1. THE PROBLEM

IAN T. T. HIGGINS

Ten years ago when the Second World Congress of Cardiology was held in Washington, the first morning was devoted to cardiovascular epidemiology, and the papers from that session were subsequently published in *World Trends in Cardiology I,* edited by Ancel Keys and Paul D. White.[1] Many of the observations on coronary disease — international and racial differences, short-term secular trends in occupied countries during World War II, differences in disease frequency related to such factors as diet, serum cholesterol, and physical activity — which have intrigued epidemiologists, were presented at that conference. The data on which these observations were based varied considerably in quality; the need for better comparability, particularly for international studies, has often been stressed.[2, 3, 4] For example, the final paragraph of the Princeton report's section on Design and Analysis of Studies[3] calls for agreement on a specified minimum set of measurements to be carried out under standard specifications and presented in the published report in a uniform format.

This section is concerned with comparability in establishing disease frequency. One chapter only is concerned with measuring an associated factor. The chapters illustrate some of the difficulties that arise and the measures that have been, or are being, taken to overcome them.

Mortality

Mortality statistics have provided most of the initial clues for hypotheses about the etiology of arteriosclerotic heart disease. But possible diagnostic errors and variation in classification and coding have always raised doubts about the reality and magnitude of the differences in frequency of the disease these statistics suggest. The obvious course is to check the diagnosis by autopsy. Postmortem study of persons apparently free from this disease whose deaths have resulted from other causes is, moreover, the only way to obtain information about the origin and development of atherosclerosis in man before it leads to complications which cause clinical manifestations. Unfortunately the increased accuracy of diagnosis resulting from autopsy is obtained at the price of an increased sample bias. This can be reduced by studying places where a very high percentage of deaths come to autopsy. Another possibilty is to make postmortem examinations of victims of fatal accidents.

Problems related to autopsies are considered in the paper by Strong and Eggen. Dr. Strong is Director of the International Atherosclerosis Project (IAP) which is a collaborative study involving 19 laboratories in the frequency and severity of atherosclerosis in populations from different parts of the world.

Morbidity

Turning now to morbidity, some information on disease frequency can be obtained from national statistics, from hospital or doctors' records, Ministry of Pensions and National Insurance, and other insurance data, sickness absence records, and so on. But the frequency of diagnosed disease will vary with standards of medical care, attitude to and awareness of the community of the disease—factors which are likely to differ in different countries. Overt disease is, therefore, an unsatisfactory index for international comparisons. The alternative is a survey screening of the community for all cases of disease.

Prevalence or Incidence?

A single survey of a population will enable us to establish prevalence. If we wish to measure incidence we must carry out at least one follow-up survey of the study population. This leads to difficulties due to emigration from the area, which are particularly frustrating in areas of high mobility. Deaths and refusal to participate are other problems. The difficulties attendant on longitudinal observations of any sample of the general population are unfortunate in that the development of disease or its incidence may be easier to detect than its prevalence, for example, on the basis of serial electrocardiograms (ECGs). Longitudinal studies also enable us to use death, including sudden death adequately defined, as a relatively simple index. Our main concern in this section is with studies designed to establish prevalence.

The measures available to screen a community for heart disease prevalence are:

1. Questions about cardiac pain, particularly angina and myocardial infarction;
2. ECGs either at rest or after exertion;
3. Clinical examination of the heart to exclude other causes of heart pain;
4. X-rays of the chest for heart size, pulmonary disease, etc.

The last two methods have not proved as useful as the first two in epidemiological studies, and I shall ignore them.

My own experience in screening a population for cardiac pain has not been a very happy one. Some years ago I questioned a random sample of men in the Rhondda Fach about chest pain, and my attempts to classify this as cardiac or noncardiac chest pain are shown in Table 1.

You will see that chest pain was recorded in a high proportion of the sample, particularly the miners and ex-miners. In only a few of these was I confident that the pain complained of was cardiac in origin. In an appreciable number I was in doubt and in the rest considered the pain to be noncardiac. Chest pain in miners is perhaps exceptionally hard to classify; nevertheless, I found the results discouraging. The questionnaire that I used was relatively unstand-

TABLE I. PREVALENCE OF PAIN IN THE CHEST (RANDOM SAMPLE OF MINERS AND NON-MINERS IN RHONDDA FACH)

Age Group (yrs.)	*Type of Pain*	*Non-miners*		*Miners and Ex-miners*	
		No.	*Per cent*	*No.*	*Per cent*
35–54	Probably cardiac	2	1.1	2	1.1
	Possibly cardiac	5	2.8	4	2.2
	Not cardiac	23	13.1	58	32.0
	All chest pain	30	17.0	64	35.3
	No chest pain	146	83.0	117	64.6
	Total in sample	176	100.0	181	99.9
55–64	Probably cardiac	3	3.5	2	2.1
	Possibly cardiac	6	7.0	12	12.7
	Not cardiac	12	14.0	25	26.6
	All cardiac	21	24.5	49	41.4
	No chest pain	65	75.6	55	58.5
	Total in sample	86	100.1	94	99.9

ardized and my interviews were conducted in the manner of the physician in the clinic. Dr. Rose, who has been working for a number of years on a questionnaire for detecting cases of cardiac pain and intermittent claudication, has been more successful than I have in selecting the precise questions that should be asked.

The next factor in diagnosis is the electrocardiogram. We started our studies of coronary disease with confidence in the ECG as a simple, objective, and painless diagnostic tool, which could be rapidly applied to the population. This confidence was rudely shaken when we looked into the the matter of agreement between different ECG readers.

In one study 537 twelve-lead, resting ECGs were classified by four readers (three cardiologists and myself) for coronary heart disease. The results are shown in Figure 1.

Twenty per cent of the tracings were considered compatible with coronary disease by at least one of the readers; two of the readers agreed in 12 per cent, three out of four in 8 per cent, and all four in less than 4 per cent. Observer variation has been confirmed in other studies. In order to differentiate recording error from inter-

pretation error, Blackburn and his colleagues[5] advocated a classification of ECGs that has been widely approved for epidemiological studies of coronary heart disease. Preliminary studies with this classification, the Minnesota Code, suggested that it was fairly reproducible. Furthermore, some observations on the Framingham study group showed that certain items carried an increased risk of an episode of coronary heart disease. Through the kindness of Dr. T. R. Dawber, I was permitted to classify all the ECGs taken at the first examination in Framingham, according to the Minnesota Code. The implications of the items considered most likely to indicate coronary disease (abnormal Q waves, ST depression, flat and in-

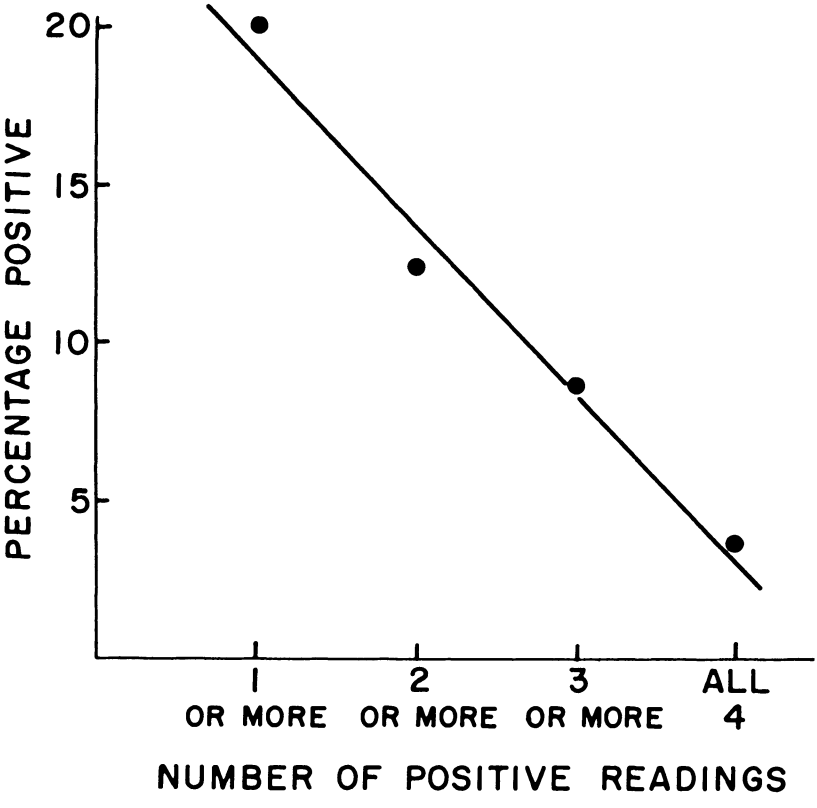

Figure 1. Percentage of ECGs read by four readers as compatible with coronary heart disease.

verted T waves) were assessed by mortality and morbidity attributed to coronary disease during the subsequent eight-year follow-up period. The findings are shown in Figures 2 and 3.

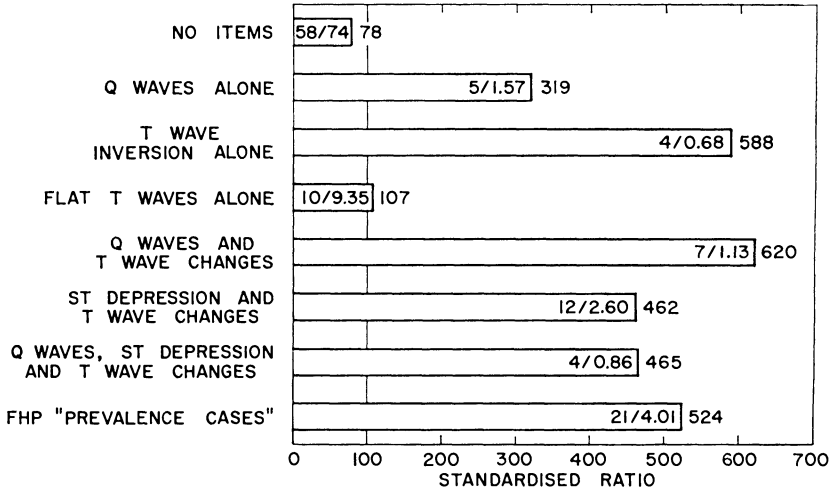

Figure 2. Risk of coronary heart disease death in eight years according to initial ECG (eight-year follow-up, Framingham Heart Project, men and women aged 30–62).

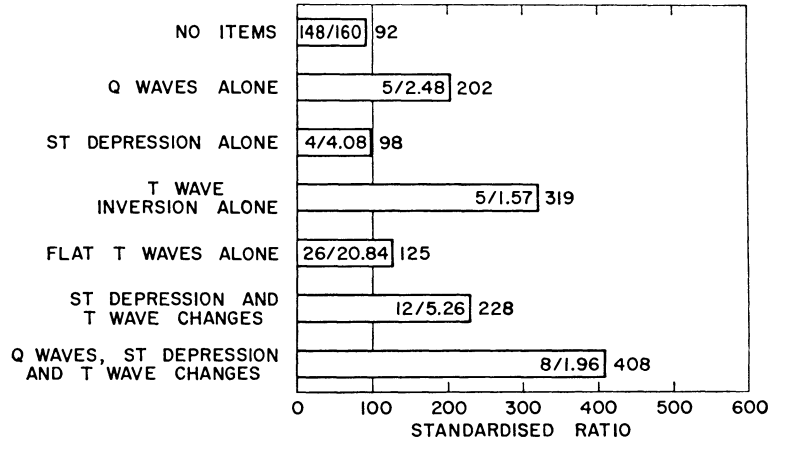

Figure 3. Risk of coronary heart disease incident in eight years according to initial ECG (eight-year follow-up, Framingham Heart Project, men and women aged 30–62).

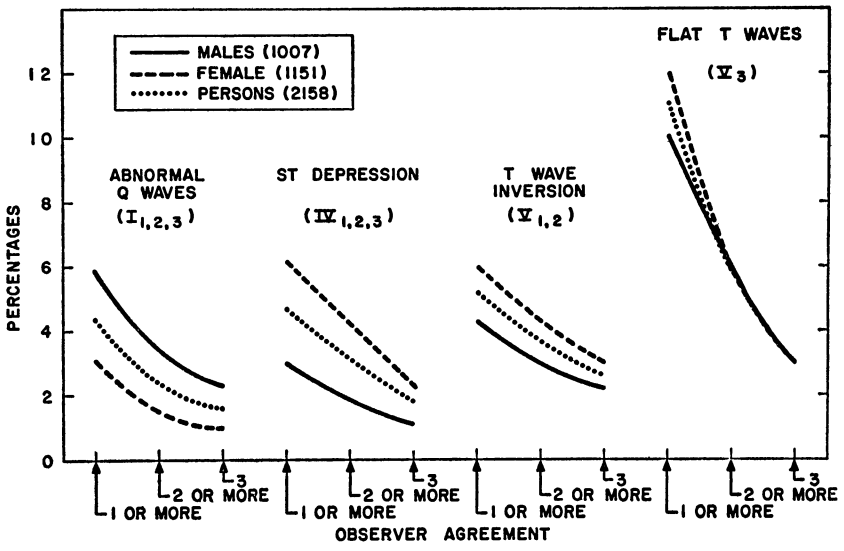

Figure 4. Observer variation in reading ECG items.

Lower than average risks of mortality and morbidity were found in those persons whose ECGs showed no reportable code items. By contrast, abnormal Q waves, T wave inversion, and ST depression when combined with T wave changes appeared to be important risk-carrying abnormalities. Neither flat T waves nor isolated ST depression appeared to carry much increased risk. The relative risk appeared to be similar in the two sexes but the absolute risk was less in women.

Additional reproducibility studies were clearly needed and these have now been carried out. About a year ago, Dr. Henry Blackburn, Dr. Leon Ostrander, and I classified according to the Minnestota Code some 2,200 ECGs from a probability subsample of the National Health Survey. Some idea of our agreement in classifying the relevant items is indicated in Figure 4.

Summarizing the findings roughly: Of those ECGs read by any of the three of us as showing Q wave abnormality, ST depression, or T wave inversion, all three of us agreed on the findings in about half. Of those ECGs read as showing flat T waves we agreed on

only about one-third. This suggests that agreement between observers even when using the code leaves something to be desired. Studies on reproducibility of this code have also been carried out by the World Health Organization, and Dr. Kagan will describe their findings.

COMPARABILITY IN RECORDING ASSOCIATED FACTORS

There are many factors possibly or probably associated with coronary disease that should be standardized. One might mention recording family history, smoking history, occupation, physical activity, and stress, and even such simple measurements as height, weight, skin-fold thickness, and blood pressure. Then there are questions of standardization that arise in the course of the analytic procedures such as age grouping and the grouping of continuous variables. We have no time to consider these in detail here. They have been considered at length in the report of the Princeton Conference[3] and in a number of reports of World Health Organization committees.[6] These papers should be consulted for further information and suggestions for collecting information about the important characteristics associated with coronary disease.

One of the factors that has received most attention has been the serum lipids, particularly the serum cholesterol. There are a number of different methods currently in use for measuring the serum cholesterol. Some of these methods give systematically different results and there is evidence of wide variation in the results from different laboratories. Dr. Cooper describes the reasons that led to the setting up of a lipid standardization laboratory in the Heart Disease Control Unit at the Communicable Disease Center, Atlanta, Georgia. The methods his laboratory employs might perhaps be considered as a model for standardizing other laboratory measurements.

The papers in this section illustrate three of the problems in comparability which arise in epidemiological studies:

1. ECGs, x-rays, and pieces of tissue or complete organs from autopsies may be collected in the field in different surveys and

studied and classified (at leisure) centrally by one or more observers. Provided the source of the material is concealed, valid comparison can be made between different places.

2. Blood samples can be collected and treated in a similar manner. But it is more convenient and practicable to process such samples in different laboratories, whose performance and accuracy are followed by regular surveillance and monitoring with samples of known concentration.

3. Questionnaires may be standardized and tested by repetition on two occasions by use of several observers and by tape-recording interviews. Differences in language, attitudes, and cultural characteristics make it especially hard to ensure comparable results in international studies.

REFERENCES

[1] Keys, Ancel, and White, Paul D. (editors), WORLD TRENDS IN CARDIOLOGY. I. CARDIOVASCULAR EPIDEMIOLOGY, New York, Paul B. Hoeber, 1956.

[2] Conference, Longitudinal Cardiovascular Studies, Brookline, Mass., 1957.

[3] Epidemiology of Cardiovascular Disease Methodology (Report of a conference at Princeton, N. J., April 24–26, 1959), Supplement to *American Journal of Public Health,* Vol. 50, No. 10, 1959.

[4] SCIENTIFIC GROUP ON COMPARABLE METHODOLOGY FOR THE EPIDEMIOLOGICAL STUDY OF HYPERTENSION AND ISCHAEMIC HEART DISEASE, Geneva, World Health Organization, 1962.

[5] Blackburn, H. A., et al., The Electrocardiogram in Population Studies; A Classification System, *Circulation,* 21, 1160–1175, 1960.

[6] Survey of the Prevalence of Ischaemic Heart Disease in Certain European Countries: Report on a Technical Meeting Convened by the Regional Office for Europe of the World Health Organization, Copenhagen, 1963, Annexe IV.