ISSN 1684-8403
Journal of Statistics
Volume 22, 2015. pp. 207-218

Identification of the Risk Factor for Colorectal Cancer in Females in Lahore

Arshia Majid[1], *Muhammad Riaz Ahmad*[2] *and Jaffer Hussain*[3]

Abstract

The main aim of this study was to investigate the risk factors of Colorectal cancer in Lahore by using the Logistic model. It was aCross-sectional case control study. For this study, the information was collected from three public hospitals of Lahore including ANMOL, General Hospital and Shaikh Zayed. The required information consisting of 124 female subjects including 62 cases and 62 controls was obtained from January 2011 to May 2012. The factors were studied descriptively and analytically. By using the Logistic Regression model, Odds Ratios and 95% confidence intervals were obtained. Five factors including age, residential area, lifestyle, constipation, and fruit were found to be significant. The Odds Ratios and 95% confidence intervals of these factors were [1.107, 1.051-1.167], [3.541, 1.040-12.052], [0.136, 0.037-0.499], [10.669, 3.031-67.553], and [0.169, 0.051-0.559], respectively. The constipation was observed to be the major risk factor of Colorectal cancer in females. The factors age, residential area and constipation were found to be directly related to Colorectal cancer. On the other hand, the lifestyle and fruit were observed to be protective factors against the Colorectal cancer.

Keywords

Logit model, Coloretal cancer, Odds ratios, Constipation

[1]Department of Statistics, Government Postgraduate College for Women,Wahdat Colony, Lahore, Pakistan.
Email: arshia196@gmail.com
[2]Department of Statistics, Government Postgraduate College, Jaranwala, Pakistan.
Email: mriaz1346@yahoo.com
[3]Assistant Professor, Department of Statistics, Government College University, Lahore, Pakistan.
Email: jafferfarooq@hotmail.com

1. Introduction

When the growth of cell is uncontrollable or irregular cells developed in the body, called cancer. Cancer has several types based on its origin and it can appear in any part or tissue of the body i.e. in lungs, colon, breast, skin, bones or nerve tissue. Whenever the growth of cells in the colon or the rectum become out of control, causes Colorectal cancer. A tumor is developed due to these cells that grow abnormally, is called malignant. When the lump and tumor is developed in the colon and the rectum it is a type of Colorectal cancer. It is also known as bowel cancer. Today, Colorectal cancer is the most common cancer in worldwide. (Cooper, 2000).

In USA, approximately 93,800 cases of colon cancer were diagnosed and is the third most usually happening of cancer in men and women and the second leading reason of cancer death (Greenlee et al., 2000). Diet and other modifiable lifestyle factors play major roles in the growth of colon cancer (Giovannucci and Smith 1994; Potter et al., 1993 and World Cancer Research Fund et al., 1997). Using national and international comparisons of disease rates, Doll and Peto (1981) conjectured that diet might account for as many as 90% of colon cancer deaths in the US. In USA, the Colorectal cancer is the 2nd most important cause of cancer related death and 3rd most important type of cancer. It has been predictable that the lifetime risk for developing Colorectal cancer about 6% for men and women (Jemal et al., 2002).

In Canada, the Colorectal cancer is the 3rd highest disease in men and women. Colorectal cancer is the 2nd top cause of cancer associated death in men and 3rd in women (Canadian Cancer Society, 2008). Every year, at least 1 million people suffered from Colorectal cancer (Cunningham et al., 2010) and 0.5 million deaths were observed globally (Merika et al., 2010). It is more common in developed countries than in developing countries (Merika et al., 2010). In worldwide, Colorectal cancer is the 3rd highest rank in incidence of cancer and 4th highest rank of death in both male and female (Parkin et al., 2005).

2. Material and methods

This was a Cross-sectional case control study which was conducted to investigate the risk factors of the Colorectal cancer. This study was conducted in three hospitals situated in Lahore and taken all the patients of the Colorectal cancer from these hospitals. The hospitals were General Hospital, Institute of Nuclear

Medicine and Oncology (INMOL) and Shaikh Zayed. In this study, all patients were studied of January 2011 to May 2012. One control was also interviewed against one case that is case control ratio was 1:1.

The sample was taken by consecutive visits of the selected hospitals and all the patients of Colorectal cancer coming to the oncology wards for the treatment were interviewed. To collect the data, a questionnaire was developed to collect the information from both cases and controls. The questionnaire was designed with the help of supervisor and medical advisors. Questionnaire of this study included almost all possible risk factors of the disease, Colorectal cancer. The data was collected through direct interview method from both cases and controls. Several questions were asked during the interview from the cases and controls like age, residential area, life style, family history of cancer, family history of Colorectal cancer, cigarettes smoking, diabetes, constipation etc. All the patients who above 25 years of age admitted in the oncology cancer wards were included in this study. This Analysis is made on 124 female subjects in which 62 are cases and 62 controls. To measure the reliability of the questionnaire used Cronbach's Alpha which was found to be 0.811.

In this study, the response variable was binary and independent variables were the nominal, ordinal and quantitative type. For the identification of risk factors in females, the Binary Logistic Regression model was run and Regression coefficients, Odds Ratio, p-value and 95% confidence intervals for Odds Ratios were computed under this model. To check the significance of the variables, the p-value was compared with predefined value of Alpha 5%. The Odds Ratio and 95% confidence interval of the Odds Ratio were used for the interpretations of the significant risk factors. For the adequacy of the model, Omnibus test and Hosmer and Lemeshow (HL) test were used. SPSS (Version-18) was used for the purpose of Statistical Analysis.

Logistic Regression is the most popular model for the Analysis of binary outcome variables (Draper and Smith, 1981). Logistic Regression model is planned to narrate a probability, which is always some number that lies between zero and one. In epidemiological studies, such a probability gives the chance of an individual developing a disease or not (Hanif et al., 2004). Logistic Regression is a mathematical modeling approach that can be used to express the relationship of several predictor variables to dichotomous response variable that is yes, no (Kleinbuam and Klein, 1994).

Let, y is binary random variable and $p = P(y = 1)$. Let $x_1, x_2, ..., x_n$ be a set of predictor variables, then the Logistic Regression of y on $x_1, x_2, ..., x_n$ estimates the parameters $\beta_1, \beta_2, ..., \beta_n$ using the Maximum Likelihood method of the following equation:

$$\log it(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n$$

Let,

$$Z = \log it(p) = \log\left(\frac{p}{1-p}\right)$$

Then,

$$Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n$$

This can be written as:

$$P = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n)} \quad \text{(Agresti, 2002)}$$

Cox and Snell $R^2$and Nagelkerke R^2 used to observe the Goodness of Fit of the model. R^2 used for interpretation and range from 0 to1 in both measures Cox and Snell $R^2$and Nagelkerke R^2. Cox and Snell R^2can never attain its maximum value 1 but Nagelkerke model can attain its maximum value 1. To measure the strength of association between predictors and dependent variables used $R^{2.}$ (Hosmer and Lemeshow, 1989).

3. Results

This study was based on 124 female subjects in which 62 cases and 62 controls. The ratio between cases and controls was 1:1. The age is taken as continuous variable. The risk of Colorectal cancer goes up with age. According to Levin et al. (2008) the risk of Colorectal cancer is more common in people over the age of 50 years. Different researchers consider the age in different ways in their studies. Younger adults can build up Colorectal cancer, but the chances increase clearly after age 50. Age is definitely the strongest non-modifiable risk factor for developing Colo-Rectal Cancer (CRC). From all 124 female subjects, 23 (18.55 %) belong to the Industrial areas and 101 (81.45 %) belong to the non-industrial areas. The counts (percentages) of the female cases and controls who are the

residents of non-industrial areas are 51 (50.50 %) and 50 (49.50%), respectively. While, the counts (percentages) of the female cases and controls who are the residents of industrial areas are 11 (47.83 %) and 12 (52.17 %), respectively. The numbers (percentages) of overall female subjects who are living in urban and rural area are 54 (43.55%) and 70 (56.45%), respectively. From 54 female subjects of urban areas, the counts (percentages) of cases and controls in urban area are 30 (55.56%) and 24 (44.44%), respectively. Similarly, the counts (percentages) of female cases and controls in rural area are 32 (45.71%) and 38 (54.29%), respectively. The situation states that the percentage of female cases in urban area is higher than the female cases in rural area. Lifestyle is divided into three categories that are sedentary, normal and active, on the basis of exercise. Out of all 124 female subjects, none of the case or control reported about the third category. Therefore, only two categories are used for Analysis which are sedentary (no exercise) and normal (about 45 minutes exercise daily). In this data, the counts (percentages) of sedentary and normal life style in the female subjects are 34 (27.42%) and 90 (72.58%), respectively. In sedentary lifestyle, the counts (percentages) of female patients and controls are 29 (85.29%) and 5 (14.71%), respectively. In the same way, in normal lifestyle the counts (percentages) of female patients and controls are 33 (36.67%) and 57 (63.33%), respectively. The percentage of female patients having sedentary life style is much higher as compared to the controls with sedentary life style. Hence, lifestyle is inversely associated with the disease.

The counts (percentages) of all who are consuming less than 10 glasses of water and 10 or more glasses of water per day are 74 (59.68%) and 50 (40.32%), respectively. The numbers (percentages) of the female cases and controls drinking less than 10 glasses of water per day are 43 (58.11%) and 31 (41.89%), respectively. Similarly, the numbers (percentages) of the female cases and controls drinking 10 or more glasses of water per day are 19 (38%) and 31 (62%), respectively. Hence, amount of water intake is inversely related to the Colorectal cancer and the persons who are drinking water 10 or more glasses of water per day have less chances of Colorectal cancer. It is suggested that higher amount of fluid consumption is a protective measure against the Colorectal cancer.

The counts (percentages) of total cases and controls with constipation and without constipation are 55 (44.35%) and 69 (55.65%), respectively. The counts (percentages) of the female cases and controls with constipation are 40 (72.73%) and 15 (27.27%), respectively. Similarly, the counts (percentages) of female cases and control without constipation are 22 (31.88%) and 47 (68.12%), respectively. According to these results, the subjects with constipation have higher risk of

developing the Colorectal cancer as compared to the subjects without constipation.

The counts (percentages) of both cases and controls in with pile and without pile are 24 (19.35%) and 100 (80.65%), respectively. The counts (percentages) of female cases and controls with pile are 20 (83.33%) and 4 (16.67%), respectively. Similarly, the counts (percentages) of female cases and controls without pile are 42 (42%) and 58 (58%), respectively. These results state that the female cases having pile are at higher risk of Colorectal as compared to without pile. It is concluded that the subjects with pile have a higher risk to develop the Colorectal cancer.

The number (percentage) of the subjects who are using fruits of low, normal and excessive categories are 55 (44.36%), 41 (33.06%) and 28 (22.58%), respectively. In cases, the number (percentage) of fruits of using low, normal and excessive categories are 40 (64.52%), 17 (27.42%) and 5 (8.06%), respectively. Similarly, in controls, the numbers (percentages) of fruits using low, normal and excessive categories are 15 (24.19%), 24 (38.71%) and 23 (37.10%), respectively. According to these results, the females who use more fruits have lesser risk to develop the Colorectal cancer as compare to those who use very less amount of fruit in their daily life.

The value of Chi-square in Omnibus test is $\chi^2 = 90.241$ with p-value 0.000. It shows that at least one of the factors is significantly affecting the response variable. It also shows that model is adequately fit of the data. The alternate test of Omnibus test is HL test. In this study, HL test with $\chi^2 = 12.210$ at p-value = 0.142 which shows that HL test is non-significant. It also shows that the model is adequately fitted. The Cox and Snell $R^2$and Nagelkerke R^2 test are used for the Goodness of Fit. Here Cox and Snell $R^2$and Nagelkerke R^2 are 0.517 and 0.689, respectively.

From Table 2, it is observed that the correctly predicted cases and controls are 55 (88.7%) and 53 (85.5%), respectively. But 09 (14.5%) controls and 07 (11.3%) cases are misclassified 09 (14.5%) as cases and 07 (11.3%) as controls, respectively. The numbers (percentages) of the whole data the correctly classified subjects with both including cases and controls are 108 (87.10%). Similarly, among all data, the numbers (percentages) of misclassified subjects with both

including cases and controls are 16(12.90%). As in this data, the number of correct classification is very high, so the fitted model is adequate.

From Table 3, the significant risk factors and their predictive strengths are observed. There are five factors which are found to be significant. The significant factors are age, residential area (R), lifestyle (LS), constipation and fruits. The Logit model is defined below.

Z=-3.721 + 0.102 * age +1.264 * R -1.998 * LS + 2.367 * constipation – 1.779 * fruits.

where

$$Z = \log it(p) = \log\left(\frac{p}{1-p}\right)$$

That factors which are found to be positively significant are directly related to Colorectal cancer. Three factors age, residential area (R) and constipation are directly related to Colorectal cancer as they are found to be positively significant. While the factors which are found to be negative significant are inversely related to Colorectal cancer. In this data, the factors lifestyle (LS) and fruits are negatively significant which shows that these factors and Colorectal cancer are inversely associated.

4. Discussion

According to Table 3, the five factors age, residential area, lifestyle, constipation and fruit were significant factors. The fruit and life style were inversely related to CRC, while age, residential area and constipation were directly related to CRC. The briefly discussion of these factors are discussed below.

4.1 Age: From Table 3, it is observed that the Odds Ratio and 95% confidence interval for Odds Ratio of age is 1.107 and (1.051, 1.167), respectively. This indicates that the older subjects have 1.107 times higher risk of getting the disease after each year. As confidence interval does not include one which shows that the effect of age is significant. In other words, this study states that the risk of Colorectal cancer goes up with the increase in age. The disease is more common in people over the age of 50 years, but also some cases occur below 50 years of age. The reason for this trend is credited to the time taken for intestinal cells to go through malignant transformation. The creation of Colorectal cancer cells requires multiple mutations which build up over time (Scholefield, 2002).

4.2 Residential area: According to Table 3, the Odds Ratio and 95% confidence interval of Odds Ratio for urban areas are 3.541 and (1.040, 12.052), respectively.

It means that urban area is directly related to Colorectal cancer with Odds Ratio is 3.541 which states that the persons who lives in urban area has 3.541 times higher risk of Colorectal cancer as compared to those who live in rural areas. Hence, the Colorectal cancer is found to be more common in urban areas. The model co-efficient is also positive which shows that there is direct association between urban areas and Colorectal cancer.

4.3 Lifestyle: Lifestyle is divided into three categories that are sedentary, normal and active, on the basis of exercise. Out of all 124 subjects, none of the case or control reported about the third category. Therefore, only two categories are used for analysis which are sedentary (no exercise) and normal (about 45 minutes exercise) daily. The Odds Ratio and 95% confidence interval for the lifestyle are 0.136 and (0.037, 0.499), respectively. It is observed that the effect of lifestyle is inversely associated with the Colorectal cancer. The Odds Ratio of lifestyle is 0.136 which means that subjects with normal lifestyle have (1-0.136=0.864) 86.4% safety against the disease as compared to the subject with sedentary lifestyle. The Logit coefficient of lifestyle (-1.998) is negative, therefore, the effect of lifestyle is inversely associated with the Colorectal cancer.

4.4 Constipation: This study states that, the constipation is observed to be positively significant with the Colorectal cancer with Odds Ratio and 95% confidence interval of Odds Ratio for the subjects with constipation are 10.669 and (3.031, 37.553), respectively. It means that the subject with constipation has 10.669 times higher risk of Colorectal cancer as compare to those without constipation. Hence, constipation has the highest Odds Ratio in this study which means that it is the major risk factor of Colorectal cancer. The Logit co-efficient of constipation i.e., 2.367 is positively significant at p-value 0.000. Therefore, the constipation is directly associated to Colorectal cancer.

4.5 Fruits: The use of fruits more than 2 days in a week is observed to be negatively significant. The Odds Ratio and 95% confidence interval of Odds Ratio for those who consumed fruits more than 2 days in a week are 0.169 and (0.051, 0.559), respectively. It is found that the subjects who are using fruits more than two days in a week is inversely associated with the Colorectal cancer. The Odds Ratio 0.169 means that (1- 0.169 = 0.831) 83.1% safety against the disease for those who are using fruits more than 2 days in a week as compared to those who used the fruits 2 or less than two days in a week. The Logit coefficient of fruits (-1.779) is negatively significant with p-value 0.000. It means that higher use of provides protection against the risk of Colorectal cancer. Current

knowledge about food consumption patterns indicates that a diet high in vegetables, fruit, and fiber is protective against certain types of cancer, but the evidence that fruit consumption is in particular associated to a reduced risk of CRC (Harshman and Aldoori, 2007).

5. Conclusion

In this study, five factors including age, residential area, life style, constipation and fruit were observed to be significant. The Age, residential area and constipation were positively significant while the lifestyle and fruit were found to be negatively significant which means that these factors were inversely related to Colorectal cancer.

6. Recommendations for further research

The important recommendations for further studies are given below:

1. A similar study may be conducted in private hospitals of Lahore because elite class of people attends the private hospital for the treatment of Colorectal cancer.
2. Such study is also conducted in the whole Pakistan.
3. The separate studies may be conducted area wise in Pakistan

Acknowledgement

We are highly obliged and acknowledge the services of the entire medical/paramedical staff and all Heads of Hospitals especially Mr. Khalid Masood, Principal Scientist, Dr. Kamran Saeed and Mr. Azhar-Uz-Zaman of INMOL Hospital, Lahore and Dr. Aman-ur-Rehman, Associate Professor, Histopathology, Sheikh Zaid Hospital, Lahore for their cooperation in filling the questionnaires from the patients.

Table 1: Classifications of different risk factors female subjects

Factors	Categories	Colorectal Cancer		
		No	**Yes**	**Total**
Social Status	less than 10000	005	021	026
	10000-20000	006	025	031
	20000 and above	051	016	067
Residential Area_2	Non-industrial	050	051	101
	Industrial	012	011	023
Residential Area_1	Urban	024	030	054
	Rural	038	032	070
Marital Status	Unmarried	028	006	034
	Married	034	056	090
Lifestyle	Sedentary	005	029	034
	Normal	057	033	090
Education	No (Illiterate)	006	043	049
	Yes (Literate)	056	019	75
Family history of cancer	No	048	043	091
	Yes	014	019	033
Family history of Colorectal cancer	No	062	061	123
	Yes	000	001	001
Personal history of cancer	No	062	053	115
	Yes	000	009	009
Cigarettes smoking	No	062	062	124
	Yes	000	000	000
Passive Smoker	No	033	014	047
	Yes	029	048	077
Use of alcohol	No	062	062	124
	Yes	000	000	000
Radiation therapy	No	060	042	102
	Yes	002	020	022
Diabetes	No	056	053	109
	Yes	006	009	015
Inflammatory bowel diseases	No	062	046	108
	Yes	000	016	016
Fluid consumption (Glasses)	Less than 10 glasses	031	043	074
	10 or More glasses	031	019	050
Hepatitis	No	062	056	118
	Yes	000	006	006
Constipation	No	047	022	069
	Yes	015	040	055
Pile	No	058	042	100
	Yes	004	020	024
Red Meat	Low	050	040	090

	Normal	012	022	034
	Excessive	000	000	000
Fast food	Low	050	052	102
	Normal	012	010	022
	Excessive	000	000	000
Fried food	Low	027	058	085
	Normal	028	003	031
	Excessive	007	001	008
Fruits	Low	015	040	055
	Normal	024	017	041
	Excessive	023	005	028
Vegetables	Low	001	001	002
	Normal	032	041	073
	Excessive	029	020	049

Table 2: Correct and incorrect classification

Observed		**Predicted**		
		Colorectal Cancer		**Percentage Correct**
		No	**Yes**	
Colorectal Cancer	No	53	09	85.5
	Yes	07	55	88.7
Overall percentage(correctly classified subjects both cases and controls)				87.1

Table 3: Model coefficients, Odds Ratios and 95% C.I for Odds Ratios

Factor	**β**	**S.E.**	**Wald**	**Sig.**	**Exp(β)**	**95 % C.I for Exp(β)**	
						Lower	**Upper**
Age	0.102	0.027	14.514	0.000	1.107	1.051	1.167
Residential Area (R)	1.264	0.625	4.093	0.043	3.541	1.040	12.052
LS	-1.998	0.664	9.046	0.003	0.136	0.037	0.499
Constipation	2.367	0.642	13.594	0.000	10.669	3.031	37.553
Fruit	-1.779	0.611	8.484	0.004	0.169	0.051	0.559
Constant	-3.721	1.522	5.976	0.014	0.024		

Note: Significance of the variables has been discussed on the basis of p-value (sig.) and the existence of p-value is on the basis of Wald test.

References

1. Canadian Cancer Society/National Cancer Institute of Canada (2008). Canadian Cancer Statistics 2008; Canadian Cancer Society: Toronto, Canada.
2. Cooper, G. M. (2000). *The Cell: A Molecular Approach*, *2nd ed*. *The Development and Causes of Cancer.* Sinauer Associates, Sunderland, USA.

3. Cunningham, D., Atkin, W., Lenz, H. J., Lynch, H. T., Minsky, B., Nordlinger, B. and Starling, N. (2010). Colorectal cancer. *Lancet*, **375(9719)**, 1030–1047. doi: 10.1016/S0140-6736(10)60353-4
4. Doll, R. and Peto, R. (1981). The Causes of Cancer: Quantitative estimates of avoidable risks of cancer in the United States today. *Journal of the National Cancer Institute*, **66(6)**, 1191-1308.
5. Draper, N. R. and Smith, H. (1981). *Applied Regression Analysis*, 2nd ed. John Wiley and Sons, New York.
6. Giovannucci E. and Willett, W. C. (1994). Dietary factors and risk of colon cancer. *Annuals of Medicine*, **26**, 443-452.
7. Greenlee, R. T., Murray, T., Bolden, S. and Wingo, P. A. (2000). Cancer Statistics. *A Cancer Journal for Clinicians*, **50**, 7-33.
8. Hanif, M., Ahmed, M. and Ahmed, A. M. (2004). *Biostatistics for Health Students with Mannual on Software Applications*, Islamic Society of Statistical Sciences, Lahore, Pakistan.
9. Harshman, R. M., and Aldoori, W. (2007). Diet and Colorectal cancer: Review of the evidence. *Canadian Family Physician*, **53(11)**, 1913-1920.
10. Hosmer, D. W. Jr. and Lemeshow, S. (1989). *Applied Logistic Regression*, John Wiley and Sons, Inc., New York.
11. Jemal, A., Thomas, A., Murray, T. and Thun, M. (2002). Cancer statistics, 2002. *A Cancer Journal for Clinicians*, **52**, 23-47.
12. Kleinbuam, D. G. and Klein, M., (1994). *Logistic Regression: A Self Learning Text*, 2nd ed. Springer-Verlag, New York.
13. Levin, B., Lieberman, D.A. and McFarland, B. (2008). Screening and surveillance for the early detection of Colorectal cancer and adenomatous polyps, 2008: A joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and The American College of Radiology. *A Cancer Journal for Clinicians*, **134**, 1570–1595.
14. Merika, E., Saif, M. W., Katz, A., Syrigos, K. and Morse, M. (2010). Review. Colon cancer vaccines: An update. *In vivo (Athens, Greece)*, **24(5)**, 607-628.
15. Parkin, D. M., Bray, F., Ferlay, J. and Pisani, P. (2005). Global cancer statistics, 2002. *A Cancer Journal for Clinicians*, **55**, 74–108.
16. Potter, J. D., Slattery, M. L., Bostick, R. M. and Gapstur, S. M. (1993). Colon cancer: A review of the epidemiology. *Epidemiology Review*, **15**, 499-545.
17. Scholefield, J. H. (2002). Screening for Colorectal cancer. *British Medical Bulletin*, **64**, 75-80.
18. World Cancer Research Fund (1997). Food, Nutrition and the Prevention of Cancer: A Global Perspective. American Institute of Cancer Research, Washington, DC.