MEDLINE as a Source of Just-in-Time Answers to Clinical Questions

Dina Demner-Fushman[1] MD, PhD, Susan E. Hauser[1] PhD, Susanne M. Humphrey[1],
Glenn M. Ford[1], Joshua L. Jacobs[2] MD, George R. Thoma[1] PhD
[1]Lister Hill National Center for Biomedical Communications,
National Library of Medicine, NIH, DHHS, Bethesda, MD
[2]John A. Burns School of Medicine, University of Hawaii

Abstract
Clinicians increasingly use handheld devices to support evidence-based practice and for clinical decision support. However, support of clinical decisions through information retrieval from MEDLINE® and other databases lags behind popular daily activities such as patient information or drug formulary look-up. The objective of the current study is to determine whether relevant information can be retrieved from MEDLINE to answer clinical questions using a handheld device at the point of care. Analysis of search and retrieval results for 108 clinical questions asked by members of clinical teams during 28 daily rounds in a 12-bed intensive care unit confirm MEDLINE as a potentially valuable resource for just-in-time answers to clinical questions. Answers to 93 (86%) questions were found in MEDLINE by two resident physicians using handheld devices. The majority of answers, 88.9% and 97.7% respectively, were found during rounds. Strategies that facilitated timely retrieval of results include using PubMed® Clinical Queries and Related Articles, spell check, and organizing retrieval results into topical clusters. Further possible improvements in organization of retrieval results such as automatic semantic clustering and providing patient outcome information along with the titles of the retrieved articles are discussed.

Introduction
It is well known that clinicians need the best available and up-to-date evidence in order to provide the best care to their patients [1], and that questions about patient care occur frequently [2]. MEDLINE, a bibliographic database maintained by the National Library of Medicine® (NLM®) can provide answers to a substantial proportion of clinical questions [3]. This capability motivates our in-depth study of MEDLINE use at the point of care. This study is enabled by MD on Tap[1], an application for handheld devices developed specifically to answer various questions with respect to satisfying information needs that clinicians have while attending to patients [4]. Providing access to MEDLINE search and collecting

aggregate user statistics, the MD on Tap project explores types of devices used by clinicians, system response times, layout and navigation principles, information retrieval, and information organization options. All these issues are important, but still secondary to the availability of relevant information.

The focus of the current study is on the relevant information that can be retrieved from MEDLINE to answer clinical questions using a handheld device in a genuine clinical situation.

The MD on Tap application installed on Palm® Treo™ 650 smartphones was used by two resident physicians enrolled in a medical informatics elective. The residents followed daily rounds in a 12-bed intensive care unit (ICU) of a community teaching hospital as observers, recorded questions asked by any member of the clinical team, and immediately searched MEDLINE to find an answer as soon as possible. The MD on Tap server recorded all residents' interactions with the system. In addition, the residents submitted detailed daily reports. These two types of reports are combined in our analysis, and permit reconstruction of the information seeking process in a clinical situation without interfering with the course of clinical events.

Background
Evidence available to clinical team members during rounds has been shown to increase the extent to which it was sought and incorporated into patient care decisions [5]. In the study conducted by Sackett and Straus a clinical team had access to an evidence cart carrying a notebook computer and various sources of evidence including MEDLINE, which was used only if the locally compiled secondary sources were insufficient. The authors note that the contents of these secondary sources could be found in MEDLINE. The restriction on the use of MEDLINE could have been a side effect of the bulkiness of the cart, which could not be taken on bedside rounds and was kept in the team meeting room.

Limited portability of the desktop computers on carts motivated the introduction of the wireless handheld

---

technology into the clinical setting. Recent studies show that although clinicians increasingly use handheld devices to support evidence-based practice and for decision support, they rarely search medical literature at the point of care [6, 7]. For example, fourteen clinicians and librarians who participated in a study exploring the impact of handheld computers on patient care had access to Ovid MEDLINE [7]. Evaluating to what extent this resource assisted in patient care, three participants responded "marginally" and another three "not at all". Despite finding the results "reasonable", participants were dissatisfied because search results were presented in large numbers and without clear ranking. Another reason for dissatisfaction was the need to HotSync® with a PC to access search results, at which point one of the participants commented he or she "might as well have used Internet MEDLINE" [7].

Despite the fact that free immediate access to MEDLINE is available on handheld devices via Web browsers, as a standalone application [4], or as one of the features of a complex clinical information system [8], so far little is known about how well these systems are suited to satisfy clinicians' information needs at the point of care. The present paper evaluates one of the components necessary for ensuring usefulness of such systems – availability of the relevant information.

Methods

In the first stage of our study, two resident physicians specializing in internal medicine were provided with a modified MD on Tap application and unlimited Internet access on their Palm® Treo™ 650 cellular phones.
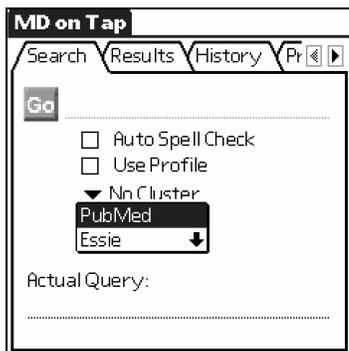

Figure 1. MD on Tap search screen

Three features were added to the experimental MD on Tap client: a user ID attached to every request sent to the MD on Tap server; MEDLINE search capability using Google API, in addition to two NLM search engines, PubMed and Essie [9], available in the publicly distributed MD on Tap client (see Figure 1); and the capability to take notes. The interactions between the experimental client and the server did not differ from previously described [4] ordinary client-server interactions.

A special Web-based desktop transaction review interface shown in Figure 2 was developed to facilitate residents' analysis of their daily activities. The review module provided detailed information about the conducted searches and the capability of viewing retrieved citations and the full text of the article, if available, and repeating the search with any of the search engines.

Figure 2. MD on Tap transaction review interface

The residents followed clinical teams consisting of an attending, chief resident, 5-8 medical residents and interns, and 2-4 medical students for two consecutive periods: 11/30/2005 through 12/16/2005 for the first resident, and 1/17/2006 through 2/9/2006 for the second resident. The second resident recorded 36 questions in the ICU of the teaching hospital and 17 questions during the non-ICU morning reports on the medical-surgical unit. The residents were unfamiliar with the MD on Tap application, but had some experience with MEDLINE/PubMed searches. Both residents had at least one formal session with a medical librarian on MD on Tap and on PubMed. They were also required to conduct searches using all provided search engines prior to participation in the rounds, and to become familiar with all available features, such as PubMed Clinical Queries[2] and MD on Tap organization of search results[10].

During the rounds, the residents were required to initiate a MEDLINE search as soon as any member of the team had a question. The residents were encouraged to use MD on Tap features at their discretion at this time.

---

[2]http://www.ncbi.nlm.nih.gov/entrez/query/static/clinical.shtml

The residents had to review their searches and file daily reports by the end of each day. For each clinical question they provided details of the clinical situation, approximate time when the question was asked, type of question (therapy, diagnosis, etiology, or prognosis), who asked the question, unique identifiers of the relevant articles, and comments about the search.

In the second stage of the evaluation an experienced MEDLINE indexer examined each relevant citation and, if needed, the full text of the article to answer two questions: 1) Whether the articles selected as relevant contained an answer to the question directly, were relevant to the question, or could have been selected as containing some interesting information not directly relevant to the question; 2) If more than one article was selected as relevant, how many of those contained novel information and were necessary to get a full answer to the question.

Search results were then analyzed with respect to the position of the relevant articles. That is, the number of titles that had to be read before seeing the first and finding the last relevant article was estimated assuming the results were displayed in a list. As shown in Figure 2, finding an answer sometimes required several interactions with the system. In such cases the search that retrieved the largest number of relevant articles for each query was selected for the analysis. In a tie, the search that presented the first relevant article earlier was selected.

In addition to the order in which citations were displayed, timestamps on the queries and citation retrieval were compared against the approximate time the question was asked. An answer was considered to be found during rounds if a citation marked as relevant, and subsequently verified as such by the indexer, was requested from our server within the round. Additional comments provided by the second resident were used to establish if an answer was found while the team discussed the patient whose condition prompted the question.

Results
Nature of questions
A total of 108 questions were recorded. The first resident collected 55 questions in 13 rounds, 4.2 questions per session on average. The second resident collected 53 questions in 15 rounds for an average of 3.5 questions per session. The questions were asked primarily by residents and attendings, with only 5 questions asked by interns in the set collected by the first resident. The distribution of questions by type

was similar for both observers (see Table 1). Most frequently asked questions pertained to therapy (48.1% of all questions). They were closely followed by etiology questions (35.2%). Only 3.7% of the questions pertained to prognosis, and 12.96% to diagnosis.

Table 1. Distribution of questions by type. Number of unanswered questions is shown in parenthesis. A=Attending, R=Resident, I=Intern

Who asked	Question type			
	Therapy	Diagnosis	Prognosis	Etiology
A1	9	3	2	8
A2	14	3		12
A total	23 (1)	6 (1)	2 (1)	20 (3)
R1	14	3	2	9
R2	12	5		7
R total	26 (2)	8 (1)	2	16 (1)
I	3	-	-	2

Finding answers
In searches conducted by the first resident only one question remained unanswered. For the remaining questions, 3.2 articles on average were selected as relevant. The second resident could not find an answer to 9 questions. For the remaining questions, 2.3 articles per question were found during the rounds (see Table 2).

Table 2. Number and distribution of articles contributing to answers to clinical questions.

	Number of answered questions	Average number of relevant articles	Average position of relevant articles in search results	
			first	last
Resident1	54	3.2	6.4	20.9
Resident2	44	2.3	8.6	20.3

The first resident found all answers using PubMed (49) and PubMed Related Articles (5). A relevant article was among the first five retrieved for 44 questions and first of the retrieval results for 19 of these. In the worst case the first relevant article was found in the 48th position and the last in the 90th.

After a short period of using no limits, this resident settled for a combination of spell check, date restricted to 1980 through present, and articles with abstracts, restricted to English and Human. This strategy was applied in 44 queries that led to an answer.

Answers to 2 of the 54 questions collected by the first resident were found after the rounds. Four of the

remaining answers, although found during the rounds were deemed only topically relevant in the secondary analysis, thus leaving 48 (88.9%) for which an answer was found during the rounds.

The second resident found answers to 17 questions using Essie, 1 question using Google, and the remaining 26 questions using PubMed. A relevant article was in the first five retrieved for 29 questions, and in 13 of these it was first in the list. In the worst case the first relevant article was found in the 43rd position, and the last relevant article was 86th.

The second resident's initial searches were restricted to articles with abstracts, and with Human and English filters set. Results were organized into strength of evidence categories. After several unsuccessful searches this strategy was changed to organizing results into subject clusters, turning the spell check on, using an appropriate Clinical Query, mostly "therapy broad", and sometimes adding English and Human filters to reduce the number of retrieved results. Spell check and topical clustering were used in 39 of 44 successful queries.

For all but one of the 43 answered questions (97.7%), the answers were found during rounds by the second resident. In several cases the comments tell us that the resident was able to find an answer immediately, which corresponds to approximately a one minute interval between the timestamps on the first query for a question and the first retrieved relevant citation.

Answer analysis
Each citation marked as relevant by the residents was subsequently evaluated by the medical indexer on a three-point scale as A – containing an answer, B – topically relevant to the question, and C – not obviously relevant to the question or the clinical scenario provided along with the question.

Table 3. Secondary evaluation of the number of questions for which answers were found in all (all A), some (A, B, C) or none of the citations.

	All A	A, B, C	No A
Resident1	31	19	4
Resident2	43		1

Table 3 presents the results of the secondary evaluation. Overall four citations selected by the first resident were rated C. In those cases other citations for the question either contained an answer, or were topically relevant. For only one question were two of the retrieved citations interchangeable. For all other questions each grade A citation provided novel information. Citations retrieved for all but one

question by the second resident were judged as containing an answer and non-redundant.

Discussion
The distribution of the collected questions is somewhat surprising and in slight disagreement with previous studies. We found the majority of questions, 48.1%, to be concerned with treatment options. These findings agree with previous studies, which report therapy questions as most frequent, ranging from 35% to 44% [2, 3]. The deviation in our findings concerns the second most frequent question type. In previous studies, 25% to 36% [2, 3] of the questions were about diagnosis. In our collection the diagnosis questions are in third place (12.96%). The second most frequent type in our collection is etiology comprising 35.2%. There might be two explanations for these differences: 1) Previously questions were collected from practicing family doctors or specialists whereas questions for our collection were collected in a teaching hospital; 2) The distinction between the differential diagnosis and etiology questions is fine and not always obvious. For example, in a scenario-based assessment of physicians' information needs a question "What is anemia?" was classified as a diagnosis type because the patient's test results showed abnormal findings indicating anemia, but "What is the cause of gastritis?" was classified as etiology because gastritis was not present in the test results [11]. By the same token, the question "What are secondary causes of erythrocytosis?" which was categorized as an etiology question by the resident, could have been assigned diagnosis based on the accompanying scenario: "58 y/o female presented with respiratory failure with Hb 16 g/dl."

The distribution of unanswered questions follows the overall question distribution. In the future, we will investigate if failures in finding answers were caused by lack of time, absence of the relevant information in the database, or other reasons.

The search strategies developed by both residents are similar with respect to goals and patterns, but differ in details. The first resident preferred a traditional list of answers, found a set of limits that permitted finding a relevant article close to the top of the list, and then used the Related Article tool. The second resident preferred topical clustering, permitting inspection of the most promising categories first, and therefore leading to a relevant article fast. For example, the Gastroenterology category was inspected first for the question "How to treat patient with upper GI bleeding and myocardial infarction?" Both residents, without being specifically instructed, looked for a comprehensive answer, i.e. continued

reviewing search results and interacting with the system after finding the first citation that contained an answer. For both residents the interactions did not involve changing the selected search strategy, but rather modification of the query. For example, when a query "secondary erythrocytosis" retrieved 262 citations, the resident added the term "cause" to the query, which reduced the total to 41, and then inspected twelve of the retrieved citations in four minutes, marking nine of these as relevant. All nine citations were graded A, and containing novel information in the secondary analysis.

Our study has some limitations. First, the residents participated in the rounds as observers thus having more time to address information needs of up to twelve team members. It is not clear how this experience reflects addressing clinicians' own information needs while taking care of patients. In the post-evaluation discussion the residents pointed out that there was not enough time to critically appraise search results even in an observing capacity. Second, we did not report if the answers were used in the clinical decision. Third, the moderate number of participants and questions provides purely qualitative information. Given the encouraging results of this study we are planning to recruit 50 or more participants to obtain quantitative results.

Conclusions and Future work
The results of our observational study demonstrate that MEDLINE, accessed via handheld devices, is a viable source of information for clinical decision support. Currently available resources accessed via handheld devices at the point of care provided answers to 86% of the clinical questions, the majority of those (88.9% - 97.7%) during rounds.

There are several issues that need to be resolved so that these resources can be used to their full capacity. Our results suggest a comprehensive answer to a clinical question will most probably be distributed among several documents. Topical organization of retrieval results is the first step toward bringing different parts of the answer to the attention of the searcher. Automatic semantic clustering of results might help find a distributed answer faster, for example, organizing results under extracted main interventions [12] to answer the question "How is Mallory-Weiss syndrome treated?" into the following categories: band ligation, endoscopic hemostasis, surgery, etc. Another approach that might facilitate evaluation of retrieved citations and decisions to pursue an article further is to provide automatically extracted patient outcome information [12] following the title of the citation, for example, "A prospective,

randomized trial of endoscopic band ligation vs. epinephrine injection for actively bleeding Mallory-Weiss syndrome. In this small study, no difference was detected in the efficacy or the safety of band ligation vs. epinephrine injection for the treatment of actively bleeding Mallory-Weiss syndrome."

We are currently exploring practical aspects of implementation of these two approaches.

References
1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ. 1996 Jan 13;312(7023):71-2.
2. Ely JW, Osheroff JA, Gorman PN, Ebell MH, Chambliss ML, Pifer EA, Stavri PZ A taxonomy of generic clinical questions: classification study. BMJ. 2000 Aug 12;321(7258):429-32.
3. Chambliss ML, Conley J. Answering clinical questions. J Fam Pract. 1996 Aug;43(2):140-4.
4. Hauser SE, Demner-Fushman D, Ford G, Thoma GR. PubMed on Tap: discovering design principles for online information delivery to handheld computers. Medinfo. 2004;11(Pt 2):1430-3.
5. Sackett DL, Straus SE. Finding and applying evidence during clinical rounds: the "evidence cart". JAMA. 1998 Oct 21;280(15):1336-8.
6. Baumgart DC. Personal digital assistants in health care: experienced clinicians in the palm of your hand? Lancet. 2005 Oct 1;366(9492):1210-22.
7. Honeybourne C, Sutton S, Ward L. Knowledge in the Palm of your hands: PDAs in the clinical setting. Health Info Libr J. 2006 Mar;23(1):51-9.
8. Chen ES, Mendonca EA, McKnight LK, Stetson PD, Lei J, Cimino JJ. PalmCIS: a wireless handheld application for satisfying clinician information needs. J Am Med Inform Assoc. 2004 Jan-Feb;11(1):19-28.
9. McCray AT, Ide NC. Design and implementation of a national clinical trials registry. J Am Med Inform Assoc. 2000 May-Jun;7(3):313-23.
10. Demner-Fushman D, Hauser SE, Ford G, Thoma GR. Organizing literature information for clinical decision support. Medinfo. 2004;11(Pt 1):602-6.
11. Seol YH, Kaufman DR, Mendonca EA, Cimino JJ, Johnson SB. Scenario-based assessment of physicians' information needs. Medinfo. 2004;11(Pt 1):306-10.
12. Demner-Fushman D, Lin J. Knowledge extraction for clinical question answering: preliminary results. In: Proceedings of the AAAI-05 Workshop on Question Answering in Restricted Domains; 2005;Jul 9-13; Pittsburgh. p. 1-10