Cox Proportional Hazard Regression Interaction Model and Its Application to Determine The Risk of Death in Breast Cancer Patients after Chemotherapy

M. Ivan Ariful Fathoni[1,2,*], Gunardi[1], Fajar Adi-Kusumo[1], Susanna Hilda Hutajulu[3] and Ibnu Purwanto[3]

[1]*Department of Mathematics, Universitas Gadjah Mada, Indonesia*

[2]*Department of Mathematics Education, Universitas Nahdlatul Ulama Sunan Giri, Indonesia*

[3]*Department of Internal Medicine, Universitas Gadjah Mada, Indonesia*

Abstract: *Introduction*: This research is based on medical record data of breast cancer patients who seek treatment at the Central General Hospital, dr. Sardjito Yogyakarta, from 2018-2021 has as many as 105 patients. Several risk factors for cancer include demographic factors, clinical factors, tumor factors, and therapy. These factors lead to different psychological states of patients, resulting in the rate of recovery and death of patients.

Objective: To determine the risk of death in breast cancer patients after chemotherapy.

Methods: The method used in this study is Cox Proportional Hazard survival analysis with an interaction model. The variables studied were age, marital status, profession, insurance, BMI, comorbidities, duration of chemotherapy, chemotherapy agent, chemotherapy type, and tumor size.

Results: The analysis results using SPSS software obtained the best hazard and survival model with four significant variables, namely the duration of chemotherapy, chemotherapy agents, chemotherapy types, and the interaction between BMI and chemotherapy types.

Conclusions: The most significant risk factor for death was palliative chemotherapy type with HR 27.195 and 3-5 chemotherapy agents with HR 4.997. Meanwhile, the long duration of chemotherapy and the interaction between lean BMI and palliative chemotherapy reduced the risk of death by HR 0.967 and 0.128, respectively.

Keywords: Survival Analysis, Cox Proportional Hazard, Breast Cancer, Chemotherapy, Risk of Death.

1. INTRODUCTION

Every woman around the world has a risk of developing breast cancer. Breast cancer is the second leading cause of death for women today. The World Health Organization (WHO) in 2020 stated that breast cancer is cancer with the newest cases in the world. WHO estimates that there will be 2,261,419 new cases and 684,996 people were dying of breast cancer in 2020 [1]. For example, in the United States, breast cancer is the most common type in women after skin cancer. Data on the American Cancer Society website [2] shows that 1 in 8 women in America has the opportunity to develop invasive breast cancer (spread to other organs), and 1 in 36 women in the country die from breast cancer. Not much different from developed countries, cancer is now the seventh deadliest disease in Indonesia. Of the many types of cancer suffered by the Indonesian population, the Ministry of Health noted that breast cancer is the most common cause [3].

This research is based on medical record data of breast cancer patients who seek treatment at the Central General Hospital, dr. Sardjito Yogyakarta. Central General Hospital dr. Sardjito is the main referral hospital in the Special Province of Yogyakarta and also functions as an academic hospital. This hospital has an outpatient cancer service, namely Tulip (Integrated Cancer Clinic). According to the Hospital-Based Cancer Registration Report, breast cancer cases accounted for 28.2%, the highest incidence of all cancers based on data collected from 2008-2017. Most of the patients who came for treatment at the TULIP installation at the Central General Hospital dr. Sardjito has entered an advanced stage (41.0% at stage 4 and 40.8% at stage 3) [4]. If breast cancer is found at an advanced stage, treatment becomes more complex and expensive, and treatment results are unsatisfactory and even accelerated death [5].

Some risk factors for breast cancer include age and race, and these factors cannot be changed. Some risk factors can change over time, especially those related to the environment and behavior, such as smoking, drinking alcohol, and eating patterns. Another factor influencing the development of breast cancer cases was investigated by Anwar [6]. This study evaluated potential determinants of awareness in breast and cervical cancer screening participation using survey data from Indonesia. Possible factors associated with greater cancer screening awareness and participation included health insurance, shorter distances to health services, and social participation. The results obtained show that there are socioeconomic differences and

*Address correspondence to this author at the Department of Mathematics Education, Universitas Nahdlatul Ulama Sunan Giri, Indonesia;
E-mail: fathoni@unugiri.ac.id

awareness in cancer screening participation in Indonesian women. The impact is the increasing number of breast cancer cases in Indonesia because it is not prevented from the start with pap smears, mammography, and breast self-examination (BSE). Based on the results of the analysis of the study, it is known that there are other factors apart from the medical/clinical side.

Several psychological factors can cause the risk of cancer. This study was conducted by Yang [7], who investigated the Fear of Progression (FoP) of cancer patients in China. The results showed that the patient's sociodemographic and psychological variables could increase FoP; as a result, the patient's disease could become more severe. It is also predicted to reduce the patient's survival rate. Psychological factors can be treated with proper treatment practices. Bail [8] investigated psychological symptoms among advanced cancer patients. Psychological factors in cancer patients are applied in this study by categorizing independent variables that are thought to affect the patient's psychological state. The variables used include demographic factors, clinical factors, as well as tumor and therapeutic factors.

2. METHODS

2.1. Survival Analysis

Survival analysis is a statistical method used to analyze data where the output variable is the length of time until the occurrence of an event by looking at the variables of concern. The incident in question is a unique event that occurs, for example, leaving the hospital. Survival time (T) is the length of time starting from the initial observation of an object until the observation ends or the occurrence of a specific event. Three factors must be considered in determining the survival time, namely:

a. Initial time is the time when an observation begins.

b. End time is the end time of observation. This time becomes the core event of the observation.

c. The time interval (in units of time) is the interval from the observation start to the observation end or the occurrence of a specific event.

Previous studies have been conducted to determine the mortality rate and survival rate of cancer patients. Bazhenova [9] conducted a study evaluated the characteristics and prognosis of neurotrophic tropomyosin receptor kinase (NTRK) fusion cancer in the real-world setting. The results showed a higher risk of death ratio for patients with neurotrophic tyrosine receptor kinase NTRK gene fusion, although the difference was insignificant. Another study analyzed the cancer trait genes responsible for the most critical phenotypic characteristics of malignant transformation and cancer progression by Nagy [10]. This study aimed to estimate the prognostic effect of cancer trait genes in several different cancer types using Cox Proportional Hazard Regression. The analysis results showed that age and tumor stage were variables that reached significance in the Cox model in most tumors.

Survival analysis has also been carried out in studies involving breast cancer patients. Ősz [11] conducted a study on selecting clinical breast cancer treatment based on an immunohistochemical determination of protein biomarkers. Cox regression and Kaplan-Meier survival analyzes were calculated to assess the predictive power of these protein biomarkers. In another study, Zengel [12] investigated the clinical course of breast cancer patients with the oligometastatic bone disease (OMBD). They evaluated patient demographic features, histopathological features with intrinsic tumor subtype, and treatment-related factors on "survival outcome" among metastatic groups. For more specific cases of breast cancer, especially triple-negative (TNBC) studied by Sarin [13]. This study aimed to analyze the epidemiology, treatment options, and survival of patients with TNBC. The survival analysis results of this study indicate the mortality rate and survival of patients with a duration of observation of more than three years.

These studies inspired the author in carrying out this research. In this study, survival analysis was used to analyze patients' mortality rate and survival at RSUP dr. Sardjito, Yogyakarta, with various predictor variables that have been discussed in the introduction. This study continues Fathoni [14] research which discusses the survival analysis of breast cancer patients in Yogyakarta. In Fathoni's study [14], the predictor variables used included basic laboratory tests, therapy, and tumor factors, with the results showing that the variables that had a significant effect on patient survival were chemotherapy regimen, hormonal therapy, and stage. As a reinforcement of the results of Fathoni [14], with the same data panel, the authors tried to re-analyze with different predictor variables. The predictor variables in this study were chosen as factors that influence the psychological state of breast cancer patients.

2.2. Survival and Hazard Function

The survival function is the probability that an individual will operate well for a particular time under specified operating conditions. Survival can be used to measure the success of a system in carrying out its

functions properly. If survival is defined as a function of the cumulative probability of a patient surviving more than time t, with $t > 0$, then the survival function $S(t)$ during the time interval $(t, t + \Delta t)$ according to Hanni & Wuryandari [15] is:

$$S(t) = 1 - F(t) = 1 - \int_0^t f(t)\, dt \quad (1)$$

The failure function (hazard) of survival time t is denoted by $h(t)$ and is defined as the probability that an individual will fail in the time interval $(t, t + \Delta t)$ given that the individual has lived for time t. According [15], the hazard function is stated in the following equation.

$$h(t) = \frac{f(t)}{1-F(t)} \quad (2)$$

2.3. Cox Regression Proportional Hazard

Various regressions are often used to analyze survival data, namely parametric regression, non-parametric regression, and semi-parametric regression. Parametric regression requires that the baseline survival follows a specific distribution. If the conditions are not met, non-parametric regression can be used. If the baseline hazard follows a non-parametric model while the independent variables follow a parametric model, semi-parametric regression is used, which is often known as Cox regression. Cox regression analysis is an analysis that is used to analyze the time data of the incident and to determine the relationship between the time of occurrence and one of the independent variables. Cox first developed Cox regression in 1972. This regression is more popularly used in research on health data, economic data, where the response variable is time (day, month, year). For example, data about the time a patient suffers from a particular disease starts from the initial admission to the hospital until certain events occur, such as death, recovery, or other special events [8].

Cox proportional hazard is a model used in survival analysis which is a semi-parametric model. Cox proportional hazard regression is used when the observed outcome is the length of time an event occurs. Initially, this modeling was used in the branch of statistics, especially biostatistics, which was used to analyze a person's death or life expectancy. However, as the times progressed, this modeling was widely used in various fields: academic, medical, social, science, engineering, agriculture, etc. [16]. The Cox model can be seen as the relationship between the independent and dependent variables, namely the survival time through the hazard. An individual's risk of death at any given time depends on the value of $x_1, x_2, \ldots, x_p$ of p independent variable $X_1, X_2, \ldots, X_p$. The set of independent variable values in the cox model is represented by a vector x, so that $x = (x_1, x_2, \ldots\ldots, x_p)$. It is assumed to be an independent variable that is independent of time. The Cox regression model can be expressed in the equation

$$h(t, x) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p) \quad (3)$$

where t is survival time, $h_0(t)$ is baseline hazard function, $\beta_1, \beta_2, \ldots, \beta_p$ is regression parameters, and $x_1, x_2, \ldots, x_p$ is the value of the independent variable $X_1, X_2, \ldots, X_p$.

The Cox regression model relies on a proportional hazard, and the effect of a given covariate is constant over time. This is very important to ensure that the covariate satisfies the proportionality assumption. If this assumption is not met, the Cox regression model is not met, and solutions such as the Interaction Cox Model or the Extended Cox Model are needed. The Cox interaction model is a model obtained by modifying the cox proportional hazard model. Modifications were made by controlling for covariates that did not meet the PH assumption. This control is done by stratifying or interacting with the covariate. Suppose there are m covariates. Suppose there is also no interaction between covariates. The Cox PH model formed is

$$h(t, X) = h_0(t) \exp[\beta_1 X_1 + \cdots + \beta_k X_k + \beta_{k+1} X_{k+1} + \cdots + \beta_m X_m] \quad (4)$$

From these m covariates, suppose there are k covariates that meet the PH assumption, and there are p covariates that do not meet the PH assumption, where $p = m - k$. Without eliminating the generality, for example, the covariate that does not meet the PH assumption is $X_{k+1}, X_{k+2}, \ldots, X_m$ after variable interaction is formed $k * p$ [3].

3. RESULTS

3.1. Research Variables

This research is applied research with a quantitative approach. This study takes or collects the necessary data and analyzes it using the Cox Proportional Hazard regression model to determine whether there is a significant influence of the factors that are thought to affect the psychological survival of breast cancer patients at dr. Sardjito Hospital Yogyakarta. The data used in this study is secondary data for breast cancer patients during 2018-2021. The amount of data obtained is patient medical record data as of May 1, 2021, with 147 data, but in this study, the data used were complete data of 105 data.

The data adequacy test using the slovin formula 5 is used to determine the minimum amount of data in the observations.

$$\boldsymbol{n = \frac{N}{1+Ne^2}} \quad (5)$$

Table 1: Independent Variables

Variables	Information	Categories	Percentage	Coding	
				(1)	(2)
Demographic Factors					
X_1	Age	-			
X_2	Marital Status	1 : Single 2 : Married	16.2% 83.8%	1 0	
X_3	Profession	1 : Profession 1 2 : Profession 2	85.7% 14.3%	1 0	
X_4	Insurance	1 : JKN PBI 2 : JKN non PBI	20.0% 80.0%	1 0	
Clinical Factors					
X_5	BMI	1 : Fat 2 : Thin 3 : Ideal	37.1% 11.4% 51.4%	1 0 0	0 1 0
X_6	Comorbid	1 : Hypertension 2 : Non Hypertension 3 : No Hypertension	17.1% 29.5% 53.3%	1 0 0	0 1 0
Tumor and Therapy Factors					
X_7	Chemo. Duration	-			
X_8	Chemo. Agen	1 : 3-5 2 : 1-2	69.5% 30.5%	1 0	
X_9	Chemo. Type	1 : Palliative 2 : Adjuvant dan Neoadjuvant	25.7% 74.3%	1 0	
X_{10}	Tumor Size	1 : 7-10 cm 2 : 3,1-6,9 cm 3 : 0-3 cm	30.5% 40.0% 29.5%	1 0 0	0 1 0

Based on the number of $\boldsymbol{N = 147}$, by determining $\boldsymbol{e = 0.052}$, we get $\boldsymbol{n = 104.04}$. So with 105 data is enough to be used in this study. The dependent variable in this study was the patient's survival time, while the independent variables were categorized as in Table **1**. The descriptive statistical analysis of categorical independent variable data also shown in Table **1** in percentage column, while descriptive analysis of continuous independent variables is shown in Table **2**.

Variables that were not categorized were age and duration of chemotherapy. Marital status is categorized into two, namely single (unmarried or divorced) and married. Professions are categorized into professions that do not have a fixed income (profession 1) such as housewives, traders, and farmers, and professions with a fixed income (profession 2) such as teachers, retirees, civil servants, and entrepreneurs. Insurance is categorized into JKN PBI for the poor and underprivileged and JKN non-PBI for the wealthy.

Body Mass Index (BMI) is calculated based on the patient's height and weight. The fat BMI category is assumed to have a greater probability of death than the lean and ideal BMI. Comorbid variables were divided into three, namely hypertension, non-hypertension, and no comorbidities. Hypertension was chosen as a separate category because hypertension is the most common comorbidity. Variable chemotherapy agents were categorized into patients who had received 1-2 chemotherapy agents and 3-5 chemotherapy agents.

Table 2: Descriptive Statistical Analysis of Continuous Data Variables

Variables	Min	Max	Average	Deviation Standard	Variance
Age (Year)	32	78	52.06	9.187	84.401
Chemo. Duration (Day)	0	273	131.7	48.742	2375.83
Treatment Duration (Day)	10	1013	699.3	263.394	69376.3

Types of chemotherapy are categorized into two, namely Palliative and a combination of Adjuvant and Neoadjuvant. Tumor size was categorized into three according to the size of the longest side, namely 0-3 cm, 3,1-6,9 cm, and 7-10 cm.

3.2. Survival Distribution Test of Patients

Testing the distribution of data is carried out using the Anderson Darling approach. An information is said to follow a distribution when the Anderson Darling value (A^2) obtained is the smallest compared to the Anderson Darling value in other distributions. If Anderson Darling (A^2) got is the smallest compared to Anderson Darling's value in other distributions, it accepts the initial hypothesis (H_0). Anderson Darling testing was done using EasyFit software.

The results of testing the distribution of the dependent variable with the hypothesis used are as follows.

H_0 : Survival time follows the Gen. Extreme Value distribution.

H_1 : Survival time does not follow the Gen. Extreme Value distribution.

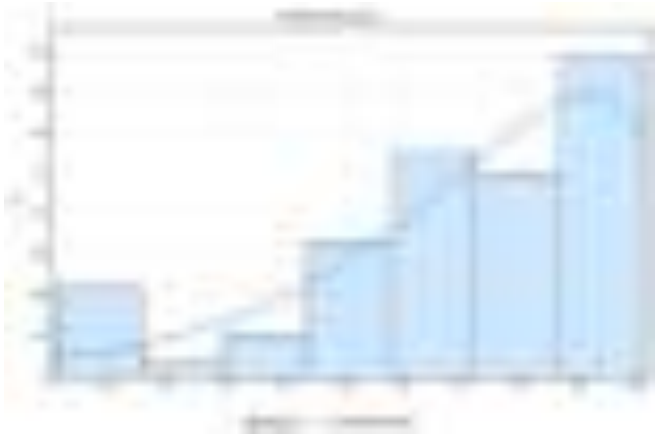

Figure 1: Gen. Extreme Value distribution PDF chart.

Based on the Goodness of Fit test results using the EasyFit application, the Anderson Darling value for the Gen. Extreme Value distribution is 0,54572. This value is the smallest Anderson Darling value of the other distributions. Based on the Anderson Darling value, the appropriate distribution is the Gen. Extreme Value distribution, as visualized in the Figure **1**.

3.3. Parameter Estimation of the Gen. Extreme Value Distribution

If the survival time data follows the Gen. Extreme Value distribution, then the function (t) is a function of the probability density of the Gen. Extreme Value distribution. The Probability Density Function (PDF) of the Gen. Extreme Value distribution is given in Figure **1** and the equation 6.

$$f(t) = \frac{1}{\sigma}\exp\left(-\left(1+k\frac{t-\mu}{\sigma}\right)^{-\frac{1}{k}}\right)\left(1+k\frac{t-\mu}{\sigma}\right)^{-1-\frac{1}{k}} \quad (6)$$

For $k \neq 0, \sigma > 0$ and domains $1 + k\frac{t-\mu}{\sigma} > 0$. Where t is survival time data, k is continuous shape parameter, σ is continuous scale parameter, and μ is continuous location parameter. The hazard and survival functions formed are as in equations 7 and 8.

$$h(t) = \frac{f(t)}{1-F(t)} = \frac{\frac{1}{\sigma}\exp\left(-\left(1+k\frac{t-\mu}{\sigma}\right)^{-\frac{1}{k}}\right)\left(1+k\frac{t-\mu}{\sigma}\right)^{-1-\frac{1}{k}}}{1-\exp\left(-\left(1+k\frac{t-\mu}{\sigma}\right)^{-\frac{1}{k}}\right)} \quad (7)$$

$$S(t) = 1 - F(t)$$

$$= 1 - \exp\left(-\left(1+k\frac{t-\mu}{\sigma}\right)^{-\frac{1}{k}}\right) \quad (8)$$

Based on the value of $k = -0{,}76004$, $\sigma = 287{,}1$ and $\mu = 669{,}65$ from the Gen. Extreme Value distribution, it is obtained

$$h(t) = \frac{f(t)}{1-F(t)} = \frac{\frac{1}{287{,}1}\exp\left(-(\mathcal{D})^{\frac{1}{0{,}76004}}\right)(\mathcal{D})^{-1+\frac{1}{0{,}76004}}}{1-\exp\left(-(\mathcal{D})^{\frac{1}{0{,}76004}}\right)} \quad (9)$$

$$S(t) = 1 - F(t) = 1 - \exp\left(-(\mathcal{D})^{\frac{1}{0{,}76004}}\right) \quad (10)$$

With

$$\mathcal{D} = 1 + k\frac{t-\mu}{\sigma}$$

$$= 1 + (-0.76004)\frac{t-669{,}65}{287{,}1} \quad (11)$$

3.4. Cox Regression Model Selection

3.4.1. Univariate Analysis

Univariate analysis was performed using Cox analysis on all independent variables. The results of the univariate analysis become a selection reference in the selection of variables that will be included in the multivariate analysis. The results of the analysis are shown in Table **3**. In the results of univariate analysis, the variables that meet the p-value $< 0{,}25$ are Profession, Insurance, BMI (1), Chemo Duration, Chemo Agen, and Chemo Type, with HR respectively 4.440 (95% CI 0.60-32.79), 1.704 (95% CI 0.72-4.06), 0.573 (95% CI 0.24-1.39), 0.982 (95% CI 0.98-0.99), 0.559 (95% CI 0.26-1.22), 6.365 (95% CI 2.89-14.03). These six variables were included in the multivariate analysis.

With the help of SPSS software, parameter estimates using the Breslow method are obtained for

Table 3: Variables in the Equation on Univariate Analysis

Variables	B	Exp(B)	p-value	95% CI for Exp(B)	
				Lower	Upper
Age	-0.001	0.999	0.959	0.957	1.043
Marital Status	-0.305	0.737	0.620	0.221	2.459
Profession	1.491	4.440	0.144	0.601	32.79
Insurance	0.533	1.704	0.228	0.716	4.056
BMI (1)	-0.558	0.573	0.219	0.235	1.394
BMI (2)	-0.188	0.828	0.765	0.241	2.849
Comorbid (1)	-0.553	0.575	0.390	0.163	2.028
Comorbid (2)	0.205	1.228	0.627	0.537	2.806
Chemo Duration	-0.019	0.982	0.001	0.975	0.988
Chemo Agen	-0.582	0.559	0.143	0.257	1.217
Chemo Type	1.851	6.365	0.001	2.888	14.03
Tumor Size (1)	0.501	1.650	0.333	0.599	4.547
Tumor Size (2)	0.228	1.256	0.660	0.456	3.460

each variable in the data of breast cancer patients in Table **3**. The estimation of the Cox model is obtained as in the equation 12 based on six selected variables.

$$h(t,X) = h_0(t)\exp(1.491X_3 + 0.533X_4 - 0.558X_5 - 0.019X_7 - 0.582X_8 + 1.851X_9) \quad (12)$$

with $h_0(t)$ as in equation 9. The overall test using partial likelihood ratio test was carried out to determine whether the model was correct, with the hypothesis:

$H_0: \beta_i = 0, i = 3,4,5,7,8,9$

(variables have no effect on the model)

$H_1: \exists\beta_i \neq 0, i = 3,4,5,7,8,9$

(variables have an effect on the model)

The rejection region H_0 is rejected if $G = -2(\ln L_R - \ln L_F) \geq \chi^2_{(0.05:6)}$ or $p - value < 0.05$. Based on the calculation with SPSS, the value of -2 log-likelihood for the Cox model without independent variables (null model) is $\ln L_R = -223.576$, and the value of -2 log-likelihood for the Cox model is $\ln L_F = -161.808$. So, we get the calculation $G = 123.536 \geq \chi^2_{(0.05:6)} = 12.59$. Based on these conditions, H_0 is rejected, it can conclude that at least one independent variable that affects the survival time or the independent variable.

3.4.2. Multivariate Analysis

The best cox model is obtained by eliminating the independent variable with the most significant p-value from each step. The process of removing the independent variables stopped at the three steps because p-value < 0,05 for all the significant variables. Using SPSS software, four variables are included in the best Cox model based on the results of backward elimination, namely BMI (1), duration of chemotherapy, chemotherapy agent, and type of chemotherapy. The duration of chemotherapy was the only significant continuous independent variable. BMI (1) is coding for BMI, with the fat category receiving a code of 1 (risk category) and 0 for the others (comparison category). Likewise, for other risk categories, chemotherapy with 3-5 agents and palliative chemotherapy types. Table **4** shows the estimation results of the best Cox model parameters based on the results of backward elimination.

Thus, the Cox model obtained based on the results of backward elimination is as follows.

Table 4: Parameter Estimation of the Best Cox Model with Backward Elimination

Variables	B	Exp(B)	SE	*p-value*
BMI(1)	-0.747	0.474	0.479	0.119
Chemo. Duration	-0.032	0.968	0.006	0.001
Chemo. Agen	1.624	5.071	0.601	0.007
Chemo. Type	2.727	15.292	0.502	0.001

$$h(t,X) = h_0(t)\exp(-0.747X_5 - 0.032X_7 + 1.624X_8 + 2.727X_9) \quad (13)$$

a partial likelihood analysis was carried out to find out which model is the best, with the following hypothesis:

$H_0: \beta_i = 0, i = 5,7,8,9$ (model null)

$H_1: \beta_i \neq 0, i = 5,7,8,9$ (model reduce)

The rejection region H_0 is rejected if $G \geq \chi^2_{(0.05:3)} = 7.815$ or $p - value < 0.05$, where $G = -2(\ln L_R - \ln L_F)$. From the output of SPSS software, it is obtained that the value of -2 log-likelihood for the Cox model without independent variables (null model) is $\ln L_R = -223.576$, and the value of -2 log-likelihood for the Cox model is $\ln L_F = -162.938$. Obtained $G = 121.276 \geq 7.815$, with p-value of reduce model is $0{,}001$. Because p-value < 0.05 so H_0 is rejected. This indicates that the model consisting of BMI (1), duration of chemotherapy, chemotherapy agent, and type of chemotherapy is the best.

3.5. Proportional Hazard Assumption

Furthermore, the Proportional Hazard Test is carried out to determine whether the significant variable meets the proportional hazard assumption. This test uses the log(-log) survival curve method. The results of plotting the log(-log) survival curve on the BMI variable are shown in Figure **2**. Based on the figure, a survival graph is obtained between the intersecting categories, so it can be concluded that BMI does not meet the proportional hazard assumption.

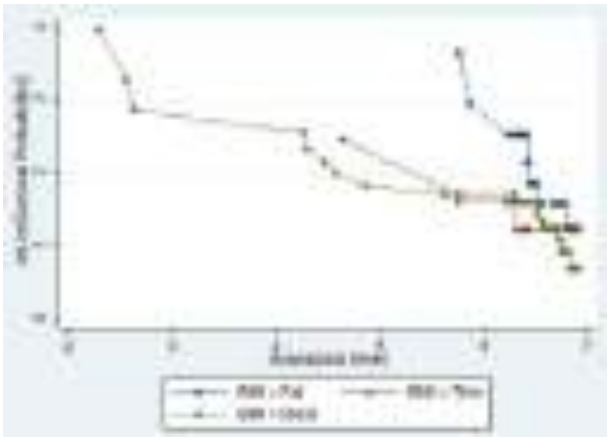

Figure 2: The log(-log) survival of BMI (proportional hazard assumption is not met).

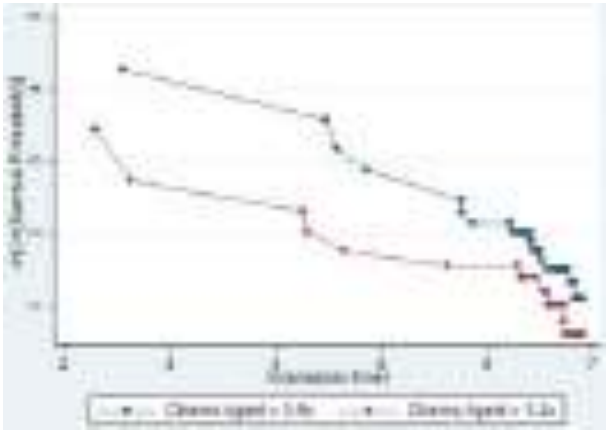

Figure 3: The log(-log) survival of Chemotherapy Agent (proportional hazard assumption is met).

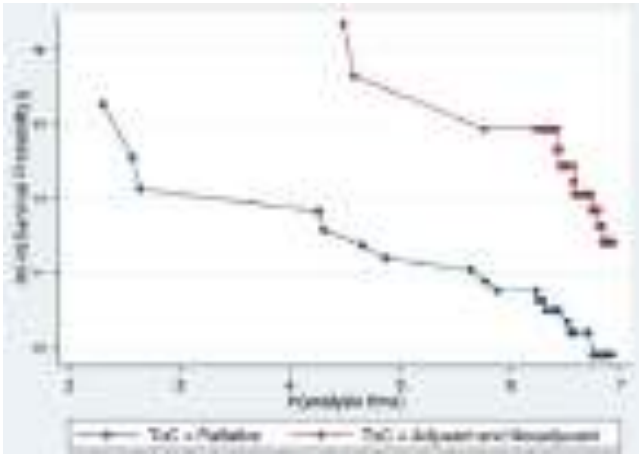

Figure 4: The log(-log) survival of Chemotherapy Type (proportional hazard assumption is met).

The variable duration of chemotherapy is a continuous variable with the smallest p-value, so it is considered essential and has a significant effect. The chemotherapy agents and types all meet the proportional hazard assumption, evidenced by the two log(-log) survival curve graphs shown in Figures **3** and **4**. The conclusion of checking the proportional hazard assumption of the four significant variables is shown in Table **5**.

Table 5: Conclusion of Checking Proportional Hazard Assumption

Significant Variables	PH Assumption
BMI(1)	Not Fulfilled
Chemotherapy Duration	Considered Important
Chemotherapy Agen	Fulfilled
Chemotherapy Type	Fulfilled

3.6. Cox Proportional Hazard Regression with Interaction Model

Based on the results of the analysis of PH assumptions, variables that do not meet the proportional hazard assumptions have interacted with other variables. With the help of SPSS software, parameter estimates using the Breslow method were obtained for each data variable, namely the duration of chemotherapy, chemotherapy agents, chemotherapy types, as well as the interaction between BMI (1) with chemotherapy duration, BMI (1) with chemotherapy agents, and BMI (1) with chemotherapy type. Four

Table 6: The Results of Cox Proportional Hazard Regression Interaction Model

Variables	B	SE	p-value	Exp(B)	95.0% CI for Exp(B)	
					Lower	Upper
X_7	-0.034	0.006	0.001	0.967	0.955	0.978
X_8	1.609	0.601	0.007	4.997	1.539	16.23
X_9	3.303	0.583	0.001	27.195	8.671	85.29
$X_5 * X_9$	-2.056	0.894	0.021	0.128	0.022	0.737

significant variables were obtained using the backward elimination method, shown in Table **6**. The interaction variable between BMI (1) and chemotherapy type was the only significant interaction variable besides duration of chemotherapy, chemotherapy agent, and type of chemotherapy.

The order of strength of the variables related to the dependent variable based on the HR value was the type of chemotherapy $(p < 0{,}001; HR = 27,195\ CI\ 95\%\ 8{,}671 - 85{,}293)$ chemotherapy agent $(< 0{,}01; HR = 4{,}997\ CI\ 95\%\ 1{,}539 - 16{,}226)$, duration of chemotherapy $(p < 0{,}001; HR = 0{,}967\ CI\ 95\%\ 0{,}955 - 0{,}978)$, and the last is the interaction of BMI(2) with chemotherapy type $(p < 0{,}05; HR = 0{,}128\ CI\ 95\%\ 0{,}022 - 0{,}737)$.

The final model of the respective hazard function and survival function that is formed is as follows.

$$h(t,X) = h_0(t)\exp(-0.034X_7 + 1.609X_8 + 3.303X_9 - 2.056(X_5 * X_9)) \quad (14)$$

$$S(t,X) = S_0(t)^{\exp(-0.034X_7 + 1.609X_8 + 3.303X_9 - 2.056(X_5 * X_9))} \quad (15)$$

with $h_0(t)$ as in equation 9, and $S_0(t)$ as in equation 10. Information:

$h(t,X)$ = hazard at a certain time

$h_0(t)$ = baseline hazard at a certain time

$S(t,X)$ = survival at a certain time

$S_0(t)$ = baseline survival at a certain time

X_5 = BMI, with a value of 1 if the category is thin and 0 if it is fat or ideal

X_7 = duration of chemotherapy starting from the start of chemo to the last chemo (days)

X_8 = hemotherapy agents, with a value of 1 when receiving 3-5 agents, and 0 when receiving

X_9 = type of chemotherapy, with a value of 1 if palliative and 0 if adjuvant or neoadjuvant

4. CONCLUSIONS

Based on the research results and discussion of the proportional hazard Cox regression method, the factors that affect the mortality and survival of patients with breast cancer can be seen. Variables that significantly affect the mortality rate and survival of patients with breast cancer are three variables: the type of chemotherapy, chemotherapy agent, and duration of chemotherapy. The three variables are tumor factors and therapy. In addition, there is also a pair of interaction variables between BMI and chemotherapy type. Based on the value of Exp (B) or the value of the Hazard Ratio, it can be concluded as follows:

1. The risk of death for patients with palliative chemotherapy was 27.95 times greater than patients with adjuvant and neoadjuvant chemotherapy. It is rational because palliative chemotherapy is given to patients at an advanced stage (stage IV).
2. Patients who have received 3-5 chemotherapy agents during chemotherapy have a risk of death 4.997 times greater than patients who have just received 1-2 chemotherapy agents.
3. The longer the duration of chemotherapy, the risk of death is 0.967 times smaller. The duration of chemotherapy depends on the patient's cancer condition. If cancer has a partial or complete response, the next cycle of chemotherapy will be longer, so the risk of death will be more negligible.
4. BMI conditions of patients categorized as underweight accompanied by the type of palliative chemotherapy can cause a 0.128 lower risk of death than BMI conditions and other types of chemotherapy. However, this condition has a negligible effect, considering that the p-value obtained is the largest among the significant variables.

ACKNOWLEDGEMENTS

The author would like to thank the Deputy for Research and Strengthening of the Indonesian Ministry of Research and Technology (National Research and Innovation Agency), who has provided research funds through Penelitian Dasar Unggulan Perguruan Tinggi tahun 2022.

REFERENCES

[1] Globocan. Breast: International Agency for Research on Cancer, WHO; 2020. Available from: https://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet.pdf.

[2] Society AC. Breast cancer 2019. Available from: https://www.cancer.org/cancer/breast-cancer.html

[3] Pangribowo S. Beban Kanker Di Indonesia. Jakarta Selatan: InfoDATIN Pusat Data dan Informasi Kementerian Kesehatan RI; 2019.

[4] FKKMK C. RKBR Maret 2018: RKBR Maret 2018 – canreg.fk.ugm.ac.id; 2018. https://doi.org/10.22146/mgi.35331

[5] Dalimartha S. Deteksi dini kanker & simplisia antikanker. Penebar Swadaya; 2004.

[6] Anwar SL, Tampubolon G, Van Hemelrijck M, *et al*. Determinants of cancer screening awareness and participation among Indonesian women. BMC Cancer 2018; 18(1): 1-11. https://doi.org/10.1186/s12885-018-4125-z

[7] Yang Y, Sun H, Liu T, *et al*. Factors associated with fear of progression in chinese cancer patients: sociodemographic, clinical and psychological variables. Journal of Psychosomatic Research 2018; 114: 18-24. https://doi.org/10.1016/j.jpsychores.2018.09.003

[8] Bail JR, Traeger L, Pirl WF, *et al*. editors. Psychological symptoms in advanced cancer. Seminars in oncology nursing; 2018: Elsevier. https://doi.org/10.1016/j.soncn.2018.06.005

[9] Bazhenova L, Lokker A, Snider J, *et al*. TRK fusion cancer: patient characteristics and survival analysis in the real-world setting. Targeted Oncology 2021; 16(3): 389-399. https://doi.org/10.1007/s11523-021-00815-4

[10] Nagy Á, Munkácsy G, Győrffy B. Pancancer survival analysis of cancer hallmark genes. Scientific Reports 2021; 11(1): 1-10. https://doi.org/10.1038/s41598-021-84787-5

[11] Ősz Á, Lánczky A, Győrffy B. Survival analysis in breast cancer using proteomic data from four independent datasets. Scientific Reports 2021; 11(1): 1-15. https://doi.org/10.1038/s41598-021-96340-5

[12] Zengel B, Kilic M, Tasli F, *et al*. Breast cancer patients with isolated bone metastases and oligometastatic bone disease show different survival outcomes. Scientific Reports 2021; 11(1): 1-12. https://doi.org/10.1038/s41598-021-99726-7

[13] Sarin R, Khandrika L, Hanitha R, *et al*. Epidemiological and survival analysis of triple-negative breast cancer cases in a retrospective multicenter study. Indian Journal of Cancer 2016; 53(3): 353.

[14] Fathoni MIA, Gunardi, Adi Kusumo F, Hutajulu SH. Survival analysis of breast ancer patients in Yogyakarta. Journal of Physics: Conference Series; 2021: IOP Publishing. https://dx.doi.org/10.1088/1742-6596/1722/1/012060

[15] Hanni T, Wuryandari T. Model Regresi Cox Proporsional Hazard pada Data Ketahanan Hidup. Media Statistika 2013; 6(1): 11-20. https://doi.org/10.14710/medstat.6.1.11-20

[16] Ernawatiningsih NPL. Analisis Survival Dengan Model Regresi Cox Study Kasus: Pasien Demam Berdarah Dengue di Rumah Sakit Haji Surabaya. Jurnal Matematika 2012; 2(2): 25-32. http://eprints.unm.ac.id/id/eprint/8351

Received on 22-09-2022 Accepted on 15-10-2022 Published on 21-10-2022

https://doi.org/10.6000/1929-6029.2022.11.13

CONTINUING REVIEW APPROVAL OF APPROVAL

Ref: KE/FK/0432/EC/2019

Ref : KE/FK/0444/EC/2020

Title of the Research Protocol	:	Analisis Efek Samping Kemoterapi, Faktor Resiko yang Memprediksi Kejadian Efek Samping, Serta Pengaruhnya Terhadap Kesintasan dan Kualitas Hidup Penderita Kanker Payudara: Studi Kohort Prospektif
Document(s) Approved and version	:	1. Study Protocol version 3.0 Januari 2020 2. Information for Subjects version 3.0 Januari 2020 3. Informed consent form version 3.0 Januari 2020
Principle Investigator	:	dr. Susanna Hilda Hutajulu, Ph.D., Sp.PD-KHOM.
Participating Investigator(s)	:	1. dr. Johan Kurnianda, Sp.PD-KHOM. 2. dr. Ibnu Purwanto, Sp.PD-KHOM. 3. dr. Kartika Widayati, Sp.PD-KHOM. 4. dr. Mardiah Suci Hardiyanti, Ph.D., Sp.PD-KHOM. 5. dr. Siswi Oktariani, Sp.PD. 6. dr. Vita Yanti Anggraeni, M.Sc., Ph.D., Sp.PD., Sp.JP. 7. dr. Anggoro Budi Hartopo, M.Sc., Ph.D., Sp.PD., Sp.JP. 8. dr. Hasanah Mumpuni, Sp.PD., Sp.JP(K). 9. Prof. Dr. dr. Budi Yuli Setianto, Sp.PD-KKV., Sp.JP(K). 10. dr. Yudiyanta, Sp.S(K). 11. Dr. dr. Irianiwati W., Sp.PA(K). 12. Prof. Lee Soo Chin 13. Dr. Matthew Allsop, B.Sc., Ph.D. 14. Prof. Dra. Yayi Suryo Prabandari, M.Si., Ph.D. 15. M. Ivan Ariful Fathoni, S.Si., M.Si. 16. Dian Caturini S., M.Sc. 17. Riani Witaningrum, M.Sc., G.TH. 18. Sunarti, A.Md. 19. Sri Mardilah Wuryani, A.Md. 20. dr. Aras Amila Husna 21. dr. Irfan Haris 22. Norma Dewi Suryani, SKM. 23. Betrix Rifana, SKM. 24. Fitrina Mahardani K., SKM., MPH. 25. Syafriani, SKM., MPH. 26. dr. Agus Jati Sunggoro, Sp.PD. 27. dr. Herindita Puspitaningtyas 28. dr. Yufi Kartika Astari 29. dr. Brilliant Winoma Jhundy 30. dr. Arief Gusnanto
Date of Approval	:	09 APR 2020 (Valid for one year beginning from the date of approval)
Institution(s)/place(s) of research	:	Rumah Sakit Umum Pusat Dr. Sardjito Yogyakarta, Instalasi Catatan Medik, Departemen Bedah, Instalasi Tulip, Departemen Radioterapi, Departemen Patologi Anatomi, Departemen Kardiologi dan Kedokteran Vaskular, Instalasi Jantung

The Medical and Health Research Ethics Committee (MHREC) states that the document above meets the ethical principle outlined in the International and National Guidelines on ethical standards and procedures for researches with human beings.

The Medical and Health Research Ethics Committee (MHREC) has the right to monitor the research activities at any time.

The investigator(s) is/are obliged to submit:
- √ Progress report as a continuing review : Annually
- √ Report of any serious adverse events (SAE)
- √ Final report upon the completion of the study

Dr. dr. Eti Nurwening Sholikhah, M.Kes., M.Med.Ed.
Panel's vice chairperson

dr. Rizka Humardewayanti A., Sp.PD-KPTI.
Panel's secretary

P.S: This letter uses signature scan of the panel's chairperson and Secretary of the Ethics Committee. The hardcopy official letter with authority's signature will be issued when it is possible and are kept as an archive of the Ethics Committee	Validation number : 5ef193c38424f (http://komisietik.fk.ugm.ac.id/validasi)