An analysis of serial Minnesota ECG code changes in the London cohort of the WHO multinational study of vascular disease in diabetes

H. Keen¹, N. Morrish², E. T. Lee³, WHO Multinational Study Group*

¹ Unit for Metabolic Medicine, Guy's Hospital, London, UK

² Bedford General Hospital, Bedford, UK

³ Centre for American Indian Health Research and Department of Biostatistics and Epidemiology, College Public Health, University of Oklahoma Health Sciences Centre, Oklahoma City, USA

Abstract

Aims/hypothesis. Deterioration and improvement in the electrocardiogram are important outcomes in cardiovascular disease progression assessment. We used a sample of serial records from the WHO Multinational Study of Vascular Disease in Diabetes (WHO MSVDD) to assess Minnesota coding variability.

Methods. A constructed subsample of 118 of the 352 paired (baseline and follow-up) and previously Minnesota-coded ECG records from the London cohort was randomised and re-read independently of the first code (respectively 11 and 0.5 years later) by the same two coders. Detailed Minnesota codes were summary coded into groups 1 (CHD unlikely), 2 and 3 (CHD possible and probable, respectively).

Results. Re-reading of the constructed sample for the baseline records (11 years later) generated 21 Summary code reassignments (2 unlikely to possible or probable; 19 possible or probable to unlikely); re-reading for the follow-up records (0.5 years later)

generated only 8 summary code reassignments (21 vs 8 $p < 0.001$) (3 unlikely to possible or probable; 4 possible or probable to unlikely; 1 probable to possible). Re-reading increased the estimated net ECG deterioration in the constructed sample from 11.8% to 25.4%. Consistency analysis showed most variability in marginal baseline abnormalities.

Conclusion/interpretation. Coding variability is now small though re-reading suggests some time-dependent coding drift. Relative over-reading at baseline suggests that the change reported in the complete WHO MSVDD cohort at follow-up was underestimated and that almost all of the reported ECG deterioration and about half of the reported ECG 'improvement' was real. [Diabetologia (2001) 44 [Suppl 2]: S72–S77]

Keywords Minnesota code, cardiovascular disease incidence, WHO Multinational Study, diabetic macrovascular disease, electrocardiogram, reproducibility.

The Minnesota code was developed [1] in the early 1960's as a standardised, semiquantitative approach to the classification of electrocardiographic abnormalities for cross-sectional epidemiological studies of the prevalence of cardiovascular disease [3]. The method, slightly modified in 1968 [4], also lends itself

to prospective studies in which specified codable ECG abnormalities at baseline are related to the later morbidity or mortality experience of the population studied. It was used in the WHO Multinational Study of Vascular Disease in Diabetes (WHO MSVDD) at baseline and follow-up examinations, in part to assess cardiovascular morbidity.

Minnesota coding instructions are as precise as possible. In part, they rely on the careful measurements in calibrated records of voltage deflections on the x axis and time intervals on the y axis which can be as small as 0.25 mm. The presence or absence of certain codable ECG abnormalities (or their degree) depend upon such measurements being greater or less than a stated reference value. As that reference

Corresponding author: Prof. H. Keen, Unit for Metabolic Medicine, United Medical and Dental School (Guy's Campus), London Bridge, London SE1 9RT, UK

* See reference 6 for list of investigators

Abbreviations: WHO MSVDD, World Health Organization Multinational Study of Vascular Disease in Diabetes.

value is approached the risk of misclassification, though small, must increase. Deviant values are needed only in specified ECG leads to be codable and certain codes suppress others. Some of the factors contributing to the risk of error have been identified [4] and even the most careful and experienced Minnesota coder will make occasional errors, especially when tired.

Coding differences between serial electrocardiograms arise from several sources. In addition to observer variation, code changes often indicate the appearance, evolution or resolution of cardiac pathology. Differences might arise from variation in positioning of the electrodes, from changes in the position of the heart and from alterations in the metabolic milieu of the myocardium. The initial and follow-up examinations of the London cohort of the WHO MSVDD included the recording of standard resting 12-lead ECGs on both occasions and provided the opportunity to measure and analyse some of these sources of variation and to evaluate their influence on estimates of cardiovascular disease progression.

Material and methods

The 497 strong London cohort of the WHO MSVDD [5] were all submitted at baseline to a resting 12-lead ECG examination. This was repeated in 352 of the survivors at follow-up after an average interval of 8.3 years (range 6.0 to 10.1 years) [6]. All of the ECG records were read independently and within a few weeks of their acquisition by the same two trained and highly experienced Minnesota coders (N. Keen, C. Rose) who coded for the whole study. On the few occasions when their coding differed in a given record, it was adjudicated by a qualified third observer.

For this analysis (as for the WHO MSVDD) the detailed (9-column) Minnesota coding (see glossary) for each record was then assigned a summary code using the Whitehall Criteria [7], indicating the relative likelihood of coronary heart disease (CHD). Minnesota codes 1–1, 1–2 and 7–1 (major and medium Q or QS waves as defined and complete left bundle branch block) were assigned to summary code 3 (CHD probable). Minnesota codes 1–3 (minor Q/QS waves), or any of 4–1, 4–2, 4–3 if accompanied by 5–1, 5–2 or 5–3 (ST segment abnormalities accompanied by abnormal T waves) were assigned to summary code 2 (CHD possible). All other Minnesota codes and normal records (1–0) were assigned to summary code 1 (CHD unlikely).

Of the 352 patients with ECG records at baseline and follow-up, a constructed sample of 118 baseline/follow-up paired records was selected for the present re-reading analysis. This sample comprised all but one of the 55 records coded as showing possible or probable CHD at baseline, summary code 2 (43 records) and code 3 (12 records) and a sample of 64 of the 297 with summary code 1 (CHD unlikely) at baseline made up of similar numbers with follow-up summary codes 1 (19 records), 2 (21 records) and 3 (24 records) (Table 1).

Three to six months after the recording and standard initial coding of follow-up ECGs, all of the 118 paired sample records were re-read and recoded independently at a single sitting by the same two observers. This recoding operation, was carried out in ignorance of the initial results and the records were as-

Table 1. Composition of constructed paired sample

	Baseline	Total pairs	Selected pairs
CHD unlikely	code 1	297	64
CHD possible	code 2	43	42 ^a
CHD probable	code 3	12	12

^a One record in this group was mislaid

signed a new, random serial number and re-read in random order. The baseline records were re-read 8 to 10 years after their standard initial coding and the follow-up records between 3 and 12 months after their initial coding.

Results

The full cohort: change between baseline and follow-up ECG summary codes. Table 2 cross tabulates baseline summary codes against follow-up summary codes for the whole cohort of 352 paired records. Of 297 with baseline summary code 1, 41 (13.8%) worsened to summary code 2, and 26 (8.8%) to summary code 3 with 230 (77.4%) remaining unchanged. Of 43 records with summary code 2 at baseline, 9 (20.9%) worsened to summary code 3, 9 did not change but 25 (58.1%) improved to summary code 1. Of the 12 records which were summary code 3 at baseline, 6 had follow-up ECGs assigned to summary code 1 and 3 to summary code 2 with only 3 remaining in the same category. In all, at follow-up there were 76 (21.6%) with worsened summary code ratings, 34 (9.7%) with improved ratings and 242 (68.7%) with no change.

The constructed sample: change between baseline and follow-up ECG summary codes. In a similar cross-tabulation of changes between baseline and follow-up in the constructed sample (Table 3), of the 64 with baseline summary code 1, 21 worsened to code 2, and 24 to code 3. As stated above, of the 42 with summary code 2 at baseline, 8 had deteriorated to code 3, and 25 improved to code 1. Of the 12 assigned to summary code 3 at baseline, 3 improved to code 2, and 6 to code 1 at follow-up.

Re-reading of ECG records: baseline ECGs – initial and re-read summary codes. Table 4 cross-tabulates the summary codes from the initial baseline ECG reading against the re-reading codes of the same records 10 to 11 years later. Re-reading generated 21 reassignments of summary code. Of the 64 with summary code 1, 62 did not change, 1 was reassigned to code 2 and the other to code 3. Of the 42 with summary code 2, 15 (35.7%) were reassigned to code 1 and none to code 3. Of the 12 with initial summary code 3, 8 did not change on re-reading but 4 were reas-

Table 2. Initial summary codes at baseline and follow-up (whole paired cohort)

		Initial follow-up code			
		Code 1	Code 2	Code 3	Total
Initial	code 1	230	41	26	297
Baseline	code 2	25	9	9	43
Code	code 3	6	3	3	12
	Total	261	53	38	352

Table 3. Initial summary codes at baseline and follow-up (constructed sample)

		Initial follow-up code			
		code 1	code 2	code 3	Total
Initial	code 1	19	21	24	64
Baseline	code 2	25	9	8	42
Code	code 3	6	3	3	12
	Total	50	33	35	118

Table 4. Initial versus re-read summary codes for baseline records (constructed sample)

		Re-read baseline code			
		code 1	code 2	code 3	Total
Initial	code 1	62	1	1	64
Baseline	code 2	15	27	0	42
Code	code 3	4	0	8	12
	Total	81	28	9	118

Table 5. Initial versus re-read summary codes for follow-up records (constructed sample)

		Re-read follow-up code			
		code 1	code 2	code 3	Total
Initial	code 1	47	2	1	50
Follow-up	code 2	2	31	0	33
Code	code 3	2	1	32	35
	Total	51	34	33	118

signed to code 1. Thus of the 21 coding reassignments, 2 were ‘deteriorations’ and 19 were ‘improvements’.

In detail (see Glossary), the 2 ‘deteriorations’ consisted of ‘new’ 5.3 and 1.2.4 Minnesota codes. Of the 19 ‘improvements’, 14 involved the ‘loss’ of 5.2 or 5.3 codes, 3 the ‘loss’ of 1.2.2, 1 the ‘loss’ of 1.3.3 and 1 the ‘loss’ of 7.1 codes.

Changes between initial and re-read follow-up summary codes. A similar cross-tabulation of initial versus re-read (within a few months) summary codes of the follow-up ECG records is shown in Table 5. This shows much less reassignment overall, with 3 ‘deteriorations’ (2 from summary code 1 to 2 and 1 from summary code 1 to 3) and 5 ‘improvements’ (2 from summary code 2 to 1, 2 from summary code 3 to 1 and 1 from summary code 3 to 2).

In detail, 2 of the ‘deteriorations’ were based on the acquisition of a 5.2 and a 5.3 code and the third of a 1.2.8 code. Of the 5 ‘improvements’, 3 resulted from loss of a 1.1.2, a 1.2.4 and a 1.3.2, one from the reassignment of a 7.1 to 7.2 but only 1 from loss of a 5.3.

Estimates of net worsening in constructed sample. Based upon the initial readings, 54 of the 118 baseline ECGs (45.8%) and 68 of the 118 follow-up ECGs (57.6%) were summary coded as abnormal (codes 2 or 3), a net deterioration of 11.8%. Based upon the re-readings of the same ECGs, 37 of the baseline records (31.4%) and 67 of the follow-up records (56.8%) were summary coded as abnormal, an estimated net deterioration of 25.4%. The great majority of this difference is due to reassignment of code 2 to code 1 in the baseline re-reading, a change not seen in the follow-up re-reading.

Table 6. Individual record summary code consistency

1 'Improvement' [34]		
Change	<i>n</i>	BL → FU
2 → 1	25	
	14	2/2 → 1/1
	9	2/1 → 1/1
	1	2/1 → 1/2
	1	2/1 → 1/3
3 → 1	6	
	1	3/3 → 1/1
	3	3/1 → 1/1
	2	3/? → 1/1
3 → 2	3	
	2	3/3 → 2/2
	1	3/1 → 2/2
2 'Deterioration' [53]		
Change	<i>n</i>	BL → FU
1 → 2	21	
	18	1/1 → 2/2
	2	1/1 → 2/1
	1	1/1 → 2/?
1 → 3	24	
	21	1/1 → 3/3
	1	1/3 → 3/3
	1	1/2 → 3/3
	1	1/1 → 3/2
2 → 3	8	
	7	2/2 → 3/3
	1	2/1 → 3/3
3 'No change' [31]		
x → x	31	
	26	x/x → x/x
	1	1/1 → 1/2
	3	2/1 → 2/2
	1	3/3 → 3/1

Under the 'Change' and 'No change' headings (1, 2, 3 above) the first figure denotes the standard initial baseline summary code and the second figure, following the arrow, the standard initial follow-up summary code. The designations 'Improvement', 'Deterioration' and 'No Change' refer to behaviour of these summary codes over the follow-up period. The *n* column details for each category of behaviour, first the total in it, then that total broken down into consistency subcategories. BL and FU head columns in which paired initial and re-read summary codes are shown for the baseline and follow-up ECG records respectively. In the 'No change' category, x represents all cases where summary codes within pairs were identical.

Individual record consistency. Each of the paired records (baseline and follow-up) in the constructed sample had two (standard initial and re-read) readings. When the codings assigned to the two readings at baseline agreed with each other and the two readings at follow-up also agreed, the paired record was classed as 'consistent'. When there were coding disagreements within either or both pairs, the paired record was classed as 'inconsistent'. Table 6 shows the detailed consistency for each of the categories, 'improvement', 'deterioration' and 'no change' reported

from the standard initial study codings of the baseline/follow-up pairs of the sample. Applying this consistency grading, 17 of the 34 paired records (50%) showing improvement over the follow up period (Table 3) were 'inconsistent'. In contrast, only 13% of the 53 paired records showing deterioration and 16% of the 31 paired records showing no change were 'inconsistent'. Thus a substantially higher proportion of the improvements over deteriorations probably resulted from observer variation and as a consequence, the true net deterioration over the follow-up period is likely to have been underestimated.

Discussion

In this study, 352 paired ECG records from the same diabetic patients taken 8 to 10 years apart were submitted to the independent double reading for Minnesota coding as part of the routine survey procedure. A subsample of 118 of these pairs was constructed and they were re-read by the same experienced coders 6 to 12 months after the follow-up records had been reported. The detailed Minnesota codes assigned to each tracing were grouped according to the Whitehall criteria, as indicated in the Methods section, into CHD probable (summary code 3, mainly Q wave items), CHD possible (summary code 2, mainly ST/T wave items) and CHD not likely (summary code 1, all other Minnesota codes and normal records). This analysis deals mainly with changes in these summary codes used as indicators of progression or regression of coronary heart disease. When appropriate the summary codes were translated back to their primary Minnesota code components.

Based upon the routine survey codings of the 352 paired records, 76 (21.6%) showed deterioration over the follow-up period whereas 34 (9.7%) showed improvement, a net deterioration of 11.9%, comparable with the simple difference of 10.3% between the 63 abnormal records at baseline (15.6%) and the 81 (25.9%) at follow-up. In the constructed sample of 118 paired records, use of the same routine survey readings showed deterioration in 53 (44.9%) and improvement in 36 (30.5%), a net deterioration of 14.4%. The simple difference between the 54 records abnormal at baseline (45.8%) and the 68 (57.6%) at follow up was 11.8%.

The results of the re-reading analysis should be considered when these estimates of ECG deterioration in the full cohort and in the constructed sample are interpreted. When the baseline ECGs were re-read 10 to 11 years later, the proportion coded abnormal dropped from 45.8% to 31.3%, mostly from the reassignment at re-reading of 19 records from summary codes 2 and 3 to summary code 1 (Table 4). When the follow-up ECGs were re-read 6 to 12 months later, only 3 of the records (2.5%) were as-

signed to a worsened summary code and 5 (4.2%) to an improved one.

Based on the re-reading, a consequence of reduction in the baseline abnormality rate is a reciprocal increase in the estimated deterioration when it is set against the follow-up abnormality rate. This would increase the simple deterioration rate between baseline and follow-up from about 11% to about 25%. Similar conclusions can be drawn from the consistency grading applied to these observations. A much higher proportion of improvements than deteriorations between baseline and follow-up were inconsistent and the consistent paired records are more likely to be valid than the inconsistent ones.

The consistency analysis further suggests that about half of the apparent ECG improvement over the follow-up period is real. Some of this improvement could represent true biological changes in the ischaemic myocardium over time. Genuine disappearance of major Q waves and ST/T changes are seen clinically and have been documented in defined population groups [8, 9, 10, 11]. However, some change in tracings could arise from variations in placement of electrodes or in position of the heart [12, 13] and, particularly in a diabetic population, some repolarisation signals could vary with changes in metabolic status.

Coding drift and recommendations. Most of the observer variation in this study arose from records with Minnesota code items with minimal abnormalities such as minor ST segment changes or flat or diphasic T waves (5.3) [8] where coding the record as positive would assign it to summary code 2 and as negative to summary code 1. Summary code change would also arise when distinguishing a minimal Q/QS wave (1.3) from a normal record (summary code 2 or 1 respectively) or from a greater than minimal Q/QS (1.2) which would qualify for a summary code 3. Whether there was a small upward deflection (R wave) immediately preceding a suspected downward Q/QS deflection could determine whether the record is assigned to summary code 3 or 1.

Such 'limit level' judgements are inescapable and expected to contribute randomly to the variability of record coding. The high re-reading consistency of the follow-up record shows the limited extent of this variability. However, the systematic difference in the treatment of these limit level decisions between the baseline and the follow-up records suggests some subconscious 'policy' change over the ten to twelve years in question by both of the independent coders. Such operational drift could be detected only if coders were regularly to reread standard sets of ECGs selected to monitor performance in the areas of difficulty defined in this and other studies. It is also advisable for independent testing of pairs of coders to detect unconscious 'collusion' in coding drift. Computerisa-

tion of Minnesota coding has long been elusive and despite recent findings [14], its role as an alternative to the human reader in longitudinal studies has yet to be validated.

In conclusion, survivors of the London Cohort of the WHO Multinational Study of Vascular Disease in Diabetes showed clear evidence of electrocardiographic deterioration over an 8.3 year time period. A planned ECG re-reading study suggests, however, that independently of the small degree of short-term observer variation, unconscious coding drift over a decade had led to some overestimation of ECG improvements and the net deterioration which had occurred was underestimated. Some variation, even in the relatively standardised Minnesota coding procedure, is inescapable and it is most likely to occur at specifiable limit levels.

Glossary

1-1-2 Q duration ≥ 0.04 s in leads 1,2,V1-V6.

1-2-3 Q duration ≥ 0.03 s and < 0.04 s in leads 1,2,V2-V6

1-2-4 Q duration ≥ 0.04 s and < 0.05 s in lead III, plus a Q wave ≥ 1.0 mm amplitude in the majority of beats in lead AVF

1-2-8 Initial R amplitude decreasing to 2 mm or less in every beat (and absence of codes 3-2,7-1-1,7-2-1 or 7-3) between any of leads V2 and V3, V3 and V4, V4 and V5, V5 and V6. All beats immediately to the right on the chest must have an initial R ≥ 2 mm.)

1-3-2 QS pattern in lead V1 and V2 (not in presence of 3-1 or 7-1-1)

1-3-3 Q duration ≥ 0.03 s and < 0.04 s plus R amplitude ≥ 3 mm in lead aVL

5-2 T amplitude negative or diphasic (positive-negative or negative-positive type) with negative phase at least 1 mm but not as deep as 5 mm in lead I or V6 or aVL when R amplitude is ≥ 5 mm, or in lead I or aVF when QRS is mainly upright, or in any of leads V2, V3, V4 or V5

5-3 T amplitude zero (flat) or negative or diphasic (negative-positive type only) with less than 1 mm negative phase in lead I or V6 or in lead aVL when R amplitude is ≥ 5 mm or in lead II (not coded in lead aVF) or in any of leads V3, V4 or V5

7-1 Complete left bundle branch block

7-2 Complete right bundle branch block

References

1. Blackburn H, Keys A, Simonson E, Rautaharju PM, Pun-sar S (1960) The electrocardiogram in population studies. A classification system. *Circulation* 21: 1160-1175
2. Keys A (1967) Epidemiological studies related to coronary heart disease: Characteristics of men aged 4-59 in seven countries. *Acta Med Scand* [Suppl 460]

3. Reid DD, Brett GZ, Hamilton PJS, Jarrett RJ, Keen H, Rose G (1974) Cardiorespiratory disease and diabetes among middle-aged male civil servants. *Lancet* i: 469–474
4. Rose GA, Blackburn H (1968) Cardiovascular survey methods. WHO Monograph Series No 56, World Health Organization Geneva
5. Diabetes Drafting Group (1985) Prevalence of small vessel and large vessel disease in diabetic patients from 14 centres. The WHO Multinational Study of Vascular Disease in Diabetes. *Diabetologia* 28: 615–640
6. Lee ET, Keen H, Bennett PH, Fuller JH, Lu M and the WHO MSVDD Investigators (2001) Morbidity follow-up of the WHO Multinational Study of Vascular Disease in Diabetes: General description and incidence. *Diabetologia* 44 [Supp 2]: S3–S13
7. Rose G, Reid DD, Hamilton PJS, McCarney P, Keen H, Jarrett RJ (1977) Myocardial ischaemia, risk factors and death from coronary heart disease. *Lancet* ii: 105–109
8. Ostrander LD (1966) Serial electrocardiographic findings in a prospective epidemiological study. *Circulation* 34: 1069–1080
9. Pyorala K, Kentala E (1974) Disappearance of Minnesota code Q-QS patterns in the first year after myocardial infarction. *Annals Clin Res* 6: 137–141
10. Karnegis JN, Matts J, Tuna N, the POSCH Group (1985) Development and evolution of the electrocardiographic Minnesota Q-QS codes in patients with acute myocardial infarction. *Am Heart J* 110: 452–459
11. Rissanen V, Raunio H, Halinen MO, Rehnberg S (1986) Changes occurring in the postinfarction ECG in relation to age, sex, and previous myocardial infarction. *Am Heart J* 111: 286–292
12. Elgrishi I, Ducimetiere P, Richard JL (1970) Reproducibility of analysis of the electrocardiogram in epidemiology using the ‘Minnesota code’. *Brit J Prev Soc Med* 24: 197–200
13. Rautaharju PM (1978) Use and abuse of electrocardiographic classification systems in epidemiologic studies. *Europ J Cardiol* 8/2: 155–171
14. Kors JA, vanHerpen G, Wu J, Zhang Z, Prineas RJ, vanBommel JH (1996) Validation of a new computer program for Minnesota coding. *J Electrocardiol* 29: 83–88