

A logistic regression model of Coronary Artery Disease among Male Patients in Punjab

Sohail Chand

Institute of Statistics, University of the Punjab,
Quaid-e-Azam Campus, Lahore email: sohail@stat.pu.edu.pk

Shahid Kamal

Institute of Statistics, University of the Punjab,
Quaid-e-Azam Campus, Lahore email: director@stat.pu.edu.pk

Abstract

This is a cross-sectional retrospective study of 308 male patients, who were presented first time for coronary angiography at the Punjab Institute of Cardiology. The mean age was 50.97 ± 9.9 among male patients. As the response variable coronary artery disease (CAD) was a binary variable, logistic regression model was fitted to predict the Coronary Artery Disease with the help of significant risk factors. Age, Chest pain, Diabetes Mellitus, Smoking and Lipids are resulted as significant risk factors associated with CAD among male population.

Keywords: Logistic Regression, Odds Ratio, Wald's test.

Introduction

Soldo and Mantom (1985) found the Coronary Artery Disease (CAD) as the leading cause of death and disability in western world and claim more lives than death due to cancer, accidents, and diabetes combined. Bhopal (2002) found that situation is not different in third world countries including Pakistan. For any given level of risk factors, the CAD risk among Asian is at least double than of Whites

Over the last few years a great deal of research and very large sums of money has been spent to discover whether it is possible to alter the incidence of heart disease by altering peoples' life style. CAD is slow, progressive condition which may develop over a patient's whole lifetime and the results of alterations in diet, exercise, living conditions, or drugs may not be seen until many years later.

Smith (2002) found major risk factors that can be changed are Smoking, High Cholesterol, Hypertension, Sedentary lifestyle, Diabetes Mellitus while contributory risk factors are Obesity and Stress

Avsar et. al (1998) assessed the predictive and discriminative value of clinical risk factors of CAD. Product moment correlation coefficients were computed to find the correlation between variables

Haq and Kiyani (1999) studied the prevalence of hyperlipidemia in patients suffering from CAD. A time-based sample of 540 patients attending the OPD of cardiac centre of CMH Multan was included in the study. The average age at which CAD occurred in this sample was 48 years

Nancy and Bohannon (1999) found the high prevalence of CAD in diabetic patients, but the magnitude of this problem is not always appreciated. Data from the 18-year Framingham Study show that the relative risk for CAD in diabetic men and women 45 to 74 years of age is 2.4 and 5.1 times greater, respectively, than for age-matched nondiabetic men and women. In the Islington Diabetes Survey, the prevalence of serious CAD increased from 9% in subjects with normal glucose tolerance to 17% in those with impaired glucose tolerance and 20% in those with diabetes

Methods and Materials

A hospital based cross-sectional 308 indoor male patients were included in the sample after their consent. The research design retrospective research. 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, Chest pain, Diabetes Mellitus, Hypertension, Smoking, Lipoproteins, Obesity, Life-style, Family History, Socio-economic class, Peripheral Vascular Disease, Stroke, Intermittent Claudication, Carotid/Renal Bruit. Life long smoking is calculated as “pack years” using the formula

$$\text{Pack Year} = \frac{\text{Cigarettes smoked daily} \times \text{Smoking Years}}{20}$$

CAD was diagnosed with the help of coronary angiography. To determine the predictive strength of risk factors associated with CAD, the logistic regression model was fitted with the help of backward elimination method. The specific form of the logistic regression model we will use is as follows

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

Where

$$g(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_{k1}$$

is the logit transformation of logistic regression model. The importance of this transformation is that $g(x)$ has many of the desirable properties of a linear regression model. The Wald test-statistic was used to test the significance of the logistic regression co-efficient. The results are shown in Table-1 below.

Table 1: Logistic Regression Model Parameters

Risk factor	$\hat{\beta}$	s.e ($\hat{\beta}$)	Wald test	d.f.	p-value	Odds Ratio
Age (AGE)	0.0840	0.0180	22.6990	1	0.0000	1.0870
Chest pain (CP)	0.5210	0.2550	4.1900	1	0.0410	1.6840
Diabetes Mellitus (DM)	1.0210	0.4530	5.0730	1	0.0240	2.7760
Smoking pack years (SMK)	0.0510	0.0190	7.0970	1	0.0080	1.0520
Lipids (LPD)	0.7110	0.3160	5.0470	1	0.0250	2.0360
Constant	-4.8950	1.1310	17.7330	1	0.0000	0.0070

As shown in above Table-1 five risk factors Age, Chest pain, Diabetes Mellitus, Smoking and Lipids are resulted as significant risk factors associated with CAD. The risk factor Age has an odds ratio 1.0870 which indicates with an increase of 1 year in age the associated risk of CAD will also be increased by 1.0870 times. Since CAD develops over a long period of time so if this odds ratio is viewed in the context of a period of 10 years then the odds ratio for 10 years period will be 2.32 which indicates that the risk of CAD increases by 2.32 times with an increase of 10 years in age while other factors remains constant.

The fitted logit model for CAD among male patients is

$$g(x) = -4.89540 + 0.0840 \text{ AGE} + 0.5210 \text{ CP} + 1.0210 \text{ DM} + 0.0510 \text{ SMK} + 0.7110 \text{ LPD}$$

The goodness of fit of the fitted logistic regression model is given in Table-2:

Table 2: Goodness of fit of logistic regression model

Observed CAD	Predictive CAD		Correct
	Normal	Positive	
Normal	14	52	21.20%
Positive	11	231	95.50%
Overall Model Correct			79.50%
Chi-Square = 58.126, p-value = 0.0000 (<0.01)			

Conclusion

Diabetes Mellitus and Lipids are found to be the most significant risk factors in developing CAD among male population. As far as screening of CAD is concerned, the lipid profile and type of chest pain should be an important part of diagnosis. It has been observed that most of the patients do not have their serum lipid ever before in life. The diagnosis of Hyperlipidemia at early stage can prevent the development of CAD. Moreover, management of diabetic patients can also reduce the chances of CAD.

References

1. Agresti, A. "An introduction to Categorical Data Analysis" John Wiley & Sons Inc. USA, 1996.
2. Avsar, A. et al "The predictive and discriminative value of lipids, lipoproteins and clinical risk factors in CAD" PHJ 1998: 31; 23-33.
3. Bhopal, R. "Heterogeneity among Indians, Pakistanis, and Bangladeshis is key to racial inequities". British Medical Journal 2002, 325(7369): 903.
4. Haq, H. and Kiyani, A.M. "The prevalence of hyperlipidemia in patients suffering from CAD" Pakistan Journal of Cardiology 1999: 10; 9-14.
5. Hosmer, D.W and Lemeshow, S. "Applied Logistic Regression" John Wiley & Sons Inc. USA, 1989.
6. Nancy, J. V. and Bohannon, M. D. "Coronary artery disease and diabetes". The practical peer-reviewed journal of primary care for physicians: 105 (2).
7. Smith, J.F. "Coronary Artery Disease" <http://www.chclibrary.org>
8. Soldo, B.J. and Mantom, K.G. "Health status and service needs of the oldest old: current patterns and future trends" Mibank Mem Fund Q Health Soc. 1985; 63: 286-319.