While this report adds little to the authors' previously published results, the high quality of the original data and the devotion of the investigators in extending their mortality followup deserves some acknowledgement.

Regrettably, there are some aspects of the analysis and discussion that are not acceptable in their present form. Since they are not essential to the points being made, it should be easy to modify the paper to deal with these problems.

It is not entirely clear why all men with cardiovascular disease are excluded from the population base. If the endpoint is coronary heart disease (CHD) death, it is perhaps reasonable to exclude persons with CHD at entry from the population at risk--although it is certainly arguable whether it is necessary to do so. But excluding men with other cardiovascular diseases is less easily defensible. We need more justification of these exclusions, as well as well as some idea as to what the distribution of cardiovascular causes in this population is and at least some gesture at indicating the effect of these exclusions on the conclusions (probably very little). And if, these same exclusions were used in analysing the all-causes mortality we need a fairly strong defence.

Similarly, the exclusion of men who subsequently died of non-CHD causes requires more justification than it is given. Presumably, this is intended as a cheap way of allowing for competing risks; but is it necessary? Does it alter the results? If so, in what way? Obviously this exclusion was not used in the analysis of all-cause mortality, but the methods are so loosely stated that it sounds as if the exclusion was used throughout the analysis.

There is a technical point about the analysis which is repeatedly misstated. The logistic regression is not a discriminant analysis and words like "discriminating" (used, e.g., in table 6 and on pages 7, 11 and 12) are inappropriate.

A major analytical difficulty lies in the use of correlation and regression coefficients to summarize inter-area data. Nothing is said in the methods section as to how these were calculated or how their standard errors were computed. To do this correctly is technically possible but exceedingly difficult. I presume that since nothing whatsoever was said about this subject in the methods section that what the authors did was to treat the data as if they were based on a simple random sample from a bivariate normal distribution. The interarea data of rates and associated mean values bears no resemblance to such a distribution and it is incorrect to assume to they do. However, I know what they are trying to get at. I think they could make their point by a table which gave age-standardized CHD death rates and mean values for the variables of interest. If the areas were listed in order of ascending CHD death rates, the argument which they are trying to make would be reasonably clear and no one could quarrel with the methodology.

The statement in the summary that "differences among the cohorts in average systolic blood pressure and serum cholesterol accounted for two-thirds of the variance in coronary death rates" cannot be allowed to stand. Not only is the statement not based on concrete statistics presented and discussed in the text, but it is presumably based on a multiple correlation statistic similar to those presented in table 2--which is difficult to define in the first place but which cannot be used in this casual fashion in the second place. Furthermore, while "accounted for" is a reasonable way to verbalize the statistical process which is

being used, the authors really should indicate that there may be some distinction between a regression "explanation" and other kinds of explanations which the non-statistician would attach to the phrase "accounted for".

The other procedure they use, of calculating expected rates from the logistic regressions and comparing them with the actual rates (table 5) is more easily defensible but if the ~~intercepts~~ constants were included in table 3 the argument would be more easily followed.

I would like to see more use made of the data in table 3. It is possible to test the coefficients in the 4 areas to see if they appear to be homogeneous and if they are to calculate a weighted average of the area coefficients. As the authors note, it is the consistency of relationships among areas that increases our assurance that the observed associations are not chance phenomena. The Japanese data constitute the main exception to the general findings, but given the large sampling error in the calculated coefficients for that group, the argument for homogeneity may still stand up to analysis.

Table 4 is certainly interesting but the use of a correlation coefficient is inappropriate and without some explanation, the use of the first order moment adds nothing to our understanding (Since I have not examined the cited authority for this statistic I cannot judge whether it would be appropriate with explanation. In any event, I can't see the need. Clearly, there is a strong regression present: the likelihood ratio statistic (where are the degrees of freedom cited on that table?) tells us that.)

Table 6 should have cited the 10-year and 15-year coefficients, not the t-values. To say that the "predictive power...did not diminish" from 10 years to 15 years would require the actual coefficients be used. By the way, where are the standardized coefficients, cited in the text?

The statement that "body weight...showed no relevance for coronary death" may or may not be true: we are not given the univariate analysis to judge for ourselves. Perhaps, not to open the door to a long discussion, it might be sufficient to add the phrase..."when the other factors are taken into account". Presumably that means that if added weight did not tend to lead to increases in blood pressure and serum cholesterol levels it would have no influence on coronary mortality, but since it does tend to lead to increases in blood pressure and serum cholesterol levels, the multivariate analysis may not be entirely germane to the question.

Finally, the discussion section is very strange, being limited to a consideration of the aberrant data from Japan. I hold no brief for a formal discussion section and since what discussion is offered on other topics is included in the Results section perhaps the Japanese data could be discussed there as well. It is a little sad that the very extensive literature on area differences in coronary mortality is not even alluded to, as the skimpy bibliography makes all too clear, but that is a task in itself and maybe the authors are justified in their restraint.

REVIEWER'S COMMENTS (2) PREVENTIVE MEDICINE

Author: Ancel Keys et al.

Title: The Seven Countries Study: 2,289 Deaths in 15 Years

The Seven-Country Study represents a cornerstone of epidemiological knowledge on coronary heart disease. This 15-year follow-up is another milestone. It reaffirms, on the basis of longer experience, previous findings and corrects an earlier statement which has caused much trouble, that smoking is not a risk factor in some parts of the world. It adds an important new finding: that the significance of risk factors does not diminish with age, as is claimed repeatedly.

There is an apparent contradiction which might be clarified with a word or two. On page 7, serum cholesterol is called the most important risk factor for the coronary death rate. On page 17, systolic blood pressure is stated to be the most predictive variable. Actually, the latter statement is based on an analysis by regions, the first on the total sample, apart from differences in the statistical methods used.

The section, starting on page 2, on cohorts, subjects, methods and follow-up is useful, as the authors state. However, it is rather long and might be set in small print, starting with the second paragraph on page 2.

May we send comments over your signature? Yes ☒ No ☐

Signature: F.H. Epstein Date: 7.6.1983

Frederick H. Epstein
University of Zurich

REVIEWER'S COMMENTS (3) PREVENTIVE MEDICINE

Author: Ancel Keys et al.

Title: The Seven Counties Study: 2,289 Deaths in 15 Years

This paper extends the follow-up of 15 of the 16 original cohorts in the Seven Counties study for a total of 15 years of mortality analysis. The main purpose of the report is the quantification of the relationship between entry levels of the major risk factors and subsequent mortality, both for deaths attributed to coronary disease and all causes. Analyses are performed in two modes--one, regression analysis of aggregated data with each cohort as the unit of study; and two, the entry characteristics of individuals as predictors of subsequent mortality by multiple logistics analysis for cohorts within regions. As summarized by the investigators, the major findings were: 1. differences among the cohorts in average systolic blood pressure and serum cholesterol accounted for two-thirds of the variance in coronary death rates across the populations; 2. the risk of coronary deaths for individuals in the U.S., and Northern and Southern Europe was directly related to age, systolic blood pressure, serum cholesterol and cigarette use. Coronary deaths among the Japanese men were too few for valid conclusions. Comparisons of the 15-year with the 10-year experience of the cohorts indicated a greater significance of the risk factors at the longer period of follow-up.

The findings are of extreme interest and importance for all students of cardiovascular disease epidemiology. They have implications not just for those interested in etiologic-oriented research, but also for the wide audience of interventionists currently carrying out or planning community intervention investigations, demonstrations and health care activities. Given the importance and potential of this information, I found several elements in the research methodology, interpretation, and reporting of the results potentially misleading. This stems primarily from the use of the logistic risk function for the analysis of risk of disease over a 15-year period. During this time, members of the original cohorts are at markedly varying time periods

of risk and the only adjustments entered into analysis were the exclusions of deaths due to non-coronary disease from the logistic analysis of coronary heart disease mortality. This resulted in 1,671 exclusions from the total of 11,579 men free of coronary heart disease at entry--a sizeable 14%. More important, the basic comparison made in the logistic analysis was a discriminant type analysis contrasting coronary decedents during the 15 years with survivors. This, in itself, may be useful information for some purposes; however, it does not permit one to make statements regarding risk of disease in relation to entry characteristics. Further, it could be misinterpreted in its applicability if someone were to extrapolate it to a new cohort on which one had just baseline information since, of course, it is impossible, a priori, to predict who would or would not subsequently die of causes other than coronary diseases. I can appreciate the desire of the investigators to use the logistic risk function since it has had such a long history of application in coronary disease. However, as applied, it is incorrect to interpret its findings as measures of risk. A more appropriate method, for this purpose, is to use some type of multivariable survival analysis making adjustments for the variable time periods of risk and removals at varying periods of time of individuals who died of causes other than coronary disease. It is becoming increasingly common to compute hazard rates using multivariable proportional hazards analyses for this purpose. The two approaches (the logistic and proportional hazards) produce generally similar results for short periods of follow-up and low incidence of events. The consequences of 15 years follow up and marked variation in coronary death rates among the cohorts in this study are unpredictable but worthy of investigation.

An additional area of concern is one that relates to incomplete analysis of the data, i.e., the association, if any, between the risk factors and non-coronary deaths. This relates to the first point made above since the investigators stated explicitly that they deleted the non-coronary deaths from the analyses in order not to dilute the measures of effect of the risk

factors on coronary mortality. While the goal is commendable, there may be very important and informative data omitted in the process. For example, the regression analysis of mean characteristics and coronary death rate among the cohorts indicated that the most important variable was the average concentration of serum cholesterol ($r = 0.87$); however, the statement is made (p. 7) that only systolic blood pressure proved to be significantly related to total mortality rates. This suggests the possibility of a negative correlation between cholesterol values at entry and non-coronary mortality. Similarily, and by extension, it would be interesting to see the logistic equation coefficients derived for individuals for all cause mortality. This is not presented directly, although it must have been calculated since the equations were used in calculating observed over the expected all cause mortality across regions as presented in Table 5.

My second criticism, i.e., investigation of the relationship between cholesterol and non-coronary mortality, represents an extension beyond the major objective of the paper and may be something the investigators may wish to defer to a separate and independent publication. I believe that at least some reference to it should be made in the text because of the interest in this problem around the world at present, and the fact that this group has the opportunity to make a major contribution to that problem. However, I believe the first criticism addressed above regarding the interpretation of data given the research design and possibly misuse of the concept of probability and risk is one that deserves more attention.

May we send comments over your signature? Yes ☑ No ☐

Signature: Herman A. Tyroler Date: 29 Jul 83

Herman A. Tyroler
School of Public Health, Univ. Of North Carolina