Alberta Heritage Foundation
for Medical Research

Accuracy and reliability of using computerized interpretation of electrocardiograms for routine examinations

Paula Corabian

January 2002

HTA 25: Series A Health Technology Assessment

This Health Technology Report has been prepared on the basis of available information of which the Foundation is aware from public literature and expert opinion and attempts to be current to the date of publication. It has been externally reviewed. Additional information and comments relative to the brief are welcome and should be sent to:

Director, Health Technology Assessment
Alberta Heritage Foundation for Medical Research
1500, 10104 – 103 Avenue
Edmonton
Alberta T5J 4A7
CANADA

Tel: 780-423-5727, Fax: 780-429-3509

ISBN 1-896956-46-7

Alberta's health technology assessment program has been established under the Health Research Collaboration Agreement between the Alberta Heritage Foundation for Medical Research and the Alberta Health Ministry.

ACKNOWLEDGMENTS

The Alberta Heritage Foundation for Medical Research is most grateful to the following persons for provision of information and comments on the draft report. The views expressed in the final report are those of the Foundation.

Dr. Israel Belenkie, Foothills Hospital, Calgary, Alberta

Dr. Pentti M. Rautaharju, Wake Forest University School of Medicine, Winston-Salem, NC

We also wish to express our gratitude for the assistance provided by Dr. Lorne B. Mitchell, Foothills Hospital, Calgary, Alberta during the preparation of the draft report.

SUMMARY

Computerized electrocardiogram (ECG) interpretation has been developed and used in an attempt to reduce the human readers' time and work demands, reduce and standardize the coding used by different ECG classification systems and minimize the inter- and intra-observer variation in human interpretation.

The past decade has witnessed a rapid increase in the use of computer programs to interpret all types of ECG tests recorded routinely. The use of computerized ECG interpretation has become accepted as providing the less experienced clinicians with an almost immediate, reasonably accurate interpretation to assist them in achieving more accurate interpretations. However, the use of computerized ECG interpretation has not resulted in major improvement in diagnostic accuracy of human interpretation.

The literature reviewed suggests that the role of computerized ECG interpretation has future potential in primary care. The immediate availability of computerized ECG interpretation has been seen as a significant improvement for practicing clinicians. However, the question whether computerized ECG interpretation can be used as an automated test for screening normal ECGs in asymptomatic adults during routine clinical examinations (without consulting with an expert in ECG interpretation) has yet to be answered.

No clinical studies determining whether computerized interpretation of resting ECG can be considered an accurate and reliable automated laboratory test for this indication were located. The methodological quality of the available published evidence is limited in several aspects and conclusive evidence could not be found:

- on the diagnostic accuracy and reliability of using computerized interpretation of resting ECG as an automated laboratory test for screening normal ECGs in asymptomatic adults; and
- on whether computerized interpretation of resting ECG can replace the interpretation by a skilled professional in an ambulatory clinical environment.

The available published evidence suggests that the computer programs with the best performance may be as accurate as the human interpretation in diagnosing normal ECGs. However, computerized interpretation of ECGs should be used with an awareness of the risk of false positive and false negative findings.

Those considering use of computerized ECG interpretation for routine clinical examination of asymptomatic adults in ambulatory clinical settings should be aware that:

- The ECG test is only one of the tests used to detect or exclude possible heart conditions and is of limited value as a stand alone screening tool in an apparently healthy population.
- Whether the use of computerized ECG interpretation actually increases physician's accuracy in ECG interpretation, saves physician time, improves quality of patient care and leads to a reduction in the costs associated with ECG interpretation have yet to be determined.
- There are different applications of computerized ECG interpretation. Different computer programs have been developed to interpret ECGs performed for different cardiac conditions within different clinical settings.
- The computer programs available on the market apply different approaches to diagnostic classification of ECGs and use different terminology.

CONTENTS

Acknowledgments i
Summary ii
Introduction 1
Scope of the Report 2

SECTION 1

ECG: Description and Interpretation 5
- General description 5
- ECG interpretation 6

Computerized ECG Interpretation 6
- Computer modeling of ECG classification 7

Advantages and Disadvantages of Computerized ECG Interpretation 9
Performance Evaluation of ECG Computer Programs 10
- Diagnostic accuracy of computerized ECG interpretation 11

The CSE Project 12
- CSE diagnostic study 13
- Other studies using the CSE database 16

SECTION 2

Available Evidence on Diagnostic Accuracy 18
- Sekiguchi et al. study 18
- Heden et al. study 20

Expert Opinion 21
Regulatory Status of Computerized ECG Interpretation in Canada 22
Discussion 23
Conclusion 25
Appendix 1: Methodology 26
References 29

INTRODUCTION

This report is a systematic review of the literature on the use of computerized electrocardiogram (ECG) interpretation for detecting heart conditions as a part of routine medical examinations in healthy adults. It has been prepared to inform the Provincial Health Authorities of Alberta and others on the available published evidence reporting on the diagnostic accuracy and reliability of using computerized interpretation of ECGs in ambulatory settings.

The use of electrocardiography for detecting heart problems in asymptomatic adults during routine clinical examinations may lead to a large number of ECGs to be interpreted yearly. Traditionally, cardiologists have performed visual ECG interpretation. To reduce the workload for electrocardiographers (cardiologists, internists, or other appropriately trained physicians), computer programs for ECG interpretation have been developed and used with or without over-reading by experienced electrocardiographers.

The first attempts to automate ECG interpretation by digital computer (automated separation of normal ECGs from abnormal ECGs done by a computer program designed to recognize and measure the wave components of the digitized ECG signal) were made by Pipberg et al. in the late 1950s [(1), (2), (3)]. Since the first commercially available programs were introduced in the early 1970s [(1;3)], computerized ECG interpretation systems have become more sophisticated and less expensive with an increasing number of available programs.

The use of computerized ECG interpretation has increased rapidly in general and hospital practice over the last 15 years [(2), (3), (4), (5), (6), (7)]. Computer programs for ECG interpretation are widely used in the United States and in the Asian and Pacific countries [(3), (5), (8), (9)]. The continued increase in the use of computerized ECG interpretation mandated the necessity of evaluating its diagnostic performance and of understanding its advantages and disadvantages.

SCOPE OF THE REPORT

This report is not intended as a comprehensive review of all the research published on the computerized interpretation for all applications of ECG tests used in all clinical settings. It considers only the use of computerized interpretation of resting ECGs performed in ambulatory settings as part of the routine clinical examinations of healthy adults. The aim of this report is to inform practitioners and other interested parties on the available published evidence on the diagnostic accuracy and reliability of using computerized interpretation of resting ECG to detect normal heart activity in healthy or asymptomatic adults.

This study does not discuss the other uses of computers in cardiology or other forms of electrocardiography using computers (such as exercise ECG or Holter monitoring). Neither does it assess the effects of computerized interpretation of other ECG tests when used by clinicians in other clinical settings, such as trained physicians and cardiologists in inpatient hospital and emergency departments. The economic aspects of the use of computerized interpretation of resting ECG performed in ambulatory settings are also not considered.

For the purpose of this study, an asymptomatic or healthy adult was defined as someone aged 18 years and older with no apparent or suspected heart dysfunction or disease (no symptoms, abnormal physical findings, or previous abnormal ECG) [(10), (11)]. Also, for the purpose of this review, ambulatory setting referred to outpatient clinical settings such as medical clinics or doctor's offices and hospital outpatient departments.

A preliminary review of the literature revealed that several computer systems and analysis models have been developed for ECG interpretations and evaluated during the past four decades [(1), (12)]. Many changes, developments and further improvements occurred in the domain of computers and computerized ECG interpretation in the late 1980 and early 1990s and since the completion of the Common Standards for Quantitative Electrocardiography project (also known as the CSE project), an international action sponsored by the European Community. Therefore, this review focused on the literature published from 1994 to April 2001.

The intent is to respond to the following questions:

- Can computerized interpretation of resting ECGs be considered an accurate and reliable automated laboratory test to detect normal activity in healthy or asymptomatic adults?
- Can computerized interpretation of resting ECG replace interpretation by a skilled professional?

- What are the advantages and disadvantages of using computerized interpretation of resting ECG as an automated laboratory test in an ambulatory clinical environment?

The report consists of two main sections. The first summarizes the information published on the advantages and limitations of using computerized ECG interpretation and on the performance evaluation of this technology. This section also summarizes the findings reported by the diagnostic study conducted as part of the CSE project.

The second section summarizes the findings reported after completion of the CSE project that provide evidence on the diagnostic accuracy and reliability of computerized interpretation of resting ECGs performed in ambulatory settings in healthy or asymptomatic adults during their routine clinical examinations. Opinion from experts in ECG interpretation is also provided. This section also includes information on the regulatory status of computerized ECG interpretation in Canada.

SECTION 1

ECG: Description and Interpretation

Electrocardiography is a non-invasive investigative method that provides information for the detection, diagnosis and therapy of cardiac conditions [11], [13],(http://www.cardionetics.com), (http://www.heartinfo.org). The electrocardiogram (ECG) is the product of the electrocardiograph, a medical device capable of recording potential differences generated by the electrical activity of the heart. Modern computerized electrocardiographs are also capable of producing ECG measurements and diagnostic classification of the ECG.

General description

Every action of the heart muscle is associated with an electrical potential, which varies throughout the cardiac cycle (http://www.cardionetics.com), (http://www.heartinfo.org). The electrical potential is detected by electrodes (leads) placed on the skin in specific locations (chest and limbs). During electrocardiography the varying electrical potentials are recorded and graphed as a series of waves.

The ECG acquisition systems used to date have been designed to obtain ECG signals that describe the cardiac electrical activity (depolarization and repolarization of the cardiac muscle cells) as accurately as possible from the body surface ECG leads. Most systems use 12 ECG leads, which are commonly recorded simultaneously [1], [2], [3], [5] [11], [12], [14], [15]. Some systems record only the three orthogonal leads (X, Y, and Z) simultaneously (known as the vectorcardiographic or VCG systems) and other (hybrid) systems record all 15 leads (12 ECG leads and the VCG leads) simultaneously [1], [2], [16], [17].

Depending on the condition being investigated, different ECG tests have been developed such as resting ECG (also referred to as the standard12-lead ECG), exercise (or stress) ECG and 24-hour ECG (or long-term ambulatory ECG for continuous monitoring of the heart tracings). These tests are designed to detect various heart problems. An individual may have one, two or three of these tests, performed in different clinical settings, depending on the presenting symptoms and the results obtained from each test.

The resting ECG typically involves recording and analysis of 12-lead ECG from a short-term (10-second) recording (http://www.cardionetics.com) [11]. It can be used to detect normal heart activity, normal and abnormal cardiac rhythms, acute and old myocardial infarctions and ischemic injury, as well as other abnormalities. However, it may not detect clinically significant, even severe, cardiac abnormalities (http://www.cardionetics. com), (Rautaharju, personal communication) [10].

ECG interpretation

The ECG interpretation is accomplished through a sequential analysis of the heart electric events. An ECG signal consists mainly of a P wave (which indicates atrial depolarization), a QRS complex (which represents ventricular depolarization), and a T wave (representing the ventricular re-polarization) (http://www.cardionetics.com), (http://www.heartinfo.org), [1], [3] [14]. The observable manifestations of the various intervals, segments and complexes recorded on an ECG (amplitude, direction and duration of the waves, and their morphological aspects) are analyzed. The information obtained is also used to detect and diagnose normal and abnormal cardiac rhythms and conduction patterns.

There are many variations and combinations of ECG features or parameters which must be measured, studied, analyzed and correlated one with another and with other available data before a definite ECG interpretation is made (http://www.cardionetics.com), (http://www.heartinfo.org), [11], [13], [14], [18], [19], [20], [21]. Each of the waveforms has its own sensitivity and specificity for detection of particular various abnormalities and may be influenced by many pathologic and pathophysiologic factors [11]. The same ECG pattern can be recorded in individuals who have different structural and pathophysiological states [11].

Therefore, ECG interpretation requires a systematic approach that includes knowledge of the patient’s age, sex and race and presenting complaint/symptom (http://www.cardionetics.com), (http://www.heartinfo.org), [11], [13], [14], [18], [19], [22], [23]. The abnormalities observed must be correlated with the data collected from the individual’s medical history and physical examination and from other tests performed.

Most clinicians who perform ECG interpretation develop skills and become experienced during residency or cardiovascular fellowship training and subsequently in clinical practice [11]. In Alberta, all physicians wanting to interpret ECGs, with the exception of cardiologists and pediatric cardiologists, must pass an examination on ECG interpretation (http://www.cpsa.ab.ca). The experts vary in their preferences of diagnostic criteria for ECG interpretation and inter- and intra-observer variations in ECG interpretation have been a concern [4], [5], [11], [15], [18], [24], [25].

COMPUTERIZED ECG INTERPRETATION

Computerized ECG interpretation has been developed in an attempt to reduce the human readers’ time and work demands, reduce and standardize the coding used by different ECG classification systems and minimize the inter- and intra-observer variation in human interpretation [1], [8], [9], [11], [15], [18], [26], [27].

Computer programs built into most ECG equipment currently available on the market are used to interpret many applications of ECG tests [(1), (6), (9), (12), (23), (28)]. These programs are designed to produce measurements and provide diagnostic interpretive statements and are mainly aimed at obtaining results comparable to those obtained by experts. They adopted a similar approach: a measurement program and an interpretive program (that interprets the clinical significance of the measurements along with a rhythm analysis algorithm) [(1), (2), (12), (13), (14)], (http://www.cardionetics.com). Data compression is applied to the signal for digital transmission and compact recording to reduce processing time and allow long-term storage.

The measurement program generally consists of signal acquisition and conditioning (which refers to transmission and storage of digital ECG data), wave detection and characterization, and feature or parameter extraction [(1), (2), (13), (14), (26), (27)], (http://www.cardionetics.com). The interpretation program is the program designed to classify the ECG parameters. Once the waveform descriptors are obtained and the adequate parameters are selected they are used to allocate the ECG to one or more diagnostic classes in the pattern classification phase.

Measurement, selection of ECG parameters, and methods used for pattern classification vary according to the program used [(1), (2), (13), (14), (26), (27)], (http://www.cardionetics.com). Different measurement programs apply various principles with respect to the analysis: some measure single heart beats and others analyze averaged beats [(1), (2)].

Computer modeling of ECG classification

There are two commonly used approaches to computer-ECG analysis: separate analysis of each ECG lead and simultaneous (global) analysis of all leads recorded simultaneously (Rautaharju, personal communication). The individual lead analysis approach measures ECG wave duration and amplitudes from all heartbeats in individual leads. Combined lead analysis, presently the most commonly used approach, first clusters individual beats into main categories according to the waveform patterns and interval measurements. These main clusters are then merged into representative complexes, either suing an average or a median of the waveform patterns.

Although adequate parameter selection is basic for any ECG pattern classification, three different approaches to diagnostic classification have been developed and used: the logical model, the statistical model, and the artificial neural networks (ANN) model [(1), (5), (12) (14), (23)].

The ***logical model*** generally uses decision-tree logical reasoning and is designed to combine expert reasoning, clinical information and knowledge related to ECG parameters [(1)]. This approach, presently the most commonly used, is amenable for using ECG data initially derived either from the individual lead or from combined lead analysis (Rautaharju, personal communication).

The logical approach is likely to be easily accepted by users as it simulates the experts' methods. However, it is plagued by the difficulty in choosing the measurement methodology of parameters and diagnostic criteria. Several diagnostic criteria for ECG classification and more than one criteria set have been developed and used for the same diagnostic problem and experts vary in their preferences [(1), (4), (5), (9), (17)]. A variety of computer programs have been developed using various sets of often substantially differing expert criteria (Rautaharju, personal communication).

In the second approach, statistical classification techniques are used to analyze prior probabilities from ECG-independent criteria (for example, a fixed set of disease prevalence) combined with a series of ECG parameters [(1)]. The ***statistical model*** is thought to be advantageous since it can resort to more complex measurements embodying more information about the ECG features. However, since the prevalence of cardiac diseases may vary according to the clinical setting, adjustments have to be made in the classifier for each particular application [(1)].

Another approach to ECG signal classification is the ***ANN model*** [(1), (12), (29)]. It has been used for ECG pattern recognition, diagnostic interpretation, rhythm analysis, and data compression. The ANN model classifying ECGs consists of three layers of neurons: an input layer fed with signal features, a hidden layer where signal processing is carried out and an output layer that yields the classification [(1), (12), (29)]. Its ability to easily adjust the networks outputs in different clinical situations has been suggested as one important advantage of this model over the logic model [(29)]. A disadvantage with ANN model is the difficulty in extracting criteria for any given classification (Rautaharju, personal communication).

The printed reports issued by the systems providing computerized ECG interpretation consist of a series of codes and diagnostic statements, which describe the rate and the rhythm, and the abnormalities of the various waves, intervals, complexes and segments found in the ECG signal. Also printed are the standard measurements (e.g., time intervals, axes of the waveforms). The length of the outputs depends on the number of ECG characteristics that are detected and the diagnostic terminology used by the system. Some list only one or two statements, others provide a list of most likely interpretations with decreasing level of likelihood, and some use quantitative certainty indices and others use probabilities [(1), (5), (23), (29)].

ADVANTAGES AND DISADVANTAGES OF COMPUTERIZED ECG INTERPRETATION

In terms of advantages associated with computerized ECG interpretation it has been stated that its use [(3), (6), (7), (8), (9), (22), (27), (30)], (http://www.cardiologyshop.com), (http://www.cardionetics.com):

- makes interpretations readily available to the clinicians,
- provides fast and reasonably accurate analysis and classification of recorded ECGs,
- reduces interpretation time for the busy clinician (the most when there are multiple diagnoses, and in cases of large numbers of ECGs to be interpreted),
- assists the clinician in ECG interpretation and improves the clinician's diagnostic accuracy,
- increases efficiency (as it provides compact and easy to read printout and speeds up reading and reporting procedures),
- facilitates automatic checking of hundreds of criteria (dependent on the patient's age and sex),
- combines experience of many experts,
- increases standardization (as the same criteria are used each time),
- facilitates handling of large numbers of ECGs,
- assists in ECG storage and retrieval via database.

However, the use of computerized ECG interpretation systems is not without faults. Several papers identified some important limitations, considered by some investigators as reasons for reluctance to accept computerized ECG interpretation (http://www.cardionetics.com), [(5), (6), (9), (13), (16), (18), (19), (20), (24), (30), (31), (32), (33), (34), (35)]:

- Computerized ECG interpretation systems have difficulty in diagnosing some cardiac conditions.
- Computerized ECG interpretation performs poorly in diagnosing cardiac arrhythmia. Failure is most frequently associated with the incapacity to detect P waves [(13), (30)].
- Computerized ECG interpretation may miss subtle changes in the ECG signal.
- The printed report issued does not offer a clinical differential diagnosis that reveals the cardiac conditions potentially responsible for the abnormalities found on the interpreted ECG or the possible heart conditions the individual might have. Some cardiac conditions may not alter the ECG and these conditions should be known when using the

information revealed by the ECG [19]. Hurst suggests that "even a normal ECG should have a differential diagnosis that includes no heart disease and a list of all types of heart disease that may be associated with a normal ECG" [19]. ECGs interpreted as normal may mask clinically important cardiac abnormalities (http://www.cardionetics.com), [32]. The issue with computerized "normal" reading is whether or not it is truly "within normal limits" (Belenkie, personal communication).

- The printed report does not offer reasons for a particular diagnostic statement (additional information on how the program arrived at that statement).
- The programs analyze the ECG signal and do not exercise clinical judgement. Computers cannot correlate the ECG findings with the clinical information as a trained and experienced human reader.
- The computerized ECG interpretation is subjected to large measurement and classification errors. It can provide erratic diagnoses with an error rate of about 20% [3], [5], [16], [19], [20] and false diagnostic suggestions may mislead the clinician.
- Technically poor ECG records may not be analyzed satisfactorily.
- The computer is dependent on good quality signal, and it does not always recognize artifacts. The diagnostic accuracy and reliability of the computerized ECG interpretation are sometimes affected by random and periodic components of signal noise (whether of biological or technical origins) or by misplacement of electrodes [22], [33], [34].
- The computer software may not express the most recent expert views in terms of ECG interpretation.

Manufacturers caution that the ECG interpretive software is designed to be of assistance to licensed clinicians who are performing ECG analysis and never replace their decisions [31], (http://www.schiller-ag.ch/ecg).

PERFORMANCE EVALUATION OF ECG COMPUTER PROGRAMS

Performance evaluation of computer programs designed to interpret ECGs is a complex and controversial issue and there is no single evaluation method [1], [2], [5], [9], [23]. The evaluation is complicated by the fact that there is no single and independent reference standard for it due to the varied nature of the ECG diagnoses.

Since the first attempt to automate ECG interpretation many studies have been reported on the performance evaluation of computerized ECG interpretation systems. These studies reflected diverse opinions, resulting from the different perspectives of those who evaluated such systems. The evaluation of computerized ECG interpretation has been viewed from different angles, such as diagnostic accuracy, consistency or diagnostic reproducibility, utility and acceptability. However, from the clinical perspective diagnostic accuracy and reliability are the most important issues [(1), (2), (5), (9), (23)].

Diagnostic accuracy of computerized ECG interpretation

ECG diagnoses have been classified in type A, type B, and type C, all of them being considered important from the clinical point of view [(1), (23)]:

- Type A diagnoses such as normal (normal individual in terms of cardiac conditions), myocardial infarction, and right and left ventricular hypertrophy, refer to the presence or absence of an anatomic lesion or pathophysiologic state. These statements can be confirmed by ECG-independent evidence (such as data obtained from medical history and clinical examination; heart surgery or autopsy findings; evidence obtained from echocardiography, ventriculography, scintigraphy, cardiac catheterization and laboratory measurements; and identification of heart disease events by monitoring hospital discharges).
- Type B diagnoses are statements referring to diagnoses of ECG changes (such as arrhythmia, and conduction disturbances). They include ECG diagnoses such as right and left bundle branch blocks and cardiac arrhythmias. The ECG itself primarily establishes them and thus an ECG-independent confirmation is usually not available. Most often the reference method used to validate these statements is the human interpretation.
- Type C diagnoses refer to the descriptive features of the ECG waveforms (such as axis deviations and non-specific ST-T changes), which usually cannot be documented by other means.

To determine the diagnostic accuracy and reliability of computerized ECG interpretation, two methods have been considered: one uses the human observer(s) as the reference and the other one uses an appropriate set of cases classified by ECG-independent evidence [(1), (23)].

The first method accepts the human interpretation as reference in order to determine the diagnostic accuracy of computerized ECG classification [(1), (23)]. The human interpretation (performed by an experienced cardiologist or a panel of cardiologists, constrained or not constrained by a fixed set of diagnostic criteria) is practically the only one available for type B and type C diagnoses. The operating conditions of the cardiologist or panel have to be specified for such an evaluation to be meaningful, as the human interpretation may use the same or different diagnostic criteria than those used by the computer program.

The second method considers non-ECG evidence as the reference in order to evaluate the diagnostic accuracy of computerized ECG interpretation [(1), (23)]. It uses a data set of ECGs performed on individuals in whom the cardiac conditions were evaluated by ECG-independent means such as: clinical examination, cardiac catheterization of the heart, heart surgery, autopsy, etc. Since early 1970s this method has been recommended as the only one to be used for independent and objective evaluation of diagnostic accuracy of computerized ECG interpretation in the case of type A diagnosis.

The literature search revealed that before 1990 most evaluations of computer programs in terms of diagnostic accuracy of their ECG interpretation have relied on the first method for all diagnosis types and only a limited number of evaluations were based on ECG-independent validated databases. Their results are difficult to compare since they are based on different databases, most of them collected within a single setting.

In 1978 the Task Force III of the American College of Cardiology defined methods to be used for evaluating computerized ECG interpretation programs, issued recommendations on computers in diagnostic electrocardiography and proposed rules for creating a reference database for accurate evaluation of type A diagnoses [(1), (23)]. As a follow-up of this activity, in an attempt to implement its outcomes on an international basis, the European Community initiated in 1984 the Common Standards for Quantitative Electrocardiography project (the CSE project), which was finalized at the beginning of the 1990s.

THE CSE PROJECT

The CSE project is considered a major breakthrough in the computerized ECG interpretation domain [(1), (2), (9), (23)]. It involved two main studies: the CSE standards study and the CSE diagnostic study. The ***CSE standards study*** was conducted to set standards for ECG measuring and for evaluating computerized ECG interpretation programs. The main objective was to reduce the variation of measurement by different computer programs. The aim was to standardize ECG measurements made by the computers, obtaining agreement on definitions of waves and of the references for the on- and offsets of the P wave, QRS complex and T wave.

As a follow-up on the CSE standard study, the ***CSE diagnostic study*** was set up to evaluate the performance of such programs in terms of diagnostic accuracy and reliability [(2), (23)]. During the CSE project a large dataset of “appropriate” ECGs (performed on patients with clinical diagnosis well documented by ECG-independent means) were collected and a team of 48 investigators [(1), (2), (23)]. Fifteen ECG processing centers from nine European and three non-European countries were organized in an international cooperative effort.

The following commentary summarizes the findings reported by the CSE diagnostic study during which the diagnostic performance of ECG interpretation by computer programs or cardiologists has been evaluated against ECG-independent evidence (as gold standard) [(1), (2), (9), (23)].

CSE diagnostic study

The data collection was based on ECG-independent clinical information used to classify all 1,220 subjects included in seven diagnostic categories: normal; left, right and biventricular hypertrophy; anterior, inferior and combined myocardial infarction [1], [2], [9], [23]. All subjects included (831 men and 389 women) were white adults (mean age of 52± 13 years). A review board, consisting of three cardiologists, has verified the clinical information for all cases. ECGs showing major conduction defects and those of "poor quality" were excluded.

Fifteen computer programs (most of the programs on the market at that time, designed by manufacturers or by university centers) and nine cardiologists (from seven different European countries) independently interpreted the ECG recordings [1], [2], [9], [23]. All fifteen ECG processing centers and all nine cardiologists were asked to apply a scheme for translating statements into a common set of diagnostic codes. Therefore, the diagnostic criteria used by the human interpreters and by the computer programs were identical. The computer programs included were respectively: AVA, Marquette, Louvain, Hannover, Hewlett Packard, IBM, Nagoya, Lyon, Glasgow, Porto, Padova, Modular ECG Analysis System (MEANS) and Leuven programs.

Nine of the programs used the 12-lead system (also referred to as the ECG system) [1], [2], [23]. Of these nine programs, seven were based on logic approach and two on statistical approach. The other six computer programs used the 3-lead system (also known as the VCG system) [1], [2], [23]. Of these six programs, two employed statistical classifiers and four employed used logic classifiers.

Of the nine cardiologists who also classified the ECG tracings, eight were ECG interpreters and five were VCG interpreters, none of them being constrained by a fixed set of diagnostic criteria [1], [2], [4], [9], [23]. Except for age and sex, no prior clinical information was provided to the processing centers or to the cardiologists. Race dependent criteria were not incorporated in the computer programs [23]. The interpretations performed by the programs and cardiologists have been compared with the clinical diagnoses as defined by the gold standard (ECG-independent evidence).

Overall, the accuracy results reported by the CSE diagnostic study showed that the human interpretation as a whole perform better than a large majority of the computer programs, at least when interpreting the standard 12-lead ECGs [1], [2], [5]. The results also showed that [1]:

- programs using the statistical model performed better than those using the logic approach, regardless of the lead system employed (median total accuracy 76.6% vs. 69.3% for ECG and 73.4% vs. 65.4% for VCG);

- for programs using the same approach for diagnostic classification, those which employed ECG system performed better than the VCG programs (76.6% vs. 73.% for the statistical approach; 96.3% vs. 65.4% for the logic approach); and
- the 12-lead statistical programs performed as well as the cardiologists (76.6% vs. 76.3%) and the 3-lead statistical programs performed better than the cardiologists (73.4% vs. 70.3%).

The ECG Working Group of the CSE diagnostic study

During the CSE diagnostic study, the ECG Working Group compared the diagnostic performance of nine computer programs, which used the ECG system, and eight cardiologists who were ECG interpreters (5). The computer programs were, respectively: Marquette, Hannover, Hewlett-Packard, Medis, Nagoya, Glasgow, Padova, MEANS, and Leuven programs. Two of these programs were based on the statistical approach (Hannover and Leuven) and the remaining seven applied the logistic approach.

The comparison also involved three combined diagnostic categories in the data set used including 382 normal subjects, 547 myocardial infarction cases, and 291 patients with ventricular hypertrophy (5), (9). In these diagnostic categories the best of the evaluated computer programs using ECG systems performed just as well as the "average" cardiologist (diagnostic accuracy: 91.5% vs. 94% for normal; 82.1% vs. 79.1% for myocardial infarction; and 67.0% vs. 59.9% for ventricular hypertrophy).

Most cardiologists had a better performance than the computer programs in confirming normality (in 92.7 to 97.6 percent of controls versus in 86.3 to 97.1 percent of controls). However, a comparison of the combined interpretations made by the cardiologists versus the combined results of the computer programs showed an approximately equal high performance in confirming normality (97.1% and 96.7%). Among the best programs, the MEANS program correctly diagnosed significantly more of the 382 normal subjects than the "average cardiologist" (5). The best program refers to the program with the best overall performance in all individual diagnostic categories evaluated (9). The "average cardiologist" refers to the cardiologist closest to the estimated median overall performance (9).

The inter-observer variability among cardiologists was significantly less than that observed among all programs considered ($p<0.001$). The intra-observer reproducibility of the cardiologists' diagnoses ranged from 76.8% to 90.4% for the 125 selected ECGs read twice by each cardiologist (5).

According to the ECG Working Group (5) their results demonstrate that the computer programs with the best performance are almost as accurate as the best cardiologists in classifying ECGs in the seven diagnostic groups and the three combined diagnostic categories considered. They suggested that by combining different computerized ECG interpretations the diagnostic accuracy might be increased as in an expert panel.

The investigators concluded that their “results demonstrate that standard ECG-reading computer programs can assist clinicians in achieving more uniform and consistent interpretations of ECGs. However, some programs perform at a considerably lower level than cardiologists and require improvement” and “all the programs studied can still be improved. Most cardiologists had a better performance than the programs in conforming normality and a higher sensitivity in diagnosing anterior myocardial infarction”.

Limitations of the CSE diagnostic study

The CSE diagnostic study was not exempted from criticism [(5), (23), (26), (28)]. Several methodological weaknesses may limit the interpretation of the findings obtained by the CSE diagnostic study and hold back the extrapolation of these findings to the general population. These limitations are summarized as it follows:

- There was little information on the selection of the patients. The selection of the ECGs included in the CSE diagnostic study was not done at random and it is not clear whether they represented consecutive cases. No information was available on how many patients were excluded and on the definition of “poor technical quality“.
- The study evaluated interpretation of ECGs performed on hospitalized patients and the normal cases included ambulatory patients and patients referred for cardiologic examination. Since the prevalence and severity of the cardiac disease in these patients may differ from that in the population at large, the results may not be extrapolated to the general population.
- The cardiologists had knowledge of the composition of the database. The designers of the computer programs could have also introduced bias into mapping of their program statements.
- The study did not measure some important cardiac conditions and it was limited to ECGs from patients with one of the seven main diagnostic categories. The data set did not encounter the whole spectrum of different disease severity and combined disease groups.
- The data set consisted mainly of single disease cases and for some diagnoses only a limited number of ECGs were available.
- The study did not evaluate some of the important aspects of electrocardiography such as rhythm and conduction statements.

Other studies using the CSE database

Arnaud et al. [33] tested the competence of cardiologists and the efficiency of computer programs included in the CSE project in the interpretation of ECGs and VCGs. Each group of interpreters, cardiologists or computer programs, for each category (ECG or VCG interpretation) was considered as a whole. By using a multivariate linear model to estimate the significance of these results, the investigators found that the cardiologists' and the programs' performance for the normal tracings (considered as tracings easy to interpret) was significantly better than that for all other diagnoses. However, the cardiologists more often correctly diagnosed the tracings indicating normality than the computer programs. The investigators commented that this may be due to the fact that the comprehensive approach applied by the cardiologists is more efficient that the logic approach of the program in such "obvious cases". For the normal tracings the performance results (for cardiologists as well as for programs) were much higher for the ECG system than for the VCG system.

Based upon the distributions of measurement errors for key intervals and accuracy figures (sensitivities, specificities, predictive values) from the CSE studies, minimum performance figures were derived by Zywietz and Willems [28]. Their analysis of the distributions for sensitivity and specificity and for positive and negative prediction values showed that the differences in the performance among the programs evaluated could be seen mainly in the sensitivity and positive predictive values.

According to Zywietz and Willems [28] the recommended sensitivity level for diagnosing normal ECG is 85% and the positive prediction value is 60%. In terms of specificity and negative predictive values, the differences among the programs were small and almost all were above 80%. The figures refer to the CSE database, which had a confined composition (30-35% normal, 20-25% hypertrophy, 45-50% infarction). The terms "sensitivity", "specificity" and "predictive values" were used according to "standard definitions" used by Willems et al in the CSE diagnostic study [5]. The investigators caution the definitions of thresholds of minimum performance requirements as described in their paper [28] are still disputable because it contains some arbitrariness.

Section 2

AVAILABLE EVIDENCE ON DIAGNOSTIC ACCURACY

The literature search revealed many studies reporting on the evaluation of computerized ECG interpretation programs used in many types of clinical settings. However, no prospective / retrospective comparative studies reporting on the diagnostic performance of computerized interpretation of ECG as an automated laboratory test used to screen heart conditions in healthy, asymptomatic adults as part of their routine examination in ambulatory settings have been located.

The literature search identified two comparative studies [8], [29] conducted to evaluate the diagnostic accuracy and reliability of the computerized interpretation of normal 12-lead ECGs as compared to ECG-independent evidence and reported on the overall ECG diagnosis. The following commentary summarizes the findings reported by these studies and presents their methodological limitations.

Sekiguchi et al. study [8]

Sekiguchi et al. compared the ECG interpretation performed by computer with that performed by physicians in training, as well as with the diagnoses provided by cardiologists. The ECGs included were performed on 1,058 Japanese adults (812 men and 246 women, mean age 49 ±19 years). The study was restricted to ECGs from patients whose diagnosis could be validated by non-ECG evidence, such as the results of cardiac catheterization, echocardiography and measurements of serum cardiac enzyme. The selected ECG recordings represented a wide variety of ECG configurations or patterns. The clinical diagnoses of ECGs followed the criteria outlined by the American College of Physicians / American College of Cardiology / American Heart Association (ACP / ACC / AHA) Task Force [36].

The ECGs were obtained from a university department between 1990 and 1996. (all standard ECG leads were recorded simultaneously). They were analyzed by 25 physicians in training (graduated from the medical school, in the post 2 years) and three experienced cardiologists. ECGs of "poor quality" were excluded from the study. To analyze the ECGs the investigators used the 12SL ECG analysis program developed by Marquette Electronics.

Of all cases in the study, 54% were diagnosed as "normal" by the computer and by the physicians, and 33% were diagnosed as "abnormal" by both. Eighty-six cases (8.1%) were diagnosed as abnormal by the computer but as normal by the physicians. Fifty-one cases (4.8%) were diagnosed as normal by the computer but as abnormal by the physicians. The specificity ranged from 99.8% to 100% for ECGs diagnosis by computer analysis when compared to the "cardiologists' assessment".

Sensitivity of computerized ECG interpretation ranged from 55.6% to 100%, with lower sensitivity (less than 80%) in diagnosing right and left bundle branch block and non-specific or specific changes in the ST-T wave. The lower true positive rate (less than 80%) exhibited by the computer analysis was seen in right bundle branch block and myocardial infarction. Interpretation of the P-R and QRS duration showed an agreement between the computer and the physicians (*the statistics were not reported)* but the interpretation of the QT interval by the computer analysis was relatively poor when compared with that evaluated by physicians.

Comparative differences of false negative (FN) and false positive (FP) rates of diagnosis of ECGs between computerized interpretation and physicians in training showed that, with the computer program, the incidence of FN reports was 10.5% while that of FP reports was 16.5%. The incidence of a FP diagnosis with the computer was 18 times higher than that found by the physicians in training.

The investigators concluded, "Such computer programs should be used only with an awareness of the risk of FP and FN findings". To ensure an accurate diagnosis a "knowledgeable" cardiologist must evaluate the ECGs. The computer programs designed to interpret ECGs may be useful "as a first draft when numerous ECGs need to be screened in a short period".

Limitations of the Sekiguchi et al. study [8]

Several methodological flaws limit the interpretation and the generalization of this study's findings. The selection of ECGs included in this study was not done at random and it is not clear whether they were consecutive cases, and whether the study was conducted prospectively or retrospectively. No information was available on how many patients had ECGs performed initially, how many were excluded, and on the definition of ECGs of "poor quality". Also, it is not clear whether the human interpretations were done blindly and independently. That is:

- whether the physicians in training and the cardiologists were blind to each other's interpretation and to the ECG-independent clinical information and to the computer interpretations; and
- whether each of the 25 physicians and three cardiologists interpreted each ECG included in the study.

No information was available on whether the reported interpretations were separate independent readings or collective readings. Because no separate reporting of cardiologists' assessment and physicians' interpretations, it was not clear whether the cardiologists' interpretations were included or not in the reported "physicians' interpretation" results. It is not clear whether the human interpreters used the same diagnostic criteria as those used by the computer program.

The study did not report on the effects of the computerized ECG interpretations on the management decisions (which can affect both the patient outcome and the use of medical resources).

Heden et al. study [29]

Heden et al. compared the diagnoses of healed myocardial infarction based on the interpretation of the 12-lead ECG performed by computer using the ANN model to those performed by experienced electrocardiographer. Their study, based on 351 healthy volunteers and 1,313 patients with a history of chest pain who had undergone diagnostic cardiac catheterization, aimed to determine the agreement between ANN and human interpreter of ECGs on this diagnosis. The healthy volunteers (with no known or suspected heart disease, lung disease or any other pathologic condition that might influence the ECG) were selected at random from a defined urban population. All ECGs recorded were analyzed for the presence or absence of healed anterior myocardial infarction, using ECG-independent methods as a reference. All patients had undergone diagnostic cardiac catheterization.

Patients with normal coronary arteries, normal findings on contrast left ventriculography, no evidence of valve dysfunction or congenital heart disease, ejection fraction ≥50% and an overall study evaluation of “normal” were identified as “catheterization-normal”. A control group was composed of healthy volunteers, patients classified as “catheterization-normal” and patients diagnosed with isolated inferior myocardial infarction. Patients with isolated anterior myocardial infarction and those with both anterior and inferior myocardial infarction were included in the anterior myocardial infarction group. “Technically deficient” ECGs and ECGs showing left bundle branch block were excluded.

A 12-lead ECG was recorded in each subject by using a computerized ECG equipment. An experienced electrocardiographer classified the ECGs into five diagnoses classes: “definitely no anterior myocardial infarction”; “probably no anterior myocardial infarction”; “possible anterior myocardial infarction”; “probable anterior myocardial infarction”; “definite anterior myocardial infarction”. No personal data, clinical findings or results obtained by the ANN interpretation of the ECGs were available for the classification procedure.

The ANN computer program was trained and tested to diagnose myocardial anterior infarction. It classified the ECGs into the five groups by using the computer output values and four different thresholds between 0 and 1. These thresholds were selected so as to give the same number of ECGs in the five diagnosis classes mentioned above as the electrocardiographer classified them. The complete agreement between the ANN interpretation and the human interpretation could be obtained only by using these thresholds. The ANN outputs were transformed to verbal statements.

The electrocardiographer classified 1,291 ECGs as "definitely no anterior myocardial infarction" and "probably no anterior myocardial infarction" (n=1,104 and n=187). Of these, 1,185 were ECGs from the control group resulting in a specificity of 94.8%. The specificity for the ANN program was also 94.8%. Overall, the agreement between electrogardiographer's classifications and those by the ANN system was established in 1,282 ECGs (77.0%).

Based on these results the investigators concluded that the ANN approach could be of value in computerized ECG interpretation in the near future. As reasons for misdiagnosis by the ANN system they mentioned the number of the ECG variables used as input values and the presence of ECGs with uncommon features.

Limitations of Heden et al. study [29]

Several methodological weaknesses limit the interpretation and the generalization of the results reported by Heden et al.:

- It is not clear whether this study is a prospective study or a retrospective analysis of a data base of ECGs performed in a hospital.
- There is not sufficient information on how the volunteers and patients were selected for this study. It is not clear whether the ECGs performed on patients were selected at random or whether they represented consecutive cases.
- No information was available on how many subjects were initially considered, how many were excluded and on the definition of "technically deficient" ECGs.
- It is not clear whether non-ECG means were used to determine presence / absence of known or suspected heart disease in the volunteers included in the study.
- Subjects' characteristics are not described in sufficient details. No information is available on the age, sex and race of the healthy volunteers and patients included in the study.
- The control group in this study did not include only healthy volunteers.

EXPERT OPINION

Professional groups have issued official policy statements (endorsed by cardiologists) declaring that there is no computer program for ECG interpretation that can replace interpretation by a skilled physician or cardiologist, and that all computerized ECG interpretations require careful over-reading by a clinician qualified to interpret ECGs [(11), (25), (36), (37)].

According to Rautaharju [9] the arguments against computerized ECG interpretation may have been biased to some degree because they largely reflect the opinions of cardiologists at university-affiliated academic institutions and the composition of physicians and patient populations is very different outside these departments. The key issue deliberated by Rautaharju was whether or not computerized ECG interpretation can be considered an automated laboratory test. He concluded that with relatively minor improvements in the computer programs, a "sizeable fraction of ECGs", particularly those performed outside the cardiology departments, can be considered for computerized ECG interpretation used as an automated laboratory test. The remaining ECGs would require more advanced software for differential diagnostic classification in combination with professional review using more comprehensive supplementary information than is presently provided in most routine electrocardiographic processing applications.
In a more recent communication, Rautaharju noted that economic issues in addition to scientific considerations continue to play an important role concerning the acceptance of computerized ECG interpretation as an automated laboratory test (Rautaharju, personal communication). Over-reading by electrocardiographers seems rational because the programs generally tend to be oversensitive in reporting ECG as abnormal and occasionally miss or misclassify clinically significant findings. However, it is unlikely that lack of over-reading of asymptomatic patients classified as normal would have a significant impact on patient management decisions at least in ambulatory settings.

According to Belenkie, whether or not computerized interpretation of resting ECGs (performed on asymptomatic people during routine examination in ambulatory settings) can substitute human interpretation remains an issue for debate since it is related to how much error is accepted (Belenkie, personal communication). The impact of missed abnormal readings can be considerable in patients screened for cardiac disease prior to other medical procedure and is still problematic when just office screening is performed in apparently healthy people, although less so.

REGULATORY STATUS OF COMPUTERIZED ECG INTERPRETATION IN CANADA

According to Health Canada and its List of Medical Device Licenses Issued (http://www.hc-sc.gc.ca/hpb-dgps/therapeut/htmleng/md_lic.html, accessed on June 4, 2001), the following companies have approval to market and use their ECG equipment in Canada: Agilent Technologies Inc., Esaote S.P.A., Fukuda Denshi Ltd., Hewlett-Packard GM BH (Medical Products Group), Nihon Kohden Corporation, Schiller AG, Spacelabs Burdick Inc. These companies were contacted by telephone and their representatives were asked whether they also applied for licenses to market their own interpretive software. All had applied for licenses to market their own interpretive software to analyze ECGs performed with their systems and received approval along with their equipment, except for Spacelabs Burdick Inc.

DISCUSSION

During the past decade, several computer systems for interpretation, storage and retrieval of ECGs have been developed and the use of computerized ECG interpretation has increased worldwide. Computerized ECG interpretation systems are currently operational in inpatient hospital and emergency departments, outpatient clinics, primary care and other clinical settings, and are used for all applications of ECG tests (to detect health problems during routine clinical examinations, or to diagnose and monitor suspected cardiac conditions).

The question whether in clinical settings without experts in ECG interpretation (such as primary care settings) a computer system can be used as an automated test to confirm normality in terms of heart condition during routine clinical examinations in asymptomatic adults has yet to be answered. No reliable conclusions could be drawn on whether computerized ECG interpretation can substitute the human interpretation for this application of resting ECG. This issue is related to that of how much error is acceptable, which is still debated.

There are no published primary research data reporting on the use of computerized interpretation of resting ECGs used to detect cardiac conditions during routine examinations of asymptomatic adults in ambulatory settings. The published primary research on the use of computerized interpretation for other applications showed that computers still cannot equal or surpass the diagnostic accuracy of the expert human ECG interpreter. The advantages and limitations of using computerized ECG interpretation with or without over-reading by an expert interpreter are still debated.

The evidence reviewed suggests that computer programs with best diagnostic performance confirm normality (in terms of heart condition) as established by non-ECG clinical evidence (the gold standard for type A diagnoses) in more than 90% of cases. It also suggests that with further improvements in diagnostic software, computerized interpretation of normal ECGs may be accepted in the future without over-reading by a human expert. However, all evidence reviewed cautions that complete reliance on computerized ECG interpretation may result in incorrect diagnoses and could lead to inappropriate management decisions.

The inexperienced physicians in their routine practice may benefit from high-quality computer software with the best performance in resting ECG interpretation, which can provide reasonably accurate diagnosis of normal ECGs and informative backup to improve the accuracy of their interpretations. However, they should not accept the computerized interpretations without questions. The over-reading by an expert is recommended when dealing with uncertain and indefinite diagnoses of normality and suspected abnormalities.

There seem to be two opposing views on the use of computerized ECG interpretation [(9), (11), (21), (22), (24), (25), (36), (37)]. One view emphasizes the medical and liability aspects of the use of computerized ECG interpretation and stresses the necessity of having expert clinicians over-read all computerized ECG interpretations, for all applications and in all clinical settings. The other view focuses on problems of economics and questions whether all ECGs must be reviewed by experts and whether it is worth the cost. Well-designed and conducted studies on diagnostic accuracy, and patient outcome and economic analyses are needed to answer these questions.

Those considering use of computerized ECG interpretation for routine clinical examination of asymptomatic adults in ambulatory clinical settings should be aware that:

- The ECG test is only one of the tests used to detect or exclude possible heart conditions and routine screening ECGs are not warranted for asymptomatic patients without a history of cardiac abnormalities [(10), (38)]. Since the ECG is not particularly sensitive and has modest specificity it is of limited value as a screening tool in an apparently healthy population (Belenkie, personal communication) [(10, 22)].
- There are different applications of computerized ECG interpretation. Different computer programs have been used to interpret ECGs which were performed on different populations, for different cardiac conditions and in different clinical settings.
- The indices of diagnostic accuracy of these programs depend on the composition of the study population and their predictive merits must be examined in relation to different prevalence values for each diagnostic category.
- The computer programs available on the market apply different approaches to diagnostic classification of ECGs and use different terminology.

CONCLUSION

The literature reviewed showed that:

- The interest in using computer programs to interpret all types of ECG tests has risen rapidly over the past decade.
- Although the programs used for computerized ECG interpretation have been recognized to be imperfect, their use has become accepted as providing the less experienced clinicians with an almost immediate, reasonably accurate interpretation to assist them in achieving more accurate interpretations.
- Computerized ECG interpretation might have made a contribution to reducing the burden of analysis of the many ECGs recorded routinely. However, its use has not resulted in major improvement in diagnostic accuracy of human interpretation by experienced reader.

No primary research studies conducted to determine whether computerized interpretation of resting ECG can be considered an accurate and reliable automated laboratory for screening heart conditions in asymptomatic adults as part of their routine clinical examination have been located.

The available evidence did not permit conclusive answers on the diagnostic accuracy and reliability of computerized interpretation of resting ECG as an automated laboratory test for screening normal ECGs in asymptomatic adults. Also the question whether it can replace interpretation by a skilled professional in an ambulatory clinical environment for this application has yet to be answered.

The available evidence (weak and limited) suggests that the computer programs with the best performance may be as accurate as the human reader in diagnosing normal ECGs. However, computerized interpretation of ECGs should be used with an awareness of the risk of false positive and false negative findings.

The literature reviewed suggests that the role of computerized resting ECG interpretation in primary care has future potential. The immediate availability of computerized ECG interpretation has been seen as a significant improvement for practicing clinicians. However, whether its use actually increases physician's accuracy in ECG interpretation, saves physician time, improves quality of patient care and leads to a reduction in the costs associated with ECG interpretation have yet to be determined.

APPENDIX 1: METHODOLOGY

Literature search

An electronic search was conducted for articles/papers pertaining to the subject, which were published in English. In addition, a search of the relevant web sites was conducted. A preliminary search of PubMed MEDLINE and scanning of all results obtained indicated that many changes and developments occurred in the computerized ECG interpretation in the late '80s and early '90s. Thus the subsequent literature search was limited to studies published between 1994 and April 2001. The literature search strategy is summarized in the following table, which lists databases/sources searched, time limits, and the keywords used:

Database/Source (time limits)	Subject Headings/ Textwords
PubMed MEDLINE (to April 19, 2001)	(automated *OR* automatic) *AND* (interpret* *OR* reading*) *AND* (EKG *OR* ECG *OR* electrocardiography) *AND* diagnosis, computer assisted
EMBASE (1994-2000 Dec)	(*autoanalysis *OR* automated.mp. *OR* *automation *OR* *automation, computers and data processing *OR* computer analysis *OR* computer assisted diagnosis) *AN*D exp *electrocardiogram *AND* (interpret$ *OR* reading$)
Best Evidence (1994-2000 Dec)	(ECG *OR* EKG *OR* electrocardiogra$) *AND* (automat$ *OR* computer *OR* machine) *AND* (interpret$ *OR* reading$)
CRD databases: HTA, EED, DARE	(ECG *OR* EKG) *AND* automat$
Cochrane Database of Systematic Reviews (2000 issue 4)	(EKG *OR* ECG *OR* electrocardiogra*) *AND* (automat* *OR* machine *OR* computer*)
HealthSTAR (1994-2000Jan)	exp *electrocardiography *AND* (exp *diagnosis, computer-assisted *OR* exp *diagnosis, computer-assisted) *AND* (interpret$ *OR* reading$)
Web of Science: SSI (1994-2001)	Automat* *AND* (ECG *OR* EKG *OR* electrocardiogra*) *AND* (interpret* *OR* reading*) *NOT* pacemaker*
National Guideline Clearinghouse	ECG *OR* EKG

Database/Source (time limits)	Subject Headings/ Textwords
ECRI website	ECG OR EKG
CCOHTA publications	Browsing
WWW	Companies websites for information regarding their software programs designed for ECG interpretation

The search was focused on articles / papers reporting results on the diagnostic accuracy or on the reliability of computerized interpretation of ECGs done at rest in ambulatory settings (outpatient clinics or doctor's offices) for healthy or asymptomatic adults (aged 18 years and over) in terms of heart problems (with no apparent or suspected heart dysfunction or disease) as part of their routine medical examinations.

During the literature search, the database selection was based upon subject matter. The search was not limited with regards to publication type.

1. Selection of material

For each citation considered, the abstract was read (when available), and articles were excluded if they were outside the scope of the review. Full articles that met the following criteria were retrieved:

- articles reporting the use of computerized / automated ECG interpretation as an automated laboratory test in ambulatory settings to screen for heart conditions in healthy asymptomatic adults during their routine clinical examination;
- articles reporting results of prospective controlled trials (randomized and non-randomized), or other prospective / retrospective comparative studies (with series larger than 10 subjects) in which diagnostic accuracy or reliability of computerized / automated ECG interpretation was compared to that of a gold standard used to diagnose normality in terms of heart condition in healthy, asymptomatic adults;
- articles reporting results of quantitative and / or qualitative reviews on the use of computerized ECG interpretation as an automated laboratory test to screen heart problems in healthy / asymptomatic adults as a part of their routine clinical examination;
- discussion papers and / or commentaries on clinical utilization, advantages, disadvantages, limitations, and other issues associated with the use of computerized ECG interpretation as an automated laboratory

test to screen heart problems in healthy / asymptomatic adults as a part of their routine clinical examination.

Studies / papers / articles were excluded from the review if:

- they reported findings on the use of computerized ECG interpretation as a diagnostic test or monitoring technique done in clinical settings other than ambulatory settings (such as inpatient hospital departments or emergency departments) for adults with suspected or diagnosed heart dysfunction;
- they reported findings on the use of computerized ECG interpretation as a diagnostic test done in ambulatory settings for adults with suspected or diagnosed heart dysfunction;
- they reported findings on the use of computerized interpretation ECG tests other than resting ECG (such as stress or exercise ECG and 24-hour continuing ECG);
- they reported only on the diagnostic accuracy of computerized interpretation of one or more ECG parameters (waveforms, complexes and segments) and not of the overall ECG pattern, including the identification of the heart rhythm;
- they reported on the use of computerized ECG interpretation in detection / diagnosis / monitoring health condition in children and adolescents (aged <18 years);
- they were published only in abstract form.

Editorials, letters, case reports and technical reports were excluded unless they provided pertinent information on the characteristics of the assessed techniques, their cost, advantages and limitations, that was not available elsewhere.

The bibliography of each of the retrieved papers was examined to identify relevant references that could be missed by the electronic search. Articles published before 1994 were quoted when appropriate.

Cardiologists with interest and expertise in ECG interpretation from Alberta were contacted for information on additional sources for material relevant to the subject.

REFERENCES

1. Abreu-Lima C, de Sa JP. Automatic classifiers for the interpretation of electrocardiograms. *Rev Port Cardiol* 1998; 17(5):415-28.
2. Willems JL. Quantitative electrocardiography. Standardization and performance evaluation. *Ann N Y Acad Sci* 1990; 601(1):329-42.
3. RuDusky BM. Errors of computer electrocardiography. *J Vascular Diseases* 1997; 48(12):1045-50.
4. Ong HT, Kuah SH, Chew SP. Reliability of computerised electrocardiographic reports. *Singapore Med* 1993; 34(1):53-54.
5. Willems JL, Abreu-Lima C, Arnaud P, van Bemmel JH, Brohet C, Degani R, et al. The diagnostic performance of computer programs for the interpretation of electrocardiograms. *N Engl J Med* 1991; 325(25):1767-73.
6. Brailer DJ, Kroch E, Pauly MV. The impact of computer-assisted test interpretation on physician decision making: the case of electrocardiograms. *Med Decis Making* 1997; 17(1):80-86.
7. Hillson SD, Connelly DP, Liu Y. The effects of computer-assisted electrocardiographic interpretation on physicians' diagnostic decisions. *Med Decis Making* 1995; 15(2):107-12.
8. Sekiguchi K, Kanda T, Osada M, Tsunoda Y, Kodajima N, Fukumura Y, et al. Comparative accuracy of automated computer analysis versus physicans in training in the interpretation of electrocardiograms. *J Med* 1999; 30(1-2):75-81.
9. Rautaharju PM. Will the electrocardiograph replace the electrocardiographer? *J Electrocardiol* 1993; 26 Suppl:158-63.
10. Health Services Utilization and Research Commission. Anatomy of a practice guideline. Tradition, science, and consensus on using electrocardiograms in Saskatchewan. *Canadian Family Physician* 1995; 41:37-49.
11. Fisch C. Guidelines for electrocardiography. A report of the American College of Cardiology / American Heart Association Task Force on assessment of diagnostic and therapeutic cardiovascular procedures (committee on electrocardiography). *JACC* 1992; 19(3):473-81.
12. Macfarlane PW, McLaughlin SC, Yang TF, Devine B. Recent developments in computer assisted ECG interpretation. *Japanese Heart Journal* 1994; 35(SUPPL. 1):37-43.

13. MacFarlane P, Glasgow Royal Infirmary. Computer applications in clinical care. Computers in cardiology. *http://students.dcs.gla.ac.uk . 2001.*

14. Holzmann C, Hasseldieck U, Rosselot E, Estevez P, Andrade A, Acuna G. Interpretation module for screening normal ECG. *Medical Progress through Technology* 1990; 16:163-71.

15. van Bemmel JH. New trends in computer ECG analysis. *J Electrocardiol* 1997; 29(Supplement):1-4.

16. Syed AZ. Methods used to interpret the 12-lead electrocardiogram: pattern memorization versus the use of vector concepts. *Clin Cardiol* 2000; 23:238.

17. Macfarlane PW. Evolution of the Glasgow program for computer-assisted reporting of electrocardiograms--1964 / 1998. *Acta Cardiol* 1998; 53(2):117-20.

18. McLaughlin SC, Aitchison TC, Macfarlane PW. Methods for improving the repeatability of automated ECG analysis. *Methods Inf Med* 1995; 34(3):272-82.

19. Hurst JW. Are you using the most accurate method for monitoring r-hirudin in patients with HIT during CPB? *Clin Cardiol* 1998; 21(3):211-16.

20. Hurst JW. Methods used to interpret the 12-lead electrocardiogram: pattern memorization versus the use of vector concepts. *Clin Cardiol* 2000; 23:4-13.

21. Grauer K, Curry RW. Computerized ECG interpretation. *Journal of Family Practice* 1991; 35(2):135-36.

22. Mirvis DM. What's wrong with electrocardiography? *J Electrocardiol* 1998; 31(4):313-16.

23. Willems JL, Abreu-Lima C, Arnard P, Brohet CR, Denis B, Gehring J, et al. Evaluation of ECG interpretation results obtained by computer and cardiologists. *Methods Inf Med* 1990; 29(4):308-16.

24. Rodney WM. Electrocardiogram interpretation. *J Fam Pract* 1992; 35(5):496.

25. Woolley D, Henck M, Luck J. Comparison of electrocardiogram interpretations by family physicians, a computer, and a cardiology service. *J Fam Pract* 1992; 34(4):428-32.

26. de Bruyne MC, Kors JA, Hoes AW, Kruijssen ACM, Deckers JW, Grosfeld M, et al. Diagnostic interpretation of electrocardiograms in population-based research: computer program research physicians, or cardiologists? *J Clin Epidemiol* 1997; 50(8):947-52.

27. Kors JA, Crow RS, Hannan PJ, Rautaharju PM, Folsom AR. Comparison of computer-assigned Minnesota Codes with the visual standard method for new coronary heart disease events. *Am J Epidemiol* 2000; 151(8):790-97.

28. Zywietz C, Willems JL. European conformance testing services for computerized electrocardiography. New procedures and standards. *J Electrocardiol* 1993; 26 Suppl(1):137-46.

29. Heden B, Ohlsson M, Rittner R, Pahlm O, Haisty WKJ, Peterson C, et al. Agreement between artificial neural networks and experienced electrocardiographer on electrocardiographic diagnosis of healed myocardial infarction. *J Am Coll of Cardiology* 1996; 28(4):1012-16.

30. Reddy BR, Taha B, Swiryn S, Silberman R, Childers R. Prospective evaluation of a microprocessor-assisted cardiac rhythm algorithm: results from one clinical center. *J Electrocardiol* 1998; 30 Suppl(4):28-33.

31. U.S.Food and Drug Administration - Center for Devices and Radiological Health. Cardiovise ECG interpretative software. *http://www.accessdata.fda.gov/. 4-19-2001.*

32. Pryor WW, Blackhurst DW, Gates K, Wagner GS. Evaluation of computerized electrocardiographic interpretations in a community hospital. *J S C Med Assoc* 1995; 91(3):105-11.

33. Arnaud P, Willems JL, Girard P, Morlet D. A new method to assess the difficulty of a medical diagnosis: application to electrocardiographic interpretation. *Methods of Information in Medicine* 1993; 32:59-65.

34. Schijvenaars BJ, Kors JA, van Herpen G, Kornreich F, van Bemmel JH. Effect of electrode positioning on ECG interpretation by computer. *J Electrocardiol* 1997; 30(3):247-56.

35. McLaughlin SC, Aitchison TC, Macfarlane PW. The value of the coefficient of variation in assessing repeat variation in ECG measurements. *Eur Heart J* 1998; 19(2):342-51.

36. Fisch C. Clinical competence in electrocardiography. A statement for physicians from the ACP / ACC / AHA Task Force on clinical privileges in cardiology. *JACC* 1995; 25(6):465-69.

37. Diamond GA, Forrester JS. Computer interpretation of electrocardiograms. *N Engl J Med* 1992; 326(24):1634-35.

38. Health Services Utilization and Research Commission. *Use of routine electrocardiograms. Guidelines review.* Beckie JL, editor. 1999. Saskatchewan, HSURC.