Bull. Org. mond. Santé
Bull. Wld Hlth Org. } 1962, 27, 667-679

Grading Atherosclerosis in Aorta and Coronary Arteries Obtained at Autopsy

WHO Trials of Macroscopic Methods

AUBREY KAGAN [1] & KAZUO UEMURA [2]

Many attempts have been made to grade atherosclerosis as found in autopsy material, but the repeatability of the methods used has seldom been tested. Most of the methods used are rough; and, where comparisons are to be made between the data obtained by different observers on different material, it is essential to know whether the differences found are due to the crudity of the method or in fact represent a real difference in the material studied. This paper describes an attempt to obtain comparable data by presenting specially prepared specimens to 14 pathologists from five laboratories in Europe and the Americas. For the purposes of the study, agreed definitions, techniques and criteria were adopted. Intra-observer, intra-laboratory and inter-laboratory disagreement was measured using both transverse- and longitudinal-section procedures. The longitudinally sectioned specimens were examined unstained and subsequently stained for lipid. The results indicate that the longitudinal-section procedure is likely to be useful in discriminating between groups of specimens, provided that certain procedural rules are observed.

For some time it has been thought desirable and possible to study atherosclerosis in autopsy material in relation to ecological patterns with a view to understanding the etiology of atherosclerosis.

In 1955 the World Health Organization convened a Study Group on Atherosclerosis and Ischaemic Heart Disease which pointed out the need for the standardization of clinical and pathological criteria and terminology. In 1957 a WHO Study Group on the Classification of Atherosclerosis Lesions met to discuss the possibility of establishing acceptable methods of grading these lesions. The reports of these two groups were published in 1957 and 1958 respectively.

Partially using the general recommendations and the methodology described in the second of these reports, a group of pathologists from the American continent commenced a combined study on the epidemiology of atherosclerosis (Inter-American Atherosclerosis Project—PIA).[3] Dr C. Tejada of Guatemala, a member of PIA, was appointed as a consultant to WHO to advise on the possibilities of commencing a similar study in Europe. In October 1960 a group of European pathologists met to discuss the most suitable way in which a combined epidemiological and pathological study could be made in Europe. They agreed that an attempt should be made to estimate atherosclerosis in aorta and coronary arteries obtained from autopsy material on an epidemiological basis.

Although quantitative studies of atherosclerosis have been attempted by numerous investigators using the unaided eye, planimetric procedures, special calipers, measurement of liquid flow, and microscopic, chemical and biochemical and radiological procedures,[4] little has been published on the accu-

[1] Medical Officer, Cardiovascular Diseases, World Health Organization, Geneva, Switzerland.

[2] Chief Statistician, Health Statistical Methodology, World Health Organization, Geneva, Switzerland.

[3] Now called the International Atherosclerosis Project. The most recent Standard Operating Protocol (1962) includes cerebral vessels. Copies of this document are obtainable from the Instituto de Nutrición de Centro América y Panamá, Guatemala.

[4] See, for instance, Yater et al. (1951), Patterson et al. (1956), Gore et al. (1957), Elkeles (1957), Tejada et al. (1958), Holman et al. (1958), Dickinson & Thomson (1959), Baker et al. (1960), Scott et al. (1961).

racy, precision or repeatability of the methods used.

This may be of less importance when one observer grades all the material if large numbers of specimens are studied. Even so it is an advantage to know the precision or repeatability of the observations. When more than one observer takes part in the grading a knowledge of repeatability becomes of considerable importance, for repeatability decreases with the number of observers. Where comparisons are made, and this is nearly always the intention, it is important to know if the differences found are likely to reflect a real difference or whether they are likely to be due to the crudity of the method.

With this in mind it was decided that the definitions and criteria arbitrarily selected and the suggested methods of measuring these factors should be subjected to trial. A preliminary test was carried out and assessed early in 1961 by five pathologists from four centres.[1] It was immediately found that some of the disagreement was due to looseness of interpretation of definitions and some to faulty technique in preparing and presenting the specimens. In the light of this experience the definitions, criteria and techniques were modified. These are detailed below and in the Annex and have been used in the trial now to be described. Although there were some differences from the definitions used in the Standard Operating Protocol of the PIA group (1960),[2] many of their definitions and procedures were included.

METHODS

PREPARATION OF SPECIMENS

Specimens were specially prepared. Fifty aortas and 100 sets of coronary arteries were removed at autopsy.[3]

Longitudinal section

The aortas and 50 sets of coronary arteries were opened longitudinally, stripped of adventitial tissue, stretched on cork boards and fixed in formalin. The aorta was removed in three sections corresponding to the ascending aorta plus arch, descending thoracic, and abdominal.[4] Each of these parts was mounted in formalin in a plastic bag and the bag was sealed and numbered. Each of the sets of three coronary arteries—left anterior descending, circumflex and right[4]—was mounted with formalin in a plastic bag. These were numbered and sealed.

Transverse section

Embedded coronary arteries. The other 150 coronary arteries (50 sets) were removed from the heart with surrounding fat.[4] The arteries were gently straightened out and fixed in formalin. Later the surrounding fat was cleared in glycerol and the specimen was attached to a nylon tubing "splint"[5] and embedded in a special preparation of gelatin. When the gel had set it was removed from its mould. The coronary artery could be seen in the gelatin and transverse section was made through the gel and the artery. Care was taken not to cut the nylon tubing "splint". The specimens were examined by bending back the gel and splint and thereby disclosing the cut surface of the artery. After use they were returned to a numbered cellulose case. The specimen was then ready for storage or transport.

Whole hearts. Twelve whole hearts were removed at autopsy and fixed in formol. The coronary arteries were sectioned transversely *in situ*, and the specimens were packed for transport in numbered plastic bags with a little formol.[6]

In order to ensure as wide a range of atherosclerosis as possible, the longitudinally sectioned specimens were quickly sorted into five groups—gross lesions, minimal lesions, just less than gross, just more than minimal, and intermediate lesions. Nine specimens of aorta and three sets of coronary arteries were taken at random by the statistician from each group.

[1] Professor T. Crawford, St George's Hospital, London, England; Professor F. Linell, Allmänna sjukhuset, Malmö, Sweden; Dr N. Sternby, Allmänna sjukhuset, Malmö, Sweden; Professor R. Vaněček, Charles University, Prague, Czechoslovakia; Professor A. M. Vikhert, Institute of Therapy, Academy of Medical Sciences, Moscow, USSR.

[2] Unpublished document of the Instituto de Nutrición de Centro América y Panamá.

[3] These specimens were prepared by Dr N. Sternby, Allmänna sjukhuset, Malmö.

[4] For definitions see Annex.

[5] Technique developed by R. M. C. Arnot.

[6] These specimens were prepared in the Department of Pathology, St George's Hospital, London.

In the trial these 45 specimens of aorta, 15 sets (45 specimens) of longitudinally sectioned coronary artery, 27 specimens of embedded coronary artery and 11 hearts were used.

Stained specimens

After examination in the unstained state the longitudinally sectioned coronary arteries and aortas were sent to PIA (Guatemala) and stained for fat by a standard method. They were then replaced in plastic bags and used for further trials.

METHODS OF EXAMINATION

The specimens, together with grading forms, copies of the definitions and criteria, diagrammatic aids and instructions were sent to the different laboratories. Each specimen had two separate grading forms for each observer and at each laboratory each observer was asked to grade the specimens independently of his colleagues. On the following day the same specimens were graded again by the same pathologists independently of their observation the previous day. Specimens of embedded coronary artery and whole hearts were seen by nine pathologists from four countries. The longitudinally sectioned specimens were examined by 14 pathologists from five laboratories in the unstained state and by 11 of the same pathologists from four laboratories after staining.

Unstained specimens

Transverse section.[1] The total amount of atherosclerosis was estimated visually for each specimen by assessing the proportion of sections in which any form of atherosclerosis was seen (units of: "none", "some to less than $^1/_3$", "$^1/_3$ to $^2/_3$", "more than $^2/_3$"). The degree of narrowing was estimated in the section showing the severest narrowing in each specimen. Reference was made to a standard diagram. The units were: "no narrowing", "some but less than ½ occlusion," "½ or more but less than pinhole", "pinhole", "occlusion by thrombus", "occlusion not by thrombus".

The "numbering of areas of narrowing" per artery was graded as: "none", "single area" (if seen only in one section), "multiple or prolonged" (if seen in more than one section).

Longitudinal section.[1] The "total amount of atherosclerosis" was estimated visually as the proportion of the surface of the specimen affected by any type of atherosclerosis.

The proportion of the atherosclerotic area involved by the different types of atherosclerosis (lipid streak and fatty plaque; fibrous plaque; complicated lesions; calcification) was then estimated in turn.

Grading was in units of "none", "some but less than $^1/_3$", "$^1/_3$ to $^2/_3$", "more than $^2/_3$".

Stained specimens

Longitudinal section. The same specimens after staining were treated as follows.

The coronaries were re-examined by 11 of the same pathologists using the same procedure as described above, except that the definition of "fatty streak" was now:

"Any intimal lesion that is stained distinctly by Sudan IV and that does not show any other type of change underlying it".[2]

ANALYSIS OF RESULTS

The number of disagreements per 100 comparisons has been calculated for comparisons of:

(*a*) first and second grading of each specimen by each observer (intra-observer disagreement);

(*b*) grading of same specimens by observers in the same laboratory (intra-laboratory disagreement);

(*c*) grading of same specimens by observers in different laboratories (inter-laboratory disagreement).[3]

This was calculated for observations made on individual specimens and for observations made on groups of specimens. The former is of interest in

[1] For definitions see Annex.

[2] This is the definition used by PIA in the 1960 Standard Operating Protocol.

[3] The comparisons of grading of each specimen were carried out as follows:

Longitudinally sectioned, unstained specimens, for instance, were examined twice by 14 observers from 5 laboratories—i.e., 3, 4, 3, 2 and 2 observers in each laboratory. In this case there were 14 intra-observer comparisons for each specimen. For intra-laboratory comparisons each observer in each laboratory was compared with each of the other observers in the same laboratory. This gave

$$\left\{\binom{3}{2}+\binom{4}{2}+\binom{3}{2}+\binom{2}{2}+\binom{2}{2}\right\}\times 2=(3+6+3+1+1)\times 2=28$$

intra-laboratory comparisons per specimen. The factor 2 was applied because each specimen was seen twice by each observer. For the inter-laboratory comparisons each observer in each laboratory was compared with each observer in every other laboratory. Thus there were

$$\{3\times(4+3+2+2)+4\times(3+2+2)+3\times(2+2)+2\times 2\}\times 2=154$$

inter-laboratory comparisons per specimen. Some observers, however, missed one or two specimens and therefore the total number of actual comparisons did not agree with the total possible number.

distinguishing the characteristics of individual specimens and the latter in distinguishing the characteristics of groups of specimens. Difference between groups is of most interest to the international study, and generally of most interest to national studies as well—for instance, difference between men and women of the same age-group, differences between men of different age-groups, differences between men of the same age but different occupations, differences between men of the same age from different environments, etc. Disagreement is less for observations on groups. Thus, if two observers grade 100 specimens there may be disagreement in categorizing 20 specimens. But in the 20 individual specimens in which there was disagreement the difference is not always in the same direction; sometimes one observer grades higher than the other and sometimes lower. Thus, for observations on the group of specimens there is less difference in the proportion assigned to each category by the two observers.

RESULTS

TRANSVERSE SECTION

Some of the observers found it difficult to identify anatomical changes in some of the embedded arteries and whole heart specimens. Because of this it was not possible to make observations on the grading of all the specimens as a group.

Table 1 shows the intra-observer, intra-laboratory and inter-laboratory disagreement rates for individual specimens. The disagreement is large. For " amount of atherosclerosis " the intra-observer disagreement (41%) was about twice as great as in the longitudinal section method (Table 2; 22%). The intra-observer disagreement rate in grading " number of areas of narrowing " and " degree of narrowing " was a little less (about 30%).

TABLE 1

PERCENTAGE DISAGREEMENT ON GRADING OF TRANSVERSE SECTIONS OF CORONARY ARTERIES (INDIVIDUAL SPECIMENS)

Method	Amount of atherosclerosis	Number of areas of narrowing	Degree of narrowing
Percentage intra-observer disagreement [a]			
Embedded coronaries	41 (184)	28 (184)	32 (95)
Whole heart	39 (230)	27 (230)	28 (159)
Percentage intra-laboratory disagreement [a]			
Embedded coronaries	46 (263)	40 (263)	37 (123)
Whole heart	34 (328)	30 (328)	28 (225)
Percentage inter-laboratory disagreement [a]			
Embedded coronaries	68 (829)	62 (827)	45 (345)
Whole heart	53 (1 054)	44 (1 052)	47 (658)

[a] The number of comparisons is shown in parentheses.

LONGITUDINAL SECTION, UNSTAINED

Table 2 shows the percentage disagreement in grading individual unstained specimens of aorta and coronary artery in longitudinal section. The upper portion shows the mean intra-observer disagreement as a percentage of comparisons made. It will be seen that the intra observer disagreement rate was of the order of 20% for " total amount of atherosclerosis ", " complicated lesion " and " calcification " in aorta and coronary artery. The centre portion shows the disagreement between observers in the same laboratory; disagreement for these three categories now increases and is about 30%. The bottom part of Table 2 shows disagreement between observers from different laboratories—approximately twice as much as the intra-observer disagreement, at 40%. Disagreement was highest for the categories " lipid streak " and " fibrous plaque ".

Table 3 refers to observations made on the *group* of specimens. The previous tables (1 and 2) refer to observations made on *individual* specimens. In distinguishing the characteristics of individuals, Table 2 is of interest; in distinguishing the characteristics of groups, Table 3 is pertinent. It can be seen that disagreement is clearly reduced in grouped specimens. Intra-observer disagreement is now of the order of 10% in all categories. Intra-laboratory disagreement is about 20% and inter-laboratory disagreement about 25%-30%.

LONGITUDINAL SECTION, STAINED

Eleven of the 14 pathologists who examined the unstained coronaries in longitudinal section examined the same specimens after staining. A comparison of the disagreement rate for intra-observer, intra-laboratory and inter-laboratory observations

TABLE 2

PERCENTAGE DISAGREEMENT ON GRADING [a] OF UNSTAINED LONGITUDINAL SECTIONS OF CORONARY ARTERIES AND AORTAS (INDIVIDUAL SPECIMENS)

Vessel	Number of comparisons	Total amount of atherosclerosis	Complicated lesion	Calcification	Lipid streak	Fibrous plaque
			Percentage intra-observer disagreement			
Aorta	627	22	15	23	33	38
Coronary artery	629	22	18	26	40	35
			Percentage intra-laboratory disagreement			
Aorta	1 252	34	20	27	44	54
Coronary artery	1 257	28	19	36	54	46
			Percentage inter-laboratory disagreement			
Aorta	6 899	43	43	34	50	58
Coronary artery	6 920	35	43	41	66	60

[a] Grading units: " none ", " >1/3 ", " 1/3 to 2/3 ", " >2/3 ".

on the individual specimen is shown in the upper portion of Table 4.

It will be seen that intra-observer disagreement is reduced with the stained specimens in the two categories " lipid streak " and " fibrous plaque ", which showed most disagreement in the unstained specimens. There is little change in the other categories. There is no decrease of inter-laboratory disagreement.

The lower portion of Table 4 compares disagreement in grading *groups* of specimens stained or unstained. There is little difference.

TABLE 3

PERCENTAGE DISAGREEMENT ON GRADING [a] OF UNSTAINED LONGITUDINAL SECTIONS OF CORONARY ARTERIES AND AORTAS (GROUPED SPECIMENS)

Vessel	Number of comparisons	Total amount of atherosclerosis	Complicated lesion	Calcification	Lipid streak	Fibrous plaque
			Percentage intra-observer disagreement			
Aorta	627	7	6	7	12	14
Coronary artery	629	10	9	11	12	11
			Percentage intra-laboratory disagreement			
Aorta	1 252	17	9	14	18	28
Coronary artery	1 257	15	11	20	28	30
			Percentage inter-laboratory disagreement			
Aorta	6 899	25	27	21	27	26
Coronary artery	6 920	18	34	23	33	31

[a] Grading units: " none ", " <1/3 ", " 1/3 to 2/3 ", " >2/3 ".

TABLE 4
PERCENTAGE DISAGREEMENT ON GRADING OF STAINED AND UNSTAINED [a] CORONARY ARTERIES (LONGITUDINAL SECTION)

Type of comparison	Number of comparisons		Total amount of atherosclerosis		Complicated lesion		Calcification		Lipid streak		Fibrous plague	
	Stained	Unstained	Stained	Unstained	Stained	Unstained	Stained	Unstained	Stained	Unstained	Stained	Unstained
Individual specimens												
ntra-observer	485	494	17	19	15	14	18	23	22	37	26	33
Intra-laboratory	1 055	1 077	27	27	23	16	24	32	33	52	43	55
Inter-laboratory	3 785	3 863	39	32	36	42	31	38	64	65	65	61
Grouped specimens												
Intra-observer	485	494	6	9	8	7	5	9	12	12	12	10
Intra-laboratory	1 055	1 077	16	15	16	9	12	16	15	26	23	27
Inter-laboratory	3 785	3 863	23	17	26	36	15	20	43	35	39	31

[a] Excluding readings of pathologists who did not examine stained specimens.

DISCUSSION

TRANSVERSE SECTION

The transverse section methods used have proved difficult, largely because of the difficulty of identifying structural and pathological changes. Grading of "amount of atherosclerosis", where tested, is less precise than when made on longitudinally sectioned specimens. The type of atherosclerosis is not estimated at all.

The main objective of this method is to grade narrowing of the lumen. Table 1 shows that the number of comparisons made on embedded coronaries was a low as 95 (out of a possible total of 243). This indicates that frequently the pathologist was unable to reach a decision.

LONGITUDINAL SECTION [1]

It is clear that this method of grading atherosclerosis is somewhat crude. A good deal of the difference found in two groups of specimens would be due to the method used. From the data available it has been possible to calculate what differences in frequency of a particular lesion found in two groups are likely to indicate that a real difference exists between the groups according to whether the specimens are examined by one observer, different observers from one laboratory or observers from different laboratories. This can best be shown on nomograms, which have been specially prepared.[2] It has been necessary to construct a separate nomogram for each type of lesion graded and according to whether aorta or coronary artery is being studied. Thus for both aorta and coronary artery there are separate nomograms for total amount of atherosclerosis, lipid streak, fibrous plaque, complicated lesion and calcification. Reference to the nomogram concerned with "total amount of atherosclerosis" in aorta illustrates their use and the scope of the method (Fig. 1).[3] For instance, if one observer

[1] In fairness it should be pointed out that some circumstances of the trial were not favourable. For instance, most of the examiners had had little previous experience of the methods used. The PIA group were used to grading stained specimens and to assessing surface areas in terms of percentages. They were asked to examine unstained specimens in the first trial and to assess in units of "none", "some but less than 1/3", "1/3 to 2/3", "more than 2/3".

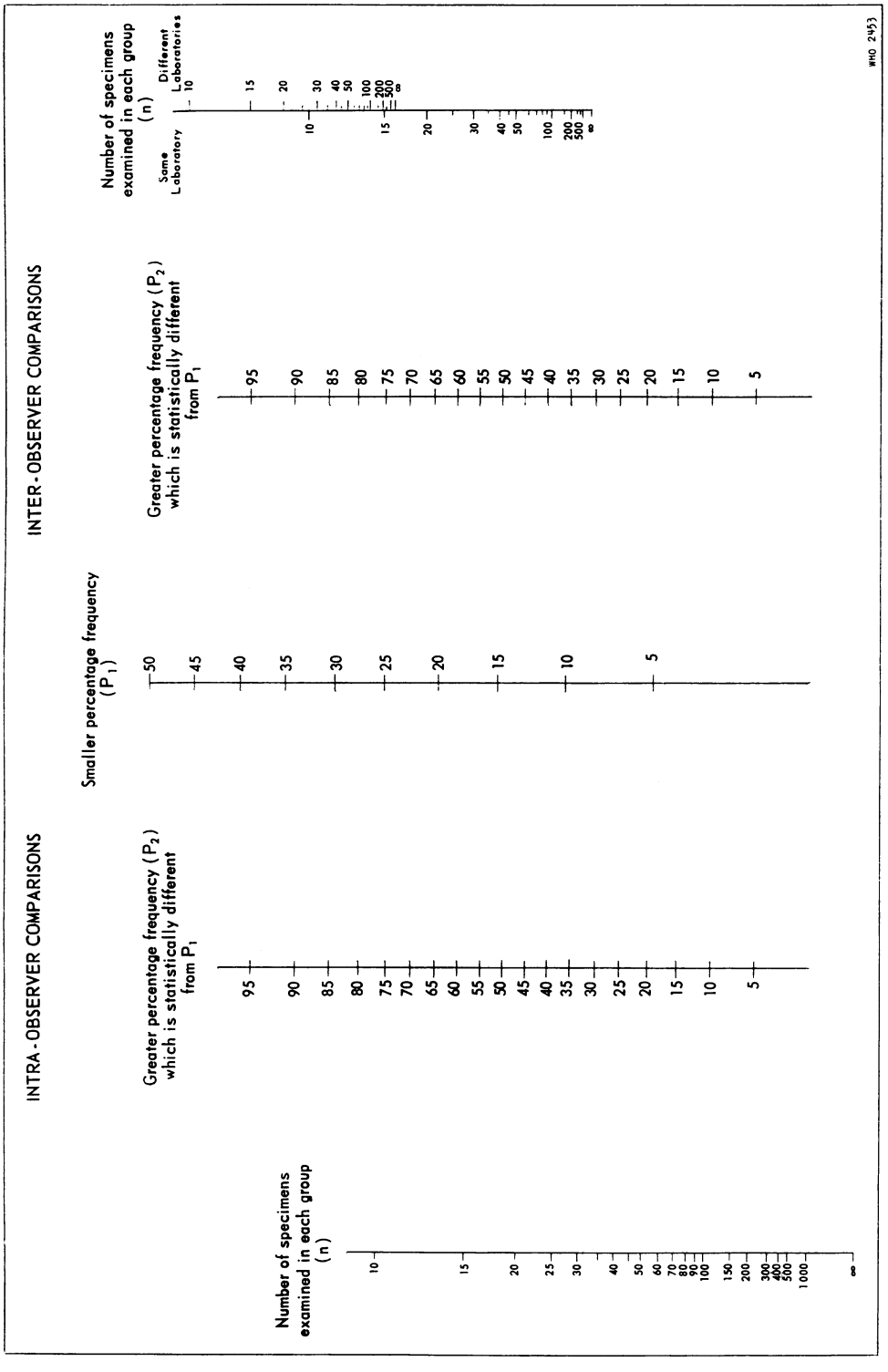
The specimens were all examined at least 28 times before they were stained and they eventually travelled a distance greater than the circumference of the world at the equator. They were inevitably damaged by this treatment and this interfered with the clarity of the staining. It does seem, however, that taking these difficulties into consideration the levels of disagreement are likely to be *less* under more usual circumstances.

[2] The nomogram is based on the results of the analysis of variance of the proportion of specimens classified in a certain grading unit (e.g. "some but less than 1/3") by the observers. The angular transformation was applied and the intra-observer, intra-laboratory and inter-laboratory components of variance were computed. The computation was made separately for each of the four grading units and the four values of each component were averaged. On this basis the sampling error was assessed for each type of comparison and the nomogram constructed.

[3] The nomograms for all the other categories are available, on request, from: Cardiovascular Diseases, World Health Organization, Geneva.

FIG. 1

NOMOGRAM FOR TESTING WHETHER THE PERCENTAGE FREQUENCIES OBSERVED IN TWO GROUPS OF SPECIMENS ARE STATISTICALLY SIGNIFICANT ($\alpha = 0.05$): AORTA, TOTAL AMOUNT OF ATHEROSCLEROSIS



examines all specimens from two groups and in the analysis it is found that 10% of the specimens from one group are graded as showing more than $^2/_3$ atherosclerosis and 20% of the specimens from the second group are graded in the same way, this result would be likely to be due to a real difference between the two groups if there were 150 specimens in each group. This would be equally true if the proportion of specimens in one group showing no atherosclerosis was 10% and the proportion of specimens showing no atherosclerosis in the other group was 20%. The procedure is to place a ruler on the figure in the middle column indicating the lower frequency of a finding in the two groups. Keeping this point fixed, the ruler is brought up to the figure in the column to the left indicating the percentage frequency of the same finding in the higher of the two groups. The column on the extreme left then indicates the number of specimens in each group that would have to be seen for the difference in prevalence to be significant at the 5% probability level.

Conversely, if the number of specimens in each group is known and the percentage frequency of a finding in the two groups is known, the nomogram will show whether this difference is statistically significant under these circumstances. Thus, for example, when 100 specimens from each of two groups are examined by one observer, a difference of 15% and 5% is significant or a difference of 50% and 65% is significant. If examination has been made by workers within a laboratory the centre column and the two columns to the right must be used (the left-hand side of the column on the extreme right). Thus a difference of 5% and 25% frequency of a grading in two groups of 500 specimens is greater than that due to the method. A difference of 5% and 45% would be sufficient to indicate a statistically probable difference with as few as 20 specimens from each group. If the examination is made by observers from different laboratories 500 specimens would be required to attach significance to a difference of as much as 20% and 70% in two groups. A difference of 90% and 35% could be shown up with as few as 20 specimens in each group.

It is clear that a single observer has better discriminating powers than several observers.

SHARING SPECIMENS

From the preceding it can be anticipated that in a combined pathological and epidemiological investigation of atherosclerosis, in which specimens obtained from several different sources are to be compared, the highest precision of the comparison will be achieved if all the specimens are examined by the same pathologist. This is because each pathologist is comparatively consistent in his grading, although the level of the grade he gives to the specimens may tend to be higher or lower than the one given by any other pathologist.

When a large number of specimens have to be studied, this will surpass the capacity of a single pathologist and the examination of specimens has to be carried out by more than one. However, the influences due to the differences of the level of grading among pathologists can be eliminated if specimens from each source are always shared among the same group of pathologists. Different pathologists may, of course, tend to give higher or lower gradings—i.e., each pathologist may have a bias of different magnitude and direction—and hence the grading of the total of the specimens will be affected by an "average" individual bias; but if specimens from each source are equally divided and allocated to each pathologist for examination, the specimens from each source will be influenced equally by the "average" bias and the between-observer bias will affect specimens from all sources equally—i.e., it will be neutralized as a source of difference between groups of specimens. That is to say, that by collecting all specimens from all groups and assigning an equal proportion of specimens from each group to each observer it is possible to cancel the inter-laboratory and intra-laboratory disagreement and to approximate the discriminating power of the method to that of a single observer, and yet retain the ability to deal with a large amount of material.

Example

In order to study the frequency of disagreement which is expected to arise when specimens are divided into groups and each group is examined by different observers, the data relating to the grading of the total amount of atherosclerosis in the aortas were used.

The 45 specimens examined in the grading trials were divided into two groups (Table 5)—namely, a group consisting of specimens numbered 1 to 22 and another group consisting of specimens numbered 23 to 45. The results of the grading of the first group by one observer and those of the second group by another observer from a different laboratory were

TABLE 5

FREQUENCY DISTRIBUTION OF GRADE OF TOTAL AMOUNT OF ATHEROSCLEROSIS IN UNSTAINED AORTAS: COMPARISON OF TWO GROUPS (SPECIMENS 1-22 AND SPECIMENS 23-45)

Observer	First Examination				Total	Second Examination				Total
	None	<1/3	1/3-2/3	>2/3		None	<1/3	1/3-2/3	>2/3	
					Group 1: Specimens 1-22					
1	—	2	7	13	22	—	2	10	10	22
2	—	1	6	15	22	—	1	3	18	22
3	—	3	4	15	22	—	4	7	11	22
4	—	4	9	9	22	—	4	8	9	21[a]
5	—	3	8	11	22	—	2	8	12	22
6	—	6	12	4	22	—	5	11	6	22
7	—	—	8	14	22	—	1	8	12	22
8	—	1	3	18	22	—	—	3	18	21[a]
9	—	1	7	14	22	—	1	5	16	22
10	—	—	7	15	22	—	—	7	15	22
11	—	7	4	11	22	—	8	7	7	22
12	—	3	7	12	22	—	1	11	10	22
13	—	7	9	6	22	—	7	10	5	22
14	—	6	8	8	22	—	6	9	7	22
					Group 2: Specimens 23-45					
1	—	4	12	7	23	—	4	13	6	23
2	—	3	8	12	23	—	5	8	10	23
3	—	4	10	9	23	—	12	7	4	23
4	—	13	7	3	23	—	8	12	3	23
5	—	7	11	4	22[a]	—	6	12	5	23
6	—	9	13	1	23	—	9	13	1	23
7	—	9	11	3	23	—	4	13	6	23
8	—	4	5	14	23	—	3	5	15	23
9	—	5	12	6	23	—	4	14	5	23
10	—	5	12	6	23	—	4	12	7	23
11	1	9	9	4	23	1	15	6	1	23
12	—	10	10	3	23	—	11	10	2	23
13	1	15	6	1	23	1	16	4	2	23
14	1	9	10	3	23	1	10	10	2	23

[a] One specimen was missed during the examination.

added and the total frequency was compared between the first and second examinations. All possible pairs of observers from different laboratories were considered, and the "group disagreement" rates were averaged.

The disagreement rate thus computed gave a figure of 8.5%, which is very close to the intra-observer disagreement rate of 7.7%. The intra-laboratory and inter-laboratory disagreement rates were, respectively, 17.6% and 24.5%.

This method would give a discrimination between observed differences of about 30% for "total amount of atherosclerosis" of aorta with as few as 30 specimens from each group.

STAINING

Agreement within and between observers in grading lipid streak and fibrous plaque in individual specimens was moderately improved with staining (Table 4), but there was no improvement with regard to grouped specimens.

It is of interest to know whether staining was associated with any *quantitative* change in grade. Fig. 2 shows this diagrammatically for aorta and coronary arteries in each of the categories—total amount of atherosclerosis, lipid streak, fibrous plaque, complicated lesion, calcification. When staining caused a quantitative change in grade it was most often in the direction of increase for the total amount of atherosclerosis and lipid streak. In the case of calcification, change in grade with staining was most often in the direction of decrease. A higher grade for the total amount of atherosclerosis and fatty streak in stained specimens was probably due to the greater ease with which red-stained fat could be detected than pale-yellow unstained areas. Calcification was detected largely by touch and graded as a proportion of the total amount of atherosclerosis. Reduction in grade could have been due to an increased assessment of the total amount of atherosclerosis without any change in assessment of the area of calcification.

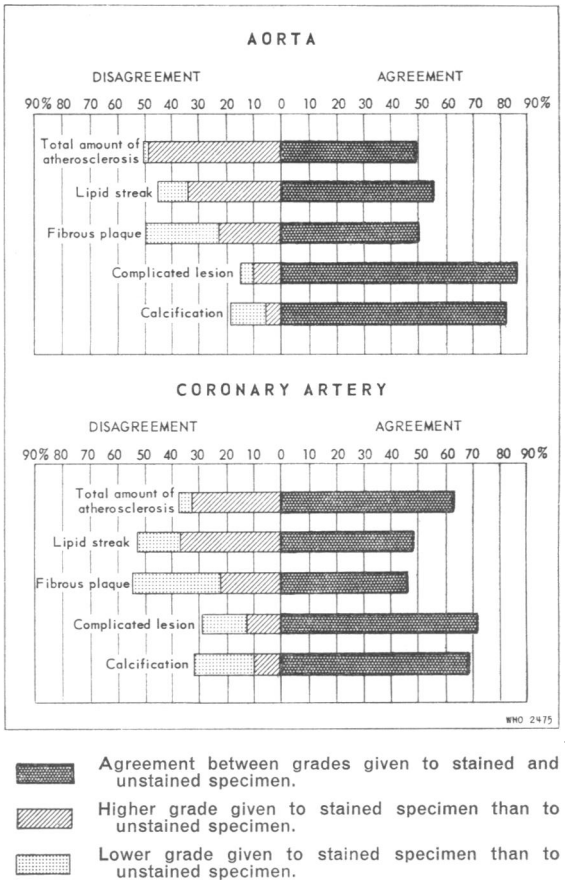
The practical conclusion is that it would be unwise to compare gradings of stained and unstained material.

APPLICATION

A group [1] met in Moscow in March 1962 and drew up a Standard Operating Protocol [2]—a working procedure for comparing atherosclerosis in groups of autopsy material—embodying the principles and guide lines indicated by the trial. In the drafting of this Protocol the 1962 Standard Operating Protocol of the PIA group [3] was kept constantly in mind and many of its recommendations were adopted. In this way it is hoped that material studied by either procedure will be comparable.

FIG. 2

COMPARISON OF GRADE OF ATHEROSCLEROSIS DETERMINED BEFORE AND AFTER STAINING OF SPECIMENS, EXPRESSED AS THE PERCENTAGES OF SPECIMENS GIVEN IDENTICAL GRADES BEFORE AND AFTER STAINING AND OF THOSE GIVEN HIGHER OR LOWER GRADES AFTER STAINING



[1] Dr G. Avtandilov, Nalchik, USSR; Dr A. Whitley Branwood, University of Aberdeen, Scotland; Professor F. Linnell, Allmänna sjukhuset, Malmö, Sweden; Professor A. L. Myasnikov, Institute of Therapy, Academy of Medical Sciences, Moscow, USSR; Dr N. Sternby, Allmänna sjukhuset, Malmö, Sweden; Professor A. Strukhov, First Sechenov Medical Institute, Moscow, USSR; Dr C. Tejada, Instituto de Nutrición de Centro América y Panamá, Guatemala; Professor R. Vaněček, Charles University, Prague, Czechoslovakia; Professor A. M. Vikhert, Institute of Therapy, Academy of Medical Sciences, Moscow, USSR.

[2] Obtainable from: Cardiovascular Diseases World Health Organization, Geneva.

[3] Obtainable from the Instituto de Nutrición de Centro América y Panamá, Guatemala.

CONCLUSIONS

The conclusions brought out by the trial, of practical application to the Standard Operating Protocol, may be summarized as follows:

Transverse section methods are unsuitable.

Longitudinal section methods as described can be recommended for discriminating between two groups of specimens if a difference of as much as 30% is likely to be found and as many as 30 specimens in each group can be obtained.

However, with more than one observer taking part, the latter would apply only if specimens from all sources are equally distributed among all the observers and if, at the time of grading, the observers are kept ignorant of the source of the specimens.

The discriminating power of the method is likely to be increased with practice, and therefore, to obtain the maximum benefit, graders should be subjected to "comparability" tests during the course of the study.

Staining of lipids under standard conditions is recommended because the grading of "total amount of atherosclerosis" and of "fatty streak" is increased under these conditions and because comparison of stained and unstained specimens would not be possible.

A "central service" is recommended so that the collection and preparation of specimens and the collection of associated data under standard conditions can be controlled, specimens can be allocated correctly for grading to the appropriate observers, and comparability tests can be applied to observers from time to time. By the same service, associated data can be separated from the specimens for grading purposes but subsequently correlated with the recorded grading of atherosclerosis.

Annex

DEFINITIONS OF TERMS USED IN THE GRADING TRIALS [1]

Aorta, abdominal: The aorta from a horizontal line drawn through the upper edge of the orifice of the coeliac artery to a horizontal line drawn through the inner surface of the bifurcation.

Aorta, ascending plus arch: That part of the aorta proximal to the descending thoracic aorta (*q.v.*)

Aorta, complete: The aorta from just above the aortic valves to the bifurcation.

Aorta, descending thoracic: The aorta from a horizontal line drawn through the first two intercostal arteries to a horizontal line drawn through the upper edge of the orifice of the coeliac artery.

Atherosclerosis: A variable combination of change of the intima of the arteries consisting of the focal accumulation of lipids, complex carbohydrates, blood and blood products, fibrous tissue and calcium deposits and associated with medial changes.

Calcification: Area in which there is calcium deposition detectable either visually or by palpation.

Complicated lesion: Area in which there is haemorrhage or thrombosis with or without calcium deposition.

Coronary, circumflex: The circumflex branch of the left coronary artery from its origin and as far as possible.

Coronary, left anterior descending: The left coronary artery from its orifice (including the ostium) and the anterior descending branch down to the apex of the left ventricle, if possible.

Coronary, right: The right coronary artery from its origin (including the ostium) and including the flexure at the margin of the posterior interventricular septum, if possible.

Fibrous plaque: Any firm, elevated lesion which is pale grey, glistening or translucent. If a lesion presents any haemorrhage, thrombosis, ulceration or calcification, such portion will be classified in that category and not as fibrous plaque.

Lipid (fatty) streak and plaque: Any intimal lesion of a distinctly yellow colour that does not show any other type of underlying change.

[1] There have been a number of changes in the 1962 Standard Operating Protocol.

ACKNOWLEDGEMENTS

The World Health Organization is indebted to the following, who gave unstintingly of their time and advice, and without whom the study would have been impossible:

CZECHOSLOVAKIA	Dr K. Kubát,[1] Charles University, Prague Dr K. Motlík, Charles University, Prague Professor R. Vaněček,[1] Charles University, Prague Professor Z. Záhor,[1] Charles University, Prague Dr T. Zemplenyi, Institute for Cardiovascular Research, Prague
FEDERATION OF MALAYA	Professor J. B. Duguid, Institute for Medical Research, Kuala Lumpur
GUATEMALA	Dr C. Restrepo,[1] Instituto de Nutrición de Centro América y Panamá Dr C. Tejada,[1] Instituto de Nutrición de Centro América y Panamá
SWEDEN	Professor F. Linnell,[1] Allmänna sjukhuset, Malmö Dr N. Sternby,[1] Allmänna sjukhuset, Malmö
UNION OF SOVIET SOCIALIST REPUBLICS	Professor A. M. Vikhert and three colleagues, Institute of Therapy, Academy of Medical Sciences, Moscow
UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND	Mr R. M. C. Arnot, Welwyn Garden City, Herts. Professor T. Crawford,[1] St George's Hospital, London Dr D. Dexter,[1] St George's Hospital, London
UNITED STATES OF AMERICA	Professor H. C. McGill, Louisana State University, New Orleans, La. Dr J. P. Strong, Louisana State University, New Orleans, La.

RÉSUMÉ

Des essais pour coter l'athérosclérose sur des aortes et des artères coronaires sectionnées longitudinalement, ainsi que sur des coronaires sectionnées transversalement, ont été effectuées par 14 pathologistes appartenant à 6 pays.

Une série de pièces anatomiques a été spécialement préparée. Les définitions, critères, techniques et méthodes de mesures ont été précisés, testés par un premier essai et ensuite modifiés. C'est après ces modifications que ces définitions, critères et techniques ont été utilisés au cours du présent essai.

L'examen des sections transversales s'est avéré décevant car il était impossible aux pathologistes d'identifier de façon parfaite les structures normales et pathologiques, et lorsqu'ils y parvenaient, la mesure quantitative de l'athérosclérose était très inférieure à celle obtenue par d'autres méthodes. En outre, cette méthode n'a pu donner d'indication sur le type d'athérosclérose.

Les pièces sectionnées longitudinalement ont tout d'abord été examinées sans être colorées. En ce qui concerne les préparations, les divergences chez le même observateur ont été de l'ordre de 20% pour « quantité totale d'athérosclérose », « lésion compliquée » et « lésion calcifiée ». Pour les « dépôts graisseux » et « plaques fibreuses », le pourcentage de divergences a été à peu près le double (environ 40%). Lorsque les mêmes coronaires ont été colorées et réexaminées par le même chercheur, le taux de divergences pour « dépôts graisseux » et « plaques fibreuses » a été réduit respectivement à 22 et 26%. Les divergences d'un chercheur à l'autre dans le même institut ou d'un institut à l'autre sont très supérieures à celles du même observateur. La coloration a augmenté quantitativement l'estimation de la « quantité totale d'athérosclérose » et des « dépôts graisseux »; ainsi elle est à recommander lorsqu'on veut faire des comparaisons.

Des nomogrammes ont été construits pour l'aorte, les artères coronaires et pour chaque type de lésion afin de montrer les rapports entre le nombre de préparations anatomiques, les méthodes d'examen (un seul observateur, plusieurs observateurs dans un même institut; observateurs appartenant à plusieurs instituts) et l'existence de différences significatives dans les constatations de deux groupes de pièces. Grâce à ces nomogrammes, l'on a pu répondre à la question de savoir si la différence relevée entre deux groupes de préparations correspondait à une différence réelle ou si elle n'était que le reflet de l'imperfection de la méthode. Celle-ci a permis d'établir un écart d'environ 30% entre les divergences d'observations pour la « quantité totale d'athérosclérose » de l'aorte avec seulement 30 pièces de chaque groupe lorsqu'un seul observateur évalue les préparations des deux groupes.

[1] Took part in the postal grading trial.

Les méthodes décrites peuvent être utilisées pour montrer une différence d'athérosclérose entre 2 ou plusieurs groupes de préparations anatomiques. Elles ne permettent pas d'obtenir des valeurs absolues pour la quantité d'athérosclérose.

Les résultats de cet essai ont été utilisés pour la mise sur pied d'un protocole standard en vue d'une comparaison de l'athérosclérose dans différents groupes d'artères coronaires et d'aortes prélevées à l'autopsie.

L'on est toujours à la recherche de méthodes améliorées de cotation de l'athérosclérose; en attendant, les limites dans lesquelles les méthodes présentes peuvent être utilisées sont maintenant tracées.

REFERENCES

Baker, A. B., Refsum, S. & Dahl, E. (1960) *Neurology*, **10**, 525

Dickinson, C. J. & Thomson, A. D. (1959) *Lancet*, **2**, 46

Elkeles, A. (1957) *Lancet*, **2**, 714

Gore, I. & Tejada, C. (1957) *Amer. J. Path.*, **33**, 875

Holman, R. L., McGill, H. C., Strong, J. P. & Geer, J. C. (1958) *Lab. Invest.*, **7**, 42

Patterson, J. C., Cornish, B. R. & Armstrong, E. C. (1956) *Canad. med. Ass. J.*, **74**, 538

Scott, R. F., Daud, A. D., Florentin, R. A., Davies, J. N. P. & Coles, R. M. (1961) *Amer. J. Cardiol.*, **8**, 165

Tejada, C., Gore, I., Strong, J. P. & McGill, H. C. (1958) *Circulation*, **18**, 92

World Health Organization, Study Group on Atherosclerosis and Ischaemic Heart Disease (1957) *Wld Hlth Org. techn. Rep. Ser.*, **117**

World Health Organization, Study Group on the Classification of Atherosclerosis Lesions (1958) *Wld Hlth Org. techn. Rep. Ser.*, **143**

Yater, W. M., Welsh, P. P., Stapleton, J. F. & Clark, M. L. (1951) *Ann. intern. Med.*, **34**, 352