INDIAN HEART JOURNAL 66 (2014) 587–592

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/ihj

Original Article

Regression methods for analyzing the risk factors for a life style disease among the young population of India

B. Ismail [a], Manjula Anil [b,*]

[a] Professor, Department of Statistics, Mangalore University, Karnataka State, India

[b] Lecturer/Statistician, Department of Community Medicine, A.J. Institute of Medical Sciences, Kuntikan, Mangalore, Karnataka State, India

ARTICLE INFO

Article history:
Received 27 June 2013
Accepted 24 May 2014
Available online 5 August 2014

Keywords:
Cardio vascular risk assessment score
Binary logistic regression analysis
Non parametric smoothing
Kernel local linear regression function
Statistical methods of risk assessment

ABSTRACT

Background: With modernization, rapid urbanization and industrialization, the price that the society is paying is tremendous load of "Non-Communicable" diseases, referred to as "Lifestyle Diseases". Coronary artery disease (CAD), one of the lifestyle diseases that manifests at a younger age can have divesting consequences for an individual, the family and society. Prevention of these diseases can be done by studying the risk factors, analyzing and interpreting them using various statistical methods.

Objective: To determine, using logistic regression the relative contribution of independent variables according to the intensity of their influence (proven by statistical significance) upon the occurrence of values of the dependent cardio vascular risk scores. Additionally, we wanted to assess whether non parametric smoothing of the cardio vascular risk scores can be used as a better statistical method as compared to the existing methods.

Materials and methods: The study includes 498 students in the age group of 18–29 years.

Findings: Prevalence of over weight (BMI 23–25 kg/m^2) and obesity (BMI > 25 Kg/m^2) was found among individuals of 22 years and above. Non smokers had decreased odds (OR = 0.041, CI = 0.015–0.107) and also increase in LDL Cholesterol (OR = 1.05, CI = 1.021–1.055) and BMI (OR = 1.42, CI = 1.244–1.631) were significantly contributing towards the risk of CVD. Localite students had decreased odds of developing CVD in the next 10 years (OR = 0.27, CI = 0.092–0.799) as compared to students residing in hostel or paying guests.

Copyright © 2014, Cardiological Society of India. All rights reserved.

1. Introduction

Tracing history of man from antiquity to present day, disease profile has shown a change. Of late diseases like coronary heart disease are seen to have emerged.[1] Coronary heart disease (CHD) is "impairment of heart function due to inadequate blood flow to the heart compared to its needs, caused by obstructive changes in the coronary circulation to the heart".[2] It is the leading cause of death in the United States of

* Corresponding author.
E-mail address: manjula_anil2006@yahoo.co.in (M. Anil).

http://dx.doi.org/10.1016/j.ihj.2014.05.027

0019-4832/Copyright © 2014, Cardiological Society of India. All rights reserved.

America and increasing by the cause of death in many developing countries. It strikes the ambitious active men in the prime of career and productivity depriving the society of a number of productive years which they would have contributed. In economic terms, the cost of care to the survivors is also high. From the practical point of view, the risk factors have been divided into major groups such as – those which are not amenable to control i.e. age, sex, family history of CHD etc. and – those which are amenable to control i.e. hyperlipidoemia, hypertension, smoking, obesity etc. However, these factors have been studied at community or group level and do not necessarily hold good at the level of an individual.[3] With modernization, rapid urbanization, industrialization and increasing level of affluence, the so called "modernization", the price that the society is paying is a tremendous load of "Non-Communicable" diseases, also referred to as "Chronic diseases" and, often, as "Lifestyle Diseases". Lifestyle is more of attitudes and behaviors, about "predispositions".[4] Coronary artery disease (CAD), one of the lifestyle diseases that manifests at a younger age can have divesting consequences for an individual, the family and society. Prevention of these deaths in young people is a nation's moral responsibility which may be achieved, by studying the risk factors, analyzing and interpreting them using various statistical methods. Also, there is a need to have few more methods to analyze and interpret such data. This study is done to determine, using logistic regression the relative contribution of independent variables according to the intensity of their influence (proven by statistical significance) upon the occurrence of values of the dependent cardio vascular risk scores. Additionally we wanted to assess whether non parametric smoothing of the cardio vascular risk scores can be used as a better statistical method as compared to the existing methods.

2. Materials and methods

The record based study includes the complete enumeration of 498 students in the age group of 18–29 years, from the Institution student database who were registered on or before January 2011 for Bachelor degree in a Medical College in South India. The ten year risk of developing coronary heart disease was calculated using Framingham Heart risk assessment tool. Non responders or participants who had missing response for covariates such as (age/gender) or desired risk factor were excluded from analysis. Binary logistic regression was applied to calculate the odds ratio to assess association between risk and covariates. The dependent variable in the model was cardio vascular risk score. Among the other independent variables included in the model – age, systolic and diastolic blood pressure, fasting blood sugar levels, body mass index, HDL and LDL cholesterol were considered as continuous variables. Logistic regression is designed to find the most parsimonious set of predictors that are effective in predicting the dependent variable. The method of Forward Wald stepwise selection method with the criteria for entry of the variable is 0.05 and removal is 0.1 has been used to select the most optimal subset of independent variable. SPSS version 16, 2013 was used to analyze the data. Also non parametric smoothing of the cardio vascular risk scores was done by fitting of kernel weighted local linear regression function. MATLAB program version 7.0, 2004 was used for non parametric smoothing of the data.

2.1. Methodology

Epidemiology and clinical research is largely grounded on the assessment of risk. When the outcome variable of interest is dichotomous, a tool popular in assessing the risk of exposure or the benefit of a treatment is a logistic regression. Over the past few decades, logistic regression is popularly used as a tool for analysis. It is often the case that the outcome variable is discrete, taking on two or more possible values.[5] In clinical situations, the status of a patient is assessed by the presence or absence of a disease. There are many factors to consider which may or may not correlate with the incidence of the disease. Finding the potential risk factors can help prevent the development of the disease. When all of the diseases and nearly all of the risk factors considered are categorical variables, Hosmer and Lemeshow, state that "the logistic regression model has become the standard method of analysis in this situation".

2.2. Binary logistic regression model

It is used to predict the probability of a change in a categorically-dependent variable, conditional on the values of independent variables. In addition to supplying an estimate of conditioned probability, the model allows one to assess the degree of influence of selected independent variables upon the occurrence of values (categories) of the dependent variable.

In simple linear regression, we assume that the mean of the dependent variable (Y) may be expressed as a linear function of an independent variable (X), in following equation

$$E(Y/X) = \beta_0 + \beta_1 X$$

where expression $E(Y/X)$ represents the expected value of Y for a given value of X. β_0 and β_1 are constants. β_0 an intercept and β_1 a slope of the regression line. This expression implies that it is possible for $E(Y/X)$ to take on any value as X ranges between $-\infty$ and $+\infty$. But in dichotomous data, the conditional mean must be greater than or equal to zero and less than or equal to 1.

We use the quantity $\pi(x) = E(Y/X)$ to represent the conditional mean of Y given X.

The specific form of the logistic regression model is[5]

$$\pi(X) = \exp(\beta_0 + \beta_1 X)/(1 + \exp(\beta_0 + \beta_1 X))$$

A transformation of $\pi(X)$ is the logit transformation. This defined, in terms of $\pi(X)$, as

$$g(X) = \ln[\pi(X)/(1 - \pi(X))]$$

The importance of this transformation is that $g(X)$ has many desirable properties of linear regression model. The logit $g(X)$ is linear in its parameters, may be continuous and may range from $-\infty$ to $+\infty$.

Download English Version:

https://daneshyari.com/en/article/5962002

Download Persian Version:

https://daneshyari.com/article/5962002

Daneshyari.com