ORIGINAL ARTICLE

How well does decision support software perform in the emergency department?

M A Graber, D VanScoy

Emerg Med J 2003;20:426–428

Objective: To determine how well general decision support systems perform given the data collected in an emergency department (ED).
Methods: A convenience sample of 25 patients was selected from those patients having a diagnostic question on presentation to the ED. All interactions with the patients were audiotaped and abstracted into a structured data form. All other data such as written notes, laboratory, and EKG results were also abstracted. All data were entered into two general diagnostic decision support programs (Quick Medical Reference (QMR Version 3.82, Knowledge Base 10–07–1998 Copyright University of Pittsburgh and The Hearst Corporation) and Iliad (Version 4.5 Copyright 1996 Applied Medical Informatics)). The diagnoses generated by the computer programs were compared with the final diagnoses of the ED attending.
Results: The final ED diagnosis was found in the differential diagnosis generated by Iliad and QMR 72% and 52% of the time respectively. The final ED diagnosis was found in the top 10 diagnoses 51% and 44% of the time and in the top five diagnoses 36% and 32% of the time for each program respectively. This approximates to the performance of these programs in other clinical settings.
Conclusions: Diagnostic decision support software has the same success in finding the "correct" diagnosis in the ED as in other clinical settings where more extensive clinical data are available. The accuracy is not sufficiently high to permit the use of these programs as an arbiter in any individual case. However, they may be useful, prompting additional investigation in particularly difficult cases.

See end of article for authors' affiliations

Correspondence to: Dr M A Graber, Department of Emergency Medicine, Pomerantz Family Pavilion, University of Iowa, 200 Hawkins Dr, Iowa City, IA 52246, USA; mark-graber@uiowa.edu

Accepted for publication 16 October 2002

Decision support algorithms and software have been developed in an attempt to improve decision making in the emergency department (ED). However, the usefulness of these tools is limited by their narrow scope, as they are usually designed to improve diagnostic accuracy in the analysis of specific problems such as chest pain or abdominal pain.[1–5] The scope of emergency medicine is wide. General decision support systems such as Quick Medical Reference (QMR) and Iliad can help the ED physician consider remote diagnostic possibilities in a time and resource efficient manner. Previous studies of these programs have focused on difficult, inpatient, or paper cases developed from a medical record after extensive clinical and laboratory data are available.[6–9] No study has assessed how these programs perform using the limited amount of data that are collected during an ED visit. The purpose of this study is to evaluate how well Quick Medical Reference (QMR Version 3.82, Knowledge Base 10–07–1998 Copyright University of Pittsburgh and The Hearst Corporation) and Iliad (Version 4.5 Copyright 1996 Applied Medical Informatics) parallel physician decision making in common ED situations.

QMR and Iliad (named for the originator, Homer Warner at The University of Utah) are "expert systems". Expert systems are designed to emulate the solutions to problems that one might expect from a human expert. In the cases of QMR and Iliad, medical symptoms, signs, and laboratory results are entered, compared to known entities, and a differential diagnosis is generated. Both QMR and Iliad allow for an adjustment to reflect the prevalence of a disease in the population. Iliad uses Bayesian analysis while QMR uses non-Bayesian algorithms.[6] For example, headache and fever may indicate meningitis. But because the prevalence of influenza in the population is higher, influenza would be listed as the more probable diagnosis.

One limitation of these expert systems is that they only encompass a finite set of illnesses, symptoms, and signs. As more entities are added to their databases, it can be expected that the accuracy and relevance of the differential diagnosis generated to improve. Additionally, as they are designed by humans, they may reflect the biases of the authors and put more weight on one finding (for example, fever) than on another (for example, headache). To minimise the risk of bias, the symptom complex representative of a disease is generally arrived at by consensus (personal communication, Homer Warner).

One scenario in which these programs are useful is when the investigator is presented with an unusual constellation of symptoms or a presentation about which they have limited knowledge. For example, the differential diagnosis of a monoarticular arthritis includes not only common illnesses such as gout, pseudogout, and gonoccoccal arthritis but also more obscure entities that may not readily come to mind such as Brucillosis. An expert system can help by expanding one's differential diagnosis and suggesting other avenues of pursuit including what questions to ask as well as the most cost effective method of evaluating the patient.

METHODS

A convenience sample of 25 patients was selected from all patients seen in our ED, a tertiary care academic medical centre. The study period was from 15 July to 15 August 2001. Exclusion criteria included age under 18, residence in a correctional facility, inability to give consent, psychiatric complaints, and inability to communicate in English. Patients were also excluded if there was no diagnostic question (for example, trauma such as lacerations, fractures, corneal abrasions, etc).

After consent was obtained, clinical information about the patient visit was prospectively collected. The basic information was recorded by ED examiners (generally a resident or supervised medical student) on a structured data form (T-System,

Table 1 Final diagnoses made by the clinician*

Diagnosis				
Abdominal wall trigger point (2)	Venous insufficiency	Unstable angina, r/o MI (3)	Chronic paroxysmal hemicrania	Myocardial infarction (2)
COPD exacerbation (2)	Post-concussion syndrome	Deep venous thrombosis	Bilateral pneumonia with hypoxia	Bacterial vaginosis
Digoxin toxicity	Vomiting with dehydration secondary to general anaesthesia	Transient ischaemic attack	Herpes simplex II	
Urticaria	Pelvic inflammatory disease (2)	Conjunctivitis	Heat exhaustion with dehydration	Congestive heart failure
Viral pharyngitis	Peripheral vertigo	Alcoholic liver disease with encephalopathy (2)	Pulmonary embolism	

*Numbers in parentheses represent the number of times this final diagnosis appeared. Some patients had more than one diagnosis.

Copyright 2001, www.tsystem.com). To capture any other data that entered into the decision making process, one of the researchers (DVS) accompanied all examiners during all contacts with each patient. All of these contacts were audiotaped and relevant clinical data from tape transcriptions were abstracted and entered into a structured data form. All other entries in the medical record including dictated notes, laboratory results, etc, were reviewed after the visit and any additional data gleaned were added to the case abstract. Only data available in the ED were included in the case abstracts. Thus, the ED physician EKG and radiograph interpretations were used in the case abstract. One of the researchers (MG) was available to assist in interpreting terminology or other data and reviewed all of the cases. The final ED discharge diagnoses from the medical records are listed in table 1.

All of the clinical information including history, physical findings, laboratory findings, radiographic findings, EKG findings, etc, was entered into QMR and Iliad. The differential diagnosis generated by the computer programs was compared with the final diagnosis of the ED physicians, all of whom were (and are) academic faculty in emergency medicine. The performance of the computer programs was determined by comparing the number of cases where the attending staffs' diagnosis was listed on the programs' differential list. As the purpose of this study was to determine how well QMR and Iliad perform using only the information collected in the ED, any information obtained while the patient was an inpatient or at subsequent outpatient visits was not considered. Likewise, the "final" diagnosis made in the hospital or at subsequent visits was not considered. This design was chosen because the purpose of this study was not to determine the accuracy of ED diagnoses but rather so see how QMR and Iliad performed compared with experienced ED physicians. The Human Subjects committee at our institution approved this protocol.

RESULTS

The results are summarised in table 2. Each case took from 20 to 40 minutes to input.

DISCUSSION

When defining success as having the ED physicians' diagnosis within the top five generated by QMR and Iliad, the programs were successful in about one third of the cases. Overall, the ED physicians' diagnosis appeared somewhere in differential diagnosis with about the same frequency that has been found in other studies (around 50%–70%).[6] Thus, while these programs may be helpful in a general sense, applying them as an arbitrator in any particular case may be problematic. The considerable length of the differential produced by the programs, often greater than 30 diagnoses, and the length of time it takes to input a case may hinder their usefulness. The strength of these programs is their ability to expand the differential diagnosis and make suggestions about further testing and evaluation.[10 11] This can be done by entering a few key findings. Further study of decision support software for this purpose in the ED is warranted.

Several limitations in the design of these programs became clear during this study. The type of information that can be entered into the programs is limited. Iliad and QMR do not take into account the drugs that a patient may be taking. The ability to input the duration of signs and symptoms is also limited, especially with QMR, and the programs are unable to account for the sequence of symptom development. Other findings are simply not in the programs' vocabulary despite trying multiple synonyms and therefore cannot be entered as part of case. Further development of these programs may mitigate these problems.

The main limitation of this study is the reliance on the ED physicians' diagnosis as the "criterion standard". The final diagnosis after hospitalisation or subsequent outpatient visits may or may not differ from the ED diagnosis. However, this is after additional information has been collected and additional time has elapsed. We concentrated on how these programs perform given the limited information collected in the ED. All of the physicians involved in the study were academic emergency physicians at a major teaching university. So, even though the final, post-hospitalisation, diagnosis may vary from that made in the ED, this study reflects the best available diagnosis in the ED.

In conclusion, diagnostic decision support software has the same success in finding the "correct" diagnosis in the ED as has been found in other clinical settings. The accuracy is not sufficiently high to permit the use of these programs as an arbiter in any individual case. However, they can be used to broaden the differential diagnosis.

Table 2 Final ED diagnosis found in program's differential

Program	Anywhere	Top ten diagnoses	Top five diagnoses
QMR	52	44	32
Iliad	72	51	36

Data shown as percentages.

Authors' affiliations

M A Graber, D VanScoy, Department of Emergency Medicine, University of Iowa College of Medicine, Iowa City, USA

M A Graber, Department of Family Medicine, University of Iowa College of Medicine

Funding: none.

Conflicts of interest: none declared.

REFERENCES

1 **Massel D**, Dawdy JA, Melendez LJ. Strict reliance on a computer algorithm or measurable ST segment criteria may lead to errors in thrombolytic therapy eligibility. *Am Heart J* 2000;**140**:221–6.

2 **Qamar A**, McPherson C, Babb J, *et al*. The Goldman algorithm revisited: prospective evaluation of a computer-derived algorithm versus unaided physician judgment in suspected acute myocardial infarction. *Am Heart J* 1999;**138**:705–9.
3 **de Dombal FT**, Dallos V, McAdam WA. Can computer aided teaching packages improve clinical care in patients with acute abdominal pain? *BMJ* 1991;**302**:1495–7.
4 **Jonsbu J**, Aase O, Rollag A, *et al*. Prospective evaluation of an EDB-based diagnostic program to be used in patients admitted to hospital with acute chest pain. *Eur Heart J* 1993;**14**:441–6.
5 **Schriger DL**, Baraff LJ, Rogers WH, *et al*. Implementation of clinical guidelines using a computer charting system. Effect on the initial care of health care workers exposed to body fluids. *JAMA* 1997;**278**:1585–90.
6 **Berner ES**, Webster GD, Shugerman AA, *et al*. Performance of four computer-based diagnostic systems. *N Engl J Med* 1994;**330**:1792–6.
7 **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.
8 **Gozum ME**. Emulating cognitive diagnostic skills without clinical experience: a report of medical students using quick medical reference and Iliad in the diagnosis of difficult cases. Proceedings of the annual symposium on computer applications in medical care, 1994:991.
9 **Grum CM**, Miller JG, Wolf FM. Computer-based problem solving for primary-care diagnosis in an internal medicine clerkship. *Acad Med* 1994;**69**:429–30.
10 **Bankowitz RA**, McNeil MA, Miller RA. Effect of a computer-assisted general medicine diagnostic consultation service on housestaff diagnostic stragery. *Meth Inform Med* 1989;**28**:352–6.
11 **Friedman CP**, Elstein AS, Wolf FM, *et al*. Enhancement of clinicians' diagnostic reasoning by computer-based consultation. A multisite study of 2 systems. *JAMA* 1999;**282**:1851–6.