Multivariate Analysis of Risk Factors for Coronary Heart Disease

By Lars Wilhelmsen, M.D., Hans Wedel, Ph. Lic., and Gösta Tibblin, M.D.

SUMMARY
For a prospective study of risk factors for coronary heart disease (CHD), defined as acute nonfatal myocardial infarction and acute fatal CHD, 973 men, all aged 50 years, were recruited from a general Swedish urban population. Of the 855 who agreed to participate, 834 who showed no signs of CHD were selected and observed for nine years and four months. Autopsies were performed on all except two of the 55 patients who died during the study. Forty-four men developed clinical manifestations of CHD during this time; 19 of them died.

Using a multiple logistic model, we analyzed nine probable risk factors. The presence of high serum cholesterol, smoking, high systolic blood pressure, dyspnea, and registration by the Temperance Board increased the risk of events related to CHD significantly. The presence of high serum triglycerides, high hematocrit readings, and increased geographic mobility did not cause a significant increase in events related to CHD. Those with sedentary jobs showed a slightly increased tendency to develop manifestations of the disease.

The predictive power of the logistic function (with cholesterol, smoking, and systolic blood pressure) was tested in another randomly selected population sample of 5,146 men aged 51-55 years and found to be very accurate. The 10% (decile) of the population that had the highest risk of clinical manifestations of coronary disease and the 10% who had the lowest risk were defined. Manifestations of coronary disease occurred 29 times more frequently in the highest decile than in the lowest decile.

Additional Indexing Words: Cholesterol, Triglycerides, Systolic blood pressure, Smoking, Dyspnea, Record of intemperance, Physical activity, Hematocrit, Geographic mobility

The Risk Factor Concept in coronary heart disease (CHD) is based on the finding of statistically significant associations between incidence of CHD and values for the variables in question. Studies performed so far have not shown that any of the risk factors are actually causative. It is likely that some of them are not causative and merely reflect increased risk via correlation with other risk factors. The investigations of the etiology of CHD call for intervention trials in man where the effects of the intervention can be compared with the effects of nonintervention. This experimental design necessitates selecting high risk individuals for treatment. Patients with a limited number of easily identifiable and treatable risk factors have to be isolated.

The multiple logistic model is a theoretically acceptable method of analysis of the risk dependence of several variables. In this article such a method has been used in a prospective population study of men of the same age. Only “hard” medical criteria—nonfatal acute myocardial infarction (AMI) or sudden deaths unquestionably caused by CHD—were accepted for diagnoses of CHD. Univariate analysis of single risk factors for AMI in this population has been published previously1,2 that disclosed associations between incidence of AMI and high values for serum cholesterol, high serum triglycerides, high blood pressure, smoking, a history of alcoholic intemperance, working at a sedentary occupation, high geographic mobility, and appearance of dyspnea on exertion.

Study Population and Methods
All inhabitants of Sweden have a national registration number that contains their date of birth and other
vital statistics. Names, addresses, and registration numbers are registered with the official county census and were accessible before the sample was drawn for the present study. The study population for the multiple logistic model was recruited from men born in 1913, who were still alive at age 50 (1963) and living in Göteborg, Sweden. All men with these criteria who were born on dates divisible by three—thus, the third, sixth, ninth, etc. days of each month—comprised the study sample. Nine hundred and seventy-three men fit these criteria. Of these 855 (88%) agreed to be examined in 1963 at Sahlgren's Hospital, Göteborg.9

Among the nonparticipants, seven had died, four were admitted to the hospital, and nine did not live in the town at the time of investigation. The remaining group (98 men) refused to come for various reasons, usually because of a negative attitude to medical care. The nonparticipants had a lower mean income, were more frequently unmarried, and were more often registered with the Temperance Board than those of the participation group. The nonparticipants had obtained sick allowance to a slightly higher degree than the others.4 The proportion of nonparticipants is considered low, and the results are probably not affected by this group to any significant degree.

This prospective analysis which allows the development of the model is based on study of 834 men. Twelve men who had had clinical AMI and eight with angina pectoris but no AMI at the first examination in 1963 were excluded from the original group of 855, as well as one man (not belonging to the group that later suffered CHD) for whom complete data for the present analysis was not available.

A detailed medical history was obtained and clinical examination performed by one physician throughout the initial study. The examination was performed in the morning, the subjects fasting until after the blood sampling.

As criteria for diagnosis of acute CHD, hospitalization with the clinical diagnosis of AMI or fresh CHD at autopsy were used. The diagnoses were established by the clinicians or pathologists, respectively. Since November, 1968, all clinical cases of AMI have been registered in a special World Health Organization myocardial infarction register.6 The clinical criteria for AMI were those adopted by the Swedish Society of Cardiology: central chest pain, shock or syncope suggesting an AMI together with a typical transaminase spectrum and/or appearance of a pathological Q wave or localized ST variations in the ECG. Criteria for CHD at autopsy were a fresh myocardial scar, or in the absence of any macroscopic scar, a total or almost total occlusion of a coronary artery together with a medical history suggesting AMI or sudden coronary death.

Between entry in 1963 and October, 1972, when this analysis began, 55 subjects died. Autopsies were performed on 53 subjects. Autopsies were not done on one subject who died in a traffic accident and on a second person whose sudden death was believed not to have been caused by coronary arterial disease. The end points characterizing the development of clinical events produced by CHD were judged to be sufficient for our purposes. There were 19 fatal cases of CHD and 25 nonfatal AMI.

Those factors which had been shown to be associated with increased risk of CHD in this series (smoking, dyspnea on exertion, physical inactivity, high blood pressure, and presence of high serum cholesterol) and some which we suspected might be related to CHD (high geographic mobility and excessive alcohol consumption) were included in the present analysis. Details of the measurements of the variables are presented elsewhere.5 Data concerning smoking habits, dyspnea on exertion, and physical activity at work were collected by the same interviewer (G.T.).

In the previous analysis,1 2 it was found that the 75 men who smoked only cigars, "cigarillos," or pipes as a group had a risk of CHD comparable to that of those smoking 15–24 cigarettes a day. Thus, it was judged most appropriate to code them according to tobacco consumption. One cigarette is estimated to be equivalent to 1 g of tobacco; 1 cigarillo = 2 g; and one cigar = 5 g tobacco. For pipe smokers, the average number of grams consumed daily was used. For the statistical analysis smoking was graded as follows: never smoked = 1; stopped smoking (= exsmoker) = 2; smoking 1–14 g/day = 3; smoking 15–24 g/day = 4; and smoking 25 g/day = 5.

Dyspnea was defined according to criteria set up by the World Health Organization.6 Persons experiencing dyspnea when hurrying on the level, or walking up a small hill, or undertaking other less taxing exertion were grouped as 1 for this variable; others were scaled 0.

Physical activity during work was scaled as follows: sedentary = 1; moderate = 2; and heavy = 3.

Blood pressure was measured by the same observer throughout the study after a 5–10 min interview with the subject in the sitting position. Both systolic and diastolic blood pressure were tested in the preliminary analysis and as there was a strong correlation between the two readings and systolic blood pressure showed a higher predictive capacity for incidence of CHD, this measure alone was used in the further analysis. Those men who had casual systolic pressures ≥ 175 mm Hg and diastolic ≥ 115 mm Hg were treated, a fact that has to be borne in mind when high blood pressure is discussed as a risk factor in this study.

A factor reflecting geographic mobility was included. Men born outside the city of Göteborg and its surroundings (scale: 1) and particularly foreigners (scale: 2) had been found to have a higher risk for CHD on univariate analysis than those born within the study area (scale: 0).

Reliable data concerning alcohol consumption for individual persons are very difficult to obtain from interviews. In Sweden, however, records of local Temperance Boards, established by a special law, provided an index of alcoholic impenance. The men in the present study (21%) who had been registered by the Temperance Board at some time in their life were scaled 1; the remainder were scaled 0. A limitation of this measure is that persons registered because of a series of minor offenses are in the same group as those with a record of chronic alcohol abuse. As this study population was a random sample of the population, the
same figure (21%) is valid for the total male population in the town and probably other large towns in Sweden.

The levels of cholesterol and triglycerides in the serum and whole blood hematocrit readings were determined by the same laboratory throughout the study, and regular checks to determine the accuracy of these methods were performed. The actual values for these factors were compared in this analysis.

The predictive power of the function obtained in this sample of men was tested in another randomly selected population sample of 5,146 men aged 51–55 years (mean age 53.2 years). The examination of this sample began in January, 1970, with men born in 1915, then with those born in 1916, and so on. The examination of those born in 1921 was completed in May, 1972. Each age group was drawn from the population register one to two weeks before the subjects were called for the investigation. The total sample was composed of 7,063 men of whom 5,361 (76%) participated in the examination. Two hundred and fifteen men had clinical manifestations of CHD according to the above criteria at examination. Thus, the sample for study included 5,146 men. Registration of new events of CHD was accomplished with the aid of the above-mentioned myocardial infarction register.

The three variables, smoking, serum cholesterol, and blood pressure, were measured using the same methods as above with the exception that the subjects were examined in the afternoon (4:30–7:00 p.m.) instead of in the morning. Some preventive measures against high blood pressure (as in the first mentioned series), high cholesterol, and tobacco smoking were instituted in this group after the examination. During a mean follow-up time of about two years, 44 new manifestations of CHD occurred.

**Statistical and Mathematical Methods**

It is widely agreed, and confirmed in this population, that a number of variables in combination are associated with increased susceptibility of a person to CHD. It is impossible to isolate the statistical "effect" of any one of these risk factors in humans. But by applying appropriate methods such as multiple logistic regression, it is possible to study the effect on CHD when one variable changes and the others are held constant. The logistic model is described by Cox and has been used in heart disease investigations by Truett, Cornfield, and Kannel, Duncan and Walker, Keys et al, and Kleinbaum. (See Appendix to this article for a discussion of the statistical model.)

For example, if we look at the probability of CHD as a function of the level of triglycerides and find that the probability increases with increasing triglyceride values, it is quite impossible to draw any conclusions as to cause and effect from this analysis. This relationship may merely reflect the positive correlation between level of triglycerides and cholesterol in the blood, if cholesterol is associated with CHD.

**Results**

The mean and standard deviations for the independent variables in groups with and without CHD for the 834 subjects are shown in table 1. Table 2 shows the percentage distribution for variables that were graded on scale ranges discussed above.

The correlation coefficients between the independent variables are shown in table 3. The largest numerical value is for the correlation between cholesterol and triglycerides (0.36).

The estimated coefficients in the logistic model are shown in table 4 performed with nine and three independent variables, respectively. The variables are ranked according to their associated t values \((b_j/s_j)\). To the left all nine variables are included and to the right only the three variables which are most closely related to CHD. The \(b_j\) coefficients give an indication of the role of each variable

**Table 1**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total group N = 834</th>
<th>No CHD N = 790</th>
<th>CHD N = 44</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>sd</td>
<td>Mean</td>
</tr>
<tr>
<td>Cholesterol (mg/100 ml)</td>
<td>246.8</td>
<td>42.6</td>
<td>245.3</td>
</tr>
<tr>
<td>Smoking</td>
<td>2.57</td>
<td>1.15</td>
<td>2.52</td>
</tr>
<tr>
<td>Systolic blood pressure (mm Hg)</td>
<td>138.2</td>
<td>20.9</td>
<td>137.7</td>
</tr>
<tr>
<td>Dyspepsia</td>
<td>0.21</td>
<td>0.41</td>
<td>0.20</td>
</tr>
<tr>
<td>Record of intemperance</td>
<td>0.21</td>
<td>0.41</td>
<td>0.20</td>
</tr>
<tr>
<td>Physical activity during work</td>
<td>1.93</td>
<td>0.86</td>
<td>1.94</td>
</tr>
<tr>
<td>Hematocrit (%)</td>
<td>44.9</td>
<td>3.3</td>
<td>44.9</td>
</tr>
<tr>
<td>Triglycerides (mmol/l)</td>
<td>1.25</td>
<td>0.71</td>
<td>1.24</td>
</tr>
<tr>
<td>Place of birth</td>
<td>0.36</td>
<td>0.58</td>
<td>0.37</td>
</tr>
</tbody>
</table>

**sd = standard deviation; N = number in group.**
Table 2

Percentage Distribution of Graded Risk Variables in Development of CHD

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total group N = 834 Grade*</th>
<th>No CHD N = 790 Grade*</th>
<th>CHD N = 44 Grade*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 1 2 3 4 5</td>
<td>0 1 2 3 4 5</td>
<td>0 1 2 3 4 5</td>
</tr>
<tr>
<td>Smoking</td>
<td>24 20 35 17 4</td>
<td>25 20 34 17 4</td>
<td>7 9 50 25 9</td>
</tr>
<tr>
<td>Dyspnea</td>
<td>79 21 80 20</td>
<td>59 41</td>
<td></td>
</tr>
<tr>
<td>Record of intemperance</td>
<td>79 21 80 20</td>
<td>61 39</td>
<td></td>
</tr>
<tr>
<td>Physical activity during work</td>
<td>2 34 32 32</td>
<td>7 32 43 18</td>
<td></td>
</tr>
<tr>
<td>Place of birth</td>
<td>69 26 5 68 27 5</td>
<td>77 16 7</td>
<td></td>
</tr>
</tbody>
</table>

*See text for value ranges assigned to each grade on the scales.

Table 3

Correlation Coefficients between the Independent Risk Variables

<table>
<thead>
<tr>
<th></th>
<th>Cholesterol</th>
<th>Smoking</th>
<th>Systolic blood pressure</th>
<th>Dyspnea</th>
<th>Record of intemperance</th>
<th>Physical activity during work</th>
<th>Hematocrit</th>
<th>Triglycerides</th>
<th>Place of birth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cholesterol</td>
<td>1.00</td>
<td>0.03</td>
<td>0.11</td>
<td>0.01</td>
<td>-0.02</td>
<td>0.04</td>
<td>0.04</td>
<td>0.36</td>
<td>-0.03</td>
</tr>
<tr>
<td>Smoking</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systolic blood pressure</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dyspnea</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Record of intemperance</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical activity during work</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hematocrit</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Triglycerides</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Place of birth</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
relative to all other variables in increasing the risk of acquiring CHD. High serum cholesterol, smoking, high systolic blood pressure, dyspnea, and alcoholic intemperance were significant. High triglyceride and hematocrit readings and geographic mobility (column head: Place of birth) did not significantly affect predisposition to CHD. Physical inactivity during work showed a slight tendency to increase the risk.

Univariate analysis of risk of CHD for persons who smoked cigarettes only produced the same results as when pipe and cigar smokers were included, according to their tobacco consumption, in the groups of smokers. Thus, we concluded that the multiple logistic function would not be affected if pipe and cigar smokers were excluded.

Both systolic and diastolic blood pressures were tested but systolic blood pressure had higher predictive capacity and high diastolic blood pressure did not change the risk when systolic blood pressure had been computed in the multivariate function.

Similarly, serum triglycerides gave no additional information useful for predicting the risk when values for serum cholesterol were available. The errors in prediction were greatly increased when the values for level of triglycerides were used in place of serum cholesterol level.

In order to evaluate the fit of the estimated logistic functions to the observed cases of CHD, the risk probability for the total population sample was divided into decile classes from the lowest to the highest risk (table 5). In part A of table 5 the risk probability was estimated by using the coefficients of the left column in table 4 (nine variables) and in B the right column (three variables) was used. One finds that the fit is nearly as good in B as in A. The three variables cholesterol, smoking, and blood pressure have the advantage of having paired correlations which are very low (table 3).

In C of table 5, the series of 5,146 persons is analyzed. The total incidence of new CHD for this age group of men followed for two years was estimated from a careful registration of CHD in this population. The estimation was based on the values for cholesterol and blood pressure and on smoking habits for each subject. It will be seen from the table that the observed and estimated numbers are very similar.

The majority of CHD cases were concentrated in the highest deciles, but some cases were found even in the lowest ones. It was not possible to dichotomize the study population into two groups, where the estimated probability of CHD was large in one and small in the other. Thus, if we tried to discriminate between subjects who developed CHD or not, we found a large probability of misclassification.

According to these results, it is possible to estimate the probability of a 50-year-old man suffering CHD before the age of 60 years when his values for smoking, cholesterol, and systolic blood pressure are known. Nomograms can also be constructed for easy use. One example is shown in figure 1 giving the probabilities of CHD for subjects with various smoking habits and cholesterol values with systolic blood pressure of 140 mm Hg. It can, for example, be seen that the probability of CHD for a man smoking 15–24 cigarettes a day is about three times the probability for an exsmoker.
RISK OF CORONARY HEART DISEASE

Table 5
Estimated and Observed Cases of Events Associated with CHD by Decile Class in the Original Study Population of 885 Men with all Nine Variables (A) and with Three Variables (B), and in Another Population of 8,146 Men With Three Variables (C)

<table>
<thead>
<tr>
<th></th>
<th>Decile class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td></td>
</tr>
<tr>
<td>No. of persons in decile</td>
<td>83</td>
</tr>
<tr>
<td>$1000 \times$ estimated risk probability*</td>
<td>6</td>
</tr>
<tr>
<td>No. of no CHD</td>
<td>83</td>
</tr>
<tr>
<td>No. of CHD</td>
<td>0</td>
</tr>
<tr>
<td>Estimated no. of CHD</td>
<td>0.5</td>
</tr>
<tr>
<td>B</td>
<td></td>
</tr>
<tr>
<td>No. of persons in decile</td>
<td>83</td>
</tr>
<tr>
<td>$1000 \times$ estimated risk probability*</td>
<td>9</td>
</tr>
<tr>
<td>No. of no CHD</td>
<td>83</td>
</tr>
<tr>
<td>No. of CHD</td>
<td>0</td>
</tr>
<tr>
<td>Estimated no. of CHD</td>
<td>0.7</td>
</tr>
<tr>
<td>C</td>
<td></td>
</tr>
<tr>
<td>No. of persons in decile</td>
<td>514</td>
</tr>
<tr>
<td>$1000 \times$ estimated risk probability*</td>
<td>1.2</td>
</tr>
<tr>
<td>No. of no CHD</td>
<td>514</td>
</tr>
<tr>
<td>No. of CHD</td>
<td>0</td>
</tr>
<tr>
<td>Estimated no. of CHD</td>
<td>0.7</td>
</tr>
</tbody>
</table>

*See table 4 for list of variables and estimated coefficients.

Discussion

High age and male sex are probably the two most important risk factors for CHD. In this study we have had the opportunity to study a sufficient number of subjects of the same sex and age.

Analysis of risk factors using the logistic model does not assume any specific distribution of the variables. The normalization of the variables (see Appendix) makes it reasonable to compare the estimated parameters $b_i$ from similar investigations. The $t$ value is however dependent on the number of individuals studied.

As in other studies using similar models,¹⁰⁻¹³ high cholesterol and high blood pressure (systolic or diastolic) have been proved important risk factors. In this study, smoking emerged as a third risk factor of great importance, as it has in the studies of Truett et al.¹⁰ and Kleinbaum et al.¹³ In other studies smoking has been found much less important than cholesterol and blood pressure.¹²

Keys et al.¹² have also estimated the coefficients in a logistic model. When they included only cholesterol and systolic blood pressure, the $b$-coefficients were 0.85 and 0.21, respectively, which is of the same magnitude found in this investigation. However, the $b$-coefficient for smoking (0.28) in the Keys study was substantially different from our calculation of 0.503 (when nine variables were used) and 0.663 (when three variables were used). The Keys analysis included the following four variables: cholesterol, systolic blood pressure, weight/height index, and cold pressor test, but as these variables were weakly correlated with smoking, the great difference between $b$-coefficients for smoking between the two studies cannot be explained by these differences in variables used. A difference that may explain varying results for the risk of smoking is the method of scaling, which has not been the same in the two studies.

High blood pressure was a significant risk factor in this multiple analysis and the association was stronger than that found in the earlier univariate analysis. This finding is of especial interest as those men who had high blood pressure in 1963 were treated and have been continuously treated since 1963. It is possible that high blood pressure would have been even more strongly connected with CHD if no treatment had been given. However, intervention studies in which hypertension is treated have not shown any significant effects on incidence of events of CHD.¹⁶

Circulation, Volume XLVIII, November 1973
In this study the level of triglycerides were studied prospectively. In previous prospective studies\textsuperscript{1, 2, 17} and one comparative study of nonselected patients with nonfatal AMI and a random sample from the general population,\textsuperscript{18} the level of triglycerides have been shown to be important predictors of CHD. Multiple logistic models were not used. Our results clearly indicate that if cholesterol readings are available, other tests than triglycerides should be chosen for predicting the risk of CHD. It must be stressed that these results cannot be extrapolated to other age groups or to women. Thus, triglycerides may have stronger predictive power in younger age groups especially. The present findings are of practical importance as triglycerides have to be determined in fasting serum. Screening can only be done in the morning.

The association between cholesterol, blood pressure, and smoking on the one hand and incidence of events associated with CHD on the other may well fit a hypothesis that these factors are causally related to CHD. The association between dyspnea on exertion and CHD probably indicates that those men who complain of dyspnea are already suffering from some kind of cardiovascular or pulmonary dysfunction. It should be stressed that the dyspnea was not associated with roentgenologically detectable increase of heart size. The mean vital capacity was, however, reduced to 96% and mean forced expiratory volume in one second to 95% of the predicted normal in the group of men with dyspnea. It is interesting to note that dyspnea before an AMI has been shown to be associated with a higher mortality after an AMI.\textsuperscript{19}

A record of intemperance and its association with CHD has not been studied prospectively before. The present findings of an association with CHD may indicate an increased risk from increased alcohol consumption when the other eight variables of this analysis are kept constant. The association may, however, also be due to a more complex variable, for example, of psychosocial kind, which in turn might be correlated to intemperance. Further studies are necessary to elucidate this aspect.

Physical inactivity during work showed a slight tendency to increase the risk of CHD. An analysis of the importance of leisure time activity has been judged of great interest, too. It was not measured until 1967, however, and is therefore not included in the present analysis.\textsuperscript{2} As a series of other risk factors, for example, smoking and a record of intemperance, are related to physical activity during work and leisure time, multivariate methods are most appropriate for such an analysis.

Hereditary factors other than those affecting the blood lipids—for example, enzymes affecting lipid metabolism of the arterial wall—may be important.
in the development of the arteriosclerotic process, but studies in this area have not been performed in this sample.

Blood coagulation and fibrinolysis were not studied in this population sample until 1967 and the possible role of these variables is not included in the present analysis. Coffee consumption, which varied between 0 and 20 cups per day, was unrelated to CHD both on univariate and multivariate analyses (unpublished data).

In the analysis above the risk probability is a function of the variables measured at the beginning of the observation time. It is known that such variables as cholesterol level, blood pressure, and smoking habits change over time and this must be taken into consideration when we extend the observation time. This is of special importance when one tries to predict the risk probability.

The predictive power of the multivariate logistic function was tested in the sample of 5,146 men. The prediction of the number of CHD cases occurring in certain decile groups of individuals in this study population was found to be very accurate. The probability of suffering CHD in the highest risk decile group, which had nearly twice as high a value for risk as the next highest decile group, is 29 times higher than for the lowest decile (table 5). On the other hand, it seems to be quite impossible to adopt the conventional terms "sensitivity" and "specificity" to these findings. Even in the group with the most severe risk only about 20% experienced a medically proven CHD event during the follow-up period of nearly ten years. Nevertheless, no better prediction function has yet been published.

It is evident from this study that the three most powerful predictors for CHD—smoking, high blood pressure and high serum cholesterol—are more or less treatable factors. This study, as well as a series of others, does not prove cause and effect relationship. Neither can it tell us whether treatment of these factors can prevent CHD. We can say, however, that these three findings are the three most important factors (in Sweden) which should be manipulated in primary preventive trials. These trials may contribute to solving the question of causation in coronary heart disease.

Appendix

Discussion of the Statistical Model

In this investigation we studied the probability of a 50-year-old man developing CHD for different values of cholesterol, smoking, blood pressure, and other factors during nine years and four months. Denote the values of these variables \( x_1, \ldots, x_p \) and the probability of developing myocardial infarction \( \{ p(x_1, x_2, \ldots, x_p) \} \) for given values of \( x_1, \ldots, x_p \).

It is reasonable to suppose that this probability increases in some variables in the formula and decreases in some other variables.

One way to satisfy these conditions is to put

\[
p(x_1, \ldots, x_p) = \frac{e^{\lambda + \sum_{i=1}^{p} \beta_i (x_i - \bar{x}_i)}}{1 + e^{\lambda + \sum_{i=1}^{p} \beta_i (x_i - \bar{x}_i)}},
\]

where \( \beta_i \) are unknown coefficients and \( \bar{x}_i \) and \( s_i \) are the means and estimated standard deviations of variable \( i \). We have normalized the variables \( x_i \) and inserted \( x_i - \bar{x}_i \) in the logistic function above.

This is done in order to make the quantity \( \beta_i \) invariant with respect to the scale of measurement of \( x_i \). The probability of a person of a certain age developing CHD within a period of time length \( t \) is of course a function of \( t \). This indicates that the parameters \( \beta_i \) depend on \( t \). However, if we suppose that the distribution of the \( x \)-variables for the CHD group and the non CHD group is invariant during a certain time, it is possible to show that \( \beta_i \) is constant in time for \( i = 1, \ldots, p \) and only \( \beta_5 \) depends on \( t \). This suggestion makes it reasonable to compare \( \beta_5 \) from similar investigations performed over longer and shorter observation periods.

An easily obtained consequence of the model is that if we change merely the variable \( x_i \) by \( s_i \) we get a change in \( \lambda \) of precisely \( \beta_i \). Thus, a large value of \( \beta_i \) means that the corresponding variable is a strong predictor of the probability for CHD.

We could of course have introduced combinations of variables, for example \((\text{cholesterol}) \times (\text{systolic blood pressure})\) or \((\text{systolic blood pressure})\). As the simultaneous confidence intervals grow very quickly when many new parameters are included we have tried to reduce the unknown parameters as far as possible. Kleinbaum et al. have shown that the only second order combinations of interest in their investigations were \((\text{diastolic blood pressure}) \times (\text{age})\) and \((\text{cholesterol}) \times (\text{age})\). All our subjects were of the same age—50 years on entry. Thus, we have reason to exclude these combinations of variables in this investigation.

There is a connection between \( p(x_1, \ldots, x_p) \) and the distribution of the \( x \)-variables. For the normal discriminant model one requires that the \( x \)-variables have a multivariate normal distribution with a change in location for the group with events associated with CHD. The logistic model holds exactly in this case and is a good approximation in many other interesting cases even when some of the \( x \)-variables are discrete, as in this investigation.

Estimation and Test of Parameters

The unknown parameters \( \beta_i \) in the logistic model are estimated by the maximum likelihood method and these estimates are denoted by \( \hat{\beta}_i \). We have used an
iterative method to find the estimates. This procedure does not assume any particular distribution of the x-variables. One way of obtaining rough estimates of the logistic parameters is to use the ordinary estimates in the normal discriminant model. We finally did not use them because of the hard restrictions in this model and the bias in these estimates but we have used them as starting values in the iterative procedure. The type of tests and level of significance can be discussed. We have chosen onedist tests at the 0.01 level. The maximum likelihood procedure implies that all t values are approximately normally distributed (t = b_i/s_i), where s_i denotes the standard deviation of b_i. Thus, we reject \( b_i \leq 0 \) if \( t > 2.3 \). As can be seen from table 4, tests on the 0.05 level, instead of on the 0.01 level, should not have affected the conclusions to any considerable extent.

Acknowledgments

Thanks are due to Bo Justusson, M.Sc., Björnettersson, M.Sc., Hans Jönrup, Ph.Lic., and Anders Odén, Ph. Lic., for valuable discussion of the statistical methods and for assistance with data processing.

References

16. Veterans Administration Co-operative Study Group on Antihypertensive Agents: Effect of treatment on morbidity in hypertension: Results in patients with diastolic blood pressure averaging 115 through 129 mm Hg. JAMA 202: 1028, 1967

Circulation, Volume XLVIII, November 1973
Multivariate Analysis of Risk Factors for Coronary Heart Disease
LARS WILHELMSEN, HANS WEDEL and GöSTA TIBBLIN

Circulation. 1973;48:950-958
doi: 10.1161/01.CIR.48.5.950
Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 1973 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/48/5/950

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation is online at:
http://circ.ahajournals.org//subscriptions/