*International Journal of Mathematics and Statistics Invention (IJMSI)*
*E-ISSN: 2321 – 4767 P-ISSN: 2321 - 4759*
*www.ijmsi.org Volume 2 Issue 5 || May. 2014 || PP-01-08*

# A Multinomial Logistic Regression Analysis to Study The Influence Of Residence And Socio-Economic Status On Breast Cancer Incidences In Southern Karnataka

Madhu B, Ashok N C and S Balasubramanian
*JSS University, Mysore – 570015, India.*
*Email id: drmadhusri@gmail.com*

***ABSTRACT :*** *Introduction: Breast Cancer is the commonest female cancer worldwide. NCRP indicates rising trends of breast cancer in India. Need to understand breast cancer burden among women in the society from different socio-economic background. Objectives: a) Is there a difference in the pattern of breast cancer cases in different socio-economic status with reference to their area of residence.b) Demonstrate the application of multinomial logistic regression analysis to examine the factors associated with breast cancer in high income, middle and low income families. Methodology: Breast Cancer cases reported to the Bharath Hospital and Institute of Oncology (BHIO) from 2007 to December 2011 were analysed. Statistical Analysis: Descriptive analysis like chi-square analysis and multinomial regression analysis is performed. Results: Out of the 909 breast cancer cases, 440 (48.2%) were from rural areas. In urban areas 64.8% belonged to middle income families whereas in rural areas 48.2% belonged to low income families. MLR analysis showed that Illiteracy, nulliparity, young women(< 40 years) belonging to nuclear families had higher odds of breast cancer in middle and low income families when compared to high income families.*

## I. INTRODUCTION

World wide, Breast cancer is the commonest female cancer with an estimated 1.38 million new cancer cases diagnosed in 2008 representing 23% of all cancers in women. [1] The incidence rates show a marked geographical variation from 27.3 per 100000 in less developed countries to 66.4 per 100000 in more developed countries.[2] However, the mortality rates due to breast cancer are high among the low income countries and survival rates with breast cancer are better in high income countries. Better survival rates in developed countries are attributed to early screening and better treatment facilities.[3] As per the recent estimates, by the year 2030 the burden of breast cancer will increase globally to over 2 million new cases per year.[1] Further more, this increase in cases will be in the developing regions of the world. In India, as per the National Cancer Registry Program ( NCRP ) reports, the burden of breast cancer and cervical cancer in North Eastern States constitute around 25%, whereas in all other states theses two cancers contributed 35.2 to 57.7%of the total cancers.[4] Thus indicating , rising trends in breast cancer incidence and declining trends of cervical cancer.[5-9]

With the rising incidence of breast cancer in India, it is essential to understand how the disease burden is shared among women in the society from different socio-economic background. Our previous analysis done on the cancer data obtained from Tertiary cancer hospital suggests that 52.4% of the breast cancer cases during the year 2007-2011 were from rural areas and 47.6% were from urban areas. Breast cancer once thought as a disease of the urban women is now affecting rural women too. There is a need for risk analysis for breast cancer with respect to residence and Socio-economic status of an individual and develop models for the prediction of breast cancer in such geographic environments. Modelling of risk processes such as risk awareness, risk identification, monitoring and reporting, planning and mitigation is essential for better planning of comprehensive breast cancer screening and treatment services.

The present paper demonstrates the use of multinomial (polytomous ) logistic regression model for predicting the risk of breast cancer in different socio-economic status.

## II. RESEARCH OBJECTIVES

a) Is there a difference in the pattern of breast cancer cases in different socio- economic status with reference to their area of residence.

b) Demonstrate the application of multinominal logistic regression to examine factors associated with Breast cancer in high income families to middle and low income families

c) Determine the significance of explanatory variables

## III. DATA AND METHODS

For the present multinominal logistic regression modelling, the breast cancer cases reported to the Bharath Hospital and Institute of Oncology ( BHIO) from January 2007 to December 2011 were scrutinised for details from the inpatient case record that is maintained by the Medical Records Department of Bharath Hospital Institute of Oncology. The Cases identified include all case records with ICD code C.50. Medical case sheets of identified cases were reviewed individually and information on socio-demographic details like name, age , sex, education, marital status, occupation, socio-economic status, health insurance facilities, details of residence like urban or rural, which taluk and district they come from. Clinical, histopathological and treatment details were noted. Duplicate cases were eliminated by cross checking name, age, sex and address. Data were entered into excel sheet and analysed using SPSS 17.0 (SPSS Inc) .The diferrence in the breast cancer cases from urban and rural areas according to high , medium and low socio-economic status was initially analysed using Chi-square tests and later Multinomial logistic regression was performed to identify the risk factors associated with the occurrence of breast cancer in urban and rural areas.

**Multinomial logistic Regression**

The multinomial (Polytomous ) logistic regression model is an extension of the binomial logistic regression model. It is used when dependent variable has more than two nominal or unordered categories. Like binary logistic regression, multinominal logistic regression uses maximum likelihood estimation to evaluate the probability of categorical membership.[10-15]

**Multinominal logistic regression instead of other techniques**

Most multivariate analysis require the basic assumptions of normality and continuous data, involving independent and /or dependent variables as aforementioned. Tabanick et al (2001) argued that multinomial logistic regression technique has a number of advantages as: i) it is more robust to violations of assumptions of multi-variate normality and equal variance and co-variance matrices across groups, ii) easily interpretable diagnostic statistics, iii) most importantly, MLR does not assume a linear relationship between the dependent and independent variables, iv) independent variables need not be interval, v) MLR does not require that the independents be unbounded and lastly vi) normally distributed error terms are not assumed.[10,11]

With the above advantages, MLR is widely used a problem solving tool , particularly in the field of psychology , mathematical finance , engineering and medicine especially for risk analysis and identifying risk factors for a given condition/ event/disease. Data analysis was carried out with aid of both descriptive and inferential analysis.[10-15]

## IV. RESULTS:

**Breast Cancer incidence in High, Middle and Low income families of urban and rural areas.**

The relationship in the occurrence of breast cancer according to socio-economic status and area of residence is described in Table 1 Out of the 909 breast cancer cases, 440 (48.2%) were from urban areas and 469( 51.2%) were from rural areas. Out of the 404 women from urban areas, majority (68.4%) breast cancer cases were from middle income families, followed by 16.8 % from high income families and 14.8 from low income families. For the breast cancer cases from the rural areas it was observed that, majority (48.4%) were from low income families, 45.2% were from middle income families and only 6.4% were from high income families. The difference observed in the occurrence of breast cancer between rural and urban areas is found to be statistically significant.

**Table 1: Breast Cancer Occurrence among High, Middle and Low income families in Urban and Rural areas**

| Socio-Economic Status | Urban | Rural | Total |
|---|---|---|---|
| High | 74 (16.8) | 30 (6.4) | **105 (11.4)** |
| Middle | 301(68.4) | 212 (45.2) | **513(56.4)** |
| Low | 65 (14.8) | 227 (48.4) | **292(32.1)** |
| | **440(48.2)** | **469( 51.2)** | **909 (100.0)** |

$\chi^2$ = 123.133 ; df = 2; p –value = .000

**Socio - demographic correlates of breast cancer according to family income in urban areas**

From table 2a, it is observed that the proportion of breast cancer in high income families of urban areas is increasing with age group and in low income families the proportion of breast cancer are more in younger age. Out of the breast cancer cases occurring in low income families in urban areas, 37.7% were illiterate and only 4.4% were having more than 10 years of schooling. In the urban areas, majority (70.9%) of breast cancer occurrence was observed in women belonging to middle income nuclear families. Nearly one fourth of the non Hindhus were from high income families and majority (71.2%) belonged to middle income families. Nearly three fourth of the of breast cancer cases did not have any form of health insurance and belonged to middle

**Table 2a: Differences in the breast cancer occurrence across high, middle and low income families according to socio- demographic variables in Urban Areas ( n= 440)**

| Variables | Socio-economic status | | | | |
|---|---|---|---|---|---|
| | High (n= 74) | Middle (n=301) | Low ( n= 65 ) | Total (n= 440) | p- value |
| **Age Groups (in years)** | | | | | |
| **< 30** | 1( 11.1) | 3 (33.3) | 5(55.6) | 9 | $\chi^2$ = 54.01, df= 10, p-value =.000 |
| **31-40** | 4 (6.2) | 39 (60.0) | 22(33.8) | 65 | |
| **41-50** | 16(13.4) | 88(73.9) | 15 (12.6) | 119 | |
| **51-60** | 19( 14.6) | 100 (76.9) | 11(8.5) | 130 | |
| **61-70** | 19 (26.4) | 45 (62.5) | 8 (11.1) | 72 | |
| **>70** | 15 (33.3) | 26( 57.8) | 4 (8.9) | 45 | |
| **Age group ( in years)** | | | | | |
| **< 40 years ( Young women)** | 5 (6.8) | 42 (56.8) | 27 (36.5) | 74 | $\chi^2$ = 35.51, df 2 , p – value < .000 |
| **> 40 years** | 69( 18.9) | 259 (70.8) | 38 (10.4) | 366 | |
| **Education** | | | | | |
| **Illiterate** | 5 (8.2) | 33 (54.1) | 23 (37.7) | 61 | $\chi^2$ = 45.33; df 4; p-value < .000 |
| **< 10 years of schooling** | 30 (13.6) | 150 (70.6) | 35 (15.8) | 221 | |
| **>10 years of Schooling** | 39 (24.7) | 112 (70.9) | 7 (4.4) | 158 | |
| **Occupation** | | | | | |
| **Unemployed** | 61(17.1) | 244(68.3) | 52 (14.6) | 357 | $\chi^2$ = .137; df 2; p-value = .931 |
| **Employed** | 13 (15.7) | 57 (68.7) | 13 (15.7) | 83 | |
| **Marital Status** | | | | | |
| **Unmarried** | 3( 10.3) | 21 (72.4) | 5 (17.2) | 29 | $\chi^2$ =2.9 ; df 4; p-value = 0.561 |
| **Married** | 60 (18.7) | 215(67.0) | 46 (14.3) | 321 | |
| **Widow/Divorced** | 11 ( 12.4 ) | 64 ( 71.4) | 14 (15.7) | 89 | |
| **Family** | | | | | |
| **Nuclear** | 33 (13.1) | 178 (70.9) | 40 (15.9) | 251 | $\chi^2$ =5.7 ; df -2 ; p-value 0.0561 |
| **Non Nuclear** | 41 (21.7) | 123 (65.1) | 25 ( 13.2) | 189 | |
| **Children** | | | | | |
| **Nulliparous** | 4(8.7) | 32 (69.60 | 10 (21.7) | 46 | $\chi^2$ =6.3 ; df - 4 ; p-value 0.172 |
| **Parous** | 35 (15.4) | 158 (69.3) | 35 (15.4) | 228 | |
| **Multiparous** | 35 (21.1) | 111 (66.9) | 20 (12.0) | 166 | |
| **Religion** | | | | | |
| **Hindu** | 55(15.1) | 261( 71.7) | 48 (13.2) | 364 | $\chi^2$ = 10.58, df =2, p value = .005 |
| **Non Hindu** | 19 (25.0) | 40 (52.6) | 17( 22.4) | 76 | |
| **Insurance** | | | | | |
| **No** | 40 (17.9) | 129 (57.8) | 54 (24.2) | 223 | $\chi^2$ = 35.00, df =2, p value = .000 |
| **Yes** | 34 (15.7 | 172( 79.3) | 11(5.1) | 217 | |
| **Clinical pattern Stage at Presentation** | | | | | |
| **Stage 1** | 8 (23.5) | 22( 64.7) | 4 (11.8) | 34 | $\chi^2$ = 1.7, df =6, p value = .945 |
| **Stage 2** | 12 (15.4) | 54 (69.2) | 12 ( 15.4) | 78 | |
| **Stage 3** | 23 ( 15.8) | 101 ( 69.2) | 22 ( 15.4 | 146 | |
| **Stage 4** | 3 (18.8) | 10 (62.5) | 3 (18.8) | 16 | |

and low income families. Marital status and parity did not have any significant difference among the three socio-economic groups in urban areas. The stage of breast cancer at first presentation did not show a significant statistical difference across the socio-economic groups.

**Socio - demographic correlates of breast cancer according to family income in urban areas**

There were 469 breast cancer cases from rural areas, out of them , 30 belonged to high income families, 212 belonged to middle income families and 227 of them belonged to low income families. From table 2b, it is observed that 79.2% of breast cancer cases aged less than 30 years were from low income families, 61% of breast cancer cases were aged between 31 to 40 years from low income families. With increase in age the proportion of breast cancer cases in low income families decreases, however in middle income families the inverse phenomenon is observed. ie the proportion of cases in young age was higher when compared to older age group. Majority (72.6%) of illiterate women with breast cancer belonged to low income families. Among breast cancer cases with more than 10 years of schooling, it was observed that 70.3% of the cases were from middle income families. Nearly 3/4$^{th}$ of breast cancer patients in rural areas were employed predominantly in agriculture sector and belonged to low income families. Out of 264 breast cancer patients without health insurance, 62.2% belonged to low income families and those who possess health insurance, 61.5% belonged to

middle income families. Marital status, type of family, parity and religion of breast cancer cases did not vary significantly in high, middle and low income families of rural areas.
Majority (70.0%) of breast cancer patients presenting at Stage 1 were from middle income families and 72.7% of patients with stage 4 breast cancer were from low income families.

**Table 2b: Differences in the breast cancer occurrence across high, middle and low income families according to socio-demographic variables in Rural Areas ( n= 469)**

| Variables | Socio-economic status | | | | |
|---|---|---|---|---|---|
| | High (n= 30) | Middle (n=212) | Low (n= 227 ) | Total (n= 469) | p- value |
| **Age Groups (in years)** | | | | | |
| **< 30** | 0( 0) | 5 (20.8) | 19(79.2) | 24 | $\chi^2$ = 44.01, df= 10, p-value =.000 |
| **31-40** | 0 (0) | 41 (39) | 64(61) | 105 | |
| **41-50** | 9(5.3) | 77(45.0) | 85 (49.7) | 171 | |
| **51-60** | 14( 13.6) | 50 (48.5) | 39(37.9) | 103 | |
| **61-70** | 4 (8.9) | 24 (53.3) | 17(37.8) | 45 | |
| **>70** | 3 (14.3) | 15( 71.4) | 3 (14.3) | 21 | |
| **Age group ( in years)** | | | | | |
| **< 40 years ( Young women)** | 0 (0) | 46 (35.7) | 83 (64.3) | 129 | $\chi^2$ = 24.30, df 2 , p – value < .000 |
| **> 40 years** | 30( 8.8) | 166 (48.8) | 144 (42.4) | 340 | |
| **Education** | | | | | |
| **Illiterate** | 2 (1.0) | 53 (26.4) | 146 (72.6) | 201 | $\chi^2$ = 99.58; df 4; p-value < .000 |
| **< 10 years of schooling** | 19 (9.3) | 114 (55.9) | 71 (34.8) | 204 | |
| **>10 years of Schooling** | 9 (14.1) | 45 (70.3) | 10 (15.6) | 64 | |
| **Occupation** | | | | | |
| **Unemployed** | 26(8.4) | 173(56.0) | 110 (35.6) | 309 | $\chi^2$ = 59.74; df 2; p-value = .000 |
| **Employed** | 4 (2.5) | 39 (24.4) | 117 (73.1) | 160 | |
| **Marital Status** | | | | | |
| **Unmarried** | 0( 0) | 6 (37.5) | 10 (62.5) | 16 | $\chi^2$ =2.1 ; df 4; p-value = 0.703 |
| **Married** | 24 (6.9) | 160(45.7) | 160 (47.4) | 350 | |
| **Widow/Divorced** | 6 ( 5.8 ) | 46 ( 44.7) | 51(49.3) | 103 | |
| **Family** | | | | | |
| **Nuclear** | 15 (5.4) | 123 (44.6) | 138 (50.0) | 276 | $\chi^2$ =1.38 ; df -2 ; p-value 0.500 |
| **Non Nuclear** | 15 (7.8) | 89 (46.1) | 89( 46.1) | 193 | |
| **Children** | | | | | |
| **Nulliparous** | 0(0) | 13 (34.2) | 25 (65.8) | 38 | $\chi^2$ =6.3 ; df - 4 ; p-value 0.178 |
| **Parous** | 15 (7.1) | 98 (46.2) | 99 (46.7) | 212 | |
| **Multiparous** | 15 (6.9) | 100 (45.9) | 103 (47.2) | 218 | |
| **Religion** | | | | | |
| **Hindu** | 26(6.2) | 185(44.0) | 209 (49.8) | 420 | $\chi^2$ = 3.2, df =2, p value = .199 |
| **Non Hindu** | 4(8.2) | 27 (55.1) | 18( 36.7) | 49 | |
| **Insurance** | | | | | |
| **No** | 14(5.3) | 86 (32.6) | 164 (62.2) | 264 | $\chi^2$ = 45.923, df =2, p value = .000 |
| **Yes** | 16(7.8) | 126( 61.5) | 63(30.7) | 205 | |
| **Clinical pattern Stage at Presentation** | | | | | |
| **Stage 1** | 0(0) | 14( 70.0) | 6 (30.0) | 20 | $\chi^2$ = 14.1, df =6, p value = .030 |
| **Stage 2** | 5 (6.1) | 47 (57.3) | 30 ( 36.6) | 82 | |
| **Stage 3** | 9 ( 5.7) | 68 (42.8) | 82 ( 51.6) | 159 | |
| **Stage 4** | 0 (0) | 3 (27.3) | 8(72.7) | 11 | |
| **Note: The number in parentheses are in percentages** | | | | | |

## V. RESULTS OF MULTI-NOMINAL REGRESSION ANALYSIS

**Null Hypothesis (H0) for the MLR** = There was no difference between the model without independent variables and the model with independent variables.
**Alternate Hypothesis (Ha) for MLR** = There is a difference between model without independent variables and model with independent variables.
First consideration was given to overall test relationship.

Secondly, strength of MLR relationship was tested to establish strength of MLR relationship and lastly, evaluating the usefulness of logistic model and relationship between the independent and independent variables.[10,12]

**Overall test of relationship**

First thing in MLR for any risk analysis is to describe the overall test of relationship between the dependent and independent variables.[10,12] Model fitting information in table (5.4.1 ), describes the relationship between the dependent and independent variables and reveals that probability of the model chi-square 455.235 was 0.000, less than the level of significance of 0.05 ( i.e $p < 0.05$).

**Table 5.4.1 Model Fitting Information**

| Model | -2 Log Likelihood | Chi-Square | df | Sig. |
|---|---|---|---|---|
| Intercept Only | 959.265 | | | |
| Final | 504.029 | 455.235 | 18 | .000 |

**Strength of multinomial logistic regression relationship**

Once the relationship is significant, the next step is to establish the strength of multinominal logistic regression relationship, MLR does compute correlation measure to estimate the strength of relationship (pseudo R square measures, such as Nagelkerke's R). Classification accuracy is also a more useful measure to assess the utility of multinominal logistic regression, which compares predicted group membership based on the logistic model to the actual, known group membership, which is value for the dependent variable.[10-12] In this case, using the Cox and Snell R Square and the Nagelkerke R square value, which provide an indication of the amount of variation in the dependent variable. These are described as pseudo R square. In the Table 5 , Cox and Snell R and Nagelkerke R square values are 0.394 and 0.466 respectively, suggesting that 39% to 46% of the variability is explained by these variables used in the model.

**Table 5.4.2 Pseudo R-Square**

| | |
|---|---|
| Cox and Snell | .394 |
| Nagelkerke | .466 |

**Evaluating utility of logistic models**

In order to characterise the model as useful, the overall classification accuracy in the predictive table is noted.[10-12]. From table 5.4.3, the overall predictive accuracy for the present model is 70.1%, suggesting that the model was useful.

**5.4.3 Prediction accuracy Table Classification**

| Observed | Predicted | | | |
|---|---|---|---|---|
| | Breast cancer in high income families | Breast cancer in middle income families | Breast cancer in low income families | Percent Correct |
| **Breast cancer in high income families** | 3 | 100 | 2 | 2.9% |
| **Breast cancer in middle income families** | 2 | 455 | 57 | 88.5% |
| **Breast cancer in low income families** | 0 | 111 | 179 | 61.7% |
| **Overall Percentage** | .6% | 73.3% | 26.2% | 70.1% |

**The classification accuracy rate was 70.1%**

**Relationship of independent and dependent variables**

There are two types of tests to identify the significant individual independent variables. The likelihood ratio test evaluates the overall relationship between an independent variable and dependent variable. While, the Wald test evaluates whether or not the independent variable is statistically significant is differentiating between groups in each embedded binary logistic comparisons.[10-12]

**Table 5.4.4 Likelihood Ratio Tests**

| Effect | -2 Log Likelihood of Reduced Model | Chi-Square | df | Sig. |
|---|---|---|---|---|
| **Intercept** | 504.029[a] | .000 | 0 | . |
| **Residence** | 533.662 | 29.633 | 2 | .000 |
| **Education** | 652.450 | 148.420 | 4 | .000 |
| **Insurance** | 552.371 | 48.342 | 2 | .000 |
| **Occupation** | 542.979 | 38.950 | 2 | .000 |
| **Family** | 508.942 | 4.912 | 2 | .086 |
| **Parity** | 514.026 | 9.997 | 4 | .040 |
| **Age** | 541.756 | 37.726 | 2 | .000 |

Likelihood ratio tests shows the contribution of each variable to model. Referring to table 5.4.4, we can conclude that the independent variables like residence, age, education, occupation, parity and health insurance are significant independent variables related to the occurrence of breast cancer for middle and high income socio-economic groups.

**Table 5.4.5 Parameter Estimates**

| Reference category-1 | | B | Std. Error | Wald | df | Sig. | Exp (B) | 95% Confidence Interval for Exp(B) | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Lower Bound | Upper Bound |
| 2 | Intercept | .647 | .363 | 3.177 | 1 | .075 | | | |
| | Urban | -.336 | .248 | 1.846 | 1 | .174 | .714 | .440 | 1.160 |
| | Rural | 0[b] | . | . | 0 | . | . | . | . |
| | Illiterate | 1.749 | .447 | 15.307 | 1 | .000 | 5.747 | 2.393 | 13.801 |
| | Less than 10 years of schooling | .987 | .273 | 13.026 | 1 | .000 | 2.683 | 1.570 | 4.584 |
| | More than 10 years of schooling | 0[b] | . | . | 0 | . | . | . | . |
| | No Health Insurance | -.504 | .228 | 4.878 | 1 | .027 | .604 | .387 | .945 |
| | With Health Insurance | 0[b] | . | . | 0 | . | . | . | . |
| | Employed | .138 | .319 | .186 | 1 | .666 | 1.148 | .614 | 2.144 |
| | Unemployed | 0[b] | . | . | 0 | . | . | . | . |
| | Nuclear Family | .543 | .251 | 4.667 | 1 | .031 | 1.720 | 1.052 | 2.814 |
| | Non Nuclear Family | 0[b] | . | . | 0 | . | . | . | . |
| | Nulliparous | 1.337 | .585 | 5.228 | 1 | .022 | 3.809 | 1.211 | 11.989 |
| | Parous | .361 | .258 | 1.955 | 1 | .162 | 1.435 | .865 | 2.379 |
| | Multiparous | 0[b] | . | . | 0 | . | . | . | . |
| | Age less than 40 years | 1.320 | .486 | 7.374 | 1 | .007 | 3.744 | 1.444 | 9.707 |
| | Age More Than 40 Years | 0[b] | . | . | 0 | . | . | . | . |
| 3 | Intercept | -2.993 | .517 | 33.551 | 1 | .000 | | | |
| | Urban | -1.324 | .295 | 20.094 | 1 | .000 | .266 | .149 | .475 |
| | Rural | 0[b] | . | . | 0 | . | . | . | . |
| | Illiterate | 4.838 | .540 | 80.301 | 1 | .000 | 126.225 | 43.811 | 363.667 |
| | Less than 10 years of schooling | 2.606 | .408 | 40.853 | 1 | .000 | 13.550 | 6.093 | 30.134 |
| | More Than 10 years of schooling | 0[b] | . | . | 0 | . | . | . | . |
| | No Health Insurance | .793 | .283 | 7.876 | 1 | .005 | 2.210 | 1.270 | 3.844 |
| | With Health Insurance | 0[b] | . | . | 0 | . | . | . | . |
| | Employed | 1.479 | .365 | 16.389 | 1 | .000 | 4.387 | 2.144 | 8.975 |
| | Unemployed | 0[b] | . | . | 0 | . | . | . | . |
| | Nuclear Family | .575 | .308 | 3.473 | 1 | .062 | 1.777 | .971 | 3.251 |
| | Non Nuclear Family | 0[b] | . | . | 0 | . | . | . | . |
| | Nulliparous | 1.848 | .651 | 8.052 | 1 | .005 | 6.350 | 1.771 | 22.764 |
| | Parous | .498 | .311 | 2.570 | 1 | .109 | 1.646 | .895 | 3.028 |
| | Multiparous | 0[b] | . | . | 0 | . | . | . | . |
| | Age Less Than 40 Years | 2.434 | .518 | 22.072 | 1 | .000 | 11.405 | 4.131 | 31.483 |
| | Age More Than 40 Years | 0[b] | . | . | 0 | . | . | . | . |

1 The reference category is: Breast cancer in high income families, 2- Breast cancer in middle income families , 3- breast cancer in low families

b. This parameter is set to zero because it is redundant.

First half of the table describes the risk factors associated with occurrence of breast cancer in middle income families. When compared to educated individuals, being illiterate had an Odds Ratio (OR) =5.7 (95% CI 2.9 to 13.8), p =.000 and in women with less than 10 years of schooling OR =2.68 (95% CI 1.5 to 4.5), p=.000. Women from nuclear families had OR 1.7 ( 95% CI 1.05 to 2.8) for the occurrence of breast cancer. Nulliparous women had an OR 3.8 (95% CI = 1.2 to11.88) and women aged less than 40 years had OR 3.7 (95% CI 1.44 to 9.7) for breast cancer

For the low income families, residing in rural area had an Odds Ratio( OR0 3.7( 95% CI 2.1 to 6.7) than for urban poor, Illiterate women had an OR 126.2 (95% CI 43.8 to 363.4) and women with less than years schooling had OR 13.5 (95% CI 6.09 to 30.13), with no health insurance the OR was 2.2 (1.2 to 3.8) and if employed OR 4.3 ( 95%CI 2.1 to 8.9) , Nulliparous women had an OR 6.3 (95%CI1.7 to 22.7) and women less than 40 years of age had OR 11.4 (4.1 to 31.4).

## VI. DISCUSSION

Breast cancer is the most common cancer diagnosed in women in the world. Breast cancer ranks second according to incidence of cancers in the world.[16] There has been a definite increase in the cases of breast cancer with 115, 251 incident cases and 315, 679 five yearly prevalence rates.[17] According to National Cancer Registry Program (NCRP)recent report for the year 2008, the load of breast cancer is around 20 to 30 % of the total cancers.[4,18] The reports of cancer registries in India, indicate a rising trends in breast cancer incidence and declining trends in cervical cancer.[4,5,18] India being a diverse country, differences in the incidence and mortality of breast cancer for rural and urban areas do exist and there is a need to identify influencing factors for occurrence of breast cancer in these areas. Current literature suggest that differences in the breast cancer diagnosis, treatment and survival is more attributed to economic disparities or disparities in health insurance coverage and neighbour hood income.[19-20] Studies suggest that the influence of socio-economic deprivation is associated with an increased risk of breast cancer mortality.[20] Therefore, there is a need to identify the gaps in the health care between rural and urban groups of Indian women. In the present study, influence of residential location and the socio-economic status of the individual on the occurrence of breast cancer is studied.

The preliminary descriptive analysis has observed that the majority (68.4%)of the breast cancer cases in urban areas were from middle income families whereas in rural areas nearly half were from low income families and another 40% were from middle income families and less than 10 percent were from high income families, indicating that breast cancer is no longer a disease of the urban population or a disease of high income families.When chi-square analysis was performed, it was observed that in both urban and rural areas young women less than 40 years belonging to low income families and illiterates had higher incidence of breast cancer. Employment, marital status and parity and stage of presentation did not differ much for different income groups in urban areas. However, in rural areas employed (as labourers)belonging to low income nuclear families with no health insurance facilities presented with late stage of breast cancer. Studies carried out by Bradley et al in 2002 identified that low socio-economic status was associated with late –stage breast cancer and uninsured women were more likely to have unfavourable outcomes.[19,20] Similar observations were also made by Rabia Ali et al in India[21] and Lantz et al in America .[22] Whereas studies from developed countries like Canada have observed that neighbourhood high income is associated with high incidence of breast cancer.[23] From these findings we can conclude that developing countries like India and rural areas are also facing the increasing burden of breast cancer.

It is essential to identify high risk target population in rural and urban areas and to target specific programs towards them to improve awareness and screening practices. Hence, the multinomial logistic regression analysis was performed. As multinomial regression analysis is a powerful statistical tool and is free from assumptions like normality or linearity, the analysis was used to assess the significant independent factors associated with the incidence of breast cancer for different income families. Factors like illiteracy, belonging nuclear families, younger age and nulliparity were significant predictors of breast cancer in middle income families. For low income families, young women, residing in rural area, being an illiterate and working as labourer with no children and absence of health insurance facilities have higher risk of breast cancer .

The significance of the predictor estimates in the present model parameter are significant. Indicating that the model is a good one and a small deviance value indicates that there is a significant fit of the logistic model.

Sometimes the multi-collinearity, interaction, categories of predictors having Zero cells and complete separation of the two groups will produce wildly improbable results. In certain situations, reports have been observed that one unit change in an independent variable increases the Odds of the modelled event by hundreds or thousands and hence interpretation and implications of the Multinomial regression analysis should be carefully considered.

## VII. CONCLUSIONS:

With reference to the research objectives, the difference in the pattern of breast cancer occurrence across different socio-economic groups of rural and urban areas is evident. The applications of multinomial logistic regression model for identifying the predictors of breast cancer in middle and low income families observed that illiteracy, nulliparity and young age ( 40 years) were significant predictors for breast cancer in MIF and LIF.Secondly, the model Chi-square test was significant at <0.001 level of significance, suggesting a significant relationship between illiteracy, nulliparity and young age for the occurrence of breast cancer.
Lastly, it is recommended that attention to be given to these target population in the form of breast cancer awareness campaigns, screening programs and health insurance facilities which would indirectly reduce the incidence as well as cancer related mortality in low income families of southern Karnataka.

## VIII. ACKNOWLEDGEMENTS:

We thank Bharath Hospital and Institute of Oncology ( BHIO) for providing the permission to obtain the breast cancer data for the present study.

## REFERENCES

[1] Jemal A, Bray F, Melissa M C, Jacques F, Elizabeth W, David Forman. Global Cancer Statistics. CA Cancer J Clin 2011; 61:69–90
[2] Risk Factors for Breast Cancer in India: an INDOX Case-Control Study. Available from http://www.indox.org.uk/research/epidemiology/casecontrol/breastcancer. Accessed on 2.4.2014
[3] Coleman MP et al. Cancer survival in five continents: a worldwide population-based study (CONCORD). Lancet Oncol 2008; 9: 730 – 56.
[4] Annual reports.1982-2008. National Cancer Registry. New Delhi: Indian Council of Medical Research; 1985-2010. Available from http://www.ncrpindia.org. Last accessed on 3.4.2014.
[5] Murthy NS, Agarwal UK, Chaudhry K, Saxena S. A study on time trends in incidence of breast cancer- Indian scenario. Euro J Cancer care (Engl) 2007; 16: 185-6.
[6] Murthy NS, Chaudhary K, Saxena S. Trends in cervical cancer incidence – Indian Scenario. Eur J Cancer Prev 2005; 14:513-8.
[7] Yeole BB. Trends in cancer incidence in female breast , cervix uteri, corpus uteri and ovary in India. Asia Pac J Cancer Prev 2008; 9: 119-22.
[8] Takiar R, Srivastav A. Time trend in breast and cervix cancer of women in India- 1990 -2003. Asia Pac J Cancer Prev 2008; 9:777-80.
[9] Dhillion PK, Yeole BB, Dikshit R, Kurkure AP, Bray F. Trends in breast, ovarian and cervical cancer incidence in Mumbai, Indiaover a 30 year period, 1976-2005: An age – period cohort analysis.Br J Cancer 2011; 105: 723-30.
[10] Chan YH. Biostatistics 305. Logistic regression Analysis. Singapore Med J 2005;46 (6): 259.
[11] Tabanick B G, Fidell L S and Osterlind S J. Using multivariate statistics.US, Allyn and Bacon Boston, 2001
[12] Bayaga A. Multinomail Logistic regression : Usage and application in risk analysis. JAQM 2010; 5(2): 288-297.
[13] Hosmer DW and Lemeshow .Applied logistic regression. New York; Wiley.1989
[14] Schwab JA (2002). Multinomial logistic regression: Basic relationships and complete problems. http://www.utexas.edu/courses/schwab/sw388r7/solvingproblems.
[15] Menard SW. Applied logistic regression analysis.UK: Sage Pubs, 2002.
[16] Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. GLOBOCAN 2008 v 2.0,Cancer incidences and Mortality Worldwide: IARC Cancer base no10.Lyon, France. International Agency for Research on Cancer; 2010.Available from; http://www.globocan.iarc.fr. [last accessed on 5 April 2014]
[17] IARC Fact sheet. Available on from http://www.globocan.iarc.fr/factsheet.asp. [Last accessed 5 April 2014]
[18] Asthana S, Chauhan S, Labani S. Breast and cervical cancer risk in India: An update . Indian J Public Health 2014; 58:5-10.
[19] Bradley CJ, Given CW and Roberts C. Race, Socioeconomic status and Breast Cancer Treatment and survival. Journal of National Cancer Institute 2002; 94 (7):490-496
[20] Vona- Davis L and Rose DP. The Influence of soco-economic disparities on breast cancer tumor and prognosis: A review. J Women Health ( Larchmt).2009 Jun ; 18(6): 883-93.doi: 10.1089/jwh.2008.1127.
[21] Rabia A, Mathew A, Rajan B. Effects of Socio-economic and demographic factors in delayed reporting and late – stage Presentation among patients with Breast cancer in a Major cancer Hospital in South India. Asian Pacific Journal of Cancer Prevention 2008; 9: 703-707.
[22] Lantz PM, Mujahid M, Schwartz K, Janz NK, Fagerlin A, Salem B, Liu L, Deapen D and Katz SJ. The influence of Race, ethnicity and Indiviual Socio-economic factors on breast cancer stage at diagnosis. Am J Public Health 2006; 96(12): 2173-2178.
[23] Borugian M J, Spinell J J, Abanto Z, Xu C L and Wilkins R . Breast Cancer Incidence and neighbourhood income. Statistics Canada, catalogue no82-003-XPE. Health reports 2011; 22 ( 2 ):1- 5.