Modeling of Breast Cancer Diagnosis Classification Based on Hospital Medical Records

M.Nadjib Bustan1*, M. Arif Tiro1 and Adiatma1

1 Department of Statistics, Universitas Negeri Makassar, Makassar, 90222, Indonesia

* mnbustan@unm.ac.id

Abstract: Accuracy of breast cancer diagnosis is required to provide appropriate action or treatment. Classification of breast cancer diagnosis can be divided into various stadiums from stages 0 through IV. The classification of the breast cancer stadiums is probably related to risk factors, clinical complaints and outcome of the given treatment. The aim of the study was to determine the most appropriate model of breast cancer classification in diagnosis based on risk factors, clinical symptoms and treatment received by the patients. Therefore, predictor factors in this study was consisting of the location of cancer, the presence of metastasis, chemotherapy, age, weight, marital status and parity. The data came from medical records of breast cancer patients treated at the hospital Dr. Wahidin Sudirohusodo, Makassar, Indonesia during the year 2016 as many as 261 where drawn 181 samples. The purpose of analysis was to determine the magnitude of the association between predictors and the classification of breast cancer diagnoses in logistic regression models. The most appropriate logistic regression model test was multinomial logistics where the values of Pearson and Deviance Chi-square Goodness of Fit showed statistically accepted, with pseudo r-square Nagelkerke 0.658. In addition, McFadden value 0.421 means this model is able to explain the variation of breast cancer stage classification by 42%. The analysis showed that the value of simultaneous test of logistic regression -2log-likelihood was 209.914 (p <0,001), where the main predictors were age and chemotherapy with \( g_0(x) = -42,903 - 0,069Age - 1,987 Chemotherapy \); \( g_2(x) = -42,646 - 0,069Age - 1,987 Chemotherapy \) and \( g_3(x) = -24,254 - 0,069Age - 1,987 Chemotherapy \). Moreover, without chemotherapy, breast cancer patients had an increased chance of getting into a higher stage with age, proportional to the stage of breast cancer diagnosis. The model explain the chances of an increased likelihood of clinical stages was related to the age and chemotherapy treatment.

1. Introduction
Breast cancer is taking position as the second-order of cancer in Indonesia with the highest prevalence compared with other cancers and non-infectious diseases[1] [2] [3]. The occurrence of breast cancer is associated with various determinants or predictor factors that may be consist of general risk factors, clinical factors, and outcomes of surgery or therapy given to cases of breast cancer. Common risk factors may include age, body weight or nutritional status, marital status, parity, and oral contraceptive use [4] [5], [6] [7]. For clinical predictors, it can be a tumor position [8], the existence of
metastases [9], and gene expression [10]. Breast cancer needs a prompt therapy to get healing and decreasing the incidence of illness. Therefore, the choice of therapy depends on the accuracy of the diagnosis, where the diagnosis of breast cancer has a certain classification. The are several different types of breast cancer diagnosis classification including clinical stages, histopathological features, radiological mammography figure [11]. The most commonly used classification is according to clinical stages consisting of 5 stages’ JCCC: Stages O, I, II (IA, IIB), III (IIIA, IIB, IIC), and IV [12] [13]. Provision of action will depend on the stage of breast cancer cases handled. In addition, treatment may include surgery/mastectomy [14], chemotherapy [15], radiotherapy, hormonal therapy and palliative [16]. Based on the above problems, this study was conducted to determine what predictor factors associated with the precision of breast cancer stage diagnosis classification by using a logistic regression approach model [17], [18] [19] [20] [21]. Ordinal logistic regression is one of the statistical methods to analyze response variables that have ordinal scale data and consist of three or more categories. The predictor variables used in this model are the categories or quantitative data [17]. The basis of ordinal logistic regression model is binary logistic model with Y is response variable that is value 1 for success event and 0 for failure event. In this model the probability is given by Y = 1 if it is known that X = x: \( \pi(x) = P(Y = 1 | X = x) \) and 1 - \( \pi(x) = P(Y = 0 | X = x) \). Binary logistic model with one explanatory variable shaped:

\[
\text{Logit} \left[ \pi(x) \right] = \log \left( \frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 x
\]

with X: explanatory variable or independent variable whereas \( \beta_0 \) and \( \beta_1 \) are model parameters. If this equation is converted to an exponential form then it will be obtained:

\[
\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}
\]

Logistic regression models for ordinal data are often referred to as cumulative logistics models. The response variable in the cumulative logistic regression model is the multilevel data represented by the numbers 1, 2, 3, ..., k. With k is the number of categorical response variables. The cumulative logistic regression model is the model obtained by comparing the cumulative probability, the probability is less than or equal to the response category j on the predictor variable p expressed in the \( x_i \) vector. \( P(Y \leq j|x_i) \), with possibly greater than category \( j \) response, \( x_i, P(Y > j|x_i) \). Then the cumulative opportunity form is defined as follows:

\[
\pi_k(x_k) = P(Y \leq j|x_i) = \frac{\exp(\beta_0 + \sum_{k=1}^{j} \beta_k x_k)}{1 + \exp(\beta_0 + \sum_{k=1}^{j} \beta_k x_k)}
\]

with \( k = 1, 2, ..., j, ..., r \)

\[
\pi_k(x_k) = P(Y \leq j|x_i) = \pi_1 + \pi_2 + \cdots + \pi_r
\]

This is the logistics distribution function in general:

\[
F(x) = \frac{1}{1 + e^{-x}} = \frac{1}{1 + e^x}
\]

If \( P(Y \leq j) \) is comparing with the probability a response of category \((j+1)\) to the category \( r \), and the result is the following:

\[
\frac{P(Y \leq j)}{P(Y > j)} = \frac{P(Y \leq j)}{1 - P(Y \leq j)} = \frac{\exp(\beta_0 + \sum_{k=1}^{j} \beta_k x_k)}{1 + \exp(\beta_0 + \sum_{k=1}^{j} \beta_k x_k)}
\]

\[
= \exp(\beta_0 + \sum_{k=1}^{j} \beta_k x_k)
\]

\[
\frac{P(Y \leq j)}{P(Y > j)} = \frac{P(Y \leq j)}{1 - P(Y \leq j)} = \frac{\pi_1 + \pi_2 + \cdots + \pi_j}{\pi_{j+1} + \pi_{j+2} + \cdots + \pi_r}
\]

Next, the transformation of logit model to logistic regression ordinal:
Logit\[P(Y \leq j)] = \log \frac{P(Y \leq j)}{1 - P(Y \leq j)} = \log \frac{\pi_1 + \pi_2 + \cdots + \pi_j}{\pi_{j+1} + \pi_{j+2} + \cdots + \pi_r}

Logit\[P(Y \leq j)] = \beta_0 + \sum_{k=1}^{r} \beta_k x_k

with the value of \( \beta_k \), for \( k = 1, 2, ..., r \) on each ordinal logistic regression model is the same. The parameters are then evaluated. For parameter testing, simultaneous test and partial test are performed. Concurrent tests are used to check the overall coefficients. In this simultaneous test, the Likelihood Ratio Test is used, this method is a model testing method by comparing the likelihood for the complete model (\( L_1 \)) and the multiplication for the model with all parameters equal to zero (\( L_0 \)). The purpose of analysis was to determine the magnitude of the association between predictors and the classification of breast cancer diagnoses in ordinal logistic regression models.

2. Material and Methods

2.1. Data Sources and Research Variable
Data source used this research is patient data of breast cancer patient at Wahidin Sudirohusodo Hospital Makassar in year 2016 as many as 181 patient. Variables used in this study consisted of response variables (Y) and predictor variables (X). Response variable used is category of stage type of breast cancer patient, which consist of 5 categories: Y = (0) Stadium 0; Y (1) Stadium I; Y = (2) Stage IIA, IIB; Y = (3) Stage IIIA, IIIB, IIIC; Y = (4) Stage IV. Research variables included 7 predictor potential variables: (1) Marital Status (1 = Unmarried; 2 = Married; 3 = Widow); (2) Chemotherapy (1 = Chemotherapy; 2 = NonChemotherapy); (3) Cancer Location (1 = (4) Patient Age (numeric), (5) Weight (numeric), (6) Parity (numerical), and (7) Metastase (1 = Yes, 2 = No ).

2.2. Step Analysis
The use stages of data analysis using ordinal logistic regression model is through the steps as follows:
1. Estimation of parameters, 2. Interpretation of parameters, 3. Paramater Model Testing; 4. Testing the suitability of the model that has been obtained. The parameter estimation is done by testing the logistic regression model and calculating the logit model parameter. Interpretation of the parameter is done by reading 3 logit functions obtained, where logit 0 is a logit function for stage 0, and logit 2 is a logit function for stage II and logit 3 as a logit function for stage III. Furthermore, from these three logit functions, there is an opportunity function for each category. Testing Paramater model is done to know the significance of parameters from the response variable as a whole. Hypothesis testing used is \( H_0: \beta_1 = \beta_2 = 0 \). Testing the suitability of the model has been obtained by testing the hypothesis. \( H_0: \) The model matches the data; \( H_1: \) Model does not match the data. Rejection region \( H_0: > \) or \( p\)-value \(< \alpha = 0.01 \). The model suitability test results are used to see if the model formed from the predictor variables above is appropriate or not in accordance with the data.

2.2.1 Parameter Estimation (Backward Method)
The parameter estimation value shows that there are two predictor variables that have significant effect on breast cancer stage, ie age and chemotherapy with \( p < \alpha (\alpha = 0.01) \). Logit model of both parameters are:

\[ g_0(x) = -42,903 - 0.069 Age - 1,987 Chemotherapy \]
\[ g_2(x) = -42,646 - 0.069 Age - 1,987 Chemotherapy \]
\[ g_3(x) = -24,254 - 0.069 Age - 1,987 Chemotherapy \]

2.2.2 Parameter Interpretation
Based on the above two logit functions, logit 0 is a logit function for stage 0 and logit 1 is a logit
function for stage I, IIA, IIB, IIIA, IIIB, IIIC. Furthermore, from both logit function is obtained the function of each category opportunities, namely:

\[ \hat{h}_0(x) = \frac{\exp(g_0(x))}{1 + \exp(g_0(x))} \]
\[ \hat{h}_2(x) = \frac{\exp(g_2(x))}{1 + \exp(g_0(x))} - \frac{\exp(g_0(x))}{1 + \exp(g_2(x))} \]
\[ \hat{h}_3(x) = \exp(g_3(x)) - \frac{\exp(g_2(x))}{1 + \exp(g_2(x))} - \frac{\exp(g_0(x))}{1 + \exp(g_0(x))} \]
\[ \hat{h}_4(x) = 1 - \hat{h}_0(x) - \hat{h}_2(x) - \hat{h}_3(x) \]

Hence, for Y=0 (Stadium 0)

\[ \hat{h}_0(x) = \frac{\exp(-42,903 - 0,069 \text{Age} - 1,987 \text{Chemotherapy})}{1 + \exp(-42,903 - 0,069 \text{Usia} - 1,987 \text{Chemotherapy})} \]

Suppose that every patient ages 1 year and the patient does not do chemotherapy, \( P(Y = 0 \mid X) \) is:

\[ \hat{h}_0(x) = \frac{\exp(-42,903 - 0,069(1) - 1,987(1))}{1 + \exp(-42,903 - 0,069(1) - 1,987(1))} \]
\[ = 2,98 \times 10^{-20} \]

For Y=2 (Stadium IIA, IIB)

\[ \hat{h}_2(x) = \frac{\exp(-42,646 - 0,069 \text{Usia} - 1,987 \text{Chemotherapy})}{1 + \exp(-42,646 - 0,069 \text{Usia} - 1,987 \text{Chemotherapy})} \]
\[ - \frac{\exp(-42,903 - 0,069 \text{Usia} - 1,987 \text{Chemotherapy})}{1 + \exp(-42,903 - 0,069 \text{Usia} - 1,987 \text{Chemotherapy})} \]

Suppose that every patient ages 1 year and patient does not do chemotherapy then \( P(Y = 2 \mid X) \) is:

\[ \hat{h}_2(x) = \frac{\exp(-42,646 - 0,069(1) - 1,987(1))}{1 + \exp(-42,646 - 0,069(1) - 1,987(1))} \]
\[ - \frac{\exp(-42,903 - 0,069(1) - 1,987(1))}{1 + \exp(-42,903 - 0,069(1) - 1,987(1))} \]
\[ = 3,85 \times 10^{-20} - 2,98 \times 10^{-20} \]
\[ = 8,73 \times 10^{-21} \]

For Y=3 (Stadium IIIA, IIIB, IIIC)

\[ \hat{h}_3(x) = \frac{\exp(-24,254 - 0,069 \text{Age} - 1,987 \text{Chemotherapy})}{1 + \exp(-24,254 - 0,069 \text{Age} - 1,987 \text{Chemotherapy})} \]
\[ - \frac{\exp(-42,646 - 0,069 \text{Age} - 1,987 \text{Chemotherapy})}{1 + \exp(-42,646 - 0,069 \text{Age} - 1,987 \text{Chemotherapy})} \]
\[ - \frac{\exp(-42,903 - 0,069 \text{Age} - 1,987 \text{Chemotherapy})}{1 + \exp(-42,903 - 0,069 \text{Age} - 1,987 \text{Chemotherapy})} \]
\[ = 8,73 \times 10^{-21} \]
Suppose that every patient ages 1 year and the patient does not do chemotherapy then, \( P(Y = 3 \mid X) \) is:

\[
\hat{r}_3(x) = \frac{\exp(-24,254 - 0.069(1) - 1.987(1))}{1 + \exp(-24,254 - 0.069(1) - 1.987(1))} - \frac{\exp(-42,646 - 0.069(1) - 1.987(1))}{1 + \exp(-42,646 - 0.069(1) - 1.987(1))} - \frac{\exp(-42,903 - 0.069(1) - 1.987(1))}{1 + \exp(-42,903 - 0.069(1) - 1.987(1))} = 3.74 \times 10^{-12} - 3.85 \times 10^{-20} - 2.98 \times 10^{-20} = 3.74 \times 10^{-12}
\]

For \( Y = 4 \) (Stadium IV)

\[
\hat{r}_4(x) = 1 - \hat{r}_0(x) - \hat{r}_2(x) - \hat{r}_3(x)
\]

Suppose that every patient ages 1 year and patient does not do chemotherapy, then \( P(Y = 4 \mid X) \) is:

\[
\hat{r}_4(1) = 1 - \hat{r}_0(1) - \hat{r}_2(1) - \hat{r}_3(1) = 1 - 2.98 \times 10^{-20} - 3.85 \times 10^{-20} - 3.74 \times 10^{-12} = 1
\]

Based on the value of the above opportunities, it can be said that if there is an additional age of patients 1 year and the patient does not perform, then the chances of an increase in stage cancer (0) equal to \( 2.98 \times 10^{-20} \).

The chance of an increased stage of cancer (IIA, IIB) is \( 3.85 \times 10^{-20} \) if there is an age increase of 1 year patient and the patient does not perform chemotherapy. Then an increase in stage cancer (IIIA, IIIB, IIIC) has an odds of \( 3.74 \times 10^{-12} \). While the chances of an increased stage of IV cancer in the case of an additional age of 1 year patient and the patient does not do chemotherapy is \( 1 \).

### 2.2.3 Parameter Model Testing

This test is done to know the significance of parameters from the response variable as a whole. The hypothesis testing used is:

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

\[ H_1 : \text{paling sedikit ada satu } \beta_i \neq 0 \]

Statistical test:

\[
G = -2 \ln \left[ \prod_{i=1}^{n} \hat{r}_i^{y_i} (1 - \hat{r}_i)^{(1-y_i)} \right]
\]

where

\[
n_1 = \sum_{i=1}^{n} y_i, \ n_0 = \sum_{i=1}^{n} (1 - y_i), \ n = n_1 + n_0.
\]

Rejection area: reject Ho if \( G > \chi^2_{(v,a)} \) with reject Ho if with degrees of freedom is the number of parameters in the model without \( \beta_0 \). Test results as follows.

| Table 1. Simultaneous Test of Ordinal Logistic Regression Mode |
|---------------------------------------------------------------|
| Model                  | -2 Log Likelihood | Chi-Square | df  | P-value |
|------------------------|-------------------|------------|-----|---------|
| Intercept Only         | 362,264           |            |     |         |
| Final                  | 209,914           | 152.350    | 9   | <0.001  |
Table 2. Partial Test of Ordinal Logistic Regression Model

|                | Estimasi  | Std. Error | Wald     | P-Value |
|----------------|-----------|------------|----------|---------|
| Stadium 0      | -42,903   | 1,746      | 604,077  | <0,001  |
| Stadium I & II | -42,646   | 1,741      | 599,994  | <0,001  |
| Stadium III    | -24,254   | 881,252    | 0,001    | 0.978   |
| Age            | -0,069    | 0,022      | 10,152   | 0.001   |
| Chemotheraphy  | -1,987    | 0,428      | 21,519   | <0,001  |

Table 1 shows the results of simultaneous tests obtained by calculating values -2log-likelihood model is 209,914. Because P-value = 0,000 is less than α (α = 0,01), Ho is rejected, meaning that there is at least one predictor variable that has a significant effect on breast cancer stage classification at Wahidin Sudirohusodo Hospital Makassar. In this case only age and chemotherapy variables had a significant effect on breast cancer stage elevation (See Table 2).

2.2.4. Model Conformity Test
The model conformity test is used to see if the model formed from the predictor variables above is appropriate or incompatible with the data. The hypothesis used is:
H0: The model matches the data; H1: Model does not match the data
α = 0.01
Test Statistics (Hosmer and Lemeshow, 2000):

\[ \hat{C}(\text{Hosmer - Lemeshow}) = \sum_{k=1}^{g} \left( \frac{\hat{C}_k - n_k \cdot \hat{n}_k}{n_k \cdot \hat{n}_k \cdot \hat{n}_k} \right) \]

Daerah penolakan H0: \( \hat{C} > \chi^2_{(g-2)} \) atau p-value <α = 0.01.

H0: The model matches the data; H1: Model does not match the data; Rejection region H0: > or p-value <α = 0.01.

Table 3. Goodness of Fit

|              | Chi-Square | Df  | p-Value |
|--------------|------------|-----|---------|
| Pearson      | 303,871    | 519 | ≈1,000  |
| Deviance     | 209,914    | 519 | ≈1,000  |

Based on Table 3 of the "Goodness of Fit" above it is found that the p-values produced for both Pearson and Deviance are more than α (α = 0.01), p-values of Pearson and Deviance are 1,000 and 1,000. So Ho is not rejected, which means the model obtained is in accordance with the data (there is no real difference between the observation results with the possibility of the predicted model). Thus, the variables that significantly influence the increase of breast cancer stage at Wahidin Sudirohusodo Hospital Makassar are age and chemotherapy variable. To know the model that is eligible to be used can be viewed from the value of R2nya. Here is the R2 value of the best model.

Table 4. Pseudo R-Square

| Model        | Value R^2 |
|--------------|-----------|
| Cox and Snell| 0,569     |
| Nagelkerke   | 0,658     |
| McFadden     | 0,421     |
Based on the table "Pseudo R-Square" value Nagelkerke R Square of 0.658. This shows that as much as 65.8% of the models formed are eligible to use. In other words, the model was produced with seven predictor variables (ie, only the age and non-chemotherapy variables were significant while the other five variables did not significantly affect the stage of breast cancer.) In addition, the model was also able to explain the variation of breast cancer stage classification at Home Sakit Wahidin Sudirohusodo by 42.1%. Moreover, without chemotherapy, breast cancer patients have an increased chance of getting into a higher stage with age, proportional to the stage of breast cancer diagnosis. The model explain the chances of an increased likelihood of clinical stages was related to the age and chemotherapy.

3. Result and Discussion

Based on ordinal logistic regression analysis obtained count value -2log-likelihood model of 209.914 with p-value <0.001. In this case only the age and chemotherapy variables that significantly influence the increase in breast cancer stage. In Partial Test of Ordinal Logistic Regression significant variables obtained that two predictor variables that have a significant effect on breast cancer stage variables, namely age and chemotherapy variables. This can be seen in p-value <0.001. Based on the three logit functions for stage (0), stage (IIA, IIB), and stage IIIA, IIB, IIC) the following conclusions are obtained. In the event of an increase in patient age and the patient not taking chemotherapy even once, an increase in stage (0) cancer. The chance of an increase in stage cancer (IIA, IIB) is greater if there is an additional patient1 years and the patient does not take chemotherapy even once. Then an increase in stage cancer (IIIA, IIB, IIC) has an opportunity of. Based on the Goodness of Fit Test it is found that the p-values produced for both Pearson and Deviance are more than α (α = 0.01), 1,000. Thus, the model obtained is in accordance with the data (there is no significant difference between the observation results and the possibility of the predicted model). Thus, the variables that significantly influence the increase of breast cancer stage at Wahidin Sudirohusodo Hospital Makassar are age and chemotherapy variable. To find the model that is eligible to be used can be seen from the value of R’nya, namely: Pseudo R-Square Cox and Snell (0.569), Nagelkerke (0.658), and McFadden (0.42). This indicates that 66% of the model is eligible to be used. In other words, the model was produced with seven predictor variables; only age and chemotherapy variables that had a significant effect while the other five insignificant variables were used to diagnose breast cancer stage elevation, and also explained variations in breast cancer stage classification in Wahidin Sudirohusodo Hospital by 42%. In summary, without chemotherapy, breast cancer patients have an increased chance of getting into a higher stage with age, proportional to the stage of breast cancer diagnosis. The model explain the chances of an increased likelihood of clinical stages was related to the age and chemotherapy.

Acknowledgments

Authors highly appreciate to thank to Research Centre of Universitas Negeri Makassar and Dean of Faculty of Mathematics and Natural Sciences Universitas Negeri Makassar who have given a very kind cooperation during the research commencement. This paper was financially supported by PNBP (Tax-free Budget) of Universitas Negeri Makassar. Therefore, especially thanks was given to Dean of Faculty of Mathematics and Natural Sciences Universitas Negeri Makassar

References

[1] Who. Cancer Country Profile Indonesia.
[2] Kesehatan K, Penanggulangan K, Nasional K. Panduan Penatalaksanaan Kanker Payudaraudara.
[3] Bustan M 2006 Manajemen Pengendalian Penyakit Tidak Menular. Jakarta: Rineka Cipta.
2015.

[4] Barlow We, White E, Ballard-Barbash R, Et Al. Prospective Breast Cancer Risk Prediction Model For Women Undergoing Screening Mammography. *Jnci J Natl Cancer Inst*; 98: 1204–1214.

[5] Bustan Mn, Coker Al, Addy Cl, Et Al 1993 Oral Contraceptive Use And Breast Cancer In Indonesia. *Contraception* 47: 241–9.

[6] Budining S and Cornainll Et Al 1995 Et Al. Epidemiological Analysis Of Risk Factors For Breast Cancer In Indonesian Females. *Risk Factors Breast Cancer*, 4.Journal Medicine Indonesia Vol 4. No.3

[7] Amandito R, Viryawan C, Santoso F, Et Al. The Characteristics Of Breast Cancer Patients In "Dharmais" Hospital National Cancer Center Jakarta Based On Occupational And Environmental Status. *Indones J Cancer*, 7.

[8] Fredholm H, Eaker S, Frisel J, Et Al 2009 Breast Cancer In Young Women: Poor Survival Despite Intensive Treatment. *Plos One* 4: E7695.

[9] Xie F, Yang H, Wang S, Et Al 2012 A Logistic Regression Model For Predicting Axillary Lymph Node Metastases In Early Breast Carcinoma Patients. *Sensors (Basel)* 12: 9936–50.

[10] Xiaobo Z, Kuang-Yu Liuab., Stephen T.C.Wongab 2004 Cancer Classification And Prediction Using Logistic Regression With Bayesian Gene Selection. *J Biomed Inform 37*: 249–259.

[11] Chhatwal J, Alagoz O, Lindstrom Mj, Et Al 2009 A Logistic Regression Model Based On The National Mammography Database Format To Aid Breast Cancer Diagnosis. *Am J Roentgenol* 192: 1117–1127.

[12] American Cancer Society Stages Of Breast Cancerhttps://Www.Cancer.Org/Cancer/Breast-Cancer/Understanding-A-Breast-Cancer-Diagnosis/Stages-Of-Breast-Cancer.Html (Accessed September 28, 2017).

[13] Kuru B, Camlibel M, Dinc S, Et Al 2008 Prognostic Factors For Survival In Breast Cancer Patients Who Developed Distant Metastasis Subsequent To Definitive Surgery. *Singapore Med J* 49: 904–11.

[14] Colleoni M, Viale G, Zahrich D, Et Al. Chemotherapy Is More Effective In Patients With Breast Cancer Not Expressing Steroid Hormone Receptors: A Study Of Preoperative Treatment.

[15] Gagliato D De M, Gonzalez-Angulo Am, Lei X, Et Al 2014 Clinical Impact Of Delaying Initiation Of Adjuvant Chemotherapy In Patients With Breast Cancer. *J Clin Oncol* 32: 735–44.

[16] Paepke S, Kiechle M. [Palliative Therapy Of Breast Cancer]. *Onkologie* 2003; 26 Suppl 7: 4–10.

[17] Hosmer Dw, Lemeshow S, Sturdivant Rx. *Applied Logistic Regression*.

[18] Agresti A 2013 *Categorical Data Analysis*. Wiley-Interscience.

[19] Bender R, Grouven U 1997 Ordinal Logistic Regression In Medical Research. *J R Coll Physicians Lond* 31: 546–51.

[20] Sharma B, Abhimanyu A, Anuradha A, Et Al 2017 Logistic Regression For Breast Cancer Analysis. *Data Min Knowl Eng* 9: 109–113.

[21] Yusuff H, Mohamad N, Ngah Uk, Et Al. Breast Cancer Analysis Using Logistic Regression. *IJRRAS*; 10.