

## Determinants of all causes of death in samples of Italian middle-aged men followed up for 25 years

A MENOTTI, S MARIOTTI, F SECCARECCIA, S TORSELLO, AND F DIMA

*From the Laboratory of Epidemiology and Biostatistics, Istituto Superiore di Sanità, Rome, Italy*

**SUMMARY** A total of 1712 men aged 40 to 59 years in two rural cohorts of northern and central Italy have been followed up for 25 years after an entry examination in 1960. Forty one individual characteristics have been considered as possible predictors of death in the next 25 years. After exclusion of 55 men with life threatening diseases (cardiovascular and cancer) and of 161 men because of missing measurements, 1495 men have been analysed for relation between entry factors and subsequent death ( $n=670$ ). Twelve factors eventually emerged as powerful predictors of future death: in hierarchical order, age, blood pressure, forced expiratory volume, cigarette smoking, xanthelasma, mother life-status, arm circumference, father life-status, shoulder-pelvis ratio, vital capacity, arcus senilis, and serum cholesterol. Discrimination as provided by logistic modelling placed 19.6% of all cases in the upper decile of the estimated risk, 36.8% in the upper quintile, 2.5% in the lowest decile, and 7.1% in the lowest quintile. Out of those located in the lowest decile of risk, 11.4% died within 25 years while the corresponding percentage in the upper decile was 87.3%. Use of the Cox model yielded slightly better coefficients than logistic function.

The long-term follow-up of population samples, even though examined only once, allows us to collect data on a large number of deaths and to study the relation of entry characteristics to the occurrence of such fatal events. When follow-up is particularly prolonged, interest in the prediction of a single disease is overwhelmed by that for survivorship *v* mortality, whatever its cause might be.<sup>1-10</sup> This apparently simple approach may produce useful information on the risk function of the most important fatal events, although there are a number of limitations.

No account is taken of the possible changes with time of the entry characteristics possibly related to mortality, nor of changes of health care which can influence the natural history of major diseases. Mathematical models that are able to include and exploit such kinds of information are starting to become available,<sup>11-13</sup> but their application is still difficult, and only exceptionally is a set of data immediately suitable for that purpose. In particular, studies with many multiple measurements of factors and high participation rates are rare, whereas the quantitative expression of changes in the standard of medical care is difficult to assess at individual level. On the other hand, a simple study of the relation between entry levels of supposed risk factors and subsequent fatal events can be extended to most of the

common chronic diseases which represent epidemic conditions today, although only in their fatal manifestations.<sup>7 8 14</sup>

Another problem rests with the initial choice of the measured risk factors, usually conditioned by the kind of major end-points defined at the beginning of the study. However, knowledge that a number of risk factors usually measured in relation to cardiovascular disease are multipotential, ie, predicting other chronic conditions, is a positive aspect, when the study was originally designed to investigate cardiovascular diseases. This means also that studies originally designed for short-term investigation of single diseases can easily be converted into long-term enterprises directed to the study of different kinds of fatal diseases and to life expectancy.

Unfortunately, only a limited number of individual characteristics (risk factors) can be used, that is, only those whose measurement techniques, already suitable for epidemiological purposes, were known and available at the start of the study. This group has already investigated such a problem employing studies with 15 and 20 years of follow-up.<sup>7-9 14</sup>

The present study is a 25-year follow-up in which the number of deaths is not very different from the number of survivors. This has advantages for analysis. The first part deals with all causes of death combined.

## Materials and methods

The analysis refers to the two Italian rural cohorts studied since 1960 within the Seven Countries Study on Cardiovascular Diseases.<sup>2 5 6</sup> Men aged 40–59 were enrolled in two rural villages located in northern and central Italy, and those examined represented 98·8% of the defined samples, including all men of that age group living in specified areas, a total of 1712 individuals.

The entry examination included: a questionnaire on personal characteristics, living habits, parents' status, smoking habits, physical activity at work, the use of "special diets"; a number of anthropometric measurements; a simple respiratory test; an ECG at rest and after a 3 minute single step test; a blood pressure measurement; a standardised questionnaire on symptoms and diseases; a guided physical examination; a semiquantitative urinalysis; and the measurement of serum cholesterol. A number of personal characteristics have been selected for the present analysis: the methods and criteria employed as well as the units of measurements are reported in Appendix 1.

Throughout 25 years all men have been followed up using different procedures, but for the purpose of this paper only the collection of mortality and causes of death is relevant. This has been ascertained by periodically obtaining the living-death status from the Local Register Offices for all individuals, including those who have emigrated, and establishing a complex information gathering system from all possible witnesses, acquaintances, the police, and the coroner.

All such data have been converted into final causes of death, using pre-defined criteria. Exceptionally, original death certificates have been used for assigning the final cause of death, ie, in cases of absolute absence of other information. The cause of death has been assigned according to the 8th revision of the WHO-ICD.

For coronary heart disease, cases of sudden and unexpected death occurring within 2 hours from the onset of symptoms have been identified and labelled. In the case of apparent multiple causes of death, the following hierarchy priorities have been used: violent causes, cancer in advanced stages, coronary heart diseases, strokes, others.

In order to identify the predictors of fatal events, 41 different individual characteristics have been considered initially. They can be grouped into 10 different categories as follows:

- 1 *demographic data* (3) age; marriage status; number of children
- 2 *family data* (7) early mortality of mother and father; family history of myocardial infarction, of other heart diseases, of hypertension, of stroke, of

diabetes

- 3 *behavioural data* (3) physical activity at work; cigarettes smoked per day; the use of a prescribed diet
- 4 *anthropometric data* (7) body mass index; relative body weight; skinfold thickness; laterality-linearity index; shoulder-pelvis ratio; trunk/height ratio; arm circumference
- 5 *biochemical data* (3) serum cholesterol; urine protein; urine glucose
- 6 *spirometric data* (2) vital capacity; forced expiratory volume in 0·75 s
- 7 *ECG data* (2) ECG aspecific abnormalities at rest (in the absence of major heart disease); significant ECG abnormalities after exercise (in the absence of major ECG abnormalities at rest and/or heart disease)
- 8 *cardiocirculatory measurements* (3) systolic and diastolic blood pressure; heart rate
- 9 *physical data* (3) xanthelasma; arcus senilis; baldness
- 10 *clinical diagnoses* (8) heart disease, cerebrovascular disease; cancer; chronic bronchitis; diabetes; kidney disease; peptic ulcer; gall bladder disease.

All fatal events occurring in the 25 years of follow-up have been considered. However, analysis has been limited to a subset of the available population since two groups of individuals have been excluded: (1) those already known to be affected at the entry examination by life-threatening conditions such as heart disease, cerebrovascular diseases, or cancer, and (2) for multivariate analysis, those with missing data, even if missing only one of the chosen characteristics. Analysis relating entry characteristics to fatal events has been conducted in the following steps:

(a) A correlation matrix has been produced including all the 41 entry characteristics; this has enabled us to identify highly correlated factors, and, consequently, some exclusions have been made.

(b) A simple preliminary univariate analysis has been performed comparing mean levels (or proportions) of each factor for both living and the dead; this has led to further exclusions of clearly unrelated factors and of those likely to produce problems in the subsequent multivariate analysis. These two procedures have reduced the number of characteristics employed in the further analysis from 41 to 29.

(c) The 25 year all-causes mortality, with the remaining 29 factors as covariates, have been analysed using the proportional hazards model<sup>15</sup>; the relevant risk factors were selected using a forward step-wise technique (tolerance 0·05).

(d) The same mortality data, without regard to the timing of events, were analysed using multiple logistic function<sup>16</sup> and employing those factors selected as significant by the proportional hazards model.

The use of two different multivariate models will be justified later.

**Results**

Of the 1712 men examined in 1960, 1495 were used for the analysis; 55 were excluded as carriers at entry of life-threatening conditions, and 161 because of missing data.

Table 1 shows mortality for all causes and for some major causes in these three groups. Mortality rates are higher in the group affected at entry, but they are also higher in the group with missing data. Why this is so is not completely clear. However, it is known that a number of those with minor impairments were unable to carry out correctly either the ventilatory measurements or the ECG exercise test or both. The excess mortality is largely cardiovascular.

The correlation matrix between the 41 variables is not shown in detail. Altogether 22 pairs of factors had a correlation coefficient greater than 0.20 and smaller than 0.50, and five pairs had a value greater than 0.50. Inspection has guided us to delete some factors from being employed in the next analysis. Relative body weight has been excluded, since it is highly correlated with body mass index, and the latter is a better indicator of possible obesity. Skinfold thickness has been retained even though it is highly correlated with body mass index ( $r=0.75$ ) since it represents a particular aspect of body mass and can be considered an estimate of the amount of body fat.

Systolic and diastolic blood pressures are highly correlated with each other and can interfere with each other in multivariate models; they have therefore been compacted into mean blood pressure (diastolic +  $\frac{1}{3}$ (systolic—diastolic)), approximating to mean pressure occurring in the humeral artery through the cardiac cycle.

Vital capacity and forced expiratory volume in 0.75 s are correlated ( $r=0.60$ ) but they represent

different aspects of spirometry and both have been retained for further analysis.

Heart rate is somewhat correlated with blood pressure but it has been retained as an expression of a different phenomenon.

For the same reason the correlations between body mass index and blood pressure; body mass index and cholesterol; and cholesterol and blood pressure have been accepted although not negligible, and these factors have been retained.

Urine glucose has been compacted into the diagnosis of diabetes, as indicated in Appendix 1.

Univariate analysis comparing means (or proportions for dichotomous variables) of cases v non-cases has also been deleted due to its volume. Factors such as age, blood pressure, cigarette smoking, and vital capacity were significantly different in those now dead from those currently alive. Other factors are so far removed from any possible significance that they have been discarded. For example, family history of cardiovascular disease and diabetes, either broken down by individual condition (infarction, stroke, etc) or aggregated into a single score, has never shown any correlation with the considered end-point.

Gall bladder disease has been excluded for the same reason. In summary, the 12 factors excluded from further analysis are: relative body weight, a measure of blood pressure (the two being compacted), five items on family history, urine glucose (compacted into diabetes), gall bladder disease, heart disease, cerebrovascular diseases, and cancer (the last three being life-threatening diseases).

There were 670 fatal events among the 1495 men at risk. Table 2 summarises the set of risk factor coefficients provided by the proportional hazards model and by the multiple logistic function. Of the 29 factors, only 12 were taken into the step-wise proportional hazards model with a tolerance of 0.05. They are listed in the order of entrance. Appendix 2 reports the coefficients of all the 29 factors when forced into the model.

The coefficients of the first 12 factors are only slightly different from those provided by the step-wise approach.

Multiple logistic function, computed on the 12 factors selected by the Cox model, shows very similar coefficients as for the other model except that xanthelasma, vital capacity, and arm circumference are no longer significant. The Spearman rank correlation between the two sets of standardised coefficients is 0.64.

In table 2 the crude coefficients are accompanied by their *t* values and the standardised coefficients (coefficient times the standard deviation of the

Table 1 *Seven Countries Study. Italian rural areas. Causes of death in 25 years in men employed in the analysis, in men excluded because of missing values of some factors, and in men excluded because of life-threatening disease (prevalence): rate per 100*

	Analysed n=1495	Missing data n=161	Prevalence n=55
All causes	44.8	58.6***	80.0***
CHD	10.5	15.4	27.3***
Stroke	5.8	8.0	7.3
Cancer	14.2	17.3	18.2
Chronic bronchitis	2.0	2.5	1.8
Violence	2.5	1.2	5.5
Other cardiovascular	1.7	5.6***	10.9***
All other causes	8.1	8.6	9.0

\*\*\*p<0.001 (against analysed)

Table 2 Seven Countries Study, Italian rural areas. Solutions of two multivariate models predicting all causes of death in 25 years with 29 factors

Factors	Mean $\pm$ SD	Cox model step-wise			Multiple logistic function		
		Coeff.	t value	Stand coeff.	Coeff.	t value	Stand coeff.
1 Age (years)	49.65 $\pm$ 5.06	0.0899	10.36	0.4549	0.1318	9.79	0.6669
2 Mean blood pressure (mmHg)	104.15 $\pm$ 13.23	0.0217	7.57	0.2871	0.0278	5.78	0.3678
3 Forced expiratory volume	14.01 $\pm$ 3.20	-0.0507	-3.18	-0.1622	-0.0770	-3.24	-0.2464
4 Cigarettes/day	8.82 $\pm$ 9.55	0.0189	4.64	0.1805	0.0281	4.41	0.2684
5 Xanthelasma	0.01 $\pm$ 0.01	0.7221	2.89	0.0794	1.0334	1.77	0.1137
6 Mother status (Score)	1.84 $\pm$ 0.74	0.1625	2.92	0.1202	0.2204	2.66	0.1631
7 Arm circumference (mm)	269.40 $\pm$ 23.40	-0.0059	-3.43	0.1381	-0.0044	-1.67	-0.1030
8 Father status (Score)	1.91 $\pm$ 0.71	0.1780	3.05	0.1264	0.2376	2.77	0.1687
9 Shoulder pelvis ratio	136.30 $\pm$ 7.15	0.0165	2.93	0.1180	0.0233	2.73	0.1666
10 Vital capacity	23.10 $\pm$ 3.56	-0.0357	-2.55	-0.1271	-0.0315	-1.48	-0.1121
11 Arcus senilis	0.14 $\pm$ 0.34	0.2463	2.48	0.0837	0.3927	2.23	0.1335
12 Serum cholesterol	5.22 $\pm$ 1.05 mmol/l 201.66 $\pm$ 40.57 mg/dl	0.0850 0.0022	2.27	0.0892	0.1197 0.0031	2.13	0.1258
Constant	—	—	—	—	-11.6970	-6.77	—

variable). The latter enables direct comparisons to be made between the different coefficients.

The statistical contribution of the individual factors is shown again in table 3, where the 12 main factors are accompanied by the levels of the improvement of chi square and of global chi square along the steps of the step-wise procedure with the proportional hazards model.

It can be seen that age contributes greatly to the prediction of events and accounts for more than half of the global chi-square (table 3).

However, the contribution of other factors is not negligible. Table 4 shows the risk of dying in the 25 years (taking the risk for mean levels equal to one) for different levels of each factor, all the others being kept constant. As a reference the mean values and the levels corresponding to  $\pm 1$  standard deviation have been considered. Adjustment has been made for some dummy variables. For men 5 years older than the mean there is an excess mortality of 36%, whereas

Table 3 Seven Countries Study, Italian rural areas. Summary of proportional hazards model step-wise results

Step variable entered	Improvement chi square	p value	Global chi square	p value
1 Age	234.732	0.0	239.674	0.0
2 Mean blood pressure	52.491	0.0	297.133	0.0
3 Forced expiratory volume	42.026	0.0	337.598	0.0
4 Cigarettes	25.647	0.0	361.970	0.0
5 Xanthelasma	7.674	0.006	373.751	0.0
6 Mother status	9.421	0.002	382.204	0.0
7 Arm circumference	9.379	0.002	392.445	0.0
8 Father status	8.115	0.004	398.911	0.0
9 Shoulder pelvis ratio	8.090	0.004	405.695	0.0
10 Vital capacity	7.207	0.007	410.318	0.0
11 Arcus senilis	5.981	0.014	421.115	0.0
12 Serum cholesterol	5.099	0.024	423.168	0.0

Table 4 Seven Countries Study, Italian rural areas.

Relative risk of death from all causes in 25 years for levels of single risk factors corresponding to the mean plus one standard deviation (SD) and the mean minus one standard deviation, everything else being equal to the mean. The following adjustments have been made in choosing the three levels for some factors:

Cigarettes: 0, 10, 20.

Father and mother status: codes 1, 2, 3 (see appendix 1) Xanthelasma and arcus senilis: absent instead of mean; present instead of plus one standard deviation.

Factor	-1 SD	Mean	+1 SD
Age	0.65	1.00	1.36
Mean blood pressure	0.79	1.00	1.20
Forced expiratory volume	1.14	1.00	0.86
Cigarettes	0.84	1.00	1.16
Xanthelasma	—	1.00	1.57
Mother status	0.89	1.00	1.11
Arm circumference	1.07	1.00	0.93
Father status	0.86	1.00	1.14
Shoulder pelvis ratio	0.91	1.00	1.09
Vital capacity	1.07	1.00	0.93
Arcus senilis	—	1.00	1.23
Serum cholesterol	0.93	1.00	1.07

smoking 20 instead of 10 cigarettes is associated with an excess mortality of 16%.

Having a serum cholesterol of 40 mg/dl above the mean corresponds to an excess risk of 7%.

Among non-modifiable factors, having a father still living compared with having a father dead from a non-violent and non-infectious cause means a difference in risk of 28%.

In general, one standard deviation of a single factor from the mean (apart from age) means an excess mortality or a protection of 10 to 20%. Combinations of various risk factors for an individual result in large variations in mortality risk.

Multiple logistic analysis allows us to discriminate between cases and non-cases by using absolute probabilities, by applying these coefficients to each individual, and locating the observed events in decile classes of the estimated risk. Table 5 reports such a computation which suggests that 131 out of 670 deaths are located in the upper decile (19.6%), 246 are located in the upper quintile (36.8%), while only 19 (2.8%) are located in the lowest decile and 48 in the lowest quintile (7.2%). As a consequence, the ratio of cases observed in the two extreme deciles is 7.84 and in the two extreme quintiles 5.12 (relative risk).

The large number of events favours a very high specificity and sensitivity of the prediction. Taking as a reference decile 10 (150 exposed to risk with 131 events) and decile 1 (149 exposed to risk with 17 events) the sensitivity of prediction is 89% and the specificity 87%. A comparison of the top with the bottom quintiles shows their levels to be 84% and 83% respectively. When the distribution is broken down into two halves, the sensitivity is 72% and the specificity is 68%.

Appendix 3 reports the levels of cumulative survival (Cox model) for each of the 25 years of observation related to individual carriers of the mean levels of the covariates (risk factors) given in Appendix 2. The above two tables and the set of coefficients enable the computation of values of risk for given levels of individual factors.

The two multivariate models employed in this analysis have given similar results, although the use of the Cox proportional hazards model has to be preferred when the follow-up, as on this occasion, is particularly long.<sup>17,18</sup> Most of the *t* values of the coefficients are larger in the Cox than in the multiple logistic model, reflecting the adjustment of the denominator operated by the former in each calendar year of follow-up. However, the advantage of the Cox

model, when the end-point is total mortality, is known to be less than in the case when a single disease is the end-point. In such a case, the multiple logistic equation compels one either to attribute to other causes of death the role of non-cases, or to exclude such cases from the denominator. In both instances the reasoning is substantially distorted, although it might be instructive for the question of competing risks.

### Comment

The attempt to identify, among a number of personal characteristics, those related to total mortality—or survivorship—has resulted in a relatively limited set of factors of differing nature.

A demographic characteristic (age), one circulatory parameter (blood pressure), two respiratory function measurements (forced expiratory volume and vital capacity), two family characteristics (mother and father life-death history), one behavioural characteristic (cigarettes smoked per day), one metabolic index (serum cholesterol), two clinical observational characteristics (xanthelasma and arcus senilis), and two musculoskeletal characteristics (shoulder-pelvis ratio and arm circumference) are the 12 factors most correlated, with different signs, to survival or death. From a biological point of view they represent a variety of indicators, ranging from inherited tendency to longevity, to behavioural characteristics such as smoking habits as the extremes. Longevity of parents, with a very simple consideration of the nature of the eventual cause of death, is clearly reflected in the longevity of offspring.

Blood pressure has been shown to be the single most important factor which, in previous analyses, was related to a number of fatal end-points such as coronary heart disease, stroke, cancer, violent death, and other causes of death.<sup>14</sup>

The meaning of the two respiratory measurements is likely to go beyond their relation to specific respiratory disease, which is rare, and they possibly represent more general physiological indicators. Serum cholesterol, xanthelasma, and arcus senilis are likely expressions of similar metabolic disorders of fats although elicited in different ways.

Arm circumference, when separated from skinfold thickness, is an indirect measurement of muscular mass which, in turn, could reflect the physical fitness of the individual.

The shoulder-pelvis ratio is a simple bone measurement index which suggests the existence of a more masculine (relatively high ratios) or a more feminine (relatively low ratios) shape. What the possible connections are between this characteristic and mortality is unknown.

Table 5 *Seven Countries Study. Italian rural areas. Distribution of observed cases of death in decile classes of risk as estimated by the multiple logistic function*

Decile class	Observed cases	% on all	% in class
1	17	2.5	11.4
2	31	4.6	20.7
3	29	4.3	19.5
4	57	8.5	38.0
5	53	7.9	35.6
6	61	9.1	40.7
7	87	13.0	58.4
8	89	13.3	59.3
9	115	17.2	77.2
10	131	19.6	87.3
All	670	100.0	—

% cases decile 10/1 = 7.84

% cases decile 9 + 10/1 + 2 = 5.12

Finally, cigarette smoking is another powerful multipotential risk factor related to coronary heart disease, stroke, and cancer.

In spite of slightly differing rank positioning of the 12 factors in the two different models, they merit priority over the rest.

The dominant predictive power of the 12 selected factors is clearly also related to some circumstantial factors which need to be remembered. Predicting factors necessarily belong to the set available at the beginning of this study and they do not necessarily represent the best absolutely, but only the best among those measured.

Other factors, not measured here, might be more predictive. Moreover, the original list of factors was conditioned by the orientation of the study, particularly toward cardiovascular diseases and coronary heart diseases. In addition, many possible measurements of interest today were not standardised and suitable for epidemiological field operations 25 years ago. Even so, it is of interest that the possible predictive power of some cardiocirculatory measurements, such as heart rate, minor ECG abnormalities, and ECG abnormalities after exercise testing, is overwhelmed by the presence of other factors.

Again, none of the morbid conditions accepted as possible predictors of death has shown a significant predicting power once individuals with initial life-threatening diseases have been excluded from the analysis. Incidentally, such conditions have shown (table 1) a very poor prognosis in the next 25 years with a fatality rate of 80%.

The most common causes of death described in table 1 are also the major killers in industrialised societies. The predicting power of some of the significant factors has already been identified, also in this study, for shorter follow-up periods, in regard to coronary heart disease (cholesterol, blood pressure, cigarette smoking, respiratory indices), to strokes (blood pressure) and to cancer (smoking) which together accounted for 68% of the total mortality.<sup>7 8 14</sup> The other significant factors seem, therefore, to carry a more specific effect on other causes and/or on the major ones.

The results of a similar analysis, using slightly different procedures, on the 20 year follow-up mortality data<sup>9</sup> revealed, as significant factors for total mortality, the same characteristics. The only difference was that mother and father life-status were, on that occasion, aggregated in the same score, whereas coronary heart disease at entry, retained as possible factors ended up as significant after shoulder-pelvis shape was excluded. This means that, by adding 5 years to the follow-up, no new factors have emerged as predictors of all causes of death, even though 41

characteristics were initially considered instead of 33.

We are therefore more and more confident that, starting from the available set of data, the 12 selected factors should represent the basic indicators and possibly determinants of future mortality and, conversely, longevity and life expectancy.

It can be argued that this approach does not take into account the possible relation between a very powerful factor and a very rare condition. On the other hand, since mortality and life expectancy are bound, on average, to the most common fatal diseases, the relevance of such a particular relationship, if any, can be neglected in this more general approach. This approach represents an attempt to summarise the contribution to mortality and life expectancy of those factors mostly related to the major health-disease problems.

The 12 major factors belong to very different biological categories. Some of them are not modifiable, such as age, parents' status, and skeletal measurements, but others can be changed and/or prevented, by appropriate life-style changes and/or drug use. It is clear from table 4 that relatively small differences in the levels of some factors capable of change can exert an appreciable difference in risk. This is particularly true for blood pressure, smoking habits, and respiratory function indices. Blood pressure seems to be the major factor, among those measured, conditioning life expectancy. From this point of view, the use of mean blood pressure is not of great practical use. However, a mean blood pressure of 104 mmHg corresponds to an average of 140 and 87 mmHg systolic and diastolic respectively, whereas by adding one standard deviation the corresponding levels are about 160 and 97: approximately the conventional cut-off limits of hypertension. This moderate increase in blood pressure carries an excess risk of death of 20%.

Measurement of the 12 selected factors takes 15 to 20 minutes plus the time needed for serum cholesterol measurement. If the reversibility of risk holds, prevention is possible. A prediction of life expectancy through the use of such measurements, at both individual and population level, might be useful: in the former case, for predicting future individual health problems and orienting preventive action; in the latter, for identifying future health problems in the community and improving health services and planning preventive programmes.

## References

- <sup>1</sup> Dyer AR, Stamler J, Berkson DM, Lindberg HA. Relationship of relative weight and body mass index to 14 year mortality in the Chicago People Gas Company Study. *J Chron Dis* 1975; **28**: 109-23.

- <sup>2</sup> Keys A, Taylor HL, Blackburn H, Brozek J, Anderson JT, Simonson E. Mortality and coronary heart disease among men studied for twenty-three years. *Arch Intern Med* 1971; **128**: 201-14.
- <sup>3</sup> Raynor WJ, Shekelle RB, Rossof AH, Maliza C, Paul O. High blood pressure and 17 year mortality in the Western Electric Health Study. *Am J Epidemiol* 1981; **113**: 371-7
- <sup>4</sup> Shurtleff D. Some characteristics related to the incidence of cardiovascular disease and death: Framingham Study, 18 year follow-up. In: *The Framingham Study. An Epidemiological Investigation of Cardiovascular Disease*. Kannel WB, Gordon T, Eds. Section 30 DHEW Publication (NHI) No 74-599, 1974.
- <sup>5</sup> Keys A, Aravanis C, Blackburn H, et al. *Seven countries. A multivariate analysis of death and coronary heart disease*. Cambridge, Mass and London, England: Harvard University Press, 1980; 1-381.
- <sup>6</sup> Keys A, Menotti A, Aravanis C, et al. The Seven Countries: 2289 deaths in 15 years. *Prev Med* 1984; **13**: 141-54.
- <sup>7</sup> Menotti A, Conti S, Giampaoli S, Mariotti S, Signoretti P. Coronary risk factors predicting coronary and other causes of death in fifteen years. *Acta Cardiol* 1980; **35**: 107-20.
- <sup>8</sup> Menotti A, Conti S, Giampaoli S, Marriotti S, Rumi A, Signoretti P. Prediction of coronary and other causes of death by some coronary risk factors. Rapporto ISTISAN 1980/14. Istituto Superiore di Sanità, Roma 1980: 1-19.
- <sup>9</sup> Menotti A, Conti S, Dima F, Giampaoli S, Giuli B, Rumi A, Seccareccia F, Signoretti P. Prediction of all causes of death as a function of some factors commonly measured in cardiovascular population surveys. *Prev Med* 1983; **12**: 318-25.
- <sup>10</sup> Menotti A. Coronary risk factors and mortality for coronary and non-coronary events. In: Lenzi S, Descovich GC, Eds. *Atherosclerosis, clinical evaluation and therapy*. Lancaster, Boston, The Hague: MTP Press, 1982; 343-51
- <sup>11</sup> Woodbury MA, Manton KG, Stallard E. Longitudinal analysis of the dynamics and risk of coronary heart disease in the Framingham Study. *Biometrics* 1979; **35**: 575-85.
- <sup>12</sup> Farchi G, Capocaccia R, Verdecchia A, Menotti A, Keys A. Risk factors changes and coronary heart disease in an observational study. *Int J Epidemiol* 1981; **10**: 31-40.
- <sup>13</sup> Hofman A, Feinleib M, Garrison M, Garrison R, Van Laar A. Does change in blood pressure predict heart disease? *Br Med J* 1983; **287**: 267-9, 1-9.
- <sup>14</sup> Conti S, Farchi G, Menotti A. Coronary risk factors and excess mortality from all causes and specific causes. *Int J Epidemiol* 1983; **12**: 301-8.
- <sup>15</sup> Cox DR. Regression models and life tables. *J Roy Stat Soc* 1972; **B34**: 187-200.
- <sup>16</sup> Walker S, Duncan DB. Estimation of the probability of an event as a function of several independent variables. *Biometrika* 1967; **54**: 167-79.
- <sup>17</sup> Woodbury MA, Manton KG, Stallard E. Longitudinal models for chronic disease risk: an evaluation of logistic multiple regression and alternatives. *Int J Epidemiol* 1981; **10**: 187-97.
- <sup>18</sup> Green MS, Symons MJ. A comparison of the logistic risk function and the proportional hazards model in prospective epidemiologic studies. *J Chron Dis* 1983; **36**: 715-24.

## Appendix 1

Units of measurement of the risk factors and other characteristics:

- Age—in years, rounded off at the nearest birthday.
- Marriage status—code: 0 = presently married; 1 = other
- Number of children—number.
- Mother status—code: 1 = mother alive or dead of violent death; 2 = mother dead of infectious disease any time, or dead of other cause after the age of 65; 3 = mother dead of other cause before the age of 65.
- Father status—code, as for mother status.
- Family history of myocardial infarction—code: 0 = absent; 1 = present (clinical history).
- Family history of other heart disease—code: 0 = absent; 1 = present (clinical history).
- Family history of hypertension—code: 0 = absent; 1 = present (clinical history).
- Family history of stroke—code: 0 = absent; 1 = present (clinical history).
- Family history of diabetes—code: 0 = absent; 1 = present (clinical history).
- Physical activity—as judged by a questionnaire and mainly in relation to physical activity at work: 1 = sedentary; 2 = moderately active; 3 = very active.
- Cigarettes—number per day.
- Diet—use of special diet suggested by personal doctor for health reasons: 0 = absent; 1 = present.
- Body mass index—weight in kg/(height in m)<sup>2</sup>.
- Relative body weight—observed weight in kg/expected weight in kg × 100 (Metropolitan Life Insurance Tables).
- Skinfold thickness—sum of two skinfolds in mm (right triceps and subscapular areas).
- Laterality-linearity index—(biacromial + bicrestal diameters in cm)/(height in cm) × 100.
- Shoulder/pelvis ratio—biacromial diameter in cm/bicrestal diameter in cm.
- Trunk/height ratio—sitting height in cm/standing height in cm × 100.
- Arm circumference—in mm (right arm) at mid-point between acromion and olecranon. Circumference cleaned from skinfold thickness of tricipital site = circumference—(skinfold thickness × 3.14).
- Serum cholesterol—mg/dl (method of Abell-Kendall as modified by Anderson and Keys) transformed into mmol/l.
- Urine protein—code: 0 = absent; 1 = present (semiquantitative).

- Urine glucose—code: 0 = absent; 1 = present (semiquantitative).
- Vital capacity—in dl/height in m.
- Forced expiratory volume—in 0.75 s in dl/height in m.
- ECG aspecific abnormality—Minnesota codes 1.2 or 1.3 or 5.1 or 5.2 or 6.1 or 6.2 or 7.1 or 7.2 or 7.4 or 8.3. 0 = absent; 1 = present.
- ECG exercise test—Minnesota codes 11.1-2 or 12.1-2 or 13.1-2 or 14.1-2 or 15.1. 0 = absent; 1 = present.
- Blood pressure—in mmHg. Systolic and diastolic (V phase of Korotkoff sounds) as mean of two consecutive measurements in supine position.  
Mean blood pressure—diastolic + (systolic-diastolic)/3.
- Heart rate—from resting ECG. Beats in 60 s.
- Xanthelasma—code: 0 = absent; 1 = present.
- Arcus senilis (or gerontoxon)—code: 0 = absent; 1 = present.
- Baldness—code: 1 to 4, according to increasing degree of baldness.
- Heart disease—any Seven Countries prevalence code from 1.00 to 10 (reference n.5: 0 = absent; 1 = present).
- Cerebrovascular disease—code: 0 = absent; 1 = present (documented clinical history and physical examination).
- Cancer—code: 0 = absent; 1 = present (documented clinical history and physical examination).
- Chronic bronchitis—code: 0 = absent; 1 = present (clinical history and/or findings from physical examination).
- Diabetes—history of overt diabetes treated with drugs or casual urine sugar: 0 = absent; 1 = present.
- Kidney disease—code: 0 = absent; 1 = present (documented clinical history).
- Peptic ulcer—code: 0 = absent; 1 = present (documented clinical history).
- Gall bladder disease—code: 0 = absent; 1 = present (documented clinical history).

### Appendix 2

Coefficients for all the 29 selected factors when forced into the proportional hazards model:

Age = 0.887 ( $t = 9.82$ ); mean blood pressure = 0.0213 ( $t = 6.83$ ); forced expiratory volume = -0.0439 ( $t = -2.68$ ); cigarettes = 0.0169 ( $t = 3.99$ ); xanthelasma = 0.7915 ( $t = 3.11$ ); mother

status = 0.1548 ( $t = 2.75$ ); arm circumference = -0.0078 ( $t = -3.08$ ); father status = 0.1799 ( $t = 3.06$ ); shoulder pelvis ratio = 0.0159 ( $t = 2.73$ ); vital capacity - 0.0354 ( $t = -2.11$ ); arcus senilis = 0.2123 ( $t = 2.11$ ); serum cholesterol = 0.0849 ( $t = 2.44$ ); skinfold thickness = -0.0175 ( $t = -2.60$ ); body mass index = 0.0377 ( $t = 1.59$ ); peptic ulcer = 0.1715 ( $t = 1.36$ ); no. of children = 0.0347 ( $t = 1.33$ ); diabetes = 0.2921 ( $t = 1.28$ ); chronic bronchitis = 0.1136 ( $t = 1.27$ ); physical activity = -0.0770 ( $t = -1.23$ ); heart rate = 0.0040 ( $t = 1.22$ ); marriage status = 0.1202 ( $t = 0.85$ ); urine protein = 0.2979 ( $t = 0.82$ ); ECG minor abnormality = 0.0922 ( $t = 0.62$ ); ECG stress test = -0.0631 ( $t = -0.34$ ); kidney disease = -0.0556 ( $t = -0.33$ ); diet = -0.0144 ( $t = -0.12$ ); laterality-linearity index = 0.0028 ( $t = 0.12$ ); baldness = 0.0040 ( $t = 0.10$ ); trunk/height ratio = 0.0002 ( $t = 0.07$ ).

### Appendix 3

Cumulative survival probabilities of all causes of death, year by year, in 25 years, as derived by the Cox model with 29 factors

Year	Yearly events	Cumulative events	Cumulative survival
1	7	7	0.9965
2	15	22	0.9902
3	15	37	0.9834
4	13	50	0.9773
5	17	67	0.9694
6	16	83	0.9617
7	16	99	0.9539
8	14	113	0.9470
9	22	135	0.9357
10	28	163	0.9208
11	19	182	0.9105
12	35	217	0.8910
13	33	250	0.8722
14	33	283	0.8530
15	28	311	0.8363
16	29	340	0.8184
17	34	374	0.7969
18	40	414	0.7710
19	43	457	0.7423
20	28	485	0.7232
21	35	520	0.6986
22	43	563	0.6677
23	36	599	0.6403
24	45	644	0.6069
25	26	670	0.5866