Brit. Heart J., 1965, **27**, 595.

THE CODING OF SURVEY ELECTROCARDIOGRAMS BY TECHNICIANS

BY

GEOFFREY ROSE

From the Department of Medical Statistics and Epidemiology, The London School of Hygiene and Tropical Medicine, Gower Street, London W.C.1

Received November 19, 1964

Epidemiological surveys of ischæmic heart disease generally depend upon electrocardiographic diagnosis. Unfortunately the insecurity of this foundation is increasingly evident: there are major difficulties both in standardizing readings and in coping with very large numbers of records. Computer analysis offers some hope of assistance; but it is not yet practicable for ordinary users, and, at least within the next few years, it is unlikely to be helpful to any but large groups (Rose, 1965). The immediate need is for a practicable system, capable of dealing with fairly large numbers of records in a standardized manner.

Problems of Standardized Reading

The criteria by which clinicians interpret electrocardiograms mostly lack precise definition, and vary moreover from one cardiologist to another. Even between expert readers there are large disagreements, both random and systematic (Thomas, Cochrane, and Higgins, 1958; Davies, 1958; Epstein *et al.*, 1961). This led workers at Minnesota (Blackburn *et al.*, 1960) to propose a coding system based on measurements rather than interpretation. In this system, now known as the Minnesota Code, records are sorted into several categories for each of various axes of classification (e.g. Q/QS items; S–T depressions; etc.). Each category is defined in terms of exact measurements in specified leads.

It seemed reasonable to expect that this descriptive, rule-of-thumb approach should eliminate the main sources of observer variation, and the originators of the code were able to show that reading disagreement between several members of their group was fairly small (Blackburn *et al.*, 1960). It appeared, however, that larger disagreements might arise between investigators who came from different units, and in 1961 a World Health Organization Scientific Group recommended formal comparative tests (World Health Organization, 1963). Preliminary results have been announced (Kagan, 1965).

Some of the results from an independent study of the same problem are shown in Table I. Seven observers from 5 countries took part, each of whom had used the Minnesota Code in surveys of international importance. Five of the observers (A to E) independently coded Xerox reproductions of the same set of 60 tracings; observer F coded 14 of the same set; observers A and G coded a different set of 70 tracings. The records were from population surveys, but enriched so that about two-thirds showed ischæmic-type changes. Observer A's frequency of coding Q/QS, S–T, and T items is taken as unity. For the other observers the figures show reporting frequencies relative to A: they are thus a measure of systematic disagreements. These prove to be alarmingly

TABLE I

RELATIVE FREQUENCY OF REPORTING VARIOUS ITEMS BY 7 PHYSICIANS FROM 5 COUNTRIES, USING THE MINNESOTA CODE

Code	Description	Relative frequency of reporting						
		A	B	C	D	E	F	G
1:1–3	Q/QS items	1·00	1·1	0·9	1·2	1·0	2·5	0·5
4:1–3	S–T depression	1·00	0·6	0·8	1·0	0·8	0·8	0·2
5:1–3	T inversion/flattening ..	1·00	1·1	1·1	0·6	1·0	1·3	1·0

big in proportion to the true between-population differences with which we are concerned, and it has to be admitted that even this descriptive approach has not achieved the degree of standardization that is essential for comparative studies by independent investigators.

CAUSES OF DISAGREEMENT

Study of the causes of disagreement between these 7 observers showed that few were due to simple oversights. Some originated in unwitting infringement of the rules of the Minnesota Code: for example, some QS patterns in leads III and aVF had been coded as though they were Q waves. But in most cases of disagreement it could not be said that anyone had plainly broken the rules: it is inherently difficult to make measurements to one-hundredth of a second or a fraction of a millimetre, and slight differences in techniques may have important effects on the coding.

Fortunately it appeared that a few recurrent situations were accounting for much of the disagreement. These included difficulties in: (1) measuring Q wave duration; (2) defining the least positive deflection in right chest leads to qualify as an R wave (thereby distinguishing S from QS); (3) precise definition of the start and end of S–T segments; (4) measuring S–T depression in the presence of baseline drift; (5) distinguishing low positive T waves from flat T waves; and (6) coding records which showed beat-to-beat variation.

REDUCING DISAGREEMENTS

A set of rules was devised to guide coders on precisely how they should deal with these and other problem situations. These rules supplement the Minnesota Code and in no way conflict with it. Their aim is to promote maximum uniformity in measurement techniques, so that coders when in difficulty may consult the rule book rather than private opinion.

Unfortunately it proves impossible to describe a standardized method wholly by a set of rules. Illustrations are necessary, and for this purpose 30 of the original 60 standard records (a stratified random sample) were set aside for use as a demonstration and training series. "Official" codings of these 30 records were prepared, together with comments on the particular issues which they raise. The remaining 30 provided a test series.

Eight physicians from six countries co-operated in an experiment to test the effect of this training system upon inter-observer agreement. Observers A, B, and C were the same as in the first test; the rest were cardiologists with little or no previous experience in the use of the Minnesota Code. Each observer first studied the supplementary rules, together with the training series of records and "official" codings, and then proceeded to code the test series (Table II). It seems that detailed and co-ordinated training has substantially reduced the extent of systematic disagreements.

CODING BY TECHNICIANS

The physicians' improved performance resulted from a system in which they no longer functioned as doctors, but simply as measurers and coders: private judgements were replaced by rules.

TABLE II

RELATIVE FREQUENCY OF REPORTING VARIOUS ITEMS BY 8 PHYSICIANS FROM 6 COUNTRIES, AFTER AGREEMENT ON SUPPLEMENTARY RULES AND STUDY OF REFERENCE TRACINGS

Code	Description	Relative frequency of reporting							
		A	B	C	H	I	J	K	L
1:1–3	Q/QS items	1·00	1·0	0·9	0·8	0·8	1·1	1·3	1·0
4:1–3	S–T depression	1·00	0·9	0·9	1·1	1·4	1·1	1·1	1·0
5:1–3	T inversion/flattening ..	1·00	1·0	1·1	0·9	1·1	0·9	1·0	1·0

Generally speaking routine technical work is done better by technicians than by doctors: they are free of preconceived ideas, more willing to accept the necessary intellectual discipline, and often temperamentally more suited to repetitive work. Training in the use of the Minnesota Code was therefore given to 4 medically-unqualified staff members in this Department: 2 were statistical assistants and 2 were nurses.

The trainees first received very brief instruction in the electrophysiology of the heart; but most of the teaching dealt with identification of waves and reference points, and techniques of precise measurement. Extensive use was made of the "training series" and notes mentioned earlier. The Minnesota system of classification appeared formidable at first, but with growing familiarity it soon came to be handled with confidence. The total amount of teaching time from start to finish of the training was 7 hours; in addition the trainees spent a longer period (12 hours or so) working on their own.

The quality of their performances after training was tested by requiring each one to code the "test series." The results for major ischæmic-type items (Table III) are again given as reporting frequencies relative to Observer A. The average performance is slightly better than that of the trained physicians reported in Table II, and considerably better than that of the groups in Table I.

Subsequently (though still before they had had much experience) 3 of these 4 technicians were included in the World Health Organization tests, which measured both between- and within-observer variation. Preliminary results (Kagan, 1965) showed that the technicians did much better than 3 cardiologists who had not been trained in the use of the Code, and at least as well as 3 other cardiologists with a large experience of it.

The technicians now undertake the routine coding of survey electrocardiograms in this Department. Each record is coded independently by 2 technicians. Agreed reports are accepted: disagreements and queries are discussed with the physician in charge, who arbitrates where necessary. From time to time also the physician reviews a sample of their agreed reports.

This system of duplicate reporting has several advantages.

(1) It provides for quality control, with a continuous measure of "random" and systematic inter-observer differences.

(2) It maintains interest and concentration, since all reports are checked.

TABLE III

RELATIVE FREQUENCY OF REPORTING VARIOUS ITEMS BY 4 TECHNICIANS, AS COMPARED WITH PHYSICIAN A

Code	Description	Relative frequency of reporting				
		A	N	O	P	Q
1:1–3	Q/QS items	1·00	1·2	1·2	1·0	1·2
4:1–3	S–T depression	1·00	1·0	1·0	0·7	1·0
5:1–3	T inversion/flattening ..	1·00	0·8	1·0	1·0	1·0

(3) It improves the quality of reporting. This is partly because one technician may report on what the other misses. But in addition the fact of disagreement spotlights the difficult records, which can then be reviewed more carefully. Tests on the W.H.O. series of tracings gave evidence of benefit from this source (Table IV). This series, comprising 100 records in all, was composed of 50 different tracings each of which appeared twice, in random positions. The series was coded independently by Observer A (a physician), Observer N (a nurse), and Observers O and P (statistical assistants). A "combined" report was then made, according to normal survey procedure: reports for 2 technicians were compared, and if they agreed, their verdict was taken: if they disagreed, Observer A gave a ruling. (Codings of each of the 3 possible pairs of technicians were taken in rotation.) It can be seen that the system based on duplicate reports has achieved almost complete repeatability.

TABLE IV

FREQUENCY OF INTRA-OBSERVER DISAGREEMENTS FOR INDIVIDUAL CODERS, AND FOR A 'COMBINED' REPORT FROM 2 TECHNICIANS

Item	Percentage disagreement on presence or absence of any codable item				
	A	N	O	P	'Combined' report
Q/QS	0	8	16	7	0
S–T depression	6	11	8	7	3
T flattening/inversion	6	0	3	0	0

CONCLUSIONS AND SUMMARY

The Minnesota Code, which substitutes descriptive for interpretative categories, represents a great advance in standardized reporting on electrocardiograms. Nevertheless, as used hitherto, it fails to provide an adequate basis for comparing the results of different workers. This situation can be improved by (a) standardized training, with supplementary rules on exact techniques of measurement; and (b) testing performance routinely. Such a system can be readily taught to medically-unqualified technicians, and the evidence confirms that they are capable of coding survey electrocardiograms at least as well as experienced cardiologists. Duplicate reporting yields a substantial gain in precision.

It may well be that other repetitive medical techniques, hitherto restricted to doctors, could sometimes with advantage be undertaken by technicians.

This investigation was supported by U.S. Public Health Service Grant HE–047 75, from the National Heart Institution.

REFERENCES

Blackburn, H., Keys, A., Simonson, E., Rautaharju, P., and Punsar, S. (1960). The electrocardiogram in population studies. A classification system. *Circulation*, **21**, 1160.

Davies, L. G. (1958). Observer variation in reports on electrocardiograms. *Brit. Heart J.*, **20**, 153.

Epstein, F. H., Doyle, J. T., Pollack, A. A., Pollack, H., Robb, G. P., and Simonson, E. (1961). Observer interpretation of electrocardiograms. Suggestions for objective interpretations. *J. Amer. med. Ass.*, **175**, 847.

Kagan, A. (1965). Comparability of electrocardiographic data, with particular reference to the W.H.O. Minnesota code trial. *Milbank mem. Fd. Quart.* In the press.

Rose, G. A. (1965). Computer analysis of electrocardiograms. In *Mathematics and Computer Science in Biology and Medicine: Proceedings of Conference held by Medical Research Council in Association with Health Departments, Oxford, July 1964.* H.M.S.O., London. In the press.

Thomas, A. J., Cochrane, A. L., and Higgins, I. T. T. (1958). The measurement of the prevalence of ischæmic heart disease. *Lancet*, **2**, 540.

World Health Organization (1963). Arterial hypertension and ischæmic heart disease: comparison in epidemiological studies. 5. Work in progress. *Chron. Wld Hlth Org.*, **17**, 97.