



# Do physicians value decision support? A look at the effect of decision support systems on physician opinion

Stephan Dreiseitl<sup>a,\*</sup>, Michael Binder<sup>b</sup>

<sup>a</sup>Department of Software Engineering for Medicine, Upper Austria University of Applied Sciences,  
Hauptstrasse 117, A-4232 Hagenberg, Austria

<sup>b</sup>Department of Dermatology, Medical University of Vienna, Währingergürtel 18-20,  
A-1090 Vienna, Austria

## KEYWORDS

Decision support  
systems;  
Interaction physician-  
decision support  
system;  
System-induced change  
in physician opinion

## Summary

**Objective:** Clinical decision support systems are on the verge of becoming routine software tools in clinical settings. We investigate the question of how physicians react when faced with decision support suggestions that contradict their own diagnoses.

**Methodology:** We used a study design involving 52 volunteer dermatologists who each rated the malignancy of 25 lesion images on an ordinal scale and gave a dichotomous excise/no excise recommendation for each lesion image. After seeing the system's rating and excise suggestions, the physicians could revise their initial recommendations.

**Results:** We observed that in 24% of the cases in which the physicians' diagnoses did not match those of the decision support system, the physicians changed their diagnoses. There was a slight but significant negative correlation between susceptibility to change and experience level of the physicians. Physicians were significantly less likely to follow the decision system's recommendations when they were confident of their initial diagnoses. No differences between the physicians' inclinations to following excise versus no excise recommendations could be observed.

**Conclusion:** These results indicate that physicians are quite susceptible to accepting the recommendations of decision support systems, and that quality assurance and validation of such systems is therefore of paramount importance.

© 2004 Elsevier B.V. All rights reserved.

## 1. Introduction

With increasing quantities of clinical data being collected in hospitals around the world, the practice of evidence-based medicine is becoming more and

\* Tel.: +43 7236 3888 2025; fax: +43 7236 3888 2099.

E-mail address: [stephan.dreiseitl@fh-hagenberg.at](mailto:stephan.dreiseitl@fh-hagenberg.at)  
(S. Dreiseitl).

more feasible. The use of clinical decision support systems (CDSSs) is based on the goal of making the hospital data repositories, and the information contained therein, available to physicians at the point of care. The expected benefits of using a CDSS are an increase in the quality of treatment by providing automated alerts and consistency checks; an increase in treatment efficacy by providing best-practice guidelines; an increase in knowledge by providing information when and where it is needed, and a cost reduction by eliminating redundant tests [1,2].

While there have been enormous advances in the data-processing aspects of CDSSs, the way how to best interface with the human users is much less understood. We believe that this is one of the reasons that so few decision support systems are now in routine clinical use. Review papers [3–7] point out that the vast majority of papers reporting on CDSSs evaluate these systems in a somewhat artificial context: as stand-alone software systems that are designed to operate *in parallel to*, but not necessarily *in support of* the physician. Although the evaluation of CDSSs is, in itself, an active research area [8–13], experimental setups that evaluate the performance of a diagnostic system *in per se*, by reporting accuracy, sensitivities, specificities, or complete ROC curves, sidestep a wide array of questions raised by the use of these systems in a real clinical environment. These questions pertain mainly to the role of the system in relation to the physician; some of these questions are: what is the output of the system (dichotomous classification/differential diagnosis/probability estimates), and does this form suit the physicians' needs? Are explanations given for the system's diagnosis, and is the form of explanation satisfactory for the physicians using the system? How well does using the system fit into the regular workflow of physicians, and how intuitive is its use? Is the system's knowledge base static, or can it be updated to include new research results, or can it be adapted to a different patient population?

The literature that addresses these and similar questions is much smaller [14–16] but, we believe, more relevant to assessing the role and potential benefits of CDSSs in real clinical settings. It is especially imperative to consider the effect of CDSSs on physician performance, i.e., to what degree physician performance can be enhanced by automated systems [17–20]. In this paper, we want to extend these ideas one step further by abstracting from performance measures and looking directly at the effect that CDSSs have on physicians, and measuring to what extent these users can be influenced by CDSSs. There is little previous research in this area,

with only a few papers reporting on the interface between physicians and decision support system [21–23].

To us, the fundamental questions in this context are to which extent CDSSs have an influence on the diagnostic tasks of physicians, and whether this influence can be quantified. When CDSSs become accepted in routine clinical use, the classical setting of a routine patient–physician encounter is augmented by a software system that can be consulted by the physician. In some sense, the role of the computer is similar to that of a laboratory test result, in that it provides a piece of information for the physician. As such, only standard test characteristics such as sensitivity and specificity need to be considered in the evaluation of a CDSS. We believe, however, that the “intelligence” incorporated in these systems distinguishes a CDSS suggestion from other pieces of information available to the physician, and that (consciously or unconsciously) the physician's reaction to these data items is different from his or her reaction to other data items. It is precisely this difference that we wanted to investigate and quantify with the small study reported on in this paper.

The motivation for conducting this study was that in our own work, we sought to provide automated diagnostic support in the domain of melanoma diagnosis [24,25], and thus faced many of the same problems and questions other researchers had before us. We were, however, not only interested in the CDSS itself, but also wanted to find out how such a system can influence the diagnostic opinion of physicians, regardless of the diagnostic accuracy the system provides. To investigate this question, we set up a mock CDSS that allows us to contradict physicians on purpose and without their knowledge. This setup is necessary as otherwise there might never be a disagreement between the system, which performs at the level of experienced dermatologists, and similarly qualified physicians. We can thus determine how physicians react when a CDSS provides concordant and discordant suggestions, respectively.

## 2. Material and methods

To investigate the question of how susceptible physicians are to suggestions of a CDSS, we conducted a small study involving the diagnosis of pigmented skin lesions. The diagnosis of such lesions is difficult even for expert dermatologists, who achieve diagnostic performances of 70–90% sensitivity and specificity only after years of training, depending on patient population and visual aids used [26]. A CDSS

with high diagnostic performance could therefore support a physician faced with the task of diagnosing a lesion as either benign or malignant.

Our study involved a web-based system that operates as follows: each study participant is presented with a sequence of 25 lesion images; after seeing a lesion image, the participant has to evaluate the malignancy of the lesion on a scale of 0 (benign) to 10 (malignant), and give a dichotomous excise/no excise recommendation. The system then provides its own suggestions, after which the study participants can change their ratings and excise recommendations. The lesion images shown were true-color images obtained by epiluminescence microscopy, the correct diagnosis (gold standard) was known for all 25 lesion images. Of these images, 14 were pigmented skin lesions, and 11 cutaneous melanomas. The images were chosen to cover a wide range of diagnostic difficulty.

Under the pretense of measuring their diagnostic performance, 52 volunteer dermatologists (third-year residents and board-certified dermatologists) participated in the study on site at the University of Vienna Medical School (now: Medical University of Vienna). The subjective experience levels of the participants ranged from novice to expert, with a median of intermediate experience. To purposely mislead the participants in some cases, we gave the wrong rating and excise recommendation for 8 of the 25 lesion images. The study participants did not know that the CDSS was only simulated, and that some of the answers they were given were false.

### 3. Results

The main motivation for this study was to measure the effect that the CDSS output had on physicians' diagnoses; we therefore do not report on the diagnostic performance of the dermatologists who enrolled in this study. Since we collected both dichotomous excise/no excise recommendations as well as ordinal ratings of lesion malignancy, we were able to quantify the effect of a contradiction (CDSS differs from physician opinion on excise recommendation) as well as small to large differences in opinion on an ordinal scale (CDSS differs from physician opinion on malignancy rating). We decided to evaluate the data we collected from the point of view of the following motivating questions, which we will then address in turn: How often did physicians change their opinion when contradicted by the CDSS? Is there a tendency to follow system suggestions more willingly if it leads to a diagnosis that is on the "safe side"? Is there a correlation between experience level and the willingness to

follow CDSS suggestions? Do physicians change their minds more often when they are less confident about their initial diagnosis?

*How often did physicians change their opinion when contradicted by the CDSS?* Over all 52 study participants and 25 lesion images, the physician's initial dichotomous excise/no excise recommendation disagreed with that of the CDSS in 357 cases. In 86 of these cases (24.09%), the physicians then changed their opinion to follow the CDSS recommendation. Looking at the individual physicians, we observed that they were contradicted 4–10 times (median 7) on the 25 lesion images. In these cases, they followed the system's suggestions 0–8 times (median 1). The fraction of times that system suggestions were followed was 0–100%, with a median of 16.67%.

For the ordinal malignancy rating on a scale of 0–10, there was obviously more chance for disagreement between CDSS and physicians. For individual physicians, the range in number of lesion images for which there was disagreement was between 13 and 22 (median 18). In these cases, the physicians changed their minds 0–94.44% of the time (median 31.58%).

*Is there a tendency to follow system suggestions more willingly if it leads to a diagnosis that is on the "safe side"?* Melanoma diagnosis is an example of a diagnostic task in which false negative results have life-threatening consequences, whereas false positive results do not. It has been observed that dermatologists incorporate these cost factors in their diagnosis; this results in a high number of false positives for each false negative diagnosis [27]. The field of melanoma diagnosis is thus suited to answer the question of whether physicians would rather err on the side of caution, if a CDSS gives them an impetus to do so.

From the material collected in this study, we constructed a  $2 \times 2$  contingency table with all 357 cases in which there was disagreement between the physician's and the CDSS's excise recommendation. In 139 cases, the CDSS's conflicting recommendation was to excise; the physicians followed this advice 32 times (23.02%). In the remaining 218 cases, the physician's initial excise recommendation was contradicted by the CDSS. Here, the CDSS suggestion was followed 54 times (24.78%). Using chi-square analysis of contingency tables, we obtain the result that there is no significant difference between following CDSS recommendations in one direction or another ( $P = 0.706$ ).

*Is there a correlation between experience level and the willingness to follow CDSS suggestions?* Prior to rating the lesion images, we asked the study participants to rate their subjective level

of experience on a scale of 0 (novice) to 5 (expert). This information enabled us to investigate whether there is any correlation between the level of experience of a physician and his or her susceptibility to CDSS recommendations. This question was motivated, in part, by other studies that have shown less experienced physicians to benefit more from CDSSs than their more experienced colleagues [17,28].

From the data material collected in this study, we were able to determine a small but significant negative correlation between a physician's experience and his or her willingness to follow the CDSS's dichotomous excise/no excise recommendation (Kendall  $\tau_b = -0.20$ ,  $P = 0.047$ ). The conclusion is similar for the change of opinion on the ordinal malignancy scale, where there is also a small negative correlation (Kendall  $\tau_b = -0.225$ ,  $P = 0.026$ ).

*Do physicians change their minds more often when they are less confident about their initial diagnoses?* Clinical decision support systems should aid physicians in arriving at a diagnosis, but should not replace them. Such systems are therefore most useful for difficult cases, where physicians might not be sure about their diagnoses. CDSSs need not be consulted in cases in which physicians are able to diagnose with high confidence.

We therefore investigated the behavior of the study participants with respect to the confidence they had in their initial diagnoses. Although we did not explicitly ask for this information, there is nevertheless a proxy information available that gives an indication of a physician's confidence in recommending a lesion excision (or not): the 10-point malignancy rating. It can be argued that a malignancy rating of 8 or more means that the physician is rather confident that the lesion in question is in fact malignant. On the other end of the scale, a rating of 3 or less indicates that the physician is confident that a lesion is in fact benign. Lesion ratings in the range of 4–7 express uncertainty in the dichotomous excise/no-excision recommendation.

We analyzed the physicians' responses to see whether there was any difference in their willingness to follow CDSS recommendations when they were confident as opposed to when they were not confident. It is also interesting to check if there is a difference in behavior depending on whether a physician's initial diagnosis is right or wrong. There are thus four distinct situations, resulting from all possible combinations of confidence (high/low) and correctness (right/wrong). While we expected to see more susceptibility to CDSS recommendations when the confidence is low, we did not expect to see a difference

**Table 1** Contingency tables showing the willingness to follow CDSS recommendations when confident and not confident, respectively, stratified by correctness of initial diagnosis

	Change	No change
Initial diagnosis correct		
Confident	24	112
Not confident	26	55
Initial diagnosis incorrect		
Confident	10	49
Not confident	26	55

depending on the correctness of the initial recommendation, since, after all, the physician did not know whether his or her recommendation was correct.

To determine whether physicians change their minds less often when confident (regardless of whether they were right or not) we set up a  $2 \times 2$  contingency table. We used the table to calculate whether 34 opinion changes when confident (out of 195 times the system advised against the initial diagnosis) are significantly different from 52 cases (out of 162) when not confident. Using chi-square analysis, this difference is highly significant ( $P = 0.0019$ ). Next, we checked whether there is a difference in behavior depending on whether the initial excise recommendation is correct or not. For this, we split the data into two tables, one for the cases where the initial recommendation was correct, and one where it was incorrect. These two contingency tables are shown in Table 1. Using the Breslow–Day test for homogeneity of odds ratios, we observe that the hypothesis of equal odds ratios in both tables cannot be rejected ( $P = 0.927$ ). As expected, there is thus no indication that there is a difference in behavior between situations where the initial diagnosis is correct or incorrect, respectively.

## 4. Discussion

Over the past decade, advances in data analysis, user interfaces and integration into physician workflow have changed the perception of CDSSs. In several areas of clinical medicine, their status has changed from mere academic exercise to software tools that are useful in real-world settings. As CDSSs are on the verge of becoming accepted in routine use, the question of the interaction between system and physician becomes more and more relevant.

Since even the best CDSSs are not 100% accurate, there will be situations in routine care where CDSS recommendations are in conflict with physician

opinion. Using the study setup reported on in this paper, we obtained three results that should be considered when deploying CDSSs:

- Based on the recommendation of a CDSS, physicians are willing to change a dichotomous decision in 24% of the cases.
- The number of times a decision is reversed correlates negatively with the experience level of the physicians using the system.
- Physicians are more willing to accept a CDSS recommendation when they are not confident of their diagnosis.

These findings emphasize the importance of strict validation and quality management in the deployment of CDSSs, since physicians *do* seem to accept the advice of such systems. This holds true especially for inexperienced physicians and those that are less confident of their diagnoses, and thus more likely to accept advice.

We are aware of that fact that the approach we chose for this study, which was to mislead physicians on purpose in some cases, does not reflect the routine use of such systems in clinical medicine. Therefore conclusions obtained from this study may not directly be transferred into daily clinical routine. Furthermore, the small sample size and the in vitro-nature of the study imply that further investigations of the results we obtained are necessary. Nevertheless, we consider the cases where there is a conflict in opinion between physicians and CDSSs to be most revealing of the way that physicians trust or distrust such systems. We chose to sometimes give wrong recommendations to increase the number of times that such conflicts occur.

We believe that there are a number of interesting questions raised by our findings:

- At present, physicians are solely responsible for their diagnoses, even if these diagnoses were provided by CDSSs. What validation measures can be taken to increase physicians' confidence in these systems [8–13]?
- How should the interface between computer and physician be designed to provide the highest degree of relevant information?

The first question raises the interesting point that while sensitivity and specificity values are indicative of CDSS performance on a set of patients, they are statistical measures of limited value for the diagnosis of individual patients. Nevertheless, these numbers can be compared with those of the physician, and therefore provide an indication of when it might be beneficial to consult CDSSs.

The second question is especially relevant for CDSSs that are black-box systems that give no explanation of their reasoning (such as the one used in our study). For the first- and second-generation expert systems in medicine, the ability to reason about the diagnostic process was of paramount importance, e.g. in CASNET [29,30], MYCIN [31] and INTERNIST-1 [32] in the 1970s, and DXplain [33] and QMR [34] in the 1980s. In recent years, there has been a focus on using probability theory for quantifying uncertainty (mainly through Bayesian networks), and on data-driven approaches (possibly using machine-learning algorithms). Examples of such systems are Pathfinder [35], PapNet [36], and Hepaxpert [37], the last being a knowledge-based, data-driven system. For black-box systems, case-based reasoning [38,39] is sometimes considered a viable approach to providing explanations.

## 5. Conclusion

Decision support systems are promising software tools for providing the benefits of evidence-based medicine at the point of care. In this paper, we investigated the reaction of physicians that are faced with a CDSS recommendation that contradicts their own judgment. We observed that in 24% of the cases where such a disagreement took place, the physicians accepted the advice of the CDSS and changed their opinion. Less experienced and less confident physicians were more likely to accept the advice. As CDSSs are also prone to error, the willingness to follow CDSS recommendations may lead to incorrect diagnoses, for which the physician will be held accountable. Careful evaluations of the performance of physicians versus computer-supported physicians will be required to assess the value of CDSSs for routine clinical use.

## References

- [1] Sim I, Gorman P, Greenes R, Haynes R, Kaplan B, Lehmann H, et al. Clinical decision support systems for the practice of evidence-based medicine. *J Am Med Inform Assoc* 2001; 8(6):527–34.
- [2] Berner E, editor. *Clinical Decision Support Systems: Theory and Practice*. Springer; 1998.
- [3] Trivedi M, Kern J, Marcee A, Grannemann B, Kleiber B, Bettinger T, et al. Development and implementation of computerized clinical guidelines: barriers and solutions. *Methods Inform Med* 2002;41(5):435–42.
- [4] Kaplan B. Evaluating informatics applications—clinical decision support systems literature review. *Int J Med Inform* 2001;64(1):15–37.



- [5] Tierney W. Improving clinical decisions and outcomes with information: a review. *Int J Med Inform* 2001;62(1):1–9.
- [6] Johnston M, Langton K, Haynes R, Mathieu A. Effects of computer-based clinical decision support systems on clinician performance and patient outcome. A critical appraisal of research. *Ann Intern Med* 1994;120(2):135–42.
- [7] Miller R. 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.
- [8] Smith A, Nugent C, McClean S. Evaluation of inherent performance of intelligent medical decision support systems: utilising neural networks as an example. *Artif Intell Med* 2003;27(1):1–27.
- [9] Altman D, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000;19(4):453–73.
- [10] Friedman C, Wyatt J. Evaluation methods in medical informatics. Springer; 1997.
- [11] Wyatt J. Quantitative evaluation of clinical software, exemplified by decision support systems. *Int J Med Inform* 1997;47(3):165–73.
- [12] Nohr C. The evaluation of expert diagnostic systems. How to assess outcomes and quality parameters? *Artif Intell Med* 1994;6(2):123–35.
- [13] Wyatt J, Spiegelhalter D. Evaluating medical expert systems: what to test and how? *Med Inform Lond* 1990;15(3):205–17.
- [14] Bates D, Kuperman G, Wang S, Gandhi T, Kittler A, Volk L, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. *J Am Med Inform Assoc* 2003;10(6):523–30.
- [15] Haug P, Rocha B, Evans R. Decision support in medicine: lessons from the help system. *Int J Med Inform* 2003;69(2–3):273–84.
- [16] McCowan C, Neville R, Ricketts I, Warner F, Hoskins G, Thomas G. Lessons from a randomized controlled trial designed to evaluate computer decision support software to improve the management of asthma. *Med Inform Internet Med* 2001;26(3):191–201.
- [17] Friedman C, Elstein A, Wolf F, Murphy G, Franz T, Heckerling P, et al. Enhancement of clinicians' diagnostic reasoning by computer-based consultation: a multisite study of 2 systems. *JAMA* 1999;282(19):1851–6.
- [18] Berner E, Maisiak R, Cobbs C, Taunton O. Effects of a decision support system on physicians' diagnostic performance. *J Am Med Inform Assoc* 1999;6(5):420–7.
- [19] 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. *J Am Med Inform Assoc* 1998;280(15):1339–46.
- [20] Elstein A, Friedman C, Wolf F, Murphy G, Miller J, Fine P, et al. Effects of a decision support system on the diagnostic accuracy of users: a preliminary report. *J Am Med Inform Assoc* 1996;3(6):422–8.
- [21] Friedman C, Gatti G, Elstein A, Franz T, Murphy G, Wolf F. Are clinicians correct when they believe they are correct? Implications for medical decision support *Medinfo* 2001; 10(Pt 1):454–8.
- [22] Berner E, Maisiak R. Influence of case and physician characteristics on perceptions of decision support systems. *J Am Med Inform Assoc* 1999;6(5):428–34.
- [23] Ridderikhoff J, van Herk B. Who is afraid of the system? Doctors' attitude towards diagnostic systems. *Int J Med Inform* 1999;53(1):91–100.
- [24] Binder M, Kittler H, Dreiseitl S, Ganster H, Wolff K, Pehamberger H. Computer-aided epiluminescence microscopy of pigmented skin lesions: The value of clinical data for the classification process. *Melanoma Res* 2000;10(6):556–61.
- [25] Dreiseitl S, Ohno-Machado L, Vinterbo S, Billhardt H, Binder M. A comparison of machine learning methods for the diagnosis of pigmented skin lesions. *J Biomed Inform* 2001;34(1):28–36.
- [26] Kittler H, Pehamberger H, Wolff K, Binder M. Diagnostic accuracy of dermoscopy. *Lancet Oncol* 2002;3(3):159–65.
- [27] Kittler H, Binder M. Risks and benefits of sequential imaging of melanocytic skin lesions in patients with multiple atypical nevi. *Arch Dermatol* 2001;137(12):1590–5.
- [28] Chang P, Li Y, Wang T, Huang S, Hsieh M, Tsui K. Evaluation of a decision-support system for preoperative staging of prostate cancer. *Med Decis Making* 1999;19(4):419–27.
- [29] Weiss S, Kulikowski C, Amarel A, Safir A. A model-based method for computer-aided medical decision-making. *Artif Intell* 1978;11(1–2):145–72.
- [30] Weiss S, Kulikowski C, Safir A. Glaucoma consultation by computer. *Comput Biol Med* 1978;8(1):24–40.
- [31] Shortliffe E. Computer-based Medical Consultations: MYCIN. Elsevier; 1976.
- [32] Miller R, Pople H, Myers J. Internist-1, an experimental computer-based diagnostic consultant for general internal medicine. *NEJM* 1982;307(8):468–76.
- [33] Barnett G, Cimino J, Hupp J, Hoffer E. DXplain. An evolving diagnostic decision-support system. *JAMA* 1987;258(1):67–74.
- [34] Miller R, Masarie F, JD JM. Quick medical reference (QMR) for diagnostic assistance. *MD Comput* 1986;3(5):34–48.
- [35] Heckerman D, Horvitz E, Nathwani B. Toward normative expert systems: Part I. The Pathfinder project. *Methods Inf Med* 1992;31(2):90–105.
- [36] Mango L. Computer-assisted cervical cancer screening using neural networks. *Cancer Lett* 1994;77(2–3):155–62.
- [37] Adlassnig K, Horak W. Development and retrospective evaluation of Hepaxpert-I: a routinely-used expert system for interpretive analysis of hepatitis A and B serologic findings. *Artif Intell Med* 1995;7(1):1–24.
- [38] Reategui E, Campbell J, Leao B. Combining a neural network with case-based reasoning in a diagnostic system. *Artif Intell Med* 1997;9(1):5–27.
- [39] Stamper R, Todd B, Macpherson P. Case-based explanation for medical diagnostic programs, with an example from gynaecology. *Methods Inf Med* 1994;33(2):205–13.