Original Article

Investigation of Breast Cancer Risk Factors in northern states of Sudan using Logistic Regression Analysis

Mohammed EA[1*] and Ahmed SM[2]

ABSTRACT

Background: Breast cancer is the most common type of cancers and leading cause of death among women worldwide. In Sudan breast cancer is the most common type of cancer and its incidence has been rising for the past two decades.

Objective: To investigate whether the breast risk factors of northern states (Northern and River Nile) are different from other states in Sudan.

Methods: A case-control study was conducted in Radiation and Isotopes Centre Khartoum RICK targeting the breast cancer patients who came to get treatment or follow-up during 2010. A total of 60 breast cancer patients from northern states and 60 control breast cancer patients from other Sudan states were interviewed using purposively designed questionnaires. Logistic regression has been used for modeling the probability that a breast cancer could be developed as a function of risk factors.

Results: This study showed that risk factors of breast cancer do not differ significantly between northern states (Northern and River Nile) and other Sudan states except the education level, are the patient alive, disease history and lactation.

Conclusion: There is no difference between risk factors of breast cancer in all Sudanese states; all Sudanese women are exposed to breast cancer with matching risk factors.

Key words: breast, risk factors, logistic regression.

Breast cancer is the most common type of cancer and leading cause of death among women worldwide. It is estimated that there was approximately 719,000 new cases and 308,000 deaths from the disease worldwide in 1985. The International Agency for Research on Cancer (IARC) estimates 1.38 million new cases and 450,000 deaths from breast cancer worldwide for the year 2008[1].

Many different factors may influence a woman's risk of developing breast cancer. The importance of some of these factors is well established, but for others, the link is more a matter of speculation than fact. In addition, some of the factors that influence risk of breast cancer cannot be identified; several involve aspects of the woman's family and reproductive history rather than her personal habits. This makes it more difficult to develop a strategy for reducing the risk of breast cancer.One of the best ways to fight breast cancer and prevent it from occurring, by identifying and controlling factors that increase a person's risk of developing the disease[3].

In Sudan, breast cancer is the common cancer type diagnosed in the Radiation and Isotope Center in Khartoum (RICK). There is an observed high registry rate of breast cancer among women coming from two Northern States of Sudan (namely the Northern and River Nile States) at RICK. Furthermore, an un-published data showed that women from Northern States are at more risk of breast cancer rather than other women in Sudan[2].

The main objective of this research was to investigate whether the breast risk factors for women of Northern States are different

1.Tropical Medicine Research Institute, National Center for Research, Sudan,
2.Sara Mohamed Ahmed, Institute of Nuclear Medicine, Molecular Biology and Oncology (INMO), Sudan.
* Correspondent: Ehab Ahmed Mohamed
E-mail: ehabfrah@hotmail.com

from other States in Sudan using logistic regression analysis.

Methods:

A case-control study conducted in the Radiation and Isotopes Centre-Khartoum (RICK) targeting the breast cancer patients who came for treatment or follow-up during 2010. A total of 60 breast cancer patients from Northern States (Northern and River Nile) and 60 control breast cancer patients from other Sudan States were interviewed using purposively designed questionnaires. All breast cancer cases registered in RICK in the period from 2000 to 2010 were also analyzed.

Logistic regression have been used for modeling the probability that a breast cancer could be developed as a function of age, occupation, education level, residence, family history of breast cancer, disease history, marital status, age at marriage, age at conception, age at menarche, age at menopause, pregnancy, contraception use, lactation, protein intake, carbohydrate intake and family history of cancers in general in the studied cases. Logistic regression methods were used to fit with the simple and multiple logistic regression models, testing for the significance of the coefficients and the confidence interval estimation. The goal of any method used was to select those variables that result in best model within the scientific based context of the problem, seeking for the best model that seems to be containing all those variables[3]. In statistical inference procedures we used chi-square distribution based on the likelihood ratio score or Wald test statistics which aregood to fit statistics[3,4].

Results:

This study included a total of 60 breast cancer patients from Northern States (Northern and River Nile) and 60 control breast cancer patients from other Sudan States attended the Radiation and Isotope Center in Khartoum (RICK) in 2010. Breast cancer cases registered in RICK during the period from 2000-2010 were analyzed comparing breast cancer cases in women from Northern States with those in women from other Sudan Statesas shown in figure (1)[5].

The demographical data of studied women with breast cancer in Northern States and other Statesare demonstrated in table (1).

The risk factors related to maternal health of the studied women in Northern States and other Statesare shown in table (2).

Table (3) represents the history of breast cancer or other cancers in the studied population.

Logistic analysis:

Logistic regression model was specified and the regression results related to the factors that influence risk factors of breast cancer were reported.

After initial univariable analysis, one

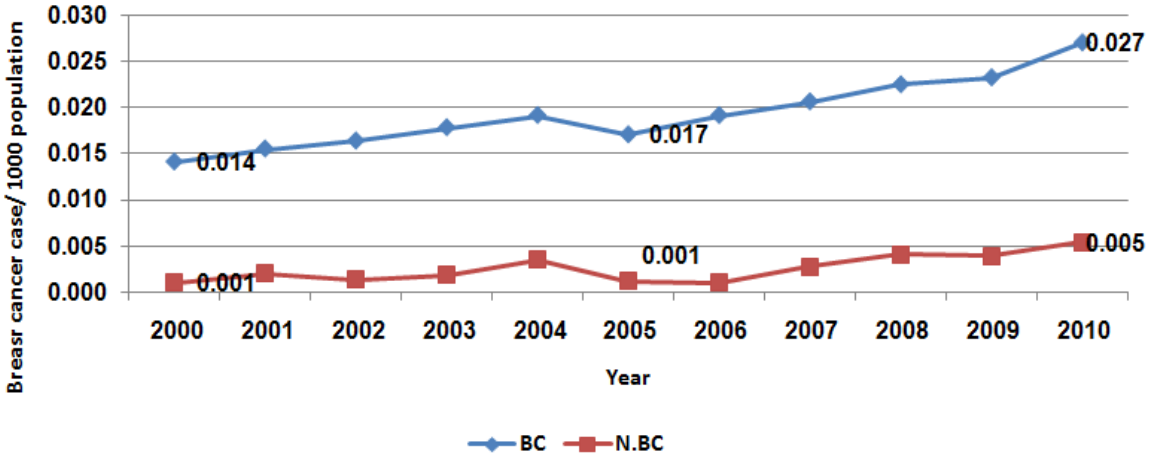

Figure 1: Breast cancer cases in whole Sudan (BC) and Northern States (N.BC)

Table 1: The demographical features of population under study

Risk factor		northern states	other states	P-value
Age	less 45 year	53.3%	59.3%	0.67
	more 45 year	46.7%	41.7%	
Occupation	Housewife	78.3%	75%	0.82
	Employer	21.7%	25%	
Education	Educated	31.7%	38.3%	0.57
	Not educated	68.3%	61.7%	
Residence	Urban	68.3%	61.7%	0.56
	Rural	31.7%	38.3%	
Social status	Married	86.3%	18.3%	1.0
	Single	13.4%	81.7%	

Table 2: The maternal health risk factors of the studied women from Northern and non-Northern States of Sudan.

Risk factor		Northern states	other states	P-value
Age at menarche	< 15	81.7%	88.3%	0.421
	>15	16.7%	10%	
Age at marriage	< 18	43.3%	48.3%	0.313
	>18	46.7%	40%	
Age at conception	< 18	25%	30%	0.67
	>18	51.7%	50%	
Number of Pregnancies	<8 times	68.3%	63.3%	0.094
	> 8 times	6.7%	18.3%	
Age at menopause	<45	50%	41.7%	0.51
	>45	23.3%	26.7%	
Contraception use		26.7%	28.3%	1.0
Lactation		66.7%	78.3%	0.22

Table 3: Cancer history in studied women from Northern and non- Northern States of Sudan

Risk	Northern States	other States	P-value
History of breast cancer	30%	11.7%	0.013
History of other cancer	20	6.7	0.04

variable was found to be significant (history of breast or other cancer).Hence at the first stage all these variables were included in the model. The results of fitting the univariate logistic regression model to these data are given in table (4). This table presents the following information which isgiven for each variable listed in the first column:

- The estimated slope coefficient for univariate logistic regression model containing only this variable (B).
- The estimated standard error of the estimated slope coefficient (S.E).

• The likelihood ratio test statistic, Wald (G), for the hypothesis that, the slope coefficient is zero. Under the null hypothesis, this quantity follows the chi-square distribution with one degree of freedom (Wald).
• The significance value, level for the likelihood ratio test (Sig.).
• The estimated odds ratio, whichwas obtained by exponentiation of the estimated coefficient (OR).

Overall percentage: This gives the percent of cases for which the dependent variables were correctly predicted from the model, increased from 50% for the null model to 73.1% for the full model.

Hosmer-Lemeshow test: To test null hypothesis there is a linear relationship between the predictor variables and the log odds of the criterion variable. Cases were arranged in order by their predicted probability on the criterion variable and since chi-square statistic is computed comparing the observed frequencies with those expected under the linear model. A nonsignificant chi-square indicates that the data fit the model well (p-value = 0.955).

Discussion:

The main purpose of the present study was to examine from an empirical point of view the factors that may be risk factors of breast cancer among two groups of Sudanese women treated at RICK; one from Northern States and the second from other States of Sudan.

The incidence of breast cancer has been rising for the past decade as shown in figure (1). This figure also illustrated the same trend of breast cancer for all Sudan and Northern Sudan States.

As shown in table (1) the demographic features of the studied population are similar to each other and all p-values are more than 0.05. Moreover, table (2) showed that the maternal health data of breast women in Northern States and other States are semi equal and there is no significant difference since p-values are

Table 4: Results of fitting the univariate logistic regression model.

Risk factor	B	S.E.	Wald	df	Sig.	OR
Age	0.886	1.135	0.608	1	0.435	2.43
Occupation	0.579	1.126	0.264	1	0.607	1.78
Education level	0.091	0.894	0.01	1	0.919	1.10
Residence	1.976	1.115	3.14	1	0.076	7.21
Family history of breast cancer	-0.984	213.166	0	1	0.996	0.37
Is alive	0.024	2.182	0	1	0.991	1.02
Disease history	-1.607	1.283	1.569	1	0.210	0.20
Marital status	19.799	28350	0	1	0.999	3.00
Age at conception	-1.31	1.01	1.682	1	0.195	0.27
Age at marriage	0.705	1.077	0.428	1	0.513	2.02
Age at menarche	-1.526	1.149	1.764	1	0.184	0.22
Age at menopause	0.977	0.919	1.13	1	0.288	2.66
Pregnancy	1.987	1.026	3.748	1	0.053	7.29
Contraception use	-1.295	1.029	1.586	1	0.208	0.27
Lactation	-21.183	40190	0	1	1.000	0.00
Lipid	0.598	0.837	0.51	1	0.475	1.82
Protein	0.19	1.062	0.032	1	0.858	1.21
Carbohydrates	-1.467	0.87	2.843	1	0.092	0.23
Family history of cancer	1.388	2.06	0.454	1	0.500	4.01
Habits	-0.722	1.602	0.203	1	0.652	0.49
Constant	-1.61	49190	0	1	1.000	0.20

less than 0.05, so they have the same chance in getting breast cancer.However, the history of breast cancer and other cancers as general in Northern States is higher than that of other States (p-value <0.05).

To validate that there is high risk to Northern State females than other states females, we introduced logistic regression analysis which is a suitable statistical tool for analysis of this type of data and hypotheses of the research. Logistic regression analysis is an increasingly popular analytic tool usually used to investigate the relationship between a categorical outcome and a set of explanatory variables. The outcome, or response, can be dichotomous (yes, no) or ordinal (low, medium, high). It enables predicting the probability that the "event of interest" will occur as a linear function of one (or more) continuous and/or dichotomous independent variables[3,4]. Logistic regression and least squares regression are almost identical. Both methods produce prediction equations. In both cases, the regression coefficients measure the predictive capability of the independent variables[6].

The results of logistic regression analysis in this study showed that there is no difference in risk factors between cases and controls, so the breast cancer risk factors are same among all Sudanese women. These results eliminate any observations or hypothesis that Northern females are more likely to develop breast cancer than others.

The limitation of this study is that it is an institution basedand of small sample size. Studies that uses statistical analysis are better to be conducted on community basis.

Conclusion:

This study revealed that there is no statistically significant difference between risk factors of breast cancer in women of Northern States and women form other Sudan States. Regional and national studies should be conducted to investigate the risk factors of breast cancer deeply.

References:

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM.Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008.Int J Cancer. 2010 Dec 15;127(12):2893-917.
2. Ahmed S. M. Incidence and risk factors of breast cancer among Sudanese females attending INMO (2003-2005), M.s.c, University of Gezeria, Sudan.
3. Hosmer D. and Lemeshow, S.Applied logistic regression.2nd edition.: John Wiley and Sons, New York. 2000.
4. Pampel F. Logistic regression: A primer. Thousand Oaks, CA: Sage Publications. 2000.
5. Radiation and Isotopes Centre Khartoum (RICK). Reports of Biostatistics and Cancer Research Unit, Khartoum, Sudan. (2000-2010).
6. 6. Agresti A. An introduction to categorical data analysis.John Wiley and Sons, Inc. 1996.