**USING SELF-REPORT MEASURES TO LOWER THE COST OF POPULATION HEART HEALTH ASSESSMENT IN NEW BRUNSWICK**

By: John Boyne, Joyendu Bhadury, and B. Christofer Balram


Made available courtesy of Canadian Evaluation Society: [http://www.evaluationcanada.ca/site.cgi?s=4&ss=2&_lang=EN](http://www.evaluationcanada.ca/site.cgi?s=4&ss=2&_lang=EN)

***Note: Figures may be missing from this format of the document***

**Abstract:**
This paper describes the development and feasibility testing of a multivariate equation that uses self-report information rather than physiological measures to estimate coronary heart disease (CHD) risk in a population sample of New Brunswick adults with no reported history of heart disease. The multivariate Framingham risk prediction model, which uses a variety of self-report add physiological measures to estimate CHD risk, was first used to calculate CHD risk in the population sample. Regression analysis was then employed to identify a linear combination of "self-reportable" variables capable of closely approximating the population risk indices derived using the Framingham model. To test its utility, the self-report equation derived from the regression analysis was applied to a small telephone survey data set drawn from a second random sample of adult New Brunswickers with no reported history of heart disease. When applied to the telephone survey data, the self-report equation yielded CHD risk estimates consistent with those from the first population sample. We concluded that the development of a self-report-based methodology for assessing the CHD risk or heart health of target populations is highly feasible. Owing to the use of self-report information, as opposed to the physiological measures employed in conventional CHD risk prediction models, a self-report-based model could significantly reduce the cost of assessing CHD risk in the target populations of community-based heart health programs. Although further research will be necessary to develop a complete self-report-based CHD risk prediction model, the results of the present study clearly indicate that this line of research has significant potential to enhance the evaluation of heart health promotion programs.

**Article:**
Coronary heart disease (CHD) is a leading cause of premature death and disability in Canada and much of the industrialized world. The major factors that place individuals at increased risk of developing CHD include male sex, age, genetic predisposition, diabetes, smoking, elevated blood cholesterol, high blood pressure, and a sedentary lifestyle. As several of these risk factors are modifiable, it is generally accepted that widespread adoption of 'heart healthy' lifestyle changes such as smoking cessation/abstinence, low-fat diet, and regular physical activity can significantly reduce the incidence of heart disease (Advisory Board of the International Heart Health Conference, 1992, p. 3).
The potential for heart healthy lifestyle changes to save both thousands of lives and millions of dollars in treatment costs has led to a proliferation of what have been termed "community" or "population" based heart disease prevention programs. These initiatives target entire populations and are characterized by intervention activities such as public education, social marketing campaigns, and community mobilization strategies that encourage all members of the population (as opposed to specific "high-risk" segments) to adopt healthier lifestyles. Proponents of this approach to heart disease prevention argue that because large populations are targeted, even modest levels of success are likely to yield major reductions in heart disease rates. The population-based approach is also thought to be more cost effective than other strategies because there are neither case finding nor individualized intervention expenses.

Although the rationale for population-based heart health promotion is persuasive, a number of factors, including a long time lag between the introduction of heart health promotion strategies and reductions in CHD morbidity and mortality, make it difficult to demonstrate clearly either the clinical efficacy or cost effectiveness of this approach (N, Smith, Murphy, & Catford, 1993). One methodology that has been employed to deal with this time lag problem involves the use of multivariate risk prediction models to forecast the impact of observed or anticipated risk factor changes on future heart disease rates. A major drawback of this methodology, however, is that conventional multivariate, risk prediction models require physiological measures such as blood cholesterol and blood pressure, which are very expensive to collect in large target populations.

The primary purpose of this paper is to demonstrate the feasibility of substituting "self-reportable" risk measures for the physiological measurements used in conventional CHD risk prediction equations. The ability to use SR measures to estimate population CHD risk suggests that it may be possible to develop a significantly less expensive, SR-based methodology for assessing the CHD risk or heart health of target populations. Further development of this alternative approach to assessing the heart health of target populations may allow multivariate risk assessment techniques to be used in the evaluation of a broader range of community-based heart health programs than is currently feasible.

EVALUATION OF HEART HEALTH PROMOTION PROGRAMS
Evaluation of community-based heart health programs often involves the tracking of changes in a variety of knowledge, behavior and, to a lesser extent, physiological risk factors in the target populations. Although this type of risk indicator tracking provides a useful index of the short-term impact of heart health promotion programs, it does not directly address the ultimate goal of morbidity and mortality reduction, nor does it take into account the interaction of multiple risk factor changes on the heart health of the target population. In order to assess the potential impact of population-based heart disease prevention programs on morbidity and mortality outcomes, epidemiologically based models have been used to forecast the potential long-term impact of risk factor changes on subsequent CHD rates (see, for example, Kottke, et al., Wu, & Park, 1988; MacLean & Petrasovits, 1990). Although the validity, generalizability, and appropriateness of this type of health risk assessment methodology have not gone unchallenged, it remains a useful tool for both clinical and economic evaluation of CHD prevention strategies (Hayes, 1992; Katz and Forman, 1993).
Perhaps the most widely used multivariate CHD risk prediction equation is the so-called Framingham model, which was developed by researchers from the Framingham heart study (Kennel & Gordon, 1971). This model has been widely used to assess the potential impact of a variety of CHD prevention strategies, including community-based heart health programs such as the Stanford Five City Project and the Pawtucket Heart Health Program (Shea & Basch, 1990). Although multivariate risk prediction functions such as the Framingham model have proven to be useful evaluation tools, their utilization requires expertise and resources that may not be available in many heart health programs. In particular, these risk prediction functions generally require information on variables such as blood cholesterol and blood pressure that can only be collected through a direct clinical examination. Although technically feasible, population sampling for these types of measures is very expensive. For example, the New Brunswick Heart Health Survey (Health & Community Services New Brunswick [HCSNB], 1989) employed an interview and clinical examination to collect cholesterol, blood pressure, and a variety of other risk factor measurements from a sample of approximately 2,000 residents in New Brunswick. The survey required several months to complete and is estimated to have cost in excess of $200,000.

A potential strategy for reducing the cost of assessing CHD risk or heart health in populations may be to develop a multivariate risk prediction methodology that employs risk index measures that can be obtained with relatively inexpensive self-report data collection techniques. Conventional multivariate MID risk prediction functions already include information such as age, smoking behavior, and history of diabetes that are collected in self-reports. Factors such as high blood pressure and elevated blood cholesterol also have behavioral or observable correlates such as obesity, a sedentary lifestyle, and having been told one has high blood pressure or elevated cholesterol. Therefore, it may be possible to develop a multivariate risk prediction function that is capable of predicting CHD risk on the basis of self-reportable measures. The present study was designed to investigate this possibility.

METHODS

Data Sets

The data employed in the present study were derived from two different samples of the general population of New Brunswick. The first data set was derived from the 1989 New Brunswick Heart Health Survey (HCSNB, 1989), which is a stratified random sample of approximately 2,000 New Brunswick residents between the ages of 18 and 74. The second data set was derived from a random telephone survey of approximately 500 New Brunswick residents aged 15 and over, conducted by the New Brunswick Heart Health Program in 1996.

In order to obtain data suitable for estimating future CHD incidence rates (i.e., CHD risk or heart health) in the New Brunswick population, three criteria were used to select records from the two population samples. First, because the study involved predicting subjects' future risk of developing CHD, only the records of those respondents who reported no previous history of heart disease were selected. Second, subjects from the 1989 HCSNB sample were excluded if their records did not include the blood cholesterol and blood pressure measurements required for use in the Framingham model. Finally, respondents in both samples whose ages were outside of the 30-74 year range were excluded because the Framingham model is designed for use with subjects between those years of age. The selection procedure yielded a 1989 data set that...
consisted of 1,204 records from the HCSNB and a 1996 data set that consisted of 391 records from the 1996 telephone survey. Records in the 1989 data set included a variety of self-reported information: knowledge, behavior, health history, and demographics as well as such physiological measurements as blood cholesterol levels, blood pressure, height, and weight (for details on the survey variables, see HCSNB, 1989). The 1996 data set included self-reported knowledge, behavior, health history, and demographic information as well as self-report measures of height and weight.

Conventional Calculation of CHD Risk Probabilities
In the first stage of the study the Framingham risk prediction model (Anderson, Wilson, Odell, & Kennel, 1991) was employed to calculate a CHD risk score for each individual in the 1989 data set. These risk scores were then used to estimate the average 10-year CHD risk in different age segments of the population, as well as the annual number of new CND cases expected in each age segment over the 10-year prediction period.

Development of the Self-Report Risk Equation
The second stage of the study employed multiple regression techniques to construct an alternative equation for calculating CHD risk scores for individuals in the 1989 data set. This involved developing a multiple regression equation that used self-reportable variables to approximate the individual risk scores previously calculated with the Framingham equation.

Evaluation of the Validity and Utility of the Self-Report Risk Equation
The validity of the self-report (SR) risk equation was evaluated by comparing the 10-year CHD risk probabilities calculated using the SR equation with those calculated using the conventional Framingham model, in different age segments in the 1989 data set.

The utility of the SR risk equation was assessed by applying it to the 1996 telephone survey data set. The SR equation results for the 1989 and 1996 data sets were compared to assess the performance of the SR-based methodology when applied to the type of telephone survey data often used in the evaluation of smaller scale heart health promotion programs.

RESULTS AND DISCUSSION
CHD Risk Estimates from the 1989 Data Set
We calculated CHD risk probabilities for individuals in the 1989 data set using the Framingham model as described by Anderson et al. (1991). In the first step of the calculations, we used the following equation to derive an individual risk score (a term) for each subject.

\[
a = 11.1122 - 0.9119 \times \log(SBP) - 0.2767 \times \text{smoking} - 0.7181 \times \log(\text{cholesterol/HDL}) - 0.5865 \times \text{ECH-LVH}.
\]

where log is the natural logarithm function, SBP is systolic blood pressure, smoking is cigarette smoking (defined as smoking one or more cigarettes per day), cholesterol I HDL is the ratio of total serum cholesterol to high-density lipoprotein, and ECH-LVH is the presence of left ventricular hypertrophy. Information on subjects' ECH-LVH status was not available in the 1989 data set, and so this variable was dropped from the equation. Omitting the ECH-LVH term was not expected to significantly influence the results because this condition should not be prevalent in subjects who have no previous history of heart disease.
Following calculation of the initial risk scores, we factored each individual's sex, age, and history of diabetes into the model and calculated each individual’s probability of developing CM over a 10-year period. We then averaged the individual CHD risk probabilities to find the mean COD risk probabilities for various age groups in the population. Finally, we calculated the number of new CHD cases expected in each age group by multiplying each subject's 10-year CHD risk probability by the number of persons they represented in the population (i.e., their weighting factor). A summation of these weighted 10-year CHD risk probabilities provided an estimate of the number of persons in each age group who were likely to develop CHD over the 10-year prediction period. The number of new cases expected annually was calculated by averaging the 10-year total for each age group.

Table 1 presents the resulting 10-year CHD risk probabilities for different age groups in the 1989 data set, as well as the annual number of new CHD cases expected in each age group over the 10-year prediction period.

The Framingham model calculations suggest that in 1989 approximately 7.2% of New Brunswickers between the ages of 30 and 74 with no previous history of heart disease were expected to develop CHD during the following 10 years. This probability translates into an expected total of approximately 2,400 new CHD cases annually. Based on findings from the Framingham Heart Study (Kennel & Feinleib, 1972), approximately 46% of these cases (1,104 cases) would be expected to present as angina pectoris, 34% (816 cases) as myocardial infarction, 12% (288 cases) as coronary or sudden death, and the remaining 8% (192 cases) as coronary insufficiency.

To assess the validity of the Framingham model estimates, we compared the predicted number of new myocardial infarction (AMI) cases to hospital admission statistics. In 1989/90 there were 1,319 admissions to New Brunswick hospitals for AMIs in individuals under the age of 75. Based on research conducted in the neighboring province of Nova Scotia (Weilgosz et al., 1992); we estimated that approximately 20% of these AYH admissions were likely to be for a recurrent AMI and the remaining 80% (1,055 admissions) for a first AMI. Findings from the Framingham heart study indicate that approximately 23% of individuals experiencing a first AMI will have previously developed angina pectoris. Therefore, we estimated that 812 individuals with no previous history of CHD would have been admitted to New Brunswick hospitals with an AMI in 1989/90. This estimated number is very close to the number of new AMI cases predicted by the Framingham model, 816. Although these hospital admission statistics provide only a crude index of CHD risk in the New Brunswick population, the similarity between the number of AMI cases

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Average 10-Year Risk Probability</th>
<th>Annual Number of New Cases Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>30-34</td>
<td>0.013</td>
<td>81</td>
</tr>
<tr>
<td>35-44</td>
<td>0.322</td>
<td>342</td>
</tr>
<tr>
<td>45-54</td>
<td>0.077</td>
<td>488</td>
</tr>
<tr>
<td>55-64</td>
<td>0.128</td>
<td>687</td>
</tr>
<tr>
<td>65-74</td>
<td>0.183</td>
<td>900</td>
</tr>
<tr>
<td>30-74</td>
<td>0.072</td>
<td>2401</td>
</tr>
</tbody>
</table>
predicted by the Framingham model and the number of new AMI cases estimated from the 1989/90 admission statistics suggests that the model provides a reasonably good estimate of population CHD risk.

**Self-Report Risk Equation**

The SR risk equation was developed by first defining the a term (risk score) from the Framingham model, which had been calculated for each individual in the 1989 data set, as a dependant variable. Stepwise multiple regression was then used to identify a set of self-reportable variables that predicted the a term. A variety of behavioral and demographic variables were tested for their predictive utility. It should be noted, however, that at least one potentially important demographic variable, family history of heart disease, was not available in the 1989 data set. In addition, the 1989 data set contained little information on eating habits and only a crude measure of leisure time physical activity. Among the available variables, the following were determined to be the most useful self-reportable predictors of the risk score:

1. Smoking behavior: Regular smoker = 1; nonsmoker = 0
2. Age: In years
3. Gender: Male = 0; female =1
4. History of high blood pressure (HBP):
   - Currently taking medication to treat HBP = 1
   - Not currently taking HBP medication = 0
5. Body mass index: BMI

Table 2 details the results from regression that was performed and gives the goodness-of-fit statistics for the SR risk equation.

The SR risk equation accounts for approximately 54% of the variance in the dependant variable and provides a relatively good fit for the Framingham risk score data. The SR risk equation has the following form:

\[
sr = 6.53 - 0.076 \text{ HBPmed} - 0.008 \text{ age} + 0.19 \text{ sex} - 366 \text{ smoke} - 0.021 \text{ BMI}
\]

where \(sr\) denotes self-report risk score. This equation is similar to the Framingham risk equation in that both employ smoking and blood pressure variables. However, in the SR equation B1111, sex, and age are substituted for the total/HDL cholesterol component of the Framingham equation.

**Predictive Validity of the SR Equation**

To assess the predictive validity of the SR equation, the individual risk scores derived using the SR equation (\(sr\) term) were substituted for the a term in the first step of the Framingham model, and a second set of 10-year CHD risk probabilities were calculated for the 1989 data set. Table 3 provides a comparison of the average 10-year risk probabilities yielded by the original Framingham model and the modified model using the SR equation.
As seen in Table 3, the SR-based risk probabilities are very close to those from the original Framingham model except in the younger age categories, where the SR-based probabilities tend to be somewhat lower. The discrepancies between the Framingham and SR risk probabilities in the two youngest age categories are probably attributable to the fact that approximately 9% of subjects 30-44 years old had untreated high blood pressure. The presence of this risk factor is reflected in the Framingham risk equation, which used actual blood pressure measurements, but not in the SR equation, where high blood pressure risk is redacted only if the subject reported currently taking medication for high blood pressure. A two-way ANOVA (age group by model) was used to compare the Framingham and SR risk probabilities. The main effect for the age group variable was statistically significant (F4,2398 = 914, p < .001), indicating real differences between the average risk probabilities in different age groups. However, neither the main effects for the model variable nor the age group by model interaction effect were statistically significant. Thus, we concluded that for the 1989 data set there was no statistical difference between the risk probabilities generated by the original Framingham model and the modified model that employed the SR equation to estimate the individual CHD risk score.

Utility of the SR Risk Scores
The comparisons in Table 3 indicate that the SR risk equation performs as well as the original Framingham risk equation when applied to the 1989 data set. It is well known, however (Frees, 1996;
Neter, Kutner, Nachtsheim, & Wasserman, 1990) that regression equations should only be applied on populations that are (statistically) identical to the one from which they were developed. An important question with respect to the SR risk equation was whether it would yield meaningful estimates of population risk when applied to a typical, small n, self-report data set. To assess its utility, we applied the SR equation to the SOUR data set derived from the 1996 telephone survey. Because this data set did not include blood cholesterol or blood pressure measurements, the results from the SR equation could not be compared with a conventional Framingham model estimate of CHD risk. However, several lines of reasoning suggested that there should be little difference in CHD risk levels for the 1989 and 1996 data sets. First, comparison of findings from the two surveys from which the data sets were derived provided no evidence of significant changes in either the prevalence of smoking or sedentary lifestyle in the New Brunswick population between 1989 and 1996. Second, given the relatively short time period between the two surveys, and the absence of mass screening or treatment programs, one would not expect average blood pressure or blood cholesterol levels in the population to have changed markedly. Finally, examination of hospital admission statistics for AMI (i.e., crude rate per 1,000 population) revealed no consistent positive or negative trend in the annual number of MU admissions between 1989 and 1996. Thus, given the lack of evidence suggesting any significant change in population CHD risk between 1989 and 1996, we assessed the ‘utility’ of the SR equation in terms of the extent to which it yielded similar estimates of population CHD risk for the 1989 and 1996 data sets.

In order to apply the SR risk equation to the 1996 data set, one modification was required. To adjust for the absence of information on diabetes in 1996 data set, subjects were assigned the mean 'diabetes' prevalence index from their age and sex category in the 1989 sample. Statistically, this is tantamount to assuming that the 'distribution' of the diabetes prevalence index in the two samples is similar. In simpler terms, this would be -equivalent to assuming, as we have, that the two samples are roughly equivalent (in a random sense) in terms of their representation of the target population in New Brunswick. Table 4 presents the 10-year, SR-based CHD risk probabilities for the 1989 and 1996 data sets.

A two-way ANOVA (age category by data set) was used to compare the CHD probability estimates from the two data sets. The main effect for age category (F4,1585 347, p < .001) was statistically significant, indicating that there were differences in risk probability levels between the various age groups. However, neither the main effect of data set nor the data set by age group interaction was significant at the .05 probability level. The consistency of the risk estimates from
the two data sets suggests that the SR equation provided useful estimates of population CHD risk even when utilized with the small telephone survey data set.

CONCLUSIONS AND LIMITATIONS
The present study demonstrated the feasibility of using self-report measures to assess CHD risk or heart health in target populations. The SR risk equation developed in this study yielded population risk estimates that closely approximated risk estimates calculated using the conventional Framingham risk model. The SR risk equation was also found to perform well when utilized with the type of telephone survey data that is likely to be available in smaller heart health promotion programs. These findings strongly suggest that with further research it may be possible to develop a relatively inexpensive SR-based methodology that could be used as an alternative to conventional multivariate risk prediction models. Development of such a methodology could be of benefit to smaller-scale heart health promotion initiatives, which may not have the resources to collect physiological measures in their target populations. An added advantage of an SR-based risk prediction model is that by focusing on behavioral measures it may provide a more direct method of demonstrating the link between typical community-based intervention activities (e.g., smoking reduction, healthier eating, increased physical activity) and population CHD risk reduction.

The present study has a number of limitations. Perhaps the most important of these is that the data set used to develop the SR equation was never intended to serve as a source of information on self-reportable CHD risk indicators. As a result, several variables that should be important predictors of CHD risk, such as family history of heart disease, leisure time physical activity, and possibly eating habits, were not included in the SR risk equation. In addition, the predictive utility of some of the information used in the equation, such as the blood pressure and diabetes information, could no doubt be improved through the use of better-designed self-report questions. Given these limitations, the SR equation developed in the present study should not be viewed as a finished product, but rather as an initial step in a line of research that may lead to the development of a viable SR-based population heart health assessment model. The objective of this work was to explore the feasibility of using SR measures to assess the heart health status of target populations. The development, within the limitations of the available data, of an SR-based risk prediction equation capable of predicting CHD risk as well as the Framingham model clearly supports the feasibility of this approach to population CHD risk assessment.

A second major limitation of this study was the inability to fully explore the generalizability of the SR equation. Although the SR equation results with the 1996 telephone survey data set

<table>
<thead>
<tr>
<th>Age Group</th>
<th>1989</th>
<th>1996</th>
</tr>
</thead>
<tbody>
<tr>
<td>30-34</td>
<td>.011</td>
<td>.009</td>
</tr>
<tr>
<td>35-44</td>
<td>.026</td>
<td>.032</td>
</tr>
<tr>
<td>45-54</td>
<td>.076</td>
<td>.074</td>
</tr>
<tr>
<td>55-64</td>
<td>.132</td>
<td>.139</td>
</tr>
<tr>
<td>65-74</td>
<td>.181</td>
<td>.173</td>
</tr>
<tr>
<td>75-78</td>
<td>.271</td>
<td>.271</td>
</tr>
</tbody>
</table>
suggest that the methodology is relatively robust, the findings from that component of the study must be interpreted with caution, as they are based on the unconfirmed hypothesis that there was no significant change in population CHD risk between 1989 and 1996. In future research, it will be important to test the utility of SR-based risk prediction models in different populations where actual CHD risk levels can be more objectively defined.

Additional research is necessary to develop a working SR-based methodology for assessing population heart health, but the results of the present study illustrate the potential viability of this approach. We estimate that using self-report measures to assess population CHD risk is at least 65% less expensive than the conventional physiological risk factor methodology. As an economical alternative to conventional multivariate risk factor equations, an SR-based methodology may eventually enable researchers involved with smaller-scale heart health promotion initiatives to more precisely model the potential long-term impact of their intervention activities and extend typical program evaluation activities beyond the level of short-term, individual-risk-factor tracking. As all provinces in Canada have completed population surveys similar to those employed in the present study, there should be widespread opportunity to further explore the potential for using self-report risk prediction equations in the evaluation of-community-based heart health promotion programs.

ACKNOWLEDGEMENTS
This work was supported in part by grant number OGP 012 5689 from the Natural Sciences and Engineering Research Council, awarded to J. Bhadury. This support is gratefully acknowledged. The authors also wish to thank the anonymous referees for their comments and suggestions.

REFERENCES


