

*Research Paper* ■

# Influence of Case and Physician Characteristics on Perceptions of Decision Support Systems

ETA S. BERNER, EdD, RICHARD S. MAISIAK, PhD, MSPH

**Abstract** **Objective:** This study examines how characteristics of clinical cases and physician users relate to the users' perceptions of the usefulness of the Quick Medical Reference (QMR) and their confidence in their diagnoses when supported by the decision support system.

**Methods:** A national sample ( $N = 108$ ) of 67 internists, 35 family physicians, and 6 other U.S. physicians used QMR to assist in the diagnosis of written clinical cases. Three sets of eight cases stratified by diagnostic difficulty and the potential of QMR to produce high-quality information were used. A  $2 \times 2$  repeated-measures analysis of variance was used to test whether these factors were associated with perceived usefulness of QMR and physicians' diagnostic confidence after using QMR. Correlations were computed among physician characteristics, ratings of QMR usefulness, and physicians' confidence in their own diagnoses, and between usefulness or confidence and actual diagnostic performance.

**Results:** The analyses showed that QMR was perceived to be significantly more useful ( $P < 0.05$ ) on difficult cases, on cases where QMR could provide high-quality information, by non-board-certified physicians, and when diagnostic confidence was lower. Diagnostic confidence was higher when comfort with using certain QMR functions was higher. The ratings of usefulness or diagnostic confidence were not consistently correlated with diagnostic performance.

**Conclusions:** The results suggest that users' diagnostic confidence and perceptions of QMR usefulness may be associated more with their need for decision support than with their actual diagnostic performance when using the system. Evaluators may fail to find a diagnostic decision support system useful if only easy cases are tested, if correct diagnoses are not in the system's knowledge base, or when only highly trained physicians use the system.

■ JAMIA. 1999;6:428–434.

Increasing interest has been shown in the use of decision support systems, including diagnostic systems, in clinical settings.<sup>1,2</sup> Previous studies by system de-

velopers and others have examined the accuracy of the diagnostic systems, but little systematic research has evaluated how practicing physicians use and judge the systems.<sup>3–30</sup> Although commercial vendors of systems may solicit user feedback, such data are not usually publicly available. Even if they were, it is not clear how accurate or generalizable user surveys are for the evaluation of complex diagnostic systems. Because of the lack of scientific data, physicians interested in determining the value of these computer systems frequently look to software reviews, which are now appearing more frequently in traditional medical journals. These reviews often provide little description of the particular clinical cases used to test the system and usually are based on only a small number of

Affiliation of the authors: University of Alabama at Birmingham, Birmingham, Alabama.

This work was supported by grant LM05125 from the National Library of Medicine.

Correspondence and reprints: Eta S. Berner, EdD, University of Alabama at Birmingham, School of Health Related Professions, Department of Health Services Administration, Health Informatics Program, 1675 University Boulevard, Room 544, Birmingham, AL 35294-3361. e-mail: (eberner@uab.edu).

Received for publication: 1/14/99; accepted for publication: 5/19/99.

cases.<sup>31</sup> A reviewer's judgment of the value of a system may be influenced by the selection of cases used to test the systems and by the reviewer's own background. The reader of the reviews or surveys thus has to rely on subjective judgments about system performance. Well-controlled field trials, based on an adequate understanding of the cases and the participants, are expensive. It might be more economically feasible to continue to rely on subjective assessments if such assessments relate to more objective performance data. Our previous research<sup>32</sup> showed that at least one decision support system, Quick Medical Reference (QMR), appeared to have a positive effect on physicians' diagnostic performance. In the present study we examine whether case difficulty, diagnostic decision support system (DDSS) information quality, and physicians' characteristics relate to physicians' perceptions of the usefulness of a DDSS and also to their confidence in their own diagnoses after using the DDSS, and whether participants' perceptions of usefulness and confidence relate to their actual diagnostic performance.

## Methods

The DDSS chosen for this study was QMR, a diagnostic decision support system for internal medicine.<sup>33</sup> The physician population and study design are described in detail in Berner et al.<sup>32</sup> Figure 1 summarizes the case selection process.

### Case Usefulness and Confidence Ratings

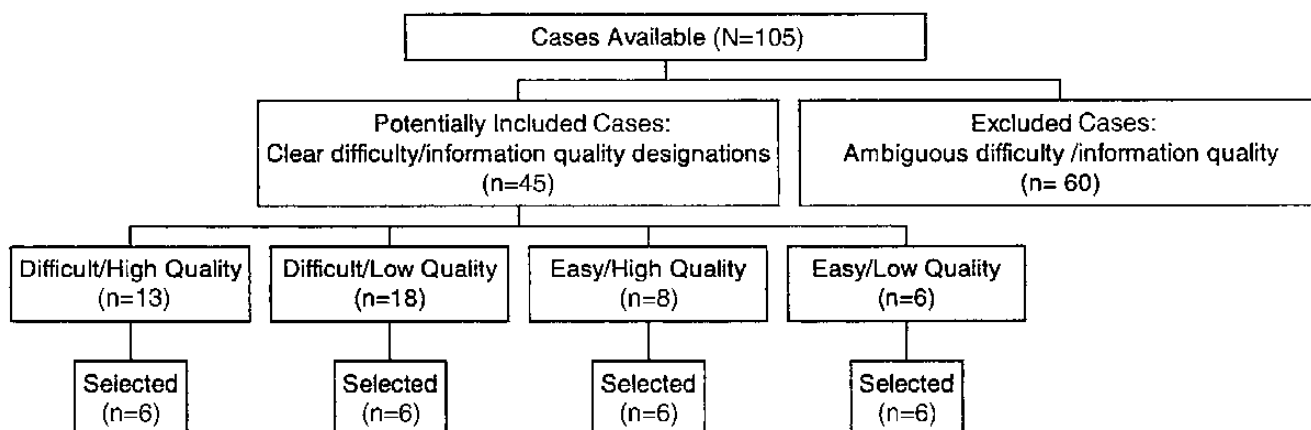
Subjects were given the set of eight cases and were asked to use QMR to assist them in formulating a differential diagnosis for each case. They were instructed to use QMR's case analysis function and any other functions they chose. After they used QMR on each

case, the physicians selected one of the descriptive statements shown in Table 1 to describe their judgment of the usefulness of QMR for that case, and one of the statements shown in Table 2 to describe how difficult they perceived the case to be and how confident they were in their own differential diagnosis. We refer to these ratings as "usefulness" and "confidence" ratings to make a clear distinction between these measures and the a priori defined case information quality and difficulty levels. The usefulness and confidence ratings were assigned values of 1 to 4 and 1 to 5, respectively, with higher numbers referring to greater usefulness and greater confidence. Although participants may not have felt they actually needed to use QMR to develop their diagnoses, using at least QMR's case analysis function was a requirement of the study.

### Performance Data

The performance data from Berner et al.<sup>32</sup> were used. Using physicians as the units of analysis, mean accuracy, relevance, and comprehensiveness scores for each of the 108 physicians on each of the eight-case sets were computed. Those scores, as discussed at length in Berner et al.,<sup>32</sup> are described below.

- **Accuracy.** The mean diagnostic accuracy score for each physician was computed as the proportion of cases for which the correct case diagnosis was listed on the physician's differential diagnosis list.
- **Relevance.** The diagnostic relevance score for a physician on a particular case was computed as the proportion of diagnoses on a physician's list that were considered appropriate for that case. Mean relevance scores were the means of the subject's individual case relevance scores.



**Figure 1** Case selection process.

Table 1 ■

## Statements to Describe the Usefulness of Quick Medical Reference (QMR)

Assigned Value	Descriptive Statement
4	Extremely useful, in that it confirmed my thinking and gave me more possible diagnoses to consider.
3	Useful, in that it confirmed my thinking, but did not give me anything new to consider.
2	Not useful, in that it gave me little useful and included too much that was irrelevant.
1	Useless, in that the suggestions were very misleading.

NOTE: The descriptive statements were given to subjects without the numeric values.

Table 2 ■

## Statements Used to Describe Confidence in Diagnosis Using Quick Medical Reference (QMR)

Assigned Value	Descriptive Statement
5	Very easy; I felt confident in my differential and did not need to use QMR for assistance.
4	Pretty easy; I used QMR mainly to confirm my thinking.
3	Somewhat difficult, but I feel confident in my final differential diagnosis.
2	Pretty difficult; I would like a subspecialist to confirm my thinking.
1	Extremely difficult; I still feel I need a subspecialist for assistance.

NOTE: The descriptive statements were given to subjects without the numeric values.

Table 3 ■

## Perceived Usefulness Ratings of Quick Medical Reference (QMR) by Case Difficulty and Quality of QMR Information\*

Case Difficulty	Mean Quality of QMR Information (SD)		
	High Information Quality	Low Information Quality	Total
Easy	3.14 (0.60)	2.66 (0.82)	2.90 (0.54)
Difficult	3.35 (0.68)	3.06 (0.77)	3.20 (0.56)
Total	3.24 (0.52)	2.86 (0.63)	3.05 (0.47)

\* $P < 0.05$ , significant main effect, repeated measures analysis of variance.

■ **Comprehensiveness.** The diagnostic comprehensiveness score for each case was computed as the proportion of appropriate diagnoses for a particular case that the physician included on the differential diagnosis.

## Data Analyses

Physicians, not cases, were the units of analysis. The primary data used were individual physician ratings across sets of cases. To examine the possibility of selection bias, differences in background characteristics between selected and unselected study subjects were tested using independent groups *t*-tests. A two-factor repeated measures analysis of covariance, adjusted for case set, was used to test for the effects of case difficulty and QMR information quality on usefulness and diagnostic confidence. Bivariate Pearson correlation coefficients (*r*) were computed to examine the relationship between the usefulness and confidence ratings and between these ratings and physicians' characteristics. They were tested for significant differences from  $r = 0$  using a *t*-test. A two-sided alpha level of 0.05 was used for these analyses. SPSS software was used for all analyses.<sup>34</sup>

To examine the relationship between participants' perceptions and their performance, again using physicians as the unit of analysis, the mean usefulness ratings and the mean confidence ratings for each of the nine cells shown in Tables 3 and 4 were correlated with the corresponding mean accuracy, relevance, and comprehensiveness scores. To control for the effect of set differences, partial correlation coefficients were computed, generating a total of 54 different correlations. With a Bonferroni correction adjusted for an alpha of 0.05, the adjusted significance level was set at 0.0009.

Table 4 ■

## Diagnostic Confidence Ratings by Case Difficulty and Quality of Quick Medical Reference (QMR) Information\*

Case Difficulty	Mean Quality of QMR Information (SD)		
	High Information Quality	Low Information Quality	Total
Easy	3.51 (0.75)	3.44 (0.79)	3.47 (0.62)
Difficult	2.55 (0.86)	2.88 (0.97)	2.71 (0.77)
Total	3.03 (0.59)	3.16 (0.73)	3.09 (0.57)

\* $P < 0.05$ , significant main effect, repeated measures analysis of variance.

## Results

### Response Rate and Respondent Characteristics

As described in Berner et al.,<sup>32</sup> a total of 120 initially selected and 70 replacement physicians were offered the opportunity to participate, and 108 completed the cases.

### Relationships among Usefulness and Confidence Ratings and Case and Physician Characteristics

The ratings of usefulness and confidence in relation to case characteristics are shown in Tables 3 and 4. The mean QMR usefulness rating was 3.05 out of 4.00, indicating that, overall, the physicians perceived QMR as useful on these cases. Perceived usefulness was significantly greater for the more difficult cases and on the cases for which QMR could produce higher quality information.

The overall mean confidence ratings were significantly lower ( $P < 0.03$ ) on the high-information-quality cases than on the low-information-quality cases. However, this effect was attributable mainly to a significant interaction ( $P < 0.0001$ ) of case difficulty and case information quality, in that the decrease in confidence ratings for the high-information-quality cases compared with the low-information-quality cases was much greater for the difficult cases than for the easy cases. In fact, the direction of the confidence scores for the easier cases was slightly reversed from that for the difficult cases. As might be expected, diagnostic confidence was, overall, significantly higher on the easier cases regardless of the quality of QMR information. The correlation between each physician's mean usefulness and confidence ratings across all cases was  $-0.50$ , indicating a moderate but significant inverse relationship between these two variables.

To better understand how physicians' characteristics affected the perception of QMR usefulness and their own diagnostic confidence, the correlations of physicians' characteristics with mean usefulness and mean confidence ratings were computed (Table 5). The results showed that physicians who began using QMR more recently tended to rate its performance on the cases as significantly more useful ( $P = 0.04$ ) than those who had first purchased it several years ago. Also, the general board-certified physicians tended to rate the usefulness of QMR on these cases significantly lower ( $P = 0.02$ ) than did other physicians, while general board-eligible physicians tended to rate QMR usefulness significantly higher ( $P = 0.04$ ). The only physician characteristic significantly associated with greater diagnostic confidence ( $P < 0.05$ ) was higher self-re-

ported skill at using certain sophisticated features of QMR.

### Relationship between Perceptions and Performance

Table 6 shows the correlations between the three measures of diagnostic performance and the two measures of perception for all strata of case difficulty and QMR information quality. For the most part, these correlations were low ( $r = -0.31$  to  $+0.29$ ). Only six correlations, five of them negative and all related to the relevance scores, were significantly different from zero. These data show that although participants' judgments of QMR usefulness and all three performance measures were both affected similarly by case difficulty and quality of QMR information, neither their perceptions of QMR usefulness nor participants' confidence in their own diagnoses were consistently related to their actual performance.

## Discussion

Theoretically, a DDSS should be perceived as more useful when the user perceived that the DDSS provided relevant information and when the user actually felt a need for decision support. Physicians with greater diagnostic skills would be less likely to perceive a need for decision support and would be expected to find a DDSS less useful than those less skilled. The results of this study lend support to these assumptions. QMR was perceived to be more useful on the high-information-quality cases, supporting the idea that users could distinguish relevant from irrelevant information. It was also perceived to be more useful on the difficult cases, on which need for support would be greater. Finally, there were inverse associations between usefulness and the ability to develop a focused and relevant differential (relevance score) as well as between usefulness and being board certified. These data again provide some support for the notion that diagnostically skilled physicians might not need the support and would find the information QMR provided less useful.

Although the relevance scores were inversely associated with perceived usefulness, usefulness and diagnostic performance as indicated by the accuracy score were not strongly associated, nor were there significant correlations with the comprehensiveness scores. It is possible that feedback on the accuracy of physicians' diagnoses might have led to stronger associations between perceived usefulness and the other measures of performance, but the design of this study did not allow for such feedback. Without feedback,

Table 5 ■

Relationship of Perceived Usefulness of Quick Medical Reference (QMR) and Diagnostic Confidence with Physician Characteristics ( $N = 108$ )

Sample Characteristics	Perceived Usefulness of QMR		Diagnostic Confidence	
	$r$	$P$ Value	$r$	$P$ Value
Stratification variables:				
Year of medical school completion	-0.05	0.58	0.09	0.34
Specialty—Internal Medicine	-0.08	0.44	0.07	0.47
Specialty—Family Medicine	0.06	0.57	0.01	0.90
Frequency of use of QMR in last 6 mo†	-0.02	0.85	0.03	0.75
Length of time using QMR‡	-0.17*	0.04	0.04	0.70
Other demographic variables:				
Year of primary residency completion	-0.13	0.20	0.11	0.29
General specialty board certified	-0.23*	0.02	0.06	0.55
General board eligible	0.20*	0.04	0.01	0.89
Other QMR experience variable:				
Number of minutes in usual QMR session	0.01	0.92	0.09	0.38
Reported comfort using QMR functions:				
Exploring disease profile/associated disorders	-0.08	0.40	0.17	0.08
Case analysis	-0.14	0.16	0.19*	0.05
Asserting diagnosis	0.01	0.88	0.33*	0.00
Work-up protocol	-0.04	0.72	0.15	0.13
Critiquing a case	-0.03	0.76	0.27*	0.00
Differential diagnosis	-0.13	0.19	0.05	0.62
Comparing two diseases	-0.11	0.26	0.15	0.12
Questions for a particular diagnosis	-0.05	0.63	0.07	0.49
Rule-in/rule-out diagnoses	-0.04	0.71	0.08	0.42
Saving a case	0.02	0.82	0.20*	0.04
Saving a case to a text file	0.09	0.34	0.10	0.33
Printing a window	-0.05	0.57	0.19	0.06
Number of QMR functions participants were comfortable using	-0.06	0.92	0.09	0.38

\*Pearson correlation coefficient,  $P < 0.05$ ,  $t$ -test, significantly different from 0.00.

†Ordinal scale with seven categories (1, use every day, to 7, have not used during last six months).

‡Ordinal scale with three categories (1, purchased QMR within last 6 mo; 2, purchased QMR 6 to 12 mo ago; 3, purchased QMR more than a year ago).

Table 6 ■

Partial Correlations ( $r$ ) Between Perceptions of Quick Medical Reference (QMR) and Physicians' Performance on Cases with Different Characteristics

Type of Case	Correlations of Perception and Performance Scores					
	Accuracy		Relevance		Comprehensiveness	
	Usefulness	Confidence	Usefulness	Confidence	Usefulness	Confidence
Difficult	-0.09	-0.14	-0.28*	0.17	0.12	-0.23
Easy	-0.10	0.12	-0.29*	0.20	0.15	-0.06
High information quality	-0.07	-0.02	-0.21	0.09	0.25	-0.20
Low information quality	-0.13	0.03	-0.26*	0.16	0.03	-0.04
Difficult, high quality	-0.05	-0.11	-0.24	0.03	0.17	-0.24
Difficult, low quality	0.02	-0.09	-0.19	0.24	-0.03	0.09
Easy, high quality	-0.02	0.03	-0.26*	0.29*	0.09	-0.13
Easy, low quality	-0.20	0.16	-0.21	0.09	0.12	0.02
Total	-0.14	-0.04	-0.31*	0.13	0.15	-0.16

physicians may have construed usefulness primarily as how well QMR satisfied their need for decision support. The lack of significant correlations is also difficult to interpret because the performance measures reflect the users' diagnoses after using QMR, and we do not know what their diagnostic performance would have been like without QMR.

The theoretical relationship between case difficulty and confidence would be expected to be the reverse of that for usefulness, especially in the absence of feedback about the accuracy of performance. Confidence would be expected to be higher on the easier cases, for which the user might be expected to be able to develop a reasonable differential without diagnostic assistance, recognize confirmatory information, and ignore irrelevant information. However, these are also the cases for which the perceived need for decision support and the perceived usefulness of QMR is less. The negative correlation between case difficulty and confidence, as well as the inverse relationship between confidence and perceived usefulness, lend support to this interpretation. These data are also congruent with the observations of Bankowitz et al.<sup>16</sup> that physicians were uninterested in seeing the results of a computer consultation if they were fairly certain of their own diagnoses.<sup>16</sup>

Although QMR information quality was directly related to perceived usefulness, the relationship between quality of information and diagnostic confidence is more complex. Usefulness is a multifaceted construct. A decision support system can be useful not only if it confirms the user's original diagnostic thinking but also if it broadens that thinking by providing more relevant diagnoses for consideration. It is possible that these two aspects of usefulness (confirming and extending the user's thinking) operate differently in the easy and the more difficult cases. When QMR provides high quality information, the confirmatory aspect is likely to be most prominent on the easy cases, while the user's thinking is more likely to be extended on the difficult cases. While confirming the user's thinking may increase confidence, broadening the user's thinking could decrease confidence in the user's original diagnoses for these cases. This interpretation provides an explanation for the observed interaction effect of case difficulty and information quality on diagnostic confidence. Becoming less confident when QMR provides useful suggestions not previously considered seems to be more plausible than becoming more confident when one receives irrelevant information.

It was not unusual to find that users who had previously purchased a decision support system tended to

find it useful or that users who felt more comfortable using the system would be more confident in its results. However, the fear that the users of these systems—especially the “early adopters” of innovations, as these participants are—would not make distinctions between more and less useful information does not appear to be warranted. If this sample of QMR users had been inappropriately biased in the direction of believing everything the computer generated and finding all of it very useful, none of the results showing differences in perceived usefulness related to QMR information quality would have been significant. The results of this study and those of Berner et al.<sup>32</sup> provide support for the notion that users can distinguish between relevant and irrelevant information produced by a DDSS and can use the relevant information to help them in their diagnostic thinking. At the same time, the difficulty of the cases on which the DDSS is used, as well as the users' own diagnostic skills, can affect users' performance and their ability to accurately assess the information provided by the DDSS.

The data also provide support for the notion that perception of usefulness of a DDSS may be related more to the user's need for support than to the user's diagnostic performance when using the DDSS. Participants may have been helped by the DDSS without realizing it or may have thought the DDSS was more helpful than it really was. The data from this study illustrate the need for less reliance on opinions and anecdotes and more systematic research on the influences of DDSS on physicians' diagnostic performance.

## Conclusions

Diagnostic decision support systems can function both to confirm and to broaden physicians' diagnostic thinking. Although both aspects can be useful, providing additional possibilities for consideration may make the user somewhat less confident and more aware of the need for further diagnostic assistance. The results suggest that users' perceptions of QMR usefulness and their diagnostic confidence were associated more with a need for support than with their actual diagnostic performance. Studies that assess users' judgments about the usefulness of DDSS and that use relatively easy cases, cases not in the DDSS knowledge base, or only highly trained specialists may yield misleading results regarding the overall usefulness of the DDSS. In particular, software reviews of system usefulness based on a single reviewer's opinion and a limited case sample may not be a reliable indicator of system functioning. Addi-

tional research on the relationships between opinions about a DDSS, how physicians use the system, and how the DDSS influences the user's diagnostic performance is still needed. The data from this study underscore the importance of understanding the characteristics of the physician users and the clinical cases that are part of evaluations of decision support systems in order to properly interpret the results of any evaluation of DDSS.

The authors appreciate the contributions of the physicians who used QMR and provided us with their data.

## References ■

1. Clayton PD, Hripczak G. Decision support in healthcare. *Int J Biomed Comput.* 1995;39:59-66.
2. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *JAMA.* 1998;280:1339-46.
3. Barness LA, Tunnessen WW Jr, Worley WE, Simmons TL, Ringe TBK Jr. Computer-assisted diagnosis in pediatrics. *Am J Dis Child.* 1974;127:852-8.
4. O'Shea JS. Computer-assisted pediatric diagnosis. *Am J Dis Child.* 1975;129:199-202.
5. Swender PT, Tunnessen WW Jr, Oski FA. Computer-assisted diagnosis. *Am J Dis Child.* 1974;127:859-61.
6. Wexler JR, Swender PT, Tunnessen WW Jr, Oski FA. Impact of a system of computer-assisted diagnosis: initial evaluation of the hospitalized patient. *Am J Dis Child.* 1975;129:203-5.
7. Osheroff JA, Bankowitz RA. Physicians' use of computer software in answering clinical questions. *Bull Med Libr Assoc.* 1993;81:11-9.
8. Waxman HS, Worley WE. Computer-assisted adult medical diagnosis: subject review and evaluation of a new micro-computer-based system. *Medicine.* 1990;69:125-36.
9. Georgakis DC, Trace DA, Naeymi-Rad F, Evens M. A statistical evaluation of the diagnostic performance of MEDAS: the medical emergency decision assistance system. *Proc 14th Annu Symp Comput Appl Med Care.* 1990:815-9.
10. Nelson SJ, Blois MS, Tuttle MS, et al. Evaluating RECONSIDER: a computer program for diagnostic prompting. *J Med Syst.* 1985;9:379-88.
11. Hammersley JR, Cooney K. Evaluating the utility of available differential diagnosis systems. *Proc 12th Annu Symp Comput Appl Med Care.* 1988:229-31.
12. Feldman MJ, Barnett GO. An approach to evaluating the accuracy of DXplain. *Comput Methods Programs Biomed.* 1991;35:261-6.
13. Heckerling PS, Elstein AS, Terzian CG, Kushner MS. The effect of incomplete knowledge on the diagnosis of a computer consultant system. *Med Inform.* 1991;16:363-70.
14. Lau LM, Warner HR. Performance of a diagnostic system (Iliad) as a tool for quality assurance. *Comput Biomed Res.* 1992;25:314-23.
15. Bouhaddou O, Lambert JG, Morgan E. Iliad and Medical HouseCall: evaluating the impact of common sense knowledge on the diagnostic accuracy of a medical expert system. *Proc Annu Symp Comput Appl Med Care.* 1995;742-6.
16. Bankowitz RA, Lave JR, McNeil MA. A method for assessing the impact of a computer-based decision support system on health care outcomes. *Methods Inf Med.* 1992;31:3-11.
17. Bankowitz RA, McNeil MA, Challinor SM, Parker RC, Kapoor WN, Miller RA. A computer-assisted medical diagnostic consultation service: implementation and prospective evaluation of a prototype. *Ann Intern Med.* 1989;110:824-32.
18. Bankowitz RA, McNeil MA, Challinor SM, Miller RA. Effect of a computer-assisted general medicine diagnostic consultation service on housestaff diagnostic strategy. *Methods Inf Med.* 1989;28:352-6.
19. Berman L, Miller RA. Problem area formation as an element of computer aided diagnosis: a comparison of two strategies within Quick Medical Reference (QMR). *Methods Inf Med.* 1991;30:90-5.
20. Middleton B, Shwe MA, Heckerman DE, et al. Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base, part II: evaluation of diagnostic performance. *Methods Inf Med.* 1991;30:256-67.
21. Miller RA, Pople HE Jr, Myers J. Internist-I, an experimental computer-based diagnostic consultant for general internal medicine. *N Engl J Med.* 1982;307:468-76.
22. Miller RA, Masarie FE Jr. The quick medical reference (QMR) relationships function: description and evaluation of a simple, efficient "multiple diagnoses" algorithm. *Medinfo.* 1992:512-8.
23. Miller R, McNeil M, Challinor S, Masarie F, Myers J. The Internist-1/Quick Medical Reference project: status report. *West J Med.* 1986;145:816-22.
24. Bacchus CM, Quinton C, O'Rourke K, Detsky AS. A randomized cross-over trial of quick medical reference (QMR) as a teaching tool for medical interns. *J Gen Intern Med.* 1994;9:616-21.
25. Sumner W II. A review of Iliad and Quick Medical Reference for primary care providers: two diagnostic computer programs. *Arch Fam Med.* 1993;2:87-95.
26. Berner ES, Webster GD, Shugerman AA, et al. Performance of four computer-based diagnostic systems. *N Engl J Med.* 1994;330:1792-6.
27. Berner ES, Jackson JR, Algina J. Relationships among performance scores of four diagnostic decision support systems. *J Am Med Inform Assoc.* 1996;3:208-15.
28. Murphy GC, Friedman CP, Elstein AS, et al. The influence of a decision support system on the differential diagnosis of medical practitioners at three levels of training. *Proc AMIA Annu Fall Symp.* 1996:219-23.
29. Elstein AS, Friedman CP, Wolf FM, et al. Effects of a decision support system on the diagnostic accuracy of users: a preliminary report. *J Am Med Inform Assoc.* 1996;3:422-8.
30. Wolf FM, Friedman CP, Elstein AS, et al. Changes in diagnostic decision making after a computerized decision support consultation based on perceptions of need and helpfulness: a preliminary report. *Proc AMIA Annu Fall Symp.* 1997:263-7.
31. Berner ES. The problem with software reviews of decision support systems. *MD Comput.* 1993;10:8-12.
32. Berner ES, Maisiak RS, Cobbs CG, Taunton OD. Effects of a decision support system on physicians' diagnostic performance. *J Am Med Inform Assoc.* 1999;6:420-7.
33. Quick Medical Reference (QMR). San Bruno, Calif.: First DataBank Corporation, 1994.
34. Norusis MJ. Statistical Package for Statistical Solutions Reference Guide. Chicago, Ill.: SPSS, 1990.