INDIAN HEART JOURNAL 66 (2014) 587–592

ELSEVIER

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.

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

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 $+\infty$.

2.3. The multiple logistic regression model

Consider the collection of p independent variables denoted by the vector $X' = (X_1, X_2, \ldots\ldots\ldots X_p)$. Let the conditional probability that the outcome is present be denoted by $P(Y = 1/X) = \pi(X)$.

The logit of the multiple logistic regression model is given by

$$g(X) = \beta_0 + \beta_1 x_1 + \beta_2 X_2 + \ldots\ldots \beta_p X_p$$

Hence the logistic regression model is[5]

$$\pi(X) = \exp(g(X))/(1 + \exp(g(X)))$$

2.4. Smoothing of the model

Over the last several decades, regression analysis has become increasingly popular as a tool for statistical modeling and data analysis. The regression has become an integral component of any data analysis concerned with describing the relationship between a response variable and one or more explanatory variable.[1] The general regression model is of the form $Y = m(X) + \varepsilon$, where Y is the response variable, $m(X) = E(Y \mathrm{l} X)$ is the mean response or the regression function and ε is the error variable. The aim of regression analysis is to produce a reasonable approximation to the unknown response function 'm'.[6]

In the linear regression, the mean response of Y is assumed to be a linear function of predictor variables i.e. $m(x) = x\beta$. Parametric regression assumes that the form of the regression function 'm' is known except for some unknown parameters and the shape of the regression function is entirely dependent on the parameters. Often in practice, it is difficult to guess the most appropriate parametric model just by looking at the data and in some cases there may not exist a suitable parametric form to express the regression function. In such situation, the non parametric regression that do not require strong assumptions about the shape of the regression function is very useful. It only assumes that the regression function is a smooth function.

The non parametric regression techniques require only qualitative information about 'm' and the philosophy is let the data speak for itself concerning the actual form of the regression function.[6] Even though over the last few decades the logistic regression model is popularly used in many fields, as a standard analysis, in the present study we propose non parametric regression and make comparison with the existing methods.

The non parametric regression model is $Y_t = m(x_t) + \varepsilon_t$ where (x_t, y_t) $t = 1\ldots n$ $(x_t \le x_{t+1})$ be a random sample from a bivariate population having density $f(x, y)$, $m(x) = E(Y \mathrm{I} X = x)$ is the unknown regression function defined on the interval [0, 1] and ε_t are independently and identically distributed normal random errors with mean 0 and variance $\sigma^2 < \infty$.

Several non parametric methods have been proposed for estimating the regression function $m(x)$, among these popular methods are kernel smoothers proposed by Gasser Muller (1979), Nadaraya and Watson (1964) and local linear regression smoother by Stone (1977). The regression estimators considered in this paper is based on fitting of kernel weighted local linear regression function Ruppert and Wand (1994). The Kernel weighted local linear regression estimator of $m(x)$ is the solution for a_0 to the following problem:

Minimize $L(a,b) = \sum[y_j - a_0 - a_1 (x_j - x)]^2 k[(x - x_j)/h]$ where k is a kernel function and h is a bandwidth.[7]

3. Results

Among 498 subjects enrolled for the study, the minimum age was 18 years with the maximum of 29 years. Majority of the students 125 (25%) were 20 year old with the mean age of 21 ± 2 years. The proportion of females 273 (54.8%) was more than the males 225 (45.2%).

From Table 1 it is observed that majority of the students are from Karnataka accounting for 292 (58.6%), followed by Kerala 129 (25.9%).

The analysis of baseline data is represented in Tables 2–4. Table 2 revealed that the body mass index of the study individuals was found to increase with the increase in age, except in subjects aged 20 years. Also the 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. The mean cholesterol was found to be highest among individuals of more than 25 years followed by 24 years. Also the mean HDL cholesterol was highest and mean LDL cholesterol was least among 25 year old subjects. The highest systolic blood pressure of 119.3 ± 12.2 was found among 23 year old subjects. However the mean risk is less than 1 till the age of 25 years but sharply increased to 1.93 for age >25 years, indicative of harmful effects of these factors.

The Table 3 depicts that only 14.6% of the study individuals used any form of tobacco. Majority of the tobacco users were males 60 (26.6%) compared to females 13 (4.8%). Use of tobacco was found to be significantly different among males and females ($p < 0.001$). Also very high significant difference was found in the duration of physical activity (per day) among males and females with the duration of more than 20 min was found more common in males 170 (75.6%) when compared among females 144 (52.5%). The consumption of Trans fatty acid (per week) was quite common among girls, however it was not statistically significant ($p > 0.05$). When Bivariate analysis was performed, the adjusted risk estimates were assessed by sex for the significant variables as mentioned in Table 3. Males were found to have marginally higher risk than

Table 1 – State wise distribution of the study subjects on the basis of their native place.

State	Total number (%)	Males number (%)	Females number (%)
Karnataka	292 (58.6)	123 (54.7)	169 (61.9)
Dakshina kannada	143 (48.9)	54 (43.9)	89 (52.6)
Others*	149 (51.1)	69 (56.1)	80 (47.4)
Kerala	129 (25.9)	58 (25.7)	71 (26.0)
Gujarath	18 (3.6)	16 (7.1)	02 (0.7)
Andhra Pradesh	15 (3.0)	08 (3.6)	07 (2.6)
Others**	44 (8.9)	20 (8.9)	24 (8.8)
Total	498 (100)	225 (45.2)	273 (54.8)

Others* include students from other parts of Karnataka Others** include students from other parts of India and certain NRI's.

Table 2 – Age wise distribution of the study subjects on the basis of the known coronary risk factors.

Age (years)	No.	BMI (kg/m²)	Cholesterol (mg/dl)	HDL (mg/dl)	LDL (mg/dl)	Systolic blood pressure (mm of Hg)	Risk***
		Risk factors (mean ± SD)					
18	56	22.6 ± 3.5	179.6 ± 37.3	52.0 ± 7.6	100.4 ± 24.6	115.7 ± 12.2	0
19	45	23.4 ± 4.4	185.16 ± 46.3	51.3 ± 9.8	100.1 ± 26.9	111.9 ± 11.5	0
20	125	21.9 ± 2.8	178.7 ± 35.2	50.1 ± 10.7	107.4 ± 27.0	114.5 ± 15.0	0.50
21	97	22.5 ± 3.1	176.9 ± 37.7	51.6 ± 8.9	102.8 ± 27.7	111.7 ± 13.4	0.21
22	67	23.1 ± 4.1	183.5 ± 45.5	50.8 ± 9.3	105.3 ± 28.8	117.3 ± 12.3	0.54
23	57	23.1 ± 3.6	180.7 ± 35.6	50.2 ± 9.7	105.6 ± 25.6	119.3 ± 12.2	0.19
24	25	23.4 ± 3.0	191.6 ± 34.8	49.6 ± 8.8	102.7 ± 29.1	113.6 ± 11.0	0.28
25	14	25.1 ± 3.1	181.2 ± 40.0	53.43 ± 6.1	99.6 ± 33.6	114.7 ± 11.5	0.57
>25	14	26.6 ± 3.5	218.4 ± 43.1	47.5 ± 11.4	110.5 ± 22.0	116.8 ± 10.4	1.93

*** represents Estimated risk of CVD in next 10 years.

females e.g. for Non tobacco users (Males odds ratio [OR] = 0.051, 95% confidence interval [CI] = 0.02–0.126 versus Females OR = 0.006, CI = 0.001–0.028) and Body mass index (Males OR = 1.607, CI = 1.370–1.886 versus Females OR = 1.276, CI = 1.139–1.429).

In Table 4, significant difference ($p < 0.05$) in the mean body mass index was found among males (23.4 ± 3.4) as compared to females (22.5 ± 3.7). However, there was no much significant difference in other biochemical characteristics.

The Quality of the model is judged by examining the overall significance of the model. This can be done by testing the hypothesis (H_0: $\beta = 0$) which means that all the regressors are insignificant on CVD. If we reject this hypothesis, it indicates that the regressors have got significant effect on CVD. The analysis indicates that at step5**** the regressors included in the model are very highly significant ($p < 0.001$).

Table 5 revealed, Hosmer and Lemeshow test***** at step5 the significance level of $\alpha = 0.05$, we accept the null hypothesis meaning the model is adequate to the data. This was tested against an alternative hypothesis that the model is not adequate to the data.

3.1. *Risk factors for cardio vascular disease*

The potential risk factors associated with cardio vascular disease (CVD) reported by the respondents are discussed using multivariate stepwise logistic regression analysis (Table 6).

Table 3 – Sex wise comparison of known life style risk factors ($n = 498$).

	Total *n* (%)	Males number (%)	Females number (%)	*p* value
1. Consumption of tobacco				
Yes	73 (14.6)	60 (26.6)	13 (4.8)	<0.001
No	425 (85.4)	165 (73.3)	260 (95.2)	
2. Consumption trans fatty acid (per week)				
Less than once	25 (5.02)	7 (3.1)	20 (7.0)	>0.05
1–4 times	106 (21.3)	45 (20.0)	61 (22.0)	
5–10 times	282 (56.6)	140 (62.2)	142 (52.0)	
More than 10 times	85 (17.1)	33 (14.7)	52 (19.0)	
3. Physical activity per day				
<10 min	88 (17.7)	28 (12.4)	60 (21.9)	<0.001
10–19 min	96 (19.3)	27 (12.0)	69 (25.1)	
20–59 min	292 (58.6)	152 (67.6)	140 (51.0)	
60 min +	22 (4.4)	18 (8)	4 (1.5)	

3.1.1. *Age*

It is evident that age is at one point three times risk on contributing significantly to the risk of developing cardio vascular disease (CVD) in the next 10 years (odds ratio [OR] = 1.32, 95% confidence interval (CI) = 1.062–1.660). The odds ratio for age indicates that every unit increase in age is associated with 32% increase in the odds of developing CVD in future ($p < 0.05$).

3.1.2. *Present stay*

The analysis reveals from the logistic regression that students residing at home had decreased odds of developing CVD (OR = 0.27, CI = 0.092–0.799) as compared to students who were residing in hostel or as a paying guest ($p < 0.05$).

3.1.3. *Tobacco use*

Also the present study depicts that non-tobacco users (OR = 0.041, CI = 0.015–0.107) had lesser risk of CVD as compared to the tobacco users ($p < 0.001$).

Further it was observed that BMI (OR = 1.42, CI = 1.244–1.631) and LDL cholesterol (OR = 1.05, CI = 1.021–1.055) had significantly increased odds of contributing towards the development of CVD ($p < 0.001$).

Even though there were many covariates for CVD such as age, gender, place of stay, tobacco use, diet, consumption of trans fatty acid per week, body mass index, total cholesterol, HDL, LDL cholesterol, fasting blood sugar, systolic blood pressure and triglycerides, while using Forward Wald stepwise logistic regression only the significant covariates were included in the model. Further, the female students had decreased odds of developing CVD as compared to their male counter parts, and it was excluded from the model as it was insignificant.

Table 7 represents results after performing stratified analysis by multiple logistic regression separately for males and females after adjustment for confounding variables. The significant determinants identified for CVD in the next 10 years were: Non Smokers (Females OR = 0.005, CI = 0.000–0.061 versus Males OR = 0.071, CI = 0.022–0.232), Body mass index (Females OR = 1.281, CI = 1.005–1.632 versus Males OR = 1.572, CI = 1.315–1.879) and LDL cholesterol (Females OR = 1.063, CI = 1.024–1.103 versus Males OR = 1.034, CI = 1.013–1.055). The age group considered in this study is a small segment (18–29 years) and hence it is difficult to interpret the risk on this young age on the basis of gender for describing trends of CVD risk.

Table 4 – Sex wise comparison of bio chemical risk factors.

Risk factors	Males ($n = 225$)			Females ($n = 273$)			p value
	Min	Max	Mean ± SD	Min	Max	Mean ± SD	
Body mass index	16.8	35.1	23.4 ± 3.4	15.1	40.6	22.5 ± 3.7	$p < 0.05$
Cholesterol	97	320	181.5 ± 41.2	98	298	181.5 ± 41.2	$p > 0.05$
HDL	15	91	51.04 ± 10.0	36	76	50.6 ± 9.0	$p > 0.05$
LDL	44	185	104 ± 28.2	43	188	104.5 ± 26.3	$p > 0.05$
Fasting blood Sugar	60	134	89 ± 11.2	63	122	90.1 ± 9.6	$p > 0.05$
Systolic blood pressure	92	148	116.1 ± 13.4	92	148	115.6 ± 12.9	$p > 0.05$
Triglyceride	42	221	88.6 ± 32.8	39	192	91.84 ± 32.2	$p > 0.05$

Logistic regression explains the contribution of different regressors to cardiac risk score whereas non parametric regression gives additional information on a smooth curve indicating a specific age group having higher risk. From the smooth curve we identify the age group which has greater risk as compared to other age groups. Hence higher risk of developing cardio vascular disease is observed among individuals in the age group between 21 and 22 years ($x = 0.62$–0.78), apart from 25 and above years ($x = 0.95$–1) (Fig. 1).

4. Discussion

In the present study, it was observed that majority of the students hailed from Karnataka accounting for 292 (58.6%). Nearly 143 (48.9%) were from Dakshina Kannada district of which 89 (52.6%) were females. Moreover 25% were 20 years old with the mean age of 21 ± 2 years. The overall consumption of tobacco use among the study participants was 73 (14.6%) and it was higher in males 60 (26.6%) than in females 13 (4.8%) and the same is statistically significant ($p < 0.001$). Similar findings were reported in studies conducted in North India[8] where the prevalence of smoking was 11.8% in males and 1.4% in females.

Table 5 – Testing model's Quality and adequacy.

Chi Square	Degrees of freedom	Significance
184.352	5	0.000****
3.086	8	0.929*****

The significance of **** implies ($p< 0.001$) and ***** implies ($p> 0.05$).

Table 6 – Strength of risk factors in the causation of cardio vascular disease (CVD).

Factors	Crude OR (95% CI)	Significance
Age	1.32 (1.062–1.660)	0.013
Present stay		
Hostel and others***	1.0	
Localities	0.27 (0.092–0.799)	0.018
Tobacco		
Smokers	1.0	
Non smokers	0.041 (0.015–0.107)	0.000
Body mass index	1.42 (1.244–1.631)	0.000
LDL cholesterol	1.05 (1.021–1.055)	0.000

CI = confidence interval, ***includes respondents staying as paying guests outside the college campus.

Our study demonstrated that the 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. These findings were comparable to study done by Rajeev Gupta et al[8] that in the age group of 20–29 years, overweight/obesity was observed in 11.5% males and 7.4% females. This difference possibly due to variation in the BMI cut off points, where in earlier study the limit for over weight was 25–30 kg/m^2, where as in our study the cut off point is 23–25 kg/m^2. Also in the present study, when the risk estimates were assessed separately on the basis of gender it was observed that significant difference exists in prevalence of BMI among males (23.4 ± 3.4 kg/m^2) and females (22.5 ± 3.7 kg/m^2) and ($p < 0.05$) with (Females OR = 1.281, CI = 1.005–1.632 Males OR = 1.572, CI = 1.315–1.879) indicating higher risk among males. Similar findings were observed in the study by Desigamani Kanniyappan, Priya Kalidhas and Rita Mary Aruna[9] among South Indian adults shows BMI significant ($p < 0.001$).

In this study, the overall prevalence of hypertension (systolic blood pressure of >140 mmHg) was as low as 5 (1%) and was comparably higher among females 3 (1.1%) than males 2 (0.8%) and ($p > 0.05$). It contradicts with the study done by Latheef SA, Subramanyam G[10] found that the prevalence of hypertension in males was 8.9% and 7.69% among females in the age group of 20–30 years of age. Also study conducted by Gupta R[11] among adults of more than 20 years of age reveals that 36.4% of males and 37.5% of females were hypertensive.

5. Limitations

This study provides baseline information and a start to a debate on CVD risk which could help with possibly creating awareness among younger generations. We could not take

Table 7 – Sex – Specific multiple logistic regression model in the causation of cardio vascular disease (CVD).

Determinants	Females	Males
	Adjusted odds ratio (95%CI)	Adjusted odds ratio (95%CI)
Tobacco		
Smokers	1.00	1.00
Non smokers	0.005 (0.000–0.061)	0.071 (0.022–0.232)
Body mass index	1.281 (1.005–1.632)	1.572 (1.315–1.879)
LDL cholesterol	1.063 (1.024–1.103)	1.034 (1.013–1.055)

CI = confidence interval.

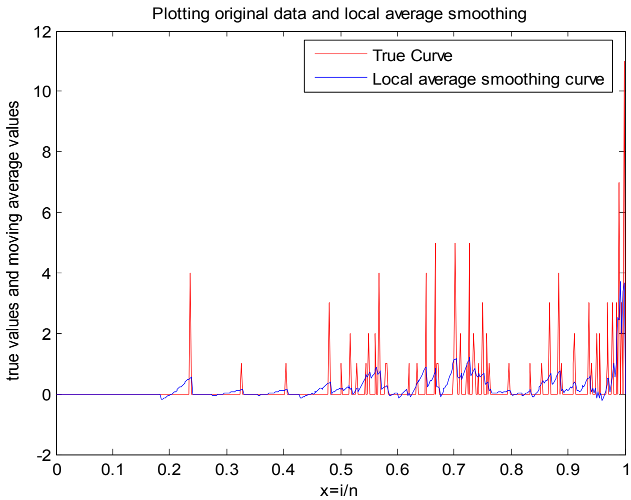

Fig. 1 – Non parametric smoothing by fitting of kernel weighted local regression function for CVD risk scores.

more information because of lack of available data from records. Also the exposure to some risk factors such as physical activity, smoking, alcohol consumption is not precisely defined .Further due to the stigma associated with alcohol and tobacco consumption, the stated figures may not be factual hence requires further elucidation. Salt intake per day, quantity and duration is also a known risk factor. Even though it is known that India suffers the highest loss in potentially productive years of life due to deaths from cardio vascular diseases in people aged 35–64 years, further exploration may be required as the present study is confined only to younger population.

6. Conclusion

The study highlights the known risk factors of cardio vascular disease (CVD) among young population. The need of the hour, is to educate the younger generation regarding the prevention of CVD. The faculties in the educational institutions should highlight regarding life style modifications so that future epidemic on CVD can be reduced. This study is exclusively based on secondary (record based) and retrospective data rather than any validation with live, primary data. Hence, future studies with better sample sizes, particularly study subjects with <25 and >25 years could provide even better insights to prove the associations of various predictors of CVD among young students.

Conflicts of interest

All authors have none to declare.

Acknowledgments

The authors gratefully acknowledge the assistance and co operation of the Head of the Department and staff, particularly the Department of Community Medicine of A.J. Institute of Medical Sciences, India. Also we would like to thank the referee for his valuable comments and suggestions which led to an improvement on an earlier version of this paper.

REFERENCES

1. Fitzpatrick RM. Sociology as applied to medicine. In: Pattick DL, Scamber G, eds. *Social and Changing Patterns of Disease*. 2nd ed. Balliere Tindall; 1986.
2. World Health Organization. *Prevention of Coronary Heart Disease: Report of a WHO Expert Committee*. World Health Organ Tech Rep Ser 678. Geneva: World Health Organization; 1982.
3. Sathe PV, Sathe AP. *Epidemiology and Management for Health Care for All*. 2nd ed. Mumbai: Popular Prakashan Pvt. Ltd; 2003.
4. Bhalwar Rajvir. *Public Health and Preventive Medicine for the Indian Armed Forces the RED BOOK*. 8th ed. Pune: Department of Community Medicine Armed Forces Medical College; 2008.
5. Hosmer David W, Lemeshow Stanley. *Applied Logistic Regression*. 2nd ed. New York: Wiley Series in Probability and Statistics; 2000.
6. Hardle W. *Applied Non Parametric Regression*. Cambridge University Press; 1992.
7. Tsay RS. *Analysis of Financial Time Series*. John Wiley and Sons; 2005.
8. Gupta R, Misra A, Vikram NK, et al. Younger age of escalation of cardiovascular risk factors in Asian Indian subjects. *BMC Cardiovasc Disord*. 2009;9.
9. Kanniyappan Desigamani, Kalidhas Priya, Aruna Rita Mary. Age, gender related prevalence of cardiovascular risk factors in overweight and obese south Indian adults. *Int J Biol Med Res*. 2011;2:513–522.
10. Latheef SA, Subramanyam G. Prevalence of coronary artery disease and coronary risk factors in an urban population of Tirupati, India. *Indian Heart J*. 2007;59:157–164.
11. Gupta R, Gupta VP, Sarna M, Bhatnagar S, Thanvi J, Sharma V. Prevalence of coronary heart disease and risk factors in an urban Indian population: Jaipur Heart Watch-2. *Indian Heart J*. 2002;54:59–66.