

Development of a cost-effective CVD prediction model using lifestyle factors. A cohort study in Pakistan

Parveen Naheeda¹, Khan Sharifullah¹, Shah Saeed Ullah², Abbas Muhammad Azeem³, Younis Shahzad¹, Waqar Kinza²

1. School of Electrical Engineering and Computer Science, National University of Sciences and Technology (NUST), Islamabad, Pakistan.
2. Department of Cardiology, Shifa International Hospital Islamabad, Pakistan.
3. University Institute of Information Technology, PMAS, Arid Agriculture University, Rawalpindi, Pakistan.

Abstract

Background: Cardiovascular diseases (CVD) such as hypertension and ischemic heart diseases cause 35 to 40% of deaths every year in Pakistan. Several lifestyle factors such as dietary habits, lack of exercise, mental stress, body habitus (i.e., body mass index, waist), personal habits (smoking, sleep, fitness) and clinical conditions (i.e., diabetes, dyslipidemia and hypertension) have been shown to be strongly associated with the etiology of CVD. Epidemiological studies in Pakistan have shown poor adherence of people to healthy lifestyle and lack of knowledge in adopting healthy alternatives. There are well validated cardiovascular risk estimation tools (QRISK model) that can predict the probability of future cardiac events. The existing tools are based on laboratory investigations of biochemical test but there is no widely accepted tool available that predicts the CVD risk probability based on lifestyle factors.

Aims: Aim of the current study was to develop alternative CVD risk estimation model based on lifestyle factors and physical attributes (without using laboratory investigation) using QRISK model as the gold standard.

Study Design: Clinical and lifestyle data of one hundred and sixty subjects were collected to formulate a regression model for predicting CVD risk probability.

Methods: Lifestyle factors as independent variables (IV) include BMI, waist circumference, physical activities (stamina, strength, flexibility, posture), smoking, general illnesses, dietary intake, stress and physical characteristics. CVD risk probability of QRISK Intervention computed through clinical variables was used as a dependent variable (DV) in present research. Chronological age was also included in analysis in addition to selected lifestyle factors. Regression analysis, principal component analysis and bivariate correlations were applied to assess the relationship among predictor variables and cardiovascular risk score.

Results: Chronological age, waist circumference, BMI and strength showed significant effect on CVD risk probability. The proposed model can be used to calculate CVD risk probability with 72.9% accuracy for the targeted population.

Conclusion: The model involves only those features which can be measured without any clinical test. The proposed model is rapid and less costly hence appropriate for use in developing countries like Pakistan.

Keywords: Chronological age; cardiovascular diseases; disease risk factor; modifiable lifestyle factors; risk score.

DOI: <https://doi.org/10.4314/ahs.v20i2.39>

Cite as: Naheeda P, Sharifullah K, Ullah SS, Azeem AM, Shahzad Y, Kinza W. Development of a cost-effective CVD prediction model using lifestyle factors. *A cohort study in Pakistan Afri Health Sci.* 2020; 20(2): 849-859. <https://doi.org/10.4314/ahs.v20i2.39>

Introduction

Cardiovascular diseases (CVD), such as hypertension and ischemic heart diseases cause 35 to 40% of deaths every year in Pakistan¹. Several lifestyle factors, such as dietary

habits², lack of exercise, mental stress, body habitus (i.e., BMI, waist), personal habits (i.e. smoking, sleep, fitness) and clinical conditions (i.e., diabetes, dyslipidemia and hypertension) have been shown to be strongly associated with the etiology of CVD^{3,4}. Epidemiological studies in Pakistan has shown poor adherence of people to healthy lifestyle and/or lack of knowledge of adopting alternatively healthy alternatives⁴.

There are well validated cardiovascular risk estimation tools^{5,6} that can predict the probability of future possible cardiac events. Results of several longitudinal studies were reported for the validity of the existing prediction mod-

Corresponding author:

Sharifullah Khan,
School of Electrical Engineering and Computer
Science, National University of Sciences and
Technology (NUST), Islamabad, Pakistan.
Email: sharifullah.khan@seecs.edu.pk

els for calculating CVD risk probability. The Framingham Heart Study is still considered outstanding research since its start in 1948 and continues to date with new versions⁶. It was a remarkable advancement in the risk assessment, understanding the causes and complexity of cardiovascular disease (CVD) and the primary prevention of cardiovascular morbidity and mortality⁶. Framingham study targeted only white male cohorts without considering ethnicity⁷ and ischemic stroke. To address the impact of gender difference in cardiovascular risk prediction, the Reynolds risk score equation was based on the same factors of Framingham in addition to a family history of cardiovascular disease and lifestyle factors⁸. Pooled Cohort Risk Equation is the clinical risk assessment score that was recommended in the year 2013 by the American College of Cardiology (ACC) and American Heart Association (AHA) joint guidelines for the assessment of cardiovascular risk⁹. It was developed to update guidelines on cholesterol, blood pressure, and overweight/obesity with the optimal clinical cardiac risk assessment model⁹.

QRISK (Q Research Cardiovascular Risk Algorithm) Intervention^{5,10} has been developed by the University of Nottingham (UK) to predict the 10-year probability of developing diabetes and cardiovascular diseases in an individual. It computes the risk of the diseases based on clinical variables, such as blood pressure, total cholesterol/HDL ratio, body mass index (BMI), family history of diabetes and cardiovascular disorders, presence of atrial fibrillation and rheumatoid arthritis. It is a well-validated algorithm⁵. The existing tools such as QRISK are based on laboratory investigations of biochemical tests. However, there is no widely accepted tool available that predicts the CVD risk probability based on lifestyle factors. The aim of the current study was to develop an alternative CVD risk estimation model based on lifestyle factors and physical attributes i.e., without using laboratory investigation. The cardiovascular risk score (probability and risk category) returned by the QRISK intervention was used as the gold standard to compare the prediction

made by the proposed model. The quality of any model is dependent on the quality of the instrument through which its output is compared. In the present work, the results of the proposed model are compared with the QRISK probabilities. The proposed model is rapid and less costly hence appropriate for use in developing countries like Pakistan.

Methods

The development of CVD prediction model using lifestyle factors was initiated by collecting data against clinical and lifestyle variables of the participants. Figure 1 shows the methodology followed for the development and comparison of the proposed model. A campaign through pamphlet, posters and social media (such as WhatsApp and Facebook) was conducted over a period of one month for the recruitment of volunteer participants. The eligibility criteria for selection of a subject was decided i.e. age > 25 years. The data was collected using a case report form (CRF) during a two days of medical camp held at National University of Science and Technology (NUST). The CRF contains two sections other than demographic data, one for lifestyle factors and the other for QRISK variables. The demographic data includes the date of birth, gender and ethnicity. Lifestyle factors include all behaviours and habits a participant is following in his routine life. Only those lifestyle factors were included that have a reported effect in terms of CVD risk and are modifiable. The medical officer, who is a certified physician at the NUST has collected the data of these lifestyle factors. Every participant has been assigned a unique case number at the time of his enrollment. All further procedures were carried out following the assigned case number. This hides the identity of a participant in the data. All the QRISK variables are clinical tests that were collected and recorded on the prescribed CRF by the trained medical representatives/staff of a registered pharmaceutical company. One hundred and sixty male participants have volunteered. However, six participants were rejected due to under-age criteria of selection.

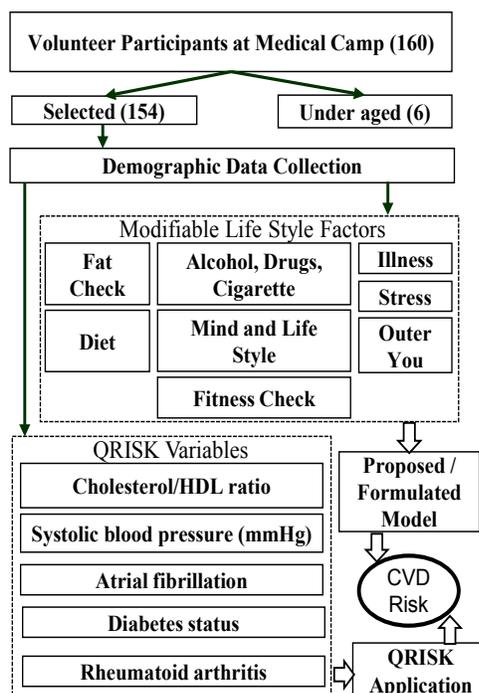


Figure 1: Methodology followed for the development of CVD risk prediction model

The collected data for the QRISK variables of each participant was fed into an online application of the QRISK. The application returns a CVD risk score (probably and category of risk i.e. high, medium or low). The proposed prediction model is a regression predictive model that performs the task of approximating a mapping function from input variables (lifestyle factors of a participant) to an output variable (the CVD risk labels returned by the QRISK). The accuracy of the proposed model can be validated by measuring how well the proposed model predicts the CVD risk using lifestyle factors. The new model was developed through multiple linear regressions by using QRISK Intervention probability as a dependent variable and all identified lifestyle risk factors as independent variables. The validity of the formulated model was carried out by comparing the results of QRISK intervention score with the results of the formulated model.

Identification of modifiable lifestyle factors

Currently obesity due to sedentary lifestyle has become one of the major concerns in developing countries. Lifestyle has many impacts on an individual's health. Unhy-

gienic food and not doing any physical activity during the day leaves very negative impacts on health which result in different diseases like metabolic problems, stomach issues, diabetes, chronic and heart diseases¹¹. Dietary intakes have an association with chronic illnesses and it has been documented in many studies¹²⁻¹⁵. Availability of fast foods is considered to be linked with sedentary, increases consumption of fat which are ultimately associated with chronic conditions¹⁶. Therefore, analysis of diet as a risk factor is a key for chronic diseases, such as diabetes, hypertension and heart diseases¹⁷. These days, primary prevention of chronic diseases emphasizes on modification of lifestyle and diet patterns because diet-related disorders are increasing chronic diseases, such as cardiovascular, diabetes and hypertension¹⁸.

Physically inactive is the fourth leading cause of death in the world and a well-known reason for the development of cardiovascular diseases¹⁹. In general, there are many evidences that exercise contributes a lot to better health and quality of life. Taking breaks during the busy day and performing some physical activities leave positive effects

on health²⁰. According to Disease Control and Prevention²¹, risk of many chronic diseases can be reduced with regularly performing physical activities which keep a person fit.

In 2008, CDC published a set of physical activity recommendations for the people aged six and older. This set contained a total amount of physical activities to achieve a range of health benefits. In 2010, the WHO also published a globally recommended set of physical activities²². There exists some other lesser known literature as well to motivate people about adopting an active lifestyle and performing physical activities.

All these guidelines emphasize on importance of carrying out physical activities for a healthy lifestyle and well-being of people. In developing countries like Pakistan, in spite of public availability of all these guidelines either people, are not aware about these guidelines or unable to understand. As a result, most people in Pakistan do not follow hygienic diet plan and lack of regular exercise also increases disease rate especially diabetes and cardiovascular diseases. Stress and use of alcohol are also modifiable risk factors for cardiovascular diseases²³. Stress contributes to high blood pressure which is ultimately a risk factor for CVD²⁴. Alcohol was excluded from the analysis of this research because all participants were Muslims and none of them was found to use alcohol.

Body mass index (BMI) is usually measured to evaluate obesity/overweight. Addition of waist circumference to BMI is a good predictor of obesity instead of using just BMI measure²⁵.

Usage of Tobacco is associated with risk of cardiovascular diseases. Smoking is termed as the most significant preventable risk factor for cardiovascular disease²⁶. The appearance of a person in relation to diseases prediction was studied in²⁷.

Participants

Case cross over design study design was utilized in our research. Case cross over was used because each participant was considered as a case and his CVD risk was calculated from QRISK and through the proposed model. For further future validation of the proposed model, risk can be calculated after some period of time by advising some lifestyle changes and observing their effects. The research study was approved from the ethics board of National University of Sciences and Technology, Islamabad. Participants were recruited with the help of campaign through pamphlets; poster and emails were sent to all departments of the university to invite faculty and staff to participate in the study. A medical camp was held in NUST University for two days to recruit participants to collect study data. One hundred and sixty (160) participants had attended the camp, six were excluded from the study due to lack of compliance with the age criteria (age is greater than 25 years). Participants came from cohorts of different regional background from all over Pakistan because NUST is located in the federal area where employees and students belong to various regions of Pakistan. The mean age was 40.9 years whereas the age range was in between 26 and 67 years.

Data Gathering

A section of the case report form (CRF) was prepared that contains all cardiovascular risk factors as mentioned in Section 2.1. The other section of the CRF contain all attributes of the QRISK Intervention. The staff of Pharmaceutical Company Pharmevo (<https://www.pharmevo.biz/>) took all medical/blood tests which were later used to calculate cardiovascular risk probability through QRISK Intervention. The staff of NUST Medical Center was responsible for monitoring the physical activities to fill the CRF.

In this research, a scoring system was used to assign values to CVD risk factors. Positive values are assigned to values which affect negatively in a person's health and negative scores have been assigned which leave positive effect on a person's health w.r.t cardiovascular diseases.

Table 1: General statistics of the participants

Variable	Mean	Std. Deviation	
Weight	74.479 kg (48.4 – 107.4)	12.307	
Height	1.687 m (1.48 – 1.86)	0.066	
BMI	26.308 kg/m ² (18.6 – 54.66)	3.449	
Waist Circumference	36.597 inches (26 – 48)	3.981	
Chronological Age	40.941 years (26 – 67)	11.388	
Systolic blood pressure	129.026 mm Hg (97 – 185)	14.669	
Cholesterol/HDL ratio	5.064 mmol/L (4.14 – 7.86)	0.6770	
QRISK Probability	13.346 (0 – 100)	20.887	
QRISK Probability Clusters	High (N=31) 45.841 (20.7 – 100)	Medium (N=26) 14.619 (10.1 – 19.8)	Low (N=97) 2.594 (0 – 9.8)

Fat Check: BMI was calculated from weight (kg) and height (m) by using standard BMI equation. The waist Circumference was calculated in inches. Waist was measured by using waist measurement tape. Participants were asked to stand straight and breathe out slowly. As provided in Table 1, the mean waist circumference was 36.597 inches (26 – 48) and mean BMI was 26.308 kg/m².

Fitness Check: To test stamina, each participant was asked to step up and step down repeatedly on a 12 inches high stair for three minutes. Immediately after the stamina test, a Likert scale rating was used to collect participant's data. The press up test was used to assess the body strength. Participants were asked to do press ups and continue until they get exhausted (muscles begin to shake). A Likert scale was used to record the number of press-ups performed by a participant. The Sit and Reach Test was used to assess flexibility. Data of the participants were recorded using a Likert scale.

Alcohol, Drugs, Cigarette: The consumption of alcohol, drugs and cigarettes were measured using a Likert scale. It was measured on the basis of an individual's consumption and marked from possible values of excessive, moderate, light, ex-drinker or ex-smoker and none. Among participants 122 were non-smokers, 6 were moderate smokers, 7 light-smokers, 3 excessive and 16 were ex-smokers.

Illness: The intensity of general illnesses was recorded by a physician using a Likert scale. For the present study, illnesses that were recorded are cold, flu, Gskin/

Eye Infection, Cold Sores (lips / nose), Mouth Ulcers, Bleeding gums, Constipation, Heartburn, Backache and general aches/pains. In addition to the above general illnesses if any major disease is already diagnosed in that individual then that information was also used to calculate health check score. If the participant was suffering from diagnosed conditions of blood pressure, diabetes, heart attack, cardiac surgery then an additional score (2) was added against each condition. An additional score (1) was added for in case a person has a family history of chronic conditions of these major diseases, angina or coronary disease but nheart attack.

Diet: Consumption of diet was measured by computing score of food items, i.e. participant's daily intake. Different regular food items were assigned positive and negative values by a cardiologist ranging between -1 to +2. Food items known for increased CVD probability were assigned the highest positive value as they affect CVD more positively. Similarly, daily intake of food items such as fruits, vegetable etc. was assigned the lowest negative values.

Mind and Lifestyle: Routine activities were recorded through a questionnaire. Discrete answers were collected against several routine activities, such as, daily walk, relaxing time, working hours, driving time, sleeping condition and frequency of having sex.

Stress: Data about stress state caused by different conditions, such as anger, isolation, anxiousness, nervousness and depression were collected using a Likert scale questionnaire.

Outer You: The physical appearance of the participants was examined by a practitioner. Examination results of teeth, eye colour, complexion, skin and hairs were recorded using a Likert scale.

QRISK: The cardiovascular risk score of QRISK intervention was calculated for every participant using QRISK online service. Table 1 shows the mean values of variables that are required for QRISK probability computation. A total of twelve participants were having diabetes. Twenty-one has reported blood pressure (BP) issue. The mean value of recorded systolic BP was 129.026 mm Hg and 5.064 mmol/L Cholesterol/HDL ratio. Cholesterol/HDL ratio was computed by dividing cholesterol number with the HDL cholesterol number. The QRISK probability was computed using all these required variables for every individual. The mean QRISK probability was 13.346 with standard deviation 20.887. QRISK also assign categorical labels to the computed probabilities as low, medium and high. Among all participants, 31 has high CVD risk probability with a mean value of 45.841. Twenty-six were assigned a medium CVD risk label with a mean value of 14.619. Most of the participants were in low CVD risk category as present participants is of young age (40.9 mean age). The low CVD risk category has a mean probability value of 2.594.

Model Formulation

In deriving the new cardiovascular risk assessment model based on modifiable lifestyle factors, the data from QRISK intervention was used as the reference. Principal component analysis (PCA), bivariate correlations and Regression analysis were used to assess the relationship among predictor variables (lifestyle factors) and the criterion variable (cardiovascular risk score).

All data were analyzed using IBM SPSS Statistics (Version 20). Probability of cardiovascular diseases was calculated by using QRISK interventions. The outcome variable was the probability of cardiovascular diseases. All data including CVD risk factors, chronological age and outcome variable were normalized. Principal component analysis (PCA) was applied to normalized data for extracting important factors which can represent the whole data. Spearman correlation was applied to check the direction of the relationship of each independent factor with CVD risk probability of QRISK intervention. Spearman correlation was used instead of Pearson as most of the factors were showing the monotonic relationship. Relationship of one independent factor i.e., waist circumference with CVD probability is shown in Figure 2.

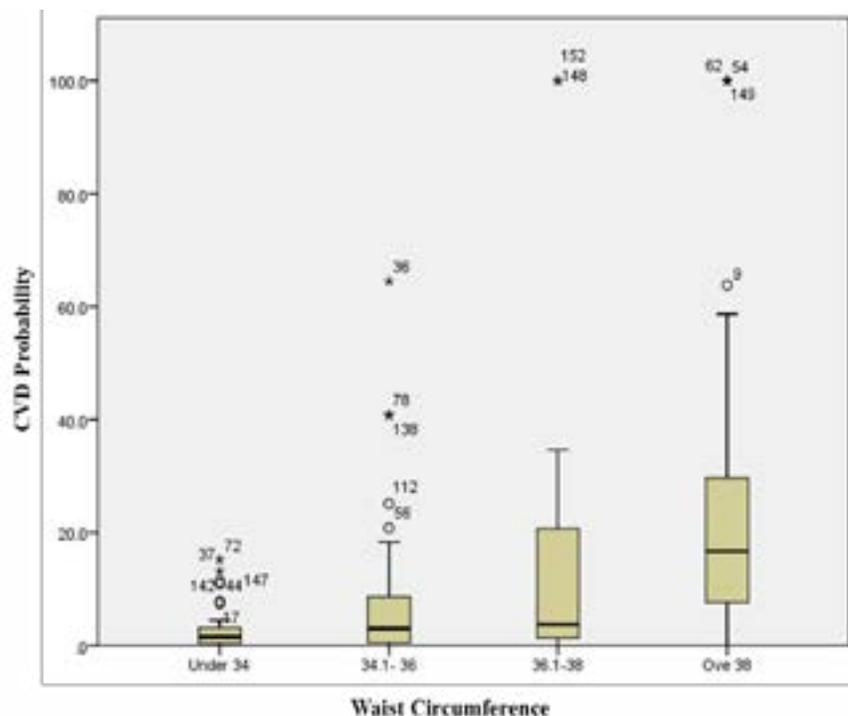


Figure 2: Monotonic Relationship of Waist Circumference with CVD Probability

Dimensionality reduction

Since there were a number of identified lifestyle factors, principal component analysis (PCA) the method was used to reduce the dimensionality of the factors for identification of highly significant ones (shown in Figure 3). Principal component analysis (PCA) was applied to figure

out factors having maximum effect. BMI, Waist Circumference, Age and Strength showed maximum variance and it can be depicted in Figure 3. Rest of factors were clustered into two groups. One Group contains Smoking, Outer, General illnesses, Posture and Stamina. Rest of the four factors i.e., Diet, Mind and lifestyle, Stress and Flexibility were grouped into the second cluster.

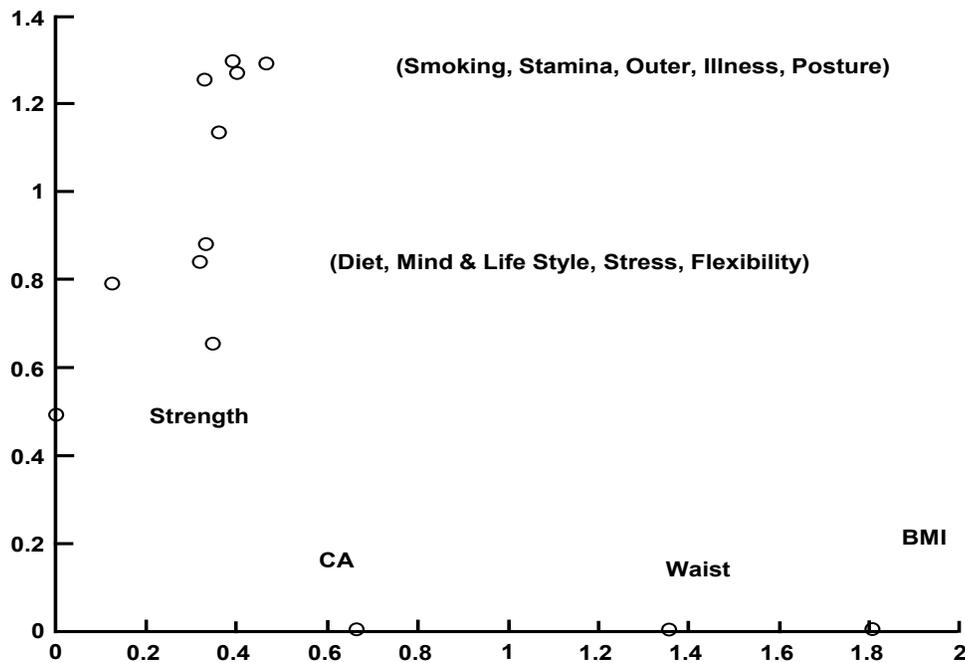


Figure 3: Principal Component Analysis

Correlation

Spearman correlation was applied to find the relationship between the outcome variable (QRISK probability) and all selected continuous and ordinal variables from the PCA. The Selected continuous variables include BMI, Waist Circumference and Chronological Age. Whereas, strength is the only ordinal variable that shows a significant correlation with QRISK probability i.e. Strength ($p < 0.012$). The relationship of the BMI ($R = 0.266$, $p < 0.01$), Waist circumference ($R = 0.507$, $p < 0.01$) and Chronological Age (0.814 , $p < 0.01$) was significantly correlated with the reference variable (QRISK probability score).

Regression

A multiple regression was run to formulate the prediction equation from both selected significant variables and

other modifiable lifestyle factors. The best prediction model achieved was $F(6,95) = 32.33$, $p < 0.005$, $R = 0.729$, $R^2 = 0.717$. Through regression analysis, Smoking ($C = -0.503$, $p < 0.046$), Diet ($C = 0.252$, $p < 0.095$), BMI (-0.027 , $p < 0.833$), Waist ($C = 2.149$, $p < 0.018$), Chronological Age ($C = 5.045$, $p < 0.000$), Strength ($C = 0.154$, $p < 0.221$) with overall constant value equal to -10.358 have resulted. The unstandardized coefficient value of a predictor variable indicates the variance in the dependent variable if all other predictors or the independent variables are held constant. For example, the effect of smoking is -0.503 . The above model summary represents that CVD risk can be measured up to 72.9% accuracy by including/span>the six most significant predictor variables. The model involves only those features which can be measured without any clinical test.

Formulated Model

Using the collected data containing lifestyle factors of one hundred and fifty-four participants and comparing it with the QRISK risk score, the statistical analysis yielded the derivation of the following cardiovascular risk assessment model:

$$\text{Regression(Probability)} = -10.358 + (-0.027 * \text{BMI}) + (2.149 * \text{WAIST}) + (0.154 * \text{Strength}) + (-0.503 * \text{Smoking}) + (0.252 * \text{Diet}) + (5.045 * \text{Age})$$

Results and Discussion

Purpose of this study was to develop a prediction model which can calculate initial CVD risk probability based on non-clinical variables. In the first step, all modifiable CVD risk factors were identified which could be measured by a common man without any laboratory test. As there already exists a prediction algorithm like QRISK Intervention, which predicts CVD risk probability of any individual for the next ten year. It was used as the gold standard in our research. The difference is, our research focuses on non-clinical variables while on the other hand QRISK intervention uses clinical tests as well, e.g., blood pressure, Cholesterol/HDL ratio. For this study, we recruited staff and students of NUST as study participants. People were belonging to different parts of the country because of the national institute. Total of 160 people was

part of the study, but 6 of them were students and excluded due to inclusion criteria of age (age>25). Average CVD risk of the studied population was 13.3% according to QRISK Intervention classification which shows medium risk people were part of the study. Among 154 participants, 97 were ranked low risk, 26 medium and remaining 31 as high risks according to QRISK Intervention. None of the participants was reported to be using drugs/ alcohol. In fact, none of them was chain smoker as well. On average, the studied population was overweight because the average BMI was 26 kg/m². Total of 63 were overweight and 8 were obese who had BMI>30. Waist Circumference is also recommended to use with BMI to check disease association. Average waist circumference was 36.59 inches (92.94 cm) and as male was part of the study so it was also showing healthy measurement because for male risk is increased when the circumference is 94 cm or above it. 86% of participants reported that stamina exercise is easy to perform, and they performed it for consecutive 3 minutes without getting out of breath. Strength test was a little difficult for people and majority people were able to do 7 to 13 press ups in one go. Very few were able to do above 30 press ups. Participants were following a healthy diet routine without any prior medical knowledge. Stress was also at a very low level in the studied population. Outer characteristics of all participants were also in good position. 89.6% of people were reported to have all outer properties in original and good form.

Table 2: Model Comparisons

Maximum Variance Features in PCA				Group#1 Features	Group#2 Features	Model Accuracy
C. Age	BMI	Waist	Strength	Smoking	Diet	72.9%
C. Age	BMI	Waist	Strength	Stamina	Diet	72.5%
C. Age	BMI	Waist	Strength	Outer you	Diet	72.1%
C. Age	BMI	Waist	Strength	Illnesses	Diet	72.1%
C. Age	BMI	Waist	Strength	Posture	Diet	72.3%
C. Age	BMI	Waist	Strength	Smoking	Mind & Life	72.3%
C. Age	BMI	Waist	Strength	Stamina	Mind & Life	72.1%
C. Age	BMI	Waist	Strength	Outer you	Mind & Life	71.7%
C. Age	BMI	Waist	Strength	Illnesses	Mind & Life	71.7%
C. Age	BMI	Waist	Strength	Posture	Mind & Life	71.8%
C. Age	BMI	Waist	Strength	Smoking	Stress	72.3%
C. Age	BMI	Waist	Strength	Stamina	Stress	72.1%
C. Age	BMI	Waist	Strength	Outer you	Stress	71.7%
C. Age	BMI	Waist	Strength	Illnesses	Stress	71.7%
C. Age	BMI	Waist	Strength	Posture	Stress	71.8%
C. Age	BMI	Waist	Strength	Smoking	Flexibility	72.9%
C. Age	BMI	Waist	Strength	Stamina	Flexibility	72.3%
C. Age	BMI	Waist	Strength	Outer you	Flexibility	72.0%
C. Age	BMI	Waist	Strength	Illnesses	Flexibility	72.0%
C. Age	BMI	Waist	Strength	Posture	Flexibility	72.1%

Regressions result showed that any person can predict his CVD Risk Probability up to 72.9% accuracy. Only 1% accuracy degradation was noticed while moving from all independent factors to important factors extracted through principal component analysis. Regression model of all factors showed an accuracy of 73.9%. Final model was achieved by running a regression on all candidate factors.

All variants from two clusters were picked one by one and accuracy was measured. Table 2 shows a combination of independent factors on which regression was executed and accordingly accuracy of each selected model is mentioned. It can be seen maximum accuracy is 72.9%. Both models have five factors in common i.e., Chronological age (C.Age), BMI, waist circumference, strength and smoking. Only the difference is flexibility and diet.

Table 3: Categorial Comparison

Risk Categories	Participants	QRISK Score (Mean)	Proposed Model Score (Mean)	Correctly Predicted Instances	Accuracy
High	31	45.25	77.25	30	0.967
Medium	26	14.62	59.32	16	0.615
Low	97	2.62	35.09	65	0.670

The accuracy of the proposed model was further verified by analyzing scores of both QRISK and proposed model for every individual under risk categories. Risk categories (High, Medium and Low) are defined by the QRISK based on its score value. Table 3 shows the descriptive results of scores computed by both QRISK and proposed model. Results show the proposed model has predicted the High category participants with high accuracy of 0.967. The proposed model has wrongly predicted only one instance in the high risk category. However, this wrongly predicted instance has given a medium risk category label by the proposed model. Data of this instance shows that the individual is taking a healthy diet, that caused the proposed model to compute a low score as compared to the QRISK. The proposed model has predicted equally for both medium and low categories with 0.615 and 0.670 accuracies respectively. Wrongly predicted instances under medium and low categories were label vice versa by the proposed model. Moreover, there was no instance from the medium or low categories that were labelled as high by the proposed model. The overall accuracy achieved the proposed model is 0.721 i.e. 43 instances in total were assigned a different category label as QRISK.

Conclusion

The development of a prediction model based on life-

style factors for cardiovascular disease (CVD) risk prediction is presented here in the present research. The proposed model predicts CVD risk using lifestyle factors as compared to other models using clinical variables. The quality of the proposed model was validated by comparing its results with a well-established risk prediction tool i.e. QRISK. Results revealed that chronological age, waist circumference, smoking and BMI showed a significant effect on cardiovascular diseases risk in the targeted populations. Strength test from physical activities showed a strong effect on the cardiovascular risk score. Different motivational programs can be started in institutes to promote physical activities during work hours to keep people active. Long sitting hours result physically inactive and obesity which ultimately causes heart-related issues. Smartphones are very common these days so different games can be developed to track whole day activities and monitor progress if a person is well-maintaining health activities. As only male participants were part of this study, a large-scale study is required by including both genders to further validate the results. The proposed model is rapid and less costly hence appropriate for use in developing countries like Pakistan.

Conflict of interest

There is no potential conflict of interest in this study.

References

1. S. Akhtar, N. Asghar, "Risk factors of cardiovascular disease in district Swat," *J Pak Med Assoc*, vol. 65, no. 9, pp. 1001–1004, 2015.
2. L. F. Hammad, N. Benajiba, "Lifestyle factors influencing bone health in young adult women in Saudi Arabia," *Afr. Health Sci.*, vol. 17, no. 2, pp. 524–531, 2017.
3. G. C. Onyemelukwe et al., "Cardiovascular risk factors in adult general out-patient clinics in Nigeria: a country analysis of the Africa and Middle East Cardiovascular Epidemiological (ACE) study," *Afr. Health Sci.*, vol. 17, no. 4, pp. 1070–1081, 2017.
4. T. H. Jafar, F. H. Jafary, S. Jessani, N. Chaturvedi, "Heart disease epidemic in Pakistan: Women and men at equal risk," *Am. Heart J.*, vol. 150, no. 2, pp. 221–226, 2005.
5. J. Hippisley-Cox, C. Coupland, J. Robson, P. Brindle, "Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database," *BMJ*, vol. 341, 2010.
6. R. Kones, "Primary prevention of coronary heart disease: integration of new data, evolving views, revised goals, and role of rosuvastatin in management. A comprehensive survey," *Drug Des. Devel. Ther.*, vol. 5, pp. 325–380, Jun. 2011.
7. P. Brindle, A. Beswick, T. Fahey, S. Ebrahim, "Accuracy and impact of risk assessment in the primary prevention of cardiovascular disease: a systematic review," *Heart*, vol. 92, no. 12, pp. 1752–1759, 2006.
8. L. Mosca, E. Barrett-Connor, N. Kass Wenger, "Sex/Gender Differences in Cardiovascular Disease Prevention," *Circulation*, vol. 124, no. 19, pp. 2145–2154, 2011.
9. D. C. Goff et al., "2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines," *J. Am. Coll. Cardiol.*, vol. 63, no. 25 Part B, pp. 2935–2959, 2014.
10. J. Hippisley-Cox, C. Coupland, Y. Vinogradova, J. Robson, M. May, P. Brindle, "Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study," *BMJ*, vol. 335, no. 7611, p. 136, 2007.
11. G. N. Healy, C. E. Matthews, D. W. Dunstan, E. A. H. Winkler, N. Owen, "Sedentary time and cardio-metabolic biomarkers in US adults: NHANES 2003–06," *Eur. Heart J.*, vol. 32, no. 5, pp. 590–597, 2011.
12. L. H. Dekker et al., "A prospective cohort study of dietary patterns of non-western migrants in the Netherlands in relation to risk factors for cardiovascular diseases: HELIUS-Dietary Patterns," *BMC Public Health*, vol. 11, no. 1, p. 441, 2011.
13. D. R. Boniface, M. E. Tefft, "Dietary fats and 16-year coronary heart disease mortality in a cohort of men and women in Great Britain," *Eur. J. Clin. Nutr.*, vol. 56, no. 8, p. 786, 2002.
14. F. B. Hu, E. B. Rimm, M. J. Stampfer, A. Ascherio, D. Spiegelman, W. C. Willett, "Prospective study of major dietary patterns and risk of coronary heart disease in men," *Am. J. Clin. Nutr.*, vol. 72, no. 4, pp. 912–921, 2000.
15. S. Voutilainen, T. H. Rissanen, J. Virtanen, T. A. Lakka, J. T. Salonen, "Low Dietary Folate Intake Is Associated With an Excess Incidence of Acute Coronary Events," *Circulation*, vol. 103, no. 22, pp. 2674–2680, 2001.
16. M. A. Mendez, R. S. Cooper, A. Luke, R. Wilks, F. Bennett, T. Forrester, "Higher income is more strongly associated with obesity than with obesity-related metabolic disorders in Jamaican adults," *Int. J. Obes.*, vol. 28, no. 4, p. 543, 2004.
17. M. Jackson, S. Walker, J. Cade, T. Forrester, J. K. Cruickshank, R. Wilks, "Reproducibility and validity of a quantitative food-frequency questionnaire among Jamaicans of African origin," *Public Health Nutr.*, vol. 4, no. 5, p. 971–980, 2001.
18. R. Wilks et al., "Diabetes in the Caribbean: results of a population survey from Spanish Town, Jamaica," *Diabet. Med.*, vol. 16, no. 10, pp. 875–883, 1999.
19. H. W. Kohl et al., "The pandemic of physical inactivity: global action for public health," *Lancet*, vol. 380, no. 9838, pp. 294–305, 2012.
20. S. J. Marshall, E. Ramirez, "Reducing Sedentary Behavior," *Am. J. Lifestyle Med.*, vol. 5, no. 6, pp. 518–530, 2011.
21. CDC, "2008 Physical Activity Guidelines for Americans," Centres for Disease Control and Prevention, 2008. Online. Available: <https://www.cdc.gov/physicalactivity/basics/index.htm>. Accessed: 01-Jan-2017.
22. W. H. Organization and others, "Global recommendations on physical activity for health," 2010.
23. C. Koopman et al., "Trends in risk factors for coronary heart disease in the Netherlands," *BMC Public Health*, vol. 16, no. 1, p. 835, 2016.
24. R. Kones, "Primary prevention of coronary heart

- disease: integration of new data, evolving views, revised goals, and role of rosuvastatin in management. A comprehensive survey,” *Drug Des. Devel. Ther.*, vol. 5, pp. 325–380, Jun. 2011.
25. I. Janssen, P. T. Katzmarzyk, R. Ross, “Waist circumference and not body mass index explains obesity-related health risk.,” *Am. J. Clin. Nutr.*, vol. 79, no. 3, pp. 379–384, Mar. 2004.
26. T. J. Anderson et al., “2016 Canadian Cardiovascular Society Guidelines for the Management of Dyslipidemia for the Prevention of Cardiovascular Disease in the Adult,” *Can. J. Cardiol.*, vol. 32, no. 11, pp. 1263–1282, 2016.
27. D. A. Gunn et al., “Facial appearance reflects human familial longevity and cardiovascular disease risk in healthy individuals,” *J. Gerontol. A. Biol. Sci. Med. Sci.*, vol. 68, no. 2, pp. 145–152, Feb. 2013.