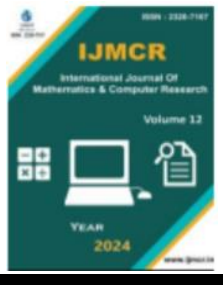
International Journal of Mathematics and Computer Research

ISSN: 2320-7167

Volume 12 Issue 03 March 2024, Page no. – 4107-4113

Index Copernicus ICV: 57.55, Impact Factor: 8.316

DOI: 10.47191/ijmcr/v12i3.05



Predictive Mathematical Model on Breast Cancer: A Study

Rajesh Shrivastava[1], Deepika Basedia[2], Keerty Shrivastava[3]

[1,2]Dr. Shyama Prasad Mukherjee Science & Commerce College, Bhopal 462022, Madhya Pradesh, India

[3]Govt. BHEL Colleges, Bhopal 462022, Madhya Pradesh, India

ARTICLE INFO

Published Online:
27 March 2024

Corresponding Author:
Deepika Basedia

ABSTRACT

In the present study, we have designed a mathematical model to analyze whether the cases of breast cancer are maximized or minimized in Madhya Pradesh. Especially to check the age range in which it's more susceptible to the disease and its means of therapy. The important data collected from Jawaharlal Nehru Cancer Hospital, Bhopal (JLNCH) and Gandhi Medical College, Bhopal (GMC) is from over ten years of reviews of the cases. Actual documentary and analytical methods were used to collect and analyze the data. It is concluded from the results that the number of cancer cases is increasing in both hospitals; its projection may reach up to 97.8% by the year 2023; the age range of 40–50 is more vulnerable to the disease. The line of treatment for breast cancer patients is surgery, chemotherapy, and radiotherapy in both hospitals.

KEYWORDS: Mathematical modeling, Breast cancer, Cancer cell, Tumor, Surgery.

I. INTRODUCTION

The malignant tumor that attacks the cells of the breast tissues results in breast cancer. A malignant tumor is a group of cancer cells that grow into adjoining tissues or permeate a remote area of the body. The disease occurs in women, but men can become vulnerable to it [1]. A survey of breast cancer cases registry designed for the collection, storage, and management of data on persons for the ten-year period contained 2.20% of males and 97.8% of females.

Women are mostly common prey of breast cancer due to poverty and unawareness and other reasons, some patients notice late which results in advanced phase of disease [1], the factors responsible for late diagnosis with advanced infection include penniless poverty stricken under privileged level, fear of surgery pain and poverty [2]. There is a low level of awareness of diagnostic mammography and screening in Madhya Pradesh. In developing countries p ublic education and awareness programs are to be developedto promote early identification and diagnosis in the preventionof breast cancer in women [2]. Problems created breast cancer in India include rare knowledge of risk factors, early diagnosis measures and early symptoms of breastcancer [2], the three major early diagnostic measures for breast cancer include breast self-examination (BSE), clinical breast examination and mammography [2]. Breastcancer is a common malignant neoplasm affecting the women; it can be diagnosed early by breast self- examination [3]. BSE remains an important investigation tool for the early diagnosis of breast cancer in our environment [3].

For improving women's health and breast cancer prediction equal educational opportunities for women seems essential, low literacy sections should be identified and educational resources may be made available to create knowledge for screening exposure [3]. Interactive media and interpersonal communication are important in health debates and knowhow for breast cancer and other health related factors [4]. Timely diagnosis a disease will result in rapid institution of effective supportive medication [4]. The tumor stage and patients response modulates the treatment [5]. Treatment involves surgery, radiation therapy, chemotherapy, hormone therapy, and targeted novel drug therapy. The aim of treatment is to "cure" the cancer if possible and/or prolong survival withthe better possible quality of life during and after treatment [7].

II. METHODOLOGY

Both traditional documentary and analysis methods are used to collect the data from two hospitals; Jawaharlal Nehru Cancer Hospital, Bhopal (JLNCH) and Gandhi Medical College, Bhopal (GMC) and assessed using the model developed. The model is formulated using regression analysis, a method that deals with the formulation of mathematical models which gives relationship among

variables. In present research work we developed mathematical model for predictions of breast cancer cases in both hospitals and interpreted by using the obtained data.

III. MATHEMATICAL MODEL FORMULATION

The cases of breast cancer may either increase or decrease but all people are not at risk of being attacked by the disease in MP.

Mathematical models are developing by used the linear regression that is concerned with the distribution of variable Time (T_i) and Number of cases (N_i) which are represented by scattered graph.

To form scatter points (T_1, N_1), (T_2, N_2), (T_3, N_3), ..., (T_n, N_n) and plotted graph taking number of cases N on Y- axis and time T on X- axis.

The linear relationship holds for the total number of cases (N_i) and time (T_i) where i = 1, 2, 3, …., n. We may obtain the numerical values of α_0 and α_1 only if we have all the possible values of these variables.

Now, the true relation between N_i and T_i is

$$N_i = \alpha_0 + \alpha_1 T_i + p_i T_i \quad (1)$$

Where,

N_i = dependent variable

T_i = independent variable

α_0 = C- intercept

α_1 = slope of the line

p_i = probability factor

μ = random error in C for observation i; i = 1, 2, 3,…, n., which depends on nature of the data (C_i) do to several factors (k_i), such as mobilization campaign on the danger of the disease in public or through medias, educational factor, economical factor, beliefs and social stigma of the people in Madhya Pradesh which may result in increase or decrease of the number of cases.

The true regression line is

$$N_i = \alpha_0 + \alpha_1 T_i \quad (2)$$

The estimate relationship is

$$\widehat{N}_i = \widehat{\alpha_0} + \widehat{\alpha_1} T_i + p_i k_i \quad (3)$$

While the estimate regression line is

$$\widehat{N}_i = \widehat{\alpha_0} + \widehat{\alpha_1} T_i \quad (4)$$

Where,

$\widehat{N}_i$= Estimated value of N_i given a specified value of T.

$\widehat{\alpha_0}$ =Estimate of the true intercept α_0

$\widehat{\alpha_1}$= Estimate of the true slope α_1

k_i =Estimate of the random error μ

Consider the diagram below:

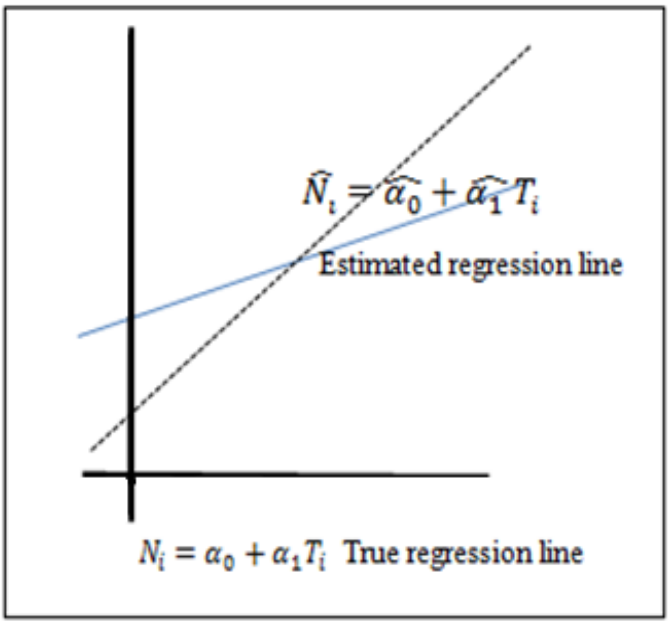


Figure 1

The true and estimated regression lines

Expressing the (e_{is}) in terms of the observed values N_i and estimated values $\widehat{N}_i$ we have

$k_i e_i = (N_i - \widehat{N}_i)$

Where p_i = 1, we have

$$e_i = (N_i - \widehat{\alpha_0} + \widehat{\alpha_1} T_i) \quad (5)$$

But $\sum_{i=1}^{n} e_i$ 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.

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (N_i - \widehat{\alpha_0} + \widehat{\alpha_1} T_i)^2 = 0 \quad (6)$$

Differentiating equation (6), partially, with respect to $\widehat{\alpha_0}$ and $\widehat{\alpha_1}$ we get,

$$-2 \sum_{i=1}^{n} (N_i - \widehat{\alpha_0} - \widehat{\alpha_1} T_i) = 0 \quad (7)$$

$$-2 t_i \sum_{i=1}^{n} (N_i - \widehat{\alpha_0} - \widehat{\alpha_1} T_i) = 0 \quad (8)$$

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

$$\sum_{i=1}^{n} (N_i - \widehat{\alpha_0} - \widehat{\alpha_1} T_i) = 0 \quad (9)$$

$$t_i \sum_{i=1}^{n} (N_i - \widehat{\alpha_0} - \widehat{\alpha_1} T_i) = 0 \quad (10)$$

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

$\sum_{i=1}^{n} N_i = n\widehat{\alpha_0} - \widehat{\alpha_1}\sum_{i=1}^{n} T_i$
(11)

$\sum_{i=1}^{n} N_i T_i = \widehat{\alpha_0} + \widehat{\alpha_1}\sum_{i=1}^{n} T_i^2$ (12)

From equation (11) above, we get

$\widehat{\alpha_0} = N - \widehat{\alpha_1} T$
(13)

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

$$\widehat{\alpha_1} = \frac{n\sum_{i=1}^{n} N_i T_i - \sum_{i=1}^{n} N_i \sum_{i=1}^{n} T_i}{\left[n\sum_{i=1}^{n} T_i^2 - (\sum_{i=1}^{n} T)^2\right]} \quad (14)$$

Having found the relationship of T and N, we canuse it to predict $\widehat{N_t}$ using the model given by equation (4). The prediction can be done by substituting the given valueof T_i in the model equation and calculate the value of $\widehat{N_t}$. However, to investigate whether the cases of breast cancer are increasing or decreasing in MP;we shall consider the model equation (4) above, for all $\widehat{N_t}$ where i = 1, 2, 3,…, n. One of the following conditions must hold.

$\widehat{N_1} < \widehat{N_2} < \widehat{N_3} < \widehat{N_n} = \text{Cases Increase}$

The slope of the graph line will rise upward (as time increases, the number of cases increases)

$\widehat{N_1} > \widehat{N_2} > \widehat{N_3} > \widehat{N_n} = \text{Cases Decrease}$

The slope of the graph line will come downward (as time increases, the number of cases decreases)

$\widehat{N_1} = \widehat{N_2} = \widehat{N_3} = \widehat{N_n} = \text{Cases Unchanged(Constant)}$

The slope of the graph line will remains parallel to the horizontal axis (T), which means the number of cases remains unchanged due to time increase (no regression).

II.

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 Addition of square error $(AS_e)$is given by

$AS_E = \sum_{i=1}^{n}(N_i - \widehat{N_i})^2$ (15)

But the addition of square regression (AS_R) is

$AS_R = \sum_{i=1}^{n}\left((\widehat{N_t} - \bar{N_t})\right)^2$ (16)

And their total AS_T is

$AS_T = \sum_{i=1}^{n}(N_i - \bar{N})^2$ (17)

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

$mS_R = \sum_{i=1}^{n}\frac{(\widehat{N_i} - \bar{N})^2}{(n-1)}$ (18)

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

$mS_E = \sum_{i=1}^{n}\frac{(N_i - \widehat{N_t})^2}{(n-2)}$ (19)

Where n is the number of periods.

We find both calculated and tabulated values of for the test as:

$F_{cal} = \frac{\sum_{i=1}^{n}(\widehat{N_t} - \bar{N_t})^2}{\sum_{i=1}^{n}(N_i - \widehat{N_t})^2}$ (20)

And then

$F_{tab} = F_{\alpha,k,(n-2)}$ (21)

Where α is the significance level of the test and k isthe number of variables

The model is said to be applicable to the collected data for the research work if the value of equation (20) is greater than equation (21) above which means the relationship exist among the variables ($N_i$and T_i).

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

$R^2 = \frac{\sum_{i=1}^{n}(\widehat{N_t} - \bar{N_t})^2}{\sum_{i=1}^{n}(N_i - \bar{N_t})^2}$ (22)

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

Parameters

The N-intercept (αo): This is the value at a point where the regression line cut the Y- axis, the value is 58.27.The slope (α_1): It is the steepness of the line with value 4.718.

The probability factor (p): Picking any of those factors, its probability of increasing the expected outcome of the number of cases is 0.5 which is similar to that of decrease.

RESULT

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

Diagnosis	JLNCH	GMC	Total	Percent
Advanced	186	111	297	48.94%
Tumor	98	42	140	23.02%
Malignant	84	87	171	28.13%
Total	368	240	608	100%

From the above general diagnostic table for both hospitals, advanced breast cancer constituted 297(48.94%) cases, Followed by breast cancer tumor with 140(23.02%), and malignant breast cancer 171 (28.13%) were presented.

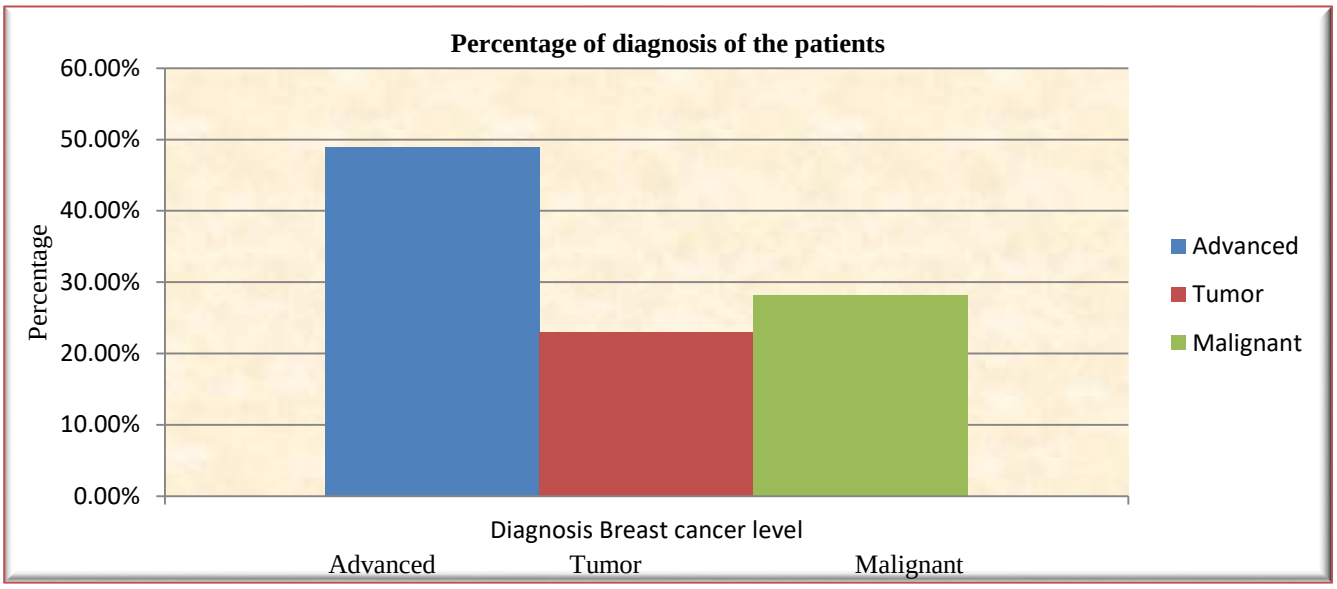


Figure 2: The graphical presentation of table1 showing the total number and percentage of different diagnosis of the patients from both hospitals.

Table 2: Showing the total number and percentage of patients based on gender from both hospitals

Gender	JLNCH	GMC	Total	Percent (%)
Male	29	14	26	7.07%
Female	416	149	565	92.93%
Total	445	163	608	100 %

From the above gender table, out of the total number of breast cancer patients from both hospitals; 565(92.93%) were women while only 26 (7.07%) were men.

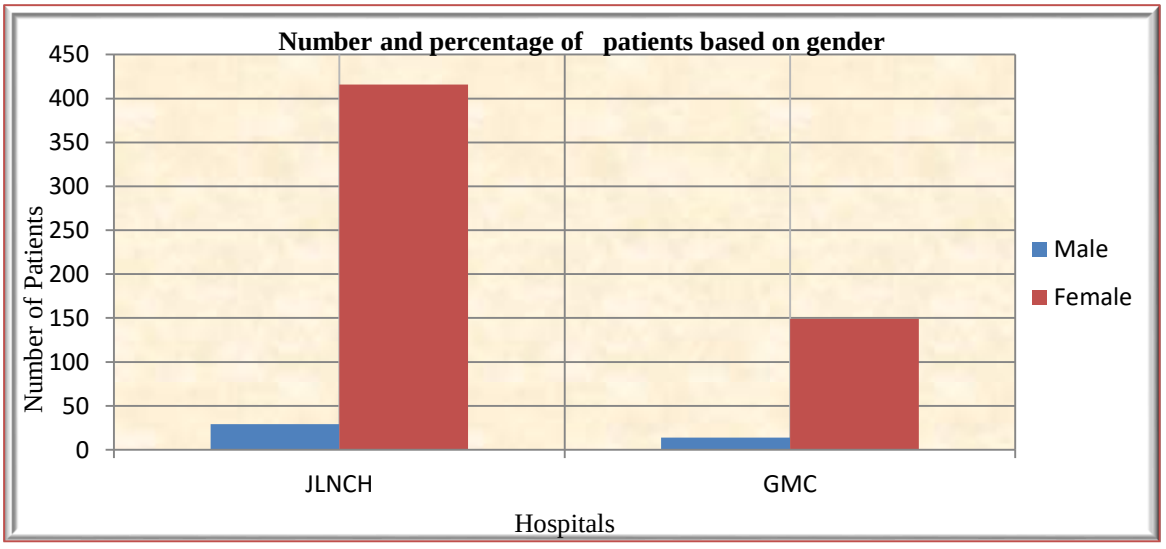


Figure 3: The graphical presentation of table 2 showingthe total number and percentage of patients based on gender from both (JLNCH & GMC).

Table 3: Showing the percentage distribution of patients based on age ranges in both hospitals

Age Range	JLNCH	GMC	Total	Percent
10 – 19	9	4	13	2.14%

20 – 29	38	12	50	8.22%
30 – 39	82	22	104	17.10%
40 – 49	228	31	259	42.60%
50 – 59	65	13	78	12.83%
60 – 69	59	8	67	11.02%
70 – 79	33	4	37	6.08%
80 +	3	0	3	.50%

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 3. The highest age range at presentation was 40-49 years constituted (42.60%),followed by 30-39 years age range (17.10%), 50-59 years age range (12.83%), 60-69 years age range (11.02%), 20-29 years age range (8.22%), 70-79 years age range (6.08%), 10-19 years age range (2.14%), and 80+ years age range (.50%) constituted the group with the least incidence at presentation.

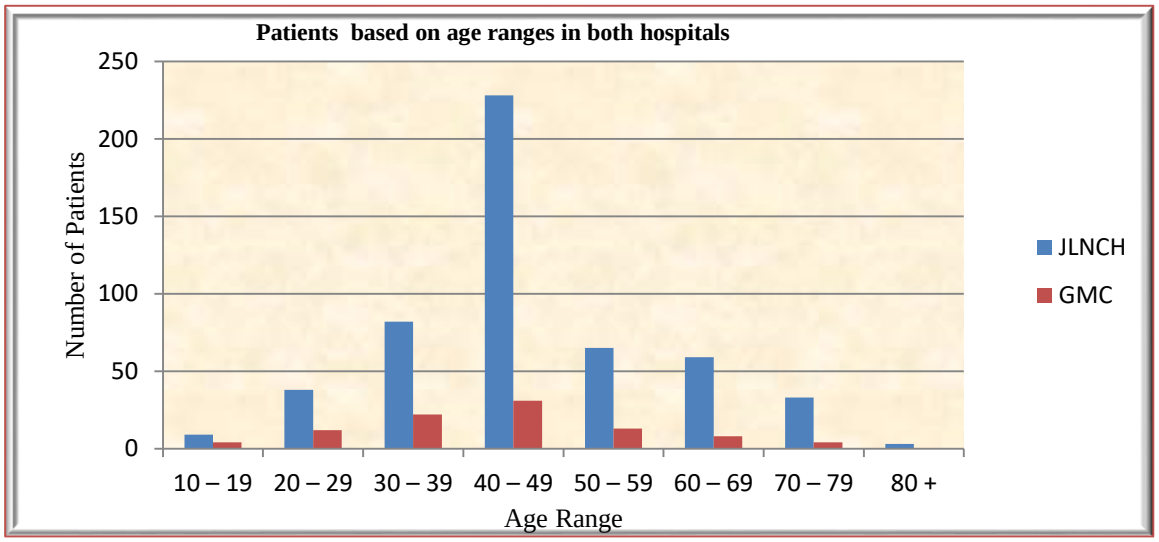


Figure 4: The graphical presentation of table 3 showingthe total distribution of patients based on age ranges in both hospitals (JLNCH & GMC).

Table 4: Showing both Jawaharlal Nehru Cancer Hospital, Bhopal (JLNCH) and Gandhi Medical College, Bhopal (GMC) prediction table

S.No.	YEAR	N_i	T_i	T_i^2	$N_i \times T_i$	$\overline{N_i}$
1	2012-13	39	-5	25	-175	31.6795
2	2013-14	38	-4	16	-136	36.3976
3	2014-15	44	-3	9	-123	41.1157
4	2015-16	48	-2	4	-88	45.8338
5	2016-17	56	-1	1	-52	50.5519
6	2017-18	59	0	0	0	55.27
7	2018-19	63	1	1	59	59.9881
8	2019-20	65	2	4	130	64.7062
9	2020-21	68	3	9	204	69.4243

10	2021-22	75	4	16	300	74.1424
11	2022-23	80	5	25	400	78.8605
Total		**498**	**0**	**110**	**519**	**607.97**

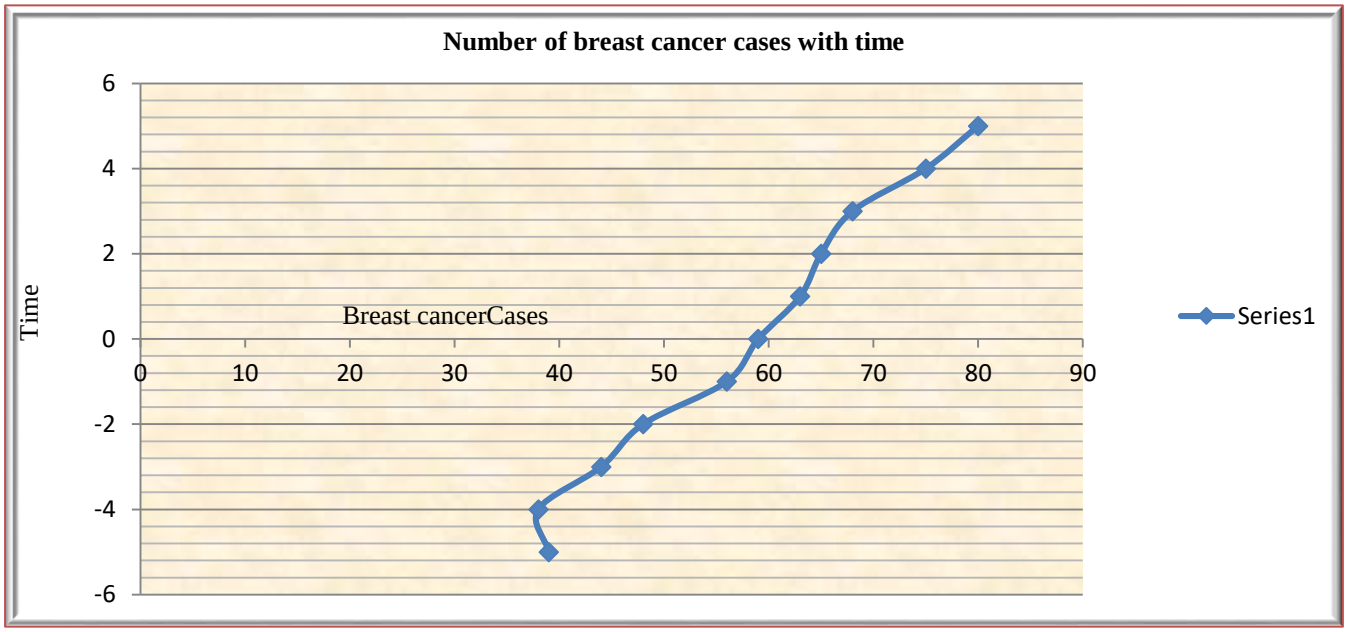


Figure 5: Showing the regression line of the breast cancer cases in both hospitals from table 4 above.

From the model given by equation (4) the predicted number of breast cancer cases in both hospitals by the year 2023 is 92.93% in females.

MODEL RELIABILITY

To test the reliability of the model: The calculated value of F = 190.11, and the tabulated value of F = 189.4.

Since, F_{cal} *and* F_{tab} is approximately same then the model is reliable or applicable to the collected data for the research work which means the relationship exist among the variables (N_i and T_i).

MODEL ACCURACY **VIII.**

For the accuracy of the model, we have: R^2 = 0.99 = 99% Hence presented a good standard model.

VII. DISCUSSION

It is apparent from the total diagnosis percentage of the patients from both hospitals that the majority of the patients appeared late for diagnosis. This analysis is supported by the fact that the percentage of advanced and malignant breast cancer cases is higher when compared to the patient's diagnosis of tumors in developed countries. Delayed contact for diagnosis results in a higher mortality rate for the patients. The graphical presentation in Figure 3 clearly indicates that the males appeared for diagnosis and were rarely found to be infected, and their risk of disease was observed to be increasing in cases in other parts of the world, against the general belief that breast cancer is a woman's disease. In the study, based on the age range of the patients shown in Figure 4 and the working groups presented in Figure 5, it was observed that the working class of women is more vulnerable to disease attachment and becomes the mode of growth and development in different spheres of life. This affected family upbringing and economic well-being.

As a whole, in India, especially in Madhya Pradesh, the generally common way of treating the disease is surgery, which is later suggested for periodic chemotherapy and radiotherapy depending on the severity of the infection. Figure 6 graphically indicates that the number of breast cancer cases may increase in 2023 with the use of the model.

CONCLUSION

The study shows the increasing number of cases of breast cancer in Madhya Pradesh by both hospitals. The working class women are more susceptible. The patients appear late for diagnosis both hospital conducted surgery and directed patients for further required therapy.

REFERENCES

1. Key Statistics for Breast Cancer 2024 Available online:https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html
2. Basic Information about Breast Cancer 2023 Center for Disease Control and Prevention. Available online: https://www.cdc.gov/cancer/breast/basicinfo/index.htm
3. Breast Cancer 2023 Available online:

https://www.who.int/news-room/fact-sheets/detail/breast-cancer

4. Albeshan, S.M., Alashban, Y.I.(2021)Incidence trends of breast cancer in India: A joinpoint regression analysis ,J. King Saud Univ. -Sci. , 33,
5. American Cancer Society 2014 Cancer Treatment & Survivorship Facts & Figures 2014 American Cancer Society Inc.1- 43.
6. Anantha L. 2014 Awareness and Practice of Breast Self-Examination among Women in South India. Int. J. Curr. Microbiol. App. Sci. 3, 391-394
7. Amoran O. 2014 Breast Cancer Screening Awareness and Practice among Women in Sagamu Local Government Area, South-Western Nigeria. British Journal of Applied Science &Technology. 4, 2320-2332
8. Okorie N. 2014 Effective Use of Information Sources for Breast Cancer Care. 101-578. De Santis C.E. A. 2015 International variation in female breast cancer incidence and mortality rates international variation in female breast cancer rates. Cancer Epidemiol. Biomark. Prev., 24, 1495–150
9. Tesfay H. 2014 Knowledge of breast cancer and its early detection measures among female students, in Mekelle University, Tigray region, Ethiopia. Science Journal of Clinical Medicine.3, 57-64.
10. Khokhar A.2012 Breast cancer in India: where we do stand and where do we go APJCP, 13(10), 4861-4866.
11. Elumelu, T. 2011 Pattern of breast cancer metastasis at the radiotherapy clinic. Rev. J. Am. Sci., 7,7.
12. Matsuno R.K. 2011 projection individualized absolute invasive breast cancer risk in Asian and pacific islander American women, Cancer Inst., 103(12) 951-961.