MEDICAL CHANNEL

Original Article

PREDICTIVE STRENGTH OF CLASSICAL RISK FACTOR OF CORONARYARTERY DISEASE AMONG MALE PATIENTS IN SELECTED HOSPITALS OF PUNJAB, PAKISTAN

⋆ Muhammad Naveed ⋆⋆ Zahid Ahmad

⋆ Department of Statistics, GC University, Lahore. ⋆⋆ Associate Professor, University of Central Punjab Lahore.

ABSTRACT

Objectives: The objective of this study is to find out the risk factor of coronary artery disease (CAD) and develop a model to predict the probability of CAD in the presence of different risk factor.

Data Sources: The sample of 281male subjects were taken from the four hospitals of Punjab.

- Punjab Institute of Cardiology, Lahore.
- Faisalabad Institute of Cardiology, Faisalabad.
- Armed force institute of cardiology, Rawalpindi.
- Ch.PervaizElahiInstitute of Cardiology, Multan.

Design of the Study: This was a cross sectional study and sampling methodology of the study is the Convenient sampling.The target population of the study was the Heart patients coming to the Angiography ward in all public sector hospitals of Punjab and sampled population was the Heart patients of Angiography ward in the four selected hospitals.

Setting: The questionnaire was used to collect the primary data from the patients of four Selected hospitals of Punjab.

Period of the Study: Period of this study was four months.

Materials and Methods: Data was analyzed using SPSS 18. The age of patients was present in the form of mean ± S.E. Binary Logistic regression model was used to estimate the probability of CAD with the help of significant risk factor. Odds Ratio were computed for each significant risk factor.

Results: Average age of male patients was 53.18 ± 10.82 years. In this study Age, Diabetes Mellitus, Hypertension, smoking, Lipoproteins, Life Style, Family History, Diet, Depression and anxiety contributed a significant role for the development of CAD. Because p-value of these risk factors are less than 0.05 level of significance.

Conclusion: Our study showed that the risk of CAD increased by increasing the age. Risk factor smoking, Family History, Hypertension are found to be most significant risk factor in developing CAD among male patients.

Key Words: Logistic Regression , Odds Ratio, Coronary Artery Disease, Risk Factor.

Correspondence Address:
DR. ZAHID AHMAD
Associate Professor
University of Central Punjab.
Lahore.
Cell #: 03004213983
Email: zahid_69@hotmail.com

INTRODUCTION

Coronary heart disease (CHD) is the one of the largest cause of death in the developed countries as well as one of the leading causes of disease burden in developing countries.(Ghziano et al).An estimated 17.3 million people died from CVDs in 2008, representing 30% of all deaths in world. Of these deaths,7.3 million were due to CHD and 6.2 million were due to stroke.(cardiovascular diseases).

The coronary heart disease (CHD) death rate however varies dramatically across the developing countries both in proportion of cardiovascular disease (CVD) deaths as well as proportion of total deaths. Coronary heart disease (CHD) is the leading cause of cardiovascular disease (CVD) death across the world.43% of all CVD deaths are attributable to CHD agreeing to the global burden of disease estimates from 2001. Globally cardiovascular disease (CVD) deaths represent about 30% of all deaths. (Gaziano et al,)By 2030, about 25 million people will die from cardiovascular disease (CVD), mainly from heart disease and stroke. These are planned to remain the single leading cause of death.(cardiovascular diseases)

Pakistan is situated in South Asia region and has highest rate of coronary artery disease. Study conducted in 2002 revealed that approximately 10 million people suffered from acute myocardial infarction (heart attack) in Pakistan. There is more chance of developing coronary artery disease in men than women in early age of Pakistan population. Study conducted in Karachi 2000-2002 showed that sixteen percent of heart attack patients were the age of forty five or less and ninety three percent of them were men. Study conducted in other cities regarding AMI were not different picture from that study. A large study conducted in 1994 the incidence of AMI patient was found to be 192.8 male per hundred thousand. This showed that both natural factor and external factor were responsible for the growth of coronary artery disease in Pakistan. The main problem in Pakistan most of the coronary artery disease patient belonged to lower income class, it has overwhelming effect of the family and society. (iqbal et al)

Cardiovascular disease is responsible for fifty percent of adult's death in all over the world and most of the deaths were due to the coronary artery disease. The people of subcontinent Asia who lived in UK had fifty percent more mortality from coronary artery disease than from UK citizen. In India incidence of coronary artery disease has been reported eleven percent in 2001. (shaniwani et al).WHO reported in 2011, coronary artery disease mortality in Pakistan reached 196258 or 15.36% of total deaths. And age adjusted death rate was 222.95 per hundred thousand, ranked Pakistan in seventeen in the world. (world health ranking).

A significant number of these deaths can be allotted to tobacco smoking, which gains the risk of dying from coronary heart disease (CHD) and cerebrovascular disease 2–3 fold. WHO report showed that Obesity is important risk factors for cardiovascular disease, which is contributing the main cause of death and reports for over 17 million deaths every year. About 7.5 million Deaths every year, or 13% of all deaths can be attributed to raised blood pressure. Unhealthy diet and Physical inactivity are other main risk factors which increase the person's risks to cardiovascular diseases.

Most cardiovascular diseases can be precluded by addressing risk factors such as tobacco use, unhealthy diet and obesity, physical inactivity, raised blood pressure, diabetes and raised lipids.

The real disaster is that obesity and overweight, their related chronic diseases, are preventable. Approximately 80% of heart disease, type 2 diabetes, stroke, and 40% of cancer could be avoided through healthy diet, avoidance of tobacco use and regular physical activity

METHODOLOGY

The sample of 281 subjects was taken from the four hospitals of Punjab. (1) Punjab Institute of Cardiology, Lahore. (2) Faisalabad Institute of Cardiology, Faisalabad (3) Ch.PervaizElahiInstitute of Cardiology, Multan. Binary Logistic regression model was used to estimate the probability of CAD with the help of significant risk factor. Odds Ratio was computed for each significant risk factor.

(a) The sampling technique to be used is convenient sampling.Four hospitals, Punjab Institute of Cardiology Lahore, Faisalabad Institute of Cardiology, Faisalabad, Armed Force Institute of Cardiology Rawalpindi and ChPervaizElahi Institute of Cardiology, Multan were selected from all government hospitals of Punjab, Pakistan. The target population of the study was the Heart patients coming to the Angiography ward in all public sector hospitals of Punjab and sampled population were the Heart patients of Angiography ward in the four selected hospitals. Two group of patients were selected from sampled population in such a way that

- Patients came to Angiography ward and diagnosed as CAD were selected as case group.
- Patients came to Angiography ward and didn't diagnose as CAD were selected as control group.

The patients with known history of CAD were excluded from the study. The logistic regression was used to develop a predictive model for the diagnosis of CAD with the help of significant risk factors. The odds ratios were computed to assess the degree of dependence of CAD on the classical risk factors. The data were collected on Age, Income group, chest pain, Diabetes Mellitus, Hypertension, Smoking, Lipoproteins, Obesity, Life-Style, Family History, diet, Sleep, Depression and Anxiety, Peripheral Vascular Disease, Intermittent Claudication. To determine the predictive strength of risk factors associated with CAD, the logistic regression model was fitted. In logistic distribution we will use the

quantity $\pi(x)$ = $E(Y/x)$ and can be expressed as

$$E(Y/x) = \pi(x) = \frac{e^{\beta_o+\beta_1 x}}{1+e^{\beta_o+\beta_1 x}}$$

A transformation of $\pi(x)$ which is called logit transformation in the logistic function is as follows

$$g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] \qquad (3.18)$$

$$g(x) = \beta_o + \beta_1 x \qquad (3.19)$$

Thus by applying this transformation we get $g(x)$ which has most of the properties of the regression equation. The model $g(x)$ has linear parameters, may be continuous, and may be ranged from $-\infty$ to $+\infty$ depending on the range of x. (Hosmer & Lemeshow, 2000).The multiple logistic regression model assumed a linear relation between the log of odds and predictor variables, and can be written as:

if then

$$g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_o + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p = x_i'\beta$$

Matrix approach is used to estimate the parameter of the multiple logistic regression model and Maximum likelihood estimate method is used in multivariate case. And the simplest form of model become as

$$\pi(x) = \frac{e^{g(x)}}{1+e^{g(x)}} = \frac{1}{1+e^{-g(x)}} = \frac{1}{1+e^{-x_i'\beta}}$$

Let $Z = g(x) = \beta_o + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p$ then $\pi(x)$ becomes

$$\pi(x) = \frac{1}{1+e^{-Z}}$$

TABLE 1
BIVARIATE ANALYSIS FOR MALE PATIENTS

Variables	Pearson χ^2	P-Value	Phi ($\pi(x) = P(Y=1)$) & Cramer's V Coefficients	Contingency Coefficients	Kendall's Tau-b	Remarks
Age	4.561	.033			0.127	S
Family Income	1.155	0.764	0.064	0.064	_____	N.S
Chest Pain	21.208	0.000	0.275	0.265	_	S
Diabetes Mellitus	6.515	0.011	0.152	0.151	_	S
Hypertension	22.422	0.000	0.282	0.272	_	S
Smoking	15.130	0.000	0.232	0.226	_____	S
Lipoproteins	27.824	0.000	0.315	0.300	_	S
Obesity	17.505	0.000	0.250	0.242	_	S
Life Style	47.198	0.000	_____	_____	-.347	S
Family History	11.634	0.001	0.203	0.199	_	S
Diet	22.422	0.000	-0.282/0.282	0.272	_	S
Sleep Habit	12.977	0.002	_____	_____	-0.079	S
Depression & Anxiety	16.734	0.000	0.244	0.237	_	S
PVD	0.027	0.870	-0.010/0.010	0.010	___	N. S
I.C	3.531	0.060	0.112	0.111	_____	N.S

Among the risk factors, age, chest pain, hypertension, smoking, diabetes, obesity, Lipoproteins, life style, family history, diet, depression and anxiety, sleeping habit, are found to be significantly associated with coronary artery disease. On the other hand there is no evidence of significant association between family income and Peripheral Vascular Disease and Intermittent Claudication, with CAD. From the table 2 we see that Lipoproteins has the highest association with CAD among the nominal variables with highest value of Phi/Cramer's V as 0.315 and Contingency Coefficient as 0.300. Among the ordinal variables age has highest association with CAD with the maximum value of Kendall's Tau-b as 0.127.

TABLE 2
TABLE FOR MULTIPLE LOGISTIC REGRESSION ANALYSIS OF MALE PATIENTS

Variable	B	S.E	Wald	d.f	Sig	Exp(B)	95% C.I for EXP(B)	
							Lower	Upper
Age	.038	.018	4.748	1	.029	1.039	1.004	1.076
Chest pain	.436	.238	3.346	1	.067	1.546	.969	2.465
Diabetes Mellitus	1.023	.397	6.637	1	.010	2.780	1.277	6.052
Hypertension	1.420	.379	14.039	1	.000	4.137	1.968	8.696
Smoking	1.449	.367	15.571	1	.000	4.259	2.074	8.748
Lipoproteins	.758	.357	4.513	1	.034	2.135	1.060	4.298
Life Style	-1.173	.233	25.321	1	.000	.309	.196	.489
Family History	1.441	.380	14.349	1	.000	4.224	2.004	8.901
Diet	-1.577	.383	16.962	1	.000	.207	.098	.438
D.A	1.115	.364	9.357	1	.002	3.049	1.493	6.227
Constant	-3.646	1.182	9.520	1	.002	.026		

As shown in the Table 2, nine risk factors age, Diabetes Mellitus, Hypertension, Smoking, Lipoproteins, Life style, Family History, Diet, D.A are significant risk factors associated with CAD,For these risk factor p-values are less than 0.05,odds ratio are also shown in Table 2.

Goodness of Fit

TABLE 3
GOODNESS OF FIT FOR THE OVERALL LOGISTIC REGRESSION MODEL

Hosmer and Lemeshow Test H_o : Observed Values = Predicted Values		
Chi- Square	**df**	**P-value**
$\chi^2 = 2.292$	8	0.971

The Hosmer-Lemeshow statistic (approximately equal to Chi Square) is 2.292, with d.f =8 as shown in Table 3 with a non-significant p-value of 0.971, which shows that model is good fit.

TABLE 4
CLASSIFICATION TABLE FOR MULTIPLE LOGISTIC REGRESSION MODEL

Classification Table				
Observed		**Predicted**		
		CAD		**Percentage Corrected**
		No	**Yes**	
CAD	**No**	63	30	67.7
	Yes	17	171	91
Overall Percentage				83.3

From the Classification Table 4**,** it is seen that out of 93 patients with non-significant CAD 63(67.7%) patients without CAD are correctly predicted by the model. Similarly out of 188 patients with significant CAD 171 (91%) patients are correctly predicted to have CAD. 30 patients not having CAD are misclassified as having CAD also, 17 patients observed to have CAD are misclassified as not having CAD. Overall 83.3% of the total numbers of patients are correctly classified.

Assessing the Model

TABLE 5:
MODEL ASSESSMENT

Omnibus Test (Significance of the Model)				**Proportion of Explained Variation**		
Chi- Square	**df**	**P-value**	**-2 Log Likelihood (Unexplained Variation)**	**Cox & Snell R-Square** $\left(R^2_{CS}\right)$	**Nagelkerke R-Square** $\left(R^2_N\right)$	**Hosmer and Lemeshow's Measure** $\left(R^2_L\right)$
$\chi^2 = 146.437$	10	0.000	210.353	0.406	0.565	0.4104

Hosmer and Lemeshow's measure $\left(R^2_L\right)$ is calculated as:

Chi-Square (model) = -2LL (model) = 210.353
Chi-Square (original) = -2LL (original) = 146.437 + 210.353 = 356.79

$$R^2_L = \frac{-2LL(\text{model})}{-2LL(original)} = \frac{146.437}{356.79} = 0.4104$$

The value of model Chi-square is 146.437 with P-value= 0.000. This is highly significant. Therefore we are 95% confident that fitted model is appropriate.
The unexplained variation of the model is 220.353 from the total variation of 356.79 and 146.437 variation explained by the fitted model.
Cox and Snell R-square is 0.406(40.6%) and Nagelkerke R-square is 0.565(56.5%) indicating the explained variation by the fitted model.

R^2_L = 0.4104 is the amount by which badness-of-fit of the model improves as a result of the inclusion of the predictor variables.

If the parameter β_i is positive, then there is positive relation between the predicted probability $\pi(x)$ for (Y=1) and the values of x_i, and for negative value of β_i, there is inverse relation between $\pi(x)$ and the values of x_i. The predicted values for the multiple logistic regression model are $\hat{\pi}(x)$.
(Hosmer and Lemeshow, 2000)
The wald test-statistic was used to test the significance of logistic regression co-efficient.

Analytic Section

Analytic section consists of two categories
(i) Bivariate Analysis
(ii) Multivariate Analysis

Bivariate Analysis
In Bivariate analysis association of every risk factor was tested with coronary artery disease by using Pearson Chi-Square test. In order to measure the strength of association, Phi (ϕ)/ Cramer's V and Contingency Coefficients were computed of the nominal significant risk factors with CAD. On the other hand Kendall's Tau-b is computed to measure the strength of association of the ordinal risk factors with CAD. Results of bivariate analysis for risk factors are presented in Table 1

Logistic Model and Predicted Probabilities
The final logistic regression model for CADis as follows

$$\hat{Y} = \frac{1}{1+e^{-z}}$$

Where

$$Z = -3.646 + 0.038*(age) + 0.436*(chestpain) + 1.023*(DM) + 1.420*(Hyp) + 1.449*(smoking)$$
$$+0.758*(Lip) - 1.173*(Lifestyle) + 1.441*(F.H) - 1.577*(diet) + 1.115*(D.A)$$

TABLE 6:
PREDICTED PROBABILITIES FOR CAD

Case No.Significant Risk Factors of the Model

	Age	Chest pain	Diabetes	Hyp	Smoking	LIP	Life style	F.H
1	< 50	Typical	Yes	Yes	Yes	No	Normal	Yes
2	> 50	Atypical	Yes	Yes	No	No	Sedentary	No
3	> 50	Non-cardiac	No	Yes	Yes	Yes	Normal	No
4	> 50	Typical	No	No	No	Yes	Active	Yes
5	< 50	Typical	Yes	No	Yes	Yes	Sedentary	Yes

RISK FACTOR FOR THE FITTED MODE

Case No	Diet	D.A	Z	Predicted Probabilities
1	Balanced	Yes	1.36	0.7958
2	Balanced	Yes	-0.755	0.3197
3	Unbalanced	Yes	0.397	0.598
4	Unbalanced	No	-2.447	0.079
5	Unbalanced	No	2.333	0.9116

RESULTS

Case 1: under case-1 there are 79.58% chances of Developing CAD
Case 2: under case-2 there are 31.97% chances of Developing CAD
Case 3: under case-3 there are 59.8% chances of Developing CAD
Case 4: under case-4 there are 7.9% chances of Developing CAD
Case 5: under case-3 there are 91.16% chances of Developing CAD

The probability of developing CAD in the presence or absence of significant risk factors may be calculated by the above model. Some of the cases are discussed below in the table 16.

DISCUSSION

The main aim of this study was to determine the predictive strength of coronary artery disease, for this multiple logistic regression model was applied and backward stepwise conditional method was used. Wald test was used to access the significance of the regression Co-efficient, later on Odds Ratio and 95% C-I was calculated.

The risk factor age was taken as continuous variable in the analysis. It is significant risk factor of coronary artery disease with p-value 0.029 and odds ratio for age is 1.039 which means that by the increasing of one year in age the risk of coronary artery disease is increased 1.039 times keeping all other factor constant.

The reference group for risk factor Diabetes Mellitus was non-diabetic patient. Its coefficient is positive and odd ratio is 2.780 which mean that patients having Diabetes Mellitus have 2.780 times more chance of having coronary artery disease than patients without Diabetes Mellituskeeping all the other factors constant. The odds ratio for risk factor Diabetes Mellitus is greater than 1 and 95% confidence interval for risk factor Diabetes Mellitus does not include 1 therefore risk factor Diabetes Mellitus has positive association with coronary artery disease and is statistically significant.

The variable Hypertension was categorized into two categoriesand non- hypertension patient was taken as a reference group. The variable shows significant positive association with CAD. Its odds ratio shows that people exposed with Hypertension have 4.137 times more chance of having CAD as compare to people unexposed with Hypertension.

The factor smoking was taken as categorical variable and non-smoker patient was taken as a reference group. Its coefficient is positive and odd ratio is 4.259 which indicate that patients with smoking have 4.259 times more chance of having coronary artery disease than patients without smoking keeping all the other factors constant. The odds ratio for risk factor smoking is greater than 1 and 95% confidence interval for risk factor smoking does not include 1 therefore risk factor Smoking has positive association with coronary artery disease and is statistically significant.

The risk factor Lipoproteins was taken as categorical variable and categorized into two categories, hyperlipidemia and non-hyperlipidemia . The reference group was taken as non- hyperlipidemia patient. Its

coefficient is positive and odd ratio is 2.135 which means that patients having Hyperlipidemia have 2.135 times more chance of having coronary artery disease than patients without Hyperlipidemia keeping all the other factors constant.

The risk factor Life Style was divided into 3 mutually exclusive ordinal groups and sedentary life style was taken as a reference group. Its coefficient is negative and odd ratio is 0.309 for life style indicate 1-.0.309=0.691 (69.1%) protection against developing CAD keeping all the other factors constant. The odds ratio for risk factor life style is less than 1 and 95% confidence interval for risk factor life style does not include 1 therefore risk factor life style has negative association with coronary artery disease and is statistically significant.

The risk factor Family History was taken as categorical variable and patient with not present Family History was taken as a reference group. Its coefficient is positive and odd ratio is 4.224 which mean that patients having Family History of CAD have 4.224 times more chance of having coronary artery disease than patients without Family Historykeeping all the other factors constant. The odds ratio for risk factor Family Historyis greater than 1 and 95% confidence interval for risk factor Family Historydoes not include 1 therefore risk factor Family History has positive association with coronary artery disease and is statistically significant.

The risk factor Diet was taken as categorical variable and reference group for unbalance diet. The coefficient shows negative association. The odds ratio of 0.207 for diet indicate that patient with balance diet have 1-0.207=0.793 (79.3%) protection against CAD keeping all other factor constant. The odds ratio for risk factor dietis less than 1 and 95% confidence interval for risk factor dietdoes not include 1 therefore risk factor diet has negative association with coronary artery disease and is statistically significant.

The risk factor Depression and Anxiety was taken as categorical variable and patient without Depression and Anxiety was taken as a reference group. The variable shows significant positive association with CAD. Its odds ratio shows that people exposed with Depression and Anxiety have 3.049 times more chance of having CAD as compare to people unexposed with Depression and Anxiety.

The odd ratio of constant gives the odd ratio of the background and baseline. The interpretation of this odd ratio is possible only if all the study variables are categorical but in our study we use age as continuous variable so we can't interpret it. but if we use age as categorical variable then the odds ratio for constant is 0.329 that shows that patient with these risk factors at the baseline have (1-0.329)100% = 67.1% protection against the patient who have all these risk factors.

CONCLUSION

The risk factors are thoroughly analyzed of CAD among Male patients. Bivariate analysis revealed that Lipoproteins has the highest association with CAD among the nominal variables with highest value of Phi/ Cramer's V as 0.315 and Contingency Coefficient as 0.300. Among the ordinal variables age has highest association with CAD with the maximum value of Kendall's Tau-b as 0.127. In Multiple Logistic regression analysis we found that by increasing age the risk of CAD also increased. The factor smoking, Hypertension and Family History are found to be the most significant risk factor in developing CAD among Male Patients.

REFERENCES

1 Agresti, A., (1996). An Introduction to Categorical Data Analysis, John Wiley & Sons, New York.

2 Gaziano, A.T., Bitton, A., Anand, S., Abrahams-Gessel, S., Murphy, A., (2010). Growing Epidemic of Coronary Heart Disease in Low- and Middle-Income Countries. currprobl cardiol,Vol.35, No.02, pp72-115

3 Hosmer, D.W., and Lemeshow, S., (2000). Applied Logistic Regression, Second Edition, John Wiley & Sons.

4 Iqbal, P.M., (2006). Hyperhomocysteinemia and Coronary Artery Disease in Pakistan. J Pak Med Assoc,Vol.56, No.06, pp282-285

5 Shiwani, H.A., Aziz, A., Shiwani, H.M., (2007). Shall We become vegetarian to minimize the risk of Coronary Heart Disease?J Pak Med Assoc, Vol.57, No.01, pp46-48

6 WORLD HEALTH RANKING. http://www.worldlifeexpectancy.com/pakistan-coronary-heart-disease

7 World HealthOrganization report2005. http://www.who.int/mediacentre/news/releases/2005/pr44/en/

8 World Health Organization report 2013. http://www.who.int/mediacentre/factsheets/fs317/en/