

COMPARING BEDSIDE INFORMATION TOOLS: A USER-CENTERED, TASK-  
ORIENTED APPROACH

by  
Rose Campbell, MLIS

A MASTER'S THESIS

Presented to the Department of Medical Informatics and Clinical Epidemiology  
and the Oregon Health & Science University  
School of Medicine  
in partial fulfillment of  
the requirements for the degree of  
Master of Science  
May 2005

School of Medicine  
Oregon Health & Science University

---

CERTIFICATE OF APPROVAL

---

This is to certify that the Masters thesis of  
Rose Campbell  
has been approved

---

William Hersh, MD

---

Joan Ash, PhD

---

Valerie King, MD, MPH

# TABLE OF CONTENTS

List of Figures .....	iii
List of Tables.....	iv
Acknowledgments.....	v
Abstract .....	vi
Introduction .....	1
Clinical Information and Evidence Based Medicine.....	1
Bedside Information Tools.....	2
The Library's Role.....	4
Evaluating Bedside Information Tools .....	5
The Evaluation Project .....	8
Methods.....	10
Database Selection .....	10
Direct Examination .....	10
User Evaluation.....	12
Trial Questions .....	12
Questionnaires.....	14
User Recruitment .....	15
Semi-structured interviews .....	17
Results.....	18
Direct Examination .....	18
Product Information .....	18
Content .....	18
Features .....	19
Summary.....	20
User Evaluation.....	26
User Demographics .....	26
Recruitment Methods.....	26
Task Measures .....	27
User Satisfaction Measures.....	29
User Perceptions of Content.....	30
User Interaction Measures .....	33
Qualitative data .....	39
ACP's PIER .....	39
DISEASEDEX.....	40
FIRSTConsult.....	40
InfoRetriever .....	40
UpToDate .....	41
General Comments .....	41

Discussion .....	43
Content Measures and Interface Measures .....	43
The Interface is the Product.....	44
“Librarian Knows Best” Model Not Best.....	45
User Testing Essential and Feasible .....	46
Limitations .....	46
Further Research.....	49
Summary .....	50
References .....	51
Appendix A: Clinical Questions .....	58
Appendix B: User Satisfaction Questionnaire .....	59
Appendix C: Background Questionnaire .....	60
Appendix D: Study Website.....	61
Appendix E: Recruitment Text.....	62
Appendix F: Consent Form.....	63
Appendix G: Data Collection Sheet (Sample).....	67
Appendix H: Interview Guide .....	69
Appendix I: Screen Shots .....	70

## List of Figures

Figure 1: Average Number of Answers Found.....	28
Figure 2: The content of the database seemed: inaccurate (1) or accurate (5). ....	30
Figure 3: Did the database give you: not enough information (1) or too much information (5).....	31
Figure 4: Did the information in the databases seem: out of date (1) or up to date (5).....	32
Figure 5: Interacting with the system was: hard (1) or easy (5).....	34
Figure 6: The layout of the screen was: confusing (1) or clear (5).....	35
Figure 7: The database was: slow(1) or fast (5).....	36
Figure 8: Overall, this system satisfied my needs: not at all (1) or completely (5).....	37
Figure 9: Percent of respondents rating this product the best.....	38
Figure 10: Percent of respondents rating this product the worst.....	39

## List of Tables

Table 1: Product Information.....	22
Table 2: Content.....	23
Table 3: Content Cont. ....	24
Table 4: Features.....	25
Table 5: Average Number of Questions Answered, Pair-wise comparisons.....	28
Table 6: Time Spent Searching, Survival Times.....	29
Table 7: Accurate/Inaccurate, Pair-wise comparisons.....	31
Table 8: Too Much Information/Not Enough Information, Pair-wise comparisons.....	32
Table 9: Up to Date/Out of Date, Pair-wise comparisons.....	33
Table 10: Easy/Hard, Pair-wise comparisons.....	34
Table 11: Clear/Confusing, Pair-wise comparisons.....	35
Table 12: Fast/Slow, Pair-wise comparisons.....	36
Table 13: Overall, Pair-wise comparisons.....	37

## **Acknowledgments**

I would like to thank the National Library of Medicine for funding my fellowship during my time at OHSU and the Northwest Health Foundation for providing funding for this project. I would also like to thank my academic advisor, Joan Ash as well as Dolores Judkins from the OHSU Library, Paul Gorman and Holly Jimison who served on my thesis advisory committee. Of course, all of the people who either participated in the evaluation or served as pilot testers need to be thanked. This project could not have completed without them. Finally, I would like to thank my partner Roland Relevo, I could not have done this without his support.

## **Abstract**

**Purpose:** To compare several bedside information tools using user-centered task-oriented measures in order to assist those making or supporting purchasing decisions.

**Methods:** Eighteen users were asked to attempt to answer clinical questions using a variety of products. Users evaluated each tool for ease-of-use and user satisfaction. The average number of questions answered, time spent searching and user satisfaction were measured for each product. A follow-up interview qualitatively captured users' experiences with these bedside information tools. This user-based information was combined with information gathered from direct examination, such as currency, coverage and subscription information.

**Results:** Results show no significant differences in the time spent searching or in user perceptions of content. However, user interaction measures show a significant preference for the UpToDate product. In addition, users found answers to significantly more questions using UpToDate. Qualitative data reveal that the simplicity of the search screens and the clear subdivision of topics were major reasons for the preference for UpToDate.

**Conclusion:** 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 highest on these measures. It may not be as difficult to engage users for this type of evaluation as is imagined.



## Introduction

### *Clinical Information and Evidence Based Medicine*

In the past decade, Evidence Based Medicine (EBM) has become mainstream (Guyatt et al. 2004; Straus et al. 2004). EBM can be defined as “the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients.” (Sackett et al. 1996) On its face, it only makes sense for the information gleaned from medical research to be used in medical practice. The sheer amount of medical research and information arising from it makes it virtually impossible for a clinician to keep up with the literature, acquiring information on a just in case manner. In its original conception, EBM did not ask that physicians keep up with the medical literature but 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 (Rosenberg et al. 1995).

If we look at Shughnessy’s equation of the usefulness of medical information where  $usefulness = (relevance * validity) / work$  (Shaughnessy et al. 1994), the traditional view of EBM produces low usefulness information simply because the work needed to get it is quite high. Physicians lack the time to perform all of the steps outlined above. In an observational study, participating physicians spent less than two minutes seeking the answer to a clinical question (Ely et al. 1999). In addition to the lack of time, many clinicians may also lack the skills needed to search, analyze and synthesize the primary literature (McColl et al. 1998). While it could be argued that clinicians could or ought to learn these skills, there are already other professionals with these skills. In the 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 and assembling, the evidence no longer needs to be gathered as the need arises. Rather, these synthesized resources can be created just in case, for clinicians to access just in time. It is access to these pre-digested forms of information that most readily supports the use of evidence in clinical decision-making (Chambliss et al. 1996; Ebell 2003). Looking at Shughnessy’s equation again, these resources can drastically decrease the work element of the equation while ensuring high values for validity, thus increasing the over-all usefulness of the same information. It is important to remember that in order to produce more useful information, synthesized information must maintain high validity and that some mechanism is needed to ensure relevance without increasing the work. Synthesizing information of poor quality or presenting it in such a way that makes accessing relevant information difficult does not increase the usefulness of information.

### ***Bedside Information Tools***

There are many electronic products that have been specifically designed to provide this sort of synthesized information to the clinician. For our purposes we will call these “bedside information tools”. These products differ from other secondary sources of information such as textbooks or review articles for a number of reasons. Because they have been designed to provide information at the point of care they tend to be action oriented. That is, they are designed to answer acute clinical information needs. Existing secondary sources are often excellent sources for background information, but are poor sources for information needs related specifically to patient care (Grandage et al. 2002).

Because these resources are electronic they can be updated much more frequently than print resources. Additionally, it can be easier to search for and share information in an electronic format. Perhaps the greatest advantage these products have over traditional secondary sources is their potential to be integrated with other electronic products. These tools can be integrated into an electronic health record, link to other electronic resources subscribed to by an institution or library or link out to material found on the World Wide Web.

As with other forms of secondary literature, there are a number to choose from. With limited financial resources, libraries and other institutions or individual practitioners cannot afford to subscribe to or purchase all available bedside information tools. Even in the case of freely available tools (ones that do not require purchase or subscription), there are still the costs associated with supporting the products. Each product has costs associated with making sure that the product is indeed up, running and properly interconnected with other resources. Additionally, each product requires that time and resources be allocated to promoting the product and training users. Even if the institution or library had unlimited financial resources to purchase, maintain and train users of these bedside information tools, there are individual costs associated with using tools such as these. Learning to navigate and use each new product incurs a cost in time and memory. Learning one interface well may be preferable to having access to many bedside information tools but not being familiar with or able to use any of them very well.

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?

## *The Library's Role*

Academic medical and hospital libraries have long been tasked with selecting which information products such as books and journals to purchase and make available to their patrons. This is often guided by a collection development policy, which states what topics and formats are collected, and to what extent. For example, a subject may be only collected at a basic level or it may be collected at a research level. The OHSU Library's collection development policy can be found on their website (2004d). These decisions are driven by the population that the library serves. This includes people such as researchers, clinicians, and students as well as what departments, disciplines and programs the library serves. This ongoing selection process is called collection development.

There are standard tools and resources used in collection development. Some are general criteria used to determine the quality of an item. Such considerations include: content (Is the content factual? Is it a topic supported by the library?), authority (Are the authors and or publishers known and respected?), scope (What aspects of the topic does it cover?), audience (For whom was this content written and does it match the library's patrons?) and timeliness (Is it current?). These criteria are usually evaluated by looking at the item directly.

In addition, librarians have traditionally used review tools to aid their decisions. For example, the Institute for Scientific Information (ISI)'s *Journal Citation Reports* contain citation analysis data, including ISI impact factors, often used in decisions about which journals to subscribe to (Black 1999). Another commonly used tool has been the *Brandon/Hill Selected Lists*. Compiled by individuals using the criteria outlined above the *Brandon/Hill Selected Lists* have long been recognized as containing the "best" core books and journals available for medicine, nursing and allied health. Although the *Brandon/Hill Selected Lists* have recently been discontinued (2004a) their rapid replacement by *Doody's Core Titles*

(2004b) is testament to how essential librarians have found these tools for collection development.

As formats and types of information products change, so must the evaluation tools and resources used by those making collection development decisions. The evaluation of bedside information tools poses a challenge, as there are few tools available to assist the evaluator. Product reviews are sparse. Published reviews were found for each of the five products being evaluated, however, more than one review was found for only four products (Badgett et al. 2002; Brown 2004; Fitzpatrick 2004; Fox et al. 2003; Fox et al. 2004; Garrison 2003; Howse 2002; Kennedy 2004; Simonson 2000; Taylor 2004; Tomasulo 2004). Even when published reviews can be found, they are often out of date, as these products update their interfaces often. Often these reviews are only for the PDA version of the product or similarly limited in scope. In addition, published articles about a product are often authored by those involved with the development of the product itself (Alper 2003; Alper et al. 2001; Badgett & Murlow 2002; Ebell et al. 1998; Ebell et al. 1999; Ebell et al. 2002). There are no centralized reviewers, as with the *Brandon/Hill Selected Lists*, or generally accepted metrics such as ISI impact factors. While the *HIMSS Solution Toolkit* (2004c) is a potential evaluation tool, bedside information tools generally fall outside of its scope. Only two of the products being evaluated here are included. Without readily available evaluations of these tools librarians and other decision makers are forced to evaluate these tools themselves (Chueh et al. 1997).

### ***Evaluating Bedside Information Tools***

When looking for evaluation methods one can turn to evaluation methods used to evaluate other information retrieval (IR) systems. (Dawes et al. 2003; Hersh et al. 1998; Smith 1996) review and summarize IR evaluations in the context of

physician information needs. IR evaluation methods can roughly be grouped into system-centered or user-centered. Common system-centered metrics for evaluating IR systems are recall and precision. These measurements are used within the context of trying to retrieve relevant documents from a pool of documents. Recall refers to the number of relevant documents in the pool that can be found in your retrieval set. Precision refers to the number of documents in your retrieval set that are in fact relevant. Recall and precision are often at odds with each other. If one wants to be sure that all relevant documents are retrieved (high recall) one must contend with a number of irrelevant documents being retrieved as well (low precision). Alternatively, if one wants everything in your retrieval set to be relevant (high precision) you must do so at the cost of missing some relevant documents (low recall). The optimal tradeoff between recall and precision depends on the particular circumstances of the search. Recall and precision measurements do not take into account the fact that the user determines the optional recall and precision levels for their purposes.

Because of their focus on document retrieval and relevance, recall and precision may not be suitable to measure whether or not a given system can answer an acute clinical information need. In a 2002 study, Hersh found no association between recall and precision measures and the ability of students to answer clinical questions using MEDLINE (Hersh et al. 2002). Recall and precision, therefore may not be appropriate for the evaluation of these bedside information tools.

Another system-centered evaluation metric is system usage. While these measurements can show how much a system is used relative to others (Abate et al. 1992; Arrol et al. 2002; Barnett et al. 2004; Connelly et al. 1990; Curley et al. 1990; Magrabi et al. 2004; Peterson et al. 2004), or compare usage before and after some intervention, such as a training class (D'Alessandro et al. 2004), the

measure is only use, not successful use. What can be extrapolated beyond that the product is being used is questionable. There many reasons why a system may get high use other than its ability to answer clinical questions. If a product is prominently linked or familiar to the user it may very well get more use than a “better” product which is more obscure or difficult to get to. For example, log-on difficulties may prevent the use of an otherwise excellent product. 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 therefore 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 to attempt a task on the system, then ask was the user able to complete the task (Hersh et al. 1996). A task-oriented evaluation of bedside information tools asks the question: Can users solve acute clinical information needs 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 (Hersh & Hickam 1998).

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? In addition to measuring success or failure for each attempted task, user satisfaction evaluations measure the user’s satisfaction with the process of completing a task. IR systems may have an equal success rate for task completion but differ in how satisfied users are with the process. By measuring user satisfaction, an evaluation can identify not only those products that 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 suit the user's needs. Therefore, a user-centered, task-oriented approach is needed to evaluate bedside information tools.

### *The Evaluation Project*

The specific products evaluated in this project are ACP's PIER, DISEASDEX from Thomson/Micromedex, FIRSTConsult from Elsevier, InfoRetriever and UpToDate. These products were selected for practical considerations. Most are produced by major medical publishers and they are the products that most libraries must choose between. In discussions with other medical librarians, these are the products that come up as products under serious consideration. These products would consider each other to be their competition. In addition, these are the products that the OHSU Library is currently considering for purchase.

The primary goal of this project was to evaluate these five bedside information tools using a user-centered, task-oriented approach. This has immediate value for the OHSU Library and others facing tough choices among information products.

An auxiliary goal was to develop a methodology for this sort of evaluation that can be used by other librarians in the future or with different products. Finally this project elicits general information about what users want in bedside information tools and why they prefer the ones that they do. This will inform



future decisions and evaluations as well as help developers to develop and refine this type of tool.

## **Methods**

### ***Database Selection***

As mentioned above the products being compared were selected on the basis of their being under consideration by the OHSU Library. The Library already subscribes to UpToDate, DISEASEDEX through its MICROMEDEX Subscription and ACP's PIER through its subscription to STAT!Ref. ACP's PIER is also available as a stand-alone product to ACP members. This study will evaluate ACP's PIER as part of STAT!Ref, as this is how most institutions have access to the product. Trial access was arranged for the remaining two databases InfoRetriever and FIRSTConsult.

### ***Direct Examination***

The initial piece of the product evaluation will be referred to as "direct examination". It is essentially a feature comparison of the various products. This methodology is most similar to traditional collection development evaluations. The information gleaned from this sort of an evaluation is useful, as some measure of quality must be assured. Additionally, in some situations a particular feature may be absolutely necessary for a product to be selected. Lord describes a similar checklist of features to examine when considering an electronic resource (Lord et al. 2001). This part of the evaluation has the added advantage of being able to be completed by a single researcher. A Microsoft Access database was created to store and present the results of this examination. One advantage of creating this database, which contains an input form and several output reports, is that the same database can be used to compare other bedside information tools or even the same tools if their features have changed. This database structure can be shared so that other evaluators can use it to gather and present information about resources being considered.

The features being compared can be grouped into three broad categories, basic information about the products, information concerning the content of the databases and information about the features of the user interface of the product.

Product information consists of the following:

- Name of the product
- Website for the product or company
- Updating frequency
- Pricing
- PDA information (Is a PDA product available? If so, does it cost extra? What operating system(s) does it work on?)
- Source of the content

Content information consists of the following:

- Links out (Does the product link out to other sources on the WWW, such as medical society pages?)
- Does the product interact with any sibling products?(Is the product accessed through a larger product or suite?)
- Does the product contain images?
- Does the product contain any audio-visual material?
- How explicitly evidence-based is the content? (What criteria or ratings are used?)
- Does it contain guidelines?
- Does it contain patient education materials?
- Does it contain calculators?
- Are there any specific specialties covered by this product?
- Comprehensiveness (How many diseases/conditions does the product cover? How many modules/articles are included?)

Features information consists of the following:

- Printing (How easy is it to print content?)
- Saving (How easy is it to save content?)
- E-mail (How easy is it to e-mail content?)
- Spell-check (Is there automatic spell-check on search terms? Does it work?)

- Does it highlight search terms in the results?
- Does it contain online help?
- Does the company provide direct user support?

Information for populating the direct examination database was gathered from the following sources, consulted in the following order: the company's product information, the company's website information, published reviews of the product, examination of the product itself, and finally, phone calls to the vendors, if needed.

### ***User Evaluation***

The second piece of this evaluation will be referred to as "user evaluation". In this part of the examination, potential users of the product were asked to evaluate each of the products. A total of twenty-four 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.

### **Trial Questions**

The test questions used for the evaluations come from a corpus of clinical questions recorded during ethnographic observations of practicing clinicians (Gorman, 1995; Gorman, 1995b). The subset of questions that were actually answered using a resource other than another person served as the starting point. 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, such as "When was

the patient's last pap smear?" which would more appropriately be answered with the patient record or the patient were dropped. Population statistic questions, such as "What is the incidence of Rabies in this community?" are more appropriately answered by the public health department and so were removed. Logistical questions, such as "Will her health insurance cover this drug?" are more appropriately answered using local documentation, and so were also dropped. Drug dosing and interaction information is more appropriately found using drug information databases, many of which are sibling databases of the ones currently under consideration. Questions concerning drug dosing and interaction were also removed from consideration. The remaining questions can be categorized at diagnosis, treatment and prognosis questions. See (Ely et al. 2000; Forsythe et al. 1992; Osheroff et al. 1991; Seol et al. 2004; Wildemuth et al. 1994) for further discussion of the classification and characterization of clinical questions. The resulting set of questions was tested by a group of pilot users. Any questions that were not answered with any of the databases were removed. Some questions were re-worded to be clearer. See Appendix A for a complete list of clinical questions used.

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 (Hersh et al. 2001a; Hersh et al. 2001b) These experiments compare two IR systems using task-oriented, user-centered methods. Each tester tests both systems with three questions, without the questions being repeated. The experimental design specifications state explicitly that the user should not see the same question twice and that the time burden on the user should be minimized. Three questions per database were determined to be sufficient for analysis.

This evaluation takes the same approach. 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 (Belkin et al. 1998). For each tester three clinical questions were systematically assigned to the databases such that no question is assigned to more than one database. The order in which databases are to be evaluated was randomly assigned. These random assignments were facilitated using the Random.org tool (Haahr, 2002). Each tester attempted to answer three questions with a specified bedside information tool, answered the user satisfaction questions, and then moved on to the next product and set of three questions, completing a total of fifteen questions (three for each database). This process was tested using pilot users. During this testing, 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.

No attempt was made to determine if questions were answered correctly. Users were instructed to use their own judgment to determine whether or not an answer to the question could be found within the time frame. They were not asked to record the answer to the questions, simply whether or not an answer was found, and how long it took them to find it.

### **Questionnaires**

The user evaluation questionnaire was developed from the literature. User evaluation questions were extracted from twelve studies (Abate et al. 1992; Bachman et al. 2003; Doll et al. 1988; Gadd et al. 1998; Jones 1999; Kupferberg et al. 2004; Kushniruk et al. 2001; Lewis 1995; Ma 2002; Pugh et al. 1994; Su 2003a; Su 2003b). 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-type scale. A final opportunity for comments was added at the end. See Appendix B for the user satisfaction questionnaire.

The background questionnaire was also guided by the literature, although less explicitly. Based on several examples of background questions (Cork et al. 1998; Hersh et al. 2000; Hersh et al. 2002; Kushniruk et al. 2001; Pugh & Tan 1994; Su 2003a; Su 2003b), 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. See Appendix C for the background questionnaire.

Both the background questionnaire and the user evaluation questionnaires were pilot tested for face validity and clarity. In order to facilitate the user evaluations, a website was developed so that the users were able to answer the questionnaires electronically if they wanted to. SurveyMonkey.com (2004e) was used to create and host the electronic questionnaires. In addition, the website contains direct links to the products being evaluated. Information about the project and contact information for the primary researcher were also provided. See Appendix D for a screenshot of the website. This website was hosted as part of the OHSU Library's website and password protected such that only participants and researchers had access to the site. This website and the electronic versions of the questionnaires were also pilot tested to ensure that they were consistent with the paper versions, previously tested.

### **User Recruitment**

After receiving approval from OHSU's Institutional Review Board, participants were recruited via a number of means. E-mails were sent to personal contacts, the communication officers of OHSU's School of Medicine and School of Nursing, the Veteran's Administration's Residents' group e-mail and the

Department of Medical Informatics and Clinical Epidemiology (DMICE) student's group e-mail. Announcements were posted on OHSU's study opportunity website (2005d) the DMICE website (2005a), the OHSU Library's website (2005b) and the School of Nursing's Opportunities and Activities Page (2005c). In addition, flyers were posted on campus and handbills were distributed at the OHSU Library's reference desk. The same IRB approved recruitment text was used for all of these recruitment strategies. See Appendix E for the recruitment text.

The pool of participants that these methods hoped to recruit was any potential user of bedside information tools. Classes of individuals targeted included physicians, residents, medical students, physician assistants, physician assistant students, and nurses, nursing students, pharmacists and informatics students. Our goal was to include at least three participants from any class of participants (Nielsen et al.; Virizi 1992).

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 participant was consented. As part of the meeting, participants received an evaluation packet. This packet consisted of: the consent form, the background questionnaire, the data collection sheets for the systems evaluations, a copy of the project's website and the username and password needed to access it, the primary researcher's card, and a recruitment flyer to pass along to any interested colleagues. See Appendix F for the consent form, Appendix C for the background questionnaire, Appendix G for a sample data collection form, Appendix D for a screenshot of the website and Appendix E for the recruitment text. Participants were asked to complete the background questionnaire and product evaluation (either on paper or electronically) and to contact the primary researcher when they were done in order to schedule the



follow-up interview. Participants were allowed to complete the evaluations at their convenience, on their own computers.

### *Semi-structured interviews*

In addition the quantitative data collected through direct examination and user evaluations, qualitative data were collected to provide context for the quantitative data. This mixed method approach is an effective way to gather user-centered data for evaluations of IR systems (Bury et al. 2004; Jennett et al. 1991; Kaplan et al. 1988; Lam et al. 2004; Pugh & Tan 1994; Zeng et al. 2004). All participants who completed the evaluations were scheduled for a follow-up interview, either in person or via telephone. These interviews were recorded, and selectively transcribed for analysis. In addition, detailed notes were taken during the interview. First users were asked to discuss what information sources they typically use to answer clinical information needs outside of this evaluation. Second, using a technique known as the critical incident technique (Pluye et al. 2004) users were asked to recall a positive or negative experiences with a bedside information tool either during the evaluation or their clinical practice. It has been noted that clinicians often have difficulty transforming an information need into a database query (Osheroff et al. 1993). In order to explore this further, interviewees were asked about this process, and in particular if the process of formulating a query was different for any of the products being tested. Respondents were asked to imagine an ideal bedside information tool and describe its features. Users were also asked about the value that these products would have on a PDA. Finally, users were asked to rank all five products from the one they liked the best to the one they liked the least. See appendix D for the interview guide that was used for these interviews.

## **Results**

### *Direct Examination*

#### **Product Information**

Despite its name, UpToDate is the least frequently updated product. Although this is a measure that most librarians would consider of paramount importance, it is also a difficult measure to make sense of. Although one would prefer a product that is updated frequently such as FIRSTConsult, it is difficult to ascertain how thorough an “update” is. A less frequent but more thorough update may be more meaningful than more frequent but minimal updates. Although prices are listed for individual subscriptions, pricing for institutions and consortiums are typically negotiated and vary widely. However, a general feel for the relative costs of these products can be gleaned from the “list price” for individual subscriptions. Nearly all of these products offer some version for a PDA. Although all of these products are authored or edited by some trusted source, only UpToDate and ACPs PIER are explicitly peer reviewed. See Table 1 for the results of the product information portion of the direct examination evaluation.

#### **Content**

All of the products offer some links out to other materials on the Internet. The most common links are links to PubMed citations. Other common links are to medical societies, guidelines and patient education materials. Some of these products have sibling products. Sibling products are products which are produced by the same publisher and often have similar user interfaces. These sibling products have distinct content and require a separate subscription. The existence of sibling products may have complicated the user evaluations as mentioned in the discussion section. Features or content (such as medical calculators) which are not present in the databases being examined may be

available in a sibling product. All of the products contain images but only FIRSTConsult contains audio-visual materials. While all of these products can be considered evidence based in a broad sense (Alper 2003), only ACP's PIER, FIRSTConsult and InfoRetriever offer explicit and systematic evidence based rating criteria and can be considered evidence based in a strict sense. Although some products may refer to themselves as evidence based, without systematic analysis of the evidence, a summary document cannot be considered evidence based. DISEASEDEX and UpToDate merely cite references and while perhaps "authoritative" they cannot be called evidence based. All of the products, with the exception of DISEASEDEX, offer treatment guidelines, or links to them. Again, all of the products, with the exception of DISEASEDEX, offer patient education materials. FIRSTConsult's patient education materials are notable, as they are customizable and available in Spanish. Only ACP's PIER and InfoRetriever offer medical calculators. UpToDate offers materials for the most specialties, but generally speaking, all of these products cover only the more primary care oriented clinical specialties and do not contain material for sub-specialties. None of these products attempt to cover all of medicine, each contain a varying number of modules or diseases. We were unable to determine the numbers of modules or diseases covered by UpToDate, even after speaking with the company on the phone. Rare or obscure conditions or diseases may not be covered by any of these information sources. See Tables 2 and 3 for the results of the content section of the direct examination evaluation.

## **Features**

Printing out the desired information from an electronic resource is often a frustrating experience. Only InfoRetriever and UpToDate offer "printer friendly" links to make printing selected material easier. None of these products allow a user to download or save any portion of the product. While one can understand the publishers of these products' concerns about users downloading and

distributing the entire database, users are prevented from legitimate downloads of a single article for later review or printing. Other electronic full-text databases such as Journals@OVID do allow for the downloading or saving of articles. Only UpToDate offers the option to e-mail the contents of an article. This option can be extremely useful if one searches the product from a shared (bedside) computer but would like to have access to the article from one's own computer. The ability to e-mail articles also makes it easy to share information with others, in fact UpToDate's e-mail feature is called "e-mail this to a colleague." All of the products offer some sort of spellchecker. InfoRetriever does not have a spell check per se, but its keyword suggestion algorithm clearly takes misspellings into account as the correct keywords are suggested for misspelled entries. The other products' spell-check features use different processes, which work to varying degrees. ACP's PIER's offers the least assistance, as it simply provides an alphabetical list of suggested terms. Depending on the nature of the original misspelling this is often not useful. Highlighting search terms can be a useful feature that helps a user locate where in the text his search term is located. Only ACP's PIER and FIRSTConsult offer this feature. The lack of this feature in the other products is most likely due to the fact that keyword searches are converted to controlled vocabulary during the search process. Although all of the products offer a help page, none of them offer any context sensitive help. Only FIRSTConsult, InfoRetriever and UpToDate have clearly identified user support websites and toll free numbers. See Table 4 for the results of the features section of the direct examination evaluation.

### **Summary**

Although no particular product stands out in this evaluation, some parts of this evaluation should be highlighted. In order to retrieve the best evidence a product needs to contain evidence based content. On the basis of this comparison alone, one would hesitate to recommend either DISEASEDEX or UpToDate. Further,

the infrequency of updates and high cost make UpToDate a poor candidate. However, UpToDate does rate quite well on features that increase usability, such as the ability to e-mail and print with ease.

## Table 1: Product Information

Product	Updates	Pricing	PDA Available? OS?	Source Material
<b>ACP's PIER</b> <a href="http://pier.acponline.org/info">http://pier.acponline.org/info</a>	"8 to 10 modules added each month" "All PIER modules are updated continually by editorial consultants."	variable; free to ACP members	Yes Palm OS & Pocket PC	American College of Physicians; All modules are peer reviewed.
<b>DISEDEX</b> <a href="http://www.micromedex.com">http://www.micromedex.com</a>	quarterly	variable	No n/a	"Edited by the MICROMEDEX staff and leading emergency physicians at Harvard, Stanford, and Oxford Universities."
<b>FIRSTConsult</b> <a href="http://www.FIRSTConsult.com">www.FIRSTConsult.com</a>	continuous rolling review (80 conditions/month), new findings, news and urgent updates as needed	individual subscriptions \$89-\$149	Yes Palm OS & Pocket PC	"Physician writers guided by the editorial advisory board and clinical reviewers synthesize materials from respected journals, & professional organizations."
<b>InfoRetriever</b> <a href="http://www.infopoems.com">www.infopoems.com</a>	quarterly \$129-\$249	individual subscriptions	Yes Palm OS & Pocket PC	Cochrane Database, 5-minute Clinical Consult, >100 Journals are reviewed.
<b>UpToDate</b> <a href="http://www.uptodate.com">www.uptodate.com</a>	every four months	individual subscriptions \$195-\$495	Yes Pocket PC	"All UpToDate topic reviews are written by leading physician educators who are experts in their sub-specialty." "Extensive peer review process."

## Table 2: Content

Name	Links	Sibling product	Images	AV	EBM rating	Guidelines
ACP's PIER	Yes PubMed and associations	Yes STAT!Ref	Yes	No	Yes rating and criteria stated in Appendix, links-out recommendations A-C and levels of evidence 1-3	Yes
DISEDEX	Yes PubMed	Yes MIRCROMEDEX™ suite	Yes	No	No no explicit ratings given	No
FIRSTConsult	Yes associations and guidelines	Yes MDConsult	Yes	Yes	Yes graded in accordance with American Academy of Family Physicians guidelines	Yes links-out
InfoRetriever	Yes PubMed	Yes InfoPOEMS®	Yes	No	Yes clearly listed, graded recommendations A-D and levels of evidence 1-4	Yes
UpToDate	Yes PubMed guidelines	Yes Lexi-Interact™ drug interactions program	Yes	No	No no explicit ratings given	Yes

**Table 3: Content Cont.**

Name	Patient Education	Calculators	Specialties	Comprehensiveness
ACP's PIER	Yes links-out	Yes links-out	Internal Medicine	over 280 disease modules
DISESEDEX	No	No	General Medicine; Emergency Medicine	over 570 disease modules
FIRSTConsult available in	Yes customizable, Spanish	No	Primary care	over 450 medical conditions
InfoRetriever	Yes links-out (familydoctor.org)	Yes	Family Medicine	775 practice guideline summaries
UpToDate	Yes	No	Internal Medicine; Cardiovascular; Endocrinology; Family Medicine; Gastroenterology; Hematology; Infectious Disease; Nephrology; OB-GYN; Oncology; Pulmonary; Rheumatology	unavailable



## Table 4: Features

Name	Printing	Save	eMail	Spell-check	Highlighted Search Terms	Online Help	User Support
ACP's PIER	No only browser functionality	No only browser functionality	No only browser functionality	Yes provides alphabetical listing	Yes	Yes static help page	No
DISESEDEX	No only browser functionality	No only browser functionality	No only browser functionality	Yes stem matching	No	Yes static help page	No
FIRSTConsult	No only browser functionality	No only browser functionality	No only browser functionality	Yes angina suggests: anguine & aniconia not angina	Yes	Yes static help page	Yes website and 800 number
InfoRetriever	Yes printer friendly link	No only browser functionality	No only browser functionality	No keyword suggestion doesn't perform spell-check	No	Yes static help page	Yes website and 800 number
UpToDate	Yes printer friendly link	No only browser functionality	Yes e-mail this to a colleague link	Yes stem matching conditoin suggests matches for condit	No	Yes static help page	Yes website and 800 number

## *User Evaluation*

### **User Demographics**

The users participating in the study range in age from 28 to 49 with the mean age being 35. The group of users is roughly split between men and women with 42% men and 58% women. The clinical occupations represented include physicians (8), pharmacists (3), medical informatics students with previous clinical experience (5), a MRI technologist and a registered nurse. Participants have been at their current profession for 1 to 20 years with a mean of 8 years. All participants are experienced computer users. 94% use a computer more than once a day, with the most common location for computer use being at work. All participants own a computer, 94% own a Windows-based PC. 70% of respondents own a Palm based PDA, no respondents own a Pocket PC based PDA. Of those respondents who own PDA's, 60% use them more than once a day. All respondents indicate that they have experience with Medline on OVID and 95% report having experience with PubMed. 82% of respondents report being familiar with the OHSU library catalog and 70% are familiar with UpToDate. 65% of respondents report having used MDConsult and MICROMEDEX in the past. Half of the testers were familiar with the Cochrane Library and 40% were familiar with STAT!Ref. InfoRetriever was only familiar to two respondents. Only one respondent indicated being familiar with either FirstConsult, ACP's PIER or DISEASEDEX.

### **Recruitment Methods**

A total of 33 people responded to the request for participants. Some of these people could not participate because they were not clinicians or were not OHSU faculty, staff or students. Some were excluded simply because they contacted the researcher after the 24 study slots were filled. The most successful recruitment

method was personal contacts with nine respondents, followed by the Medical Informatics Student listserv with seven contacts. The residents' group e-mail, the OHSU study opportunity newsletter got five responses and the flyers yielded four participants. The posting on the library website yielded three respondents and a single participant was recruited through the OHSU Library's reference desk.

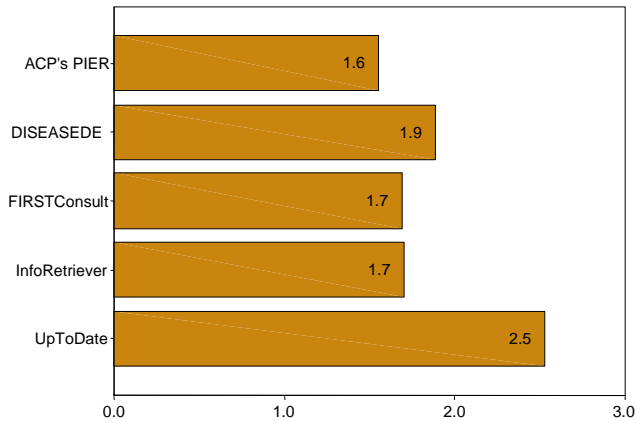
Of the original 24 consented participants, 18 completed the study within the time frame of this study. Participants who failed to complete the study did not differ from those that completed the evaluation in terms of recruitment method, or clinician type (as ascertained from the initial meeting with these participants). Drop-out rates may have been high because participants were asked to perform the evaluations on their own time. A laboratory like evaluation may have reduced drop-out rates, but may have also reduced the number of volunteers in total.

## **Task Measures**

### *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).

Figure 1: Average Number of Answers Found



Friedman Test: N=16 p<.0001

In order to determine which databases may differ in the average number of answers found, a Wilcoxon Signed Ranks Test was applied to each pair-wise combination. Only pairs comparing UpToDate with other databases were found to be significant at a level of  $p < .005$  (denoted with \*\*). Comparisons of DISEASEDEX with other products were significant at a level of  $p < 0.05$  (denoted with \*) (See Table 5).

Table 5: Average Number of Questions Answered, Pair-wise comparisons.

Wilcoxon Signed Ranks Test					
Databases Compared	p-value	N	Databases Compared	p-value	N
ACP's PIER & DISEASEDEX	.014*	18	DISEASEDEX & InfoRetriever	.025*	17
ACP's PIER & FIRSTConsult	.564	16	DISEASEDEX & UpToDate	.003**	17
ACP's PIER & InfoRetriever	.317	17	FIRSTConsult & InfoRetriever	.317	16
ACP's PIER & UpToDate	.001**	17	FIRSTConsult & UpToDate	.000**	16
DISEASEDEX & FIRSTConsult	.014*	16	InfoRetriever & UpToDate	.001**	17

### *Time Spent Searching*

In this analysis Kaplan-Meier analysis was used to calculate mean and median survival time. This method accounts for the fact that three minutes could be the recorded time because the question was answered in three minutes or because the attempt to answer the question failed. As these are self recorded times, rounded to the nearest minute (values could be 0,1,2,3) they are only very rough estimates of the amount of time participants spent searching any product. The mean survival rate for all products was two minutes and the median survival time for all products was three minutes. They do differ in the 95% confidence intervals for these estimates (See Table 6).

Table 6: Time Spent Searching, Survival Times

Kaplan-Meier Survival Analysis				
Database	Mean Survival Time (minutes)	95% CI Upper	Median Survival Time (minutes)	95% CI Upper
		95% CI Lower		95% CI Lower
ACP's PIER	2	2	3	.
		3		.
DISEASEDEX	2	2	3	3
		2		4
FIRSTConsult	2	2	3	2
		3		4
InfoRetriever	2	2	3	2
		2		4
UpToDate	2	1	3	2
		2		2

### **User Satisfaction Measures**

In all of the user evaluation questions using a Likert-type scale, the questions were phrased such that 1 was correlated with a desirable trait such as speed and 5 was correlated with an undesirable trait such as slowness (See Appendix B). For the presentation of the results in graphical form, it is more intuitive for the desirable traits to be associated with higher numbers. The original results were subtracted from six in order to transform the data before presentation. The

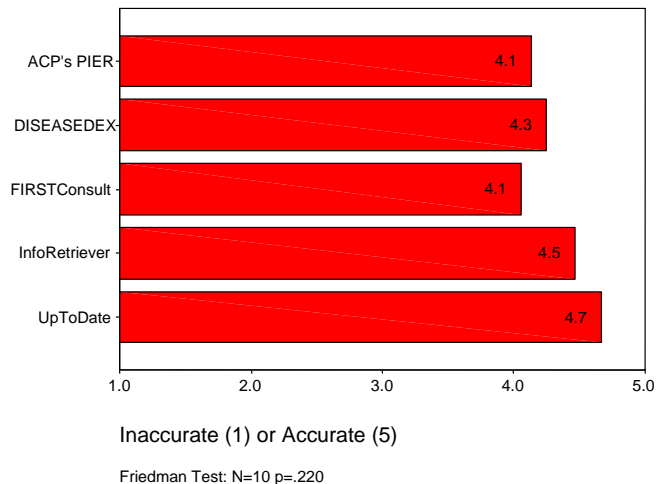
sample size for these analyses is considered to be the total number of respondents.

## User Perceptions of Content

*The content of the database seemed: accurate/inaccurate*

All products rated at least a 4 on this scale, and none rated a perfect 5. UpToDate had a mean rating of 4.62; InfoRetriever received a mean rating of 4.54; DISEASEDEX rated a 4.14; ACP's PIER had a mean rating of 4.07 and FIRSTConsult had the lowest rating of 4.06. The Friedman Test was used to test the hypothesis that all databases received the same ratings. This hypothesis was not rejected with a  $p=.220$  ( $N=10$ ) (See figure 2).

Figure 2: Did the content of the database seem:



In order to determine which pairs of databases may differ with respect to users' ratings of accuracy, a Wilcoxon Signed Ranks Test was applied to each pair-wise combination. No pairs were found to be significant at a level of  $p<.005$ . Only the comparison of UpToDate with ACP's PIER (the highest ranked, compared with one of the lowest ranked) was found to be significant at a level of  $p<0.05$  (denoted with \*) (See Table 7).

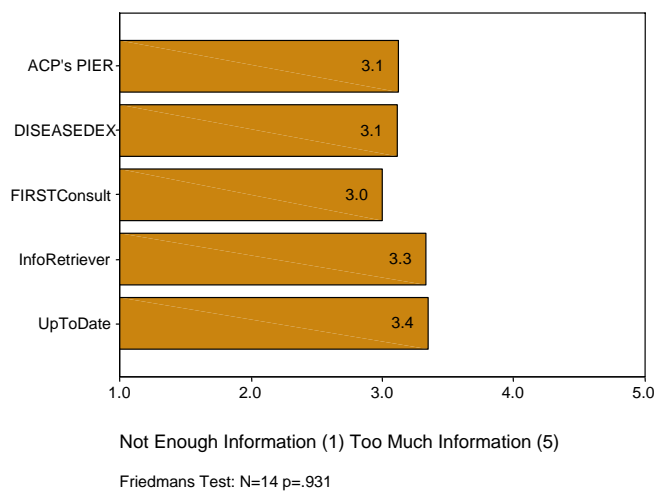
Table 7: Accurate/Inaccurate, Pair-wise comparisons.

Wilcoxon Signed Ranks Test					
Databases Compared	p-value	N	Databases Compared	p-value	N
ACP's PIER & DISEASEDEX	1.000	14	DISEASEDEX & InfoRetriever	.527	14
ACP's PIER & FIRSTConsult	.480	14	DISEASEDEX & UpToDate	.102	14
ACP's PIER & InfoRetriever	.180	13	FIRSTConsult & InfoRetriever	.096	14
ACP's PIER & UpToDate	.046*	13	FIRSTConsult & UpToDate	.084	14
DISEASEDEX & FIRSTConsult	.862	15	InfoRetriever & UpToDate	.083	14

*Did the database give you: too much/not enough information?*

All of the products rated near 3, which would be just the right amount of information. FIRSTConsult had a mean rating of 3.0. Both ACP's PIER and DISEASEDEX had a mean rating of 3.1. InfoRetriever had a mean rating of 3.3 and UpToDate a mean rating of 3.34. The Friedman Test was used to test the hypothesis that all databases received the same ratings. This hypothesis was not rejected with a  $p=.931$  ( $N=14$ ) (See figure 3).

Figure 3: Did the Database Give You



A Wilcoxon Signed Ranks Test was applied to each pair-wise combination. No pairs were found to be significant a level of  $p<.05$  (See Table 8).

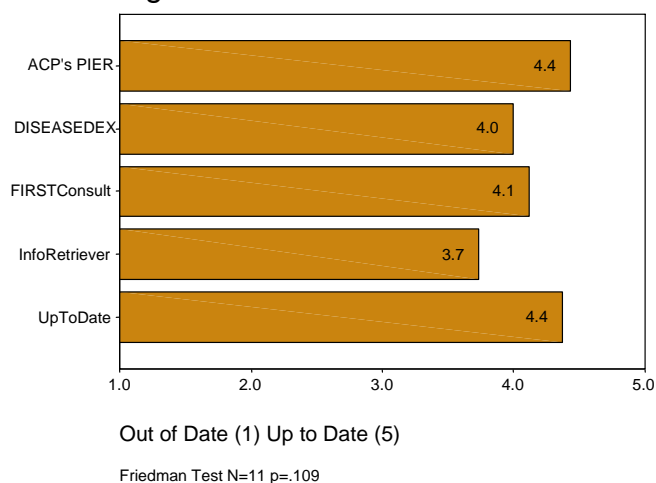
Table 8: Too Much/Not Enough Information, Pair-wise comparisons.

Wilcoxon Signed Ranks Test					
Databases Compared	p-value	N	Databases Compared	p-value	N
ACP's PIER & DISEASEDEX	.581	15	DISEASEDEX & InfoRetriever	.546	17
ACP's PIER & FIRSTConsult	.623	15	DISEASEDEX & UpToDate	.539	17
ACP's PIER & InfoRetriever	.717	16	FIRSTConsult & InfoRetriever	.547	17
ACP's PIER & UpToDate	.372	15	FIRSTConsult & UpToDate	.334	16
DISEASEDEX & FIRSTConsult	.876	16	InfoRetriever & UpToDate	.948	17

*Did the information seem: up to date/out of date?*

All of these products rated well on this scale with mean rankings near 4. ACP's PIER and UpToDate were rated the best with a mean score of 4.4. FIRSTConsult rated a 4.1. DISEASEDEX had a mean of 4.0 and InfoRetriever had a rating of 3.7. The Friedman Test was used to test the hypothesis that all databases received the same ratings. This hypothesis was not rejected with a  $p=.109$  ( $N=11$ ) (See figure 4).

Figure 4: Did the Information Seem:





A Wilcoxon Signed Ranks Test was applied to each pair-wise combination. No pairs were found to be significant a level of  $p < .05$  (See Table 9).

Table 9: Up to Date/Out of Date, Pair-wise comparisons.

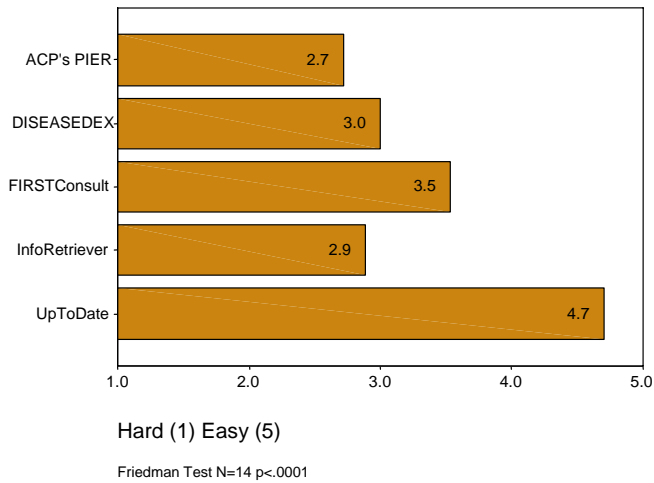
Wilcoxon Signed Ranks Test					
Databases Compared	p-value	N	Databases Compared	p-value	N
ACP's PIER & DISEASEDEX	.317	14	DISEASEDEX & InfoRetriever	.271	14
ACP's PIER & FIRSTConsult	.705	13	DISEASEDEX & UpToDate	.262	15
ACP's PIER & InfoRetriever	.083	13	FIRSTConsult & InfoRetriever	.142	14
ACP's PIER & UpToDate	1.000	13	FIRSTConsult & UpToDate	.623	15
DISEASEDEX & FIRSTConsult	.260	15	InfoRetriever & UpToDate	.061	14

### User Interaction Measures

*Interacting with the system was: easy/hard*

UpToDate rated much higher than any other product on this scale, receiving a near perfect score of 4.7. FIRSTConsult rated a mean of 3.5; InfoRetriever rated a mean of 2.9. DISEASEDEX rated a mean of 3.0 and ACP's PIER rated a mean of 2.7. The Friedman Test was used to test the hypothesis that all databases received the same ratings. This hypothesis was rejected with a  $p < .0001$  ( $N=15$ ) (See figure 5).

Figure 5: Interacting with the System Was:



In order to determine which databases may differ with respect to users' ratings, a Wilcoxon Signed Ranks Test was applied to each pair-wise combination. Only the comparison of UpToDate with other products was found to be significant at a level of  $p < 0.005$  (denoted with \*) (See Table 10).

Table 10: Easy/Hard, Pair-wise comparisons.

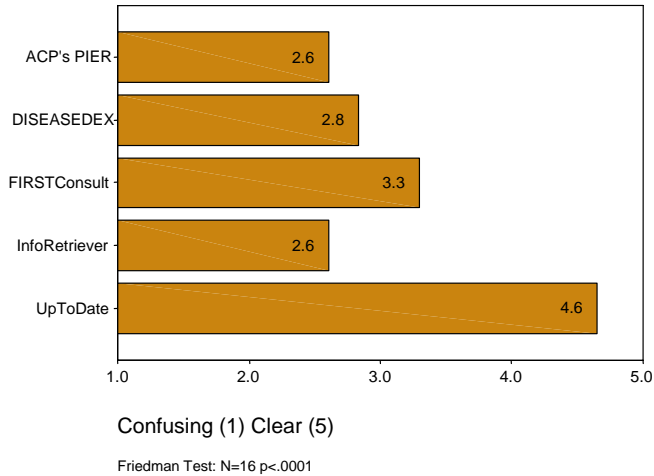
Wilcoxon Signed Ranks Test					
Databases Compared	p-value	N	Databases Compared	p-value	N
ACP's PIER & DISEASEDEX	.239	17	DISEASEDEX & InfoRetriever	.583	17
ACP's PIER & FIRSTConsult	.056	17	DISEASEDEX & UpToDate	.002*	16
ACP's PIER & InfoRetriever	.635	18	FIRSTConsult & InfoRetriever	.200	17
ACP's PIER & UpToDate	.002*	17	FIRSTConsult & UpToDate	.004*	16
DISEASEDEX & FIRSTConsult	.166	16	InfoRetriever & UpToDate	.001*	17

*The layout of screens was: clear/confusing*

UpToDate rated much higher than any other product on this scale, receiving a mean score of 4.6. FIRSTConsult rated a mean of 3.3, DISEASEDEX rated a mean of 2.8. ACP's PIER and InfoRetriever rated a mean of 2.6. The Friedman Test

rejected the null hypothesis that all databases received the same rating, with a  $p < .0001$  (N=16) (See figure 6).

Figure 6: The Layout of the Screens Was:



A Wilcoxon Signed Ranks Test was applied to each pair-wise combination. Only the comparison of UpToDate with other products was found to be significant at a level of  $p < 0.005$  (denoted with \*) (See Table 11).

Table 11: Clear/Confusing, Pair-wise comparisons.

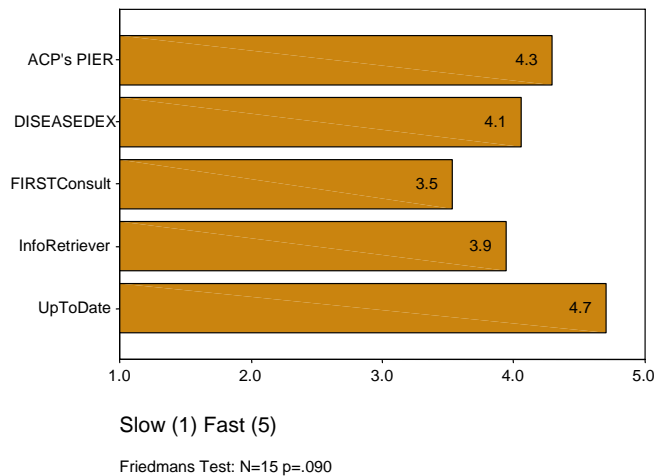
Wilcoxon Signed Ranks Test					
Databases Compared	p-value	N	Databases Compared	p-value	N
ACP's PIER & DISEASEDEX	.465	18	DISEASEDEX & InfoRetriever	.441	18
ACP's PIER & FIRSTConsult	.140	17	DISEASEDEX & UpToDate	.002*	17
ACP's PIER & InfoRetriever	.975	18	FIRSTConsult & InfoRetriever	.167	17
ACP's PIER & UpToDate	.001*	17	FIRSTConsult & UpToDate	.004*	16
DISEASEDEX & FIRSTConsult	.342	17	InfoRetriever & UpToDate	.001*	17

*The database was: fast/slow*

UpToDate rated higher than any other product on this scale, receiving a score of 4.7. ACP's PIER rated a mean of 4.3. DISEASEDEX rated a mean of 4.1.

InfoRetriever rated a mean of 3.9 and FIRSTConsult rated a mean of 3.5. The Friedman Test was used to test the hypothesis that all databases received the same ratings. This hypothesis was not rejected with a  $p=.090$  ( $N=15$ ) (See figure 7).

Figure 7: The Database Was:



A Wilcoxon Signed Ranks Test was applied to each pair-wise combination. Only the comparison of UpToDate with other products was found to be significant at a level of  $p<0.05$  (denoted with \*) (See Table 12).

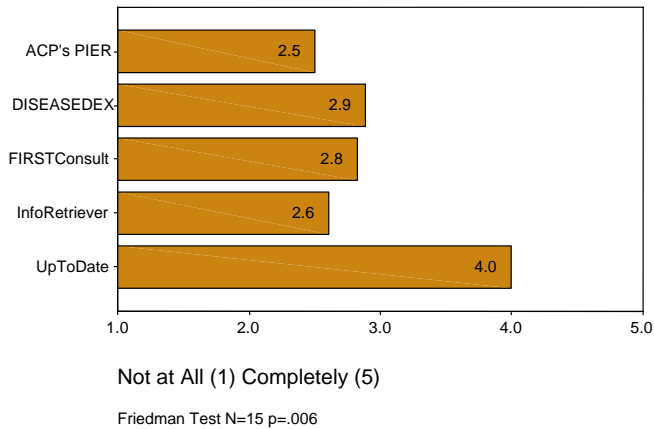
Table 12: Fast/Slow, Pair-wise comparisons.

Wilcoxon Signed Ranks Test					
Databases Compared	p-value	N	Databases Compared	p-value	N
ACP's PIER & DISEASEDEX	.350	17	DISEASEDEX & InfoRetriever	.700	18
ACP's PIER & FIRSTConsult	.080	16	DISEASEDEX & UpToDate	.046*	17
ACP's PIER & InfoRetriever	.130	17	FIRSTConsult & InfoRetriever	.446	17
ACP's PIER & UpToDate	.142	16	FIRSTConsult & UpToDate	.009*	16
DISEASEDEX & FIRSTConsult	.395	17	InfoRetriever & UpToDate	.018*	17

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 8).

Figure 8: Overall, did this Database Satisfy Your Needs:



A Wilcoxon Signed Ranks Test was applied to each pair-wise combination. Only the comparison of UpToDate with other products was found to be significant at a level of  $p < 0.02$  (denoted with \*) (See Table 13).

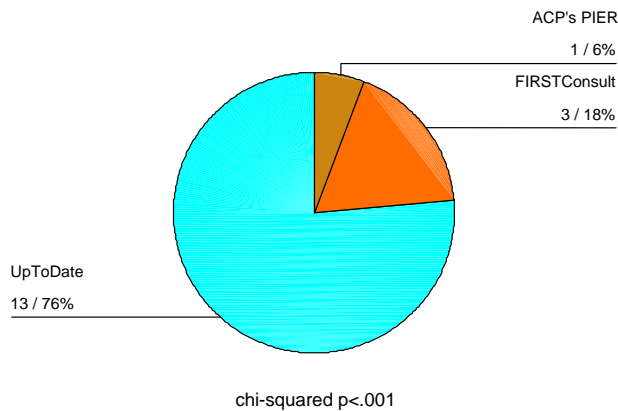
Table 13: Overall, Pair-wise comparisons.

Wilcoxon Signed Ranks Test					
Databases Compared	p-value	N	Databases Compared	p-value	N
ACP's PIER & DISEASEDEX	.327	18	DISEASEDEX & InfoRetriever	.369	18
ACP's PIER & FIRSTConsult	.638	16	DISEASEDEX & UpToDate	.010*	17
ACP's PIER & InfoRetriever	.785	18	FIRSTConsult & InfoRetriever	.936	16
ACP's PIER & UpToDate	.003*	17	FIRSTConsult & UpToDate	.007*	15
DISEASEDEX & FIRSTConsult	.904	16	InfoRetriever & UpToDate	.003*	17

*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 9).

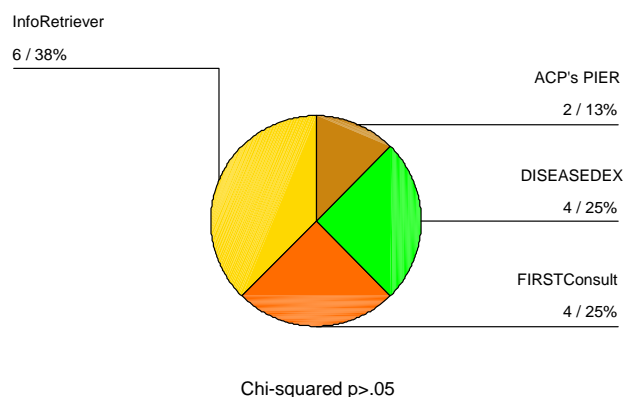
Figure 9: Percent of Respondents who Ranked this Product the Best



*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 10).

Figure 10: Percent of Respondent who Ranked this Product the Worst



### *Qualitative data*

Comments from the evaluation questionnaire were combined with notes from the semi structured interviews. Common themes and comments about each database are reported here. As many of the comments relate to the search interface of these products Appendix I contains screen-shots of all five bedside information tools being evaluated here.

### **ACP's PIER**

Participants usually ranked this product somewhere near the middle. Many users had difficulties returning to the search results after viewing an article from the results. The browser's back button returns the user to the search screen, rather than the results. Users also expressed frustration with the alphabetical listing of diseases, as they found it a difficult way to browse. One tester reported that it was "hard to search through more categories to find information." Others noted that some searches were difficult because the product seemed disease based rather than symptom based. On the other hand, some users were quite pleased with the evidence-based content of the product and the clear labeling of rationale and evidence.

## **DISEASEDEX**

Many users were familiar with this product's sibling database MICROMEDEX. The search interface for MICROMEDEX and DISEASEDEX are the same, however a MICROMEDEX search returns information in an entirely different format than the results of a DISEASEDEX search. Most users were quite happy searching MICROMEDEX for drug information. However, few users extended this feeling to DISEASEDEX. The most common complaint about this database is that the information is presented in long blocks of text, noting that there was "too much text to read, in long paragraphs". Most users found it difficult to find the exact information they were searching for within any article. Others noted that the product is not structured to support browsing, as one must enter a search term in order to "get in" to the database in the first place.

## **FIRSTConsult**

Although few of the testers were familiar with this product, it rated quite high with users. Only UpToDate was consistently ranked higher, and most testers were quite familiar with UpToDate. Users found that the structure of FIRSTConsult supported browsing and offered information in appropriately sized chunks. One tester noted that FIRSTConsult had "great information, divided up well, to allow quick access to just how I need this information presented." Users also found the differential diagnosis tool as a useful way to find information. However, several users experienced server errors or very slow response times. While this could be the result of many factors, it should be noted that this is the only tool for which such problems were reported.

## **InfoRetriever**

This product seemed to have the steepest learning curve. Many users mentioned that the product might have been able to answer the question, if they had known how to use the product better. Comments such as "I'm not sure I figured out



how to use this” were common. A contributing factor may have been the fact that this was the only product that required the user to specify the type of search before beginning the search. One user noted, “The keyword matching step took time but did not seem to add much relevance. “ These findings are consistent with (Fitzpatrick 2004) review of InfoRetriever which noted that the initial search screen was not intuitive and with (Lam et al. 2004) in which users of InfoRetriever complained that it took too much time to learn how to use the product.

### **UpToDate**

Users love UpToDate. They find the simple google-like search interface appealing, they like the presentation of information with a navigation frame to the left, and they like the linked references. Users noted that the articles are structured in a consistent and clear manner and that this is similar to the way they themselves think through a disease or problem. Although many of the testers were already familiar with UpToDate, people who had never used UpToDate also sang its praises.

### **General Comments**

In general testers indicated that both searching and browsing were important ways to access information. They wanted a simple and clear search interface, but they also wanted to be able to move both within a topic and between topics in a manner that makes sense to them. Being able to browse by disease or symptom and having topics broken into clearly labeled and consistent sub-sections were mentioned by most testers as desired features of an ideal bedside information tool. Testers disliked having to wade through long chunks of text, but also worried that bullet point answers had been oversimplified. Testers expressed a desire to have information available that would expand or collapse in an interactive manner. They wanted something that would allow them to access a

lot of information on some occasions, and more summarized information in other situations. Users did not demand systematic evidence-based ratings but did insist that information be referenced. Virtually none of the participants expressed any concerns about the timeliness of the information or what peer review process this information may have been vetted through. Users wanted the information to be presented in a way that made it easy to access the content, but seemed to trust that the content of the database was correct. Only one respondent expressed concern about the content of an answer noting that it was “poorly written.”

## Discussion

### *Content Measures and Interface Measures*

When looking at the user interaction measures and user ranking of products, UpToDate is consistently rated high. These findings are further supported by the qualitative results from the follow-up interviews. Users found interacting with UpToDate to be easy, the layout of the screen to be clear, the system to be fast, and overall it met their needs. Users also found a significantly higher proportion of answers in UpToDate compared to other products.

However, if you look at the user evaluation questions that focus on content (did the content seem accurate or inaccurate, did the database give too much or not enough information, and did the information seem up to date or out of date) the differences between products are not significant. Similarly if you look at the mean time spent searching UpToDate's scores were not found to be significantly different from the other products.

In terms of interacting with the product, users clearly prefer UpToDate. When one examines the results of the direct examination UpToDate is not obviously superior. Of all the products examined, it is updated the least frequently, it doesn't offer any medical calculators and while it does provide citations, it doesn't contain evidence based content. However, UpToDate does rate high in the features category, it is one of two products providing printer-friendly links and the only product offering e-mail. Once again, when examining the user-interaction aspects of these products, UpToDate scores quite high.

These findings are consistent with a number of studies that tested what bedside information tools users selected when they have access to more than one tool.

These use studies, while not usually exploring the reasons why people use the products they use, do indicate that when given a choice users pick UpToDate over print resources or electronic textbooks very frequently (Peterson et al. 2004; Schilling et al. 2005).

### ***The Interface is the Product***

That these data show significance only in terms of user interaction scores, and a user's ability to find answers indicates that these types of products must be evaluated in terms of user interaction as well as their content. A product selected because of excellent content alone 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.

With more traditional library materials the differences between products were primarily content differences. The usability of a book is a relative constant. With the first electronic resources such as searchable bibliographic databases and e-journals, the interface was similarly not much of an issue because the benefits of the electronic version far outweighed those of print and the differences between the electronic versions were insignificant in comparison.

As electronic formats become the expected or only 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 this type of bedside information tool, the interface is the product as much if not more than the content itself. This only underscores that the traditional content focused evaluation methods familiar to librarians may no longer be sufficient in the evaluation of this new breed of information tools. For bedside information tools

features such as accessibility and usability become the driving factors for performance.

### ***“Librarian Knows Best” Model Not Best***

It is somewhat disconcerting that the high rates with user satisfaction are not aligned with those factors of the direct examination that librarians traditionally focus on such as frequency of updates or the explicit leveling of evidence based content. It has been argued that unlike librarians, users make choices based on the convenience of the product rather than the content of the product (Koonce et al. 2004).

As far as the collection development dilemma goes, this is nothing new. There has long been a tension between giving the users what they want and giving the users what librarians believe they should have (Pearl 1996). On the one hand, the “librarian knows best model” alienates users (Dickenstein et al. 2000; Travis et al. 2002), and can lead to the selection of tools and products that users will never use because the user interface is poor. However, having professionals evaluate the quality of information resources and make recommendations based on factors such as how often a product is updated and who is providing the content is the hallmark of a curated collection of information resources. A high level of validity must be assured in order for the information retrieved to be useful. It has been shown that when evaluating resources librarians look at different factors than users (Kupferberg & Jones 2004). The user’s perceptions of a product must be included in the evaluation. There needs to be some sort of balance between easy interfaces with questionable information (the internet) and information that is of high quality but difficult to access (high quality print journals in an inconvenient location).

### ***User Testing Essential and Feasible***

This evaluation shows that user preferences play an important role in deciding what products are worth purchasing. The good news is that eliciting user feedback may not be as difficult as librarians often imagine. Recruiting participants into this study was not difficult, in fact more people volunteered than were even used in the study. This may be unique to this institution. However, at any academic medical center there may well be a similar set of library users who would volunteer for a study because they are interested in the topic, or because they are researchers themselves and simply want to help with someone else's research. Perhaps recruitment efforts were improved because the project was co-sponsored by both the Library and the Department of Medical Informatics. Other academic health sciences libraries may consider partnering with medical informatics professionals if there are any on campus. Libraries often make trial versions available to their users, and may have difficulty eliciting feedback. A more structured evaluation method is probably needed. Targeted recruitment, specific evaluation tasks and response forms may help to engage users for testers. Although financial incentives were not used as a recruitment method in this study, a small stipend may well be worth the financial outlay (Edwards et al. 2002) in order to have reliable data concerning user preferences before considering the purchase of any electronic product.

### ***Limitations***

There are a number of limitations worthy of discussion. Although efforts were made to recruit all potential library 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. Alternately, a more homogeneous group of users, such as only physicians, would have produced more reliable 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 (Haynes et al. 1994; Hersh et al. 2000; Kupferberg & Jones 2004; Schwartz et al. 2003), no training in these products 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. It can 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 the product to other options. One must consider both the costs to the user in training, the costs to the trainer and the inevitability of the product changing and training having to be repeated or revised.

With this sample it is difficult to determine exactly what effect a user's previous experience with a product may have had. Because so many users are familiar with UpToDate, it is difficult to draw conclusions from the group of users not familiar with UpToDate simply because it is so small. Only five participants were not already familiar with UpToDate. If familiarity were the only factor accounting for its high scores, one would expect familiarity to be correlated with

positive ratings for all products. This is not the case. If we look at the two products which few of the participants were familiar with and the OHSU Library does not currently subscribe to (InfoRetriever and FIRSTConsult) we do not see universally poor ratings. In fact, FIRSTConsult is consistently ranked high (it was ranked best by three participants and it ranks closely behind UpToDate on a number of scales) while InfoRetriever is one of the most poorly rated products (50% of respondents ranked this product the worst). Some amount of the favorable ratings for UpToDate may be due to the participant's familiarity with the product, but familiarity alone is not sufficient to produce the results shown in this study.

Some product specific concerns include the fact that ACP's PIER was evaluated using the STAT!Ref interface. ACP's PIER is also available through other vendors and directly from the ACP with different interfaces. ACP's PIER could very well have scored higher if a different interface had been chosen for evaluation. However, the STAT!Ref interface is how most institutions access the ACP's PIER product, accessing the product directly through ACP is reserved for individual subscribers. Similarly, the evaluations of the DISEASEDEX product may soon be obsolete as MICROMEDEX plans on completely redesigning the interface soon.

Midway through the evaluation process trial access to InfoRetriever expired. Some people were stalled in their evaluations because of this. It is unclear how this may have affected InfoRetriever's ratings. However, many users had commented negatively on InfoRetriever's interface before the problem with the trial access.

Although the five products evaluated were all products the library either subscribed to or were considering subscribing to, there are more candidates in this genre of tool that were not evaluated including DynaMed, Medweaver,



Clinical Medicine, eMedicine and Access Medicine to name a few. Most of these products are newer entrants to the marketplace or were not under consideration by the library for various reasons. Evaluation of these lesser-known products may have revealed a gem, which few users or librarians are familiar with.

The self reported measurements for time are very rough, and may have not been able to detect actual differences in time spent. A laboratory like setting would have allowed more precise measures of time.

### *Further Research*

The results of this study suggest several avenues for further research. Due to time constraints the qualitative data were not examined in depth. Both broad and strong themes could be identified. More thorough analysis of these data may reveal more themes and relationships. This analysis is planned for further research. This analysis could reveal and characterize the barriers to use that clinicians and students face when trying to use electronic resources at the point of care. By identifying and perhaps understanding those features which facilitate use and those which hinder use, product developers will have a framework to consult when building user interfaces for content designed to address acute clinical information needs.

This study did not address the correctness, or usefulness of the answers found. By doing so, content differences were likely not experienced by the users. Because of this, the importance of the interaction measures may have been magnified. By addressing whether or not the question was answered correctly as measured against some gold standard, differences in content could be further explored. Similarly, a study design where users tested the same question on each database, may result in users having stronger perceptions about differences in content.

## **Summary**

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. Including users in any evaluation of products is essential if one wishes to select a product that users will actually use. It may not be as difficult to engage users for this type of evaluation as is imagined by librarians.

## References

- Mount Sinai School of Medicine Web Development Team. Brandon/Hill Selected Lists>Introduction>Overview>Important Announcement. [10/20/2004]  
<http://www.mssm.edu/library/brandon-hill/index.shtml>.
- DOODY ENTERPRISES, INC. Doody Enterprises to publish core titles list. [10/20/2004]  
<http://www.doodyenterprises.com/news06012004.htm>.
- Healthcare Information and Management Systems Society (HIMSS). HIMSS Solution Toolkit. [10/20/2004] <http://www.solutions-toolkit.com>.
- Medical Informatics and Clinical Epidemiology Home Page. [2/28/05]  
<http://www.ohsu.edu/dmice/>.
- OHSU Library Web Managers. OHSU Library. 2005 [2/28/05]  
<http://www.ohsu.edu/library/>.
- OHSU Library Collection Development Committee. OHSU Library Collection Development Policy. 2004 [11/2/2004]  
<http://www.ohsu.edu/library/depts/colldevcomm/cdcpolicy.shtml>.
- OHSU School of Nursing Web Managers. OHSU School of Nursing Student Opportunities and Activities. 2005 [2/28/05]  
<http://www.ohsu.edu/son/student/opportunities.shtml>.
- OHSU Research Integrity Office. OHSU Study Opportunities. 2005 [2/28/05]  
<http://www.ohsu.edu/research/rda/so/>.
- SurveyMonkey.com because knowledge is everything. [11/29/2004]  
[www.surveymonkey.com](http://www.surveymonkey.com).
- Abate MA, Shumway JM, Jackowitz AI. Use of two online services as drug information sources for health professionals. *Methods of Information in Medicine* 1992; 31 (2):153-8.
- Alper BS. Practical evidence-based internet resources. *Family Practice Management* 2003; July/August:49-52.
- Alper BS, Stevermer JJ, White DS et al. Answering family physicians' clinical questions using electronic medical databases. *The Journal of Family Practice* 2001; 50 (11).
- Arrol B, Pandit S, Kerins D et al. Use of information sources among New Zealand family physicians with high access to computers. *The Journal of Family Practice* 2002; 51 (8).
- 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):351-7.

Badgett RG, Murlow CD. Welcome, PIER, a new physicians' information and education resource. *Annals of Internal Medicine* 2002; 136 (7):553-4.

Barnett G, Barry MJ, Robb-Nicholson C et al. Overcoming information overload: an information system for the primary care physician. *MEDINFO* 2004; 2004:273-6.

Belkin NJ, Carballo JP, Cool C et al. Rutgers' TREC-6 Interactive Track experience. In: EM Voorhees; DK Harman, editors, translator and editor *The Sixth Text REtrieval Conference (TREC-6)*. Gaithersburg, MD; 1998; p. 597-610.

Black S. An Assessment of social sciences coverage by four prominent full-text online aggregated journal packages. *Library Collections, Acquisitions, & Technical Services* 1999; 23 (4):411-9.

Brown ML. DISEASEDEX General Medicine System. *Journal of the Medical Library Association* 2004; 92 (2):280-2.

Bury J, Hurt C, Roy A et al. A quantitative and qualitative evaluation of LISA, a decision support system for chemotherapy dosing in childhood Acute Lymphoblastic Leukaemia. *MEDINFO* 2004; 2004:197-201.

Chambliss ML, Conley J. Answering clinical questions. *Journal of Family Practice* 1996; 43 (2):140-5.

Chueh H, Barnett GO. "Just-in-time" clinical information. *Academic Medicine* 1997; 72 (6):512-7.

Connelly DP, Rich EC, Curley SP et al. Knowledge resource preferences of family physicians. *The Journal of Family Practice* 1990; 30 (3):353-9.

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):164-76.

Curley SP, Connelly DP, Rich EC. Physicians' use of medical knowledge resources: preliminary theoretical framework and findings. *Medical Decision Making* 1990; 10:321-41.

D'Alessandro DM, Dreiter CD, Peterson MW. An evaluation of information-seeking behaviors of general pediatricians. *Pediatrics* 2004; 113 (1):64-9.

Dawes M, Sampson U. Knowledge management in clinical practice: a systematic review of information seeking behavior in physicians. *International Journal of Medical Informatics* 2003; 71:9-15.

Dickenstein R, Mills V. Usability testing at the University of Arizona Library: how to let the users in on the design. *Information Technology and Libraries* 2000; September 2000:144-51.

Doll WJ, Torkzadeh G. The Measurement of end-user computing satisfaction. *MIS Quarterly* 1988; June 1988:259-73.

Ebell MH. Point-of-care information that changes practice: it's closer than we think. *Family Medicine* 2003; 35 (4):261-3.

Ebell MH, Barry H. InfoRetriever: rapid access to evidence-based information on a hand-held computer. *M.D. Computing* 1998; 15 (5):289,91-97.

Ebell MH, Messimer SR, Barry H. Putting computer-based evidence in the hands of clinicians. *Journal of the American Medical Association* 1999; 281 (13):1171.

Ebell MH, Slawson D, Schaughnessy A et al. Update on InfoRetriever software. *Journal of the Medical Library Association* 2002; 90 (3):343.

Edwards P, Roberts I, Clarke M et al. Increasing response rates to postal questionnaires: systematic review. *British Medical Journal* 2002; 324:1-9.

Ely JW, Osheroff JA, Ebell MH et al. Analysis of questions asked by family doctors regarding patient care. *British Medical Journal* 1999; 319 (7206):358-61.

Ely JW, Osheroff JA, Gorman PN et al. A taxonomy of generic clinical questions: classification study. *British Medical Journal* 2000; 321:149-432.

Fitzpatrick RB. InfoRetriever basics: a resource for evidence-based clinical information. *Journal of Electronic Resources in Medical Libraries* 2004; 1 (1):103-11.

Forsythe DE, Buchanan BG, Osheroff JA et al. Expanding the concept of medical information: an observational study of physicians' information needs. *Computer and Biomedical Research* 1992; 25:181-200.

Fox GN, Moawad NS. UpToDate: a comprehensive clinical database. *The Journal of Family Practice* 2003; 52 (9):706-10.

Fox GN, Moawad NS, Music RE. FIRSTConsult: a useful point-of-care clinical reference. *The Journal of Family Practice* 2004; 53 (6):466-72.

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.

Garrison JA. UpToDate. *Journal of the Medical Library Association* 2003; 91 (1):97.

Grandage KK, Slawson DC, Shaughnessy AF. When less is more: a practical approach to searching for evidence-based answers. *Journal of the Medical Library Association* 2002; 90 (3):298-304.

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; 329:990-1.

Haynes RB, Walker CJ, McKibbin A et al. Performances of 27 MEDLINE systems tested by searches with clinical question. *Journal of the American Medical Association* 1994; 1 (3):285-95.

Hersh W, Crabtree MK, Hickam D et al. Factors associated with successful answering of clinical questions using an information retrieval system. *Bulletin of the Medical Library Association* 2000; 88 (4):323-31.

Hersh W, Hickam D. How well do physicians use electronic information retrieval systems? A framework for investigation and systematic review. *JAMA* 1998; 280 (15):1347-52.

Hersh W, Over P. Interactivity at the Text Retrieval Conference (TREC). *Information Processing and Management* 2001a; 37:365-7.

Hersh W, Pentecost J, Hickam D. A task-oriented approach to information retrieval evaluation. *Journal of the American Society for Information Science* 1996; 47 (1):50-6.

Hersh W, Turpin A, Price S et al. Challenging conventional assumptions of automated information retrieval with real users: Boolean searching and batch retrieval evaluation. *Information Processing and Management* 2001b; 37:383-402.

Hersh WR, Crabtree MK, Hickam DH 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):283-93.

Howse D. InfoRetriever 3.2 for Pocket PC. *Journal of the Medical Library Association* 2002; 90 (1):121-2.

Jennett PA, Kishinevsky M, Parboosingh IT et al. Responses to non-emergency questions in rural medicine: their usefulness to practice decisions. *Medical Education* 1991; 25:238-42.

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.

Kaplan B, Duchon D. Combining qualitative and quantitative methods in information systems research: a case study. *MIS Quarterly* 1988; December 1988 (4):571-86.

Kennedy J. InfoRetriever 2004. *Journal of the Medical Library Association* 2004; 92 (3):381-2.

Koonce TY, Giuse NB, Todd P. Evidence-based databases versus primary medical literature: an in-house investigation on their optimal use. *Journal of the Medical Library Association* 2004; 92 (4):407-11.

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.

Kushniruk AW, Patel C, Patel ML et al. 'Televaluation' of clinical information systems: an integrative approach to assessing web-based systems. *International Journal of Medical Informatics* 2001; 61 (1):45-70.

Lam WWT, fielding R, Johnston JM et al. Identifying barriers to the adoption of evidence-based medicine practice in clinical clerks: a longitudinal focus group study. *Medical Education* 2004; 38:987-97.

Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction* 1995; 7 (1):57-8.

Lord J, Ragon B. Working together to develop electronic collections. *Computers in Libraries* 2001:40-4.

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):567-80.

Magrabi F, Westbrook JI, Coiera EW et al. Clinicians' assessment of the usefulness of online evidence to answer clinical questions. *MEDINFO* 2004; 2004:297-300.

McCull A, Smith H, White P et al. General practitioners' perceptions of the route to evidence based medicine: a questionnaire survey. *British Medical Journal* 1998; 316 (7128):361-7.

Nielsen J, Landauer TK. A Mathematical model of finding of usability problems. In *Interchi '93*, 24-29 April 1993. 206-13 p.

Osheroff JA, Bankowitz RA. Physicians' use of computer software in answering clinical questions. *Bulletin of the Medical Library Association* 1993; 81 (1):11-9.

Osheroff JA, Forsythe DE, Buchanan BG et al. Physicians' information needs: analysis of questions posed during clinical teaching. *Annals of Internal Medicine* 1991; 114:576-81.

Pearl N. Gave 'em what they wanted. *Library Journal* 1996:136-8.

Peterson MW, Rowat J, Kreiter C et al. Medical students' use of information resources: is the digital age dawning? *Academic Medicine* 2004; 79 (1):89-95.

Pluye P, Grad RM. How information retrieval technology may impact on physician: an organizational case study in family medicine. *Journal of Evaluation in Clinical Practice* 2004; 10 (3):413-30.

Pugh GE, Tan JKH. Computerized databases for emergency care: what impact on patient care? *Methods of Information in Medicine* 1994; 33 (5):507-13.

Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. *British Medical Journal* 1995; 310 (6987):1122-6.

Sackett DL, Rosenberg WMC, Gray JAM et al. Evidence based medicine: what it is and what it isn't. *British Medical Journal* 1996; 312 (7023):71-2.

Schilling LM, Steiner JF, Lundahl K et al. Residents' patient-specific clinical questions: opportunities for evidence-based learning. *Academic Medicine* 2005; 80 (1):51-6.

Schwartz K, Northrup J, Israel N et al. Use of on-line evidence-based resources at the point of care. *Family Medicine* 2003; 35 (4):251-6.

Seol Y-H, Kaufman DR, Medonca EA et al. Scenario-based assessment of physicians' information needs. *MEDINFO* 2004; 2004:306-10.

Shaughnessy AF, Slawson D, Bennett JH. Becoming an information master: a guidebook to the medical information jungle. *Journal of Family Practice* 1994; 39 (5):489-502.

Simonson MT. Product Review: UpToDate. *National Network* 2000; 24 (2):16-7.

Smith R. What clinical information do doctors need? *BMJ* 1996; 313:1062-8.

Straus SE, Jones G. What has evidence based medicine done for us?: It has given us a good start, but much remains to be done. *British Medical Journal* 2004; 329:987-8.

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* 2003a; 54 (13):1175-92.

Su LT. A Comprehensive and systematic model of user evaluation of web search engines: II. an evaluation by undergraduates. *Journal of the American Society for Information Science and Technology* 2003b; 54 (13):1193-223.

Taylor MV. FIRSTConsult. *Journal of the Medical Library Association* 2004; 92 (2):285-7.

Tomasulo P. PIER - Evidence-based medicine from ACP. *Medical Reference Services Quarterly* 2004; 23 (3):39-48.

Travis TA, Norlin E. Testing the competition: usability of commercial information sites compared with academic library web sites. *College and Research Libraries* 2002:433-48.

Virizi RA. Refining the test phase of usability evaluation: how many subjects is enough? *Human Factors* 1992; 34 (4):457-68.



Wildemuth BM, DeBleik R, Friedman CP et al. Information-seeking behaviors of medical students: a classification of questions asked of librarians and physicians. *Bulletin of the Medical Library Association* 1994; 82 (3):295-304.

Zeng QT, Kogan S, Ngo L et al. Relationships among different subjective measurements of consumer health information retrieval performance. *MEDINFO 2004*; 2004:1328-2.

## Appendix A: Clinical Questions

1. *Is an annual urinalysis adequate to test for diabetic nephropathy in a diabetic?*
2. *What is the current thinking on treatment for rheumatoid arthritis?*
3. *What are the current guidelines for cholesterol levels?*
4. *What diagnostic tests are recommended for a male with atypical chest pain?*
5. *How do you treat Chondromalacia Patellae?*
6. *How do you evaluate angina?*
7. *How much bleeding is normal in postmenopausal women?*
8. *Is it appropriate to use nitroglycerin for episodes of atrial fibrillation?*
9. *What is the clinical course for following a child who is knock-kneed (genu valgum)?*
10. *Find an overview of renal hypertension.*
11. *What is the best approach to treating a woman with a history of abnormal paps?*
12. *When is a pelvic ultrasound indicated for a patient having pelvic pain?*
13. *What is the appropriate work-up for ADD?*
14. *What are the parameters for hyponatremia?*
15. *What's the treatment for gingivitis?*

## Appendix B: User Satisfaction Questionnaire

Please answer the following questions about your search experience by circling the appropriate number, or *Don't know/Not applicable* where appropriate.

Interacting with the system was:

*Easy 1 2 3 4 5 Hard*  
*Don't know/ Not applicable*

The layout of the screens was:

*Clear 1 2 3 4 5 Confusing*  
*Don't know/ Not applicable*

The database was:

*Fast 1 2 3 4 5 Slow*  
*Don't know/ Not applicable*

The content of the database seemed:

*Accurate 1 2 3 4 5 Inaccurate*  
*Don't know/ Not applicable*

Did the database give you:

*Too much information 1 2 3 4 5 Not enough information*  
*Don't know/ Not applicable*

Did the information in the database seem:

*Up-to-date 1 2 3 4 5 Out-of-date*  
*Don't know/ Not applicable*

Overall, this system satisfied my needs:

*Completely 1 2 3 4 5 Not at all*  
*Don't know/ Not applicable*

Did you spend any time familiarizing yourself with the product before you began the evaluation?

*Yes/No*

Do you have any other comments about your search experience?

## Appendix C: Background Questionnaire

Participant#

Please answer the following questions:

(optional) What is your age?

(optional) What is your gender?

What is your profession?

How many years have you been at your current profession?

If you are a student, what is your academic program?

What year are you in your academic program?

Do you use a computer: more than once a day/once a day/once a week/once a month or less? (circle one)

Do you own a computer?

If so, is it a PC/Mac/Other? (circle as many as apply)

Is it a laptop/desktop? (circle as many as apply)

Please circle the location where you use a computer the most:

home/work/computer lab/library/café/other

Do you own a PDA?

If so, is it a PocketPC/Palm/Other? (circle as many as apply)

Do you use a PDA: more than once a day/once a day/once a week/once a month or less? (circle one)

Please circle all of the following products which you have used before:

PubMed

MEDLINE on OVID

OHSU Catalog

MDConsult

FirstConsult

UpToDate

InfoRetriever

STAT!Ref

ACP Pier

Micromedex

Diseasedex

Cochrane Library

# Appendix D: Study Website

[OHSU Library](#)

## Bedside Information Tools

Welcome to the study and thank you for your participation. This site contains links to the Bedside Information Tools you are being asked to evaluate, as well as links to electronic versions of the evaluation forms. Please use the appropriate Bedside Information Tool to answer the clinical questions found in your packet. The questions for each tool as well as the order they are to be evaluated are different for each person. Please evaluate the tools in the order listed in your packet, not the order listed here.

Here are some things to keep in mind as you complete your evaluations:

- You don't have to evaluate all of the tools at one sitting. However, please complete the evaluation of a single tool at one time.
- Feel free to familiarize yourself with a tool before you begin the evaluation. If you would prefer to begin the evaluation immediately this is fine as well.
- Do not spend more than three minutes looking for the answer to any particular question.
- If you cannot find the answer to a question, that is perfectly fine. We're testing the tools, not your ability to use them!
- If you would rather complete your evaluations on the paper forms found in your packet, feel free. Simply return the paper forms to Rose Campbell, OHSU mail code BICC.
- If you have any question please feel free to contact Rose Campbell at [campbros@ohsu.edu](mailto:campbros@ohsu.edu) or 503-418-3742.

[Background Questionnaire](#)

[Evaluation Tool](#)

[ACP's PIER © on campus](#) | [ACP's PIER © off campus](#)

Select ACP's PIER: Physicians' Information and Education Resource from the list of resources, it should be the second one listed. By default all of the resources are highlighted, selecting a single resource will highlight that resource.

[DISEASEDEX™ on campus](#)

Please note that DISEASEDEX™ is not available off campus. Select the Search by database link on the right hand side. Select DISEASEDEX™ Emergency Medicine and DISEASEDEX™ General Medicine from the list of databases.

[FIRSTConsult on campus](#) | [FIRSTConsult off campus](#)

Please use the following logon information to access the product: Username=campbrosstudy & Password=study

[InfoRetriever® on campus](#) | [InfoRetriever® off campus](#)

Select the InfoRetriever® link from the left hand menu.

[UpToDate® on campus](#)

Please note that UpToDate® is not available off-campus.

[Back to the top](#)

*Last updated February 23, 2005 by the [OHSU Library Web Managers](#). Please send comments, questions, and reports of problems to [library@ohsu.edu](mailto:library@ohsu.edu) or use the [Library's customer support form](#).*

## **Appendix E: Recruitment Text**

We are looking for students and clinicians to evaluate electronic resources designed to answer clinical questions. You will be asked to find answers to clinical questions using various electronic resources. Afterwards you will be asked to complete a survey about your searching experience and participate in a follow-up telephone interview. The evaluation will take about one hour and the interview will take about twenty minutes. The evaluations can be done on your own time, from any computer on campus. The interview can be arranged at any time convenient to you. This project is jointly sponsored by the OHSU Library and the Department of Medical Informatics and Clinical Epidemiology. Contact Rose Campbell at [campbros@ohsu.edu](mailto:campbros@ohsu.edu) or 503-418-3742 for more information.

## Appendix F: Consent Form



**Oregon Health & Science University**  
Consent and Authorization Form

**IRB#: e493**

**Protocol Approval Date: 1/11/2005**

---

### OREGON HEALTH & SCIENCE UNIVERSITY Consent Form

**TITLE:** *Comparing Bedside Information Tools*

**PRINCIPAL INVESTIGATOR:** Joan Ash, PhD (503) 494-4540

**CO-INVESTIGATOR:** Rose Campbell, MLIS (503) 418-3742

**PURPOSE:**

You have been invited to be in this research study because you are a potential user of the computer systems being studied. The purpose of this study is to find out how users feel about various computer systems for answering clinical questions. The people making the decisions about what computer system to purchase are often not the same people who would be using the computer system. By asking the people who will be using the computer system their opinions about which ones work best, the people making purchasing decisions can make better choices.

The study will take place over the course of 8 weeks. The total time you spend evaluating the products will be about one hour. Afterwards, you will be interviewed over the telephone for about one hour. This interview will be audiotaped.

There will be 21 participants from OHSU enrolled in the entire study.

**PROCEDURES:**

You will be asked to use various computer systems to answer 3 questions. After you have used the system to answer the questions and recorded your answers to the questions you will be asked to rate your experience with the system. You will be asked to answer three different questions using 5 different

computer systems. This part of the study will take you about one hour. After you finish this task you will be called on the telephone for a follow-up interview. You will be asked about your experiences with the computer systems and what you would like to see in an ideal computer system. This part of the study will last about one hour. This interview will be audiotaped.

A copy of the form you will be using for the evaluations is attached. It includes the questions you will be using the computer system to answer as well as the questions you will be asked about the computer systems. It also includes some background questions that will help us learn how different types of people feel about these computer systems.

### **RISKS AND DISCOMFORTS:**

As evaluating these products will take about one hour, it may interrupt your normal activities at work or home. Similarly, the telephone interview may disrupt normal activities at work or home. You may use a computer product which you would not normally have access to. Even if you evaluate it favorably, it may not be selected for purchase.

### **BENEFITS:**

You may or may not personally benefit from being in this study. However, by serving as a subject, you may help us learn how to benefit computer users in the future. You may have access to computer products that you would not normally have access to. You may use this access to answer personal and professional questions other than those used in the study. You will have a chance to learn about computer systems that you may not be familiar with. Your evaluations will effect what computer products will be available to you in the future. You will have an opportunity to have your opinions heard. By participating in this study you may encourage future computer system purchasing decisions to include user opinions.

### **ALTERNATIVES:**

You may choose not to be in this study. Suggestions for computer system purchases can be made directly to the OHSU library or your own clinical or academic department. You can learn about these computer systems without being in this study by contacting the OHSU library.

### **CONFIDENTIALITY:**

We will not use your name or your identity for publication or publicity purposes.



**COSTS:**

There are no costs to you nor will you be paid for your participation.

**LIABILITY:**

If you believe you have been injured or harmed while participating in this research and require immediate treatment, contact Rose Campbell (503) 418-3742.

The Oregon Health & Science University is subject to the Oregon Tort Claims Act (ORS 30.260 through 30.300). If you suffer any injury and damage from this research project through the fault of the University, its officers or employees, you have the right to bring legal action against the University to recover the damage done to you subject to the limitations and conditions of the Oregon Tort Claims Act. You have not waived your legal rights by signing this form. For clarification on this subject, or if you have further questions, please call the OHSU Research Integrity Office at (503) 494-7887.

**PARTICIPATION:**

If you have any questions regarding your rights as a research subject, you may contact the OHSU Research Integrity Office at (503) 494-7887.

You do not have to join this or any research study. If you do join, and later change your mind, you may quit at any time. If you refuse to join or withdraw early from the study, there will be no penalty or loss of any benefits to which you are otherwise entitled.

You may be removed from the study if the investigator stops the study or if you do not follow instructions.

If you chose to withdraw from the study you will lose access to any computer systems you would not normally have access to.

The participation of OHSU students or employees in OHSU research is completely voluntary and you are free to choose not to serve as a research subject in this protocol for any reason. If you do elect to participate in this study, you may withdraw from the study at any time without affecting your relationship with OHSU, the investigator, the investigator's department, or your grade in any course.

We will give you a copy of this signed form.

**SIGNATURES:**

Your signature below indicates that you have read this entire form and that you agree to be in this study.

OREGON HEALTH & SCIENCE UNIVERSITY  
*INSTITUTIONAL REVIEW BOARD*  
  
*PHONE NUMBER (503) 494-7887*  
CONSENT/AUTHORIZATION FORM APPROVAL DATE  
**Jan. 11, 2005**  
  
***Do not sign this form after the***  
**Expiration date of: 1/10/2006**

---

Participant Printed Name                      Signature                      Date

---

Person Obtaining Consent                      Signature                      Date  
Printed Name

## Appendix G: Data Collection Sheet (Sample)

Participant # 10

Please complete this evaluation *fifth*.

Please answer the following three clinical questions using ACP's PIER© (logon instructions below). Please record the time you started using ACP's PIER© to look for an answer and the time that you found an answer to the clinical question. For the purposes of the evaluation, you do not need to record the answer to the question, just the time it took for you to find an answer. If you cannot find an answer in less than three minutes, simply note that an answer was not found. After you have answered all of the clinical questions please answer the questions about your searching experience.

Logon Instructions: From the Study's homepage:

<http://www.ohsu.edu/library/ref/campbros> select the ACP's PIER© link.

Select PCP's PIER: Physicians' Information and Education Resource from the list of resources, it should be the second one listed. By default all of the resources are highlighted, selecting a single resource will highlight that resource.

=====

1. *What is the current thinking on treatment for rheumatoid arthritis?*

Start time:                      Stop time:                      Answer found? Yes/No

2. *What is the clinical course for following a child who is knock-kneed (genu valgum)?*

Start time:                      Stop time:                      Answer found? Yes/No

3. *What is the best approach to treating a woman with a history of abnormal paps?*

Start time:                      Stop time:                      Answer found? Yes/No

=====

Please answer the following questions about your search experience by circling the appropriate number, or *Don't know/Not applicable* where appropriate.

Interacting with the system was:

*Easy 1 2 3 4 5 Hard*  
*Don't know/ Not applicable*

The layout of the screens was:

*Clear 1 2 3 4 5 Confusing*  
*Don't know/ Not applicable*

The database was:

*Fast 1 2 3 4 5 Slow*  
*Don't know/ Not applicable*

The content of the database seemed:

*Accurate 1 2 3 4 5 Inaccurate*  
*Don't know/ Not applicable*

Did the database give you:

*Too much information 1 2 3 4 5 Not enough information*  
*Don't know/ Not applicable*

Did the information in the database seem:

*Up-to-date 1 2 3 4 5 Out-of-date*  
*Don't know/ Not applicable*

Overall, this system satisfied my needs:

*Completely 1 2 3 4 5 Not at all*  
*Don't know/ Not applicable*

Did you spend any time familiarizing yourself with the product before you began the evaluation?

*Yes/No*

Do you have any other comments about your search experience?

## Appendix H: Interview Guide

Participant #

Date:

Time:

Location:

Telephone or In Person

In general, resources used to answer clinical information needs.

Positive and negative experiences with bedside information tools.

Process of turning information need into a query.

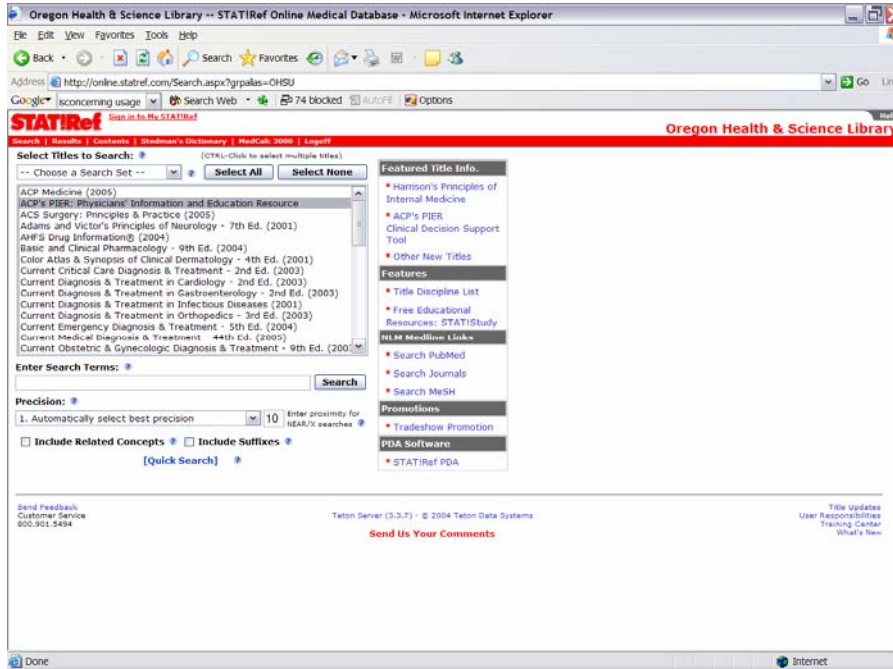
Features of an ideal bedside information tool.

Probe for how valuable having this information on a PDA would be.

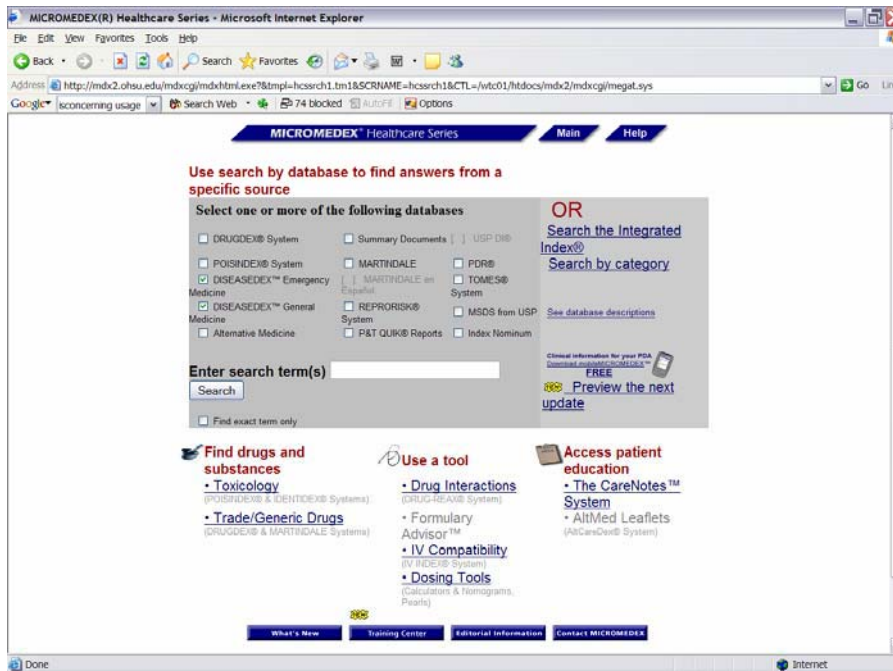
Rank the five resources from best liked to least liked.

# Appendix I: Screen Shots

## ACP's PIER



## DISEASEDEX



# FIRSTConsult

The screenshot shows the FIRSTConsult website in a Microsoft Internet Explorer browser window. The address bar displays [http://www.firstconsult.com/home/framework/1s\\_main.htm?](http://www.firstconsult.com/home/framework/1s_main.htm?). The page features a navigation menu on the left with categories like 'Differential Diagnosis', 'Medical Conditions', 'Procedure Files', 'Reference Centers', and 'Patient Education'. The main content area includes a 'Welcome back...' message, a 'FIRSTConsult Drug Watch' section with a 'New Feature!' announcement, and a 'Full Text Search' box. There are also several promotional banners for 'POCKETConsult' and 'What users are saying'.

# InfoRetriever

The screenshot shows the InfoRetriever search interface in a Microsoft Internet Explorer browser window. The address bar displays <http://www.infopoems.com/rsearch/rsearch.cfm>. The page features a navigation menu on the left with categories like 'Getting Started', 'Downloads', 'Register', and 'InfoRetriever®'. The main content area includes a search form with a dropdown menu for 'Keyword search + text in titles only', a text input field for 'STEP 2: Type in the medical term(s) to search for:', and a 'Find matching Keywords' button. There are also several promotional banners for 'POCKETConsult' and 'What users are saying'.

# UpToDate

