IJBPAS, October, 2017, 6(10): 1835-1841

ISSN: 2277–4998

International Journal of Biology, Pharmacy and Allied Sciences (IJBPAS)

'A Bridge Between Laboratory and Reader'

www.ijbpas.com

STATISTICAL-ANALYSIS OF SOCIO-ECONOMIC RISK FACTORS FOR BREAST CANCER IN WOMEN: A CASE-CONTROL STUDY IN PUNJAB, PAKISTAN

KOUSAR N[1*], HUSSAIN M[2], IMRAN MA[2], ALI A[2], FATIMA N[2] AND HASHMI MH[3]

1: Department of Statics, Government College University, Faisalabad, Pakistan

2: Institute of Agricultural and Resource Economics, University of Agriculture, Faisalabad, Pakistan

3: Centre of Agricultural Biochemistry and Biotechnology (CABB), University of Agriculture, Faisalabad

***Corresponding Author: Kousar N: Research Associate, Institute of Agricultural and Resource Economics, University of Agriculture, Faisalabad, Pakistan; Email: mehranhussain@hotmail.com**

Received 20th Nov. 2016; Revised 27th Jan. 2017; Accepted 14th April 2017; Available online 1st Oct. 2017

ABSTRACT

Objective: To identify the significant socio-economic risk factors of breast cancer and to predict probability of suffering from breast cancer due to significant risk factors.

Study Design: Cross sectional, Metric formula and Statistical techniques like (Wald test statistic and logistic regression). A case-control study was conducted during the academic session's 2013-14 in GC University, Faisalabad.

Materials and Methods: The data set comprised of 547 breast cancer patients and 1020 controls in all with 1:2 ratio between cases and controls. The data for breast cancer patients were collected from the four leading cancer hospitals as Allied Hospital Faisalabad (AHF), Institute of Nuclear Medicines and Oncology Lahore (INMOL), Punjab Institute of Nuclear Medicine (PINUM) Hospital Faisalabad and Mayo Hospital Lahore. A structured questionnaire was prepared in English and Urdu is used to interview the cases and controls. The dependent variable is presence of breast cancer in binary form. The questionnaire is hunting the information from cases and controls about the socio-economic variables (independent variable) like age at marriage, breastfeeding history, menarche age, use of oral contraceptives, consanguinity etc.

Results: Metric formula was used to develop different categories and these categories also coded.

This study results shows that the estimated values of variables i.e. breastfed, education, body mass index, consanguineous marriage and ethnicity have highly significant values of logistic regression analysis of all married women. The odds ratio(O.R)results shows that risk of breast cancer decreased by number of children increases, awareness of the disease rises, age at menarche advances from age of thirteen, physical activity and women living in urban areas. Similarly risk of breast cancer increased by taking the oral contraceptive, getting married in family and job status of working. The results of cox and snell R^2 and Nagelkerke R^2shows that model is reliable means model is good fit.

Conclusions: It is concluded that the married women with consanguineous marriage and negative history of breastfeeding are at significantly increased risk of breast cancer. Early age at menarche is not a significant risk factor for this population. Higher number of pregnancies is a significant protective and higher body mass index and ethnicity is also directly related with the risk of breast cancer.

Keywords: Breast cancer, Wald test, odds ratio, logistic regression

INTRODUCTION

According to World Health Organization (WHO) estimates, breast cancer incidence is far more in America and Europe than in Asia. Breast cancer incidence is increasing in all regions of the world with majority of rise seen in developing countries due to progressive westernization of social, cultural and reproductive trends [1].

The incidence of breast cancer is highest in Pakistan compared to women from neighboring countries. Women of all geographic areas, races and ethnicities are affected. The available population based cancer registry data from South Asia shows that Pakistan has the highest Age Standardized Rate of breast cancer at 69 per 100000 in this region [2] and [3]. Every year 40,000 women die of breast cancer and this incidence is increasing in young girls [4] and [5].

Cancer is one of the disastrous diseases and to a great part of our population a constant physical and mental anguish. Breast cancer is world's most corporate cancer of females accounting for nearly a quarter of all cancers in the women. Breast cancer incidence is increasing in all regions of the world with majority of rise seen in developing countries like Pakistan. A case-control study is conducted at province (Punjab) level to identify the potential risk factor lactation for breast cancer among females. The objective of this study is to identify the significant socio-economic risk factors of breast cancer and to predict probability of suffering from breast cancer due to significant risk factors. The findings of this study will be a good contribution and addition in the medical research and will be a pathway for future

Biostatisticians, medical researchers and epidemiologists.

MATERIAL AND METHOD

Design, Setting and Participants

A case-control study was conducted during the academic session's 2013-14 in GC University, Faisalabad. The hospitals remained the best option for the researchers to approach the cases. The data for breast cancer patients were collected from the four leading cancer hospitals as Allied Hospital Faisalabad (AHF), Institute of Nuclear Medicines and Oncology Lahore (INMOL), Punjab Institute of Nuclear Medicine (PINUM) Hospital Faisalabad and Mayo Hospital Lahore. The data set comprised of 547 breast cancer patients and 1020 controls in all with 1:2 ratio between cases and controls. The main objective of this study was to identify the significant socio-economic risk factors of breast cancer and to predict probability of suffering from breast cancer due to significant risk factors.

Data Collection Tools/Instruments

Cross sectional data were collected from the respondents by using purposive sampling technique was used to collect the information with the help of a well-structured questionnaire and this questionnaire prepared in both languages (English and Urdu) as well used to interview the cases and controls. Some information related to the patients' history also collected from the records (history sheet) of the hospitals. The questionnaire was designed after consulting doctors/specialists in the relevant field and on basis of their recommendations. In addition to basic information about demographic and socio economic factors i.e. age, education, number of child's, job status and physical activity etc. The questionnaire also sought the information from cases and controls about the variables like information about full term pregnancies and births, age at marriage, breastfeeding history, menarche age, and use of oral contraceptives, education, physical activity, body mass index, ethnicity, job status and consanguinity. The data were restricted to married women. For this case-control study, the interviewer was not blind to case-control status of the respondents; therefore the presence of interviewer's bias cannot be excluded.

Software and Data Analysis

Statistical Package for Social Science (SPSS) software of version 20.0 was used to analyze the collected data. Binary logistic regression model and other statistical tools (odds ratios and confidence intervals) were calculated for different risk factors. Significance of parameter estimated associated with different risk factors in the logistic regression model were test by Wald test. A tremendous

treatment of generalized linear models is presented in Dupont and Page [3], [6], [7] and [8].The dependent variable is presence of breast cancer. All the data were coded and entered into computer. As and when needed editing of the data was done. For advanced analysis all the data were scrutinized. The outcome variable for this study was binary. The data were collected for 547 cases and 1020 control for different possible risk factors of breast cancer. A total of 1567 women were studied out of which 547 were cancer patients admitted in the above mentioned hospitals. These 1020 women were control women not suffering from breast cancer. The Table 1 describes the study population of some risk factors collected through questionnaire very clearly. Steven`s measurement system for scale of measurements is used to decide the variables. The Quetelet index proposed by astronomer and statistician Adolphe Quetelet for a person of weight W(kg) and height H(meters) is simply:

$$BMI = W/H^2$$

This formula is called the Metric formula. It is a ratio of weight in kilograms and square of height in meters. Some variables are divided as positive and negative. The positive category is coded as 1 and negative as 0 [9], [10], [11] and [12]. Similarly ever is coded as 1 and never is 0 before analysis. The age at menarche below 13 age is coded 1 and above 13 age 0. As age at menarche below 13 is a prognostic risk factor for our study population. The age at marriage 25 and above is risk factor coded as 1. The body mass index 28 and above is coded 1 as if it is a risk factor.

RESULTS AND DISCUSSION

Each risk factor' number and their percentages in data to analysis is given in the table1 (given in annexure).For the overall data, all married women logistic regression was fit, results are shown in Table2 and Table3. The Table2represents the estimate value of parameter, standard error, Wald test value and their significance or P-value. The Table3 represents odds ratio and their 95% confidence interval also. The estimate value -0.372 means that increase\decrease in number of children results in decrease\increase by 0.372 in log odds value of presence of breast cancer. The increase\decrease by breastfed year results 0.359 decrease\increase in log odds value of presence of breast cancer. The increase\decrease of education causes the 0.275 decrease\increase in log odds value of presence of breast cancer. The increase\decrease of body mass index and ethnicity causes the 0.275 and1.06 decrease\increase in log odds value of presence of breast cancer respectively. Similarly, the estimate value of remaining variables can be interpreted.

As the other estimate values of variables i.e. breastfed, education, body mass index, consanguineous marriage and ethnicity have high significant value less than 0.001 against the null hypothesis. Thus these are a highly significant values of logistic regression analysis of all married women. The variable age at menarche has relatively less significant P-value. The remaining variables like intake oral contraceptive, age at marriage, job status and physical activity have insignificant P-value. The interpretation of estimate value in terms of odds ratio is discussed in Table3.

The interpretations of odds ratio in Table3 is described here. The odds ratio 0.689 interprets that risk of breast cancer decreases by 31% as number of children increases. Similarly the interpretation of 0.801 is that risk of breast cancer decreases by 19% as age at menarche advances from age of 13. The variable education has the odds ratio 0.760 interprets that the risk of breast cancer decreases as awareness of the disease rises. The 0.R.=1.060 implies that there is 6% more chances/risk of breast cancer in case of in taking the oral contraceptive. Similarly, the values of odds ratio of getting married in family increases the risk of breast cancer by 47%. The job status of working has odds ratio 25% more than odds ratio of breast cancer. It may be because of having less number of children. The physical activity decreases the risk of breast cancer by 23%. The women living in urban areas have less risk than women living in rural areas by 80% as O.R = 0.201.To check the goodness of fit against the hypothesis that model is good fit, the Hosmer and Lemeshow test is applied in SPSS. The test-statistic value is 8.656 and P-value is equal to 0.372 which is greater than 0.05. It concludes that model is good fit. The results of cox and snell R Square and Nagelkerke R Square values were also in agreement.

Table1: Some Risk Factors of Study Population

Risk Factors		Total (1567)		Cases (547)		control (1020)	
		No.	Percentage	No.	Percentage	No.	Percentage
Smoking History	Ever	168	10.7	07	1.2	161	15.8
	Never	1430	91.3	540	98.7	883	86.6
Consanguineous Marriage	Positive	809	51.6	244	44.6	565	39.5
	Negative	643	41.0	286	52.3	357	60.5
Age at Menarche	Below 13	65	4.1	142	25.9	62	6.0
	13 and Above	1146	73.1	405	74.0	741	72.6
Age at Marriage	25 and Above	436	27.8	202	37.0	234	22.9
	Below 25	119	7.6	44	8.0	75	7.4
Body Mass Index	28 and Above	486	31.0	72	13.2	414	40.6
	Below 28	1081	69.0	475	86.8	606	59.4

Source: Authors own calculations

Table 2: Variables in the Logistic Regression Fit for Married Women

Variable	Estimate	S.E.	Wald test value	P-value
Age at Marriage	0.039	0.025	2.422	0.120
Number of Child	-0.372	0.037	99.364	<0.001**
Breastfed	-0.359	0.038	88.823	<0.001**
Age at Menarche	-0.221	0.074	8.840	0.003
Education	-0.275	0.065	18.045	<0.001**
Intake Oral Contraceptive	0.058	0.111	0.270	0.603
Job Status	0.226	0.335	0.454	0.500
Physical Activity	-0.269	0.404	0.445	0.505
Body Mass Index	-0.209	0.019	125.354	<0.001**
Consanguineous marriage	0.387	0.081	22.838	<0.001**
Ethnicity	-1.602	0.188	72.626	<0.001**

Source: Authors own calculations

Table 3: Odds Ratio and 95% C.I. for O.R. for all married women

Variables	ODD RATIO (O.R)	95% C.I. for Odd Ratio	
		Lower	Upper
Age at marriage	1.040	0.990	1.092
No. Of child and stillbirth	0.689	0.640	0.741
Breastfed	0.698	0.648	0.753
Age at menarche	0.801	0.693	0.927
Education	0.760	0.669	0.862
Intake oral contraceptive	1.060	0.852	1.318
Job status	1.253	0.650	2.415
Physical activity	0.764	0.346	1.685
Body mass index	0.812	0.783	0.842
Consanguineous marriage	1.473	1.257	1.727
Ethnicity	0.201	0.139	0.291
Chi-square =8.656	Cox and Snell R^2= 0.353	Nagelkerke R^2= 0.480	

Source: Authors own calculations

CONCLUSION

This study of female subjects at Punjab, Pakistan, showed that the married women with consanguineous marriage and negative history of breastfeeding are at significantly increased risk of breast cancer. Early age at menarche is not a significant risk factor for this population. Higher number of pregnancies is a significant protective and higher body mass index and ethnicity is also directly related with the risk of breast cancer. This study suggest that the risk of breast cancer can also be avoided or reduced by doing the physical activity, encouraging early marriages, avoiding cousin marriages and government should create awareness of usefulness of breast feeding.

REFERENCES

[1] Butt, Z., Shahbaz, U., Naseem, T., Ashfaq, U., Khan, U. A., Khan, M. R., Hashmim, A. T. and Bukhari, M. H., Reproductive risk factors for female Breast cancer: a case control study, 2009.

[2] Moore, K. L., Clinically Oriented Anatomy, 5th Edition. Lippincott Williams & Wilkins; 289, 2009, ISBN-0781736390.

[3] Dupont, W. D., and Page, D. L., Risk factors of breast cancer in women with proliferative breast

disease, National England Journal of Medicines, 31(2), 1985, 146-51.

[4] National Breast Cancer Awareness, Global facts and figures, 2011.

[5] Toor, I.A. and Perveen, R., Factors Influencing Girls Primary Enrolment in Pakistan, The Lahore Journal of Economics, 9(2), 2004, 141-155.

[6] Althuis, M. D., Dozier, J. M., Anderson, W. F., Devesa, S. S. and Brinton, L. A., Global trends in breast cancer incidence and mortality 1973-1997, International Journal of Epidemiology, 34(2), 2005, 405-412.

[7] Franca-Botelho, A. C., Ferreira, M. C., Franca, J. L., Franca, E. L., and Franca, A. C. H., Breastfeeding and its Relationship with Reduction of Breast Cancer, 2012.

[8] Leady, P.D., Practical research Planning and Design (4th Ed.), Macmillan, 1989, New York.

[9] Menard, S., Coefficients of determination for multiple logistic regression analysis, The American Statistician 54, 2000, 17-24.

[10] Padmavathi, J., Logistic regression in feature selection in data mining. International Journal of Scientific & Engineering Research. 3(8), 2012, ISSN 2229-5518.

[11] Press, S.J. and Wilson, S., Choosing between logistic regression and discriminant analysis, J. Amer. Statist. Assoc, 73, 1978, 699-705.

[12] Tritchler, D., An algorithm for Exact Logistic Regression. Journal of the American Statistical Association, 79, 1984, 709-711.