

Analysis of Lethal Factors in Patients with Breast Cancer

Jiayu Kong*

Liaoning Province Shiyen High School, Shenyang, 110000, China

* Corresponding author: kongxiangqian0518@outlook.com

Abstract. Breast cancer is a common malignant tumor in women. The relatively early age for the occurrence of breast cancer is generally 20 years old. The largest number of people are between 45 and 50 years old. Breast cancer is the most significant female killer. China's breast cancer cases occupy 30% of the world's total amount, showing an increasing trend nowadays. To inhibit the increase in the mortality rate of breast cancer, it is particularly urgent to explore the lethal factors of breast cancer. In this paper, the author uses the collinearity test to eliminate the closely related variables in the independent variables. Then, the author uses the binary logic regression method to obtain the relationship coefficient between each factor and breast cancer death. Finally, the author finds that the main factors of breast cancer death are age, T stage, N stage, differentiate, Estrogen Status, and Survival Months. This article provides new ideas for breast cancer prevention and treatment, hopes that breast cancer patients can prolong their lifespan, and hopes to contribute to breast cancer research worldwide.

Keywords: Breast cancer; medical technology; collinearity test; binary logic regression.

1. Introduction

Breast cancer is a neoplasm, which is often found in the epithelial layer of the breast [1]. According to the incidence ranking, its main manifestations are lack of vitality, hair loss, mental stress, sweating, and anxiety. The patients with breast cancer are mainly women, and a small number of men will also appear. In 2022, an estimated approximately 2.309 million new cancer cases occurred globally, ranking second. In 2022, the global cancer mortality rate is estimated to be about 666,000, ranking fourth. The crude mortality rate is 11.3 per 100,000, and the age-standardized mortality rate (ASMR) is 12.6 per 100,000 [2, 3]. According to Global Cancer Epidemiology Statistics (GLOBOCAN), breast cancer is recognized as a seriously widespread carcinoma among Chinese females. Due to the loss of normal cellular characteristics, cancer cells are loosely connected and easily detached. Once cancer cells are shed, they can flow throughout the body with blood and lymphatic fluid, forming a tendency to metastasize and endangering life [4]. Breast cancer has a high incidence rate and mortality. Its symptoms bring great mental pressure and physical injury to patients. With the increase in human medical level and people's attention to physical conditions year by year, the cost of treatment and post-care for breast cancer has gradually increased. Therefore, to curb the deterioration of disease damage and reduce social and economic pressure and psychological burden on personnel, it is particularly important to study the factors that have effect on breast cancer [5].

In the research on the elements impacting breast cancer, many factors have been explored, including induction factors and late treatment factors. Inducing factors include unhealthy lifestyle habits like smoking and drinking, genetic inheritance, and emotional stress [6-8]. The exploration of later influencing factors includes the discovery of real-time fluorescent quantitative PCR (PT-qPCR), which shows that SNHG3 is a potential target for breast cancer diagnosis and heal, and the three monomer compounds of Pu'er tea elements, which were explored through MTT colorimetric test, proteomic test, apoptosis detection and other tests, have inhibitory effects on the proliferation and metastasis of breast cancer cells, thus impeding the continuous deterioration of breast cancer [9, 10]. During the treatment of breast cancer, Liu Haifeng, Liang Lingling, and others conducted a comparative experiment and concluded that positive psychology combined with multimodal exercise intervention can improve patients' psychology and coping style, alleviate fatigue, and improve immune function [11].

This paper is aimed to find out the inducing factors of breast cancer and the influencing factors of post-treatment through data analysis, so as to better inhibit the spread of breast cancer, guide medical treatment, and provide security for human health and, the social economy.

2. Methods

2.1. Data Sources and Description

This dataset of breast cancer patients was sourced from the November 2017 update of the SEER Program of the National Cancer Institute (NCI). The SEER Program offers population-based cancer statistics. The dataset encompassed female patients suffering from infiltrating duct and lobular carcinoma breast cancer (SEER primary site recode NOS histology codes 8522/3) diagnosed between 2006 and 2010. Patients with unknown tumor size, those with examined regional lymph nodes (LNs), positive regional LNs, and patients whose survival months were less than one month were excluded. As a result, a total of 4024 patients were ultimately included.

2.2. Variable Selection

The detailed meaning of the rest continuous variables is listed in Table 1 shows:

Table 1. Different types of variables

Term	Type	Range
Age	Numeric	30 to 69
Race	Categorical	0-“white”, 1-“black”, 2-“other”
Marital Status	Categorical	0-“Married”, 1-“Divorced”, 2-“single”, 3-“widowed”, 4-“Separated”
T Stage	Categorical	0-“T1”, 1-“T2”, 2-“T3”, 3-“T4”
N Stage	Categorical	0-“N1”, 1-“N2”, 2-“N3”
5th Stage	Categorical	0-“IIA”, 1-“IIIA”, 2-“IIB”, 3-“IIIB”, 4-“IIIC”
differentiate	Categorical	0-“Well differentiated”, 1-“Moderately differentiated”, 2-“Poorly differentiated”
Grade	Categorical	1 to 3
A Stage	Categorical	0-“Regional”, 1-“Distant”
Tumor Size	Numeric	1 to 140
Estrogen Status	Categorical	0-“Regional”, 1-“Distant”
Progesterone Status	Categorical	0-“Regional”, 1-“Distant”
0 Node Examined	Numeric	1 to 61
Regional Node 0	Numeric	1 to 46
Survival Months	Numeric	1 to 107

The data in this article includes 1853 samples and 15 independent variables, with no missing data among them. This article has several categorical variables, such as race, marital status, T stage, N stage, 5th Stage, differentiate, grade, A stage, estrogen status, and progesterone status, which are replaced with virtual numbers in the article.

2.3. Method Introduction

The main method used in this article is binary logistic regression. It is an effective classification algorithm that solves the problem of binary variables.

Logistic regression is primarily employed for regression analysis when the dependent variable is of a categorical nature (for instance, yes or no). The independent variable can be either a categorical or a continuous one. It can pick out independent variables that have an influence on the dependent variable from multiple independent variables and supply prediction formulas for making predictions.

When the dependent variable is binary, it is called binomial logistic regression, which is usually used when the explanatory variables are binary quality variables of 0 and 1.

3. Results and Discussion

3.1. Data Preprocrss

Given that binary logistic regression requires the removal of independent variables with strong correlations, this study first used the Spearman correlation method to test the correlation between independent variables. The results are shown in figure 1.

This article initially uses a correlation of 0.5 as the threshold, with a correlation greater than 0.5 indicating a severe relation. The association between the 6th Stage and Nth Stage is 0.58, the relationship between 6th Stage and T Stage is 0.70, and the correlation between Regional Node 0 and N Stage is 0.83. These values are all greater than 0.5, indicating a strong correlation between these independent variables, and further removal of some of them is needed.

Collinearity refers to the high correlation between independent variables, which may lead to unstable estimation results in the regression model, making it difficult to explain and infer. The purpose of the Spearman-related visualizations is to determine the extent of relatedness among independent values, to take corresponding measures, such as keeping variables with weak correlation in the article, conducting variable transformation, etc., to prevent poor fitting and inaccurate regression models.

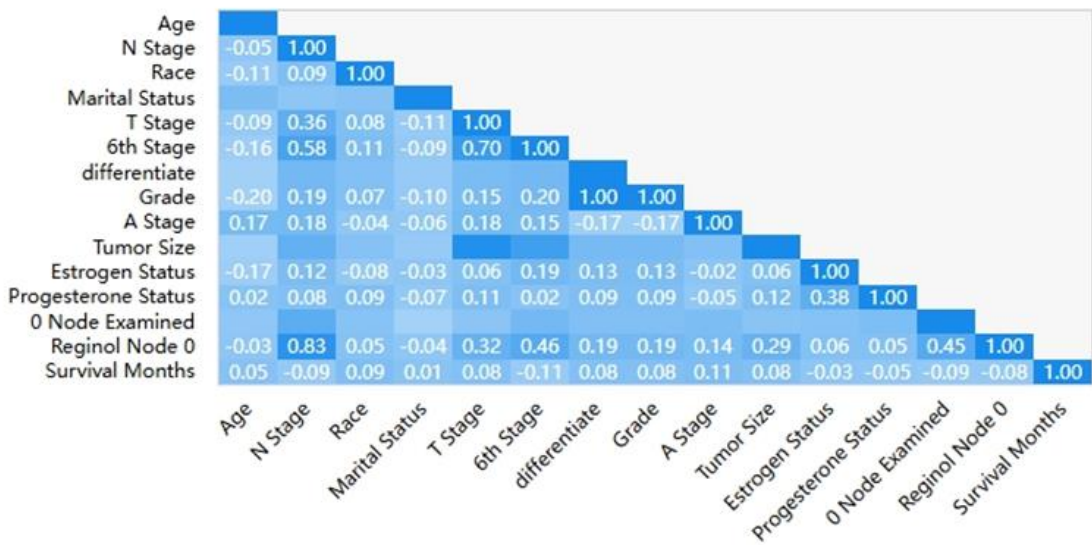


Fig. 1 Spearman-related visualizations

3.2. Logistics Results

Age, Race, Marital Status, T Stage, N Stage, 6th Stage, Different, Grade, A Stage, Tumor Size, Estrogen Status, Progesterone Status, 0 Node Examined, Regional Node 0, and Survival Months serve as explanatory variable, while Status serves as the response variable for binary logistic regression analysis. After automatic recognition by the model, a total of 7 items including Age, T Stage, Grade, Estrogen Status, 0 Node Examined, Regional Node 0, and Survival Months were left in the model. There are 1881 samples in the model. Participate in the analysis and there are no missing data.

After removing grade and differentiation, this article conducted binary logistic regression, and the regression results analysis is shown in the following figure. Firstly, holistic analysis of comprehensive performance of the model is conducted. From the table above, the inaugural presumption for model testing was whether the model quality remains when the independent variables (Age, T Stage, N Stage, Differentiation, Estrogen Status, Survival Months) were included. The p-value here is less than

0.05, indicating rejection of the original supposition. This implies that this article incorporates efficacious independent variables and a model building with strong meaningfulness.

Table 2. Fit the selected variables and coefficients beta

Independent variable	Variable representation	coefficient
Age	β_1	0.035
T Stage	β_2	0.310
N Stage	β_3	0.692
differentiate	β_4	0.451
Estrogen Status	β_5	0.917
Survival Months	β_6	-0.059

From the above table 2, it can be seen that Age, T Stage, N Stage, Differentiation, Estrogen Status, and Survival Months can explain the change in mortality.

$$\ln\left(\frac{p}{1-p}\right) = -1.234 + 0.035\beta_1 + 0.31\beta_2 + 0.692\beta_3 + 0.451\beta_4 + 0.917\beta_5 - 0.059\beta_6 \quad (1)$$

Where p stands for the likelihood of Status being dead and 1-p represents the probability of Status being alive. Summary analysis shows that Age, T Stage, N Stage, Differentiation, and Estrogen Status have a significant positive impact on Status, while Survival Months have a grave negative impact on Status.

Table 3. Evaluation of the model fitting effect

p	AIC	BIC
0.000	1091.648	1130.425

To test the assumption, statistical measures are represented by p-values. This method can calculate the probability that the sample results are manifested or more extreme cases occur when the null hypothesis is true (Table 3). In practical applications, the author usually sets a significance level (such as 0.05). If the p-value of a regression coefficient is less than this significance level, the author considers the regression coefficient statistically significant, that is to say, explanatory variable's variety can largely determine the response variable's situation. The p-value in this article is equal to 0 and less than 0.05, indicating that the age, T stage, N stage, differentiation, Estrogen Status, and Survival Months studied in this article have a significant impact on the mortality rate of the outcome variable. The smaller the numeric magnitude of AIC and BLC, the better. In this article, these two values are relatively large and need to be improved.

4. Conclusion

This article uses collinearity testing to identify 6 independent variables among the 17 original variables that are less correlated with mortality. Among them, age, T stage, N stage, differentiation, and Estrogen Status are positively correlated with mortality, while Survival months are negatively correlated with mortality. By calculating the breast cancer's lethal factor, a new idea is provided for the treatment and prevention of breast cancer. Patients can be stratified by lethal factors to determine which patients belong to high-risk groups, and thus develop more personalized treatment and follow-up plans for them. For high-risk patients, closer monitoring and proactive treatment measures can be taken to promptly identify and address potential issues, thereby improving patient survival rates. Exploring the lethal factors of breast cancer has an important research impact on improving the diagnosis level, promoting treatment innovation, strengthening individual management, and improving public health awareness, which brings new hope for defeating breast cancer.

References

- [1] Yao Y T. Breast cancer Prevention and Community Rehabilitation Nursing, Chinese ethnic folk medicine, 2010, 19: 90.
- [2] Ma L, Wang S P. Study on the Related Symptoms of breast cancer Patients and Their Impact on Quality of Life, Chinese Journal of Evidence Based Medicine, 2007.
- [3] Wu H, Lv Q. Epidemiological trends and prevention and control enlightenment of breast cancer in the world and China: interpretation of the Global Cancer Statistical Report 2018-2022. Chinese Journal of General Surgery Fundamentals and Clinical Medicine, 2024.
- [4] Xie S Z. The harm of breast cancer and its prevention knowledge promotion. Science and Technology, and Innovation, 2018, 11: 6-9.
- [5] Li X Y, Huang Q, Wu Y M, Hu S. Narrow Spectrum Summary and Prospects of the 2022 Global Cancer Statistics Report. Cancer Prevention and Treatment Research, 2024, 5: 307-312.
- [6] Jiang J Q. Beware of life factors inducing breast cancer. Family Medicine, 2001, 6: 9.
- [7] Wang Y D. Risk factors of breast cancer. Chinese Medical Guide, 2008, 18: 104-106.
- [8] Tao F J, Dai Z, Deng Q. Study on the inducing factors and preventive measures of breast cancer. Chinese Science and Technology Journal Database (Full Text Edition) in Medicine and Health, 2020, 10: 56- 58.
- [9] Chen Y X, Yu H M, Liu S Y, Liao D Y, Yu T, Yang S Y, Zeng T, Zhang Y. SNHG3 mediated autophagy promotes the proliferation, migration, and invasion of breast cancer cells. Chin Pharmacol Bull, 2024.
- [10] Yang R, Zhao X P, He M J, Liu M, Luo R, Chen C L, Pan S K, Ding Z G. Study on the effect of puer tea element on four kinds of breast cancer cells. Tea Science, 2024.
- [11] Liu H F, Liang L L, Ma Y Z, Hou Y, Lin B. Application of positive psychology combined with multimodal exercise intervention in patients with breast cancer undergoing chemotherapy. Journal of Nursing, 2024, 11: 28-32.