Mathematical modeling on Breast Cancer: A Case Study

Moharram A. Khan, Murtala Sani

1Department of Mathematics & Computer Science, Umaru Musa Yar’adua University, Katsina, Nigeria
2Department of Education, Yusufu Bala Usman College of Legal and General Studies, Daura, Katsina, Nigeria

ABSTRACT

A predictive mathematical model is presented to analyze whether the cases of breast cancer are increasing or decreasing in Katsina state. Further to investigate the age range in which it is more attacked by the disease and its modes of treatment. The relevant data will be collected from Katsina General Hospital (KGH) and Federal Medical Centre Katsina (FMCK) over eleven years review of the cases. Both historical documentary and analytical methods were used to collect and analyzed the data. The obtained result shows that the number of the cases was increasing in both hospitals which would reach up to 97 by the year 2020, while 40-49 is the age range more attacked by the disease. Surgery is the common treatment given to the patients in both hospitals and then referred them to either Aminu Kano or Zaria teaching hospitals for chemotherapy and radiotherapy treatments.

Keywords: Mathematical modeling, breast cancer, chemotherapy, radiotherapy, surgery.

1. INTRODUCTION

Breast Cancer is a malignant tumor that started in the cells of the breast. A malignant tumor is a group of cancer cells that grow into (invasive) surrounding tissues or (metastasize) to distant area of the body. The disease occurs almost in women, but men can get it too [5]. The common belief among Nigerians is that breast cancer is only a disease of women and it cannot affect men [3]. A report from Teaching Hospital Ilorin, there was 568 new cases of breast cancer registered at the cancer registry during the 10-year period, comprising of 7(12.23%) males and 561(98.77%) females [4].

Breast cancer is the commonest cancer in Nigerian women. As a result of poverty and ignorance among other factors, some patients present late usually with locally advanced disease [9], the factors responsible for late presentation with advanced disease include low social economic level, fear of mastectomy and poverty [8]. There is a low level of awareness of mammography and mammographic screening in Nigeria. Public education and awareness programs should be developed to promote early detection and diagnosis in the prevention of breast cancer in women in Nigeria and other developing countries [6]. Problems associated with breast cancer in Africa include poor knowledge of risk factors, early detection measures and early warning signs of breast cancer [13], the three major early detection measures for breast cancer include breast self-examination (BSE), clinical breast examination and mammography [1]. Breast cancer is the most common malignancy affecting the women; it can be detected early by breast self-examination [7]. BSE remains an important investigation tool for the early diagnosis of breast cancer in our environment due to lack of facilities and screening programs [14].

In order to improve women’s health and breast cancer prognosis, providing equal educational opportunities for women seems necessary and more effort should be made to reach out to the low literacy groups in our society, to create awareness and the know-how in carrying out these screening exercises [12]. Mass media and interpersonal networks of communication are essential in health interventions for breast cancer and other health related issues [10]. Early recognition through diagnosis of metastatic disease will allow rapid institution of effective palliative treatment [11]. Breast cancer treatment was based on the tumor stage and patient’s performance status [2]. Treatment may also involve surgery, radiation therapy, chemotherapy, hormone therapy, and targeted therapy. The goals of treatment are to “cure” the cancer if possible and/or prolong survival and provide the highest possible quality of life during and after treatment [5].

2. METHODOLOGY

Both historical documentary and analysis methods were used to collect the data from the two hospitals; Katsina General Hospital (KGH) and Federal Medical Centre Katsina (FMCK) and analyzed using the model developed. The model is formulated using regression analysis, which is a method that deals with the formulation of mathematical models that gives relationship among variables. It developed to make prediction and other analysis by using the obtained data for the research work.

2.1 Formulation of the Model

2.1.1 Basic Assumptions

Breast cancer cases are either increasing or decreasing but not all people are at risk of being attacked by the disease in Katsina state.

For the model: We consider the simple linear regression that concerned with the distribution of two variables: cases and time or period \((t_i, C_i)\) which can be presented graphically on a scatter diagram. To form a scatter points \((t_1, C_1), (t_2, C_2), (t_3, C_3), \ldots, (t_n, C_n)\) and plotted on conventional graph with \(C\) values on vertical axis and \(t\) value on the horizontal axis.
The linear relationship hold for the total number of cases \( C_i \) and time or period \( t_i \) where \( i = 1, 2, 3, \ldots, n \).

We could obtain the numerical values of \( \beta_0 \) and \( \beta_1 \) only if we could have all the possible values of these variables. Now, the true relation between \( C_i \) and \( t_i \) is

\[ C_i = \beta_0 + \beta_1 t_i + k_i t_i = \beta_0 + \beta_1 t_i + \mu \]

Where,
- \( C_i \) = dependent variable
- \( t_i \) = independent variable
- \( \beta_0 \) = C- intercept
- \( \beta_1 \) = slope of the line
- \( k_i \) = probability factor

\( \mu \) = random error in \( C \) for observation \( i; i = 1, 2, 3, \ldots, n \), which depends on nature of the data \( (C_i) \) do to so many factors \( (k_i) \), such as mobilization campaign on the danger of the disease in public or through medias, educational factor, economical factor, believes and social stigma of the people in Katsina which may increase or decrease the number of cases.

The true regression line is

\[ C_i = \beta_0 + \beta_1 t_i \]

The estimate relationship is

\[ \hat{C}_i = \hat{\beta}_0 + \hat{\beta}_1 t_i + k_i t_i \]

While the estimate regression line is

\[ \hat{C}_i = \hat{\beta}_0 + \hat{\beta}_1 t_i \]

Where,
- \( \hat{C}_i \) = estimated value of \( C_i \) given a specified value of \( t_i \).
- \( \hat{\beta}_0 \) = estimate of the true intercept \( \beta_0 \)
- \( \hat{\beta}_1 \) = estimate of the true slope \( \beta_1 \)
- \( k_i \) = estimate of the random error \( \mu \)

Consider the diagram below:

The true and estimated regression lines

\[ \hat{C}_i = \hat{\beta}_0 + \hat{\beta}_1 t_i \]

\[ C_i = \beta_0 + \beta_1 t_i \]

**Figure 1**

Expressing the \( (e_i) \) in terms of the observed values \( C_i \) and estimated values \( \hat{C}_i \), we have

\[ k_i e_i = (C_i - \hat{C}_i) \]

Where \( k_i = 1 \), we have

\[ e_i = (C_i - \hat{\beta}_0 - \hat{\beta}_1 t_i) \]

But \( \sum_{i=1}^{n} e_i^2 = 0 \) by definition of (least square model), the best solution is to take summation and square to both side of equation (5) above and equate it to zero.

\[ \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (C_i - \hat{\beta}_0 - \hat{\beta}_1 t_i)^2 = 0 \]

Differentiating equation (6), partially, with respect to \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \), we get

\[ -2 \sum_{i=1}^{n} (C_i - \hat{\beta}_0 - \hat{\beta}_1 t_i) = 0 \]

\[ -2 t_i \sum_{i=1}^{n} (C_i - \hat{\beta}_0 - \hat{\beta}_1 t_i) = 0 \]

Simplifying equation (7) and (8), we get

\[ \sum_{i=1}^{n} (C_i - \hat{\beta}_0 - \hat{\beta}_1 t_i) = 0 \]

\[ t_i \sum_{i=1}^{n} (C_i - \hat{\beta}_0 - \hat{\beta}_1 t_i) = 0 \]

To expand equation (9) and (10), we have

\[ \sum_{i=1}^{n} C_i = n \beta_0 + \beta_1 \sum_{i=1}^{n} t_i \]

\[ \sum_{i=1}^{n} C_i t_i = \hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} t_i^2 \]

From equation (11) above, we get

\[ \beta_0 = \bar{C} - \beta_1 \bar{t} \]

We first multiply equation (11) by \( \sum_{i=1}^{n} t_i \), (12) by \( n \) and then, subtract it to get

\[ \beta_1 = \frac{n \sum_{i=1}^{n} C_i t_i - \sum_{i=1}^{n} C_i \sum_{i=1}^{n} t_i}{n \sum_{i=1}^{n} t_i^2 - (\sum_{i=1}^{n} t_i)^2} \]
Having found the relationship of \( t \) and \( C \), we can use it to predict \( \hat{C}_t \) using the model given by equation (4). The prediction can be done by substituting the given value of \( t_i \) in the model equation and calculate the value of \( \hat{C}_t \).

However, to investigate whether the cases of breast cancer are increasing or decreasing in Katsina state; we also consider the model equation (4) above, for all \( \hat{C}_i \) where \( i = 1,2,3,\ldots,n \). One of the following conditions must hold:

- \( \hat{C}_1 < \hat{C}_2 < \hat{C}_3 < \hat{C}_n = I(\ast) \)
- \( \hat{C}_1 > \hat{C}_2 > \hat{C}_3 > \hat{C}_n = D(\ast\ast) \)
- \( \hat{C}_1 = \hat{C}_2 = \hat{C}_3 = \hat{C}_n = C(\ast\ast\ast) \)

For the above three basic conditions:

- \( \ast \) the slope of the graph line will rise upward (as time increases, the number of cases increases)
- \( \ast\ast \) the slope of the graph line will come downward (as time increases, the number of cases decreases)
- \( \ast\ast\ast \) the slope of the graph line will remain unchanged do to time increase (no regression).

To test the reliability (applicability) of the model to the collected data for the research work using F test, we consider equation (6) above, the sum of square error (SSE) is given by

\[
SS_E = \sum_{i=1}^{n}(C_i - \hat{C}_i)^2
\]

But the sum of square regression (SSR) is

\[
SS_R = \sum_{i=1}^{n}(\hat{C}_i - \bar{C})^2
\]

And their total (SST) is

\[
SS_T = \sum_{i=1}^{n}(C_i - \bar{C})^2
\]

From equation (16) mean of square regression is given by

\[
MS_R = \frac{\sum_{i=1}^{n}(\hat{C}_i - \bar{C})^2}{1}
\]

Also from equation (17) the mean of square error is given by

\[
MS_E = \frac{\sum_{i=1}^{n}(C_i - \hat{C}_i)^2}{(n - 2)}
\]

Where \( n \) is the number of periods.

We find both calculated and tabulated values of \( F \) for the test as:

\[
F_{cal} = \frac{\sum_{i=1}^{n}(\hat{C}_i - \bar{C})^2}{\sum_{i=1}^{n}(C_i - \hat{C}_i)^2}
\]

And then

\[
F_{tab} = F_{a,k-1,(n-2)}
\]

Where \( a \) is the significance level of the test and \( k \) is the number of variables

The model is said to be applicable to the collected data for the research work if and only if equation (20) is greater than (21) above which means the relationship exist among the variables (\( C_i \) and \( t_i \)).

To measure the adequacy of the model, we have to measure the amount of variability in the data accounted for by the model using coefficient of determination (\( R^2 \)) which is given by:

\[
R^2 = \frac{\sum_{i=1}^{n}(\hat{C}_i - \bar{C})^2}{\sum_{i=1}^{n}(C_i - \bar{C})^2}
\]

Equation (22) can take values between 0 and 1 of the variability in the yield data which should be converted to percent in order to indicate the fitness of the model.

2.1.2 Parameters

- The C-intercept (\( \beta_0 \)): This is the value at a point where the regression line cut the C axis, the value is 45.27.
- The slope (\( \beta_1 \)): It is the steepness of the line with value 4.72
- The probability factor (k): Picking any of those factors, its probability of increasing the expected outcome of the number of cases is 0.5 or \( \frac{1}{2} \) which is similar to that of decrease.

3. RESULT

Table 1: Showing the total and percentage of different diagnosis of the patients from both hospitals

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>FMCK</th>
<th>KGH</th>
<th>Total</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced</td>
<td>154</td>
<td>86</td>
<td>240</td>
<td>48.19%</td>
</tr>
<tr>
<td>Tumor</td>
<td>103</td>
<td>28</td>
<td>131</td>
<td>26.31%</td>
</tr>
<tr>
<td>Malignant</td>
<td>80</td>
<td>47</td>
<td>127</td>
<td>25.50%</td>
</tr>
<tr>
<td>Total</td>
<td>337</td>
<td>161</td>
<td>498</td>
<td>100%</td>
</tr>
</tbody>
</table>

From the above general diagnostic table for both hospitals, advanced breast cancer constituted 240 (48.19%) cases, followed by breast cancer tumor with 131 (26.31%), and malignant breast cancer 127 (25.50%) were presented.
Four hundred and twenty eight cases were studied; the minimum and maximum ages at presentation were 15 and 90 years respectively. The percentage of the age ranges of the population studied at presentation is as shown in the above table 13. The highest age range at presentation was 40-49 years constituted (21.89%), followed by 60-69 years age range (17.87%), 30-39 years age range (17.27%), 20-29 years age range (16.87%), 50-59 years age range (13.25%), 70-79 years age range (7.63%), 10-19 years age range (3.21%), and 80+ years age range (2.01%) constituted the group with the least incidence at presentation.

Table 4: Showing both Federal Medical Center and Katsina General Hospital prediction table

<table>
<thead>
<tr>
<th>S/N</th>
<th>C_i</th>
<th>t_i</th>
<th>t_i^2</th>
<th>C_i × t_i</th>
<th>ĉ_i</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>-5</td>
<td>25</td>
<td>-125</td>
<td>21.68182</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>-4</td>
<td>16</td>
<td>-96</td>
<td>26.4</td>
</tr>
<tr>
<td>3</td>
<td>31</td>
<td>-3</td>
<td>9</td>
<td>-93</td>
<td>31.11818</td>
</tr>
<tr>
<td>4</td>
<td>34</td>
<td>-2</td>
<td>4</td>
<td>-68</td>
<td>35.83637</td>
</tr>
</tbody>
</table>
On gender presentation of the cases which was presented graphically in figure 3 indicated that, based on the percentage there were rear males cases and their risk of been attacked by the disease were increasing in the study area as in the case of other parts of the world, despite their common belief that breast cancer is only a disease of women and it cannot affect men.

For the total and percentage distribution of patients based on age ranges as shown in figure 4 and 5, mostly the disease attacked the working class women of the area which affected their growth and development in different aspect of live, especially family care and economic well been.

In general in Nigeria, particularly in Katsina state, in case of treatment of the disease, both hospitals were given surgery to their patients and referred them to Aminu Kano or Zaria teaching hospitals for chemotherapy and radiotherapy treatments, but in some cases FMC give them part of chemotherapy depending on the stage of the disease. Lastly, the predicted number of the cases by the year 2020 using the model was 97, showing that the cases were increasing as graphically presented in figure 6.

7. CONCLUSION

The number of breast cancer cases in the study area was increasing and mostly attacked the working class women with the habit of late clinical presentation of the disease. Both hospitals were conducting surgery to their patients and referred them to Zaria or Kano teaching hospitals for the remaining part of the treatments.

ACKNOWLEDGEMENT

This work is the part of M.Sc thesis of Murtala Sani, UMYU Katsina (Session 2013/2014) under the supervision of Prof. Moharram Ali Khan

REFERENCES


\[ R^2 = 0.99 = 99\% \]

Hence, indicated a very good fit of the model.

5. MODEL ADEQUACY

For the adequacy of the model, we have:

\[ R^2 = 0.99 = 99\% \]

Hence, indicated a very good fit of the model.

6. DISCUSSION

Based on the total and percentage of different diagnosis of the patients from both hospitals as graphically presented in figure 2, most of the patients presented the cases very late considering the percentage taken by advanced and malignant breast cancer types compared to early presentation of tumor as in the cases of other developed countries and areas, which was among the factors responsible for increasing the mortality rate of the patients in the study area and worldwide.


