Brit. J. prev. soc. Med. (1970), **24**, 197–200

REPRODUCIBILITY OF ANALYSIS OF THE ELECTROCARDIOGRAM IN EPIDEMIOLOGY USING THE 'MINNESOTA CODE'

I. ELGRISHI, P. DUCIMETIERE

AND

J. L. RICHARD

Groupe d'Etude sur l'Epidémiologie de l'Athérosclérose, 27 rue Lacordaire, Paris 15 ème

Use of the electrocardiogram is necessary to advance knowledge of the epidemiology of ischaemic heart disease. Analysis of the electrocardiogram in epidemiological practice, however, may be difficult since the clinical method of interpretation cannot be used because of its deficiencies in standardization and reproducibility (Blackburn, 1968; Burgess, Fejfar, and Kagan, 1963; Davies, 1958; Higgins, Kannel, and Dawber, 1965; Rose, 1965; Rose and Blackburn, 1968; Scandinavian Committee on ECG Classification, 1967).

Therefore, the epidemiologist must avoid any interpretation of the electrocardiogram. He has to give a description of it which should be as reproducible as possible and yet relevant to the diseases under study. Such a description has to be based on a set of measurements and not on the qualitative evaluation of certain patterns.

The method of classification known as the Minnesota Code (Blackburn, Keys, Simonson, Rautaharju, and Punsar, 1960; Blackburn, 1968) answers this requirement. Since 1966 the Groupe d'Etude sur l'Epidémiologie de l'Athérosclérose (GREA) has used a particular protocol for this code; details have been published elsewhere* (Elgrishi, Ducimetière, Richard, and Gelin, 1969). The present study examines the reproducibility of an analysis carried out in this way.

Materials and Methods

Four technicians were trained under the direction of one of us (I.E.) in the use of the version of the Minnesota Code adopted by the GREA. Their training lasted about three months. They attend regularly at collective sessions for training and revision.

Each technician coded, on two occasions (although not aware of this), 80 electrocardiograms selected from among those of a cardiovascular survey then in progress; 60 had been recorded at rest and 20 after exercise. Twenty-five, which were within normal limits, were chosen at random. The 55 abnormal ones were selected so that the main electrocardiographic features analysed in the code were represented in the sample. The technical quality of the tracings was ignored in making the selection. In this way, each electrocardiogram was coded eight times, a total of 640 classifications. Their comparison is the subject of the present study.

The tracings recorded at rest include leads V1, V2, V3, V4, V5, V6, I, II, III, aVL, aVF, and aVR. Standing immediately after exercise (sub-maximal exertion on a bicycle ergometer) leads V6, V4, V2, I, II, III, aVR, aVL, aVF, V6 and V4 were recorded, in that order.

In contrast to the original Minnesota Code protocol, the findings after exertion were coded independently from the findings at rest.

From the formulae used by the GREA it is possible to obtain a classification of the tracings according to the international conventions of the Minnesota Code**. The results presented in this study are therefore expressed with these international conventions.

The eight readings made for each electrocardiogram were compared with a reference interpretation. The most common code given by the technicians was selected for each tracing; it is therefore called the majority reference code (MRC). In no case was an arbitration made between divergent interpretations.

The results were analysed according to two criteria, accuracy (validity with regard to the MRC) and reproducibility.

1. Accuracy was estimated by comparing the interpretation of each technician with the MRC.

*The main modifications concern three sections of the Minnesota Code: The Q wave items are presented in the text in topographical order.
A code is required for each corresponding group of leads.
A code is required for each relevant lead in the J point and T wave sections.
**A program in Fortran IV is available.

This measured the variability of an observer with reference to a typical codification.

2. Reproducibility was estimated by comparing two interpretations of the same tracing carried out independently by the same technician. This measured the variability between two readings for each observer.

Both criteria are expressed as a percentage: in the case of accuracy, this refers to the frequency of agreement of an observer's interpretation with the MRC; in the case of reproducibility, this refers to the frequency of agreement between two interpretations of the same observer.

Results

Interpretation of ECG according to 'Normal or Abnormal' Classification

Twenty-three electrocardiograms were considered 'normal' in the majority of readings (absence of any codable abnormality in any section of the code; corresponding codification, 0); for these 23 tracings there were 184 readings, of which 163 (89%) conclude a 'normal tracing'. The conclusions of the 21 divergent readings are: presence of a Q wave (9 cases); presence of a 'flat' T wave (6 cases); tachycardia (4 cases); and left QRS axis deviation (2 cases). For the remaining 57 tracings there were 456 analyses, of which 13 (3%) conclude a 'normal tracing' and are therefore in disagreement with the MRC. The disagreements concerned the presence of a Q wave (3 cases); a Wolff-Parkinson-White syndrome (3 cases); a first-degree atrioventricular block (4 cases); and high-amplitude R wave (3 cases).

The specificity of the discrimination between normal and abnormal tracings with reference to the MRC can be estimated as 89% and its sensitivity as 97%.

It should, however, be pointed out that the abnormalities which lead to a given tracing being judged 'abnormal' may differ from one interpretation to another.

Interpretation of Different Sections of the Code

For each section of the Code, the mean coefficient of reproducibility and accuracy lies between 93% and 98% in the presence of a negative finding (MRC = 0). The detailed results are not presented here.

The table gives the results for each section in the presence of a positive finding (MRC different from 0). The 'ventricular conduction defect' section is not included in the table. One case of complete right bundle-branch block and one case of incomplete right block received majority coding. In the first case the eight readings are in agreement, and in the second case one investigator coded a complete block in both readings.

Discussion

The problems of reproducibility and comparability of the readings of electrocardiograms have been studied by many authors. A comparison with these studies is difficult, for two reasons:

1. The methods of reading are not always comparable. The use of the Minnesota Code has not been systematic. The original protocol of the Minnesota Code has been subjected to individual modifications which were sometimes considerable.
2. Accuracy and reproducibility have not been expressed by the same coefficients as are employed here.

An analysis of electrocardiographic tracings by seven cardiologists allowed Rose (1965) to verify the marked variations in the estimation of J point

Table
Accuracy and Reproducibility of ECG Analysis

Section of Minnesota Code	No. of ECGs	No. of ECGs with MRC Different from 0	Mean Coefficient of Accuracy	Mean Coefficient of Reproducibility
Q.QS patterns Codes 1, 2, 3	60	13	73% (58–81)	71% (54–85)
Left axis deviation of QRS Code 1	80	8	81% (69–94)	69% (50–88)
High-amplitude R waves Code 1	60	5	70% (40–100)	65% (40–80)
J point and ST depression Codes 1, 2, 3	80	23	77% (65–87)	71% (61–78)
T wave item Codes 1, 2, 3	80	38	88% (82–92)	87% (79–95)
Atrioventricular conduction defect Codes 3, 4, 5	80	6	69% (58–84)	63% (50–66)
Arrhythmias Codes 1 to 9	80	19	89% (81–95)	86% (80–89)

The range of the technicians' individual scores is given in parentheses

depression and T wave amplitude when the reading was made according to the personal criteria of each specialist. Besides, he showed that this variability is reduced by the application of precise rules of measurement.

Interpretations of 31 electrocardiograms with the Minnesota Code performed by Higgins *et al.* (1965) are in agreement in 39% of cases. When the detail of the analysis is considered, agreement was achieved in 87% of cases in regard to the Q wave, in 77% of cases in regard to the J point and in 74% of cases in regard to the T wave amplitude.

A double interpretation of 440 electrocardiograms was carried out by Higgins *et al.* (1965). The interpretation showed a difference in 93 (21%).

The order of magnitude of the variability of the interpretation indicated by previous authors (Blackburn, 1968; Burgess *et al.*, 1963; Davies, 1958; Most, Hornsten, Hofer, and Bruce, 1968) is also found in the present study.

The simple discrimination between normal and abnormal tracings is fairly good since the specificity and the sensitivity are respectively 89% and 97%. However, as has already been pointed out, the abnormalities which lead to the conclusion 'abnormal tracing' may be different from one observer to another. Examination of the results shows that the differences in interpretation between observers are of the same order as those found for the same observer. In this connection it should be remembered that the technicians were trained by the same doctor and that they participated in regular common training sessions. On the other hand, the coefficients of accuracy and reproducibility are not independent; the better the reproducibility, the greater is the chance of a large number of interpretations in agreement with the majority interpretation. In contrast, for each section the results are clearly different according to whether the MRC does or does not include a codable abnormality. The accuracy and reproducibility are less high in the presence of an abnormality (majority codification other than 0). In this case, the appraisal of the importance of the abnormality (codification 1, 2, 3 . . .) introduces an extra variability into the interpretation.

The most important sections in the Minnesota Code in regard to ischaemic heart disease (ST depression and T wave) gave very satisfactory results.

It is, however, difficult to code these sections with precision. The quality obtained appears to be related to the following facts:

1. The rules for the use of the Minnesota Code in regard to these sections are very precise.

2. An important part of the technicians' training was devoted to the study of ventricular repolarization.

3. The presentation of the questionnaire was planned in such a way that the investigators had to give an individual answer for each electrocardiographic lead. This limits the consequences of inadvertency.

The results are slightly less good for Q waves and QS patterns. This is perhaps explained by the fact that in the questionnaire we use, this section is presented in a less analytical way than the ones concerning the T waves or the J point. (A response is required for a set of leads and not for each lead.) Added to this, the technicians were trained to code Q waves or QS patterns conservatively.

If one groups together different sections of the code, an overall estimation of the mean accuracy and reproducibility of the analysis is possible for all the sections. One hundred and fourteen abnormalities were coded as representing the MRC. These majority patterns were noted 744 times out of the 919 corresponding analyses, i.e., 82%. For all patterns coded 0 or not codable (486 analyses), the corresponding proportion is 97%. For the 114 abnormalities recorded as majority interpretations, the coefficient of reproducibility for the two readings was 79%. This reproducibility is 96% in respect of codes 0.

The incidence of the various abnormalities observed in this study does not, of course, reflect the morbidity in the population investigated, since the electrocardiograms were selected to include an over-representation of electrocardiographic signs of ischaemia and infarction. In considering the above overall estimations one can say that the value of the coefficient of accuracy in the course of an investigation in a real population is somewhere between 82% and 97% and the value of the coefficient of reproducibility is between 79% and 96%. In both cases the real value is proportionately nearer the upper limit than the lower limit as the incidence of electrocardiographic abnormalities is low in the population under study.

It seems that improvement in the quality of the interpretation could be achieved by

(*a*) repeating periodically, with short series, studies of this type. The results should be communicated with comments to the investigators;

(*b*) repeating as often as possible the collective sessions for training in such a way as to create a homogeneous team;

(*c*) a further examination of the presentation of the most controversial sections of the code (Q

waves—disturbances of ventricular conduction);

(*d*) requiring regular work of the investigators, and perhaps imposing a maximum number of electrocardiograms to be interpreted each day, the object being to avoid mistakes due to fatigue and lapses of attention;

(*e*) measurement magnification devices.

In practice, it seems desirable to have the electrocardiogram interpreted by at least two technicians. When the readings disagree, arbitration is necessary. We suggest that this should be done, under the direction of the responsible physician, by the group of investigators themselves. This method has the obvious advantage of providing only one interpretation for each electrocardiogram and, besides, of contributing to improvement in the standardization of reading.

Although the reproducibility of the interpretation is still imperfect and much effort is still required, the Minnesota Code, with its procedural application, remains at the present time the only means of achieving a standardization of the manual coding of electrocardiograms in the course of epidemiological investigations. Much of the element of reproducibility is improved by computer programming for ECG coding but there is still a large commitment of many population studies with conventional paper tracings.

This study was carried out by the Groupe d'Etude sur l'Epidémiologie de l'Athérosclérose with the participation of the Institut National de la Santé et de la Recherche Médicale, the Préfecture de Paris (DGASS) and the Clinique Cardiologique de l'Hôpital Boucicaut, Paris.

We are grateful to Mesdames Couly, Cresson, Forge, Gelin and Richard, who kindly carried out the interpretation of the electrocardiograms, to M. Myquel for the programming, and to Associate Professeur Himbert for much helpful advice in the preparation of the text.

REFERENCES

BLACKBURN, H. (1968). The exercise electrocardiogram: differences in interpretation. Report of a technical group on exercise electrocardiography. *Amer. J. Cardiol.*, **21**, 871.

——, KEYS, A., SIMONSON, E., RAUTAHARJU, P., and PUNSAR, S. (1960). The electrocradiogram in population studies. A classification system. *Circulation*, **21**, 1160.

BURGESS, A. M., FEJFAR, Z., and KAGAN, A. (1963). Hypertension artérielle et cardiopathies ischémiques. Comparabilité des études épidémiologiques. World Health Organization, Geneva.

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

ELGRISHI, I., DUCIMETIÈRE, P., RICHARD, J. L., and GELIN, J. (1969). L'analyse de l'électrocardiogramme en épidémiologie. Protocole d'emploi du code du Minnesota. *Path. et Biol.*, **17**, 175.

HIGGINS, I. T. T., KANNEL, W. B., and DAWBER, T. R. (1965). The electrocardiogram in epidemiological studies. *Brit. J. prev. soc. Med.*, **19**, 53.

MOST, A. S., HORNSTEN, T. R., HOFER, V., and BRUCE, R. A. (1968). Exercise ST changes in healthy men. *Arch. intern. Med.*, **121**, 225.

ROSE, G. A. (1965). The coding of survey electrocardiograms by technicians. *Brit. Heart J.*, **27**, 595.

——, and BLACKBURN, H. (1968). Cardiovascular Survey Methods. *Wld Hlth Org. Monogr. Ser.*, no. 56.

SCANDINAVIAN COMMITTEE ON ECG CLASSIFICATION (1967). The 'Minnesota code' for ECG classification. Adaptation to CR leads and modification of the code for ECGs recorded during and after exercise. *Acta med. scand.*, Suppl. 481.