Ramadayanti Kusthika Putri[1], Umu Sa'adah[1], Achmad Efendi[1*]

[1)] Department of Statistics, Faculty of Sciences, Universitas Brawijaya, Malang

*Corresponding author: a_efendi@ub.ac.id

Abstract. Survival analysis can be used to examine events that may occur more than once in one individual or is called a recurring event. This event can be seen in cases of stroke recurrence. The occurred recurrence shows a more severe recurrence stage than before (recurring events are not identical). The Stratified Cox model with Conditional I, Conditional II, and Marginal methods can be used to analyse recurring events, especially in recurrent events that are not identical. The used data are secondary data from the medical record archives of stroke patients at Chasbullah Abdulmadjid Hospital Bekasi City from March 1, 2016 to April 30, 2016. Based on the Wald test and the backward elimination steps, we found two explanatory variables that influence the recurrence rate of stroke which were hypertension and hypercholesterolemia. The model from the three methods is a model with a combination of variables of hypertension and hypercholesterolemia. Patients suffering from hypertension or hypercholesterolemia have the potential to experience a recurrence of stroke compared to patients who do not have hypertension or hypercholesterolemia. Based on the AIC value of each method, the best method for recurring data on stroke patients is the Conditional I method.

Keywords: Conditional I, Conditional II, Marginal, Recurrent Event.

1. Introduction

The time period from the beginning of the study until an event occurs is called survival time, which can be in the form of years, months, weeks, days, hours, minutes or even seconds. The observed events can be death, relapse or disease recovery, damage or re-operation of a machine, and events in individuals who stop or get a job [1]. The difference in survival analysis with other statistical analysis lies in the concept of censorship. The censored data are obtained from individuals who did not experience the incident until the research ended. In addition, the censored data can also come from the loss of the individuals studied when the research is still ongoing.

Cox regression model is one of regression analysis that can be used to analyze survival data. It is also known as the Cox Proportional Hazard regression model because in this model, it is assumed that all explanatory variables fulfill proportional hazard assumptions. Proportional hazard assumptions are met if the risk ratio of an event is constant or individual risk function is proportional to the risk function of another individual [2]. Survival analysis can be used to analyze recurrent events, both identical and not identical repetitive events. In the medical field research, recurrent events are categorized as identical and not identical, for example based on the severity suffered by the patient. If the incidence rate of each patient's disease is the same, then the incident is said to be an identical

recurring event, on the other hand if a recurring event indicates a more severe stage than before then it is said that the recurring event is not identical. The Stratified Cox model can be used to analyze recurring events, especially for recurrent events that are not identical [1]. The Stratified Cox model is divided into three methods, namely Conditional I, Conditional II, and Marginal one.

Non-identical recurring events can be seen in cases of stroke recurrence. In stroke, the occurred recurrence shows a more severe recurrence stage than before. Stroke is the leading cause of death after heart disease and cancer. Stroke or Cerebro Vascular Accident (CVA) is loss of brain function caused by the cessation of blood supply to parts of the brain [3]. This study will apply the three methods above (Conditional I, Conditional II, and Marginal one) to recurrent data on stroke recurrence with gender, family history, hypertension status, hypercholesterolemia status, obesity status, diabetes mellitus status, cardiac abnormalities status, and therapy (using *mannitol* or not) given in patient as an explanatory variable. Besides that we want to find out what factors have a significant effect on the rate of non-identical recurrence of stroke, and want to know the best model obtained.

2. Survival Analysis

One of the purposes of survival analysis is to model the relationship between risk factors and the time of occurrence of a particular event. Because in general in observing the symptoms of an illness it takes a long time until the symptoms of the disease recur, then survival analysis is widely used in the health field [7].

2.1. Recurrent Events

Survival analysis that is commonly used only considers single events experienced by individuals or each individual only experiences one event, but does not rule out the possibility that researchers want to examine events that may occur more than once in one individual [5]. When the studied individual experiences an event several times or repeatedly during the study period, then the situation can be said to be a recurrent event. Recurrent events consist of identical and not identical recurring events [1]. If each repetition does not cause a different effect, it can be said to be an identical recurring event. For example in the case of an asthma attack, both the first, second and subsequent attacks are considered the same and there are no different effects. While recurring events can be said to be not identical if the effects of each recurrence have a different effect on the individual. An example is in stroke cases, where each recurrence having a different effect. The used analysis in the recurring event especially for non-identical recurring events is the Stratified Cox model [1]. There are three methods in the Stratified Cox model, namely Conditional I, Conditional II, and Marginal one.

2.2. Stratified Cox

In a recurring event, the event variable is very dependent on time, so that the variable does not meet the proportional hazard assumption, so one solution to overcome this is to use the Stratified Cox model. Modeling on Stratified Cox is done by forming strata on the response variable based on the number of repetitions of events during the study [4]. Suppose that a Stratified Cox model is created from p explanatory variables with u variables that fulfill proportional hazard assumptions $(X_1, X_2, \dots, X_u)$ and v variables that do not meet proportional hazard assumptions $(Z_1, Z_2, \dots, Z_v)$. The first thing to do is to define the variable of Z^*, and then form the strata of each variable of Z^*, and then combinations are formed. The number of combinations is expressed as v^*. The general model of Stratified Cox regression is

$$h_{ij}(t, \mathbf{X}) = h_{0j}(t) exp\left[\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p\right] \quad (2.1)$$

where:
i: individual at the stratum-j; j: 1, …, v^*
h_{0j} : risk base line function for each stratum

The variables included in the model above are only variables that fulfill proportional hazard assumptions $(X_1, X_2, \dots, X_u)$. The values of the coefficients are the same for each $\beta_1, \beta_2, \dots, \beta_u$. While

the risk baseline function for each stratum (h_{0j}) is likely to be different. Equation (2.1) can be solved by a numerical method in the form of Newton-Raphson iterations to maximize the partial likelihood function [6].

2.2.1. Conditional Models

Based on differences in the arrangement of data on the time of events in the life table, the Conditional method is divided into two, namely the Conditional I and Conditional II methods. The focus of both methods is the same, finding out the survival time between two events. In contrast to structuring data on survival analysis in general, structuring data on the events in the Conditional I method uses the actual time series since the individual was first examined. Time sequences are recorded continuously without experiencing the initial conditioning each time a recurrence occurs [1].

Table 1. Example of structuring recurrent event data using the Conditional I method

Individual (i)	Event (j)	Status	Time	
			Start (t_0)	Event (t_j)
1	1	1	0	2
1	2	1	2	10
2	1	1	0	16
2	2	1	16	22
2	3	0	22	25

Meanwhile, the arrangement of data on the time of event for the Conditional II method undergoes conditioning every recurrent event. This implies that at the beginning of the study will always start with 0 (zero) until the recurrent events and the time of the incident is the length of time until the next event [1].

Table 2. Example of structuring recurrent event data using the Conditional II method

Individual (i)	Event (j)	Status	Time	
			Start (t_0)	Event (t_j)
1	1	1	0	2
1	2	1	0	8
2	1	1	0	16
2	2	1	0	6
2	3	0	0	3

2.2.2. Marginal Models

Structuring the time event data in the Marginal method does not include the time of the history of previous events [1]. For the analysis of repetitive data that is not identical, the Marginal method focuses on total of survival time, starting from the first time the research was conducted until the end

of the study. In this method structuring the time data of events is carried out without the occurrence of repetition of events, so that what is recorded in the life table is only the time of occurrence of the event.

Table 3. Example of structuring recurrent event data using the Marginal method

Individual (i)	Event (j)	Status	Event time (t_j)
1	1	1	2
1	2	1	10
1	3	1	10
2	1	1	16
2	2	1	22
2	3	0	25

The procedures used in selecting the best model are forward selection (backward selection), backward selection (backward selection), and stepwise selection [8].

3. Data and Methodology

The data used in this research is from RSUD (Region Hospital) of Chasbullah Abdulmadjid. The observation was conducted on 1 March – 30 April 2016. The response variable is survival time (days) while the predictor variables are status (recurrent; going home/moving to other hospital/dead), gender (man; woman), family stroke history (yes, no), hypertension status (yes, no), hypercholesterolemia (yes, no), obesity (yes, no), diabetes mellitus (yes, no), heart abnormalities (yes, no), and treatment (*Mannitol*; No).

Steps in descripting and analyzing data are: presenting recurrent and un-identical data, censoring identification, estimating the estimate of β using *Stratified Cox* model, Wald significance test, model selection using *backward selection* with AIC values, forming strata in the method of Conditional I, Conditional II, and Marginal (Table 1, 2, and 3), Stratified Cox regression modeling with the three methods, hazard ratio estimation particularly for the significant predictors, and the last step is the selection of the best model.

Stroke prone conditions in Indonesia are increasing because of combined physical changes, environment, habits, lifestyle, and types of diseases that develop suddenly causing cumulative risk of stroke-affected people in Indonesia increasing from 10 to 15 times or certainly far greater compared to previous times [10]. A person suffers from a stroke because of having risk factors for stroke [9]. Stroke risk factors are divided into two, namely risk factors that cannot be changed and the one that can be changed.

4. Data Analysis and Result

This section discusses the description of the results of the descriptive analysis and also the results of data modeling. In addition, a discussion of the results of the obtained analysis is also found in this section.

4.1. Descriptive Analysis

The data came from repeated recurrence of stroke patients in Chasbullah AM Hospital in Bekasi City from March 1, 2016 to April 30, 2016. There were 122 recurrences during the study period with 53 patients having the first recurrence, 45 patients experiencing a second recurrence, and 24 patients

experience a third recurrence. Of 71 patients (40 men and 31 women), 76% had a history of stroke in the family, 63% were positive for hypertension, 56% positive for hypercholesterolemia, 48% positive for obesity, 72% positive for diabetes mellitus, and 30% positive for cardiac abnormalities .

Besides being caused by internal factors (gender, family history, hypertension status, hypercholesterolemia status, obesity status, diabetes mellitus status, and cardiac abnormalities), recurrence of stroke patients is also influenced by external factors such as the type of drug or therapy given for a patient. Therefore, it is considered the type of *Mannitol* therapy as one of the factors that affect recurrence of stroke. Based on the medical record, it was found that 65% of 71 patients received *Mannitol* therapy and the rest received other therapies.

4.2. Data Analysis

Table 4. Parameter Estimates of β

Explanatory var.	$\hat{\beta}_p$	*Robust Std. Err.*	*Wald Stat.*	*p-value*
Gender	-0.009	0.193	-0.05	0.963
Family history	0.433	0.240	1.81	0.071
Hypertension	0.420	0.151	2.79	0.005
Hypercholesterolemia	0.509	0.154	3.31	0.001
Obesity	-0.018	0.203	-0.09	0.931
Diabetes mellitus	0.190	0.253	0.75	0.451
Heart Disorders	0.174	0.173	1.00	0.316
Therapy	0.354	0.282	1.26	0.209

Based on the estimation of the parameter in Table 4, information is obtained that the statistically significant explanatory variable (has a p-value less than = 0.05) are the variable of hypertension and hypercholesterolemia. Hence, on this partial test result, the variables that influence in the model are hypertension and hypercholesterolemia.

4.2.1. Stratified Cox Regression Model with Conditional I Method. The estimated value of parameters for hypertension and hypercholesterolemia in the Stratified Cox regression model with the Conditional I method are as following:

$$h_{ij}(t) = h_{0j}(t)\exp(0.516HT_{ij} + 0.468HC_{ij})$$

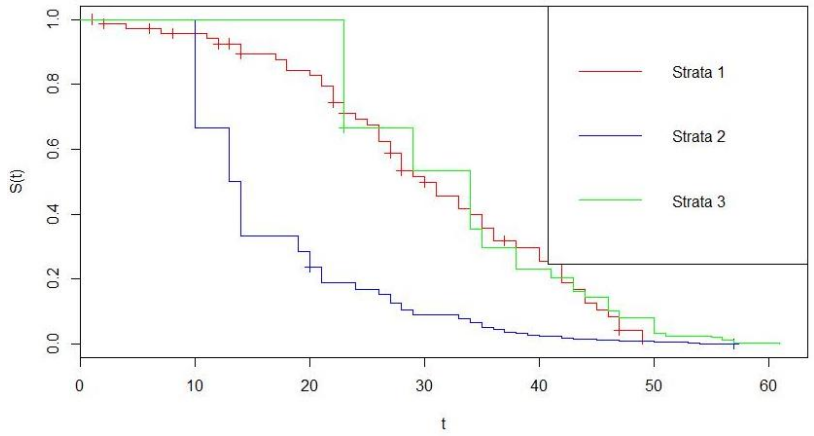

Figure 1. Kaplan-Meier Curve for Conditional I Method

4.2.2. Stratified Cox Regression Model with Conditional II Method. The estimate value of the parameter for variable of hypertension and hypercholesterolemia in the Stratified Cox regression model with the Conditional II method is as follows:

$$h_{ij}(t) = h_{0j}(t)\exp(0.491HT_{ij} + 0.634HC_{ij})$$

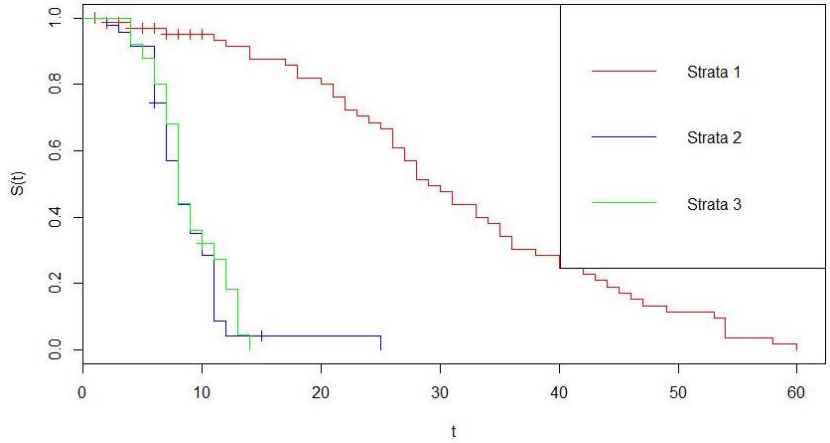

Figure 2. Kaplan-Meier Curve for Conditional II Method

4.2.3. Stratified Cox Regression Model with Marginal Method. The predictive value of the β parameter variable hypertension and hypercholesterolemia in the Stratified Cox regression model with the Marginal method is as follows:

$$h_{ij}(t) = h_{0j}(t)\exp(0.627HT_{ij} + 0.548HC_{ij})$$

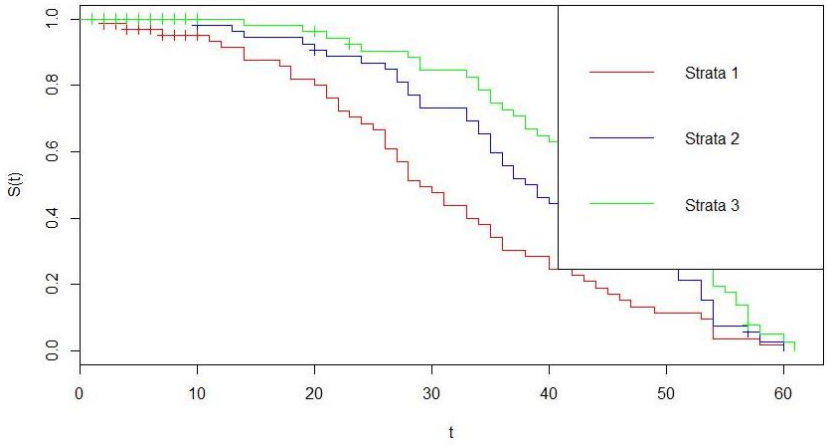

Figure 3. Kaplan-Meier Curve for Marginal Method

4.2.4. Comparison of the results from three methods. Based on the comparison of the results of the three methods in Table 5, information was obtained that from the three methods showed almost the same results. In order to find the best method, we can use the AIC value from each method. The method with the smallest AIC value is the best method. The AIC values obtained from the three methods are 587.949, 692.450, and 926.946 for Conditional I, Conditional II, and Marginal one respectively. We found that the best method for repeated recurrence of stroke patients in Chasbullah Abdulmadjid. Hospital, Bekasi City, is the Conditional I method because it has the smallest AIC value.

Table 5. Comparison of the three methods

Variable		*Conditional* I	*Conditional* II	*Marginal*
Hypertension	Coeff.	0.516	0.491	0.627
	P-value	0.002	0.007	0.008
	Hazard Ratio	1.675	1.634	1.872
Hypercholesterolemia	Coeff.	0.468	0.634	0.548
	P-value	0.001	0.001	0.049
	Hazard Ratio	1.597	1.885	1.730

5. Conclusion

The Stratified Cox model formed on non-identical recurrent data of stroke in Chasbullah Abdulmadjid Hospital, Bekasi City, was with the Conditional I method. Factors that influence the rate of non-identical recurrent stroke were hypertension and hypercholesterolemia. Patients suffering from hypertension or hypercholesterolemia have the potential to experience a recurrence of stroke compared to patients who do not have hypertension or hypercholesterolemia.

References

[1] Kleinbaum, D. G. and Klein, M. 2012. *Survival Analysis: A-Self Learning Text. Third Edition*. Springer, New York
[2] Guo, S. 2010. *Survival Analysis*. Oxford University Press, Inc, New York.

[3] Brunner and Suddarth. 2002. *Buku Ajar Keperawatan Medikal Bedah*. Translated to Indonesian by Agung Waluyo *et al*. EGC, Jakarta.
[4] Ata, N. dan Sozer, M. T. 2007. Cox Regression Model with Nonproportional Hazards Applied to Lung Cancer Survival Data. *Hacettepe Journal of Mathematics and Statistics, Volume 36 (2)*. Hacettepe University, Turkey.
[5] Hosmer, D. W., Lemeshow, S., dan May, S. 2008. *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*. *Second Edition*. John Wiley and Sons, Inc, New Jersey.
[6] Collet, D. 2003. *Modelling Survival Data in Medical Research*. *Second Edition*. Chapman and Hall, London.
[7] Colton, T. 1974. *Statistics in Medicine*. Little Brown and Company, Boston.
[8] Lee, E. T. dan Wang, J. W. 2003. *Statistical Methods for Survival Data Analysis. Third Edition.* John Wiley and Sons, Inc, New Jersey.
[9] Pinzon, R. dan Asanti, L. 2010. *Awas Stroke! Pengertian, Gejala, Tindakan, Perawatan, dan Pencegahan.* CV. Andi Offset, Yogyakarta.
[10] Yayasan Stroke Indonesia. 2006. *Hidup Sehat dan Cegah Stroke Edisi Oktober*. http://www.yastroki.or.id/read.php?id=307. Accessed on 24 September 2018.