Impact of a Computer-Based Diagnostic Decision Support Tool on the Differential Diagnoses of Medicine Residents

Mitchell J. Feldman, MD
Edward P. Hoffer, MD
G. Octo Barnett, MD
Richard J. Kim, BSE
Kathleen T. Famiglietti, BS
Henry C. Chueh, MD, MS

Abstract

Background Computer-based medical diagnostic decision support systems have been used for decades, initially as stand-alone applications. More recent versions have been tested for their effectiveness in enhancing the diagnostic ability of clinicians.

Objective To determine if viewing a rank-ordered list of diagnostic possibilities from a medical diagnostic decision support system improves residents' differential diagnoses or management plans.

Method Twenty first-year internal medicine residents at Massachusetts General Hospital viewed 3 deidentified case descriptions of real patients. All residents completed a web-based questionnaire, entering the differential diagnosis and management plan before and after seeing the diagnostic decision support system's suggested list of diseases. In all 3 exercises, the actual case diagnosis was first on the system's list. Each resident served as his or her own control (pretest/posttest).

Results For all 3 cases, a substantial percentage of residents changed their primary considered diagnosis after reviewing the system's suggested diagnoses, and a number of residents who had not initially listed a "further action" (laboratory test, imaging study, or referral) added or changed their management options after using the system. Many residents (20% to 65% depending on the case) improved their differential diagnosis from before to after viewing the system's suggestions. The average time to complete all 3 cases was 15.4 minutes. Most residents thought that viewing the medical diagnostic decision support system's list of suggestions was helpful.

Conclusion Viewing a rank-ordered list of diagnostic possibilities from a diagnostic decision support tool had a significant beneficial effect on the quality of first-year medicine residents' differential diagnoses and management plans.

Editor's Note: The online version of this article contains the sample resident surveys used in this study.

Mitchell J. Feldman, MD, is Assistant Clinical Professor of Pediatrics, Harvard Medical School, and Physician at Laboratory of Computer Science, Massachusetts General Hospital (LCS-MGH); **Edward P. Hoffer, MD,** is Senior Computer Scientist at LCS-MGH; **G. Octo Barnett, MD,** is Professor of Medicine, Harvard Medical School, and Senior Scientific Director at LCS-MGH; **Richard J. Kim, BSE,** is Programmer, LCS-MGH; **Kathleen T. Famiglietti, BS,** is Project Manager, LCS-MGH; and **Henry C. Chueh, MD, MS,** is Director, LCS-MGH, and Assistant Professor of Medicine at Harvard Medical School.

Funding: This work was conducted with support from Harvard Catalyst | The Harvard Clinical and Translational Science Center (NIH Award UL1 RR 025758) and financial contributions from Harvard University and its affiliated academic health care centers.

Additional funding support for this study was provided by a grant from the Partners-Siemens Research Council. The Partners-Siemens Research Council did not participate in the design of the study, in the collection, analysis or interpretation of the data, in the writing of the report, or in the decision to submit the article for publication. The researchers were independent from the Partners-Siemens Research Council.

The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard Catalyst, Harvard University and its affiliated academic health care centers, the National Center for Research Resources, or the National Institutes of Health.

Introduction

Computer-based medical diagnostic decision support systems (MDDSSs) have been available for 50 years.[1,2] While many initial evaluations of these systems focused on the ability of the system itself to arrive at the "correct" diagnosis,[3–6] more recent studies have evaluated whether the systems enhance the diagnostic ability of clinicians.[7–10] Working in partnership

Other disclosures: M.J.F., E.P.H., G.O.B., R.J.K., K.T.F. and H.C.C. are employed by the Laboratory of Computer Science at Massachusetts General Hospital, the developers of the DXplain diagnostic decision support system. During the previous 3-year period, the Laboratory of Computer Science has received licensing royalties related to the DXplain system from approximately 40 hospitals and medical schools, and from Merck & Co, Inc, and Epocrates, Inc.

The authors thank Hang Lee, PhD, for statistical assistance, and the 2010 Massachusetts General Hospital Internal Medicine Interns for their participation.

Corresponding author: Mitchell J. Feldman, MD, Massachusetts General Hospital, Laboratory of Computer Science, 50 Staniford Street, 7th Floor, Boston, MA, 02114, 617.724.7991, mfeldman@partners.org

Received July 27, 2011; revision received November 9, 2011; accepted November 28, 2011.

DOI: http://dx.doi.org/10.4300/JGME-D-11-00180.1

with an information source has been theorized to help individuals perform better than when unassisted.[11] In regard to MDDSSs, Miller wrote, "the ultimate unit of evaluation should be whether the user plus the system is better than the unaided user."[7] While evaluations of MDDSSs in vitro can be useful as these systems develop, evaluations that measure the ability of a system to help clinicians or improve clinicians' performance are preferable.[1]

DXplain (General Hospital Corp, Boston, MA) is an MDDSS developed at the Massachusetts General Hospital (MGH) Laboratory of Computer Science; it has evolved and expanded and been used continually since 1987. The knowledge base contains more than 2400 disease profiles, more than 4800 clinical findings (history, physical examination, and laboratory terms) and hundreds of thousands of data points that describe the relationships between them. Diseases included in the knowledge base encompass the specialties and subspecialties of internal medicine, surgery, pediatrics, and obstetrics and gynecology.[12] Previous studies have shown that the system performed well in suggesting appropriate diseases.[13,14]

Prior studies have examined how computer-based decision support systems, including other MDDSSs, influence user performance.[8,15–18] This is the first study to determine whether the differential diagnoses and management plans of first-year internal medicine residents could be improved by viewing the rank-ordered list of diagnostic considerations from DXplain.

What was known

Computer-based medical diagnostic decision support systems can enhance the diagnostic effectiveness of clinicians.

What is new

Use of medical diagnostic decision support systems improved the differential diagnosis in 20% to 65% of participating interns, depending on the case. The time commitment was reasonable at 15 minutes for all 3 cases.

Limitations

Single site, small sample. Cases were selected to maximize impact if a diagnosis were omitted from initial consideration. Subjects were interns, and the effect would likely be less pronounced for more experienced clinicians.

Bottom line

Viewing a rank-ordered list of diagnostic possibilities from a diagnostic decision support tool has a positive effect on first-year medicine residents' differential diagnoses and management plans. Effects may be more modest for more experienced clinicians.

Methods

We identified 3 cases from actual patients with high-risk diagnoses (ovarian cancer, colon cancer, and appendicitis) by using a research patient data registry developed in our laboratory, and abstracted and deidentified these into case summaries. Summaries only included information from medical notes before the high-risk diagnosis had been established. For each case, relevant case terms were entered into DXplain to observe the rank order of the high-risk diagnosis on DXplain's list. Relevant case terms were demographic findings, high information findings, and other findings in the summary relevant to the patient's presenting complaint. We chose cases for which DXplain successfully raised the consideration of the high-risk diagnosis, given the selected case terms. For the case descriptions shown to the residents, we used deidentified narrative text from patient notes entered by the clinician into the electronic medical record. Irrelevant negative findings were omitted for brevity. The study received Institutional Review Board approval.

Seventy-five MGH first-year internal medicine residents were recruited via e-mail after approval by the residency program director. A gift certificate to Amazon.com was offered for participation. After consenting to participate, residents were sent individual login information to the survey website. Three cases were presented to each participant. After viewing each case description, residents were asked to indicate the most likely diagnosis and then to list up to 2 more diagnoses that also would be reasonable to consider given this information. For each of these diagnoses, they were asked to list what further action should be taken (eg, colonoscopy, abdominal computed tomography or ultrasonography, surgical or gynecological referral). Residents were instructed not to use any external references as they completed the survey.

DXplain's disease list for the case was then displayed, and each resident was again asked to list the most likely diagnosis and up to 2 additional diagnoses that should be considered and any further actions for each. For each case, DXplain's disease list contained 15 to 20 potential diagnoses, several of which were supported or strongly supported (indicated by a + or ++). In all 3 cases used, the patient's ultimate diagnosis was highest on DXplain's list. Finally, participants were asked the question "Do you feel that DXplain's list of possible diseases was helpful in expanding your diagnostic focus?" (5-point Likert scale: Of no use = 1, Minimal help = 2, Somewhat helpful = 3, Helpful = 4, Very helpful = 5). The sample cases used in this study are provided as online supplemental material.

Outcome Measures

Diagnosis "Off" List to "On" List Whether the case diagnosis was absent from the resident's differential diagnosis before seeing DXplain's suggestions, but present

TABLE **Percentage of Residents Achieving Outcome Measure and Survey Question Responses**

Outcome Measure	Case 1: Ovarian CA	Case 2: Colon CA	Case 3: Appendicitis
Dx off list to on list Residents achieving measure, % (*P* value per McNemar test)	65 (<.001)	20 (.046)	25[a]
Dx off list or alternate Dx to No.1 Dx Residents achieving measure, % (*P* value per McNemar test)	30[b]	35 (.008)	35 (.008)
Further action off list to on list Residents achieving measure, % (*P* value per McNemar test)	65 (<.001)	15[c]	20 (.18)
Survey question,[d] response ± SD	3.7 ± 1	3.8 ± 1	4.0 ± 0.6

Abbreviations: CA, cancer; Dx, diagnosis.

[a] McNemar test not applicable because all residents listed the case diagnosis after the intervention.
[b] McNemar test not applicable because no residents listed the case diagnosis as a first choice before the intervention.
[c] McNemar test not applicable because all residents listed a further action after the intervention.
[d] "Was DXplain helpful in expanding your diagnostic focus?": 1 = Of no use; 2 = Minimal help; 3 = Somewhat helpful; 4 = Helpful; 5 = Very helpful.

on the differential diagnosis after the resident viewed DXplain's suggestions.

Diagnosis "Off" List or Alternate Diagnosis to No. 1 Diagnosis Whether the case diagnosis moved from being "off" the resident's differential before DXplain's suggestions were seen to being first on the differential after the resident viewed DXplain's suggestions, or moved from the third or second position on the resident's differential before DXplain's suggestions were seen, to the first position after the resident viewed DXplain's suggestions.

Further Action "Off" List to "On" List Whether a recommended management plan that was not listed by the resident before seeing DXplain's suggested differential diagnosis was included after those suggestions were viewed by the resident.

Overall Impression of Usefulness of DXplain's Suggestions Based on a 5-point Likert scale: 1 = Of no use, 5 = Very helpful.

Time to Complete all Cases Time for each resident to complete all cases.

Time Spent Considering DXplain's List and Entering Follow-up Differential Diagnosis Time between submitting the initial differential diagnosis until submitting the second differential after viewing DXplain's suggestions.

The McNemar test was used to compare the before and after results to determine whether viewing DXplain's suggestions had an effect. The SAS program[19] was used to compute the test results.

Results

Of the 75 MGH first-year internal medicine residents invited to participate, 20 accessed the web-based questionnaire and completed all 3 cases. The average time (SD) to complete the 3 cases was 15.4 (5) minutes. The average time (SD) for all residents and all cases, between entering the initial differential diagnosis and entering the subsequent differential after reviewing DXplain's suggestions was 2.3 (1.3) minutes.

For case 1, 13 of 20 residents (65%) who had not considered the case diagnosis before viewing DXplain's suggestions did indicate this diagnosis as a possibility after seeing DXplain's list. This was statistically significant ($P < .001$) with the McNemar test. Three of 20 residents (15%) listed the case diagnosis both before and after viewing DXplain's list, and 4 of 20 (20%) did not list the case diagnosis either before or after viewing the list. Six of 20 (30%) of residents who, before viewing DXplain's suggestions, had either not listed the case diagnosis, or listed it as an alternate possibility, moved the case diagnosis to the primary consideration after viewing DXplain's list. Thirteen of 20 (65%) who had not initially listed a "further action" (laboratory test, imaging study, or referral) considered correct (eg, recommended management in standardized texts such as UpToDate[20]) before viewing DXplain's list did include a correct management option after seeing the list. This result was also statistically significant ($P < .001$) with the McNemar test. In both case 2 and case 3, the residents either added the correct diagnosis to the differential and/or moved it to the top of their lists a significant minority of the time (TABLE). The McNemar test could not be performed in a few situations because of the degeneracy case (eg, in case 3 when 25% of residents who had initially not listed the case diagnosis did so after seeing DXplain's list, since all the residents [100%] listed the case diagnosis after the intervention). On the 5-point Likert scale, residents on average found viewing DXplain's list to be "helpful."

Discussion

Our results show that the differential diagnoses of residents improved after viewing suggestions provided by the DXplain MDDSS. Since the average time to complete all 3 cases was 15.4 minutes, and the average time between entering the initial differential diagnosis and entering the subsequent differential after reviewing DXplain's suggestions was only 2.3 minutes, there was little time for the residents to consult other resources. Therefore, it is likely that the only factor triggering a change in the residents' differentials and management plans was DXplain's list of suggested diagnoses. A study by Friedman et al[21] showed that 41% of resident participants were overly confident, and did not seek assistance in establishing a diagnosis, when listing diagnoses later determined to be incorrect during the generation of differential diagnoses for a set of cases, and 36% of students and 32% of faculty were similarly overconfident. It is reassuring that our results suggest that the residents were receptive to the decision support offered. That the residents changed their differentials in response to the diagnostic suggestions suggests that "anchoring" and overconfidence were not insurmountable effects.[21]

The cases used in this study were real and the patients' diagnoses had already been established. We purposely chose cases for which the diagnosis was not immediately obvious, but for which DXplain listed the case diagnosis high on its differential diagnosis list when the case findings were entered. There was a significant difference in the effect that the residents' viewing DXplain's suggestions had in case 1 (ovarian cancer) compared with cases 2 (colon cancer) and 3 (appendicitis), regarding how often the case diagnosis went from being off the list to being included on the list. This is likely due to the case 1 description, for which findings of urinary frequency along with suprapubic discomfort and abdominal bloating and pain, steered the residents toward diagnoses of the genitourinary and gastrointestinal systems, respectively, for their initial differentials (before seeing DXplain's differential). Current medical literature supports a consideration of ovarian cancer given these symptoms,[22,23] and DXplain's knowledge base reflects this information. Cases 2 and 3 were less subtle, and this may explain why the MDDSS was less influential than in case 1. However, viewing DXplain's suggestions was still helpful in these cases, increasing by 20% to 25% the degree to which residents included the case diagnosis in their differentials as compared with before viewing DXplain's list.

We sent a questionnaire one time to each repeat user of DXplain. From 2001–2011, a total of 4301 of 6063 unique users who submitted questionnaires responded that "DXplain suggested appropriate diseases in its differential that had not been considered." This question and the response results are similar to the survey question regarding residents' acceptability of DXplain in the current study, and thus add face validity to the results. While our results are encouraging, there are some possible pitfalls.

Limitations

In addition to selecting cases for which the diagnosis was not immediately apparent from the visit note, we chose cases and diagnoses for which DXplain's utility could be maximal if it raised that diagnosis for consideration. In clinical practice, many patients will ultimately be diagnosed with less serious conditions or with diagnoses for which decision support would be neither helpful nor necessary. If MDDSSs are integrated into clinical workflows, it will be important to balance the thresholds at which an MDDSS agent generates an alert so as to minimize false positives, and consequently reduce alert fatigue.[24]

In each of the 3 cases, the actual case diagnosis was listed first on DXplain's list. This could be a source of case-selection bias. Further research would be needed to determine the effect on residents' differential diagnoses of the actual case diagnosis being lower on DXplain's list.

The participants in our study were MGH first-year medical residents. Others[9] have shown that the degree to which diagnostic decision support can be helpful varies inversely with the experience of the clinician. Therefore, our results may be better than what might be achieved with more experienced clinicians. However, even a more modest improvement in suggesting for consideration diagnoses not previously considered would be beneficial to diagnostic accuracy.

Conclusion

Our data support the notion that viewing a rank-ordered list of diagnostic possibilities from a diagnostic decision support tool has a significant beneficial effect on the quality of first-year medicine residents' differential diagnoses and management plans. Further study is needed to establish whether integrating computer-based decision support tools with clinical workflows may allow for earlier or improved diagnosis and more appropriate management.

References

1 Miller RA. Computer-assisted diagnostic decision support: history, challenges, and possible paths forward. *Adv Health Sci Educ Theory Pract*. 2009;14(suppl 1):89–106.

2 Miller RA. Medical diagnostic decision support systems—past, present, and future: a threaded bibliography and brief commentary. *J Am Med Inform Assoc*. 1994;1(1):8–27.

3 Barness LA, Tunnessen WW Jr, Worley WE, Simmons TL, Ringe TB Jr. Computer-assisted diagnosis in pediatrics. *Am J Dis Child*. 1974;127(6):852–858.

4 de Dombal FT, Leaper DJ, Horrocks JC, Staniland JR, McCann AP. Human and computer-aided diagnosis of abdominal pain: further report with emphasis on performance of clinicians. *Br Med J*. 1974;1(5904):376–380.

5 Miller RA, Pople HE Jr, Myers JD. Internist-1, an experimental computer-based diagnostic consultant for general internal medicine. *N Engl J Med*. 1982;307(8):468–476.

6 Berner ES, Webster GD, Shugerman AA, Jackson JR, Algina J, Baker AL, et al. Performance of four computer-based diagnostic systems. *N Engl J Med*. 1994;330(25):1792–1796.

7 Miller RA. Evaluating evaluations of medical diagnostic systems. *J Am Med Inform Assoc*. 1996;3(6):429–431.

8 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(10):824–832.

9 Friedman CP, Elstein AS, Wolf FM, Murphy GC, Franz TM, Heckerling PS, et al. Enhancement of clinicians' diagnostic reasoning by computer-based consultation: a multisite study of 2 systems. *JAMA*. 1999;282(19):1851–1856.

10 Ramnarayan P, Kapoor RR, Coren M, Nanduri V, Tomlinson AL, Taylor PM, et al. Measuring the impact of diagnostic decision support on the quality of clinical decision making: development of a reliable and valid composite score. *J Am Med Inform Assoc*. 2003;10(6):563–572.

11 Friedman CO. A "fundamental theorem" of biomedical informatics. *J Am Med Inform Assoc*. 2009;16(2):169–170.

12 Hoffer EP, Feldman MJ, Kim RJ, Famiglietti KT, Barnett GO. DXplain: patterns of use of a mature expert system. *AMIA Annu Symp Proc*. 2005;2005:321–324.

13 Feldman MJ, Barnet GO. An approach to evaluating the accuracy of DXplain. *Comput Methods Programs Biomed*. 1991;35(4):261–266.

14 Feldman M, Barnett GO. Pediatric computer-based diagnostic decision support. In: *Scientific Program, Section on Computers and Other Technologies*. Elk Grove Village, IL. American Academy of Pediatrics; 1991:15.

15 Roy PM, Durieux P, Gillaizeau F, Legall C, Armand-Perroux A, Martino L, et al. A computerized handheld decision-support system to improve pulmonary embolism diagnosis. *Ann Intern Med*. 2009;151:677–686.

16 Ramnarayan P, Roberts GC, Coren M, Nanduri V, Tomlinson A, Taylor PM, et al. Assessment of the potential impact of a reminder system on the reduction of diagnostic errors: a quasi-experimental study. *BMC Med Inform Decis Mak*. 2006;6:22.

17 Ramnarayan P, Winrow A, Coren M, Nanduri V, Buchdahl R, Jacobs B, et al. Diagnostic omission errors in acute paediatric practice: impact of a reminder system on decision-making. *BMC Med Inform Decis Mak*. 2006;6:37

18 Ely JW, Graber ML, Croskerry P. Checklists to reduce diagnostic errors. *Acad Med*. 2011;86(3):307–313.

19 *SAS* [computer program]. Version 9. Cary, NC: SAS Institute Inc; 2003.

20 UpToDate.com. http://www.uptodate.com/index. Accessed November 9, 2011.

21 Friedman CP, Gatti GG, Franz TM, Murphy GC, Wolf FM, Heckerling PS, et al. Do physicians know when their diagnoses are correct: implications for decision support and error reduction *J Gen Intern Med*. 2005;20(4):334–339.

22 Goff B, Mandel L, Melancon C, Muntz H. Frequency of symptoms of ovarian cancer in women presenting to primary care clinics. *JAMA* 2004;291:2705–2712

23 Hamilton W, Peters TJ, Bankhead C, Sharp D. Risk of ovarian cancer in women with symptoms in primary care: population based case-control study. *BMJ*. 2009;339:b2998.

24 Shah NR, Seger AC, Seger DL, Fiskio JM, Kuperman GJ, Blumenfeld B, et al. Improving acceptance of computerized prescribing alerts in ambulatory care. *J Am Med Inform Assoc*. 2006;13(1):5–11.