Pak J Med Res
Vol. 57, No. 2, 2018

Original Article

Risk Assessment Model of Rural and Urban Breast Cancer Patients of Khyber Pakhtunkhwa Province, Pakistan

Salahuddin[1]*, Arifullah[2], Najma[3], Mehreen[4]
Institute of Management and Information Sciences, CECOS University of IT and Emerging Sciences[1], Hayatabad, Higher Education Department[2], Department of Statistics, Shaheed Benazir Bhutto Women University[3], Hayatabad Medical Complex[4], Hayatabad, Peshawar.

Present Address: *Department of Family and Community Medicine, College of Medicine, University of Dammam, Kingdom of Saudi Arabia.

Abstract

Background: Adult females are usually affected by the most common category of cancer known as the cancer of the breast.
Objectives: To assess various risk factors leading to high incidence of women breast cancer, in urban-rural regions of the Khyber Pakhtunkhwa (KPK) province.
Patients and Methods: Total of 500 patients were enrolled for the study visiting the cancer hospital, Institute of Radiotherapy and Nuclear Medicine (IRNUM), Peshawar. Among them, 270 were breast cancer patients. Logistic regression models were estimated through the backward elimination procedure to determine the risk factors of breast cancer.
Results: Factors responsible were menopausal status, reproductive status as well as the interaction of family history and diet were the significant risk factors of rural regions. In an urban region, menopausal status and the interaction of breast feeding and diet were the significant risk factors.
Conclusion: This study concludes that reproductive, menopausal status, along with family history, diet and the joint effects of breast feeding were the significant risk factors. Our recommendation from this study was that, the breastfeeding is a precautionary measure against breast cancer. It is, therefore, suggested that mothers, particularly in urban areas, should be encouraged to feed their children by breast feeding.
Key words: Breast cancer, risk factors, backward elimination, brown test, logistic regression, urban and rural regions.

Introduction

Cancer cause the cells to get out of control and most types of cancer cells form a tumor and are cited after the region of the upper body where the tumor grows. Cancer of breast starts in the tissue of the breast. The breast cancer disease may arise from either the glands or the channels of the breast. If cancer develops from the glands, it is called lobular carcinoma. While if it occurs in the canals of the breast it is known as ductal carcinoma.[1] Worldwide, breast cancer is the most frequent cancer in adult females. Nearly, half of these instances are in developed countries: around 220,000 in Europe and around 180,000 in North America, annually.[2] The breast cancer disease is responsible for 26.5% of all new cancer cases among women in Europe and 17.5% of cancer deaths. The highest occurrence rates are witnessed in North America, while the lowest danger of breast cancer disease is witnessed in Asia and Africa.[3] A total of 9556 fresh cases of cancer of the breast was identified in Australia in 1996 and 2,623 expiries

Corresponding Author:
Salahuddin
Department of Family and Community Medicine
College of Medicine, University of Dammam
Kingdom of Saudi Arabia .
Email: salahuddin_90@yahoo.com

Received: 12 November 2016, **Accepted:** 27 March 2018, **Published:** 28 June 2018

Authors Contribution

S, A, N and M have done the conceptualization of project. S did the data collection. S and N also did the literature search and study design. A did the statistical analysis. Drafting, revision and writing of the manuscript were done by S and N. M was also involved in various medical aspects of this research work and in the discussion of results.

credited with cancer of the breast.[4] There are various elements linked with the incidence of the cancer of the breast, such as: menopause, age, childbirth, breastfeeding, consumption of alcohol and type of diet. Several surveys have been conducted indicating an association amongst reproductive factors and the hazard of breast malignant neoplastic disease.[5,6] It has been shown that the data regarding breast cancers and smoking of tobacco are unpredictable.[7,8] A study was conducted to examine the relationship between body mass index (BMI) and breast cancer in African American women. It was found that BMI at reference date and change in BMI were associated with increased risk of breast cancer in African American women.[9] It has been observed that eating of meat (especially red) might raise the risk of breast cancer.[10] Some past studies indicate that utilization of oral contraceptives might raise the risk of cancer of the breast in postmenopausal female, particularly those female with a history of more than ten years of its use.[11] It has been calculated that having one first-degree relative, nearly 80% risk of the cancer of breast increases, nearly 3 fold risk increases having two first-degree relatives, and 4 fold risk increases having 3 or more first degree relatives.[12]

A study was conducted with the objective to assess risk factors leading to high occurrence of women breast cancer in various hospitals of Lahore, Pakistan. The authors concluded that women from the middle class family background, with higher body mass index and a high ratio of incomplete pregnancies were the significant risk factors of breast cancer.[13] A brief review has been given to provide awareness about breast cancer as well as an updated knowledge about its occurrence and risk factors: age, family history, early menarche, intake of combined estrogen and progestin menopausal hormones, alcohol consumption, physical inactivity, low socioeconomic status and lack of awareness of breast cancer in Pakistan.[14] A study was conducted to estimate rate and risk ratio of breast cancer in rural and urban regions. The objective of this study was to compare breast cancer incidence rates in rural, urban and metro regions of India. The study reveals that the lifestyle operative in rural area is protected against the risk of developing breast cancer.[15]

Several other studies have been conducted in Pakistan, aiming to evaluate different risk factors causing breast cancer in women.[16-19]

Institute of Radiotherapy and Nuclear Medicine (IRNUM), Peshawar, is the only Government Cancer Hospital in KPK province. The patients from all over KPK are coming to this hospital for cancer treatments. Therefore, it was more appropriate to collect the data from IRNUM for this study.

In this work, an attempt has been made to assess various risk factors leading to high incidence of women breast cancer in urban-rural regions of the KPK Province, and to estimate appropriate models of the occurrence of breast cancer in the population of both regions.

Statistical package SPSS was used for the analysis.

Patients and Methods

The data were collected from an aggregate of 500 patients, among them, 270 cases were of breast cancer patients visiting to the IRNUM, Peshawar, during 2005-2007. In 270 breast cancer patients, 107 belong to the urban areas and 163 belong to the rural areas. All the information was collected from the patient's record provided by hospital administration. The risk factors are: Menopausal status (M), Breastfeeding (F), Oral contraceptives (C), Reproductive status (R), Diet (D) and Family history (H), while the dependent variable is the Breast cancer (B).

Generalized linear models include linear regression and analysis of variance, logit and probit models for control responses; log-linear and multinomial response models for counts and some frequently uses models as special cases.[20] Different kinds of the Generalized Linear Model are now frequently utilized and enforced in various statistical computer packages. Logit Models, Logistic Regression, Poisson Regression and Log-Linear models are different kinds of the generalized linear model. Throughout this investigation, the dependent variable is dichotomous (yes, no); thus, the logistic regression analysis is a suitable technique. In logistic regression analysis, the probability of an event can be directly estimated as they do fall in the interval 0 and 1, and the association amongst the independent and dependent variables are not a linear function. Logistic regression was, therefore, used to develop risk assessment models.

Results

The percentages of invasive ductal carcinoma, invasive lobular carcinoma and other kinds of cancer are greater in rural area than urban areas.

The variable age is measured in years, the height is in centimeter and the weight is in kilograms. Descriptive Statistics of breast cancer patients by residence is given in Table-1.

Table 1: Descriptive statistics of breast cancer patients by residence.

	Urban			*Rural*		
	Age	*Height*	*Weight*	*Age*	*Height*	*Weight*
Freq.	107	107	107	163	163	163
Mean	46.96	155.48	61.69	47.16	154.93	61.34
Median	45	154	62	46	155	62
Mode	40	152	60	35	155	68
Range	52	37	42	71	41	42
SD	10.8	6.35	8.86	11.88	6.31	9.04
Skew-ness	0.483	0.631	-0.284	0.532	-0.057	-0.439
Kurtos-is	0.267	1.809	0.223	0.292	2.222	0.416

In an urban area, the standard deviations for age, height and weight are 10.8, 6.35 and 8.86 while in rural areas, the standard deviations for age, height and weight are 11.88, 6.31 and 9.04, respectively. It is clear that there is not much difference between these measures of the both regions.

To test the hypothesis, regarding equality of mean ages and mean weights, for rural and urban areas, we set our hypothesis as:

i. The Mean ages of breast cancer patients are equal in the rural and urban areas, and
ii. The Mean weights of breast cancer patients are equal in the rural and urban areas.

An assessment of the normality of data is a prerequisite for many statistical tests, like t-test, because normal data is an underlying assumption in parametric testing. We used Kolmogorov-Smirnov normality test. The test rejects the null hypothesis that the population is normal, if the *p*-value of the test is less than chosen α (e.g., 0.05). Thus, if $p < 0.05$, the population is not normal.

Table 2: Test statistics.

	Age	*Weight*
Mann-Whitney U	8685	8636
Asymp. Sig.(2-tailed)	0.955	0.893

We applied Kolmogorov-Smirnov normality test of the data on the age and weight distributions for urban and rural areas. In all cases, the calculated $p < 0.001$, the result was highly significant, the data was not normal. We cannot apply t-test and therefore, refer to any non-parametric test. A non-parametric test, Mann-Whitney U test, was used and the results are given in Table-2.

Since the age variable *p*-value is equal to 0.955, greater than 0.05. It is highly insignificant, we therefore accept the null hypothesis of equal ages and conclude that the ages of breast cancer patients are equal in both rural and urban areas. Similarly, for weight variable, *p*-value is equal to 0.893, greater than 0.05. It is also highly insignificant; we therefore accept the null hypothesis of equal weights and conclude that the weights of breast cancer patients are equal in both rural and urban areas.

The height distributions of urban and rural areas are normal. The 95% confidence interval estimate for mean height of breast cancer patients in urban and rural areas, are given in Table-3.

Table 3: 95% confidence interval estimate for mean height.

Groups	*Sample size*	*95% Confidence Interval for Mean height*	
		Lower bound	*Upper bound*
Urban area	107	154.26	156.69
Rural area	163	153.95	155.90

Now, to test the hypothesis that the mean heights of breast cancer patients are equal in both urban and rural areas, we can use an independent sample t-test, as the distribution of height is normal. These are given in Table-4.

Since *p*-value is equal to 0.485, greater than 0.05. It is highly insignificant; we therefore accept the null hypothesis of equal mean heights of breast cancer patients in both areas and conclude that the heights of breast cancer patients are equal in both rural and urban areas.

The patients who visit IRNUM were the residence of urban and rural areas of KPK Province. The proposed risk factors are breastfeeding (F), reproductive status (R), family history (H), menopausal status (M), diet (D) and oral contraceptives (C), whereas the dependent (response) variable is breast cancer (B).

Out of the total 270 samples, 107 belong to urban areas. These 107 samples were used in the model fitting process. Brown method selects an initial model having factors D, H, R, C, F and M. The backward elimination method chooses factor M and one two-factor-interaction effect (D*F) for the final model at step 2. The results (SPSS output) are given in Table-5.

The final selected model is:

$$\text{Logit} = -1.009 + 2.145M + 1.985(D*F). \quad (1)$$

Table 4: Independent samples test.

Height	t-test for Equality of Means						
	T	d.f.	Sig.(2-ailed)	Mean difference	Std. Error difference	95% C.I. of the difference	
						Lower	Upper
Equal variances (assumed)	0.699	268	0.485	0.55	0.79	-1.0	2.10
Equal variances (not assumed)	0.698	226	0.486	0.55	0.79	-1.0	2.10

Table 5: Variables in equation, p-value and 95% confidence interval for urban/rural area.

Analysis for	Variable	B	S.E.	Wald	d.f.	p-value	Exp (β)	95% C.I. for Exp (β)	
								Lower	Upper
Urban Area	D*F	1.985	0.853	5.416	1	0.020	7.280	1.368	38.739
	M	2.145	0.698	9.454	1	0.002	8.546	2.177	33.550
	Constant	-1.009	0.229	19.411	1	0.000	0.365	---------	---------
Rural Area	D * H	2.262	1.151	3.864	1	0.049	9.600	1.007	91.551
	R	1.525	0.447	11.636	1	0.001	4.597	1.914	11.044
	M	2.984	0.776	14.769	1	0.000	19.767	4.315	90.551
	Constant	-1.078	0.201	28.643	1	0.000	0.340	--------	----------

In this model, the odds ratio for menopausal status is 8.546, this suggests that those females who have periods after 50 years of age are almost 8.546 times more likely to have breast cancer than those who have no periods after 50 years of age. The odds ratio for the interaction of diet and family history is 7.280, this suggests that females, having breast cancer family history and who use high fat diet, are almost 7.28 times more likely to have breast cancer than those who have no family history and who use low fat diet. Since all calculated *p* values are less than 0.05, all results are therefore highly significant in urban areas.

Out of the total 270 samples, 163 belong to the rural area. These 163 samples were used in the model fitting process. Brown method selects an initial model having factors D, H, R, C, F and M. The backward elimination method selects the final model having factors R and M, and one two-factor-interaction effect (H*D) at step 5. The results (SPSS output) are given in Table-6. The final selected model is given as:

Logit = -1. 078+2.984 M + 1.525 R + 2.262 (H*D). (2)

In analysis according to the residence, we observe that model for urban area (model given by equation 1) indicates the importance of the (variable) menopausal status and the interaction of (two variables) diet and breast feeding. While the model for rural area (model given by equation 2) indicates the importance of the variables M, R and interaction of two variables D and H. Since all calculated *p* values are less than 0.05, all results are therefore highly significant in rural areas.

Discussion

The main purpose of this study was to investigate and compare the occurrence of breast cancer in urban areas with the incidence of breast cancer in rural areas of KPK Province, Pakistan and to estimate suitable models of the occurrence of breast cancer in the population of both areas. The occurrence of breast cancer were considered in relation to diverse risk factors: menopausal status (M), reproductive status (R), oral contraceptives (C), breast feeding (F), family history (H) and diet (D) of 500 persons coming to the cancer hospital, IRNUM, Peshawar. Other variables under consideration were marital status, residence, age, weight and height. Out of 270 female breast cancer patients, 107 (39.63%) belonged to urban areas, 163 (60.37%) belonged to rural areas.

We found that the count of both, invasive ductal carcinoma and invasive lobular carcinoma were greater in rural parts than the urban part. The age, height and weight of the breast cancer patient were equal in both rural and urban parts.

To fit a model, breast cancer (β) was used as a binary response variable. We considered, β equal to 1 for patients having cancer of the breast and β equal to 0 for patients without cancer of the breast.

For choosing an initial model, Brown method was applied. The method of backward elimination was used for selection of a better model.

To compare the occurrence of cancer of breast patients in rural and urban areas, separate models were fitted using the data of the patients belong to rural and urban areas. The initially selected independent variables were D, C, R, M, F and H. For

rural parts, backward elimination method selects the model having factors R and M, and the interaction factor (H*D). This shows that reproductive status, menopausal status, and the joint risk factor (H*D) were the significant risk factors for the cancer of the breast. In an urban area, we got the final model contained the factors M and (D*F). Hence, the menopausal status and the interaction factor (D*F) were the significant risk factors.

We concluded, that for rural parts reproductive status, menopausal status, along with family history and joint effects of diet were the significant risk factors for the cancer of the breast in women.

In an urban area, the authors concluded that menopausal status, diet and the joint effects of breast feeding were the significant risk factors.

This study demonstrates the importance of breast feeding in dropping the risk of breast cancer. Our main finding is that the mothers who breastfeed lower their risk of breast cancer. This finding is in line with other researchers, who shows that breastfeeding, significantly reduces the risk of breast cancer compared to those who never breastfed.[21,22]

From the analysis of the collected data, it clearly indicates that living in a rural area decrease the risk of developing breast cancer. The present work findings strongly recommends that breastfeeding is a precautionary measure against breast cancer. It is, therefore, suggested that mothers, particularly in urban areas, should be encouraged to feed their children by breast feeding.

Acknowledgement

The authors are thankful to the referee's valuable comments for the improvement of the article.

Conflict of interest: The authors declare that there is no conflict of interest.

References

1. Ghafoor A, Samuels A, Jemal A. Breast Cancer Facts and Figures 2003 - 2004. American Cancer Society publications. (Accessed on 14th September 2011) Available from URL:http://www.cancer.org/acs /groups/content/@nho/documents/document/caff2003brfpwsecuredpdf.pdf
2. Park K. Park's Textbook of Preventive and Social Medicine, 21st Edition. Banarsidas Bhanot Publishers: Jabalpur, Madhya Pradesh, India; 2011.
3. Parkin DM, Bray FI, Devesa SS. The Cancer burden in the year 2000: The global picture. Eur J Cancer 2001; 37(Suppl 8): 4-66.
4. Condon JR, Armstrong BK, Barnes A, Cunningham J. Cancer in Indigenous Australians: a review. Cancer Causes Control 2003; 14(2): 109-21.
5. Gao YT, Shu XO, Dai Q, Potter JD, Brinton LA, Wen W, et al. Association of menstrual and reproductive factors with breast cancer risk: results from the Shanghai Breast Cancer. Int J Cancer 2000; 87(2): 295-300.
6. Clavel-Chapelon F, E3N-EPIC Group. Differential effects of reproductive factors on the risk of pre-and postmenopausal breast cancer. Results from a large cohort of French women. Br J Cancer 2002; 86(5): 723-7.
7. World Health Organization: International Agency for Research on Cancer. Tobacco smoke and involuntary smoking /IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans; 83, 2004: Lyon, France. (Accessed on 14th October 2005) Available from URL: Available from: http://monographs.iarc. fr/ENG/Monographs/ vol83/
8. Kropp S, Chang-Claude J. Active and passive smoking and risk of breast cancer by age 50 years among German Women. Am J Epidemiol 2002; 156(7): 616-26.
9. Zhu K, Caulfield J, Hunter S, Roland CL, Payne-Wilks K, Texter L. Body mass index and breast cancer risk in African American women. Ann Epidemiol 2005; 15: 123-8.
10. Dai Q, Shu XO, Jin F, Gao YT, Ruan ZX, Zheng W. Consumption of animal foods, cooking methods, and risk of breast cancer. Cancer Epidemiol Biomarkers Prev 2002; 11(9): 801-8.
11. Van Hoften C, Burger H, Peeters PH, Grobbee DE, Van Noord PA, Leufkens HG. Long-term oral contraceptive use increases breast cancer risk in women over 55 years of age: the DOM cohort. Int J Cancer 2000; 87(4): 591-4.
12. Hamajima N, Hirose K, Tajima K, Rohan T, Calle EE, Heath CW Jr, et al. Alcohol, tobacco and breast cancer – collaborative reanalysis of individual data from 53 epidemiological studies, including 58515 women with breast cancer and 95067 women without the disease. Br J Cancer 2002; 87(11): 1234–45.
13. Rabia T, Sadia H, Marriam ZB, Fatima A. Molecular stress or geography, risk factors leading to high prevalence of Breast Cancer. J App Pharm 2013; 4: 706-12.
14. Asif HM, Sultana S, Akhtar N, Rehman JU, Rehman RU. Prevalence, risk factors and disease knowledge of breast cancer in Pakistan. Asian Pac J Cancer Prev 2014; 15(11): 4411-6.
15. Nagrani RT, Budukh A, Koyande S, Panse NS, Mhatre SS, Badwe R. Rural urban differences in breast cancer in India. Indian J Cancer 2014; 51: 277-81.
16. Bhurgri Y, Kayani N, Faridi N, Pervez S, Usman A, Bhurgri H, et al. Patho-epidemiology of breast cancer in Karachi, 1995-1997. Asian Pac J Cancer Prev 2007; 8: 215-20.

17. Butt Z, Shahbaz U, Naseem T, Ashfaq U, Khan UA, Khan MR, et al. Reproductive risk factors for female breast cancer: a case-control study. Annals 2009; 15: 206-10.
18. Butt Z, Haider SF, Arif S, Khan MR, Ashfaq U, Shahbaz U, et al. Breast cancer risk factors: A comparison between pre-menopausal and postmenopausal women. J Pak Med Assoc 2012; 62: 120-4.
19. Zahra F, Humayoun F, Yousaf T, Khan NA. Evaluation of risk factors for carcinoma breast in Pakistani women. J Fatima Jinnah Med Assoc 2013; 1: 34-8.
20. McCullagh P, Nelder JA. Generalized Linear Models. 2nd Edition (Chapman & Hall / CRC Monographs on Statistics & Applied Probability) Aug 1, 1989.
21. Akbari A, Razzaghi Z, Homaee F, Khayamzadeh M, Movahedi M, Akbari ME. Parity and breastfeeding are preventive measures against breast cancer in Iranian women. Breast Cancer 2011; 18(1): 51-5.
22. Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies, including 58,209 women with breast cancer and 101,986 women without the disease. Lancet. 2001; 358(9291): 1389-99.