

# Clinician Performance and Prominence of Diagnoses Displayed by a Clinical Diagnostic Decision Support System

Eta S. Berner, EdD, Richard S. Maisiak, PhD, MSPH,  
Gustavo R. Heudebert, MD and K. Randall Young, Jr., MD

University of Alabama at Birmingham, Birmingham, AL

## ABSTRACT

*Clinical decision support systems (CDSS) can impact both diagnostic and therapeutic decision-making, but physicians sometimes fail to heed the appropriate CDSS advice, or become influenced in a negative way by the CDSS. This study examined the relationships among clinicians' prior diagnostic accuracy, the performance of a diagnostic CDSS, and how the CDSS influenced the accuracy of the clinician's subsequent diagnoses. Results showed that (1) clinicians who already were considering the correct diagnosis prior to using the CDSS were more likely to get the CDSS to produce the correct diagnosis in a prominent position than those not considering it initially; (2) physicians are strongly anchored by their initial diagnoses prior to using the CDSS; and (3) changes in the clinicians' diagnoses after using the CDSS are related to presence or absence of the correct diagnosis in the top 10 diagnoses displayed by the CDSS.*

## INTRODUCTION

Clinical decision support systems (CDSS) have been shown to decrease medication errors and improve diagnostic decision-making.<sup>1-3</sup> Groups such as Leapfrog and others have advocated the use of reminder and alerting systems and interest in computerized provider order entry with clinical decision support is increasing.<sup>4</sup> Although these systems are clearly capable of influencing physicians in a positive manner, there are also reports in the published literature that physicians may fail to alter their original wrong decisions, or worse, may change correct decisions to incorrect ones after using CDSS.<sup>2,5,6</sup> For instance, Teich et al. found that some decisions were easier to influence than others. In particular, they commented that it is difficult to get physicians to change plans they had already made.<sup>5</sup> Galanter et al. published an article describing a situation where a CDSS provided an alert of a medication overdose that the attending clinician repeatedly ignored.<sup>6</sup> On the other hand, Friedman et al., in an article documenting the effectiveness

of two diagnostic decision support systems mentions that sometimes their subjects switched from the correct diagnosis to an incorrect one after using the system.<sup>3</sup> Although Friedman et al.'s subjects improved more when the CDSS provided the correct diagnosis, it was not immediately clear what factors related to what Friedman et al. termed "negative consultations." It is possible that the CDSS provided misleading information that caused the subjects to change, or that the subjects did not attend to or understand what they were viewing in spite of accurate CDSS information. Similarly, in other studies where the CDSS provided useful information that was not acted upon, it is not clear whether the subjects were anchored to a prior course of action or whether there was some other reason for failing to heed the suggestions. As CDSS move out of the research arena and into practice settings, it becomes important not just to document when they are effective, but also to systematically examine the situations in which they do not work as intended. Such data are important for improving the CDSS as well as for instructing users on the most appropriate ways to utilize the CDSS output.

This paper provides an analysis of the impact of CDSS performance on diagnostic decision-making. Studies of the impact of diagnostic CDSS have tended to examine how the physician performs without examining in detail the what the CDSS actually provided, other than the correct diagnosis being present on the CDSS list.<sup>2,3</sup> Only by examining the CDSS performance can we begin to understand how it influences clinician performance. In the present study we examined the effect on physician diagnostic performance of the presentation of the correct diagnosis in a prominent or less prominent position in the output of a diagnostic CDSS. Although we focused on a diagnostic system, the issue of the factors that lead a CDSS to change, or fail to change, clinician decision-making is important for all kinds of CDSS.

## METHODS

The subjects were 70 internal medicine residents who took part in a study examining the effects of training and practice on use of a diagnostic decision support system. The diagnostic decision support system was Quick Medical Reference (QMR™), version 3.8.5 marketed by First DataBank. It was available via the Internet on a server running Citrix Metaframe, version 1.8 and Windows NT, version 4.0. These residents had all been given two to four hours of training on the use of QMR, had QMR available online for use for a two-month period and then used QMR to assist them with the diagnosis of four diagnostically challenging paper and pencil cases. The cases were selected from a previous sample of diagnostically challenging cases.<sup>7</sup> Although all of the cases were rare and/or atypical, the correct diagnosis (based on definitive test results) was in QMR's knowledge base and, with optimal use of QMR, appeared within the top 20 of QMR's suggestions. Subjects reviewed the case, prepared their own differential diagnosis, used QMR in any manner they chose, and then revised their differential diagnosis. A maximum of 20 diagnoses were allowed. For each case, subjects saved the last QMR screen they were looking at as a text file. This permitted determination of whether the correct diagnosis was either being actively considered by the resident and/or presented to them as a possibility by QMR. Thus, there were a total of 280 cases analyzed. For eight cases, the screens were inadvertently not saved and those cases are excluded from further analyses, making the total analyzed 272.

For each case we determined the resident's unaided diagnosis, the position of the correct diagnosis on the last QMR screen saved for that case, and the resident's final diagnosis.

Unaided diagnosis—For each case, this was scored correct if the correct diagnosis was included on the resident's initial differential prior to using the CDSS and incorrect if the correct diagnosis was not included.

Position of the correct diagnosis on the last screen saved—Rank of the correct diagnosis as displayed on QMR's screen. The top diagnosis was rank 1. We grouped the ranks into strata of five as shown in Tables 2 and 3. We considered display within the top two strata (e.g, first ten diagnoses on the list) as a prominent position. Previous research testing QMR performance in a more artificial situation has shown that when QMR displays the correct diagnosis it is highly likely to be in the top ten diagnoses.<sup>7</sup>

Final diagnosis-- For each case, this was scored correct if the correct diagnosis was included on the resident's final differential after using the CDSS and incorrect if the correct diagnosis was not included on the differential.

The proportion of four cases that were correctly diagnosed by each physician was calculated for the unaided and aided conditions. The proportion of the four cases for which the correct diagnosis was listed on the CDSS screen in the top ten position or in the other positions was also calculated for each physician. The associations among the three factors (aided, unaided, position of correct diagnosis on QMR's screen) were then tested with correlation and with multiple linear regression analysis in which unaided correct score and prominent position screen score were independent factors and aided correct score was the dependent factor. An alpha level of .05 was used. SPSS software was used for statistical tests.<sup>8</sup> A case-based analysis was also conducted.

## RESULTS

The proportion of cases with correct diagnoses both prior to, and after using the CDSS, was 55%. The mean proportion of QMR screens where the correct diagnosis was prominently displayed was 17%.

Correlations among unaided correct diagnoses, final correct diagnosis and position of correct diagnosis on the screen, are presented in Table 1.

	Unaided Correct Dx (resident)	Aided Correct Dx (resident)	Correct Dx displayed prominent position (QMR)
Unaided correct Dx	1.000	0.633*	0.454*
Aided correct Dx		1.000	0.523*

\*p<.001

The correlations show that the three factors were significantly and directly correlated with each other. Regression analysis showed that the standardized

beta coefficients for unaided diagnosis correct score (.498,  $p < .001$ ) and prominent QMR screen score (.297,  $p = .004$ ) were each significantly and directly associated with the aided correct score.

Tables 2 and 3 show the case-based analyses, which amplify the results from the physician-based analyses.

Rank of correct diagnosis on last QMR screen seen by subjects	Number of cases with correct diagnosis on QMR screen by unaided diagnosis					
	Cases with Correct Unaided Diagnosis (n=151)		Cases with Incorrect Unaided Diagnosis (n=121)			
	#	(%)	Cum. %	#	(%)	Cum. %
1-5	23	(15)	15	5	(4)	4
6-10	14	(9)	25	3	(2)	7
11-15	8	(5)	30	5	(4)	11
16-20	8	(5)	35	7	(6)	17
21-25	4	(3)	38	5	(4)	21
26-30	2	(1)	39	4	(3)	24
>30	3	(2)	41	10	(8)	32
Not on list	89	(59)	100	82	(68)	100

As shown in Table 2, when the unaided case diagnosis was correct prior to using QMR, QMR produced the correct diagnosis within the top 10 diagnoses in 37 of the 151 cases (25%). However, for 89 of the 151 cases (59%), QMR did not display the correct diagnosis. For the remaining cases, QMR produced the correct diagnosis, but it was not in a prominent position. For the 121 cases where the correct diagnosis was not considered prior to using QMR, QMR prominently displayed the correct diagnosis for only 8 cases (7%), and it was not displayed at all in 82 of the 121 cases (68%). Thus, initially considering the correct diagnosis prior to using the CDSS influences how one interacts with it, which in turn influences how the CDSS performs.

As shown in Table 3, after using QMR, 130 of the 151 cases (86%) that were initially correct still contained the correct diagnosis. Similarly, after using QMR, in only 20 of the cases that initially failed to consider the correct diagnosis, was the correct diagnosis included on the final differential; 83% were unchanged from their unaided diagnosis. When there was congruence between the unaided diagnosis and what appeared in a prominent position on the QMR screen, the unaided diagnosis was unlikely to change after using QMR. When there

was dissonance between the unaided diagnosis and what appeared prominently on the QMR screen, the unaided diagnosis was less likely to remain as the aided diagnosis.

Position of correct diagnosis on last QMR screen seen by subjects	Number and percentage of cases with correct aided diagnosis by unaided diagnosis and QMR diagnosis rank	
	Cases with Correct Unaided Diagnosis (n=151)	Cases with Incorrect Unaided Diagnosis (n=121)
1-5	23 (100%)	4 (80%)
6-10	14 (100%)	3 (100%)
11-15	8 (100%)	0 (0%)
16-20	7 (88%)	1 (14%)
21-25	3 (75%)	0 (0%)
26-30	1 (50%)	1 (25%)
>30	3 (100%)	2 (20%)
Not on list	71 (80%)	9 (11%)
TOTAL	130 (86%)	20 (17%)

Although the overall proportion of correct diagnoses was the same prior to and after using the CDSS, and although most cases did not change, it is instructive to look at those instances where there were changes. That is, there were 20 cases where the differential initially did not have the correct diagnosis, but the resident included it after using QMR. Conversely, there were 21 cases in which the unaided diagnosis was correct, but the final diagnosis after using QMR was incorrect. As Table 3 shows, when the unaided diagnosis was incorrect, but QMR displayed the correct diagnosis in a prominent position, in all but one of the cases (7/8 or 88%), subjects added the correct diagnosis to their final differential. The remaining correct diagnoses came from other positions, but most of the time if the correct diagnosis was not considered prior to using QMR and was not in the top ten diagnoses displayed, there was no change to a correct diagnosis after using QMR. Similarly, none of the 21 deletions of the correct diagnosis after using QMR occurred when the diagnosis was displayed prominently by QMR; almost all of them were a result of QMR failing to display the correct diagnosis.

## DISCUSSION

Anchoring biases have been documented before, but the results of this study expand them to a new

context.<sup>9</sup> The results illustrate that prior anchoring both modulates how one interacts with a CDSS and also influences how one interprets the results of the CDSS. Teich et al. in analyzing the effects of reminders and alerts commented that physicians are more receptive to CDSS that do not require them to change their initial plans.<sup>5</sup> Diagnostic CDSS may frequently suggest changes in initial plans, but as this study shows, such suggestions may not be heeded.

There are some limitations in using the last screen as a proxy for describing how QMR performed. Obviously the variations in what the CDSS displayed were a function of the data selected to be entered as well as how the residents used QMR. What was entered, in turn, was undoubtedly dependent on the residents' prior thinking about the case. Initially considering the correct diagnosis most likely led to identification of salient signs and symptoms to enter into the CDSS, which in turn led to better CDSS performance and vice versa for those not initially considering the correct diagnosis.

It is not surprising then, that QMR performed better when subjects were considering the correct diagnosis prior to using it. However, it should be noted that there were almost 17% who were not considering the correct diagnosis initially who did get QMR to display it. These data highlight that systems that rely on the users to select data to enter, rather than automatically extracting all available patient data, are likely to be subject to variable performance depending on the users' preconceived notions of relevant data. This is also relevant to alerting systems that rely on information in the medical record to produce the appropriate alerts. Both CDSS and clinicians need appropriate data to reach accurate decisions and incomplete data can degrade CDSS performance.

It is also possible that, even though the last screen that was saved did not show the correct diagnosis, prior screens were more helpful or the last screen did show related diagnoses that still aided the subjects' thinking. By targeting only the correct diagnosis and only on the last screen saved, we may be underestimating the usefulness of the CDSS. That the CDSS might have produced other useful information that steered subjects to the correct diagnosis might explain why there were a few changes to the correct diagnosis even when the CDSS did not display the correct diagnosis in a prominent position on the last screen and may also explain why more subjects did not abandon their previously correct diagnosis even if the CDSS did not appear to reinforce it.

Another limitation that is not unique to this study is that by studying improvement from CDSS use by asking subjects to commit in writing their prior unaided differential diagnosis we may have inadvertently promoted the anchoring effect. Other studies have also used this methodology.<sup>3</sup> Prior studies by two of the authors (RSM and ESB) showed that subjects performed better when the correct diagnosis was in QMR's knowledge base than when it was not, but in those studies the subjects did not prepare an unaided differential.<sup>2</sup>

An issue related to all studies of diagnostic systems concerns the selection of the physician's inclusion of the correct diagnosis as the primary outcome measure. Diagnosis is an intermediate step that is presumed to direct treatment choices. Because appropriate treatment decisions can often be made in the absence of a "correct" diagnosis, one cannot directly determine the impact of the observed performance of either the residents or the CDSS on actual treatment decisions.

The importance of the CDSS of not just providing useful information, but displaying it in a prominent position is important information for CDSS developers. CDSS developers and researchers have usually assumed that the user will consider all of the information the CDSS provides and CDSS may often provide more than ten diagnostic suggestions. Cognitive psychologists have stressed for a long time that there are limits to human information processing abilities.<sup>10</sup> Physicians typically develop on their own a fairly limited differential and it may be very difficult for clinicians to adequately evaluate the sometimes lengthy lists of suggestions that CDSS supply. Such processing limitations are even more likely in a fast-paced clinical environment.

In an earlier paper comparing the performance of four diagnostic CDSS, the authors examined the percentage of correct diagnoses appearing at different cut-off levels.<sup>7</sup> The CDSS that was best at displaying the correct diagnosis within the top ten was worst when the cut-off criterion was displaying the correct diagnosis anywhere on the list of diagnoses. At the time the study was done, it was not known what cut-off point was the most appropriate to use to judge the system's performance. Two of the authors of the present paper (ESB and RSM) noted that while using a restricted number of diagnoses to examine physician performance may be appropriate, there were reasons why it might not be appropriate to use a restricted cut-off point, such as the top ten diagnoses, for evaluating CDSS performance.<sup>11</sup>

Coupled with the earlier studies,<sup>2,3,7,11</sup> the results from the present study provide data that suggest that, because physicians are less likely to be influenced by diagnoses that appear lower than the top ten, displaying the diagnosis within the top ten diagnoses produced by a CDSS may be the best choice for a cut-off point for evaluating the accuracy of CDSS performance.

The results of this study are also relevant for examining the impact of other types of decision support systems in addition to those that focus on diagnosis. Decision support systems by definition are aimed at influencing, not making, clinician decisions. The importance of the anchoring effect of prior plans and the fact that simply presenting information to clinicians does not guarantee that it will be attended to and understood is important to recognize for the development and evaluation of any type of decision support system.

### CONCLUSIONS

The present study provides information on some of the factors that lead physicians to attend to the advice of a diagnostic CDSS.

There is a strong relationship between physicians' unaided diagnoses, how well the CDSS performs, and what the physicians' diagnoses are after using the CDSS. Physicians are strongly anchored by their prior diagnoses when using a CDSS, but the CDSS is most likely to change physician diagnostic choices when the CDSS displays or fails to display the diagnosis within the top ten diagnoses. Furthermore, physicians who are already considering the correct case diagnosis are more likely to get a diagnostic CDSS to display the correct diagnosis in a prominent position. Failure to heed the advice of the CDSS or worse performance after using the CDSS may be attributable to the physician only considering the CDSS suggestions that are prominently displayed. Developers of CDSS that provide a lengthy list of suggestions for users need to recognize that users may only attend to the most prominent suggestions.

### Acknowledgements

This study was supported by grant LM 05125 from the National Library of Medicine. The authors acknowledge the assistance of Ms. Tonya La Lande with the data collection and analysis.

### REFERENCES

1. Hunt D, Haynes R, Hanna S, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *JAMA* 1998; 280(15):1339–1346.
2. Berner ES, Maisiak RS, Cobbs CG, Taunton OD. Effects of a decision support system on physician diagnostic performance. *J Am Med Informatics Assoc* 1999; 6:420-7.
3. Friedman CP, Elstein AE, Wolf FM et al. Enhancement of clinicians' diagnostic reasoning by computer-based consultation. A multisite study of 2 systems. *JAMA*. 1999; 282:1851-6.
4. The Leapfrog Group. [www.leapfroggroup.org](http://www.leapfroggroup.org)
5. Teich, JM, Merchia, PR, Schmiz, JL, Kuperman, GJ, Spurr, CD, Bates DW. (2000) Effects of computerized physician order entry on prescribing practices *Arch Intern Med*. 2000; 160(18):2741-7.
6. Galanter WL, DiDomenico RJ and Polikaitis A. Preventing exacerbation of an ADE with automated decision support. *J. Healthcare Information Management* 2002; 16:44-9.
7. Berner ES., Webster GD, Shugerman, AA., et al. Performance of four computer-based diagnostic systems. *N. Engl J Med*. 1994; 330: 1792-6
8. Norušis, MJ. *SPSS 11.0 Guide to Data Analysis*. Upper Saddle River, NJ: Prentice Hall, 2002.
9. Kahneman, D, Slovic, P and Tversky, A. *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press. 1982.
10. Miller, GA. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol Review*; 63:81-97.
11. Maisiak RS and Berner ES. Comparison of Measures to Assess Change in Diagnostic Performance Due to a Decision Support System. *Proc., AMIA Fall Symposium*, 2000: 532-6.