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

M A Graber, D VanScoy

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.
Table 1  Final diagnoses made by the clinician*

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>QMR</th>
<th>Iliad</th>
<th>Top five diagnoses</th>
<th>Top ten diagnoses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdominal wall trigger point (2)</td>
<td>Venous insufficiency</td>
<td>Unstable angina, r/o MI (3)</td>
<td>Chronic paroxysmal hemicrania</td>
<td></td>
</tr>
<tr>
<td>COPD exacerbation (2)</td>
<td>Post-concussion syndrome</td>
<td>Deep venous thrombosis</td>
<td>Bilateral pneumonia with hypoxia</td>
<td></td>
</tr>
<tr>
<td>Digoxin toxicity</td>
<td>Vomiting with dehydration</td>
<td>Transient ischemic attack</td>
<td>Herpes simplex II</td>
<td></td>
</tr>
<tr>
<td>Urticaria</td>
<td>Pelvic inflammatory disease (2)</td>
<td>Conjunctivitis</td>
<td>Heat exhaustion with dehydration</td>
<td></td>
</tr>
<tr>
<td>Viral pharyngitis</td>
<td>Peripheral vertigo</td>
<td>Alcoholic liver disease with encephalopathy (2)</td>
<td>Pulmonary embolism</td>
<td></td>
</tr>
</tbody>
</table>

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

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%)." Thus, while these programs may be helpful in a general sense, applying them as an arbiter 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.

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.