Methods of Information in Medicine

J. L. Willems[1], P. Arnaud[2], J. H. van Bemmel[3], R. Degani[4], P. W. Macfarlane[5], Chr. Zywietz[6] for the CSE Working Party

Common Standards for Quantitative Electrocardiography: Goals and Main Results

[1]CSE Coordinating Center, Division of Medical Informatics, University of Leuven, Belgium
[2]INSERM Unité 121, Hôpital Cardio-Vasculaire, Lyon, France
[3]Department of Medical Informatics, Erasmus University Rotterdam, The Netherlands
[4]Istituto per Ricerche di Dinamica dei Sistemi e Bioingegneria, LADSEB-CNR, Padova, Italy
[5]Department of Medical Cardiology, Glasgow, Scotland
[6]Abteilung für Biosignalverarbeitung, Medizinische Hochschule, Hannover, Federal Republic of Germany

Abstract: Computer processing of electrocardiograms (ECGs) has over the last 15 years increased rapidly. Still, there are at present no standards for computer ECG interpretation. Different techniques are used not only for measurement and interpretation, but also for transmission and storage of data. In order to fill these gaps, a large international project, sponsored by the European Commission, was launched in 1980 to develop "Common Standards for Quantitative Electrocardiography (CSE)". The main objective of the first CSE study was to reduce the wide variation in wave measurements currently obtained by ECG computer programs. The second study was started in 1985 and aimed at the assessment and improvement of diagnostic classification of ECG interpretation programs. To this end reference libraries of well documented ECGs have been developed and comprehensive reviewing schemes devised for the visual and computer analysis of ECGs. This task was performed by a board of cardiologists in a Delphi review process, and by 9 VCG and 10 standard 12-lead programs developed by university research groups and by industry. A third action was started in June 1989 to harmonize acquisition, encoding, interchange and storing of digital ECG data. The action thus performed have become internationally recognized milestones for the standardization of quantitative electrocardiography.

Key-Words: Automated ECG Analysis, Wave Recognition, Diagnostic Classification, Standardization, Quality Assurance Testing.

1. Introduction

Although the first attempts to automate ECG analysis by digital computer were made as early as 1957 by Pipberger and coworkers [1], it took considerably more time to develop operational computer programs than originally anticipated. However, over the last 15 years computer processing of ECGs has increased rapidly [2–5]. The first ECG computer programs were mainly developed by university research groups. In the last decade development has shifted to industry. Until recently, there were no standards in this field. There were no common definitions of waves, no standards for measurement or diagnostic classification, and no uniform terminology for reporting. Even more, data acquired by one ECG system can up to today, due to lack of standards not be transmitted and processed by another ECG computer system in routine practice. This has created a situation whereby large differences in measurement results by different computer programs hamper the exchange of diagnostic criteria and interpretation results [6, 7]. In addition, more and more microcomputer-based interpretative ECG machines are being put on the market without any prior independent validation.

In order to overcome some of these problems, a concerted action was started in the European Community (EC) in June 1980, striving towards "Common Standards for Quantitative Electrocardiography" (abbreviated: CSE) [8, 9]. The present paper will briefly summarize the objectives, current results and future actions of this project. In other papers presented in this special issue devoted to the CSE project, more detailed information on this action will be given.

A plea for such action had been made at several international conferences organized under the aegis of the International Medical Informatics Association (IMIA) [2–5] and the American College of Cardiology [10].

2. Objectives and General Scope of the Actions

ECG computer processing can basically be reduced to three principal stages:

1. Acquisition, transmission and storage of digital ECG data,

2. Pattern recognition and measurement,
3. Diagnostic classification.

In each of these stages there are important needs for standardization. At the start of the CSE project the following objectives were formulated:

1. Standardization of ECG measurement procedures in quantitative terms; comparative studies of measurements performed by different programs; drawing of guidelines, definitions and standards for measurement.
2. Assessment of the performance of diagnostic classification of computer programs and algorithmic documentation of their operation.
3. Establishment of modest ECG libraries to reach these goals [8, 9].

Over the past ten years a steady course has been followed to achieve these goals. From the beginning the project was divided into two logically related studies, dealing first with measurements and next with diagnostic interpretation. The order of these studies has been determined from the inception of the project by the logic that diagnostic criteria can only be exchanged if programs provide the same basic measurement results when analyzing identical ECG recordings. Investigators from 25 institutes in the European Community are participating in the project. They constitute the CSE Working Party. Also investigators from six North-American and one Japanese center collaborate, either by processing data or as consultants (see Appendix). These studies have been performed as a concerted action within the frame of the 2nd, 3rd and 4th Medical and Public Health Research Programmes of Directorate General (DG) XII of the European Commission [11]. As of June 1989 a third action was started, aiming at standards for digital ECG data transmission, encoding and storing, an essential requirement to interconnect different systems. This project is carried out under the umbrella and with support of the Advanced Informatics in Medicine Programme (AIM) and the Conformance Testing Services Programme (CTS) of DG XIII of the European Commission.

3. Standardized ECG Measurement

3.1. Main Objectives of the First CSE Study

The main objective of the first part of the CSE project was to reduce the variation of measurements by different ECG computer programs. The aim was to standardize computer-derived ECG measurements, obtaining agreement on definitions of waves and of references for the onsets and offsets of P, QRS and T waves. The intention was, when an ECG of a patient is recorded for instance in Rome with program X, that the same interpretation results will be printed as when his ECG would have been recorded and analyzed in London with program Y. The means and variances of various programs analyzing a common database should for that reason fall within acceptable ranges. Only then can diagnostic criteria for myocardial infarction and other cardiac diseases be exchanged and possibly standardized.

3.2. CSE Measurement Reference Libraries

One of the major results of the first CSE study, which has now been completed, was the establishment of a reference database for ECG measurement. In fact, two reference databases have been established in the first CSE study. The first database consists of 250 original and 310 so-called artificial ECGs in which four groups of three leads were recorded simultaneously, as was common practice until recently [12]. Because different ECG measurement programs apply various principles with respect to the analysis (for example, some programs measure single beats, whereas other ones analyze averaged beats), artificial ECGs have been constructed by selecting one beat from each lead-group of the original recordings and making strings of identical beats. The second database [13] involves 250 original and 250 artificial ECGs in which not groups of three, but all leads – the 12 standard leads (12-SL) plus the Frank XYZ leads – were recorded simultaneously, as is being done more and more in the newest microcomputer-based electrocardiographs. The development of these CSE libraries has been described extensively in eight annual CSE Progress Reports and in various publications [12–18].

The libraries represent a wide variety of ECG configurations; 25% were normal, the remainder abnormal. All ECGs are sampled at 500 Hz, with a resolution of at least 10 bits and a maximal quantization step of 5 microvolts. The original and corresponding artificial ECGs are randomly divided into two sets, containing nearly equal samples of the various pathologic entities. The results of one set have been released; the results of the other set are kept secret in the CSE Coordinating Center for independent testing.

3.3. Visual ECG Analysis

A group of five cardiologists, each from a different Member State of the EC, has performed a detailed determination of the onsets and offsets of the various ECG waves (P, QRS and T), on very enlarged signals (see Fig. 1). In view of the well-known inter- and intra-observer variability in determining wave-recognition points, an elaborate reviewing scheme was devised to establish the CSE 3-lead reference library [12, 14]. By using a modified Delphi approach, individual referee "outliers" (i.e., point estimates which differ considerably from the median referee result) were eliminated in four successive steps, as illustrated in Fig. 2. In this way a database was established with well-defined wave reference points. Reproducibility tests proved that the final estimates resulting from this group analysis, are very stable [12].

3.4. Analysis by Different Programs

The same digital ECG recordings were also analyzed by 9 different vectorcardiographic (VCG) and 10 different standard 12-lead ECG programs, from Europe, the US, Canada, and Japan, including those of major industrial companies [14]. Various graphical and statistical analysis pro-

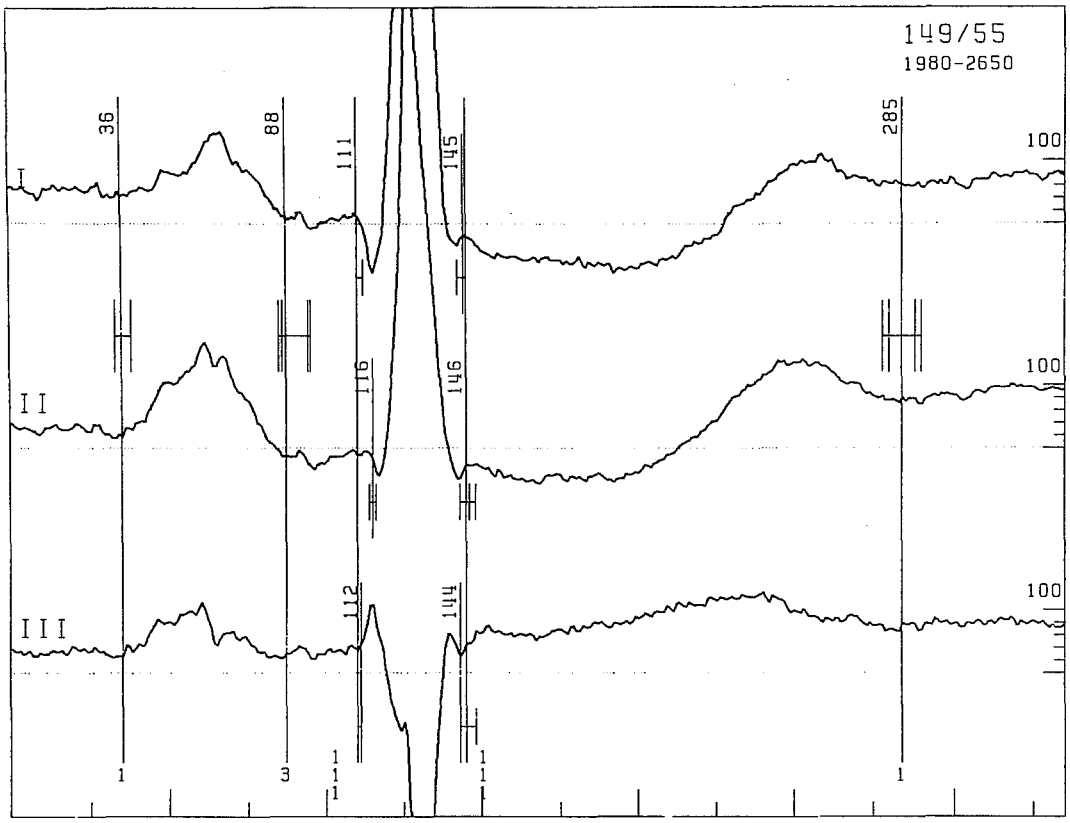

Fig. 1 Example of wave-onset and offset determination by the referees. The long vertical lines denote the median, the short ones the individual estimates of the five referees (note that individual referees estimates may overlap). The adjacent numbers denote time locations (in sample points) with reference to the beat onset. The figures at the bottom of the vertical lines indicate the final reviewing round in which the measurements were obtained.

grams have been developed in the Coordinating Center for program-to-program and referee-to-program comparison. Figure 3 depicts average differences and variances of different individual programs against the final referee results. The reference is indicated by long vertical lines and the variance of the computer results by horizontal bars. From this analysis it is obvious that some programs deviate systematically from the reference, and some have a significantly higher variance for certain measurements, which means that their algorithms can be further improved. However, the median wave recognition results of the 9 VCG and the 10 standard 12-lead ECG analysis programs, were almost identical to the final visual estimates obtained by the referees. The median of the program results presented a significantly lower variance than did individual program estimates. This median could therefore be used as a substitute in the determination of the multilead CSE reference library [13]. In general, the performance of XYZ and the so-called 12-simultaneous-lead (12-SL) programs, which record all standard leads simultaneously, was better than that of conventional 3-simultaneous-lead (3-SL) programs, which nowadays are increasingly less available on the market.

3.5. Comparison of Derived Interval Measurements

As could be expected from the differences in the basic P, QRS and T wave onsets and offsets, some programs demonstrated significantly different P, and QRS durations as well as PR and QT intervals, as compared to others. The duration of the P wave and QT interval derived by the XYZ programs was significantly shorter than that computed by the standard 12-lead programs [13]. In contrast, the QRS duration derived from both lead systems was on the average nearly identical, although some individual results were widely divergent. Among the different standard 12-lead programs, the results were also divergent. Some multilead programs showed statistically shorter interval measurements and other ones longer intervals than the median. This may be explained by the use of different thresholds or template-matching algorithms, the application of correction procedures, or even definition problems. Indeed, as demonstrated in Table 1, five programs reported Q-wave durations that were on the average 6 ms longer in all leads than the median duration, because these programs include initial iso-electric segments in the Q or QS wave, when present, whereas others did not. Such differences may lead to significantly different diagnostic results when the same criteria are used in different programs, for example for myocardial infarction. The inclusion of terminal iso-electric segments in the R wave also explains, at least partially, the difference in R wave and QRS durations between various programs and hence also differences in diagnostic statements on ventricular conduction defects.

3.6. CSE Measurement Standard

The CSE measurement reference libraries, have become an international standard for the evaluation and

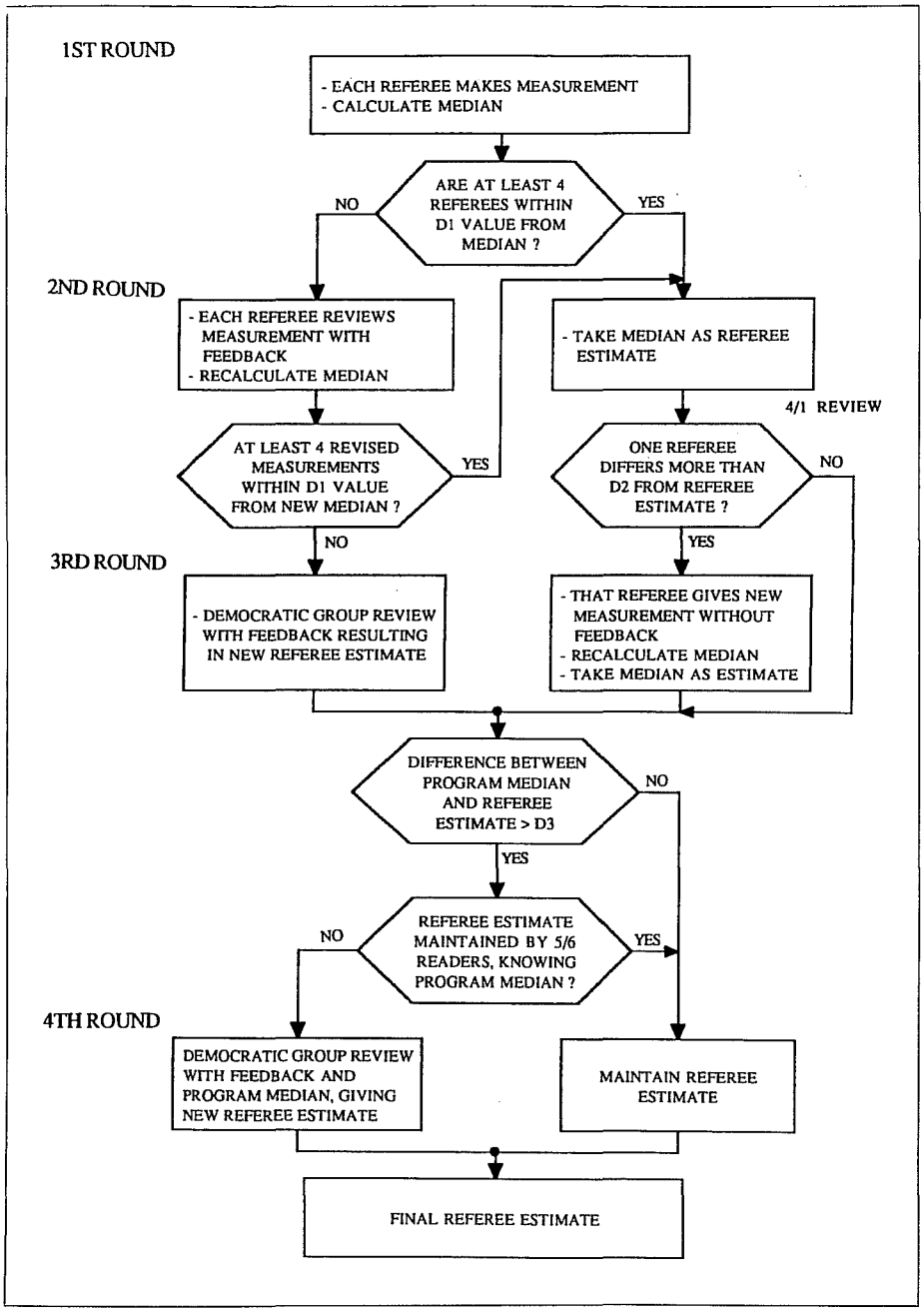

Fig. 2 Summary of the different reviewing rounds and the final determination by the referees of P and QRS-onsets and offsets and for T-end. The limits D1 to D3, used to estimate deviations of individual versus median referee results, were respectively as follows: D1: 10, 10, 6, 10 and 26 ms; D2: 12, 14, 8, 14 and 36 ms; D3: 12, 12, 6, 10 and 28 ms.

improvement of ECG and VCG measurement programs. The libraries have also been a useful instrument in the establishment of recommendations for more precise measurement rules and definitions [15]. Also noise testing has been performed to study results under degraded operational conditions [16, 17]. As a result of the CSE study, modifications have been made in the wave recognition and measurement programs in several participating institutes. The CSE databases have also been used by other institutes inside and outside Europe, including Eastern Europe, to improve the accuracy of existing programs or to develop new algorithms.

As outlined elsewhere [2–5, 15, 18] several algorithms may lead to similar solutions in pattern recognition. For that reason the CSE Working Party does not wish to propose any formal algorithm as the standard for ECG wave recognition. However, the Working Party recommends the CSE reference data-bases as standards for 3-SL and 12-SL ECG measurement programs. In summary, the meaning of the CSE standard is twofold:

1. individual program measurements should approach the reference (i.e., the median of the referees) as closely as possible; the mean difference should be close to zero, and
2. the standard deviation of the differences of a program's results with respect to the reference should not exceed certain limits. These limits have been published in a paper with CSE recommendations on ECG analysis [15].

4. Assessing Diagostic Program Performance

4.1. Main Objectives and Stages of the Second CSE Study

The second part of the CSE project aims at the "Assessment of the diagnostic performance of ECG computer programs". The study was started in 1985 and will be finalized by the end of 1990 [19–21]. The diagnostic CSE study aims essentially at consumer protection, both of the patient and of the medical community. Ultimately, a quality label involving minimum performance requirements, should be given to ECG computer programs [22]. This part of the CSE project will reveal guidelines of how to give such quality labels to ECG interpretation systems. To this end a library of ECGs is being collected from patients in which the clinical diagnosis is well documented by ECG-independent means, such as cardiac catheterization, coronary arteriography, echocardiography, and other techniques. Furthermore, statistical methods are being developed to compare various results provided both by computer and visual ECG interpretation [23–25].

The Working Party agreed in 1985 on a common protocol, details of which have been described previously [19]. This protocol has been used to

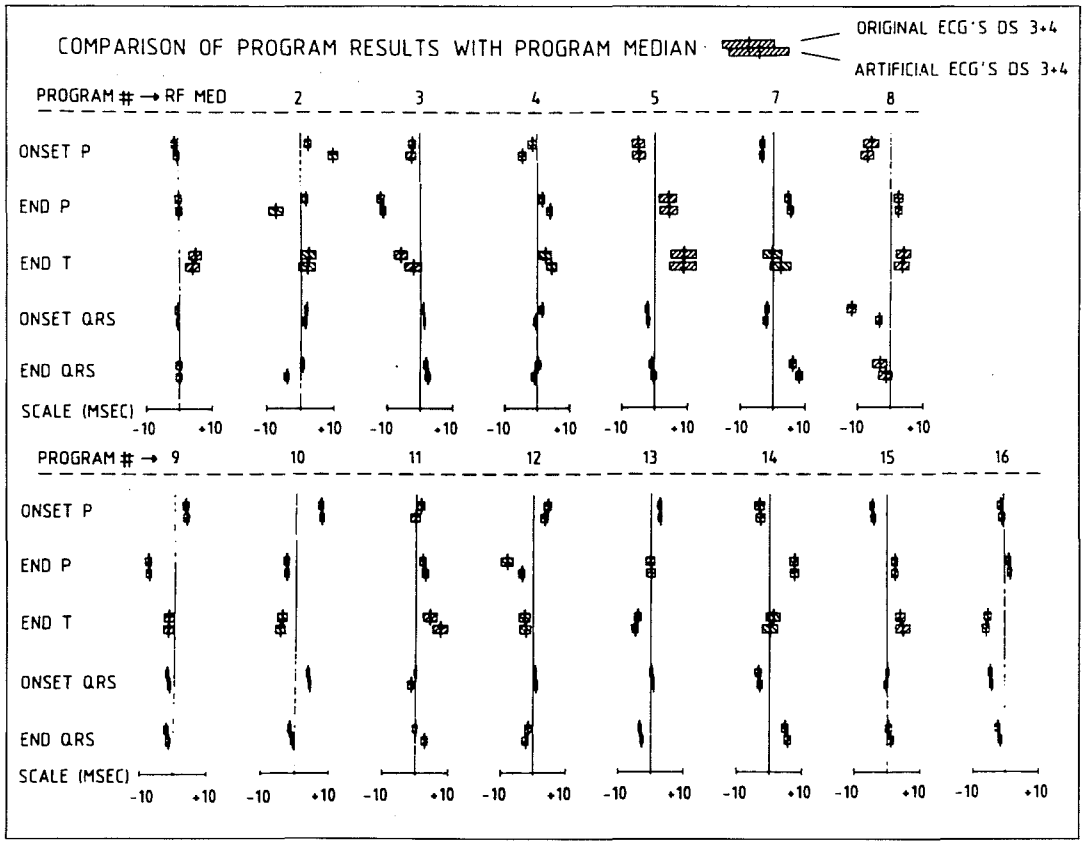

Fig. 3 Differences between (i) individual program results numbered 2 to 16, and median referee results (RF MED), and (ii) median program results in the multilead CSE Library (Data Sets 3 and 4).

perform a pilot study, based on a first set of 250 clinically validated cases [19]. The diagnostic CSE database was expanded during 1987–1988 to 500 cases and by the end of 1989 to over 1,000 cases, as initially planned. An interim analysis was performed on the first 500 cases. Results of this preliminary analysis are reported elsewhere in this issue [26]. Results of the final analysis of the whole diagnostic CSE database will be reported at the end of 1990. The CSE diagnostic database consists of well-validated multilead recordings comprising seven diagnostic groups, i.e., normal (NL) or no structural abnormalities (NSA); left, right and biventricular hypertrophy (LVH, RVH and BVH); and anterior, inferior and combined myocardial infarction (AMI, IMI and MIX). Also a limited number of cases with both myocardial infarction and ventricular hypertrophy have been included.

Table 1 Mean differences and corresponding standard deviations (in ms) between individual programs and median Q, R and S duration results*.

CSE Program Number	Q duration (N = 2,420)**		R duration (N = 5,619)**		S duration (N = 4,176)**	
	Mean	SD	Mean	SD	Mean	SD
2	5.7	11.2	2.4	10.1	1.9	9.6
4	–0.1	6.7	–1.3	7.0	1.7	7.1
5	0.5	11.2	1.4	12.4	–2.1	8.7
7	–1.1	12.1	3.9	12.5	9.5	13.4
8	6.8	16.6	2.5	15.2	–3.5	13.3
11	–1.2	7.4	–2.6	7.9	0.3	7.2
12	4.5	13.4	2.3	15.2	6.3	14.0
13	0.8	8.5	–1.7	7.7	0.4	6.9
14	5.8	11.8	2.6	10.4	2.0	8.1
15	–0.1	9.3	0.7	10.4	1.4	10.0
16	5.7	14.9	2.7	12.3	0.4	12.7

* The differences (Prog-Median) were only obtained from cases where a result was present for both the individual program and the median. The results are cumulated over all standard leads and all recordings (i.e., the original and artificial electrocardiogram) (N = 492 ECGs times 12 leads).

** N = Number of Q, R and S waves in the median of all 12-lead programs.

4.2. Material and Methods Used in the Diagnostic CSE Study

The normals in the CSE diagnostic database are free of significant cardiopulmonary disease on the basis of a health screening examination (in 75% of the cases) or invasive cardiac studies (in 25% of the cases). The diagnosis of LVH, RVH, or BVH was based on cardiac catheterization and/or echocardiographic findings. Akinesia or dyskinesia in seven different segments of the ventriculogram, coded according to American Heart Association rules, was used as the main selection criterion for the infarct classes [26]. Also some cases with a typical history and enzyme changes of acute myocardial infarction were included.

The ECGs were recorded on digital tape, 15 leads simultaneously, i. e.,

the standard 12 leads plus leads X, Y and Z, with a sampling rate of 500 Hz. The data were collected in five centers (Dublin, Glasgow, Leiden, Leuven and Louvain). A review board, consisting of three investigators, has checked the clinical information for all cases. ECGs with major ventricular conduction defects, i.e., complete left or right bundle branch block and WPW were excluded.

The ECGs have been analyzed by 15 different computer programs (see Table 2) and by seven cardiologists coming from six different EC Member States. Five of the referees have also interpreted in an independent way the VCGs (i.e., the vector-loops and the scalar XYZ leads). Nine of the programs used the standard 12-lead ECG, six the VCG. The programs included in this study were: the AVA, Marquette, Louvain, Hannover, HP, Medis-IBM, Nagoya, Lyon, Glasgow, Padova, Porto, MEANS, and the Leuven program. Descriptions of these programs have been published on various occasions [2–5]. The Hannover and the MEANS programs have separate packages for the interpretation of both the ECG and the VCG (see Table 2). The AVA results have been derived using clinical prior probabilities, as required by the program developers. Except for age and sex, no prior clinical information was provided to the other processing centers or to the referees. When a program listed minor abnormalities without making reference to any of the seven primary categories listed above, the case was classified in the NSA or "so-called normal" group.

All individual program and referee results have been compared with the true disease category. Different misclassification matrices have been computed, firstly in various bi-group analysis, i.e., normal vs abnormal, LVH vs non-LVH, etc, and next in multigroup analysis (see [26]). Based on these results sensitivity, specificity, positive and negative predictive values, and total and partial accuracy figures, have been determined for each program and each reader, as well as for the combined program and referee results. In addition, the information content and other measures of performance, such as ROC curves, plots of predictive values versus incremental prevalence figures, and McNemar statistics [23–25], were computed for each diagnostic category and each program individually [26].

Table 2 Participating processing centers in CSE diagnostic study.

		Principal investigator	ECG leads	
European ECG Computer Programs				
B	Louvain	Brohet	XYZ	–
FRG	HES	Zywietz	XYZ	12 leads
FRG	Medis	Pöppl	–	12 leads
F	Lyon	Arnaud, Rubel	XYZ	–
UK	Glasgow	Macfarlane	–	12 leads
I	Padova	Degani	–	12 leads
NL	MEANS	van Bemmel	XYZ	12 leads
B	Leuven	Willems	–	12 leads
Port	Porto	Abreu-Lima	XYZ	–
Non-European Programs				
USA	AVA	Pipberger	XYZ	–
USA	IBM	Bonner, Pöppl	–	12 leads
USA	HP	Monroe	–	12 leads
USA	Marquette	Rowlandson	–	12 leads
Japan	Nagoya	Okajima, Ohsawa	–	12 leads

Table 3 Total and partial accuracy obtained by the different cardiologists and computer programs on the first 500 cases of the diagnostic CSE database*.

Cardiologist Number		Accuracy Total	Partial	Program Number	Accuracy Total	Partial
ECG	1	68.8	79.5	A	71.3	77.9
	2	76.7	86.4	B	64.1	71.8
	3	75.6	82.6	C	74.4	84.4
	4	74.3	81.9	D	70.6	76.4
	5	72.3	79.7	E	69.1	80.7
	6	73.4	80.8	F	69.9	77.2
	7	72.3	78.3	G	67.0	74.6
	COMB	77.7	85.6	H	67.0	70.0
				I	72.4	83.4
				J	75.1	83.2
VCG	1	63.6	73.8	K	62.2	70.1
	2	71.5	79.6	L	64.7	72.2
	3	65.1	73.2	M	62.3	70.0
	4	67.6	80.4	N	68.8	76.2
	5	71.1	81.0	O	76.3	85.5
	COMB	71.3	80.9	COMB	76.6	84.9

* Results obtained without additional testing for BVH or MIX in the coordinating center.
COMB: Combined results of the cardiologists from the ECG and VCG readings respectively (See Methods)

4.3. Preliminary Results of the CSE Diagnostic Study

Different tables have been prepared comparing program and referee results to the clinical truth for the first 500 cases of the diagnostic CSE database. Total and partial accuracy figures derived from the full 7 × 7 classification matrices are presented in Table 3, both for the programs and the referees. Other results are reported in detail elsewhere in this issue [26].

4.4. Discussion of Preliminary Results of the CSE Diagnostic Study

The aims of the CSE diagnostic study are to evaluate and eventually improve diagnostic ECG systems, to compare different diagnostic strategies, and to establish an ECG-independent and clinically validated library to carry out these objectives. The development of such a database cannot be performed by a single center, but needs a joint international effort. The CSE Working Party and

foreign consultants agreed that there was an urgent need for such a study, since industry has nowadays to a large extent overtaken development of ECG interpretative systems in academic centers and is placing "cheap" but unvalidated computerized ECG machines on the market. In many cases these systems lack adequate validation and could potentially pose a real health hazard in the hands of inexperienced users (e.g., general practitioners).

According to a predefined protocol a set of cases was collected from 15-lead recordings with so-called "type A" anomalies, which can be validated by ECG-independent clinical information [10]. From the results obtained so far it is obvious that the referees definitely perform better than several computer programs, at least when interpreting the standard 12-lead ECG. The VCG results of the referees, similarly as the VCG programs, were less accurate mainly due a lower specificity. An interesting finding of the diagnostic pilot study was the higher total accuracy of the combined ECG results over all individual referee results. Although less clear-cut, this was also the case for the program results [21]. These findings support the superiority of combined opinions from a group over individual interpretations.

The results obtained so far demonstrate that the performance of several diagnostic ECG and VCG computer programs can still be improved. However, it may also be noted that the best programs perform almost as good as the best of seven cardiologists in classifying seven common abnormalities.

4.5. Further Prospects

It should be underlined that the results reported in this paper are preliminary, since they are only based on part of the database and because program improvements were still being made in different processing centers. In addition, it should be underlined that several performance indices of diagnostic accuracy depend heavily on the composition of the test groups and the fraction of test cases in each group. Moreover, for some diagnostic categories (RVH and BVH) only a limited number of cases were available. Other comments and remarks on the diagnostic CSE study are reported elsewhere in this issue [26].

A standardization and performance testing task such as the one undertaken by CSE, is difficult and can only be done by means of the international collaboration of clinicians, engineers and computer specialists. By including contributions and expertise from different centers with international reputation, results will not be restricted to a special "ECG School", but will be accepted by the medical community as a whole, as well as by industry.

5. Standards for Transmission, Encoding and Storage of Digital ECG Data

5.1. Standardization Needs

In addition to the needs and objectives listed above, there is a growing need to harmonize and standardize the technical and telecommunication areas of acquisition, transmission, encoding and storage of computer-processed ECG data [20, 27]. Until recently, transmission of ECGs was mainly undertaken with analog techniques. Almost all newer electrocardiographs offer features of digital recording and interpretation. These stand-alone microcomputer-based machines can be connected to data management systems for long-term storage and serial comparison of ECGs. There is now even the possibility of storing ECG data on personal microchip cards. However, standards are also lacking in this information technology (IT) area.

Indeed, various manufacturers use different techniques not only for measurement and interpretation, but also for transmission and storage of data. As a result, equipment cannot be linked and monopoly situations loom in and over hospital and regional authorities, buying equipment and trying to link these electrocardiographs with departmental ECG management systems equipped with mass storage facilities. It is in the general public interest if users are not restricted in their options by incompatible technical features and services of different systems and manufacturers. Digital acquisition and transmission of data offers the possibility of increased fidelity and flexibility, but at the same time it requires the development of efficient encoding, compression and even encryption schemes.

5.2. Action Plan and Prospects

To meet these objectives, a proposal was drafted for the "Conformance Testing Services" (CTS) and the "Advanced Medical Informatics" (AIM) Programmes of Directorate General XIII of the European Community (EC). The CTS project aims to establish standards and to set up testing services in at least two laboratories in two different EC Member States. These centers should be capable of implementing, executing and, if necessary, further refining IEC specifications with regard to performance and safety testing of computer-based electrocardiographs, along the various technical steps from signal acquisition to parameter measurement. This is an area in the IT field, which is in line with other activities supported by CTS and the European Standardization Bodies CEN/CENELEC [27].

In the CTS and the complementary AIM project, which both started on June 1, 1989, an investigation will be made of the complete chain of acquisition, transmission and storing of signals and measurement data to determine at each step:

- the standards and specifications which can be applied;
- the tests which have to be executed according to existing IEC recommendations as well as eventually supplementary tests;
- the work to be done to move from this stage to an operational testing service.

The AIM project specifically aims at the development of a "Standard Communications Protocol for Computerized Electrocardiography". There are three interrelated main workpackages in this project. The first workpackage aims at the development of a universal protocol for digital ECG data transmission. The second focuses

on finding an optimal data compression and coding scheme, and the main objective of the third workpackage is to develop a conceptual reference model for digital ECG data storage.

IEC recommendations with respect to electrocardiographic devices data back to 1977 and definitely need to be updated. ISO or CEN/CENELEC standards will be followed with respect to the IT specific issues. Specifications will be drafted on the basis of consensus of all interested parties, i.e., academia, industry, and professional standardization bodies.

6. Conclusion

Medical technology is developing at an increasingly greater speed. Progress in medical informatics and computing will in the future depend more and more on the adoption of standards for interpreting and communicating medical data, signals and images. The basic aim of the standards already developed in CSE or planned for the future, are to increase the availability and compatibility of objective decision support by means of Quantitative Electrocardiography throughout the European Community and the world at large. The importance of the development of quantitative test procedures and reference libraries for assessing the precision and accuracy of ECG computer programs has been emphasized on several occasions, as early as 1971 [2, 10]. However, prior to the CSE study no one had proceeded with some action in this direction. Performance evaluation tests are needed for consumer protection, both for the patient and the medical community. Ultimately, a quality label or conformity certificate of minimum performance requirements should be given to all medical computer programs, before they are put on the market. In CSE, the groundwork has been performed for this to happen in the near future.

Acknowledgments

The research presented in this paper has been supported by the European Commission within the frame of its 2nd, 3rd and 4th Medical and Public Health Research Programmes of DG XII, as well as by the AIM and CTS programmes of DG XIII and by various funding agencies in ten EC Member States.

This paper has partially been presented at the Symposium of the New York Academy of Sciences on "Electrocardiography: Past and Future", held in Nice, September 7–9, 1989.

The authors are grateful to Ms. I. Tassens for secretarial assistance and to Mr. I. Schoolmeesters for computer programming help.

REFERENCES

[1] Pipberger HV, Arms RJ, Stallman FW. Automatic screening of normal and abnormal electrocardiograms by means of a digital electronic computer. Proc Soc Exp Biol Med 1961; *106:* 130–2.

[2] Zywietz Chr, Schneider B, eds. *Computer Application in ECG and VCG Analysis.* Amsterdam: North Holland Publ Comp, 1973.

[3] van Bemmel JH, Willems JL, eds. *Trends in Computer Processed Electrocardiograms.* Amsterdam: North Holland Publ Comp, 1977.

[4] Wolf HK, Macfarlane PW, eds. *Optimization of Computer ECG Processing.* Amsterdam: North Holland Publ Comp, 1980.

[5] Willems JL, van Bemmel JH, Zywietz Chr, eds. *Computer ECG Analysis: Towards Standardization.* Amsterdam: North Holland Publ Comp, 1985.

[6] Willems JL. A plea for common standards in computer-aided ECG analysis. Comp Biomed Res 1980; *13:* 120–31.

[7] Willems JL, Pardaens J. Differences in measurement results obtained by four different ECG computer programs. In: Ostrow HG, Ripley KL, eds. *Proceedings of Computers in Cardiology 1977.* Long Beach, California: IEEE Computer Society, 1977; 115–21.

[8] Official Journal of the European Communities. 1980; *L78:* 4–28.

[9] Willems JL, Arnaud P, Degani R, Macfarlane PW, van Bemmel JH, Zywietz Chr. *Protocol for the Concerted Action Project "Common Standards for Quantitative Electrocardiography".* 2nd R & D programme in the field of Medical and Public Health Research of the EEC (80/344/EEC), CSE Ref. 80-06-00. Leuven: Acco Publ, 1980.

[10] Rautaharju PM, Ariet M, Pryor TA, Arzbaecher RC, Bailey JJ, Bonner R, et al. Task Force III: Computers in diagnostic electrocardiology. Am J Cardiol 1978; *41:* 158–70.

[11] Willems JL. Assessment of the concerted action method in European Medical Research: A case study. In: Brown BH, van Noort R, eds. *Biomedical Engineering: A Strategy for the Future,* Proceeding of a Workshop, help in Lisbon, Portugal, October 26–29th, 1988. Sheffield: University Press, 1989; 115–33.

[12] Willems JL, Arnaud P, van Bemmel JH, Bourdillon PJ, Degani R, Denis B, Harms FMA, Macfarlane PW, Mazzocca G, Meyer J, Ritsema van Eck HJ, Robles de Medina EO, Zywietz Chr. Establishment of a reference library for evaluating computer ECG measurement programs. Comp Biomed Res 1985; *18:* 439–57.

[13] Willems JL, Arnaud P, van Bemmel JH, Bourdillon PJ, Degani R, Denis B, Graham I, Harms FMA, Macfarlane PW, Mazzocca G, Meyer J, Zywietz Chr. A reference database for multilead electrocardiographic computer measurement programs. JACC 1987; *6:* 1313–21.

[14] Willems JL, Arnaud P, van Bemmel JH, Bourdillon PJ, Brohet C, Dalla Volta S, Damgaard Andersen J, Degani R, Denis B, Demeester M, Dudeck J, Harms FMA, Macfarlane PW, Mazzocca G, Meyer J, Michaelis J, Pardaens J, Pöppl SJ, Reardon BC, Ritsema van Eck HJ, Robles de Medina EO, Rubel P, Talmon JL, Zywietz Chr. Assessment of the performance of electrocardiographic computer programs with the use of a reference database. Circulation 1985; *71:* 523–34.

[15] The CSE Working Party. Recommendations for measurement standards in quantitative electrocardiography. Eur Heart J 1985; *6:* 815–25.

[16] Willems JL, Zywietz Chr, Arnaud P, van Bemmel JH, Degani R, Macfarlane PW. Influence of noise on wave boundary recognition by ECG measurement programs. Comp Biomed Res 1987; *20:* 543–62.

[17] Zywietz Chr, Willems JL, Arnaud P, van Bemmel JH, Degani R, Macfarlane PW. Stability of computer ECG measurements in the presence of noise. Comp Biomed Res.

[18] Willems JL (CSE Project Leader). *Annual CSE Progress Reports.* Leuven: Acco Publ 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989.

[19] Willems JL, Abreu-Lima C, Arnaud P, van Bemmel JH, Brohet C, Degani R, Denis B, Graham I, Van Herpen G, Macfarlane PW, Michaelis J, Moulopoulos S, Pöppl S, Zywietz Chr. Testing the performance of ECG computer programs. The CSE diagnostic pilot study. J Electrocardiol 1987; *20:* S298–301.

[20] Willems JL. *8th CSE Progress Report, Common Standards for Quantitative Electrocardiography.* Leuven: ACCO Publ, 1988; 1–284.

[21] Willems JL, Abreu-Lima C, Arnaud P, van Bemmel JH, Brohet C, Degani R, Denis B, Graham I, van Herpen G, Macfarlane PW, Michaelis J, Moulopoulos S, Pöppl S, Zywietz Chr, for the CSE Working Party. Effect of combining electrocardiographic interpretation results on diagnostic accuracy. Eur Heart J 1988; *10:* 1348–55.

[22] Willems JL. A review of computer ECG analysis: time to evaluate and standardize.

CRC Critical Reviews in Medical Informatics 1986; *1:* 165–207.
[23] Bailey JJ, Campbell G, Horton MR, Shrager RI, Willems JL. Determination of statistically significant differences in the performance of ECG diagnostic algorithm: an improved method. J Electrocardiol 1988; *21:* S188–S192.
[24] Michaelis J, Willems JL. Performance evaluation of diagnostic ECG programs. In: Ripley KL, ed. *Computers in Cardiology 1987.* Long Beach Cal: IEEE Computer Society, 1988; 25–30.
[25] Willems JL, Campbell G, Bailey JJ. Progress on the CSE diagnostic study. Application of McNemar's test revisited. J Electrocardiol. 1989; *22:* S136–141.
[26] Willems JL, Abreu-Lima C, Arnaud P, Brohet C, Denis B, Gehring J, Graham I, van Herpen G, Machado H, Michaelis J, Moulopoulos S. Evaluation of ECG interpretation results obtained by computer and cardiologists. Meth Inform Med 1990; *29:* 308–316.
[27] Willems JL. An example of standardization need: Common Standards for Quantitative Electrocardiography. *Proceedings of the CEN/CENELEC Workshop on European Standardization of Medical Devices,* held in Brussels, December 12–14, 1988. Publication of CEN/CENELEC. The Joint European Standards Institution, Brussels, 1989; 341–50.
[28] Willems JL. Computerized electrocardiography: setting the example for standardization. In: Barber B, Cao D, Gin D, Wagner G, eds. *MEDINFO 89.* Amsterdam: North Holland Publ Comp, 1989; 538–42.

Address of the authors:
Jos L. Willems, M.D., Ph.D.,
CSE Project Leader,
Univ. Dept. Hosp. Gasthuisberg,
49, Herestraat,
3000 Leuven,
Belgium

Appendix

Participating Investigators in the CSE European Working Party

Belgium: C. Brohet (Univ. of Louvain), J. L. Willems (Univ. of Leuven)
West-Germany: J. Meyer and J. Michaelis (Univ. of Mainz), S. J. Pöppl (Inst. Medical Data Processing, München), Chr. Zywietz (Univ. of Hannover)
Denmark: J. Damgaard Andersen (Univ. of Copenhagen)
France: P. Arnaud (INSERM U121 Lyon), B. Denis (Univ. of Grenoble), P. Rubel (INSA, Lyon)
Great-Britain: P. J. Bourdillon (Univ. of London), P. W. Macfarlane (Univ. of Glasgow)
Greece: S. Moulopoulos (Univ. of Athens), E. Skordalakis (Tech. Univ. of Athens)
Ireland: I. Graham (Univ. of Dublin)
Italy: S. Dalla Volta (Univ. of Padova), R. Degani (Ladseb CNR, Padova), G. Mazzocca (Univ. of Pisa)
The Netherlands: J. H. van Bemmel and J. Kors (Erasmus Univ. Rotterdam), H. J. Ritsema van Eck (Rotterdam), G. van Herpen (Univ. of Leiden)
Portugal: C. Abreu-Lima (Univ. of Porto), H. B. Machado (Univ. Hosp. Lisbon)

Consultants

J. J. Bailey (N.I.H., Bethesda, Md, U.S.A.) and H. V. Pipberger (George Washington Univ., Washington DC, USA), P. M. Rautaharju (Univ. of Dalhousie, Canada), E. O. Robles de Medina (Univ. of Utrecht, The Netherlands)

Non-European Participants

USA: C. Monroe, S. Charlesworth (Hewlett-Packard), K. Michler (Telemed), I. Rowlandson (Marquette), D. W. Mortara (Mortara Instruments)
Canada: P. M. Rautaharju, P. Macinnis (Univ. of Dalhousie)
Japan: M. Okajima, N. Okamoto, M. Yokoi (Univ. of Nagoya), N. Ohsawa (Fukuda Denshi)

Representing BME-COMAC

J. Marques-Montes (Univ. of Madrid, Spain) and U. Faust (Univ. of Stuttgart, FRG)

CSE Coordinating Center

Division of Medical Informatics, Univ. of Leuven, Leuven, Belgium