The Effects of Computer-assisted Electrocardiographic Interpretation on Physicians' Diagnostic Decisions

STEVEN D. HILLSON, MD, MSc, DONALD P. CONNELLY, MD, PhD, YUANLI LIU, MS

Objective. To evaluate the effect of computer-assisted interpretation of electrocardiograms (ECGs) on diagnostic decision making by primary care physicians. *Design*. Randomized controlled trial. *Setting*. Primary care physicians' outpatient clinics in or near the Minneapolis/St. Paul metropolitan area. *Participants*. Forty family physicians and general internists who were members of either of two large consortia of clinics. *Intervention*. Subjects evaluated ten clinical vignettes accompanied by ECGs and reported their diagnostic impressions. The vignettes were based on actual patient visits. Half of the subjects received ECGs with computer-generated reports, the other half received the same ECGs without reports. *Main outcome measures*. ECG reading time; agreement with the clinical diagnosis; agreement with the computer report; diagnostic confidence. *Results*. The subjects receiving the reports were more likely to agree with the clinical diagnoses of the original cases, particularly for two vignettes in which the diagnoses were uncommon and were mentioned in the reports. The subjects receiving the reports were also more likely to make diagnoses that were consistent with the reports, even when the reports were erroneous. Those receiving the reports spent, on average, 15 seconds less time looking at each ECG, a 25% decrease. *Conclusions*. In simulated cases, primary care physicians appear to use computer interpretations of ECGs when available, as shown by enhanced diagnostic accuracy and modestly reduced time spent reading the tracings. However, erroneous reports may mislead physicians. Since the effects of computerized ECG interpretation do not appear to have been uniformly favorable in this mock-clinical setting, it will be important to evaluate the effects of this technology in actual practice. *Key words*: computer-assisted diagnosis; computer-assisted decision making; physician decision making; electrocardiogram. **(Med Decis Making 1995;15:107–112)**

Computer assistance in medical test interpretation has a variety of goals. Two of the most important are improving the accuracy of medical diagnosis and reducing physicians' time and work demands. Computer-assisted test interpretation (CATI) is particularly well developed for electrocardiograms (ECGs), and it has been suggested that the time needed by physicians for ECG interpretation is indeed reduced when CATI is available for ECGs.[1] However, ECG interpretation time has not been examined in clinical settings, where non-cardiologists frequently evaluate ECGs in the context of patient encounters. Furthermore, while several studies indicate important limitations in the accuracy of CATI systems for interpreting ECGs,[2–7] the effects of the CATI interpretations on physician decisions have not been reported. This randomized prospective trial was designed to examine the effects of CATI on physician diagnostic-decision making in simulated clinical problems.

Received December 20, 1993, from the Department of Medicine, Ramsey Clinic, St. Paul, Minnesota (SDH), and the Institute for Health Services Research (SDH) and the Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis (DPC); and Harvard University, Boston, Massachusetts. Revision accepted for publication July 18, 1994. Presented in part at the annual meeting of the Society for Medical Decision Making, October 1992. Supported entirely by Federal funds under HCFA Cooperative agreement 99-C-99169/5. The statements contained in this paper are solely those of the authors and do not necessarily reflect the views or policies of the Health Care Financing Administration.

Address correspondence and reprint requests to Dr. Hillson: Department of Medicine, Ramsey Clinic, 640 Jackson Street, St. Paul, MN 55101.

Methods

SUBJECTS

We contacted primary care physicians who identified themselves as either family practitioners or general internists, and who were members of either of two large consortia of clinics in the Minneapolis/St. Paul metropolitan area. The subjects were informed that the study was examining physician clinical decision making and that all responses were confidential. Those who agreed to participate were scheduled for a face-to-face 30-minute appointment with one of the inves-

tigators (YL). Participants were compensated for the time they spent on the study.

CASE STUDIES

We constructed ten clinical vignettes based upon actual episodes of care. The vignettes were selected using a predetermined list of ECG findings that might be encountered in a primary care setting. Each vignette consisted of a few sentences describing pertinent history, physical examination, and relevant laboratory findings. For each vignette, the ECG obtained at the time of the clinical encounter was obtained, including the original computer-generated report.

Three of the ten vignettes were cases in which the computer-generated reports were wrong—that is, the reports misinterpreted some features of the ECGs. We used these erroneous reports in the vignettes in order to duplicate the information that would have been immediately available to a clinician in practice. This allowed us to examine the potential impact of imperfections in the CATI-generated interpretations.

PROTOCOL

The subjects were randomly assigned to a CATI-supported or an unsupported condition. In the CATI-supported condition, subjects received the clinical vignettes, the accompanying ECGs, and the computer interpretations of the ECGs (displayed at the top of each ECG in a typical fashion). In the unsupported condition, subjects received the same clinical vignettes and ECGs, but the interpretations were removed. Each ECG was on a separate page stapled to, and immediately following, the corresponding vignette. This effectively forced the subject to turn the page in order to see the ECG.

The subjects were asked to examine each clinical vignette and accompanying ECG, and then to answer three questions about the vignette:

1. What diagnosis do you think explains the patient's problem?
2. How confident are you in this diagnosis? [Marked on a ten-point Likert scale where 1 = "Not at all sure" and 10 = "Very sure"]
3. What other diagnoses are you considering?

The interviewer recorded the total time, in seconds, that each subject spent looking at each ECG.

At the conclusion of the ten case vignettes, the subject was asked about his or her experience with computer-generated ECG reports. In addition, demographic information was recorded about each subject.

Two internists, blinded to the subject's group (CATI vs non-CATI), independently scored the subject's stated diagnosis (the "subject's diagnosis") as agreeing or disagreeing with the final diagnosis from the actual clinical encounter (the "clinical diagnosis"). They also scored the subject's diagnosis as being consistent or inconsistent with the computer-generated interpretation (the "CATI report"). The two scorers then met to reconcile differences in scored agreements.

The protocol was reviewed and approved by the University of Minnesota's Committee on the Use of Human Subjects in Research.

ANALYSIS

Student's t-test was used to compare times spent reading the ECGs and diagnostic confidence under the two conditions. Subsequently, multiple linear regression was used to control for vignette number, physician specialty, gender, years since graduation from medical school, prior experience with CATI, and attitude toward CATI. To account for non-normality of the time measurement, a logarithmic transformation was used.

Pearson's chi-square was used to examine whether receiving the computer interpretation affected agreement between the subject's diagnosis and either the clinical diagnosis or the CATI report. Logistic regression was then used to re-evaluate the effects of CATI, controlling for the other variables noted above.

The standard errors for both the logistic and the linear regression models were corrected for cluster effects using a modification of the Huber standard-error calculation.[8]

Results

PARTICIPANTS

Forty-two potential subjects were contacted. One refused to participate because of time constraints; another agreed to participate, but left for a vacation before the study session could be completed. Characteristics of the 40 participants are listed in table 1. Of the 40 participants, 31 were male, and 38 were specialty-board–certified (30 in family medicine, eight in internal medicine). All but five were full-time practitioners. The participants ranged in age from 28 to 62 years; an average of 11 years had elapsed since their graduation from medical school.

ECG READING TIME

The physicians spent from as little as 5 seconds to as long as 248 seconds examining the individual ECGs. The mean time spent was 54 seconds per ECG (table 2). The physicians who received the computer-generated reports with the ECGs spent an average of 46.7 seconds per ECG, while those who did not receive

reports spent 61.9 seconds ($p < 0.001$). This 15-second decrease was unchanged after controlling for physician characteristic and vignette number, and for cluster effects.

DIAGNOSTIC CONFIDENCE

The electrocardiogram interpretations did not appear to alter the physicians' confidence in their diagnoses. The mean confidence score for the physicians receiving reports was 5.8, while the mean confidence score for those not receiving reports was 6.0 ($p > 0.3$).

Confidence was unrelated to whether the subject's diagnosis was consistent with the computer report. Among those who received the CATI reports, confidence scores were 5.8 for those whose interpretations were consistent with the CATI reports, and also 5.8 for those whose interpretations were not consistent with the CATI reports. Confidence was also unrelated to whether the subject's diagnosis agreed with the clinical diagnosis.

AGREEMENT WITH CLINICAL DIAGNOSIS

Overall, the proportion of cases in which the physician's first-listed diagnosis agreed with the clinical diagnosis was rather low, at 23%. This may have reflected the very sparse information provided in the clinical vignette. The subjects who received CATI support agreed with the clinical diagnoses in 63 of 209 cases (30.1%), while those who did not receive support agreed with the clinical diagnoses in only 29 of 190 cases (15.3%). This difference persisted after controlling for physician characteristics, cluster effects, and the vignette number, and was significant at $p = 0.004$.

Agreement with the clinical diagnosis was highest when subjects received accurate CATI reports (54 agreements of 146 cases, 37%), intermediate when subjects received either no report (29 agreements of 190 cases, 15%) or inaccurate CATI reports (9 agreements of 63 cases, 14%) ($p < 0.001$).

Table 1 • Baseline Characteristics of 40 Primary Care Physicians Who Interpreted ECGs with and without Accompanying Computer-generated Reports

	Computer-assisted Test Interpretation		
	Yes ($n = 21$)	No ($n = 19$)	p
Age (mean ± SD)	38.6 ± 7.2 years	38.1 ± 4.4 years	NS
Female gender	6/21 (29%)	3/19 (16%)	NS
Specialty			
Family practice	18/21 (86%)	11/19 (58%)	NS
Internal medicine	3/21 (14%)	8/19 (42%)	
Period since graduation (mean ± SD)	11.8 ± 6.5 years	10.2 ± 4.5 years	NS
Report experience*	17/21 (81%)	15/19 (79%)	NS

* The physician at least sometimes receives reports on electrocardiograms in his or her usual practice.

Table 2 • Effects of Computer-assisted Test Interpretation on Major Outcomes of 40 Primary Care Physicians' Interpretations of Electrocardiograms

	Computer-assisted Test Interpretation (CATI)		
Outcome	Yes ($n = 21$)	No ($n = 19$)	p
Time per ECG interpretation (mean ± SD)	46.7 ± 25.7 sec	61.8 ± 41.6 sec	<0.001
Confidence*	5.8 ± 2.3	6.1 ± 2.3	NS
Agreement with clinical diagnosis			
CATI correct	54/146 (37.0%)	29/190 (15.3%)	<0.001
CATI erroneous	9/63 (14.3%)		
Agreement with computer report	109/209 (52.1%)	64/190 (33.7%)	<0.001

* The physician's own confidence in the proposed diagnosis, measured on a ten-point Likert scale, where 1 = "Not at all sure" and 10 = "Very sure."

AGREEMENT WITH CATI REPORT

Overall, the physicians' first-reported diagnoses were consistent with the CATI reports in 173 of 399 cases (43.3%). The diagnoses reported by those who received CATI support were more likely to be consistent with the CATI reports (109 of 209 cases, 52.2%) than were the diagnoses of those who did not receive such support (64 of 190, 33.7%). Again, this effect persisted after controlling for other characteristics ($p < 0.0001$).

MISLEADING COMPUTER REPORTS

Computer-generated ECG interpretations vary in accuracy. Incorrect assessment of the cardiac rhythm is a particularly common difficulty for many systems. In three of the case vignettes the initial computer reports were in error and required subsequent modification by an electrocardiographer. We intentionally gave the initial, erroneous reports to the study subjects, and analyzed these three cases individually to determine whether physicians might have been misled by the erroneous reports.

In the first case, the computer interpreted the rhythm as junctional, when the correct rhythm was sinus. Among the 21 subjects receiving the erroneous CATI report, only five correctly indicated that the rhythm

Table 3 • Agreements with the Clinical Diagnoses, by Vignette, of 40 Primary Care Physicians' Interpretations of Electrocardiograms

Clinical Diagnosis	Agreement CATI*	Non-CATI†	p
Vignette 1 Pulmonary hypertension	10/21 (48%)	6/19 (32%)	NS§
Vignette 2 Supraventricular tachycardia‡	5/21 (24%)	3/19 (16%)	NS
Vignette 3 Hyperkalemia‡	2/21 (10%)	4/19 (21%)	NS
Vignette 4 Wolff–Parkinson–White syndrome	8/21 (38%)	0/19 (0%)	0.004
Vignette 5 Left ventricular aneurysm and ventricular tachycardia	1/20 (5%)	0/21 (0%)	NS
Vignette 6 AV nodal tachycardia‡	2/21 (10%)	1/19 (5%)	NS
Vignette 7 ASCVD, atrial fibrillation, and hyperkalemia	7/21 (33%)	6/19 (32%)	NS
Vignette 8 ASCVD with pacemaker	9/21 (43%)	8/19 (42%)	NS
Vignette 9 Pericarditis	15/21 (71%)	1/19 (5%)	0.001
Vignette 10 Subarachnoid hemorrhage	4/21 (19%)	0/19 (0%)	NS

*Computer-assisted test interpretation.
†No computer-assisted test interpretation.
‡Vignettes 2, 3, and 6 had erroneous CATI reports.
§NS = statistically nonsignificant, $p > .05$.

was sinus. Among the 19 subjects who did not see a report, 11 indicated sinus rhythm. This difference is statistically significant ($p < 0.05$).

In the second case, the computer reported the electrocardiogram as probably normal for age, failing to note very prominent T waves, a possible clue to the correct diagnosis of hyperkalemia. Two of the 21 subjects who received the computer report made the correct diagnosis; four of the 19 who were not thus supported made this diagnosis. This difference is not statistically significant ($p > 0.3$).

In the third case, the CATI report was "supraventricular tachycardia" when the rhythm was atrial flutter. Only three participants specifically reported the rhythm as atrial flutter, one in the CATI-supported group and two in the unsupported group ($p > 0.4$ by Fisher's exact test).

ALTERNATIVE DIAGNOSES

The physicians reported zero to nine alternative diagnoses (in addition to their leading diagnoses), with a mean of 2.2 alternative diagnoses per case vignette. Those who received CATI support generated 2.3 alternative diagnoses per vignette, compared with 2.0 alternatives for those who did not receive this support ($p < 0.05$). After adjusting for physician characteristics and vignette number, the physicians who received CATI support reported 0.68 more alternative diagnoses than did those who did not receive CATI support.

ANALYSIS OF INDIVIDUAL VIGNETTES

Examination of the individual vignettes showed that the effects of the computer interpretations on agreement with the clinical diagnoses varied by vignette. Table 3 shows the marked difference in agreements and effects of CATI support. The most notable impact of CATI occurred for vignettes 4 (Wolff–Parkinson–White syndrome) and 9 (pericarditis). In each case, the correct diagnosis was mentioned in the CATI report. For these two vignettes combined, only one correct diagnosis was made without support, while 23 correct diagnoses occurred with support.

ATTITUDES AND EXPERIENCE WITH COMPUTER ECG INTERPRETATION

Most of the subjects were familiar with receiving ECGs that already had interpretations placed on them. Eight of the subjects stated that they never received such reports, while 17 said they always did, six usually did, and nine sometimes did. Among the 32 subjects who at least sometimes received reports with their ECGs, 26 said that these reports were generated by computers.

When asked whether having a report on the ECG makes it easier to care for patients, 31 of the participants said yes, three said no, and six were unsure. Many expressed the opinion that the report could sometimes make care easier, but that it could also be misleading.

Discussion

The use of computers to interpret ECGs has risen rapidly over the last 15 years.[9] This technology makes interpretations immediately available to the clinician. In the absence of this technology formal an interpretation may be available later from a physician—often a cardiologist—who is experienced in reading ECGs, but such interpretations may be delayed by several days, limiting their usefulness for making clinical decisions. Although the algorithms used in generating the reports are recognized to be imperfect,[3] their immediate availability has been seen as a significant improvement for practicing physicians.[2]

This study demonstrates that the effects of computer-assisted interpretation on clinical practice and decision making may be complex.

In the group of primary care physicians we studied, CATI allowed the physicians to reduce the time spent reading the individual ECG by about 15 seconds, a 25% decrease compared with working without the CATI reports. Although the effect is strongly statistically significant, its clinical meaning is not clear. Certainly if clinicians read dozens of ECGs each day the time savings attributable to CATI could be important. However,

a clinician who reads only a few ECGs could expect only a modest reading-time benefit, perhaps less than a minute a day. This study did not assess the potential time savings attributable to the automated printing of the report.

Perhaps more important than its effect on ECG reading time is CATI's potential effect on physician diagnostic accuracy. In this study the physicians who received CATI reports were more than twice as likely to arrive at the correct diagnoses as the physicians who did not receive the interpretations. The effect was particularly prominent for two case vignettes where the correct diagnoses were missed by nearly everyone who did not receive the interpretations. Indeed, subset analysis showed no significant effect of CATI on diagnostic accuracy for the other eight vignettes, individually or combined. The two case vignettes that were substantially affected by CATI represented somewhat unusual syndromes that were correctly identified by the CATI reports (Wolff–Parkinson–White syndrome and pericarditis). This suggests that one role of a CATI system may be to remind clinicians of uncommon diagnoses, a function common to other medical computing applications.[10]

The effects of CATI on diagnostic accuracy may not be entirely benign, however. Three of our case vignettes included CATI reports that were misleading. These reports were the actual CATI statements made on real ECGs, and represent true limitations in the accuracy of CATI algorithms. In one of these three cases the subjects who received the erroneous reports were significantly more likely to make a corresponding diagnostic error. When such errors occur there is a clear potential for unnecessary or inappropriate testing and treatment. Various investigators have examined the types and frequencies of errors made by CATI systems.[3,5,11,12] It will be important to consider the risks imposed by these errors when developing policies regarding expert oversight of CATI reports. Of course, even expert electrocardiographers disagree about ECG interpretations, making a "perfect" CATI system virtually impossible to design.[7]

It is important to recognize that this study is limited by the artificial and simplified nature of the vignette task. In practice, a physician sees a real patient with complex cues that are elicited over several minutes or hours. The physician can consult colleagues or medical references. The diagnostic decisions may be made while writing or dictating a comprehensive note, or the clinician may choose not to state a diagnosis at all following the initial contact, but wait for further information. In the vignettes, the subjects had limited information and could not ask for more. They received an ECG whether they wanted it or not, and although there was no formal time limit, most participants were quite busy clinicians and could not allow more than 30 minutes to complete all ten vignettes.

This limitation is important given the relatively low agreement between the subjects' diagnoses and the clinical diagnoses. We suspect that the subjects' diagnostic performances on the real patients corresponding to the clinical vignettes would have been much better. This would attenuate the potential benefit from CATI in real practice. In fact, a clinician who performs perfectly without the reports would not benefit from CATI at all (though erroneous CATI could still induce errors).

Although the differences between clinical vignettes and real patients are important, there are substantial similarities as well. In a series of studies of diagnostic decision making, Elstein et al. showed surprising similarities in the cognitive processing of clinical information regardless of whether real patients, actors, videotapes, or case vignettes were used.[13] In particular, the generation of diagnostic hypotheses seemed to occur similarly in all settings. Thus, although our findings need to be validated in actual clinical practice, they may represent real phenomena that reflect the role of CATI in determining physician decision making.

This study used case vignettes that included some unusual clinical entities, and did not attempt to reproduce the actual range of cardiac and non-cardiac problems that primary care clinicians are faced with in ordinary practice. Further, it is not likely that as many as three of ten CATI reports are significantly in error in routine clinical practice. Thus, these findings must be seen as illustrating the potential effect of CATI, but not necessarily the actual effect.

It must also be recognized that while the physician's diagnostic decision is extremely important, the management decision is also of great interest, since this can affect both patient outcome and use of medical resources. This study did not attempt to assess how CATI may affect such management decisions. This study also did not assess effects of CATI on the decisions of other types of clinicians, such as emergency physicians and cardiologists, who may routinely encounter computer interpretations of ECGs.

We conclude that for clinical case vignettes, physicians appear to use computer-generated ECG reports when they are available, even when those reports are in error. This is reflected in the findings that physicians spend less time looking at the ECGs and arrive at different diagnoses when CATI reports are present. Since these effects are not uniformly favorable, it may be important to assess the impact of such test reports in clinical practice.

The authors gratefully acknowledge Leonard A. Schlossberg, MD, for technical assistance, Willard G. Manning, PhD, for assistance with analytical modeling, and Lawrence Widman, MD, and two anonymous reviewers for comments on early versions of the manuscript.

References

1. Sridharan MR, Flowers NC. Computerized electrocardiographic analysis. Mod Conc Cardiovasc Dis. 1984;53:37–41.
2. Grauer K, Kravitz L, Ariet M, Curry RW, Nelson WP, Marriott HJL. Potential benefits of a computer ECG interpretation system for primary care physicians in a community hospital. J Am Board Fam Pract. 1989;2:17–24.
3. Kudenchuk PJ, Ho MT, Weaver WD, et al. Accuracy of computer-interpreted electrocardiography in selecting patients for thrombolytic therapy. J Am Coll Cardiol. 1991;17:1486–91.
4. Thomson A, Mitchell S, Harris P. Computerized electrocardiographic interpretation: an analysis of clinical utility in 5110 electrocardiograms. Med J Aust. 1989;151:428–30.
5. Willems JL, Abreu LC, Arnaud P, et al. The diagnostic performance of computer programs for the interpretation of electrocardiograms. N Engl J Med. 1991;325:1767–73.
6. Zatti M, Guidi G, Marcolini F. Computer assisted laboratory diagnosis: a ten-year experience. Int J Biomed Comput. 1988;23:125–35.
7. Woolley D, Henck M, Luck J. Comparison of electrocardiogram interpretations by family physicians, a computer, and a cardiology service. J Fam Pract. 1992;34:428–32.
8. Huber PJ. The behavior of maximum likelihood estimates under non-standard conditions. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. 1967;1:221–33.
9. Drazen E, Mann N, Borun R, Laks M, Bersen A. Survey of computer-assisted electrocardiography in the United States. J Electrocardiol. 1988;suppl:S98–S104.
10. Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. DXplain. An evolving diagnostic decision-support system. JAMA. 1987;258:67–74.
11. Elko PP, Weaver WD, Kudenchuk P, Rowlandson I. The dilemma of sensitivity versus specificity in computer-interpreted acute myocardial infarction. J Electrocardiol. 1992;24 suppl: 2–7.
12. Shirataka M, Miyahara H, Ikeda N, Domae A, Sato T. Evaluation of five computer programs in the diagnosis of second-degree AV block. J Electrocardiol. 1992;25:185–95.
13. Elstein A, Shulman L, Sprafka SA, et al. Medicine Problem Solving. Cambridge, MA: Harvard University Press, 1978.