Risk Modelling and Solution of Cardiovascular Risk Score Problems

The Framingham Heart Study Perspective

Rimpi Datta
Electronics & Communication Engineering
Narula Institute of Technology, Agarpara
Kolkata -700109, India

D. P. Goswami
Department of Mathematics
Narula Institute of Technology, Agarpara
Kolkata – 700109, India

***Abstract*—Framingham Heart Study is a large scale and long term study that predicts the risk for men and women who have not already had a heart attack or any heart disease. This study was carried out in the town Framingham in the United States to develop and validate several risk prediction algorithms for different types of specific Cardiovascular Diseases such as coronary heart disease (CHD), Stroke, heart failure etc. Later this study was extended to evaluate general Cardiovascular Risk Score with minimal set of laboratory tests. But this scoring system suffers from some technical difficulties. We have constructed two synthetic counter examples where in one example the physical descriptions and score evaluating parameters are almost identical but the existing scoring system identifies one of them as a low risk profile and another as an intermediate risk profile. Similarly, the other examples categorize one person in the high risk category and another person in the intermediate risk category despite their almost similar descriptive features. These two examples show that some technical modifications are needed in order to make the algorithm applicable in real scenario and not to produce counter-intuitive results as this chart shows some irrational behavior and generates absurd results at some points that are intuitively unacceptable. In this study we have pointed out four such cases numerically and demonstrated how this problem can be removed with a simple mathematical trick.**

Keywords— ***Cardiovascular disease, risk models, framingham study***

I. INTRODUCTION

The Framingham Risk Score can estimate or predicts future development of the Cardiovascular Disease (CVD). The Framingham Risk Score is one of a scoring system, which is used to measure an individual's probability of developing cardiovascular disease. The number of this scoring system is available in this reference [1]. This CVD scoring systems give us a best prediction of the probability that a person will become the victim of cardiovascular disease within the next 5 or 10 years. They are very useful for both the patient and for the diagnosis to identify the life prediction and to take preventive medical treatment.

Framingham score system is an extraordinary piece of work [1] which required a dedicated long study for several decades. It also has a very strong theoretical foundation based on regression for calculating risk scores. But the final outcome that has been popularized in the tabular form [1] exhibits some problems in the application domain. It suffers from the so called 'boundary problem' and give some erratic results in some cases.

There is a great limitation [2] in the Framingham Risk Score. In the recent data or algorithm we see that risk factors for men and women are assigned according to the different categories like age, total cholesterol, HDL cholesterol, blood pressure, smoking etc [3, 4]. If we go through it we can points out that there is a big problem for risk factors that is if we consider a man at the age of 34 then we see that from FRS his Risk Factor is assigned as 0. After just 12 month in the case of that man his Risk Factor will be assign as 2. There is a huge difference between the Risk Factors of than man's present age and just 12 months after. In this 12 months there is a huge difference between that person's Risk Score. This boundary problem is the limitation of Framingham Risk Score.

Perhaps this problem can be solved in appropriate theoretical setting, but here we have suggested a procedure which is very simple, rational, intuitive and mathematically meaningful and will require no attention in the basic formulation rather. This can be regarded as a minimal add-on.

To eliminate the 'boundary problem' the most natural thing what we can do is to remove the 'sharp boundary' because the sharp boundaries [5] give rise to sharp transitions. If a single real number is assigned against an interval, these types of sharp transitions become unavoidable. So we have assigned the value at the midpoint of the interval because if some property remains invariant over an interval, the midpoint of the interval also satisfies this property and if a single point is to be selected from the interval to satisfy this property, the midpoint is the most faithful one. With this intuitive framework we can now proceed to make table that is constituted of the midpoints of the given intervals along with the score points.

II. RISC SCORE ASSESSMENT TECHNIQUE

A. Framingham Risc Score

Many US doctors [3] point out a person's future development of heart disease and also give the details of the prediction of the risk of heart disease by using a risk assessment which is based on the observations from a long-term study conducted in Framingham [1]. This is called the Framingham risk or the Framingham risk score. The

ISSN: 2278-0181
Vol. 4 Issue 01,January-2015

Framingham risk calculator is used for measurement for men and women who have not already had a heart attack or any heart disease. There are two Framingham Risk Scores, one for men and one for women. This risk score [1] is the risk of having a heart attack or dying from heart disease within the time period of 10 years. It is referred as the followings-

Low risk = 0% to 6% chance

Intermediate risk = 6% to 20% chance

High risk = more than 20% chance

They are useful for taking a prevention in medical treatment, and for patient's education. In the US the Cardiovascular disease (CVD) is the great cause of death and illness. In the year 1948, in the Health Research, the Framingham Heart Study - under the direction of the National Heart Institute (now that is known as the National Heart, Lung, and Blood Institute or NHLBI) – has started an ambitious project. At that time, it has less effect on the heart disease and stroke, but the Cardiovascular Disease had been increasing the death rates since the beginning of the century and had been taken an epidemic form in America. The Framingham Heart Study became a joint project of the NHLBI and Boston University. The main objective of the Framingham Heart Study was to identify the factors and characteristics that contribute to CVD by following its development over a long period of time. To study the pattern of CVD Development [4, 5] the scientists are considered 5,209 men and women between the ages of 30 and 62 from the town of Framingham, Massachusetts, and began for the first round of extensive physical examinations and lifestyle interviews. Then the observations are continued by recruiting different persons for laboratory tests. By very carefully monitoring of the Framingham Study population has been pointed out the CVD risk factors, likes - high blood pressure, high blood cholesterol, smoking, diabetes, and physical inactivity.

B. Outcome of FRS Assessment Study

While finding the Study's observed research results, the NHLBI and the Framingham investigators are enhancing their research into other areas in CVD. Framingham investigators also collaborate with the top researchers in many countries worldwide on projects in stroke and dementia, osteoporosis and arthritis, nutrition, diabetes, eye diseases, hearing disorders, lung diseases, and genetic patterns of common diseases etc. Framingham Heart Study has made more than a half century of research success.

Risk prediction estimates for the risk of various cardiovascular disease outcomes in different time horizons [2, 3] are available as score charts. The risk prediction depends on components like: cardiovascular outcome, population of interest, time horizon and risk factors. From this score profile many diseases can be estimated like: general cardiovascular disease, coronary heart disease, diabetics, hypertension, hard coronary heart disease etc.

III. CASE STUDY

A. Typical Case Study

Suppose the age of A is 39 years. At his age we consider his total cholesterol is 280 mg/dl. His systolic blood pressure is 119 mm hg. and HDL cholesterol is 49 mm./dl. In another case we choose a man whose age is 39 years and 9 months i.e. his age is 39.9 years. At his age the total cholesterol is 279 mg/dl. His systolic blood pressure is 120 mm hg. and HDL cholesterol is 50 mm./dl. Now from the scoring of global assessment for Framingham Risk Score we can get them the points of their Risk Score of the chance for developing the Cardiovascular Disease within next 10 years. To make it more understandable we are making a chart of their

Table I. - Case Study Profile

RISK FACTORS	MAN A	MAN B	SCORE POINTS FOR RISK FACTORS OF MAN A	SCORE POINTS FOR RISK FACTORS OF MAN B
AGE	39	40	2	5
TOTAL CHOLESTEROL	279	280	3	4
HDL CHOLESTEROL	49	50	0	-1
BLOOD PRESSURE (UNTREATED)	119	120	-2	0
SMOKER	YES	YES	4	4
DIABETIC	YES	YES	3	3

Now adding up all the risk points of man A we get

FOR MAN A:
[2+3+0+ (-2)+4+3]=10

FOR MAN B:
(5+4+ (-1) +0+4+3)=15

Now for these points we can get the absolute values of 10 years risk percentage of cardiovascular disease.
For the point of 10, the percentage risk score CVD is 9.4%
For the point of 15, the percentage risk score CVD is 21.6%

So we can see here that there is a large difference between the two risk score percentage of man A and man B. This difference is 12.2 though the age difference between man A and man B is just one year. This is a great problem. We can points out an another problem that in the risk category the10 years absolute risk for development of the Cardiovascular Disease

B. Risc Categories:

Table II. - Risk Category identification based on risk factor

RISK CATEGORY	10 YEARS ABSOLETE RISK (%)
HIGH	>20
INTERMEDIATE	6-20
LOW	0-6

From this table it is evident that the physical descriptions that affect the score points are nearly identical. They are so close that even a single person can exhibit these two set of results if the measurements are performed at different times and slightly different physiological conditions which occurs even in normal situations due to complex human physiological interactions. Now if the CVD risk scores are calculated for this two persons man A, man B from the previous score charts it becomes 9.4% for man A and 21.6 for man B. This means that man A is in the INTERMEDIATE zone and man B is in the HIGH zone which is totally counter intuitive that their physical parameters are all the same in all practical purposes.

C. Risc Points and It's Importance

We have also observed from CVD prediction Score and Heart age in the following that for man A and for man B there is a huge difference between their heart age i.e. man A's heart age is 54 and 72 heart age is for man B. So there is a huge difference i.e. 18. So can conclude that man A and man B's physical description are almost similar but the heart age of them has large difference. From this study, we found that not only CVD risk scoring system suffers from this boundary problem, all the Framingham risk scoring systems including diabetes risk scores, hypertension risk scores suffer from the same problem. A critical and careful look can easily identify the problems everywhere in this type of scoring systems. So a solution to this present problem will open a new avenue towards solving this class of problems in general. So this will be a very general solution that will be applicable in numerous situations successfully.

IV. PROBLEM FORMULATION AND SOLUTION

Perhaps this problem can be solved in appropriate theoretical setting, but here we have suggested a procedure which is very simple, rational, intuitive and mathematically meaningful and will require no attention in the basic formulation rather. This can be regarded as a minimal add-on.

To eliminate the 'boundary problem' the most natural thing what we can do is to remove the 'sharp boundary' because the sharp boundaries give rise to sharp transitions. If a single real number is assigned against an interval, these types of sharp transitions become unavoidable. So we have assigned the value at the midpoint of the interval because if some property remains invariant over an interval, the midpoint of the interval also satisfies this property and if a single point is to be selected from the interval to satisfy this property, the midpoint is the most faithful one. With this intuitive framework we can now proceed to make table that is constituted of the midpoints of the given intervals along with the score points.

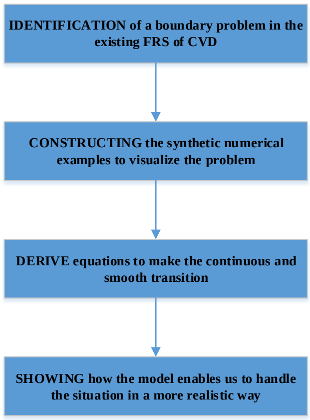


Fig. 1. - Process Flowchart

A. Preperation of applicable score chart

Table III. - Risk points and midpoints of the given interval

Risk category	Midpoints of given Intervals
Age for Men	(32, 0), (37, 2), (42, 5), (47, 6), (52, 8), (57, 10), (62, 11), (67, 12), (72, 14), (75, 15)
Age for Women	(32, 0), (37, 2), (42, 4),(47, 5), (52, 7), (57, 8), (62, 9), (67, 10), (72, 11), (75, 12)
Total Cholesterol for Men	(160, 0), (179.5, 1), (219.5, 2), (259.5, 3), (280, 4)
Total Cholesterol for Women	(160, 0), (179.5, 1), (219.5, 3), (259.5, 4), (280, 5)
HDL for Men	(60, -2), (54.5, -1), (47, 0), (39.5, 1), (35, 2)
HDL for Women	(60, -2), (54.5, -1), (47, 0), (39.5, 1), (35, 2)
SBP Treated for Men	(120, 0), (124.5, 2), (134.5, 3), (149.5, 4), (160, 5)
SBP Treated for Women	(120, -1), (124.5, 2), (134.5, 3), (144.5, 5), (154.5, 6), (160, 7)
SBP Not Treated for Men	(120, -2), (124.5, 0), (134.5, 1), (149.5, 2), (160, 3)
SBP Not Treated for Women	(120, -3), (124.5, 0), (134.5, 1), (144.5, 2), (154.5, 4), (160, 5)

From the above lists we can now tell the risk scores at some discrete points but we also need to assign values at the intermediate points to prepare a practically meaningful and applicable score chart. This seems a difficult task at the first glance but we shall show how a simple mathematical construction makes it easy. We have fitted a straight line through the points representing the linear trend of the event. This magic line really does the miracle. It fills all the gaps between the points and we can get distinct scores at all points without any sharp transition and abrupt change and helps us to get rid of the intuitively unacceptable embarrassing situation. The following few plots show the linear trends with the fitted straight line through the points. Equations of the straight lines along with their fitting errors have been shown with each plot.

B. Solution by streight line Fitting

To solve this boundary problem in our constructed numerical example, where the existing scoring system generates inconsistent and absurd results, we use the least square curve fitting and give a more appropriate result in this awkward situation. At first we have fitted the straight line with the help of the midpoints of the given intervals along with the score points.

The following equations are the derived fitted straight line for each risk factor for both men and women.

In the case of Men

(A) Age-

$$y = -9.94838+0.336066x \quad (1)$$

(B) Total Cholesterol-

$$y = -4.75981+0.0307683x \quad (2)$$

(C) HDL Cholesterol-

$$y = 7.21373-0.152833x \quad (3)$$

(D) Systolic Blood Pressure not treated (SBP) –

$$y = -14.0953+0.108172x \quad (4)$$

(E) Systolic Blood Pressure treated (SBP) –

$$Y=0.1082x-12.1 \quad (5)$$

In the case of Women

(A) Age-

$$y= -7.61515+0.265472x \quad (6)$$

(B) Total cholesterol-

$$y = -6.27013+0.0403738x \quad (7)$$

(C) HDL Cholesterol-

$$y = 7.21373-0.152833x \quad (8)$$

(D) Systolic Blood Pressure not treated-

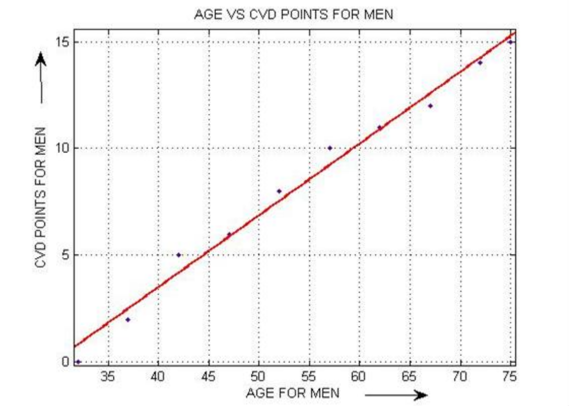
$$y = -22.6132+0.172648x \quad (9)$$

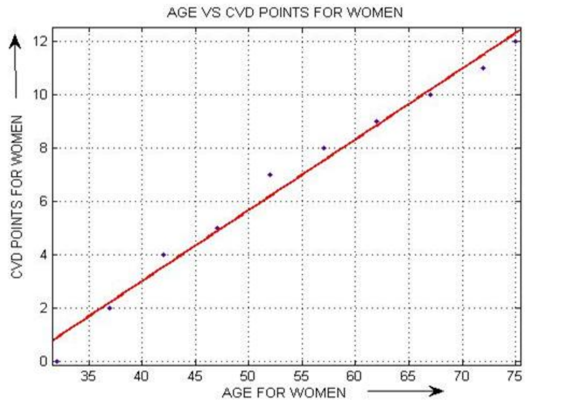
(E) Systolic Blood Pressure treated-

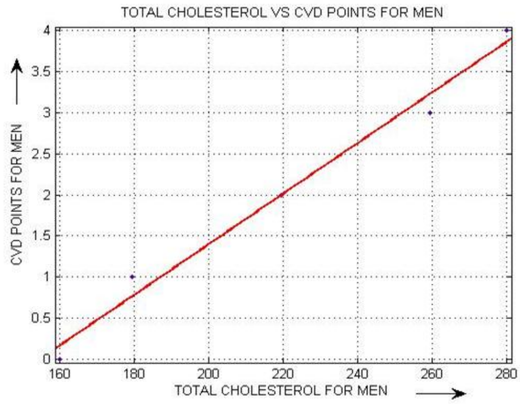
$$y=0.1764x-20.97 \quad (10)$$

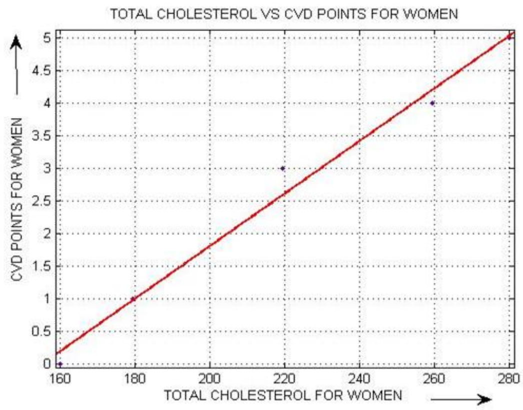
C. Resust for all risk factors

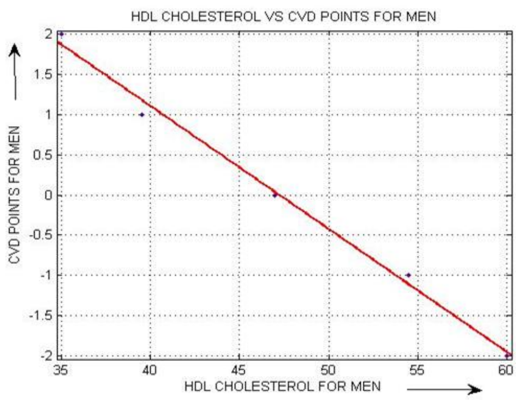
Th

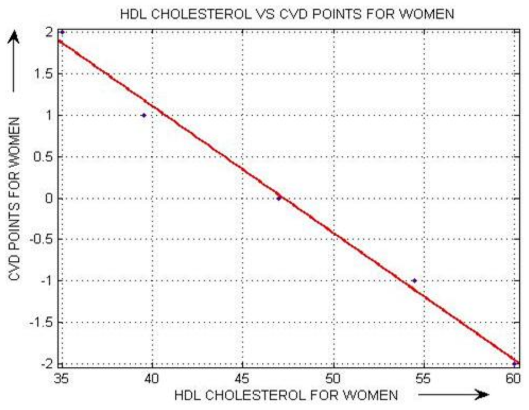


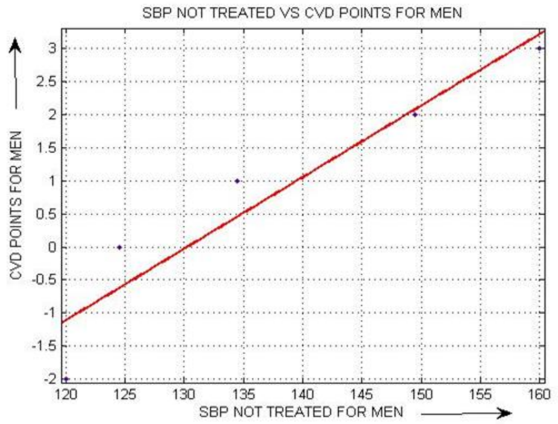


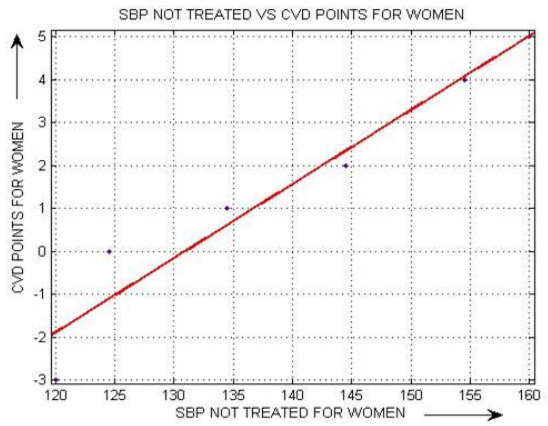


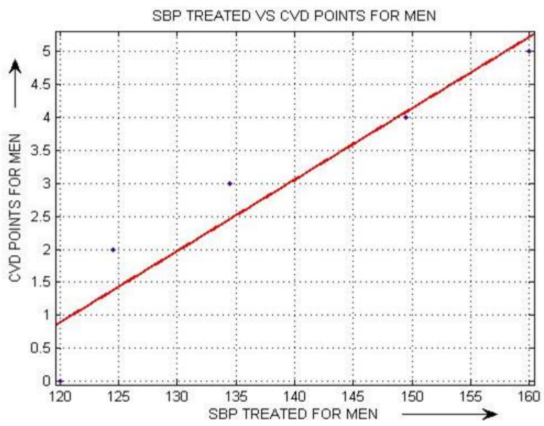


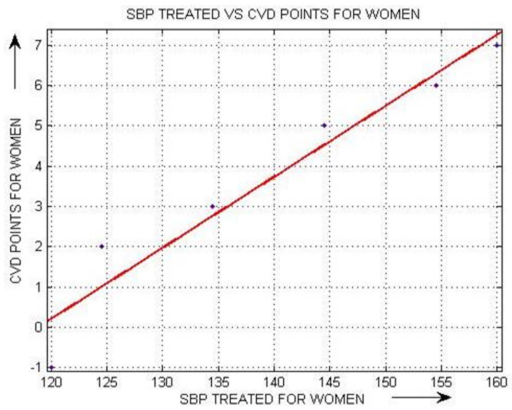












One natural question arises here. Why the straight line? Will it not be better if we fit a nonlinear curve, say a higher degree polynomial and do the same thing. Yes this is a genuine question since higher degree polynomial show a greater flexibility than straight lines. Primarily the straight lines are ok as far as they give good results for two reasons, one is the simplicity and the other is comprehensibility. Fitting a straight line is much simpler than higher degree nonlinear polynomials and more over linear trends are easily intuitively comprehensible.

V. FINAL OUTCOME AND CONCLUSION

In the following tables we have compared the results obtained with straight line fitting

Table IV. - Case study profile for men

RISK FACTORS	MAN A	MAN B	SCORE POINTS FOR RISK FACTORS OF MAN A	SCORE POINTS FOR RISK FACTORS OF MAN B
AGE	39	40	3.16	3.49
TOTAL CHOLESTEROL	279	280	3.82	3.86
HDL CHOLESTEROL	49	50	-.27	-.42
BLOOD PRESSURE (UNTREATED)	119	120	-1.22	-1.12
SMOKER	YES	YES	4	4
DIABETIC	YES	YES	3	3

Table V. - Case study profile for woman

RISK FACTORS	WOMAN C	WOMAN D	SCORE POINTS FOR RISK FACTORS OF WOMAN C	SCORE POINTS FOR RISK FACTORS OF WOMAN D
AGE	39	40	2.74	3.00
TOTAL CHOLESTEROL	279	280	4.99	5.03
HDL CHOLESTEROL	44	45	0.49	0.34
BLOOD PRESSURE (UNTREATED)	149	150	3.11	3.28
SMOKER	NO	NO	0	0
DIABETIC	NO	NO	0	0

Let us summaries what we have got so far. We have developed a new way of presenting the Framingham risk scores that was in some sense 'inevitable' due to the inconsistent results produced by the existing score charts. We have tried to solve the problem in the most natural way that is intuitively appealing. This new presentation style differs from the previous representation in a qualitative way. Previously, a look up table would tell us different scores at different points, instead now we have a mathematical function relating two variables. So a long look up table is no longer needed, single equation will take care of that. This situation is much more pleasing and less clumsy.

VI. CONCLUTION AND FUTURE WORK

Framingham study is an extensive, important and extremely valuable study in the history of medical science. But the Framingham score chart, originated from this study suffers from some technical problems at least in its present form. We pointed out four situations where the existing scoring system generates inconsistent and absurd results. We have used very simple mathematical technique based on so called least square curve fitting that eliminates the 'boundary problem' in the Framingham score system and makes the continuous transition to make it more realistic and usable.

It is the time to stop for a while and take a look back. What we have achieved so far? We have crossed the first handle and reached the primary target of our journey. It apparently seems that are done. It is true that we have eliminated the 'boundary problem' and resolved the counter-intuitive behavior of the previous score chart. We have also proposed a new way of looking at the old problem. This is indeed a happy situation. But still the journey has not been ended, we have a lot more to go. In science and particularly in mathematics we always try to generalize a solution. This means we want to cover a wider area with a bigger single umbrella so that some smaller particular solutions can be brought under it. In our present problem the existing score chart consisted of several step functions that gave rise to the boundary problem which we replaced with a smooth function representing gradual transition. These two solutions are the two extreme ends and one should enjoy enough freedom to lie between these two ends.

In our study we have demonstrated the 'boundary problem' present in the Framingham risk score and found a way to resolve this. But the risk zones also suffer from the same difficulty in the Framingham score system low risk zone is classified as 0% to 6% chance of risk, Intermediate risk zone as 6% to 20% chance of risk and high risk zone as greater than 20% chance of risk. A sharp boundary can readily be identified here. This event can successfully be modelled with the use of fuzzy set theory. Since the words 'Low', 'Intermediate', 'High' represent uncertainty in linguistic terms, fuzzy set theory becomes inevitable to model them. Suitable membership functions for each of the linguistic variables will do the job successfully and represent some overlapping regions that will remove the sharp transitions from one risk zone to another.

REFERENCES

[1] D'Agostino RB Sr, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, Kannel, "General cardiovascular risk profile for use in primary care: the Framingham Heart Study" WB. Circulation. 2008 Feb 12.

[2] Reddy S, Yusuf S., "Emerging epidemic of Cardiovascular Disease in Developing Countries" Circulation. 1998 pp:596-601.

[3] Bulatao RA, Stephens PW, "Global estimates and projections of mortality by cause" Washington DC: Population, Health and nutrition Department; World Bank; pre-working paper 1007, 1992

[4] Philip Greenland, Laurie LaBree, Stanley P. Azen, Terence M. Doherty, Robert C. Detrano, "Coronary Artery Calcium Score Combined With Framingham Score for Risk Prediction in Asymptomatic Individuals", JAMA. 2004; 291(2):210-215. doi:10.1001

[5] Suzanne G. Haynes, Sol Levine, Norman Scotch, Manning Feinleib, William B. Kannel, "The Relationship of Psychosocial Factors to Coronary Heart Disease in the Framingham Study", American Journal of Epidemiology,Volume 107 Issue 5, pp. 362-383.