UNIVERSITY OF Minnesota

SCHOOL OF PUBLIC HEALTH • LABORATORY OF PHYSIOLOGICAL HYGIENE STADIUM GATE 27 • MINNEAPOLIS, MINNESOTA 55455

February 7, 1971

MEMO TO: C. Meinert, C. Klimt, W. Krol, Y. Ohno, R. MacGregor, S. Tominaga, J. Stamler, P. Canner, T. Landau

FROM: Henry Blackburn

SUBJECT: Comparisons of test-retest results with different ECG coding procedures.

We constantly seek to improve the quality of ECG readings which, from the nature of the beast, is an important problem. After the Coordinating Center reported results of its external quality control, we tried to think of procedural changes which would 1) improve reader agreement and reduce "technical" error 2) while not affecting "validity" or introduce a secular change in the reported frequency of findings.

One of the questions in any ECG coding system, manual or computerized, is the appropriate selection of heart beats to be measured and classified. The Minnesota procedure follows the rules, with a few specified exceptions, that 1) the majority of recorded beats in a given ECG lead must meet or exceed the criterion, and 2) when there is doubt about choice between subclasses, code to the "lesser" class. Most computer programs, in contrast, select a unique beat, one which meets certain criteria for "representativeness," and measures the detail in that beat. We examined an alternative rule, selecting the next to last full sinus beat of the record.

The accompanying table gives results for three separate series of paired readings on three different batches of CDP scheduled ECGs, by three different pairs of readers.

The first consideration is the percentage-disagreement, defined by the formula in which  $t_1$  is the total positive cases of one observer,  $t_2$  the same of the second observer, a is the number of agreed negative cases and T is the total number of records examined.

Pair I comparisons suggest little or no improvement due to coding by the unique beat system. For the two items where there is a particularly serious problem of disagreement, ST and P-R interval codes, there is clearly no substantial improvement in reader agreement.

Pair II and III comparisons showed a higher order of agreement generally, in other batches of records and with different observers. However, the same ECG items proved unreliable, ST and PR codes. There was no evidence that selection of a given heart beat produced an improvement or deterioration in observer variation over the majority of beats rule.

The next columns of the table compare the frequency of agreed positive codes, which might be similar to the agreed frequency of items officially reported by Baltimore. There is no evidence of a significant systematic difference in sensitivity of the one-beat versus majority-beat procedures.

Another procedure was used in the Pair III column to examine whether the A and B coding rules produced systematic reading differences. Still another CDP batch was independently read twice by the same reader, using the A and the B coding procedure. No important systematic difference in the frequency of reported ECG items was observed between the majority beat or the single beat procedure.

On the basis of these paired tests, it appears that we will have to look for other means of improving reliability of certain ECG codes or accept them as they are. Moreover, due to the relatively low frequency and the very poor repeatability of P-R interval codes this finding should be considered unreliable.

The absence of important systematic differences in frequency of items between the two methods suggests that it is not necessary to re-read, by the routine procedure, the single batch (27) of CDP records inadvertently read by the trial method in November-December 1970.

For the moment, no change in CDP reading procedure is counseled, and the routine majority beat rule should be in force.

Dictated from Geneva

HB/rs