The Application of Extended Cox Proportional Hazard Method for Estimating Survival Time of Breast Cancer

The Application of Extended Cox Proportional Hazard Method for Estimating Survival Time of Breast Cancer

Hartina Husain [1], Sri Astuti Thamrin[1], Sulaiha Tahir [1], Ahmad Mukhlisin [1], Mirna Apriani M [2]

[1,]Statistics, Faculty of Mathematics and Natural Sciences, Hasanuddin University
[2]Department of Epidemiology, Faculty of Public Health, Hasanuddin University

E-mail: hartina.husainn@gmail.com

Abstract. Breast cancer is one type of cancer that is the leading cause of death worldwide. This study aims to model the factors that affect the survival time and rate of cure of breast cancer patients. The extended cox model, which is a modification of the proportional hazard cox model in which the proportional hazard assumptions are not met, is used in this study. The maximum likelihood estimation approach is used to estimate the parameters of the model. This method is then applied to medical record data of breast cancer patient in 2011-2016, which is taken from Hasanuddin University Education Hospital. The results obtained indicate that the factors that affect the survival time of breast cancer patients are malignancy and leukocyte levels.

1. Introduction

Preventing the death rate from illness is the main goal of treatment. Prevention is done by considering the external and internal factors that affect the pattern of human life. Non-infectious diseases are still the main concern of one cancer. Cancer is the leading cause of death and disability worldwide, affecting more than 14 million people each year [1]. One of cancer types that is breast cancer. According to the [2] each year there are 1.7 million cases of breast cancer and 552,000 died. According to data Globocan (IARC) [3] breast cancer has a percentage of 43.3% with a mortality rate of 12.9%. Unlike lung cancer with a percentage of 34.3% with a 30.0% death rate. Breast cancer (Carsinoma mammae) occurs due to abnormal growth of breast cells (Sutjipto, 2006). In South Sulawesi, breast cancer cases are ranked first among many cancer diseases suffered by women. Based on data from medical record of Dr. Wahidin Sudirohusodo Makassar, the numbers of patients treated from 2010 to 2013 are 132, 360 and 573 cases of breast cancer, respectively.

Some of the factors that are suspected to be the trigger of breast cancer are affected by five behavioral and dietary risks, namely high growth mass index, less consumption of fruits and vegetables, less physical activity, cigarette use, and excessive alcohol consumption. These five factors have a rate of causing 30% cancer death rate. This unhealthy behavior that causes cancer risk becomes high. The existence of these factors indicates that there is a correlation between breast cancer disease and survival of patients [4]. Therefore, it is needed to see the relationship between survival time and the factors that its influence.

The survival analysis is a series of useful statistical procedures to analyze the timing of events. Explains that the survival analysis is a statistical analysis that model the length of time until an event occurs [5]. The survival analysis is used with emphasis on individual survival time or duration of individual survival time until an event occurs (time to event data). Such events may be the development of a disease, the response to treatment, the recurrence of the disease, death or even other occurrences

determined by the researcher. In the hazard ratio model, the resulting value is no longer time-dependent so that the ratio of the two objects remains at all times proportional hazard. The proportional cox regression model emphasizes the fulfillment of the proportional hazard assumption which means that the ratio between individual hazard functions of one and other individual hazard functions is constant. However, in some cases, the covariate hazard ratio is not over time or not fixed. Therefore, proportional hazard assumptions are not met. To see the effect of a covariate that does not meet the proportional hazard assumption, the covariate must be converted into a time-consuming covariate, called a folded proportional hazard. This model is referred to as an extension of the proportional cox model containing the time-bound or multiplication variable of the independent variable with the time function. Some of the time functions that can be used in the extendend model are g (t) = 0, g (t) = t and g (t) = log (t) . This model aims to determine how much influence of variables on the survival time of a breast cancer patient.

Research related to non-proportional hazard has been done by Feriana and Dwi [6]. Modeled lung cancer data using cox stratification model [6]. Furthermore, Arini and Isna [7] has done research on extended cox model-to-model patient data of kidney transplanted. This study found that patients with catheters were more rapidly infectious than those with percutaneous instances. According to the study, Iskandar and Bayu [8] examined the cox proportional hazard model on the Joint occurrence in the case of traffic accidents in the United States obtained results that age and use of seat belts increase the risk of accidents. In addition, Pahlevi et al. [9] has conducted research on cox stratified regression model on unspecified Hemorrhagic Stroke. In addition, Vitriana et al. [10] examined the procedure for establishing the extended Cox model on joint events and its application in the case of individuals quitting work. Provided by Cox extended model with significant variables, namely age variable, marital status, age bound time and marital status bound time. Based on these reviews, this paper is interested to model the breast cancer patients' survival time based on the factors influenced using the cox extended proportional hazard.

This paper outlines in section 2 a materials and method such as survival analysis, proportional hazard model, estimated cox proportional hazard model and data collect. Section 3, we illustrate the performance of our models with a breast cancer data set from the Unhas Education Hospital database. Then, section 4 concludes the paper with a summary and some possible areas of extension.

2. Materials and Methods

2.1 Study Area

The total population of Makassar City based on data from Statistics Center Bureau called Badan Pusat Statistik (BPS) in 2015 was recorded at 1,449,401 inhabitants. Based on sex ratio, female population was more than male population with sex ratio of 97.84. Which means that every 100 souls of female population there are 97 male population. South Sulawesi case of breast cancer is the first rank of cancer which suffered by many women. The number of patients treated during 2010 found 132 cases of breast cancer, in 2011 found 360 cases of breast cancer, in 2012 there was an increase to 573 cases of breast cancer and in 2013 as many as 592 cases of breast cancer.

2.2. Data Collection

Data of this research is medical record of patient of Breast Cancer at RSUP Universitas Hasanuddin Makassar City with number of patient counted 101 people. The covariates of this study were age, medication, occupation, marital status, malignancy, histology, stage, leukite, hemoglobin, and sensor indicators with 6.94% censored data and 93.06% uncensored data. Interval age of breast cancer patients is ranged in age 21-75 years with the average patient age is 43.77 years.

2.3 Survival Analysis

Survival analysis is a statistical technique used to analyze data. This technique aims to determine the results of variables that affect an event beginning to the end of the event, such as time (day, week, month or year). Forms of initial events for example the initial infected patients and the end events such as death or recovery of patients [11]. Let X denote the time of failure or death time, then T can be viewed as a nonnegative random variable. Survival function is the chance that an individual can survive till time (experience after time t) (Thamrin, 2013a). The function of Survival is defined as $S(t) = P(T>t) = \int_x^\infty f(t)dt$ and $f(t) = -\frac{dS(t)}{dt}$, where $F(x) = P(X \leq x)$ is a cumulative distribution function.

According to Miller (1998), the life function *S* (t) is a function that does not rise or monotone down with properties:

1. *S(t)* = 1 for *t* = 0, it means individual opportunities to live at the moment *t* = 0 is 1
2. *S(t)* = 0 for *t* = ∞, it means individual opportunities to live at the moment *t* = ∞ is 0

The other functions related to the function of survival is the hazard function. The hazard rate function is defined as the conditional failure rate i.e. the limit of the probability of an individual failing to persist in very short intervals of time *t* to *t* + Δ*t*, if the individual has survived until time t [4] This function is defined as $h(t) = \lim_{\Delta t \to 0} \frac{P[t \leq X < t+\Delta t | T \geq t]}{\Delta t}$ where the density function is probability

$$f(t) = h(t).S(t) \quad (1)$$

The proportional hazard model depends on the values $x_1, x_2, ..., x_p$ form *p* explanatory variables, $X_1, X_2, ..., Xp$. The values of the explanatory variables in the proportional hazard model are expressed in vector form $\boldsymbol{x}$, so that $\boldsymbol{x} = (x_1, x_2, ..., x_p)$'. Suppose $h_0(t)$ is a hazard function for an individual that all vector explanatory variables **x** has a value of zero, then the function $h_0(t)$ called baseline hazard function. The general proportional hazards model is as follows

$$h_i(t) = \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)\, h_0(t) \quad (2)$$

can also be expressed in terms of log equations is

$$\log \frac{h_i(t)}{h_0(t)} = (\beta_1 xi_1 + \beta_2 xi_2 + \cdots + \beta_p xi_p), \quad (3)$$

Cox proportional model is used to determine the effect of variables on survival of cancer patients. This model makes it possible to isolate variables that have little effect on survival. In addition, the model allows to estimate the risk or danger of death for an individual based on the prognostic variable.

The Weibull distribution is one of the continuous distributions in probability theory and statistics. Weibull distribution can be used if in a data survival hazard function increases or decreases monotonically with an increase in survival time. Cox proportional hazard model uses Weibull distribution:

$$h(t, X) = \lambda p t^{p-1}, \quad (4)$$

where $\lambda = \exp\left[\sum_{i=1}^{p} \beta_i X_i\right]$ and

$$h_o(t) = p t^{p-1}, \quad (5)$$

with the shape of Weibull hazard is a parametric model with form parameters (λ), scale parameters (p) and covariate (β_i) not known.

2.4 Extended Cox

In time dependent covariate is, $X_b(t) = X_b.g_b(t)$, where X_b is a covariate that has no time dependency. However, because it does not meet the assumption Proportional Hazard, X_b must be interacted with the time function. If all p covariate does not meet the proportional hazard assumption, then the number of p covariates must be interacted with time. The time function for covariate-b is defined as $g_b(t)$, obtained by the form of extended cox as follows:

$$h(t, X(t)) = h_o(t) \exp\left[\sum_{b=1}^{p} \beta_b X_b + \sum_{b=1}^{p} \delta_b X_b g_b(t)\right] \quad (6)$$

where available p_1 covariate that meets the proportional hazard assumptions and exists p_2 which does not meet the proportional hazard assumption with $p_1 + p_2$, then obtained the model as follows:

$$h(t, X(t)) = h_o(t) \exp\left[\sum_{a=1}^{p_1} \beta_a X_a + \sum_{b=p_1+1}^{p_2} \beta_b X_b + \sum_{b=p_1+1}^{p_2} \delta_b X_b g_b(t)\right] \quad (7)$$

with time function $g_b(t)$ which can be used is: $g_b(t) = 0, g_b(t) = t$, $g_b(t) = \ln t$

3. Results

The formation of Cox proportional hazard model was conducted to determine the relationship between survival time and the variables suspected to influence survival time. Estimation of Cox proportional hazard model parameters was done using Breslow method approach.

Table 1. Parameter Estimation Results Cox Proportional Hazard Model of Breast Cancer.

Variable	Coefficient	exp (β_j)	SE
Age	0,018915	1,019095	0,009109
Treatment	0,153104	1,165447	0,113618
Work	0,397254	1,487734	0,251960
Marital status	0,827081	2,286634	0,350766
Degree of malignancy	0,053600	1,055062	0,163627
Histology	0,096296	1,101081	0,144184
Stadium	0,122131	1,129902	0,116122
Haemoglobin levels	0,052927	1,054352	0,207227
Leukocyte levels	0,146134	1,157351	0,247730

Based on Table 1 we get the Cox proportional hazard model, that is:

$$h_i(t_i) = h_0(t)\exp(0{,}018915X_1 + 0{,}153104X_2 + 0{,}397254X_3 + 0{,}827081X_4 + 0{,}0536X_5 + 0{,}096293X_6 + 0{,}122131X_7 + 0{,}052927X_8 + 0{,}146134X_9)$$

Furthermore, based on the model formed will be tested partial likelihood ratio log obtained, ln L(0) = -353,4687 and ln L($\hat{\beta}$) = -335,8804 so that the value can be obtained G, that is 35,1766. Earned value G = 35,1766 $\geq \chi^2_{(0,05;9)}$ with value $\chi^2_{(0,05;9)}$ = 16,9190, then H_0 rejected and concluded that there is at least one influential variable in the model established.

All covariates affecting breast cancer are tested for parameters. Partial test results parameters with Wald test using Breslow method approach as follows:

Table 2. Partial test results parameters.

Variable	Coefficient	SE	Z^2_{hitung}
Age	0,020912	0,008823	5,6169
Treatment	0,1376	0,1115	1,52276
Work	0,3263	0,2173	2,256
Marital status	0,544	0,319	2,90703
Degree of malignancy	-0,08573	0,15004	0,32604
Histology	0,1959	0,1273	2,37468
Stadium	0,1578	0,1015	2,41802
Haemoglobin levels	-0,1168	0,1922	0,36966
Leukocyte levels	0,3217	0,2126	2,28917

Based on Table 2, it is found that the overall variables affect the survival time because of the value $Z^2_{hitung} \geq \chi^2_{(0,05;1)}$ with $\chi^2_{(0,05;1)}$ = 0,004. So obtained model:

$$h_i(t_i) = h_0(t)\exp(0{,}020912X_1 + 0{,}1376X_2 + 0{,}3263X_3 + 0{,}544X_4 \pm 0{,}08573X_5 + 0{,}1959X_6 + 0{,}1578X_7 \pm 0{,}1168X_8 + 0{,}3217X_9) \quad (8)$$

3.1 Testing Proportional Hazard Assumption

Testing proportional hazard assumption in this study using Schoenfeld residual as follows:

Table 3. Correlation, χ^2_{hit}, and p-value variable is free

Variable	Correlation	χ^2_{hit}	p-value
Age	0,00742	0,00395	0,95
Treatment	-0,0657	0,528	0,468
Work	0,00522	0,0028	0,958
Marital status	0,106	1,06	0,302
Degree of malignancy	**-0,197**	**4,45**	**0,035**
Histology	0,137	1,57	0,211
Stadium	-0,161	2,94	0,0863
Haemoglobin levels	-0,0504	0,252	0,616
Leukocyte levels	**0,284**	**6,44**	**0,0111**

Table 3 shows that age, medication, occupation, marriage, histology, stage, and Hemoglobin levels variables meet assumptions Proportional Hazard with p-value > α (0,05). Meanwhile, the variables Degree of malignancy and Leukocyte levels have p-value respectively, that is 0,035 and 0,0111 smaller than α (0,05) so it does not fulfill the Proportional Hazard assumption. Based on the test using Schoenfeld residual obtained the conclusion that the variable degree of malignancy and Leukocytes did not meet the assumption of proportional hazard so this study was continued to the extended proportional Hazard cox model.

3.2 Cox Extended Proportional Hazard (PH) model

Based on the assumption of proportional hazard test, it is found that the variable that does not meet the assumption, that is the variable of malignancy degree and the leukocyte level does not meet the proportional hazard assumption so that new model is needed to overcome the variable that does not meet the cox proportional hazard that is Cox Extended. Therefore, added the time function $(t) = t$ on variable Degree of malignancy and Leukocyte levels, that is as follows:

$$h_i(t_i) = h_0(t)\exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 g(t) + \beta_6 X_6 + \beta_6 + \beta_8 X_8 + \beta_9 X_9 g(t)) \quad (9)$$

Then, this model estimated parameter, with $g(t) = t$ using the Breslow method approach

Table 4. Estimation of Model parameters *Cox Extended* with $g(t) = t$ using the Breslow method approach

Variable	Coefficient	exp (β_j)	SE
Age	0,01649	1,01663	0,00929
Treatment	0,15967	1,17313	0,11561
Work	0,44844	1,56587	0,25622
Marital status	0,88150	2,41451	0,35139
Degree of malignancy	0,06355	1,06561	0,17211
Degree of malignancy g(t)	**0,27402**	**1,31524**	**0,53929**
Histology	0,05199	1,05337	0,14669
Stadium	0,14433	1,15526	0,11566
Haemoglobin levels	0,06206	1,06402	0,20781
Leukocyte levels	-0,06370	0,93829	0,27107
Leukocyte levels g(t)	**1,52563**	**4,59806**	**0,69343**

Based on Table 4 we get the Cox extended model equation with time function $g(t) = t$ as follows:

$$h_i(t_i) = h_0(t)\exp[0{,}01649X_1 + 0{,}15967X_2 + 0{,}44844X_3 + 0{,}88150X_4 + 0{,}06355X_5 + 0{,}27402X_5 g(t) + 0{,}05199X_6 + 0{,}14433X_7 + 0{,}06206X_8 - 0{,}06370X_9 + 1{,}52563X_9 g(t)].$$

From the model obtained is done partial likelihood logging, so obtained ln L($\hat{\beta}$) = $-353{,}4687$ and ln L($\hat{\delta}$) = $-333{,}3845$, so that the value can be obtained G, that is $40{,}1684$. Earned value $G = 40{,}1684 \geq \chi^2_{(0{,}05;11)}$ with value $\chi^2_{(0{,}05;11)} = 4{,}575$, then H_0 rejected and concluded that there is at least one influential variable in the model established. Then, then this model is tested partially with Wald Test, as follows

Table 5. Parameter Test Results Partially with Wald Test

Variable	**Coefficient**	**SE**	$\mathbf{Z^2_{hitung}}$
Degree of malignancy g(*t*)	-0,1248	0,8827	0,082944
Leukocyte levels g(*t*)	1,4206	4,1397	6,431296

Based on Table 5 it is found that the malignant degree variable is bound by time and affects the survival time with $\mathbb{Z}_{hit}^2$ = 0,082944 > $\chi_{(0,05;1)}^2$ with $\chi_{(0,05;1)}^2 = 0{,}004$ other than that variable Leukocyte levels Time bound and effect on survival time with $\mathbb{Z}_{hit}^2$ = 6,431296 > $\chi_{(0,05;1)}^2$ with $\chi_{(0,05;1)}^2 = 0{,}004$.

4. Conclusion

In this paper, we modeled survival time for breast cancer patients using extended cox proportional hazard. Based on the results of the data processing interpretation of the single unit, the individual has the risk of death is equal to $\exp(0{,}06355 + 0{,}27402) = 0{,}27757$, So each addition of one degree will decrease the risk of dying in breast cancer patients by 0,2. Every individual has a leukocyte level < 4500 /mm^3 has a risk of unity is as big $\exp(0{,}06370 + 1{,}5463) = 0{,}33757$. Therefore, individuals who have leukocyte levels > 4500 /mm^3 has a lower risk of death rate of 0,3.

Acknowledgments

The authors acknowledge to the Directorate of Research and Community Service, Directorate General Strengthening Research and Development Ministry of Research, Technology and Higher Education with Student Creativity Program (PKM) with the grant number 11153/UN4.1.3/PM.05/2017.

Appendix

```
cox<-
coxph(Surv(Dirawat,Status)~Usia+Pengobatan+Pekerjaan+Pernikahan+Keganasan+Histologi+Stadium+
Hb+wbc, method="breslow",data=DataFIX)
summary(cox)
Call:
coxph(formula = Surv(Dirawat, Status) ~ Usia + Pengobatan + Pekerjaan +
   Pernikahan + Keganasan + Histologi + Stadium + Hb + wbc,
   data = DataFIX1, method = "breslow")
```

For full ODC files contact the author directly.

References

[1] World Health Organisazation. 2013a Cancer Control: A Global Snaptshot in 2015 http://www.who.int/cancer/cancer-snapshot-2015/en/ Accesed on 9 November 2016
[2] World Health Organisazation 2013b Breast Cancer Awareness Month: increased awareness, equitable access to early diagnosis and timely, effective, and affordable treatment needed globally http://www.who.int/gho/publications/world_ health_statistics/2013/en/ Accesed on 9 November 2016
[3] Kementrian Kesehatan Republik Indonesia 2015 *Infodatin: Pusat Data dan Informasi Kementrian Kesehatan RI*. http://www.depkes.go.id/resources/ download/pusdatin/infodatin/infodatin-kanker.pdf Accesed on 12 Oktober 2016
[4] Basu S and Tiwari R C 2010 Breast cancer survival, competing risk and mixture cure model: A Bayesian analysis *Journal of the Royal Statistical Society* **173**(2) 307–329
[5] Kleinbaum D C and Klein M 2005 *Survival Analysis* (New York: Springer)
[6] Feriana and Dwi A 2011 *Model Cox Stratifikasi* (Jakarta: Universitas Indonesia)

[7] Arini and Isna N 2011 *Extended Cox Model untuk Time-Independent Covariate yang Tidak Memenuhi Asumsi Proportional Hazard pada Model Cox Proportional Hazard* (Jakata: Universitas Indonesia)
[8] Iskandar and Bayu M 2015 *Model Cox Proportional hazard Pada Kejadian Bersama* (Jojakarta: Universitas Negeri Yogyakarta)
[9] Pahlevi, Mohammad R, Mustafid, Triastuti W 2016 Model Regresi Cox Stratified pada Data Ketahanan *Junal Gaussian,* **5**(3) 455-464
[10] Vitriana, Anita N, Rosita K 2016 Model Cox Extended dengan g(t) = t untuk Mengatasi Nonproportional Hazard pada Kejadian Bersama Jogjakarta
[11] Kleinbaum D C and Klein M 2011 Survival Analysis (New York: Springer)