Enhancement of Clinicians' Diagnostic Reasoning by Computer-Based Consultation: A Multisite Study of 2 Systems

Charles P. Friedman; Arthur S. Elstein; Fredric M. Wolf; et al.


http://jama.ama-assn.org/cgi/content/full/282/19/1851

Correction

Correction is appended to this PDF and also available at http://jama.ama-assn.org/cgi/content/full/jama;285/23/2979

Contact me if this article is corrected.

Citations

This article has been cited 47 times.

Contact me when this article is cited.

Related Articles published in the same issue

November 17, 1999

Enhancement of Clinicians’ Diagnostic Reasoning by Computer-Based Consultation
A Multisite Study of 2 Systems

Charles P. Friedman, PhD  
Arthur S. Elstein, PhD  
Fredric M. Wolf, PhD  
Gwendolyn C. Murphy, PhD  
Timothy M. Franz, PhD  
Paul S. Heckerling, MD  
Paul L. Fine, MD  
Thomas M. Miller, MD  
Vijoy Abraham, MS

Computer-based decision support systems (DSSs) seek to improve the quality of health care by providing accurate, useful, and timely advice to clinicians. In a typical DSS, an explicit representation of medical knowledge is applied to the specific circumstances of a case to provide advice to clinicians regarding the diagnosis or management of that case. Over the past 3 decades, several DSSs have addressed medical diagnosis; however, the value of these systems to clinical medicine remains an open question. In part, the question remains because diagnostic DSSs have been evaluated with emphasis on the computer system itself, without placing clinicians in the role of direct users. With this focus, the question of primary interest is whether consultation with DSSs improves the diagnostic hypotheses of clinicians.

Context  Computer-based diagnostic decision support systems (DSSs) were developed to improve health care quality by providing accurate, useful, and timely diagnostic information to clinicians. However, most studies have emphasized the accuracy of the computer system alone, without placing clinicians in the role of direct users.

Objective  To explore the extent to which consultations with DSSs improve clinicians’ diagnostic hypotheses in a set of diagnostically challenging cases.

Design  Partially randomized controlled trial conducted in a laboratory setting, using a prospective balanced experimental design in 1995-1998.

Setting  Three academic medical centers, none of which were involved in the development of the DSSs.

Participants  A total of 216 physicians: 72 at each site, including 24 internal medicine faculty members, 24 senior residents, and 24 fourth-year medical students. One physician’s data were lost to analysis.

Intervention  Two DSSs, ILIAD (version 4.2) and Quick Medical Reference (QMR; version 3.7.1), were used by participants for diagnostic evaluation of a total of 36 cases based on actual patients. After training, each subject evaluated 9 of the 36 cases, first without and then using a DSS, and suggested an ordered list of diagnostic hypotheses after each evaluation.

Main Outcome Measure  Diagnostic accuracy, measured as the presence of the correct diagnosis on the hypothesis list and also using a derived diagnostic quality score, before and after consultation with the DSSs.

Results  Correct diagnoses appeared in subjects’ hypothesis lists for 39.5% of cases prior to consultation and 45.4% of cases after consultation. Subjects’ mean diagnostic quality scores increased from 5.7 (95% confidence interval [CI], 5.5-5.9) to 6.1 (95% CI, 5.9-6.3) (effect size: Cohen d = 0.32; 95% CI, 0.23-0.41; P < .001). Larger increases (P = .048) were observed for students than for residents and faculty. Effect size varied significantly (P < .02) by DSS (Cohen d = 0.20; 95% CI, 0.08-0.32 for ILIAD vs Cohen d = 0.45; 95% CI, 0.31-0.59 for QMR).

Conclusions  Our study supports the idea that “hands-on” use of diagnostic DSSs can influence diagnostic reasoning of clinicians. The larger effect for students suggests a possible educational role for these systems.
the extent to which the system improves the diagnostic hypotheses of clinicians, not the extent to which its advice is "correct."

This approach mirrors the evolving concept of the role DSSs can play in clinical practice. In the 1970s and 1980s, these systems were largely conceived as "oracles," with clinicians seen as passive recipients of the systems' advice.15,16 Over time, however, this view of DSSs was seen as too narrow and mechanistic.15 Diagnoses would continue to be made by people, not machines, and a successful DSS must establish a productive partnership with the clinician. From this perspective, several new issues arise to direct the evaluation of DSSs.

First, a clinician's own medical knowledge plays a critical role in a DSS consultation. Revised diagnoses resulting from consultations are a joint function of what the clinician knows and whatever information is provided by the DSS.17,18 Diagnostic DSSs may or may not have the "intelligence" to offer suggestions useful to experienced clinicians on difficult cases that these clinicians cannot diagnose when unassisted. It is unclear how useful such systems will be to medical students who must integrate a system's advice with their own limited knowledge base.

Second, variation in the ways clinician-users might interact with the DSS becomes important. The system's advice, and thus its potential value, depends on how users can convey to the DSS their personal understanding of a case by selectively entering clinical findings and choosing specific system features.19,20

Third, a DSS consultation may have both beneficial and detrimental effects on a clinician's reasoning. The DSS may offer persuasive advice in the form of an appealing but incorrect diagnosis. If this incorrect advice is accepted or even seriously considered by the clinician, the system's effect may actually be detrimental.

Recent systematic reviews21,22 of computer-based DSSs indicate that previous studies have largely subscribed to the oracle model and have not addressed the issues listed above. With DSSs now distributed commercially on CD-ROM and over the Internet, it is important to deepen understanding of their potential to assist clinicians more directly.23,24

The central question guiding our investigation was: To what extent can consultations using diagnostic DSSs, with clinicians in training or practice as system users, improve the quality of diagnostic hypotheses over a set of challenging cases? We examined this question with 2 mature DSSs and subjects at 3 levels of experience from 3 medical centers; none of the centers was the development site of either system. We hypothesized that effects of DSS consultations would depend on the subjects' experience levels, with greater effects for less experienced subjects. We had no a priori expectations regarding the DSSs, but recognized that substantial differences in their design could lead to differing effects. The nature of the research design required that we also consider the extent to which observed effects on diagnostic reasoning were attributable to DSS advice vs rethinking about the case.

METHODS

We used an experimental procedure to obtain clinicians' diagnostic hypotheses on an assigned set of cases both before and after DSS consultation. The effect of the consultation was determined by comparing subjects' preconsultation and postconsultation diagnostic hypotheses. A balanced research design allowed exploration of these effects in relation to clinicians' experience level and DSS used. New quantitative measures of the quality of diagnostic hypotheses, designed to be sensitive to subtle but potentially significant changes in diagnostic reasoning, were developed and used in this work.

Study Sites

The study was based at 3 academic medical centers: the University of Illinois (Chicago), the University of Michigan (Ann Arbor), and the University of North Carolina (Chapel Hill). Researchers at each site included a principal investigator experienced in medical informatics, a general internist coinvestigator, and collaborators responsible for subject training, recruitment, and data collection. (Two of the principal investigators and 1 collaborator moved to new institutions prior to completion of the study. These relocations had no effect on data collection procedures.)

Case Materials

The physician coinvestigators, working as a team, directed the identification and wrote the summaries of 36 cases (12 from each site) used in the study. All cases were based on actual patients. We sought difficult cases so that DSS consultations would have potential to engender improvement.

Across the sites, the coinvestigators identified 58 cases with known diagnoses based on a definitive test, clinical follow-up, or autopsy, and perceived to be diagnostically challenging. Four of these cases had final diagnoses outside the DSS knowledge bases and were eliminated. Further considerations of breadth and redundancy of diagnoses narrowed the candidate list to 43 cases.

For each remaining case, a clinician coinvestigator at the site of origin wrote a 2- to 5-page summary including patient history and physical findings, laboratory results, and radiological and other diagnostic studies. The summaries also included ample nonsalient data to avoid cueing. The person who wrote the abstracts deleted from the summary any known findings, such as a positive biopsy, that would have made the diagnosis trivial for clinicians and probably for the DSSs as well. These deletions (mean: 1.7 items per case) were made without reference to either DSS.

The 3 coinvestigators subsequently reviewed and rated all case summaries for perceived difficulty using a 7-point scale. Review of the averaged ratings led to elimination of 6 cases judged to be insufficiently challenging, and 1 judged to be too difficult. The 36 cases remaining for use in the study were divided into 4 balanced clusters of 9 cases each. The clusters contained 3 cases from...
each site, included a variety of organ system etiologies, and were equated for perceived difficulty using the aggregated ratings.

Decision Support Systems
We selected ILIAD and Quick Medical Reference (QMR), 2 diagnostic DSSs that were mature, well described in the literature, and available commercially for use by physicians in training and practice.5,8,11,16,25,26,27 Both systems offered sophisticated graphical interfaces to facilitate their use and generated user interaction logs as automated tools for data collection.

ILIAD's knowledge representation derives from statistical data used in concert with expert knowledge of clinicians that were expressed as rules.8 In consultation mode, users enter clinical findings about a case using an interface that allows both free-text and menu-based entry. The system generates a rank-ordered list of diagnostic hypotheses, each with estimated probability. ILIAD can suggest next steps in a work-up that would clarify the differential. Users can also browse ILIAD's representation of each disease. The version of ILIAD (4.2 for Macintosh) used in this study contained explicit representations of 920 diseases.

Knowledge representation of QMR is derived from systematic review of the published literature supplemented by the expert knowledge of selected clinicians.3 Disease representations of QMR are not statistical; relationships between findings and diseases are expressed on heuristic 5-point scales. The QMR can be used in a case analysis mode to generate a ranked list of potential diagnoses for an entered set of case findings. The system offers several special functions, such as comparison and contrast of pairs of diseases, designed to help clinicians refine their diagnoses. The version of QMR (3.7.1 for Windows) used in this study contained explicit representations of 623 diseases.

Subjects and Training
Data were collected from 1995 to 1998. At each of the 3 sites and for each of the 2 DSSs, we recruited 12 faculty physicians, all general internists; 12 internal medicine residents, either late in their second training year or early in their third year; and 12 fourth-year medical students. The faculty had at least 2 years of postresidency clinical experience (mean, 11 years; range, 2-32 years). Subjects were offered modest stipends commensurate with experience level ($200 for faculty, $150 for residents, $50 for students). All subjects were volunteers. By self-report on a questionnaire completed prior to data collection, 7 of the 216 subjects (3 faculty, 4 residents) reported regular use of DSSs. At one institution (Michigan) individual subjects were formally consented into the study. At the other institutions, the research was considered exempt and thus implied consent was obtained from those agreeing to participate.

Each subject was individually trained on his/her assigned DSS. A standardized training protocol, with a checklist of competencies, was used at all sites. The trainers documented mastery of all competencies and assisted subjects in working 3 practice cases prior to the start of data collection.

Procedure
Each subject was assigned randomly to a case cluster, with assignment balanced such that clusters were equally represented across sites and experience levels. For each case in the assigned cluster, the subject was first asked to read the summary and then to generate a list of up to 6 ordered diagnostic hypotheses. The subject then used the DSS to explore the case in any way he/she considered potentially helpful. After using the DSS, the subject generated another diagnostic hypothesis list. The subject then moved onto the next assigned case. We retained computer log files of user entries and the diagnoses proposed by the DSS for each case.

All data collection sessions were proctored and no time limits were imposed on subjects. Median time to complete the initial work on a case, without the DSS, was 8 minutes (semi-interquartile range, 5-10 minutes). Median time for the second iteration, using the DSS, was 22 minutes (semi-interquartile range, 15-30 minutes).

Scoring Metrics
We used 2 measurements for assessing subjects’ diagnostic hypothesis lists. The first, a binary measure, credited the subject with a correct diagnosis if the correct diagnosis—or 1 considered almost synonymous (eg, polymyalgia rheumatica vs giant cell arteritis)—appeared anywhere in the subject’s list.

We also developed and validated a continuous diagnostic quality score that would be sensitive to more subtle but potentially important changes in the quality of subjects’ diagnostic reasoning. For a subject’s ordered list of diagnostic hypotheses, the quality score was composed of 2 components. The first component awarded up to 7 points based on the plausibility of each diagnosis listed, whether correct or incorrect, as judged by consensus of the clinical coinvestigators. The second awarded up to 6 points based on the location of the correct diagnosis, if present. The resulting metric awarded a maximum of 13 points for a perfect hypothesis list comprising only the correct diagnosis, and 1 point for a list comprising only irrelevant diagnoses. This metric, described in detail elsewhere, has been found to be reliable and valid.28

Analysis
Three methods of analysis provide complementary views of the results. The first used the binary measure with cases as the unit of analysis to provide a directly interpretable portrayal of DSS effects on the subjects’ diagnostic reasoning. We calculated the fractions of cases in which the correct diagnosis appeared anywhere in the subjects’ hypothesis lists, and compared these fractions before and after DSS consultation. These data were analyzed separately for the 3 clinical experience levels and each DSS.

The second method used the more sensitive diagnostic quality scores as the outcome variable and subjects as the unit of analysis. Over the 9 cases each sub-
subject completed, we averaged separately the preconsultation and postconsultation quality scores and conducted on these data a 3-way mixed-model analysis of variance. For this analysis, occasion (preconsultation vs postconsultation) was a within-subjects factor. Clinical experience level and assigned DSS were between-subjects factors. We selectively used paired t tests to explore differences between subgroups. Effect sizes were computed as standardized differences between group means and were expressed using the Cohen d statistic. Statistical power to detect medium effects (Cohen d = 0.5; P < .05) was estimated at 0.95 for the all-subjects comparison of diagnostic accuracy scores before and after consultation, at 0.74 for differences across experience levels, and at 0.90 for differences between DSSs. While subjects' sites and assigned case clusters were also factors in the experiment, these factors did not affect the primary results and were not included in the analysis reported here.

The third analysis sought to elucidate whether the effects of consultations might be attributed to each DSS's advice. Using system log files, we reproduced the diagnostic hypothesis list displayed by each DSS during each subject's work on each case. We differentiated cases in which the DSS displayed the correct diagnosis among its top 20 hypotheses from cases in which the correct diagnosis was not displayed. In the former cases, the correct diagnosis was there to be seen; thus, the DSSs had significant potential to be helpful. In the latter cases, the value of each DSS's advice was more doubtful. We then compared preconsultation with postconsultation changes in diagnostic quality scores for these 2 subsets of cases. If the changes were comparable for the 2 subsets, this would suggest that the changes were due to sources other than each DSS's advice. If the changes were greater for the cases where the DSSs had greater potential to be helpful, this would argue that the particular system's advice was the causal factor.

RESULTS
Based on Presence of the Correct Diagnosis

The complete data set included 1934 cases generated by 215 subjects. (All data that were generated by 1 faculty subject and the data for 1 student were not properly recorded and were thus lost to the analysis.) A correct diagnosis appeared in subjects' hypothesis lists for 764 cases (39.5%) before DSS consultation, increasing to 879 cases (45.4%) after consultation (Table 1). Positive consultations, where the correct diagnosis was present after consultation but not before, were observed for 232 cases (12.0%); negative consultations, where the correct diagnosis was present before consultation but not after, were observed in 117 cases (6.0%). The overall consultation effect (net gain) is 115 cases (5.9%). Preconsultation performance, based on subjects' personal knowledge only, increased with experience level. The largest consultation effects were observed for the students, with smaller effects for residents and faculty. Larger consultation effects were observed in subjects using QMR.

### Based on the Diagnostic Quality Score

Table 2 provides mean preconsultation and postconsultation diagnostic quality scores and consultation effect sizes, broken down by subjects' experience level and DSS used. Statistical main effects and interactions are discussed below.

**Occasion.** Scores after consultation exceeded scores before consultation across the entire experiment (test of main effect for occasion: F1,209 = 48.0; P < .001), with an effect size (Cohen d) of 0.32.

**Experience.** A significant interaction of occasion with experience level (F2,209 = 3.1; P = .048) indicates that the consultation effect sizes (faculty, 0.25; residents, 0.31; students, 0.59) varied by experience level. The effect sizes within each experience level are statistically significant (P < .005 by 3 paired t tests). Both before and after consultation, quality scores were higher for subjects with greater levels of clinical experience (test of main effect for experience level: F2,209 = 48.7; P < .001).

**Decision Support System.** A significant interaction of occasion with DSS (F1,209 = 5.6; P < .02) indicates that the consultation effect size was greater for QMR (d = 0.45) than ILIAD (d = 0.20). The effect sizes for each DSS are statistically significant (P < .001 by 2 paired t tests). All other effects in the analysis model were not significant.
Analysis to Elucidate the Source of DSS Effects

As used by subjects, the DSSs displayed the correct diagnosis for 785 (40.6%) of 1934 cases (39.7% for faculty; residents, 43.2%; students, 38.8%). For those cases where the DSS displayed the correct diagnosis, diagnostic quality scores expressed as mean (95% confidence interval [C1]) increased significantly from 6.86 (95% CI, 6.47-7.25) to 8.14 (95% CI, 7.77-8.51) (P<.001 by paired t test). For the 1149 cases where DSSs did not display the correct diagnosis, mean quality scores were essentially unchanged: 5.15 (95% CI, 4.90-5.40) before consultation and 4.95 (95% CI, 4.70-5.20) after consultation. This implies that the DSSs were influential in generating the measured increases. Cases where the DSS did and did not display the correct diagnosis were associated with significantly different mean preconsultation quality scores (P<.001 by t test). Therefore, cases that were harder for the subjects were also harder for the DSSs.

COMMENT

Across the full sample of clinicians and cases, DSS consultation had a modest positive effect on diagnostic reasoning. The overall increase in diagnostic quality scores (Cohen d = 0.32) was between the effect size typically considered small (0.2) and medium (0.5) in magnitude.3 The cases used in the study proved, as designed, to be diagnostically challenging. Experienced faculty subjects identified correct diagnoses (without DSS support) for less than 50% of cases.

The positive consultation effects were obtained even though the DSSs, as used by the study subjects, generated the correct diagnosis in 41% of cases. By studying the consultation model where each DSS’s advice is filtered through human cognition, rather than the oracle model, we see how a DSS does not have to be invariably correct to be helpful. However, this generates a process that works in both directions. While clinicians could selectively incorporate sound advice and ignore unhelpful advice, more often than the reverse, negative consultations did occur in a small percentage of cases. The negative consultations were equally prevalent across levels of experience.

As hypothesized, the magnitude of consultation effects was related to clinical experience, although positive effects for all 3 levels were statistically significant. The larger effects for students suggest a possible role for these DSSs in undergraduate medical education; for example, DSS consultations could be illuminating to students who are researching cases and preparing presentations. Students, despite their smaller personal knowledge base, were equally likely as faculty to use the DSSs in ways that induced correct diagnoses from the systems.

Although a head-to-head comparison of systems was not a primary intent of this study, we observed consultation effects that were larger for QMR than ILIAD. This effect could have multiple causes, since the tested DSSs differ profoundly in their models for representing medical knowledge and algorithms to generate advice, as well as the interfaces used to control them.

This study used a repeated-measure research design that models the process of clinical consultation. Repeated-measure designs can generate rethinking effects whereby subjects’ performances after consultation increase as a consequence of additional time spent with the case. The significant quality score increases seen in the cases where the DSSs generated the correct diagnosis, with no increase seen in the other cases, offers evidence to support the view that most of the observed effect was due to the DSS consultation rather than rethinking.

Several limitations of this study derived from deliberate choices made in structuring the work. Our subjects reported little experience with diagnostic DSSs prior to the training we provided. In this regard, we believe they typified physicians in training and practice across the country, but DSS users more experienced than those in our study may have generated different results with these cases. The subjects were based in either a manifestly academic setting or had strong ties to the academic setting through clinical appointments with teaching responsibilities. As such, the study results may not generalize to other practice venues. The DSSs and cases address only the domain of difficult cases in internal medicine. Similar work applied to other medical specialties may yield different results. The DSSs chosen for the study, while considered the most mature and available of those extant at the time, may not be the most effective DSSs available today. Updated versions of the 2 DSSs we tested may outperform, or underperform, their predecessors studied here. Also, the report used only 1 lens through which DSS effects might be examined. Diagnostic hypothesis formation is but 1 aspect of the clinical rea-

©1999 American Medical Association. All rights reserved.

Table 2. Decision Support Consultation Based on Diagnostic Quality Scores*

<table>
<thead>
<tr>
<th>Experience Level</th>
<th>DSS</th>
<th>No. of Subjects</th>
<th>Diagnostic Quality Score, Mean (95% Confidence Interval)</th>
<th>Effect Size by Cohen d (95% Confidence Interval)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Without DSS</td>
<td>With DSS</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consultation</td>
<td>Consultation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students</td>
<td>72</td>
<td>4.5 (4.2-4.8)</td>
<td>5.1 (4.8-5.4)</td>
<td>0.59 (0.37-0.81)</td>
<td>.001</td>
</tr>
<tr>
<td>Residents</td>
<td>72</td>
<td>6.0 (5.7-6.3)</td>
<td>6.4 (6.1-6.7)</td>
<td>0.31 (0.13-0.49)</td>
<td>.001</td>
</tr>
<tr>
<td>Faculty</td>
<td>71</td>
<td>6.5 (6.2-6.8)</td>
<td>6.8 (6.5-7.1)</td>
<td>0.25 (0.07-0.43)</td>
<td>.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QMR</td>
<td>107</td>
<td>5.7 (5.4-5.9)</td>
<td>6.1 (5.9-6.3)</td>
<td>0.20 (0.08-0.32)</td>
<td>.001</td>
</tr>
<tr>
<td>Students</td>
<td>107</td>
<td>5.6 (5.3-5.9)</td>
<td>6.2 (5.9-6.5)</td>
<td>0.45 (0.31-0.59)</td>
<td>.001</td>
</tr>
<tr>
<td>All subjects</td>
<td>215</td>
<td>5.7 (5.5-5.9)</td>
<td>6.1 (5.9-6.3)</td>
<td>0.32 (0.23-0.41)</td>
<td>.001</td>
</tr>
</tbody>
</table>

*DSS indicates decision support system; QMS, Quick Medical Reference.
ASONING PROCESS. The DSSs may be more useful in other ways, such as by suggesting tests and other next steps in a patient evaluation. Other limitations to this work derive from its conduct in the laboratory. The case summaries used, although comprehensive and based on real patients, were not complete medical records. In the practice of medicine, clinicians would typically have access to more information about the patients than the summaries provided. The motivation and thus the performance of the subjects could have been somewhat different if the cases had been patients under the clinicians’ own care.

In summary, DSS consultation modestly enhanced the diagnostic reasoning of subjects using the tested systems. Smaller effects for more experienced physicians indicated that any case difficult enough to challenge an experienced internist will likely also challenge the systems we studied. Although these systems are clearly not infallible oracles, they may have useful roles in the evolving world of computer-based information resources.

**Funding/Support:** This work was supported in part by grant R01-LM-05630 from the National Library of Medicine.

**Acknowledgment:** We gratefully acknowledge the assistance of the following individuals: Judith Miller, MS, for her contribution to the development of the training methods and work in data collection; James Sisson, MD, for identifying and writing summaries for several cases; David Potts, PhD, and Kevin Bisoi, PhD, in preparing cases; Sema Barlas, PhD, Keith Cogdill, PhD, Macy Ng, MA, Winston Sieck, MA, Alice Nambalamba, MA, and Xiao Mei, MA, for data collection.

**DIAGNOSTIC REASONING ENHANCED WITH COMPUTERS**

**REFERENCES**


groups. Among non-Hispanic whites, nonnative born mothers have slightly higher FD rates. The racial differentials (regardless of nativity) are much larger than the nativity differentials (regardless of race/ethnicity). These differentials persisted after subdividing these groups by maternal age (data not shown). For most of the race/ethnic/nativity groups, age of 30 years or more raised the rate of FD slightly, but these rates were not significantly different than for mothers younger than 30 years.

Comment. The direction and size of the relationships reported here (nonnative born groups experiencing more favorable outcomes than their native-born counterparts, except in the case of non-Hispanic whites, for whom this relationship is reversed) are consistent with previous US nativity studies that focused on other pregnancy outcomes such as preterm birth, low birthweight, and infant mortality. While the overall FD rates (by race/ethnicity) in this study are comparable with those from other US studies, data quality is a continuing concern in studying fetal demise. Underreporting of FDs probably occurred to some extent in this study and may vary by nativity and race/ethnicity.

Jeffrey E. Kallan, PhD
Lawrenceville, NJ


CORRECTION

Table Discrepancies: In the Original Contribution entitled “Enhancement of Clinicians’ Diagnostic Reasoning by Computer-Based Consultation: A Multisite Study of 2 Systems” published in the November 17, 1999, issue of THE JOURNAL (1999;282:1851-1856), there were discrepancies in a table. On page 1854, the following table replaces Table 1.

Table 1. Decision Support Consultation Effects Based on Presence of the Correct Diagnosis*

<table>
<thead>
<tr>
<th>Experience Level</th>
<th>No. of Cases</th>
<th>Correct Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Without Decision Support System</td>
<td>With Decision Support System</td>
</tr>
<tr>
<td></td>
<td>No. (%) of Consultations</td>
<td></td>
</tr>
<tr>
<td>Students</td>
<td>647</td>
<td>168 (26.0)</td>
</tr>
<tr>
<td>Residents</td>
<td>648</td>
<td>293 (43.7)</td>
</tr>
<tr>
<td>Faculty</td>
<td>639</td>
<td>314 (49.1)</td>
</tr>
</tbody>
</table>

Decision Support System

<table>
<thead>
<tr>
<th></th>
<th>No. of Cases</th>
<th>Correct Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ILIAD</td>
<td>390 (40.1)</td>
</tr>
<tr>
<td></td>
<td>QMR</td>
<td>375 (39.0)</td>
</tr>
<tr>
<td>All cases</td>
<td>1934</td>
<td>765 (39.6)</td>
</tr>
</tbody>
</table>

*Within each row of the table, percentages are based on the total number of cases in that row. QMR indicates Quick Medical Reference.
†A consultation was deemed positive when clinicians generated the correct diagnosis as part of their differential diagnosis after a decision support system consultation but did not include it on their initial preconsultation diagnosis list.
‡A consultation was deemed negative when clinicians included the correct diagnosis in their initial preconsultation diagnosis list but excluded it from their postconsultation diagnosis list after using a decision support system.
§Values represent the difference between the number of correct cases after a decision support system consultation and the number correct before consultation.

©2001 American Medical Association. All rights reserved.