THE ANALYSIS OF FACTORS AFFECTING STAGE LEVEL OF BREAST CANCER PATIENTS AT RSUD DR. SOETOMO SURABAYA

Destri Susilaningrum¹, Nur Azizah², Mutiah S³ and Mukti Ratna D⁴

Sepuluh Nopember Institute of Technology, Indonesia¹,²,³ and ⁴
Email: destr_its@yahoo.com¹, nur.azizah.1907@gmail.com², mutiahsalamah@yahoo.com³ and Mukti_Ratna@its.ac.id⁴

Keywords
Binary Logistic Regression, Breast Cancer, Stage

Abstract
Breast cancer is the cancer with the largest number in the world, and ranks as the fifth largest cause of death, which is 6.9% of the total 9,958,133 million cases of cancer deaths. Breast cancer is one of the main killers of women in the world and in Indonesia. Breast and cervical cancer dominate cancer cases in East Java. Throughout 2019, RSUD Dr. Soetomo Surabaya received 167,000 cancer patients, and the highest case was breast cancer. Binary logistic regression is a method of analysis that used to find out the relationship between the response variables (Binary or dichotomous) with the predictor variables that are polychotomous. The high number of breast cancer sufferers makes this disease requires special attention. The cancer staging or staging system is based on whether the cancer has spread from the breast to other parts of the body. Cancer stage is divided into two categories, namely early stage (stages 0 to III A) and late stage (stages III B to IV). This study will examine the factors that affect the level of cancer stage in breast cancer patients at RSUD Dr. Soetomo Surabaya in 2019. Using the binary logistic regression analysis method, it was found that the factors that significantly influence the stage of breast cancer are Grade and Obesity with a classification correctness level of 82.5%.

Introduction
Cancer is one of the main causes of morbidity and mortality in the world. According to the World Health Organization (WHO) in 2020 (Bray, Laversanne, Weiderpass, & Soerjomataram, 2021), breast cancer is the first cancer with the highest number in the world (Cao, Chen, Yu, Li, & Chen, 2021). Followed by lung cancer, colon cancer, prostate cancer, and stomach cancer (Cho, Park, & Kim, 2021). A survey conducted by WHO states that 45% of breast cancer incidences are in Asia (Sohn, Chang, & Miles, 2021).

Breast cancer also known as Carcinoma Mammmae (Ca Mammmae) is a malignant tumor that grows in breast tissue (Hwang, Peregrina, Maglalang, & Yoo, 2021). This cancer grows because of the abnormal growth of breast cells (Shaikh, Krishnan, & Thanki, 2021a). The cause of breast cancer is not known with certainty. So far it has been associated with female reproductive hormones (Bonfiglio & Di Pietro, 2021). This hormone is estrogen which plays a role in the growth and development of female reproductive organs, including breasts (Kumar et al., 2021). Breast cancer can spread or move to surrounding tissues, spread to lymph nodes (Shaikh, Krishnan, & Thanki, 2021b), enter blood vessels to other organs such as bones, lungs, liver, even the brain and cause malfunction of these organs which can lead to death (Mortezaee, 2021). The stage...
The Analysis of Factors Affecting Stage Level of Breast Cancer Patients at RSUD Dr. Soetomo Surabaya

Destri Susilaningrum¹, Nur Azizah², Mutiah S³, and Mukti Ratna D⁴

of breast cancer must be confirmed before the diagnosis is completed and treatment is selected (Metzger Filho et al., 2021). This process can determine whether the cancer has spread from the breast to other parts of the body. Cancer stages can be categorized into two, namely the Early stage which includes stages 0 to III A and the Last stage which consists of stages III B to IV.

RSUD Dr. Soetomo is a General hospital owned by the government of East Java Province, which is the largest hospital in East Java as well as a referral hospital for East Java and Eastern Indonesia. There is a poly that handles cancer, namely Poly Oncology where cancer treatment is carried out in an integrated manner and focused on organs, one of which is breast cancer. In 2019, Dr Soetomo Hospital was recorded 167,000 cancer patients, and the highest case was breast cancer. Cancer is also a high cause of death in Dr Soetomo Hospital Surabaya, from the 10 cancer cases hospitalized the highest are breast cancer patients, followed by cervical cancer and blood cancer.

This study aims to determine the factors that are thought to affect the level of breast cancer stage in breast cancer patients at Dr. Soetomo Hospital Surabaya in 2019. The analysis used was binary logistic regression. The benefit of this research is that knowing these causative factors, the patient can control these factors so that the development of staging can be controlled.

Research Method

Binary Logistic Regression

Binary logistic regression is a method of analysis that used to find out the relationship between the response variables (Binary or dichotomous) with the predictor variables that are polychotomous. The response variable (y) consists of two categories: "success" (y = 1) and "failure" (y = 0). The model logistic regression given as

\[
\pi(x_i) = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip})} \quad (1)
\]

The Equation 2.1 can be explained as a logit model: by logit transformation of \( \pi(x_i) \) as:

\[
g(x_i) = \ln \left( \frac{\pi(x_i)}{1 - \pi(x_i)} \right) \quad (2)
\]

\[g(x_i) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_j x_{ij} + \ldots + \beta_p x_{ip} = \sum_{j=0}^{p} \beta_j x_{ij}\]

The test is conducted to get the best model which was built by the significant parameters. Parameters were first tested simultaneously and then tested partially to get the significant parameters.

Simultaneously Test of Parameters

Hypothesis:

\[H_0: \beta_1 = \beta_2 = \ldots = \beta_p = 0\]

\[H_1: \text{ at least one of } \beta_j \neq 0 ; j = 1, 2, \ldots, p\]

Test Statistic:
\[ G = -2 \ln \left[ \frac{\left( \frac{n_i}{n} \right)^{n_i} \left( \frac{n_0}{n} \right)^{n_0}}{\sum_{i=1}^{n} \hat{\pi}_{ij}^{n_i} (1 - \hat{\pi}_{ij})^{(1-n_i)}} \right] \]  

(3) 

With \( n_i = \sum_{j=1}^{n} y_{ij} \); \( n_0 = \sum_{j=1}^{n} (1 - y_{ij}) \); \( n = n_i + n_0 \)

Reject \( H_0 \) if \( G > \chi^2_{(\alpha, df)} \)

**Partial Test of Parameters**

Hypothesis:

\( H_0 : \beta_j = 0 \)

\( H_1 : \beta_j \neq 0 ; j = 1, 2, ..., p \)

Test Statistic:

\[ W_j^2 = \frac{\hat{\beta}^2_j}{[SE(\hat{\beta}^2_j)]^2} \]  

(4) 

Reject \( H_0 \) if \( W_j^2 > \chi^2_{(\alpha, df)} \)

**Goodness of Fit Model**

The test for overall fit of a Binary logistic regression model with the hypothesis:

\( H_0 : \) Model Fit

\( H_1 : \) Model Not Fit

Test Statistic:

\[ \hat{C} = \sum \left( \frac{\sum_{j=1}^{n_i} y_{ij} - \sum_{j=1}^{n_i} \hat{\pi}_{ij}}{\left( \sum_{j=1}^{n_i} \hat{\pi}_{ij} \right) \left( 1 - \frac{\sum_{j=1}^{n_i} \hat{\pi}_{ij}}{n_i} \right)} \right) \]  

(5) 

\( y_{ij} \): the Binary outcome for observation \( j \) in group \( i \) of the partition; \( i = 1, \ldots, g; j = 1, \ldots, n_i \)

\( \hat{\pi}_{ij} \): the corresponding fitted probability for the model fitted to the ungrouped data.

Reject \( H_0 \) if \( \hat{C} > \chi^2_{(g-1, \alpha)} \) or P-value \( < \alpha \) or P-value \( < \alpha \)

**Classification Procedure**

This procedure was used to evaluate the result of prediction value given by the best model to compare with the observation value, give an evaluation on classification procedure to see the probability of miss classification. It is measured by apparent error rate (APER). APER value stated the proportion value of miss classification sample by the function of classification. If the subject only classified as two groups, \( y_1 \) and \( y_2 \), then determination of classification errors can be known through the classification table described in Table 1.
Table 1. Classification Table.

<table>
<thead>
<tr>
<th>Observation</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td>$n_{11}$</td>
</tr>
<tr>
<td>$y_2$</td>
<td>$n_{12}$</td>
</tr>
<tr>
<td>$y_1$</td>
<td>$n_{21}$</td>
</tr>
<tr>
<td>$y_2$</td>
<td>$n_{22}$</td>
</tr>
</tbody>
</table>

Description:
- $n_{11}$: The number of the subject of $y_1$ on the correct classified as $y_1$
- $n_{12}$: The number of the subject of $y_1$ on the misclassified as $y_2$
- $n_{21}$: The number of the subject of $y_2$ on the misclassified as $y_1$
- $n_{22}$: The number of the subject of $y_2$ on the correct classified as $y_2$

\[
APER(\%) = \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}} \times 100\% \quad (6)
\]

And the proportion of the correct classification = \(1 - APER(\%)\) = \(\frac{n_{11} + n_{22}}{n_{11} + n_{12} + n_{21} + n_{22}}\) 100\% 

Source of Data and Research Variables
This research used secondary data from the medical records of patients suffering breast cancer at RSUD Dr. Soetomo Surabaya in 2019 as many as 217 patients. The variables used consist of two kind variables: Respond variable, and Predictor variable.

Respond Variable
Respond Variable ($Y$) in this research is the level stage as given in Table 2.

Table 2. Respond Variable.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Level stage statue</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y$</td>
<td>$Y = 1$ = Early stage or $Y = 0$ = Last stage</td>
</tr>
</tbody>
</table>

Predictors Variable ($X$)
Predictor variable ($X$) gives in Table 3.

Table 3. Research Variable.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Variable</th>
<th>Categorical</th>
<th>Data Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>Grade</td>
<td>1= Grade 3, 2= Grade 2, 3= Grade 1</td>
<td>Ordinal</td>
</tr>
<tr>
<td>X2</td>
<td>Age</td>
<td>-</td>
<td>Ratio</td>
</tr>
<tr>
<td>X3</td>
<td>Obesity History</td>
<td>1= Obesity, 2= No obesity</td>
<td>Nominal</td>
</tr>
<tr>
<td>X4</td>
<td>Comfort Level</td>
<td>1= Pain, 2= No Pain</td>
<td>Nominal</td>
</tr>
<tr>
<td>X5</td>
<td>Patient’s Psyche</td>
<td>1= Anxiety / Depression, 2= Surrender</td>
<td>Nominal</td>
</tr>
<tr>
<td>X6</td>
<td>History of Anemia</td>
<td>1= Yes, 2= No</td>
<td>Nominal</td>
</tr>
<tr>
<td>X7</td>
<td>Treatment</td>
<td>1= Operation (MRM, BCS), 2= Chemotherapy</td>
<td>Nominal</td>
</tr>
<tr>
<td>X8</td>
<td>Family History of Breast Cancer</td>
<td>1= Yes, 2= No</td>
<td>Nominal</td>
</tr>
<tr>
<td>X9</td>
<td>Metastasis</td>
<td>1= Yes, 2= No</td>
<td>Nominal</td>
</tr>
</tbody>
</table>
The operational definition of each variable is as follows.

A. Grade Breast cancer

The WHO using criteria, namely the Nottingham Grading system (also called Elston-Ellis modification of the Scarff-Bloom Richardson grading system). This rating scale looks at 3 different cell images and is classified each score from 1-3. The classifications are:

Table 4. Breast Cancer Grade

<table>
<thead>
<tr>
<th>Grade</th>
<th>Score</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade I</td>
<td>Score 3-5</td>
<td>Low grade with well differentiated cancer (well differentiated) where the cancer cells do not grow quickly and do not appear to spread.</td>
</tr>
<tr>
<td>Grade II</td>
<td>Score 6-7</td>
<td>Cancer with moderate differentiation (moderately/intermediate differentiated) which has a picture between grades 1 and 3.</td>
</tr>
<tr>
<td>Grade III</td>
<td>Score 8-9</td>
<td>Cancer that is poorly differentiated (poorly differentiated or undifferentiated) in which cancer cells grow rapidly and are more likely to spread.</td>
</tr>
</tbody>
</table>

Another benefit of determining the grading is to determine the response to therapy that will be given (Crotty et al., 2021). At a poor degree of differentiation, where cell growth and spread is considered to be more rapid or aggressive (Redmond, McCarthy, Buchanan, Levingstone, & Dunne, 2021), additional therapy is needed besides definitive, namely by administering chemoradiation. The relationship between stage and grade of breast cancer is that the stage in breast cancer serves to describe the condition of the cancer, namely the location and spread of cancer cells and how far it affects other body organs.

B. Comfort Level

Pain is a common complaint after treatment of breast cancer patients, even years after treatment (De Baets et al., 2021). Cancer pain is often found in patients who first come for treatment (Drewes, Kuhlman, Andresen, & Olesen, 2021).

C. Patient’s Psyche

Reactions in some people with cancer vary widely, such as shock, fear, anxiety, feelings of grief, anger, sadness, and even withdrawal. Anxiety is a mental condition filled with worry and fear of what might happen, both related to limited problems and strange things.

D. Anemia

Anemia in cancer patients can cause fatigue and decreased quality of life and increased mortality. In general, the presence of anemia in cancer patients can increase mortality by up to 65%. Patients are said to be anemic if they have hemoglobin < 12 g/dL.

E. Family History of Breast Cancer

Family history of inherited breast cancer is a risk factor for breast cancer. Family history of breast cancer will increase the development of breast cancer at a young age.

F. Treatment

Breast cancer treatment consists of surgery, radiation therapy, chemotherapy and hormone blocking drugs.
G. Metastases
Cancer cells can infiltrate the surrounding tissues and spread (metastasize) through blood vessels and lymph vessels. Metastases are cancer cells that have spread beyond the organs or tissues from where the cancer first appeared. Spread of cancer cells to other parts of the body is affected by many things, such as the type of cancer, the severity of cancer stage, and the location of the cancer originating [15].

Result and Discussion
The Characteristics of Breast Cancer Patient at RSUD Dr Soetomo

Picture 1. the characteristic of (a) The Age, (b) The Stage Level, and (c) The Grade Level.

Picture 1. a) show that the Age with high percentage of sufferer breast cancer at RSUD Dr. Soetomo Surabaya are the ages between 45 to 60 years old about 50% of the breast cancer patients, while between the ages of 27 - 44 and more than 60 years each with almost the same percentage, namely 26% and 25%.

Picture 1. b) show that majority the breast cancer patients at RSUD Dr. Soetomo Surabaya in Last stage condition about 75%. And figure 1 c) show that the greater Grade is Grade 3 (Cancer that is poorly differentiated) about 56%, than Grade 2 with 37% and the rest are Grade 1 (7%).

The Factors That Affect the Stage Level of Breast Cancer
The result of the binary logistic regression analysis, with the full logit model which include all variables gives as below.

\[ g(x_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_9 x_{i9} = \sum_{j=1}^{9} \beta_j x_{ij} \]

Simultaneously Test of Parameter
To find out which parameters are significant, the simultaneously test with the hypothesis as in subsubsection 2.1.1 had done and give the result on Table 5.

Table 5. Simultaneous Test Results.

<table>
<thead>
<tr>
<th>$\chi^2$</th>
<th>Df</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>85.615</td>
<td>9</td>
<td>0.000</td>
</tr>
</tbody>
</table>

For $\alpha = 0.10$, it showed that P-value = 0.000 less than $\alpha$ indicated that reject $H_0$, the partial test was then conducted to find out the variables that significant in the model.
Table 6. Partial Test Results.

<table>
<thead>
<tr>
<th>Variables</th>
<th>B</th>
<th>Wald</th>
<th>Df</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>X₁: Grade</td>
<td>1.303</td>
<td>14.877</td>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td>X₃: Obesity History</td>
<td>-0.760</td>
<td>3.244</td>
<td>1</td>
<td>0.072</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.428</td>
<td>5.285</td>
<td>1</td>
<td>0.022</td>
</tr>
</tbody>
</table>

Table 6 shows that the P-value less than α = 0.10, this indicate the variables are significant to the model, and the variables are: Grade and Obesity History. The logit Binary logistics Regression model gives as below.

\[ g(x) = -1.428 + 1.303x_1 - 0.76x_3 \]

3.2.3 Goodness of fit model

The hypothesis of goodness of fit model as in subsubsections 2.1.3, shown in Table 7.

Table 7. Goodness Fit of Model Results.

<table>
<thead>
<tr>
<th>( \chi^2 )</th>
<th>Df</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.935</td>
<td>8</td>
<td>0.654</td>
</tr>
</tbody>
</table>

Table 7. showed that P-value is more than α = 0.10, indicated the test cannot reject \( H_0 \), and so the model fit.

Classification Procedure

The subject only classified as two groups, with \( Y_1: \) Secure and \( Y_2: \) Insecure. The determination of classification errors can be known through the classification table described in Table 8.

Table 8. Classification Table.

<table>
<thead>
<tr>
<th>Observed</th>
<th>Prediction</th>
<th>Percentage Corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Early Stage</td>
<td>Last Stage</td>
</tr>
<tr>
<td>Early Stage</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>Last Stage</td>
<td>8</td>
<td>154</td>
</tr>
<tr>
<td>Overall Percentage</td>
<td>82.5</td>
<td></td>
</tr>
</tbody>
</table>

Based on the results above, the factors affecting stage level of breast cancer patients at RSUD Dr. Soetomo Surabaya are Grade and Obesity History with the correct classification between predicted and observation 82.5%.

**Conclusion**

The factors that affect the level stage of breast cancer patients at RSUD Dr. Soetomo Surabaya are Grade and Obesity History.

**References**

Bray, Freddie, Laversanne, Mathieu, Weiderpass, Elisabete, & Soerjomataram, Isabelle. (2021). The ever-increasing importance of cancer as a leading cause of premature death worldwide. *Cancer*.


Copyright holder:
Destri Susilaningrum¹, Nur Azizah², Mutiah S³, Mukti Ratna D⁴ (2021)

First publication right:
Devotion: Journal of Research and Community Service
This article is licensed under: