What ECG Computer Program to Choose for Clinical Application

The Need for Consumer Protection

Although first attempts to automate ECG analysis go back to the late 1950s,1-3 serious clinical application is a development of only the recent past. It appears that at present the greatest need for automated ECG interpretation exists in large city hospitals where the ever-increasing number of ECG recordings is overtaxing the relatively fixed supply of trained electrocardiographers and in rural areas where specialists with adequate expertise are frequently spread out too thin. In these situations, assistance through automation appears desirable and deserves support.

The number of commercial firms providing partially or completely automated ECG interpretations in response to this need has mushroomed to the point where it is estimated that there are close to 50 at this time in the U. S. alone. A variety of competing commercial ECG computer programs is being offered, accompanied by claims of excellent diagnostic performance and embellished with statements on the "infallibility" of the computer. The physician considering computerized ECG analysis must choose between one of the available systems in the midst of these many claims and counter-claims. Conferences on ECG computer analysis have most often increased rather than decreased the prevailing confusion. A further problem presented to potential users deals with so-called "certified" programs. This term is usually used when computer programs are translated for use on computers of different makes. The "certification" indicates that the different program versions perform in a practically identical manner. This may imply in some instances that the errors made by the original program will be faithfully reproduced after program translation.

Certain shortcomings which are common to all presently available ECG computer programs are rarely pointed out. One of them is the inability to identify reliably complex arrhythmias. Simpler rhythm disturbances and ectopic beats can be recognized rather accurately.

Comparisons between old records and most recent ones represent an essential part of clinical electrocardiography. The need for automated comparisons is not filled by providing old computer printouts together with recent ones. To our knowledge, a program for rapid retrieval of old tracings leading to automated quantitative comparisons with interpretation of differences between records has not yet been made available.* Such differences need to be evaluated in the light of day-to-day or year-to-year variability4 of the normal electrocardiogram. Computer reports on comparisons should provide only data which significantly exceed limits of normal variability.

What is needed most urgently at this time are easily comprehensible performance standards which should be used as common yardsticks for systems evaluation. As pointed out very clearly by the late Dr. Charles K. Friedberg, "The ultimate test is diagnostic accuracy."5 Unfortunately, the great majority of published ECG evaluations have not used the Friedberg criterion and are therefore inadequate and often misleading. Most designers of ECG computer programs had to make their own, somewhat arbitrary selection of diagnostic rules which was considered by them, but not necessarily by others, as optimal. Almost without exception, performance tests of these programs were done by their own designers using the same decision rules. As one might expect, "diagnostic accuracy" of such studies reaches 80-95%.6 The relatively small failure rate has to be attributed to technical errors in automatic wave recognition and/or measurement programs. Since the same diagnostic decision rules were followed both by the computer and the human reader, it is clear that "diagnostic accuracy" is not tested at all in such studies.

A second less frequently used type of system evaluation compares results of a computer-based set

*Since preparation of this Editorial, comparison programs have been reported by Pryor et al. (Comput Biomed Res 5: 709, 1972) and by Dunn et al. (Annual Meeting AHA Council on Epidemiology, New Orleans, Louisiana, March 11, 1973).
of diagnostic rules with results of human readers using what may be different diagnostic rules. As had to be expected, percentages of disagreements rose sharply and reached almost 50% in one reported example. Failure of agreement was attributed to various sources such as technically poor recordings, physician error, imprecise measurements, differences in opinion, and imprecise criteria.

A different type of performance evaluation was used by Simonson et al. who tested 10 expert electrocardiographers. They were asked to interpret 105 unknown ECGs from patients whose diagnoses were known and documented by independent means such as cardiac catheterization, cardiac surgery, or autopsy. Correct ECG interpretations achieved averaged only 54%. The question which then has to be raised in interpreting results of a computer-physician comparison is: Whose criteria and opinions were right and whose were wrong, which leads to the most crucial problem: What represents a correct ECG interpretation or diagnosis?

To try to reconcile opinions and criteria of different observers in order to test the performance of ECG computer systems appears to us as an exercise in futility, particularly when more reliable evaluation methods are readily available. Aside from postmortem findings which, although valuable, are not free of shortcomings, many invasive and noninvasive diagnostic procedures are available which can be used as objective evidence for cardiac abnormalities. Cardiac catheterization, ventriculograms, angiograms, and direct examinations during cardiac surgery are in such wide use that lack of objective correlates for ECG computer system evaluation cannot be accepted any more as reason for using exclusively subjective test methods.

Some commercial firms are using human readers to edit and/or modify computer results. In the light of the known limitations in consistency of human interpreters such revised reports are least suitable for objective testing and undesirable.

In order to provide objective yardsticks for evaluation of presently available systems, and to protect the consumer from exaggerated and frequently misleading performance claims, a library needs to be established by an impartial, disinterested organization comprising ECG records from patients with a wide range of cardiac abnormalities whose diagnoses were established by objective, generally accepted nonECG methods. Sensitivity and specificity of any computer program could then be evaluated by comparing the two sets of diagnoses. If desired, independent physician interpretation of the same tracing could be incorporated in the comparison.

To our knowledge, the Cooperative Study of the Veterans Administration on Cardiovascular Data Processing has been the only group so far which has made an attempt to establish an extensive ECG library with documentation of patients' diagnoses derived exclusively from independent information. Since this study deals primarily with orthogonal electrocardiograms, a similar effort for conventional 12-lead records is also needed at this time. Needless to say, both types of records should be obtained at the same time for testing purposes.

The clinical implications of unnecessary overdiagnoses and underdiagnoses need no further elaboration here for those who are involved in direct patient care. It is hoped that one of the main contributions of the computer to clinical electrocardiography will be to provide a more realistic and quantitative evaluation of the diagnostic information content of the ECG which is, because of its indirectness, limited. Improved documentation of these limitations pertaining both to human interpretation and computer analysis can be obtained most efficiently by large-scale tests using the computer as a test tool.

Although rules for objective program evaluation are in principle simple and straightforward, it is surprising how frequently they have been disregarded even in most recent studies where either collaborators of the program designer or a group of independent cardiologists served as judge and jury. As expected, agreement with the computer in the former case was a relatively high 82%. In the second study, evaluation of three different programs led to agreement in 51, 49, and 70%, respectively, and it was concluded that these programs are "not yet ready for routine clinical use." Considering the accuracy rate of slightly more than 50% obtained by expert electrocardiographers in the cooperative study by Simonson et al., one may wonder what the real meaning of such a comparison is. The question which needs to be asked again is: Who was right and who was wrong? Investigations which lack an objective reference standard will never provide us with an answer.

Hubert V. Pipberger
Jerome Cornfield
References


What ECG Computer Program to Choose for Clinical Application: The Need for Consumer Protection

HUBERT V. PIPBERGER and JEROME CORNFIELD

_Circulation_. 1973;47:918-920
doi: 10.1161/01.CIR.47.5.918

_Circulation_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 1973 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/47/5/918.citation

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in _Circulation_ can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to _Circulation_ is online at:
http://circ.ahajournals.org//subscriptions/