Predictive Accuracy of a Cardiovascular Disease Risk Prediction Model in Rural South India – A Community Based Retrospective Cohort Study
Farah N Fathima1, Deepa Raghunath2, Shailendra Kumar Hegde3, Twinkle Agrawal4, Dominic Misquith5, Pruthvish Sreekantiah6
1,4Assistant Professor, 5Professor, Department of Community Health, St John's Medical College, Bangalore; 6Professor, Department of Community Medicine, MS Ramaiah Medical College, Bangalore

Abstract
Background: Identification of individuals at risk of developing cardiovascular diseases by risk stratification is the first step in primary prevention. Aims & Objectives: To assess the five year risk of developing a cardiovascular event from retrospective data and to assess the predictive accuracy of the non-laboratory based National Health and Nutrition Examination Survey (NHANES) risk prediction model among individuals in a rural South Indian population. Materials & Methods: A community based retrospective cohort study was conducted in three villages where risk stratification was done for all eligible adults aged between 35-74 years at the time of initial assessment using the NHANES risk prediction charts. Household visits were made after a period of five years by trained doctors to determine cardiovascular outcomes. Results: 521 people fulfilled the eligibility criteria of whom 486 (93.3%) could be traced after five years. 56.8% were in low risk, 36.6% were in moderate risk and 6.6% were in high risk categories. 29 persons (5.97%) had had cardiovascular events over the last five years of which 24 events (82.7%) were nonfatal and five (17.25%) were fatal. The mean age of the people who developed cardiovascular events was 57.24 ± 9.09 years. The odds ratios for the three levels of risk showed a linear trend with the odds ratios for the moderate risk and high risk category being 1.35 and 1.94 respectively with the low risk category as baseline. Conclusion: The non-laboratory based NHANES charts did not accurately predict the occurrence of cardiovascular events in any of the risk categories.

Key Words
Cardiovascular disease; Non-laboratory based risk score; Risk stratification; Retrospective cohort study; NHANES charts

Introduction
Cardiovascular diseases (CVDs) are one of the leading causes of death worldwide (1), with 80% of cases occurring in low-income and middle-income countries including India (2-3). The prevalence of CVD is estimated to be between 7 and 11% in urban and 1 and 6% in rural India (4). As the healthcare resources in developing countries such as India are limited, there is a need to identify feasible, effective and low cost interventions which can then be integrated into the existing public health delivery system. A well-established primary prevention strategy is to use prediction rules or risk scores to identify those at higher risk to target specific behavioral or drug interventions (5).

Previous studies that we conducted have shown the prevalence of cardiovascular risk factors to be high in...
rural populations of South India (6-7). In one study we recorded a prevalence of hypertension and diabetes mellitus at 23.5% and 4.75% respectively; 30% used tobacco and 26% were overweight (6). Another study showed that 6.9% of the population was hypertensive, 5.2% were diabetic, 15% were current smokers and 20.6% were overweight (7).

Studies such as the Framingham Heart Study (8) and the Multiple Risk Factor Intervention Trial (9) have shown that the coexistence of multiple risk factors confers a magnified risk which is multiplicative rather than additive (10-11). The demonstration of such multiplicative risk has given rise to the concept of “comprehensive cardiovascular risk” or “total risk”, quantifying an individual’s overall risk of developing cardiovascular disease resulting from the confluence of risk factors. This is particularly relevant among Indians because of the clustering of risk factors. A number of methods have been devised to calculate individual risk based on risk factor levels (12-24). Two recent studies (5,17) used non-laboratory-based low-information risk scores and concluded that these scores perform as well as the more cost-intensive laboratory based scores. These non-laboratory based NHANES (National Health and Nutrition Examination Survey) risk scoring charts predict an individual’s five year risk of developing a first time fatal or nonfatal adverse cardiovascular outcome. This risk prediction score was developed and tested in western populations and hence there is a need to study its accuracy and applicability in an Indian setting.

Using data from two studies done by the authors (6-7) five years ago (2004-05), as the baseline, we conducted a follow up study to assess the predictive accuracy of the non-laboratory based NHANES risk prediction model. This would be useful to predict the five year risk of developing an adverse cardiovascular outcome.

Aims & Objectives

1. To estimate the five year risk of developing a first time cardiovascular event among individuals in a rural South Indian setting.
2. To assess the predictive accuracy of the non-laboratory based NHANES risk prediction model among individuals residing in a rural South Indian setting.

Material and Methods

A community based retrospective cohort study was designed and conducted in three villages located in the rural field practice areas of Departments of Community Medicine of two medical colleges of Bangalore. These areas have a combined population of 7261 people living in 1282 households.

Ethical clearance was obtained from the institutional ethical review board prior to the onset of the study. The study was conducted in two phases. In phase one, data from the two previous studies (6-7) conducted in the same population in the year 2004-05 was used as the baseline. Using this data a study cohort was established which comprised of all adults in the age group of 35 – 74 years at the time of initial assessment (2004-05). The colour coded non-laboratory based NHANES risk scoring charts were used to compute the five year risk of having a first time adverse cardiovascular outcome for every individual in the study cohort. (5) The variables used in the NHANES risk prediction charts are gender, age, history of diabetes mellitus, current smoking status, systolic blood pressure and body mass index. Based on the above mentioned factors, these charts stratify the population into low (<10%), moderate (10-20%) and high (>20%) risk of developing an adverse cardiovascular event in the next five years. These risk categories formed the exposure variables in the study.

Those individuals with a history of a previous adverse cardiovascular outcome at the time of initial assessment were excluded from the study cohort. Also, those individuals in whom outcome variables could not be determined were excluded from the study analysis.

In phase two of the study, individuals in the study cohort were contacted through household visits done by trained physicians who were blinded to the risk category of the individuals to be followed up. This follow up visit was made five years (2010) after the initial assessment (2005). All cardiovascular events (both fatal and nonfatal) that occurred during the five year period formed the outcome variable. After obtaining written informed consent, data on mortality and morbidity due to cardiovascular disease, medical history and current health status were collected using a structured interview schedule. Medical records, pathology reports, electrocardiographs and death certificates were reviewed by the investigators. All cardiovascular events that occurred between baseline survey and the current follow up date were reviewed wherever possible with the help of discharge documentation.
Causes of death were verified by death certificates wherever available. In cases where a death certificate was not available, an attempt was made to elicit the probable cause of death by verbal autopsy conducted by trained doctors. Operational case definitions were used and data were collected on the occurrence of endpoints for cardiovascular disease namely death, non-fatal myocardial infarction, transient ischemic attacks, angina, congestive heart failure, peripheral vascular disease, and coronary re-vascularisation including coronary artery bypass grafting and percutaneous transluminal coronary angioplasty.

Sample size estimation: We estimated the sample size based on positive predictive values and negative predictive values reported in three different risk categories by Gazziano et al. (5) We found that a total of 400 subjects are required to estimate these positive and negative predictive values with a precision of 10% and at 95% confidence interval.

Analysis: We did the statistical analysis using standard statistical software package (Epi Info 7) and statistical significance was determined at 5%. We cross tabulated the different risk categories and the occurrence of cardiovascular events and computed the incidence of events in each risk category. The relative risk of developing an event was calculated for moderate and high risk categories taking the low risk category as baseline. Chi square statistic for trend was obtained. The actual number of cardiovascular events that had occurred in each risk category were then compared to those predicted by the NHANES charts.

**Results**

A total of 521 individuals fulfilled the eligibility criteria and hence formed the study cohort. Of these, 486 (93.28%) individuals could be traced after a five year period. The mean age of the study population was 54.25 ± 10.75 years and males constituted 47.95% (233) of the study population. A total of 29 persons (5.97%) had had cardiovascular events over the last five years out of which 24 events (82.75%) were nonfatal and five (17.25%) were fatal [Table 1].

Table 2 depicts the distribution of the study population (from retrospective data) according to their level of risk of developing a fatal or a non-fatal cardiovascular event over five years using the non-laboratory based NHANES charts. We found that 56.79% of the people were in the low risk category. This means that their risk of developing a cardiovascular event was <10%, where as 36.62% and 6.58% were in the moderate risk (10-20% risk) and high risk (>20% risk) categories respectively. We found that the overall five year incidence of cardiovascular events in the study population was 5.97%. The incidence of events in the low, moderate and high risk category was 5.1%, 6.7% and 9.4% respectively. The mean age of the individuals who developed a cardiovascular event was 57.24 ± 9.09 years.

With low risk category as the baseline the relative risk of developing a cardiovascular event in the moderate and high risk category was 1.4 and 1.8 respectively. The odds ratios for the three levels of risk showed a linear trend.

Table 3 depicts the predicted and actual number of cardiovascular events in the study population. The table shows that the actual number of events did not fall within the predicted range in any of the risk categories. Our study also shows that on follow up of the cohort over five years, the non-laboratory based NHANES charts did not accurately predict the occurrence of cardiovascular events in any of the risk categories. The same has been depicted in Figure 1.

**Discussion**

Primary prevention is a well-established strategy to combat the rising incidence of cardiovascular diseases. Risk prediction scores can be used to identify those at higher risk of developing a cardiovascular event and controlling the risk factors in such individuals is crucial in the implementation of a high risk strategy program for the control of CVD. In our study, we recorded that more than one third (43.2%) of the adults above the age of 35 years in a rural community were at moderate to high risk of developing a fatal or nonfatal cardiovascular event over five years.

There is a paucity of published studies across India that have used risk scores to predict the cardiovascular risk. A hospital based study done by Kanjilal et al (25) in two urban areas of India used ten different laboratory based risk scores for risk stratification.

Their results showed that most of the subjects were in the low risk category according to all the risk scores used. In the above study the Framingham and the Joint British Societies’ systolic blood pressure based coronary heart disease risk scoring systems defined 5.32% and 3.7% of the cohort to be in the high risk category. An additional 14.85% and 12.78%
of people were at intermediate risk for CVD in the forthcoming ten years with the remaining being in the low risk category (25).

Despite searching extensively, we did not find any published data on the stratification of CVD risk using non-laboratory based risk scores or any community based study done in a rural setting. In our study, using the NHANES non-laboratory based risk score we found that 36.62% of the study population were at moderate risk while 6.58% were at high risk of developing a fatal or nonfatal cardiovascular event over the forthcoming five years. This is a pointer towards the high burden of cardiovascular diseases that we can expect over the years to come in a rural Indian population.

We also found that cardiovascular events occur at a young age among Indians with the mean age of occurrence of CVD being 57.24 ± 9.09 years. This finding was similar to a study done in rural Andhra Pradesh by Joshi et al, where the mean age of reported CVD was 54 years (26). The above study also found that one half and one third of all cardiovascular deaths occurred before 70 and 60 years of age, respectively.

The occurrence of CVD among the economically productive age group of India implies reduced productivity and high treatment costs. These costs are borne mainly by out of pocket payments due to lack of health insurance, lack of facilities in public hospitals and high costs in the private sector. This indicates an urgent need towards development of effective, low cost screening methods to identify such individuals so that preventative interventions can be directed at them.

Data from office of the Registrar General of India show that 19% of all deaths in India (25% for South India) are CVD related. In our study 5 out of a total of 24 deaths (17%) were due to CVD. Though the proportions seem comparable, the small variation can probably be attributed to the small numbers in our study and calls for similar studies in larger population groups (27).

**Conclusion**

Our study showed that 56.79% of the people were in the low risk category, where as 36.62% and 6.58% were in the moderate risk (10-20% risk) and high risk (>20% risk) categories respectively. Our study also shows that on follow up of the cohort over five years, the non-laboratory based NHANES charts did not accurately predict the occurrence of cardiovascular events in any of the risk categories.

**Recommendation**

This study reinforces the need for alternative approaches to risk stratification that can be applied to a South Asian population. This will help to adopt low cost and effective tools to predict cardiovascular risk in individuals and in populations to ensure policy changes for adequate life style modifications and treatment.

**Limitation of the study**

In our study, a total of 35 people could not be traced even after best efforts by the investigators and hence the cardiovascular outcomes in them could not be ascertained. Though the percentage of such persons is very small (6.71%) and is unlikely to affect the results, the authors acknowledge this to be one of the limitations of the study. In addition, the sample size and power calculations were done based on the reported sensitivity of the NHANES chart in the lowest risk category. Hence the current results may not be sufficient to draw conclusions for the general Indian population at all levels of risk.

Assessment of risk category which is the exposure variable in this study was based on reported history of diabetes mellitus. Studies in India have shown that only 60-70% of the diabetics are aware of their diabetic status (28-30).

In addition, the NHANES tool uses a BMI cut off of 25, whereas the revised cut offs in the Indian population classify BMI greater than 23 as overweight. These two factors which are inherent features of the NHANES tool could have led to underestimation of the exposure variable in the study resulting in some amount of misclassification error in measurement of exposure.

Despite these limitations, the main strength of this study is that it is a rural community based retrospective follow up study with a high proportion (93.3%) of the study population being followed up after five years. In addition, the follow up in the field was done by physicians that enabled the investigators to pick up cardiovascular events with a high degree of accuracy. The fact that the physicians who did the follow up were blinded to the exposure variable minimises the interviewer bias in the study.

**Relevance of the study**

This study shows that on follow up of the cohort over five years, the non-laboratory based NHANES charts
did not accurately predict the occurrence of cardiovascular events in any of the risk categories. This could be because the NHANES non-laboratory based risk scores were developed using a Western population base and hence may not accurately reflect the risk pattern in a South Asian population. Disproportionate estimation of risk may lead to risk-reducing interventions not being offered to those who need them the most.

This calls for follow up of larger cohorts and to record cardiovascular outcomes in order to develop risk predictive scores applicable to the Indian population.

Authors Contribution
FNF was involved in Concept, Design, Definition of intellectual content, Literature search, Data acquisition, Data analysis, Statistical analysis, Manuscript preparation, Manuscript editing, Manuscript review and will act as Guarantor. DR was involved in Definition of intellectual content, Data acquisition, Data analysis and Manuscript review. SKH was involved in Design, Definition of intellectual content, Literature search, Data acquisition, Data analysis, Statistical analysis, Manuscript editing, and Manuscript review. TA was involved in Design, Definition of intellectual content, Manuscript editing, and Manuscript. TP was involved in Data acquisition, Data analysis and Manuscript review. DM was involved in Design, Definition of intellectual content, Manuscript editing, and Manuscript review. PS involved in Design, Definition of intellectual content, Manuscript editing, and Manuscript review

Acknowledgement
The authors would like to thank Dr Sharan Holyachi and Dr Tony Pius for their immense support to ensure the completion of this study.

References
6. Fathima NF, Pruthvish S, Murthy NS. Study of Risk Factors for Cardiovascular Disease in two village populations of Chintamani Taluk, Dissertation Submitted to Rajiv Gandhi University of Health Sciences, 2006 [Unpublished]


Tables

<table>
<thead>
<tr>
<th>TABLE 1 TOTAL NUMBER OF PEOPLE FOLLOWED UP, FATAL AND NONFATAL CARDIOVASCULAR EVENTS AMONG THE STUDY POPULATION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variables</strong></td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Total population base</td>
</tr>
<tr>
<td>Total number of households</td>
</tr>
<tr>
<td>Study Cohort (individuals between 35 and 74 years available from the two previous studies)</td>
</tr>
<tr>
<td>Total number of people who fulfilled the eligibility criteria</td>
</tr>
<tr>
<td>Lost to follow up (Untraceable)</td>
</tr>
<tr>
<td>Total number followed up finally (%)</td>
</tr>
<tr>
<td>No Cardiovascular Event (%)</td>
</tr>
<tr>
<td>Cardiovascular Events (%)</td>
</tr>
<tr>
<td>Non-fatal Cardiovascular events (%)</td>
</tr>
<tr>
<td>Fatal Cardiovascular events (%)</td>
</tr>
<tr>
<td>Total deaths due to all causes (%)</td>
</tr>
<tr>
<td>Non Cardiovascular deaths (%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 2 RISK STRATIFICATION OF THE STUDY POPULATION, NUMBER OF EVENTS AND RELATIVE RISK</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Risk Category</strong></td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Low</td>
</tr>
<tr>
<td>Moderate</td>
</tr>
<tr>
<td>High</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Chi square for linear trend = 1.215; ‘p’ value = 0.27036
TABLE 3 PREDICTED AND ACTUAL NUMBER OF CVD EVENTS IN THE STUDY POPULATION

<table>
<thead>
<tr>
<th>Risk Category</th>
<th>Number of people</th>
<th>Actual Events</th>
<th>Minimum number of Predicted Events</th>
<th>Maximum number of Predicted Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very low (&lt;5%)</td>
<td>75</td>
<td>5</td>
<td>0</td>
<td>3.75</td>
</tr>
<tr>
<td>Low (5-10%)</td>
<td>201</td>
<td>9</td>
<td>10.05</td>
<td>20</td>
</tr>
<tr>
<td>Moderate (10-20%)</td>
<td>178</td>
<td>12</td>
<td>17.8</td>
<td>35.6</td>
</tr>
<tr>
<td>High (20-30%)</td>
<td>19</td>
<td>1</td>
<td>3.8</td>
<td>5.7</td>
</tr>
<tr>
<td>Very High (&gt;30%)</td>
<td>13</td>
<td>2</td>
<td>3.9</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>486</td>
<td>29 (5.97)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figures

FIGURE 1 PREDICTED AND ACTUAL NUMBER OF EVENTS IN THE STUDY POPULATION

Box – depicts the range of predicted events in a risk category
Dot – depicts the actual number of events in each category