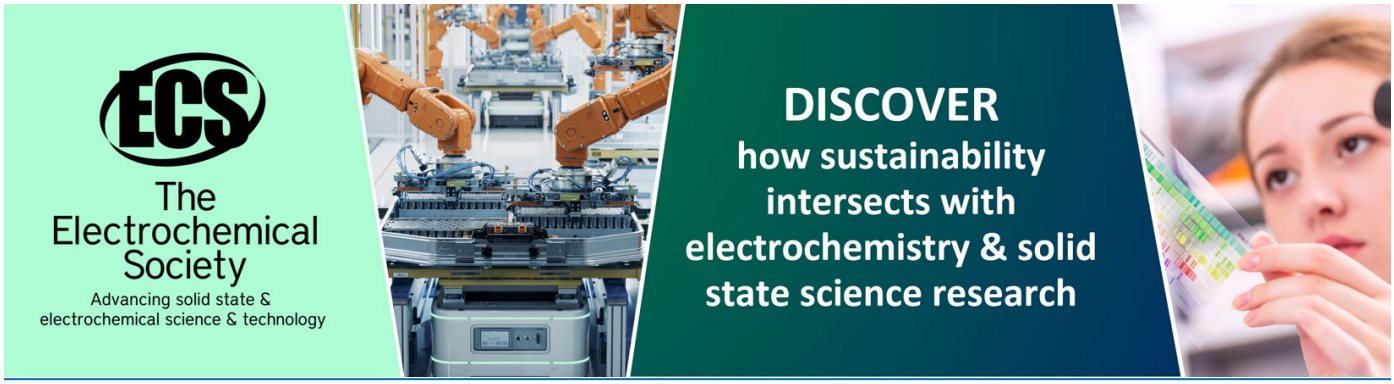
Extended Cox model for breast cancer survival data using Bayesian approach: A case study



Extended Cox model for breast cancer survival data using Bayesian approach: A case study

R Arsyad, S A Thamrin* and A K Jaya

Department of Statistics, Faculty of Mathematics and Natural Sciences, Hasanuddin University, Makassar, Indonesia, 90245
Email*: tuti@unhas.ac.id

Abstract. Breast cancer (*carsinoma mammae*) is one type of cancer that occurs due to abnormal breast cell growth. Some of the factors that are thought to trigger breast cancer include the unhealthy lifestyles. The existence of these factors indicates that there is a correlation between breast cancer and patient survival. One of method for analyzing survival data is Cox proportional hazard. Cox proportional hazard model implies that each covariate is proportional. But in reality, there are often cases where there is a disproportionate covariate, in the sense that there is a relationship with the time, called time dependent covariate. In this case an extended of the Cox proportional hazard model needs to be done. Therefore, the aim of this paper to determine the relationship between the breast cancer patients' survival time and the factors that influence it using extended Cox model with Bayesian approach. This methodology is applied to breast cancer survival data from Hasanuddin University hospital in Makassar, Indonesia, for the period 2005-2018. The result shows the factors that substantially affect the breast cancer patients' survival time are marital status, histology, and leukocyte levels.

1. Introduction

Cancer is the leading cause of death and disability worldwide, affecting more than 14 million people each year, one type of cancer is breast cancer. Every year there are 1.7 million breast cancer cases and 552,000 who die [1]. There is 43.3% people suffering from breast cancer with a mortality rate of 12.9%. Breast cancer (*carsinoma mammae*) occurs because of abnormal breast cell growth. In South Sulawesi, breast cancer cases are ranked first among many cancers suffered by women [2]. Some of the factors that is trigger breast cancer are influenced by five dietary behaviors and risks, namely high growth mass index, lack of fruit and vegetable consumption, lack of physical activity, cigarette use and excessive alcohol consumption. This unhealthy behavior causes the risk of cancer becoming high. The existence of these factors indicates that there is a correlation between breast cancer and patient survival [3]. Therefore, a method is needed to see the relationship between survival time and the factors that influence it.

Survival analysis is one method of analysis regarding the period of time from the observation process to the occurrence of an event by looking at other things that affect the event. One of the objectives of survival analysis is to see the relationship between explanatory variables and survival time. One of method for analyzing survival data is Cox regression, that was introduced by Cox and Oakes (1972) [4]. In this model the accompanying variables were included in the model as independent variables and survival time as non-independent variables [5]. By applying the Cox regression model, it will be known the form of the relationship between variables where the form of the relationship represents the phenomenon under study and can produce or link what is desired with what is studied. This regression model is also known as the proportional hazard model because the proportional assumptions on the hazard function. In general, the Cox regression model is faced with situations where the possibility of individual failure at a time is influenced by one or more explanatory variables [5].

The Cox proportional hazard model implies that each covariate is proportional. But in reality, there are often cases where there is a disproportionate covariate, in the sense that there is a correlation with the time rating, so that in this case the model needs to be expanded. The extended Cox model is an extension of the model of the Cox proportional hazard model, which contains time-dependent covariates or multiplications of the covariates with functions over time.

The approach that will be used in this paper is Bayesian. In the Bayesian method population parameters are seen as variables that have an initial distribution (prior). Before drawing samples from a population, information is sometimes obtained about the parameters to be estimated. This information is then combined with information from samples to be used in estimating population parameters. It seems that the Bayesian method is more promising because there is additional information to infer population characteristics.

There is some research have been done that related to extended Cox. Husain et al. [6] estimated survival time of breast cancer patients using extended Cox proportional hazard, Isik et al. [7] chose the time function and the differences between time functions using extended Cox regression model. Kurniawan et al. [8] also used extended Cox model to determine the durability of debtor efforts on credit risk. Meanwhile, Saegusa et al. [9] considered to use extended Cox model for modelling the time varying treatment effect and score test statistics. Therefore, based on these research review, this paper will model the effect of breast cancer patients' survival times based on the factor influenced using extended Cox model with Bayesian approach.

2. Material and Methods

2.1 Data Source

The breast cancer data used in this study is taken at Hasanuddin University Hospital in Makassar, Indonesia from 2005-2018. The response variable of this study was the length of stay of patients until they were declared home either in a state of improvement, recovery, referral or even death. Meanwhile, the independent variables were age, treatment, occupation, marital status, malignancy, hemoglobin level, stadium, histology, leukocyte levels. While the censoring indicator is indexed based on survival status.

2.2 Survival Analysis

Survival function $S(t)$ is defined as the opportunity for an individual to survive with a survival time up to time *t*, which is as follows:

$$S(t) = P(T > t) \tag{2.1}$$

Survival function in equation (2.1) can also be expressed in the form of a density function in the following:

$$S(t) = P(T > t) = \int_t^{\infty} f(t)dt \tag{2.2}$$

Hazard function or also known as hazard rate expressed by $h(t)$. Mathematically can be written as follows:

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{P(t \le T < (t + \Delta t) | T \ge t)}{\Delta t} \\ &= \frac{F\prime(t)}{S(t)} = \frac{f(t)}{S(t)} \end{aligned} \tag{2.3}$$

2.3 Cox Proportional Hazard Regression

In general, the forms of the Cox proportional hazard model are as follows [10]:

$$h(t, X) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 +_{\dots} + \beta_k X_k) \tag{2.4}$$

where $h_0(t)$ is baseline hazard function, $\beta_1, \beta_2, \dots, \beta_k$ is regression parameter and $X_{1,}, X_2, \dots, X_k$ is the independent variables.

2.4 Weibull Distribution

The Weibull distribution is one of the continuous distributions in statistical probability theory because of its ability to approach several types of data distribution. According to Walpole and Myers in Beren et al. [11], Weibull distribution can also be used on reliability issues and life time testing such as the time span until the failure of a component occurs or the durability of a component is measured from a certain time to damage. The probability density function of the Weibull distribution is written as:

$$f(t;\alpha,\theta) = \frac{\alpha t^{\alpha-1}}{\theta^{\alpha}} \exp\left\{-\left(\frac{t}{\theta}\right)^{\alpha}\right\}, t \geq 0, \alpha > 0, \theta > 0 \quad (2.5)$$

Parameters α and θ determine the shape and scale of distribution. If $\theta^{-\alpha} = \mu$ then the Weibull distribution probability density function becomes:

$$f(t;\alpha,\theta) = \alpha\mu t^{\alpha-1} \exp(-\mu t^{\alpha}) \quad (2.6)$$

2.5 Extended Cox

The extended Cox model is an extension of the model of the Cox proportional hazard model in equation (2.4), which contains covariates that are time-dependent or multiply from the covariate with a function of time. The general form of the extended Cox model as:

$$h(t, X(t)) = h_0(t) \exp\left[\sum_{b=1}^{p} \beta_b X_b + \sum_{b=1}^{p} \delta_b X_b g_b(t)\right] \quad (2.7)$$

If there is a p_1 covariate that meets the PH assumption and there is p_2 that does not meet the PH assumption, then $p_1 + p_2$ then the following model is obtained as follows:

$$h(t, X(t)) = h_0(t) \exp\left[\sum_{b=1}^{p_1} \beta_b X_b + \sum_{b=p_{1+}1}^{p_2} \beta_b X_b + \sum_{b=p_{1+}1}^{p_2} \delta_b X_b g_b(t)\right] \quad (2.8)$$

2.6 Parameter Estimation

2.6.1 Bayesian Approach

In estimation theory, there are two approaches, namely the classical statistical approach and the Bayesian statistical approach. The classical statistical approach relies entirely on inference processes on data samples taken from the population. While the Bayesian approach, in addition to utilizing data samples taken from the population also takes into account an initial distribution called priors [12]:

2.6.2 Likelihood Function

The likelihood function is a joint density function $f(x_1, x_2, \ldots, x_n\ ;\ \theta)$ of random variables $X_1, X_2, \ldots, X_n$. For example there are n observations of $x_1, x_2, \ldots, x_n$which each has a function of probability density $f(x, \theta)$, the likelihood function of a function of θ which is denoted by $L(\theta)$, namely $L(\theta) = f(x_1, x_2, \ldots, x_n\ ;\ \theta)$, namely:

$$L(\theta) = \prod_{i=1}^{n} f(x_i\ ;\theta) \quad (2.9)$$

2.6.3 Prior Distribution

The main problem in the Bayesian method is the selection of the prior distribution $\pi(\theta)$ for a parameter. The prior indicates uncertainty about the unknown parameter θ. Prior distributions are grouped into two groups based on the shape of the likelihood function [13] as follows:

1. Relating to the form of distribution of the results of identification of the data pattern
   a. The distribution of the prior conjugate (conjugate) refers to the reference model analysis especially in the formation of its likelihood function so that in determining the prior conjugate it is always thought about determining the pattern which has a conjugate form with its likelihood builder density function.

b. The prior distribution is not a conjugate (non-conjugate), when giving priors to a model does not heed the pattern forming the likelihood function.

2. Regarding the determination of each parameter in the prior distribution pattern
   a. Prior informative distribution, referring to the giving of parameters from the prior distribution that have been selected, either prior conjugate or non-conjugate distribution, giving parameter values to the prior distribution will greatly influence the form of posterior distribution to be obtained from the information data obtained.
   b. The prior non-informative distribution, if the prior distribution distribution is not based on previous information. If knowledge of priors is very weak, priors can have normal distribution with zero mean and large variance. The impact of prior use with a zero average is the estimated parameter smoothed to zero. But because this refinement is done by variance, the smoothing can be reduced by increasing the variance.

2.6.4 Posterior Distribution

Posterior distribution is a conditional density function θ if X observation value is known, can be written:

$$f(\theta|x_i) = \frac{f(\theta,x_i)}{f(x_i)} \tag{2.10}$$

If θ is continuous, the prior and posterior distributions θ can be expressed by the density function. The conditional function of a random variable if it is known that the value of the second random variable is a function of density along with two random variables divided by the second random variable.

The function of the shared density needed can be written in the form of a prior distribution and the likelihood function is given as follows [14]:

$$f(\theta, x_i) = f(\theta)f(x_i;\theta) \tag{2.11}$$

where $f(\theta)$ is the prior distribution and $f(x_i,\theta)$ is the likelihood function. Then, marginal function is known as follows:

$$f(x_i) = \int_0^\infty f(\theta)f(x_i;\theta)\, d\theta \tag{2.12}$$

The posterior density function for continuous random variables can be written as:

$$f(\theta|x_i) = \frac{f(\theta)f(x_i;\theta)}{\int_0^\infty f(\theta)f(x_i;\theta)d\theta} \tag{2.13}$$

Whereas for discrete random variables, the posterior probability density function is given as follows:

$$f(\theta|x_i) = \frac{f(\theta)f(x_i;\theta)}{\sum f(\theta)f(x_i;\theta)} \tag{2.14}$$

3 Result

There are two methods that can be used to see whether a data follows the Weibull distribution, namely by using a comparison plot $\log(t)$ and $\log(-(\log)(S\,(t))$ and Mann Test. A data is called following the Weibull distribution if $\log(t)$ is linear with $\log(-\log(S(t)))$. Linearity makes it possible to see the suitability of using the Weibull model through a plot.

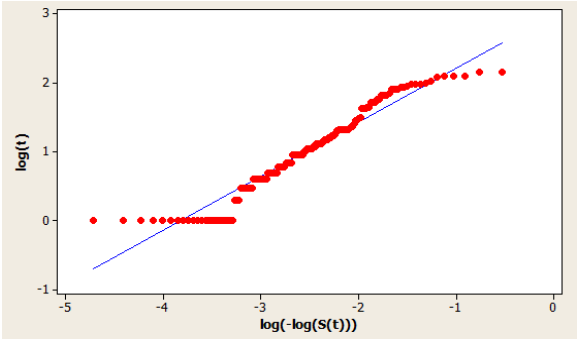


Figure 1. Comparison plot $\log(t)$ and $\log(-\log(S(t))$

Table 1 Estimated parameters of the Cox proportional hazard model for breast cancer survival data from Hasanuddin University hospital, Makassar, Indonesia during the period 2015-2018

Variables	Coefficient (β_j)	Standard Error	*p-value*
Age (X_1)	0.0203	0.0080	0.0118
Treatment (X_2)	0.1516	0.1002	0.1303
Occupation (X_3)	0.1396	0.1962	0.4768
Marital Status (X_4)	0.3140	0.3572	0.3794
Malignancy (X_5)	0.2325	0.1638	0.1554
Hemoglobin levels (X_6)	-0.0488	0.1514	0.7470
Stadium (X_7)	0.0360	0.0897	0.6883
Histology (X_8)	-0.2494	0.1618	0.1231
Leukocyte levels (X_9)	-0.0704	0.1985	0.7226

From Figure 1 it can be seen that $\log(t)$ and $\log(-\log(S(t)))$ indicate a linear relationship, this can be seen with the red dots that spread around the linear line. This shows that the data for patients with breast cancer followed the Weibull distribution. However, in terms of diagnosing the Weibull distribution using a plot, it is not an effective because it contains a subjective view in looking at the plot. To ensure that, the Weibull distribution data can be checked by using statistical procedures, namely the Mann Test.

Table 2. The proportional hazard statistical test result for each independent variable for breast cancer survival data

Variable	Rho	Chisq	p-value
Age	0.0449	0.2264	0.6342
Treatment	-0.0234	0.0939	0.7592
Occupation	-0.0544	0.4561	0.4994
Marital status	0.0276	0.1079	0.7425
Malignancy	0.0825	1.1029	0.2936
Hemoglobin level	-0.0172	0.0435	0.8347
Stadium	-0.0335	0.1772	0.6738
Histology	-0.2229	9.3493	0.0022
Leukocyte level	0.2088	6.4235	0.0112

The null hypothesis (H_0) is represented that data is distributed Weibull and vice versa. The value of M is -1.34 while the F value of the table at the significance level $\alpha = 0.05$ with $df_1 = k_1 = 76$ and $df_2 =$

$k_2 = 75$ is 1.46. Because the value $M = -1.34 < F_{0.05,76,75} = 1.46$, it was decided that H_0 was accepted. It means that the data used is followed by the Weibull distribution.

The fitting of the Cox proportional hazard model was conducted to find out the relationship between survival time and variables that affect survival time. The parameter estimates of the Cox proportional hazard model are obtained in Table 1. Based on Table 1, the Cox proportional hazard model as follows:
$h(t,X) = h_0(t)\exp\,(0.0203\,X_1 + 0.1516\,X_2 + 0.1396X_3 + 0.3140\,X_4 + 0.2325\,X_5 - 0.0488\,X_6 + 0.0360\,X_7 - 0.2494\,X_8 - 0.0704\,X_9)$

The results of checking the proportional hazard assumption with the schoenfeld residual approach are given in Table 2. The p-value of histology and leukocyte level less than 0.05, which mean that there is the correlation between the covariate and the survival time until the patient is declared cured. Therefore, the histology and leukocyte levels do not meet the assumption of proportional hazard.

The likelihood function for uncensored data is given as follows:

$$L(t_i|X_i,\delta_i) = \big(f(t_1|X_1,\delta_1)\big)^{\gamma_1}\big(f(t_2|X_2,\delta_2)\big)^{\gamma_2} \dots \big(f(t_n|X_n,\delta_n)\big)^{\gamma_n} = \prod_{i=1}^{n}\big(f(t_i|X_i,\delta_i)\big)^{\gamma_i}$$

The likelihood function for right censored data is as follows:

$$L(t_i|X_i,\delta_i) = \big(S(t_1|X_1,\delta_1)\big)^{1-\gamma_1}\big(S(t_2|X_2,\delta_2)\big)^{1-\gamma_2} \dots \big(S(t_n|X_n,\delta_n)\big)^{1-\gamma_n} = \prod_{i=1}^{n}\big(S(t_i|X_i,\delta_i)\big)^{1-\gamma_i}$$

Then the full likelihood function for the extended Cox model is as follows:

$$L(t_i|X_i,\delta_i) = \prod_{i=1}^{n}(\alpha {t_i}^{\alpha-1}(\exp(\beta_1X_{1i} + \beta_2X_{2i} + \cdots + \beta_9X_{9i} + \delta_8X_{8i} + \delta_9X_{9i}))^{\gamma_i} \exp(\beta_1X_{1i} + \beta_2X_{2i} + \cdots + \beta_9X_{9i} + \delta_8X_{8i} + \delta_9X_{9i})\,{t_i}^{\alpha}$$

The selection of the prior distribution is related to the parameters in the distribution pattern. There are two ways in determining the prior distribution, namely the informative prior and the non-informative prior. Prior used in this study is informative prior for each covariate which is followed by the normal distribution, that is $(\beta \sim N(\mu,\sigma^2))$, $(\delta \sim N(\mu,\sigma^2))$ and α has Gamma distribution, $\alpha \sim \Gamma(1,1)$.

$$f(\beta_i) = \frac{1}{\sigma_i\sqrt{2\pi}}\exp\left(-\frac{(\beta_i-\mu_i)^2}{2{\sigma_i}^2}\right), i = 1,2,\dots,9$$

$$f(\delta_i) = \frac{1}{\sigma_i\sqrt{2\pi}}\exp\left(-\frac{(\delta_i-\mu_i)^2}{2{\sigma_i}^2}\right), i = 8,9$$

$$f(\alpha) = \frac{1}{\Gamma(1)}\exp\,(-\alpha)$$

Assuming $\mu_i = 0$ and ${\sigma_i}^2$=100, the prior distribution is obtained for the parameters $\beta_1, \beta_2, \dots, \beta_9, \delta_8, \delta_9$ as follows:

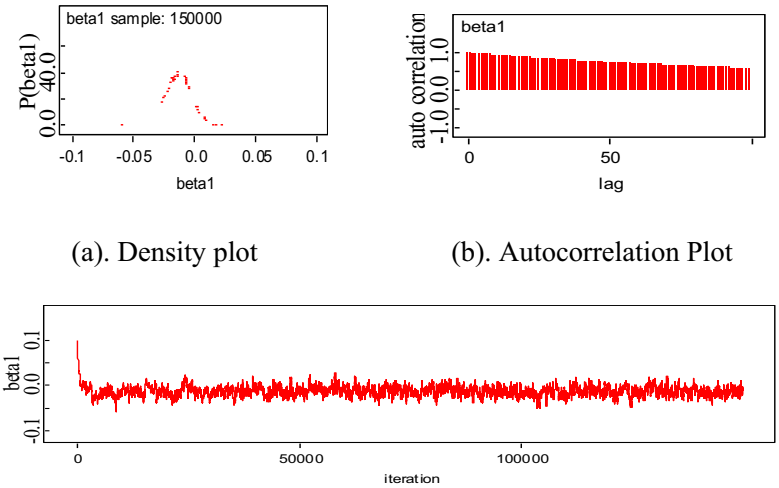
$$f(\beta_i) = \frac{1}{10\sqrt{2\pi}}\exp\left(-\frac{{\beta_i}^2}{200}\right), i = 1,2,\dots,9$$

$$f(\delta_i) = \frac{1}{10\sqrt{2\pi}}\exp\left(-\frac{{\delta_i}^2}{200}\right), i = 8,9$$

The posterior distribution is expressed by a comparison between the function of joint density and marginal density function. The joint density function can be written in the form of the likelihood function and prior distribution. Marginal density function is written by integrating the shape of the likelihood function and the prior distribution.

$$f(\boldsymbol{\beta}, \boldsymbol{\delta}, \alpha \,|t_i, X_i) = \frac{A}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} A\, d\alpha\, d\beta_1\, d\beta_2 \dots d\beta_9\, d\delta_8\, d\delta_9}$$

To get the appropriate parameter estimation results, it is necessary to check the convergence of each parameter. A review of the convergence of each parameter can be seen based on the diagnostic plot of the results of each parameter. The density, autocorrelation and history plots obtained for age are given in Figure 2.



(a). Density plot (b). Autocorrelation Plot

(c). History Plot

Figure 2. Diagnostic plots for Age

Table 3 Summary of posterior estimated parameters breast cancer patients' survival data from Hasanuddin University Hospital in Makassar during 2015-2018

Variable	Mean	95% Credible Interval (CI)
Age	-0.0123	$(-0.0350; 0.0086)$
Treatment	0.1811	$(-0.1018; 0.4392)$
Occupation	-0.2882	$(-0.7617; 0.2060)$
Marital Status	-1.2160	$(-2.447; -0.1986)$
Malignancy	0.0516	$(-0.2999; 0.3975)$
Hemoglobin level	-0.1946	$(-0.6054; 0.1938)$
Stadium	0.0533	$(-0.1641; 0.2692)$
Histology	0.4660	$(0.0969; 0.8250)$
Histology g(t)	-0.6536	$(-0.8466; -0.4715)$
Leukocyte level	-0.3433	$(-0.8705; -0.3531)$
Leukocyte level g(t)	-0.1222	$(-0.2301; -0.0228)$
Form parameter (α)	0.9994	$(-0.2301; -0.0228)$

Figure 2 shows the history plot of each parameter do not perform a pattern or trend and it have stabilized after running the Markov Chain Monte Carlo (MCMC) algorithm with 150,000 iterations, discarding as 10,000 as burn-in a. Besides that, it can also be seen that the plot density for each covariate parameter has been smooth which indicates that the parameters have converged. Moreover, the autocorrelation plot is truncated only in the first lag and towards a value close to zero. This shows that the sample obtained in the autocorrelation plot has less than 1 autocorrelation. It means that the algorithm reaches convergence and the sample is already in the target distribution and autocorrelated.

The posterior estimated summary for each parameter can be seen in Table 3. From Table 3, the variables that substantial effect the survival time of patients with breast cancer are marital status, histology, histology time dependent, leukocyte levels, leukocyte levels time-dependent. This can be seen from the credible intervals of each parameter. These variables do not contain zero in their credible intervals.

In addition to the covariate, the posterior mean of shape parameter α for the Weibull distribution is 0.9994 with a credible interval $(-0.2301;\ -0.0228)$. Because of the parameter α does not contain zero at at a credible interval, the parameter α has an influence on the model.

Based on equation (2.8), the extended cox model is obtained as follows:

$$h\big(t, X(t)\big) = 0.9994\ t^{0,9994-1} \exp\big(-1.2160\ X_4 + 0.4660\ X_8 - 0.6536\ X_8\ \mathrm{g(t)}\ - 0.3433\ X_9 - 0.1222\ X_9\ \mathrm{g(t)}\big)$$

From this extended model, it can be seen that the marriage status covariate value has a hazard ratio of $\exp(-1.2160) = 0.296$, indicating the risk of unmarried patients suffering from breast cancer by 0.296 times than married patients. It can be said that married patients will recover faster than unmarried patients. For histological variable has a hazard ratio of $\exp(0.4660 - 0.6536) = 0.829$, indicating that each histological increase will increase the risk of death by 0.829 times. While time-dependent leukocyte level covariates have a hazard ratio of $\exp(-0.33433 - 0.1222) = 0.628$, indicating the risk of patients who have leukocyte levels $< 4300\ /uL$ has a death risk of 0.628 times than patients who has a leukocyte level of 4300 /uL, $11300\ /uL$ and $> 11300\ /\ uL$. It means that patients who have leukocyte levels $>\ 4300/uL$ will recover faster.

4 Conclusion

In this paper, the relationship between the breast cancer patients' survival time and the factors that influence it have been determined using extended Cox model with Bayesian approach. Based on the results of the analysis and discussion that has been conducted, the breast cancer patients' survival times model has been constructed. The factors that affect substantially the patients' survival times of breast cancer in Hasanuddin University Hospital from 2005 to 2018 are marital status, histology, histology time dependent, leukocyte levels and leukocyte levels time dependent.

References

[1] 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.

[2] Gatra.com. (2019). Most Breast and Cervical Cancer in Indonesia. https://www.gatra.com/detail/news/388960-Most-Breast-Cancer-and-Neck-Cervix-in-Indonesia. Accesed on June 20, 2019.

[3] 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.

[4] Collet. (2004). *Modelling Survival Data in Medical Research*. 2nd. ed. Chapman & Hall/CRC.
[5] Thamrin, S.A., Mengersen, K., and Gree, J.M. (2012). *Bayesian Weibull Survival Model for Gene Expression Data.* In Case Studies in Bayesian Statistical Modelling and Analysis by Alston, C.L., Mengersen, K., and Pettitt, A.N. United Kingdom. John Wiley and Sons, Ltd.
[6] Husain, H., Thamrin, S.A., Tahir, S., Mukhlisin, A., and Apriani, M.M. (2018). The Application of Extended Cox Proportional Hazard Method for Estimating Survival Time of Breast Cancer. *Journal of Physics: Conference Series*, 979, 012087: 1-8.
[7] Isik, H., Tukun, N. A., Karasoy, D. (2017). Extended Cox Regression Model: The Choice of Time Function. *AIP Conference Proceedings*, 1863, 240004
[8] Kurniawan, I., Kurnia, A., Sartono, B. (2015). Survival Analysis with Extended Cox Model about Durability Debtor Efforts on Credit Risk. *Forum Statistika dan Komputasi: Indonesian Journal of Statistics*, 20(2): 85-95
[9] Saegusa, T., Di. C., and Chen, Y.Q. (2014). Hypothesis Testing for an Extended Cox Model with Time-Varying Coefficients. *Biometrics*, 70(3): 619-628.
[10] Kleinbaum DC and Klein M. (2005). *Survival Analysis A Self-Learning Text*. 2nd. ed. New York: Springer.
[11] Beren, C., Thamrin, S.A., and Jaya, A.K. (2016). Estimation of Exponentiated Modified Weibull Extention Distribution Parameters. *Proceedings the National Seminar*, Oktober 2016. ISBN: 978-602-99837-2-2
[12] Rahmi, S. (2016). *Spatial Survival Modeling Using Bayesian Approach*. Skripsi. Hasanuddin University.
[13] Box, G.E.P. dan Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. Philippines: Addition-Wesley Publishing Company.
[14] Soejoeti, Z. dan Soebanar. (1998). *Bayesian Inference*. Karunika Universitas Terbuka: Jakarta.