

# Comparing Bedside Information Tools: A User-Centered, Task-Oriented Approach

Rose Campbell, MILS, Joan Ash, PhD

Department of Medical Informatics and Clinical Epidemiology, School of Medicine  
Oregon Health Sciences University, Portland, Oregon

## ABSTRACT

*There are a number of electronic products designed to provide information at the point of care. These bedside information tools can facilitate the practice of Evidence Based Medicine. This paper evaluates five of these products using user-centered and task-oriented methods. Eighteen users of these products were asked to attempt to answer clinical questions using a variety of products. The proportion of questions answered, and user satisfaction were measured for each product. Results show that proportion of questions answered was correlated with user satisfaction with a product. When evaluating electronic products designed for use at the point of care, the user interaction aspects of a product become as important as more traditional content-based measures of quality. Actual or potential users of such products are in the best position to identify which products rate the best in these measures.*

## INTRODUCTION

In the past decade, Evidence Based Medicine (EBM) has become mainstream (1). EBM can be defined as “the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients” (2). In its original conception, EBM asked practitioners to be able to access the literature at the point of need. This process included four steps. First, formulate a clear clinical question from the patient’s problem. Second, search the primary literature for relevant clinical articles. Third, critically appraise the evidence for validity and finally, implement these findings in practice (3).

Physicians lack the time to perform all of the steps outlined above. In an observational study, participating physicians spend less than two minutes seeking the answer to a clinical question (4). In addition to the lack of time, many clinicians may also

lack the skills needed to search, analyze and synthesize the primary literature (5). While it could be argued that clinicians could or ought to learn these skills, there are already other professionals with these skills. In a modern team-based approach to healthcare it is natural for others to do the information synthesis and analysis. In the McColl study cited above, 37% of respondents felt the most appropriate way to practice EBM was to use “guidelines or protocols developed by colleagues for use by others.” Only 5% felt that “identifying and appraising the primary literature” was the most appropriate. With others doing searching and synthesizing, assembling the evidence no longer needs to be done as the need arises. It is access to these pre-digested forms of information that most readily supports the use of evidence in clinical decision-making.

There are many electronic products, which have been specifically designed to provide this sort of synthesized information to the clinician. For our purposes we’ll call these bedside information tools. Because they have been designed to provide information at the point of care, they tend to be action oriented. They are designed to answer questions directly related to patient care rather than background questions. Because these resources are electronic they can be updated much more frequently than print resources and it can be easier to search for and share information.

As with other forms of secondary literature, there are a number to choose from. With limited financial resources, libraries and other institutions cannot afford to subscribe to or purchase all available bedside information tools. Providing access to all the bedside information tools is not a practical solution. The question then is, which bedside information tools should be selected and supported within an institution or library? In order to make this decision we need a method for evaluating these products.

When looking for evaluation methods one can turn to evaluation methods used to evaluate other information retrieval (IR) systems. IR evaluation methods can roughly be grouped into system-centered or user-centered. Popular system centered metrics for evaluating IR systems are recall and precision, but because of their focus on document retrieval and relevance, recall and precision may not be suitable at all to measure whether or not a given system can answer clinical questions. In a 2002 study, Hersh found no correlation between recall and precision measures and the ability of students to answer clinical questions using MEDLINE (6). Recall and precision, therefore are not appropriate for the evaluation of these bedside information tools.

Another System-Centered evaluation is system usage. While these measurements can show how much a system is used relative to others, the measure is only use, not successful use. What can be extrapolated beyond that the product is being used is questionable. Usage data can also only be collected on products available for use, so one cannot collect data on products only being considered for purchase. System usage is also inappropriate for the evaluation of these bedside information tools.

User-Centered evaluations look at the user and the system as a whole when evaluating IR systems. Task-Oriented evaluations ask users attempt a task on the system, then asks was the user able to complete the task. A Task-Oriented evaluation of bedside information tools asks the question "Can users find answers to clinical questions using this tool?" By having users attempt to complete these tasks and measuring their success we can answer this question for these products. By looking at both the system and the user these evaluations measure whether or not an IR system can do what it is designed to do.

User-satisfaction goes one step further and asks not only "Did the system do what it was designed to do?" but also "Was the user satisfied with the way it did it?" By measuring user satisfaction an evaluation can identify not only those products, which successfully answer clinical questions, but also those products that the users enjoy using the most.

Evaluating and selecting products for others to use can pose a challenge. Evaluators need to be explicit in including users in these evaluations. Although products must meet minimum standards of quality, by not addressing users in an evaluation, one risks choosing a product that does not suite the user's needs. Therefore, a user-centered, task-oriented

approach is needed to evaluate bedside information tools.

The specific products evaluated in this project are ACP's *PIER*, *DISEASEDEX* from Thomson Micromedex, *FIRSTConsult* from Elsevier, *InfoRetriever* and *UpToDate*. These products were selected for practical considerations. These products are market peers, and would consider each other to be their competition.

## METHODS

Eighteen potential users of the product were asked to evaluate each of the products. Users were given test clinical questions to attempt to answer using the products. After using each product, they were asked to evaluate their user experience by answering a user evaluation questionnaire. They were also asked to answer a background questionnaire designed to measure differences in occupation, occupational experience, computer experience and searching experience. Users also participated in a semi-structured follow-up interview. Because of space considerations, the qualitative results of these interviews are not reported here.

### Trial Questions

The test questions used for the evaluations come from a corpus of clinical questions recorded during ethnographic observations of practicing clinicians (7, 8). In order to create a set of questions for which these products would be the appropriate resource to find the answers a number of categories of questions were removed from consideration. Patient data questions, population statistic questions, logistical questions and drug dosing and interaction questions were removed from consideration. The remaining questions can be categorized at diagnosis, treatment and prognosis and treatment questions. This set of questions was then tested by a group of pilot users. Any questions that were not answered with any of the databases were removed. Some questions were reworded to be clearer.

### Experimental Design

The experimental design of the allocation of questions, databases and testers is based on the experimental design specifications set out by the TREC-6 Interactive Track (9, 10). Although five databases are being compared rather than two, the time commitment for the tester is about the same. Users were instructed to spend no more than three minutes on each question, significantly less time than the 20 minute time limit used in TREC-6 experiments

(11). For each tester three clinical questions were pseudo-randomly assigned to the databases such that no question is assigned to more than one database. The order that the databases are to be evaluated was also randomly assigned. Each tester attempted to answer three questions with a specified product, answered the user satisfaction questions, and then moved on to the next bedside information tool and set of three questions. Users were only asked if they found the answer to a question, the answers that they found were not recorded and no attempt was made to verify the correctness of the answer. This evaluation process was tested using pilot users. It was determined that the time commitment for the process was about an hour and that the instructions for logging onto the products and progressing through the evaluations were clear.

### Questionnaires

The user evaluation questionnaire was developed from the literature. User evaluation questions were extracted from 11 studies (12-22). These questions were then sorted such that variant forms of the same question were grouped together. The seven questions that were asked most frequently were transformed into questions that could be answered using a 5-point Likert Scale. A final opportunity for comments was added at the end.

The background questionnaire was also guided by the literature, although less explicitly. After examining several examples of background questions (6, 14, 21-24), a questionnaire was developed to address what would be likely sources of variation within users: age, gender, profession, years at profession, experience with computers in general and experience with searching.

### User Recruitment

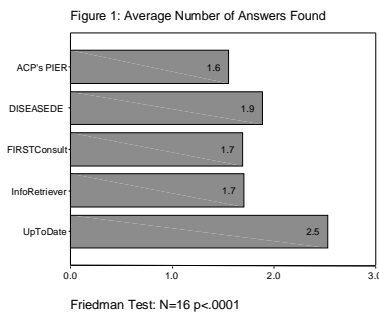
After receiving approval from OHSU's Internal Review Board, participants were recruited via personal contacts, posting on listservs and websites as well as flyers were posted on campus. Classes of individuals targeted included physicians, residents, medical students, physician assistants, physician assistant students, and nurses, nursing students, pharmacists and informatics students.

As individuals agreed to participate in the study, an investigator met with them personally. At this meeting the evaluation protocol and background to the study were discussed and the participants were consented.

## RESULTS

### Average Number of Answers Found

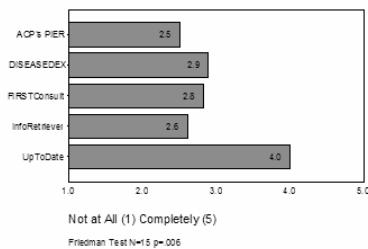
In this analysis, the total number of answers found by each participant in each database was recorded. Each participant found answers to 0, 1, 2 or 3 questions. On average, participants found answers to 2.5 questions in UpToDate, 1.9 questions in DISEASEDEX, 1.7 questions in both FIRSTConsult and InfoRetriever and 1.6 questions in ACP's PIER. The Friedman Test was used to test the null hypothesis that the average number of questions answered was the same for each database. This hypothesis is rejected with a p-value of <0.001 (N=16). (See Figure 1).



### Overall this system satisfied my needs: completely/not at all

UpToDate rated much higher than any other product on this scale, receiving a mean score of 4.0. DISEASEDEX rated a mean of 2.9; FirstConsult rated a mean of 2.8; InfoRetriever a mean of 2.6 and ACP's PIER a mean of 2.5. The hypothesis that these databases received the same rating was rejected by the Friedman Test (N=15 p=.006) (See Figure 2).

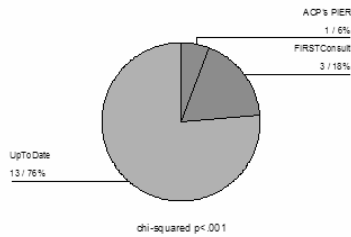
Figure 2: Overall, did this Database Satisfy Your Needs:



### Percent of users who ranked the products the best

76% of respondents (13) ranked UpToDate the best. 18% of respondents (3) ranked FIRSTConsult the best and 6% (1) respondent ranked ACP's PIER the best. No respondents rated DISEASEDEX or InfoRetriever to be the best. A Chi-squared analysis shows these values to be different at a level of p<.001 (test statistic  $35.65 > X^2_{4, .001} = 18.47$ ) (See Figure 3).

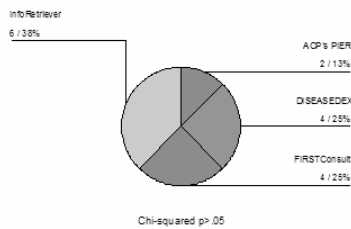
Figure 3: Percent of Respondents who Ranked this Product the Best



### Percent of users who ranked the products the worst

38% of respondents (6) ranked InfoRetriever to be the worst. 25% (4) ranked DISEASEDEX to be the worst and the same proportion of respondents ranked FIRSTConsult to be the worst. ACP's PIER had 13% of respondents (2) ranking it to be the worst. No respondents ranked UpToDate to be the worst. A Chi-squared analysis does not show these values to be different at a level of  $p < .05$  (test statistic  $6.18 < X^2_{4, .05} = 9.49$ ) (See Figure 4).

Figure 4: Percent of Respondent who Ranked this Product the Worst



## DISCUSSION

When looking at the user interaction measures and user ranking of products, UpToDate is the clear winner. Users found that overall, UpToDate met their needs, most users ranked UpToDate to be the best product, and none rated it the worst.

These findings are consistent with studies that tested what tools users selected when they have access to more than one tool. These use studies, while not usually exploring the reasons that people use the products they use, do indicate that when given a choice, users pick UpToDate as a resource very frequently (25, 26).

Evaluations that focus on the content of an information resource rather than on the user, as is the case for much of traditional collection development evaluation models, would have not revealed the strong user preference for UpToDate. These types of products must be evaluated in terms of user

interaction as well as their content. A product selected because of excellent content may be rendered useless by a difficult user interface. To use an analogy from more traditional formats, the best textbook in terms of content is not a good buy if it is sold with its pages glued shut.

As electronic formats become the expected form for information, as is the case of the products being evaluated here, the interface takes on greater significance. In the user's mind the presumed similarity of content becomes insignificant compared to differences in the user interface. In the case of these types of bedside information tools, the interface is the product as much, if not more, than the content of the product. This only underscores that the traditional content focused evaluation methods may no longer be sufficient in the evaluation of this new breed of information tools.

There are a number of limitations worthy of discussion. Although thorough efforts were made to recruit all potential users, the evaluators were essentially a convenience sample of volunteers. These volunteers may not represent the total user population, as they tended to have an interest in these types of products to begin with. Their responses may not be typical of other users who, having no interest in the technology, are still "forced" to use these types of products for one reason or other. Another result of relying on pure volunteerism is that the users in various classes were not evenly represented. Input from other types of clinicians, including more nurses, physician's assistants, dentists etc may have changed the results. Although these products are designed for use at the point of care they also have great value for students. The only students included in this study were medical informatics students. Medical and other clinical specialty students may have had very different opinions about the products.

Unlike some other studies (6, 13, 23) no training was provided. Users may have had different opinions about some products if they had been trained on their optimal use. However, users were allowed to familiarize themselves with the product before the evaluation if they wished, and previous experience with products was recorded with the background questionnaire. A user's previous experience or familiarity with the product did not bias their opinion of the product. It can also be argued that sitting down to a resource with no training whatsoever is standard practice. This study replicated that. If a product requires training in order for users to be successful with it, those training costs must be considered when comparing to other options

## REFERENCES

1. Guyatt G, Cook D, Haynes RB. Evidence based medicine has come a long way: the second decade will be as exciting as the first. *British Medical Journal* 2004;329:990-991.
2. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *British Medical Journal* 1996;312(7023):71-72.
3. Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. *British Medical Journal* 1995;310(6987):1122-1126.
4. Ely JW, Osheroff JA, Ebell MH, Bergus G, Levy BT, Chambliss ML, et al. Analysis of questions asked by family doctors regarding patient care. *British Medical Journal* 1999;319(7206):358-361.
5. McColl A, Smith H, White P, Field J. General practitioners' perceptions of the route to evidence based medicine: a questionnaire survey. *British Medical Journal* 1998;316(7128):361-367.
6. Hersh WR, Crabtree MK, Hickam DH, Sacharek L, Friedman CP, Tidmarsh P, et al. Factors associated with success in searching MEDLINE and applying evidence to answer clinical questions. *Journal of the American Medical Informatics Association* 2002;9(3):283-293.
7. Gorman PN. Information needs of physicians. *Journal of the American Society for Information Science* 1995;46(10):729-736.
8. Gorman PN, Helfand M. Information seeking in primary care: how physicians choose which clinical questions to pursue and which to leave unanswered. *Medical Decision Making* 1995;15(2):113-119.
9. Over P. TREC-6 Interactive Track Home Page. In: TREC-6; 2000.
10. TREC-6 Interactive Track Specification. In.
11. Belkin NJ, Carballo JP, Cool C, Lin S, Park SY, Reih SY, et al. Rutgers' TREC-6 Interactive Track experience. In: Voorhees EM, Harman DK, editors. *The Sixth Text REtrieval Conference (TREC-6)*. Gaithersburg, MD; 1998. p. 597-610.
12. Bachman JA, Brennan Pf, Patrick TB. Design and evaluation of SchoolhealthLink, a web-based health information resource. *The Journal of School Nursing* 2003;19(6):351-357.
13. Kupferberg N, Jones L. Evaluation of five full-text drug databases by pharmacy students, faculty, and librarians: do the groups agree? *Journal of the Medical Library Association* 2004;92(1):66-71.
14. Kushniruk AW, Patel C, Patel ML, Cimino JJ. 'Televaluation' of clinical information systems: an integrative approach to assessing web-based systems. *International Journal of Medical Informatics* 2001;61(1):45-70.
15. Ma W. A database selection expert system based on reference librarian's database selection strategy: a usability and empirical evaluation. *Journal of the American Society for Information Science and Technology* 2002;53(7):567-580.
16. Gadd CS, Baskaran P, Lobach DF. Identification of design features to enhance utilization and acceptance of systems for internet-based decision support at the point of care. *Proceedings/AMIA Annual Fall Symposium* 1998:91-5.
17. Jones J. Development of a self-assessment method for patients to evaluate health information on the internet. *Proceedings/AMIA Annual Fall Symposium* 1999:540-4.
18. Doll WJ, Torkzadeh G. The Measurement of end-user computing satisfaction. *MIS Quarterly* 1988;June 1988:259-273.
19. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction* 1995;7(1):57-58.
20. Abate MA, Shumway JM, Jacknowitz AI. Use of two online services as drug information sources for health professionals. *Methods of Information in Medicine* 1992;31(2):153-8.
21. Pugh GE, Tan JKH. Computerized databases for emergency care: what impact on patient care? *Methods of Information in Medicine* 1994;33(5):507-13.
22. Su LT. A Comprehensive and systematic model of user evaluation of web search engines : I. theory and background. *Journal of the American Society for Information Science and Technology* 2003;54(13):1175-1192.
23. Hersh W, Crabtree MK, Hickam D, Sacharek L, Rose L, Friedman CP. Factors associated with successful answering of clinical questions using and information retrieval system. *Bulletin of the Medical Library Association* 2000;88(4):323-331.
24. Cork RD, Detmer WM, Friedman CP. Development and initial validation of an instrument to measure physicians' use of, knowledge about and attitudes toward computers. *Journal of the American Medical Informatics Association* 1998;5(2):164-176.
25. Schilling LM, Steiner JF, Lundahl K, Anderson RJ. Residents' patient-specific clinical questions: opportunities for evidence-based learning. *Academic Medicine* 2005;80(1):51-56.
26. Huan, G, Sauds, D, Loo T. Housestaff use of medical references in ambulatory care. *AMIA proceedings* 2003.