Evaluating a federated medical search engine

Tailoring the methodology and reporting the evaluation outcomes

D. Saparova[1]; J. Belden [2]; J. Williams[3]; B. Richardson[1]; K. Schuster[1]

[1]School of Information Science and Learning Technologies, University of Missouri, Columbia, MO, 65211; [2]Department of Family and Community Medicine, University of Missouri , Columbia, MO 65212; [3]MedSocket, Columbia, MO 65211

Summary

Background: Federated medical search engines are health information systems that provide a single access point to different types of information. Their efficiency as clinical decision support tools has been demonstrated through numerous evaluations. Despite their rigor, very few of these studies report holistic evaluations of medical search engines and even fewer base their evaluations on existing evaluation frameworks.

Objectives: To evaluate a federated medical search engine, MedSocket, for its potential net benefits in an established clinical setting.

Methods: This study applied the Human, Organization, and Technology (HOT-fit) evaluation framework in order to evaluate MedSocket. The hierarchical structure of the HOT-factors allowed for identification of a combination of efficiency metrics. Human fit was evaluated through user satisfaction and patterns of system use; technology fit was evaluated through the measurements of time-on-task and the accuracy of the found answers; and organization fit was evaluated from the perspective of system fit to the existing organizational structure.

Results: Evaluations produced mixed results and suggested several opportunities for system improvement. On average, participants were satisfied with MedSocket searches and confident in the accuracy of retrieved answers. However, MedSocket did not meet participants' expectations in terms of download speed, access to information, and relevance of the search results. These mixed results made it necessary to conclude that in the case of MedSocket, technology fit had a significant influence on the human and organization fit. Hence, improving technological capabilities of the system is critical before its net benefits can become noticeable.

Conclusions: The HOT-fit evaluation framework was instrumental in tailoring the methodology for conducting a comprehensive evaluation of the search engine. Such multidimensional evaluation of the search engine resulted in recommendations for system improvement.

Correspondence to:

Dinara Saparova
111 London Hall, Columbia, MO 65211
Telephone: (573) 884–2737
E-mail: ds754@mail.missouri.edu

1. Introduction

Federated search engines that focus on medical content are examples of health information systems (HIS). These systems provide a single access point to different types of information retrieved from a variety of sources [1] including organizational resources and electronic patient data. The range of information retrieved by federated search engines from disparate locations can help physicians address their information needs that arise in a typical workday by conducting a single search.

There are a number of HIS that serve as federated search engines such as Federated Drug Reference [2], Clinical Focus [3], Quick Clinical [4], EBM Search [5, 6], LaneConnex [7], and InfoRetriever [8]. In clinical practice these systems have assisted physicians providing patient care by increasing physicians' use of evidence at the point of care [3, 9, 10], their confidence in answers to clinical scenarios [11], and accuracy of their answers to clinical questions [12]. In an educational context, these systems have been found to help lower the barriers to information resources [5] and increase the use of evidence-based resources for learning [8].

Rigorous evaluation of HIS is essential for better understanding of their true potential and value. Considering the many ways a HIS could be evaluated, it is important for researchers to adopt a framework that allows testing the essential characteristics and capabilities of HIS. An established and comprehensive framework devised to evaluate HIS in health care is the HOT-fit framework [13]. The HOT-fit framework considers the technological characteristics, demands and preferences of users, and organizational settings in order to evaluate the relevance and appropriateness of an HIS. It was devised to address the gaps in HIS evaluations identified through critical appraisal of prior relevant HIS literature and built on the Information Systems Success Model and the Information Technology Organization Fit Model [14].

The HOT-fit evaluation framework places equal emphasis on the importance of human, organization, and technology factors because each enables the demonstration of a system's net benefits (e.g., effects on clinical practice, job efficiency and effectiveness, quality of decisions, error reduction, communication, and clinical outcomes) during its evaluations. According to this framework, human-fit factor helps understand the nature of system use and user satisfaction when interacting with the system. Organizational-fit factor explains how an HIS fits in the current structure and environment of the organization, including its culture, processes, management, and financing sources. Technology-fit factor demonstrates system quality, quality of information it produces, and quality of service provided to maintain the system.

Several studies have evaluated federated medical search engines. However, when compared to the tenets of the HOT-fit evaluation framework, these evaluations focused on only one of the factors, i.e., human, or organization, or technology. For example, a number of studies evaluated federated medical search engines from the perspective of human-fit by focusing primarily on user satisfaction [3,6] or the patterns and nature of system use [11, 12, 15, 16]. Other studies evaluated systems' technology-fit by focusing on the quality of evidence available through these systems [17]. Even when studies demonstrated net benefits of federated search engines through their multidimensional evaluations, these studies did not clearly outline the evaluation frameworks used in their research approaches [9, 18, 19]. Despite the availability of methodological variety and rigor reported in these studies as well as significant findings, there is still a deficit of research that focuses on holistic evaluations of federated medical search engines.

In this article we demonstrate the application of the HOT-fit evaluation framework during the assessment of a federated medical search engine, MedSocket. This approach allowed us to employ a combination of qualitative and quantitative methods of data collection and analysis that altogether provided a comprehensive picture of the efficiency of Medsocket. The rigorous evaluation of the federated medical search engine resulted in a number of recommendations and implications that could be useful not only for physicians and informaticists but also system developers and evaluators.

2. Methods

2.1. Purpose of study

This study intended to demonstrate the net benefits and the potential fit of MedSocket to an organizational setting with established habits for clinical information acquisition. This goal was achieved by 1) tailoring the HOT-fit framework to the specific context, 2) deriving appropriate evaluation approaches, and 3) applying the newly developed methodology to the evaluation of MedSocket.

2.2. Context

MedSocket of Missouri Inc. created MedSocket, currently known as 1-Search. With a single search box, the system provides access to a variety of electronic information resources, e.g., textbooks, journals, evidence-based medicine sources, drug and image databases, handouts, guidelines, news as well as personal and organizational content. It offers custom aggregation based on an institution's subscriptions along with open source resources. MedSocket was built using Microsoft SharePoint and FAST search server. It crawls and indexes some sources and performs a live search query for other sources. MedSocket also uses MeSH terms to improve the search engine's accuracy by providing end users with suggested terms through auto-completion. In order to enhance the visibility of search results, MedSocket organizes retrieved resources one under the other, with a preview, indication of the source, and highlighted query (▶ Figure 1). Taken together, the system's design allows MedSocket to incorporate a number of features that allow users to perform searches in a specific information resource or simultaneously in several types of information resources.

A brief comparison of MedSocket with other federated medical search engines demonstrates the following similarities and differences among the systems. MedSocket, similar to Federated Drug Reference [2] and Clinical Focus [3], can be integrated with electronic health record systems. When compared with a Web Portal like Clinical Focus [3], MedSocket presents users with just one search box. As opposed to Federated Drug Reference [2], which federates various types of pharmaceutical information, MedSocket aggregates and filters search results in order to meet physicians' information needs at the point of care by aggregating results based on guidelines, patient handouts, drug information, EBM, images, journals, textbooks, personally stored content, and organizational content such as institution's subscriptions along with open source resources. MedSocket is similar to EBM Search [5, 6] and LaneConnex [7]. Both MedSocket and LaneConnex [7] allow users to customize their clinical profile configuration based on domain expertise (e.g., family medicine, pediatrics, dermatology etc.). However, LaneConnex [7] functions more like a library portal and, thus, is not optimal for point of care use. Also, while EBM Search [5, 6] and Information Retriever [8] focus on different levels of clinical information modeled on the evidence-based pyramid, they do not provide access to other types of information needed at the point of care like handouts and internal treatment protocols.

2.3. Research approach

The hierarchical structure of the HOT-factors allows for the measurement of human, organizational, and technological components of HIS through a variety of evaluation metrics [13]. For example, human factor is comprised of system use and user satisfaction. In its turn, system use is comprised of metrics such as the amount and duration of use, actual versus reported use, purpose of use, motivation to use, etc. While user satisfaction is comprised of satisfaction with specific functions, overall satisfaction, perceived usefulness, etc. Similar subdivision exists in the organization and technology factors.

In the context of this study, human fit of the system was evaluated through (a) user satisfaction with the system and (b) patterns of system use. Evaluation of technology fit was based on measurements of (a) user performance on time-on-task and (b) accuracy of their answers. Organization fit was evaluated from the perspectives of system fit to the existing organizational structure.

Our research approach included three phases. During phase 1 we identified a research strategy, prepared materials for use during system evaluations, and prepared a simulated experimental en-

vironment. System evaluations with perspective users took place during phase 2. Phase 3 consisted of developing approaches for data analysis.

2.4. Phase 1 – Preparing materials for evaluations

In clinical practice, physicians' information needs are usually very broad [20] and may include a combination of clinical, patient-specific [21, 22], and organizational [23] questions. Questions that arise during a patient visit are unique in each case and, for the most part, unpredictable. In order to reflect the complexity of physicians' information needs, this study used a paper form to collect questions from four family and community medicine clinics (▶Supplement Appendix A). The forms asked physicians to document their information needs that resulted from patient encounters. Furthermore, the form asked whether physicians pursued the information need presented by the question. When physicians pursued their information needs, they were asked to report the outcome, the time it took to complete their searches, and the type of information resource they used.

Fifty-one clinicians (11 resident physicians, 2 fellow physicians, 36 faculty physicians, and 2 who did not indicate their status) agreed to document their information needs over a one-week period during May, 2012. After the forms were returned, two physicians – a board-certified faculty family physician informaticist with 30 years of clinical experience and a volunteer resident first-year family physician – reviewed them and identified 138 information needs. Based on the initial review of their semantics, questions were grouped into three categories of patient-specific, clinical, and organizational (▶Table 1).

Guided by their working definitions, the reviewers independently categorized all collected information needs and after adjudicating individual differences agreed on the following count:
- 100 clinical questions (72% from the total count)
- 11 patient-specific questions (8% from the total count)
- 26 organizational questions (19% from the total count)[1].

Only 126 out of 138 questions collected had a clear indication of the action taken. Questions that the volunteer physicians did not provide answers to were predominantly clinical and organizational. Thus, we concluded that clinical and organizational questions were deemed by physicians as too difficult and/or time-consuming to answer. The perceived difficulty and time commitment these questions posed made them ideal for evaluating MedSocket with users (▶Supplement Appendix B).

Physicians are known to experience time constraints when addressing their information needs during patient visits [24, 25]. For example, physicians typically spend about two minutes searching for an answer to a question in presence of a patient. Analysis of the data reported through the paper forms in our study demonstrated that participating physicians spent on average three minutes pursuing an answer to a question during patient encounters. Since the calculated three-minute duration of the search time was only in one-minute proximity to the duration of searches previously reported [24,25], we decided to proceed with an allotment of three minutes per search. A pilot test confirmed that this time was sufficient to find an answer to each of the identified search scenarios.

2.5. Phase 2 – User evaluations of MedSocket

2.5.1. Participants

With the permission of the University's Institutional Review Board, we recruited a convenience sample of ten practicing physicians (5 males, 5 females) from the Department of Family and Community Medicine at a state university located in the Midwestern United States. The following factors supported our decision to recruit only ten participants to evaluate the system. Previous studies devoted to HIS evaluations and guided by the HOT-fit framework reported recruiting 15 participants [13]. Furthermore, Kushniruk, Patel & Cimino (1997) mention that as many as 8–10 subjects can

---

1 One question was left without identification of its type because it was not feasible to decipher the handwriting.

lead to identification of up to 80% of the surface level usability problems with an information system [26]. Thus, by limiting the sample size to ten participants we believe were able to receive meaningful feedback regarding system usability. Six participants were in the age group of 41–50 years old; the other four were in the age groups of 25–30, 31–40, 51–60, and 60+ years old. Six physicians had been in practice for more than ten years and four physicians for less than ten years. Participation in the study was voluntary, and no monetary compensation was offered to the physicians who chose to participate.

2.5.2. Procedures

Ten questions, divided into two randomized sets, comprised the evaluation scenarios. Each set of five questions contained a combination of clinical and organizational questions. The ten participants were randomly assigned to two groups. Participants in group 1 were asked to use MedSocket to find answers to the first set of five questions and then use their preferred search methods to find answers to the next set of five search scenarios. The order of the search scenarios for group 2 was reversed, which allowed us to counterbalance scenarios and avoid any order bias.

Data collection took place during June 2012. MedSocket was available to all medical professionals in the department since 2011 for an annual licensing fee paid by the department. Participants could participate in the study in either a laboratory setting or in the setting of their practice. Cable Internet connection was used in the laboratory setting and a wireless hospital internet in clinic. Eight sessions were conducted using Firefox 9.0.1, and two sessions – with Internet Explorer 9.0. All sessions were completed on Dell laptops, Intel (R) Core (TM) i7 CPU, Windows 7 Enterprise. Each participant was allowed a maximum of three minutes to conduct a search and provide an answer in a separate Word document. If the search exceeded three minutes, the participant was asked to cease his/her search regardless of whether or not he/she was able to find an answer. All the sessions were facilitated by two of the authors.

After each search task, each participant was surveyed about his/her confidence in the accuracy of the answers they found and satisfaction with his/her search experience (on a 1–5 Likert scale with 1=not at all satisfied/confident, 2=not quite satisfied/confident, 3=satisfied/confident, 4=quite satisfied/confident, 5=very satisfied/confident). Sauro & Dumas (2009) experimentally found that one-question usability Likert scale is a sensitive tool for measuring usability, particularly with small samples size (e.g., 10–12 users) [27]. Additionally, end-of-session semi-structured interviews were conducted to gain an overall impression of the physicians' perceptions of MedSocket. The use of one-question-questionnaire and semi-structured interviews provided a valid means to gauge user satisfaction with MedSocket.

2.6 Phase 3 – Data analyses approaches

Data analyses included several phases. We conducted descriptive analysis of metrics such as participants' confidence in the accuracy of the found answer, their satisfaction with the search experience, and time-on-task. To verify participants' confidence in the accuracy of the found answer, we developed a Gold Standard. Also, recordings from each session were transcribed, open-coded, and analyzed in order to assess physician feedback regarding system usability.

The Gold Standard was developed as a result of the combined expertise of a family physician who served as a domain expert, and a medical librarian who served as an information retrieval and medical reference expert. To derive the Gold Standard, a physician and the medical librarian independently performed searches for the best acceptable answers to the identified search scenarios. Initial searches were performed using MedSocket and, when necessary, supplemented with searches in other highly reliable, evidence-based information sources such as Micromedex, National Guideline Clearinghouse, Up-To-Date, DynaMed, Family Practice Notebook, and Cochrane Database of Systematic Reviews[2]. Upon finding unambiguous answers, the physician and librarian documented the

---

2 Micromedex (http://www.micromedex.com); National Guideline Clearinghouse (http://www.guideline. gov/); Up-To-Date (http://www.uptodate.com/home); DynaMed (https://dynamed.ebscohost.com/); Family Practice Notebook (http://www.fpnotebook.com); and Cochrane Database of Systematic Reviews (http://www.cochrane.org/).

URLs of the source and the excerpts containing the answer. These URLs and excerpts were used to compare their answers, adjudicate any differences, and construct the Gold Standard answers.

To apply the Gold Standard during data analysis, participants' answers were reviewed for accuracy by two family physicians – one board-certified faculty physician informaticist with 30 years of clinical experience and one volunteer resident first-year physician – in accordance with the prepared answer correctness criteria (▶Supplement Appendix C) and graded. The accuracy of answers was evaluated with the help of a 4-point grading scale that we developed by analogy with the grading scale developed by Thiele et al. [28]. Our choice of the 4-point scale was justified by the possibility of having several degrees of answer accuracy, e.g., participants arrived at the correct answer (4 points), partially correct answer (3 points), no answer (2 points), and incorrect answer (1 point).

3. Results

3.1. Human fit

3.1.1. User satisfaction

On average, participants reported that their search experience was satisfactory when using Med-Socket (*Mean (M) = 3.22, Standard Deviation (SD) = 1.41*) and their preferred search methods (*M = 3.56, SD = 1.43*). Despite the overall satisfaction level, participants' satisfaction with individual searches varied from question to question (▶Figure 2).

In terms of satisfaction with specific features, a number of participants liked having quick access to patient handouts available through MedSocket's left side menu (▶Figure 1). As one participant commented: "*...because MedSocket can search for handouts, that would be my first choice, ... because I know I could quickly focus MedSocket search to just patient handouts, which is a marvelous thing.*"

Some participants reported that MedSocket's left menu was helpful because it provided quick access to different types of resources; however, its functionality was not intuitive for other participants. One participant noted: "*I noticed that there were different topics like evidence-based medicine, drugs, but I don't know what that does for me ... Do I search generally or does it do me some good to click on one of those subtopics?*"

3.1.2. System use

The participants' answers to the demographic survey suggested that two participants used Med-Socket more than once a week; four physicians reported using the system more than once a month, and four physicians reported rarely or never using MedSocket. Participants expected MedSocket to be quick and efficient in finding information, which was not always the case. Observations revealed that in some instances it took some participants several attempts to generate a query that allowed them to retrieve the needed information. MedSocket also did not always recognize the misspelled queries thus resulting in two or three query reformulations. One participant complained: "*...for example, elevated liver enzymes. It took me forever. I did not actually find it. If you put it in other search engines that I am familiar with, automatically you will find it. It would take me a second. So, I guess I need to find the specific wording ... Every wording you put should bring you to the answers as quickly as possible.*"

Participants expected the presentation of search results to be concise; instead, continuous use of MedSocket generated more "clutter". For example, on one search occasion, the result list contained at least five results from the DynaMed database. As one participant commented: "*MedSocket is a little bit overwhelming in that it gives you so many resources and it really doesn't seem to tailor to the key words that I put in, at least initially.*"

3.2. Technology fit

3.2.1. System quality

The participants reported that MedSocket was easy and intuitive to use. There was no discernible learning curve in participants' interactions with the search engine. The actual system use, however, could have been improved by awareness of certain features, e.g., search filters or personalized fea-

tures. As one of the participants commented: "*I don't know if I was using the most effective search strategies ... like knowing whether to search all types or knowing to limit the search to a certain type of literature.*" Only one participant noticed and used the embedded Google search feature.

Time-on-task measurements showed that when the participants used MedSocket, they were able to find answers 1.092 minutes faster to four out of ten questions. These questions were about patient information (Q1), differential diagnosis (Q3), therapy (Q5), and medication doses (Q10). On the other hand, Medsocket was 4.088 minutes slower in finding answers to the other six questions. These questions included organizational questions (Q2, Q7), drug interaction (Q4), patient information (Q6), diagnosis (Q8), and disease (Q9) (▶Figure 3). Overall, it took participants 2.996 minutes longer to find the answers to the ten questions when using MedSocket (M = 2.17, SD = 0.90) than through their preferred search method (M = 1.879, SD = 0.98).

3.2.2. Information quality

The accuracy of answers provided a metric for the quality of information retrieved during a search. This measure consisted of participants' self-reported confidence in the accuracy of the answers and the accuracy of the answers assigned by the two reviewers after applying the Gold Standard. Similar to user satisfaction, the participants, on average, reported being confident in the accuracy of the answers found through MedSocket (M = 3.54, SD = 0.70) and their preferred search methods (M = 3.72, SD = 1.41); nevertheless, there were differences in the level of participants' confidence with regards to certain questions depending on the search method (▶Figure 4).

The application of the Gold Standard produced different results. The initial inter-rater reliability between the reviewers' was around 60% (k=0.621, p< 0.001). Further discussion resolved the majority of the remaining discrepancies and resulted in the 90% final agreement (k=0.9, p<0.001). The reviewers concluded that answers to questions about treatment (Q5, Q10) were most accurate when found through MedSocket. Answers to questions about diagnosis (Q3, Q8), drug interaction (Q4), and pharmacy hours (Q7) were most accurate when found through participants' preferred search methods. Answers to question about patient information (Q1), organization question (Q2), patient information (Q6), and question about disease evaluation (Q9) were most accurate regardless of the search method (▶Table 2).

3.3. Organization fit

The findings revealed that in some instances an established organizational culture prevented physicians from adopting and using MedSocket. Observations of the participants' preferred search methods revealed that they mostly used UpToDate and DynaMed to find clinical information. For organizational questions, the participants preferred to look for answers through Google searches.

Despite these organizational barriers physicians found MedSocket's capability to search through internal resources (e.g., call schedules, protocols, and policies) useful for finding organizational information relevant to their practice. As for clinical information, physicians expected MedSocket to suggest information sources that were familiar to them or frequently used in their institution. As one participant commented: "*Usually [I use MedSocket] if I'm trying to find out a piece of administrative information. Like, if something is covered [by insurance] under this or how do I need to do a surgical procedure. [These are] administrative, usually not clinical questions. I usually go somewhere else for clinical resources ... Honestly, I use UpToDate for clinical questions*".

During the interviews a few participants commented that they could not easily locate the link to MedSocket among the organizational intranet resources. Better visibility of the MedSocket link on the departmental resources page could have facilitated access to and possibly more frequent use of the system. The participants added that access to MedSocket was also complicated on shared computers in some clinics. They bookmarked the MedSocket link on computers in clinic; however, when they practiced from a different clinic location, access to their MedSocket bookmark was not available.

4. Discussion

4.1. Discussion of the methodology

We presented an evaluation of MedSocket in an effort to understand its net benefits. This study was necessary to understand the potential of an alternative information retrieval tool being introduced to the Department of Family and Community Medicine at a large university.

Compared to previously reported evaluations of federated medical search engines [2–8], this study utilizes an existing comprehensive evaluation framework, the HOT-fit. The major benefit of relying on this evaluation framework was the possibility to tailor and derive an evaluation methodology that was most suitable to the context of the study. HOT-fit thus allowed us to test the functional and technological aspects of MedSocket based on user experiences and needs.

The derived methodology consisted of several phases, each serving a particular purpose and contributing to the overall methodology strength. Asking physicians to report their information needs at the point of care during phase 1 allowed us to identify a new category of questions – organizational – and include them in system evaluations, which has not been previously reported. Including organizational and clinical questions as search scenarios resulted in more comprehensive evaluations of the search engine by evaluating its ability to retrieve information beyond clinical information sources.

Phase 2 included evaluations of MedSocket with ten representative users – practicing family physicians. The uniqueness of this user study was in evaluating the efficiency of the system in simulated environments resembling real-world working situations of family physicians during patient visits. Such experimental conditions facilitated opportunities for capturing patterns of physicians' perspective use of the system in clinical context. Additionally, verbal feedback from participants allowed us to identify usability issues with the system, which was extremely important as usability evaluations become widely recognized as critical to the success of adopting interactive health information systems [30].

Use of several comparative techniques embedded in the methodology provided us with an opportunity to obtain a more comprehensive, complete picture of MedSocket's efficiency. First, comparing physicians' performance when using their preferred search methods against MedSocket allowed us to understand whether the search engine was able to meet physicians' information needs as they might encounter them in their current practice. Second, comparing physicians' self-reported metrics of MedSocket's efficiency with the developed Gold Standard allowed us to obtain a more objectified, complete picture of information quality retrieved by the HIS. Combined expertise used in the development of the Gold Standard and scoring of the participants' answers for accuracy allowed us to evaluate the system against both the high quality information resources and practical utility.

The methodology developed for this study contributes to knowledge of evaluative methodologies applicable to federated search engines. Furthermore, it will add to existing research on understanding the utility of HIS in general medical practice and the factors affecting their adoption. Although the derived methodology was applied to a specific case study of MedSocket, we believe it can be replicated in the evaluation of other federated medical search engines.

Critical appraisal of the developed methodology resulted in identifying a number of opportunities for its improvement. First, physicians' questions were self-reported through the paper forms. Had the questions been collected through physician observations during their practice and documented by researchers, the final pool of collected questions could have been distributed differently. Second, despite the fact that questions for user evaluations were selected based on the suggestions of the two reviewers, there could still be a possibility for bias. Third, having a separate group of reviewers subsequent to the finalization of methods for coding against the Gold Standard could have added to methodology strength. There is a need for more studies addressing evaluations of federated medical search engines holistically, to further validate the usefulness of the derived methodology.

4.2. Discussion of the evaluation results

This study confirmed the importance of performing usability evaluations of HIS before implementation [26]. The human fit of MedSocket was determined by the actual and expected use of the sys-

tem and user satisfaction; on average participants reported that while using MedSocket they were satisfied with their search experience and confident in the accuracy of found answers to queries. At the same time, several opportunities for improvement emerged. For example, physicians who are required to perform quick searches expected the system to automatically recognize misspelled queries, which did not always happen. Additionally, the participants were often frustrated by the slow download speed of the federated search engine, which ultimately affected their performance on tasks in terms of time. Still, on average MedSocket searches were within the 2-minute interval (M=2.17, SD=0.471), a time that previous research has demonstrated to be an acceptable benchmark for conducting searches [24, 25]. Finally, the participants wanted a more accurate and precise presentation of the results to make the process of selecting the most appropriate search results more efficient.

The identified issues negatively affected user perceptions of relevance, access, and speed of Med-Socket, which are critical factors to retrieving clinical information [28]. This, in its turn, suggested a tight bond and a causal relationship between the human and technology factors [13]. If left unattended, these issues will turn into significant barriers for technology acceptance [30] and adoption [31].

The findings of the study also suggest that to facilitate the acceptance and adoption of a novel system, certain modifications and customization need to take place. If the new system significantly differs from the one used currently, then users may show resistance and reject the innovation [32]. In other words, an innovation needs to be compatible with the established habits and behaviors of individuals to be assimilated into their work flow [31]. In this study, for example, physicians, being a part of a certain organizational culture, demonstrated established patterns of information behavior and strong preferences for certain information resources, which hindered their perceptions of Med-Socket's utility.

We also concluded that even if the system is perceived easy and intuitive to use, training remains necessary to promote realistic expectations on the part of the users [33]. Training sessions can also be helpful in demonstrating the system's potential impact on physicians' clinical practice. Although it was not the purpose of this study to demonstrate an impact of MedSocket on organizational level (e.g., cost reduction, fewer medication errors, and other clinical outcomes), we believe MedSocket could have a positive impact on an individual level.

We believe that if and when the identified shortcomings are properly addressed, there is a potential for MedSocket to function more efficiently. This will result in an increase of physicians' confidence in the accuracy of found information, allow them to conduct faster searches, and provide access to multiple types of information through a single access point. By making a positive change on an individual level, use of MedSocket can eventually lead to improvements on the level of the whole organization.

5. Conclusions

This study evaluated MedSocket in a clinical simulated setting. The results of the study suggested several opportunities for system improvement and identified the system's potential to positively affect clinical practice. It is our hope that once the improvements are made, the use of MedSocket could improve physicians' work effectiveness at the point of care.

Clinical Relevance Statement
Successful adoption of a new system will largely depend on its human, technology, and organization fit. Implementation of a new system should account for individual preferences and established behaviors as well as existing organizational culture.

Human Subjects Protections
The study was approved by the Institutional Review Board of the University of Missouri (IRB Project #1201075).
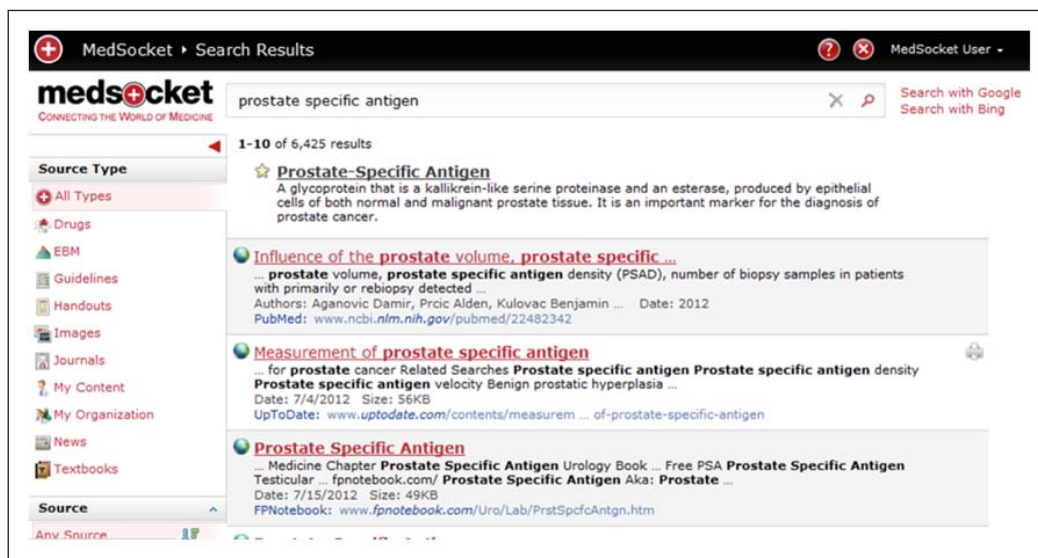
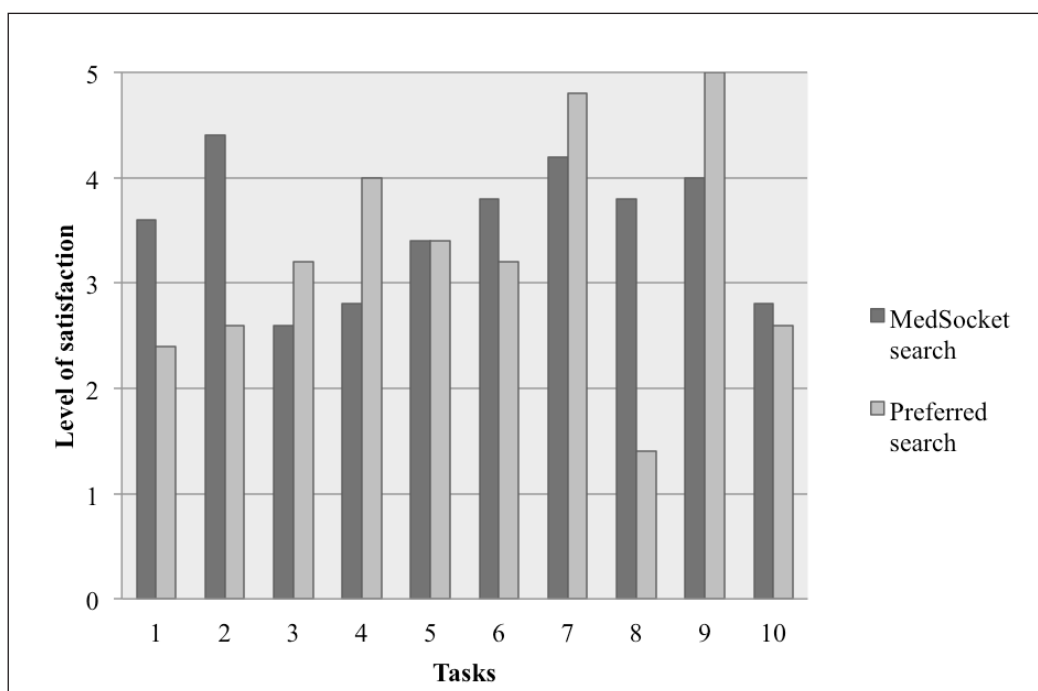**Fig. 1**   MedSocket interface

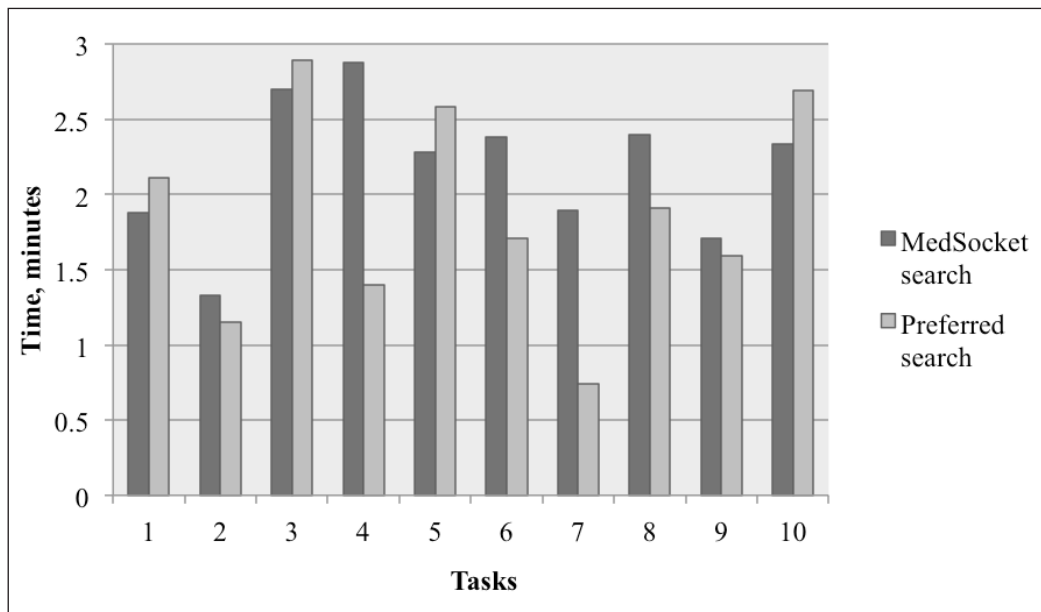**Fig. 2**   Comparison of participants' satisfaction with searches
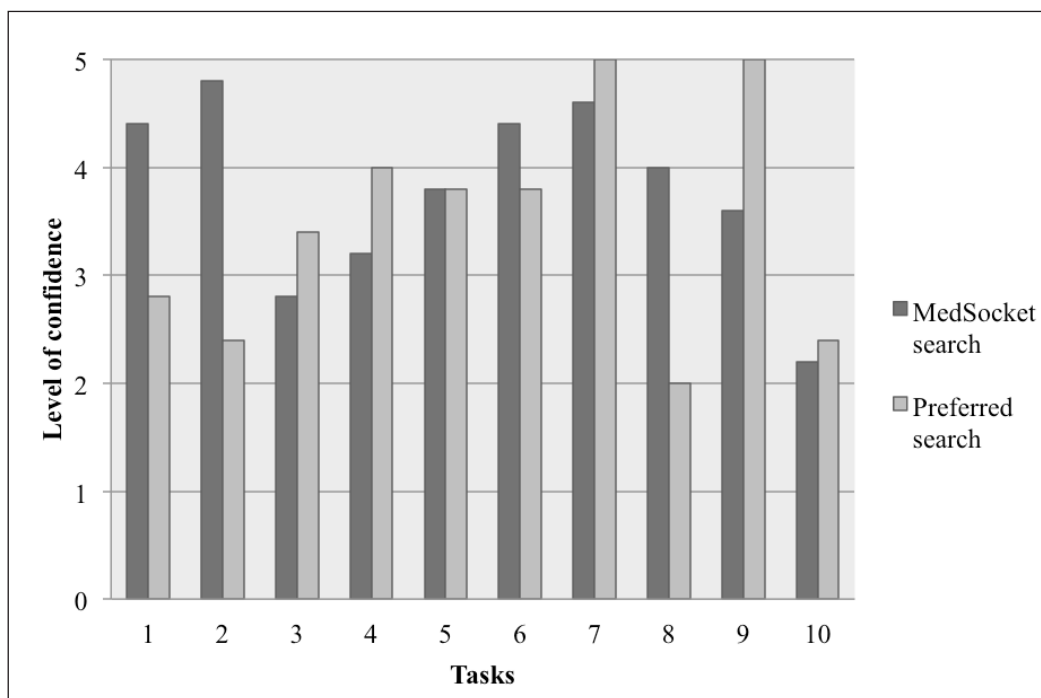
Fig. 3   Time-on-task comparison

Fig. 4   Comparison of participants' confidence in the accuracy of answers

Table 1 Categorization of questions

Type	Definition	Source(s)	Scope	Example(s)
Patient-specific	A healthcare-related information need that is tied to a specific person getting treatment in a healthcare setting.	EHR; HIE	Specific; Usually 1 person	Did this patient ever get the shingles vaccine before? What was this patient's last potassium level? What medications is this patient taking?
Clinical	A prevention, etiology, symptom, diagnosis, therapy or prognosis related information need that could be applied to a broader population of people, who may or may not be getting treatment in a healthcare setting.	Best medical sources (e.g. UpToDate, DynaMed, PubMed, ePocrates), but may also be from less reputable sources from the web, news sites, or Wikipedia.	General; Could be applied to more than 1 patient	Should I give the shingles vaccine to a patient that has already had Shingles? What are the side effects of nortriptyline? What are the best drugs to use for lowering BP in a patient with diabetes and heart failure?
Organizational	A healthcare-related administrative or financial information need that could be applied to a population of patients that are affiliated with certain healthcare related systems.	Intranet; Insurance company database	Generally applied to more than one patient with a relationship to specific systems (health systems, insurance company)	Will this patient's insurance cover the shingles vaccine? What is the local policy on how to restrain an incompetent combative patient? Which neurologist in our health system should I refer a patient with uncontrolled migraines to?

Table 2 Reviewers' evaluations of answer accuracy: Interpretation: First, we calculated an average accuracy score per search method for ten participants (column (b) for MedSocket; column (c) for preferred search). Then, we calculated the difference between the accuracy scores (column (d)) by detracting values in column (b) from values in column (c). If the difference score was a positive value (e.g., 0.6), we concluded that MedSocket search was superior in terms of answer accuracy. If the difference score was a negative value (e.g., –0.6), we concluded that MedSocket search was at disadvantage in terms of answer accuracy. Columns (e), (f), and (g) demonstrate superiority of the search method in terms of answer accuracy.

Q, #	Average answer accuracy by		Score difference	Searches favoring		
	MedSocket	preferred search		MedSocket search	preferred search	both search methods
(a)	(b)	(c)	(d)	(e)	(f)	(g)
1	3.8	3.8	0			Both
2	4	4	0			Both
3	3.2	3.8	-0.6		Preferred	
4	3.4	3.8	-0.4		Preferred	
5	2.8	2.2	0.6	MedSocket		
6	2.4	4	-1.6			Both
7	3.2	3.6	-0.4		Preferred	
8	3.6	4	-0.4		Preferred	
9	3.6	3.6	0			Both
10	4	3.2	0.8	MedSocket		

References

1. Daily G. Case study-A Case of Clustered Clarity-Vivisimo, Inc. helps University of Pittsburgh Health Sciences Library System patrons effectively search more than 300 health and biomedical titles in ebook. EContent-Digital Content Strateg Resour 2005; 28(10): 44–46.

2. Ketchell DS, Ibrahim K, Murri N, Wareham P, Bell D, Jankowski T, A. Architecture for a Federated Drug Reference in a managed care environment. Proc AMIA Annu Fall Symp 1996; 413–417.

3. Tannery NH, Epstein BA, Wessel CB, Yarger F, LaDue J, Klem M Lou. Impact and User Satisfaction of a Clinical Information Portal Embedded in an Electronic Health Record. Perspect Heal Inf Manag 2011; Fall(8(Fall)): 1d.

4. Coiera E, Walther M, Nguen K, Lovell NH. Architecture for knowledge-based and federated search of online clinical evidence. J Med Internet Res 2005; 7(5): e52.

5. Bracke PJ, Howse DK, Keim SM. Evidence-based Medicine Search: a customizable federated search engine. J Med Libr Assoc 2008; 96(2): 108.

6. Keim SM, Howse DK, Bracke PJ, Mendoza K. Promoting evidence based medicine in preclinical medical students via a federated literature search tool. Med Teach 2008; 30(9–10): 880–884.

7. Ketchell DS, Steinberg RM, Yates C, Heilemann HA. LaneConnex: an integrated biomedical digital library interface. Inf Tech Lib 2013; 28(1): 31–40.

8. Leung GM, Johnston JM, Tin KY, Wong IO, Ho L, Lam WW, Lam T. Randomised controlled trial of clinical decision support tools to improve learning of evidence based medicine in medical students. BMJ 2003; 8(327(7423)): 1090.

9. Magrabi F, Coiera EW, Westbrook JI, Gosling AS, Vickland V. General practitioners' use of online evidence during consultations. Int J Med Inf 2005; 74(1): 1–12.

10. Van Duppen D, Aertgeerts B, Hannes K, Neirinckx J, Seuntjens L, Goossens F, Van Linden A. Online on-the-spot searching increases use of evidence during consultations in family practice. Patient Edu Couns 2007; 68(1): 61–65.

11. Westbrook JI, Gosling AS, Coiera EW. The impact of an online evidence system on confidence in decision making in a controlled setting. Med Decis Mak 2005; 25(2): 178–185.

12. Westbrook JI, Coiera EW, Gosling AS. Do online information retrieval systems help experienced clinicians answer clinical questions? J Am Med Inf Assoc 2005; 12(3): 315–321.

13. Yusof MM, Kuljis J, Papazafeiropoulou A, Stergioulas LK. An evaluation framework for Health Information Systems: human, organization and technology-fit factors (HOT-fit). Int J Med Inf 2008; 77(6): 386–398.

14. Yusof MM, Papazafeiropoulou A, Paul RJ, Stergioulas LK. Investigating evaluation frameworks for health information systems. Int J Med Inf 2008; 77(6): 377–385.

15. Westbrook JI, Gosling AS, Coiera EW. Do clinicians use online evidence to support patient care? a study of 55,000 clinicians. J Am Med Inf Assoc 2004; 11(2): 113–120.

16. Westbrook JI, Gosling AS, Westbrook M. Use of point-of-care online clinical evidence by junior and senior doctors in New South Wales public hospitals. J Intern Med 2005; 35(7): 399–404.

17. Magrabi F, Westbrook JI, Coiera EW, Gosling AS. Clinicians' assessments of the usefulness of online evidence to answer clinical questions. In: Fieschi M et al., editor. MEDINFO 2004. Amsterdam: IOS Press; 2004; 297–300.

18. Westbrook JI, Coiera EW, Gosling AS, Braithwaite J. Critical incidents and journey mapping as techniques to evaluate the impact of online evidence retrieval systems on health care delivery and patient outcomes. Int J Med Inf 2007; 76(2–3): 234–245.

19. Coiera EW, Westbrook JI, Rogers K. Clinical Decision Velocity is Increased when Meta-search Filters Enhance an Evidence Retrieval System. J Am Med Inf Assoc 2008; 15(5): 638–646.

20. Covell DG, Uman GC, Manning PR. Information needs in office practice: are they being met? Ann Intern Med 1985; 103(4): 596–599.

21. McConaghy JR. Evolving medical knowledge: moving toward efficiently answering questions and keeping current. Prim Care 2006; 33(4): 831–837.

22. Flynn MG, McGuinness C. Hospital clinicians' information behaviour and attitudes towards the "Clinical Informationist": an Irish survey. Heal Info Libr J 2011; 28(1): 23–32.

23. Hughes B, Wareham J, Joshi I. Doctors' online information needs, cognitive search strategies, and judgments of information quality and cognitive authority: how predictive judgments introduce bias into cognitive search models. J Am Soc Inf Sci Technol 2010; 61(3): 433–452.

24. Ramos K, Linscheld R, Schafer S. Real-time information-seeking behavior of residency physicians. Fam Med KC 2003; 35(4): 257–260.

25. Ely JW, Osheroff JA, Ebell MH, Bergus GR, Levy BT, Chambliss ML, Evans ER. Analysis of questions asked by family doctors regarding patient care. BMJ 1999; 7(319(7206)): 358–361.
26. Kushniruk AW, Patel VL, Cimino JJ. Usability Testing in Medical Informatics: Cognitive Approaches to Evaluation of Information Systems and User Interfaces. Proceedings of the 1997 AMIA Fall Symposium. 1997; 218–222.
27. Sauro J, Dumas JS. Comparison of Three One-Question, Post-Task Usability Questionnaires. Proceedings of CHI 2009. Boston, MA; 2009.
28. Thiele RH, Poiro NC, Scalzo DC, Nemergut EC. Speed, accuracy, and confidence in Google, Ovid, PubMed, and UpToDate: results of a randomised trial. Postgr Med J 2010; 86: 459–465.
29. Bennett NL, Casebeer LL, Kristofco RE, Strasser SM. Physicians' Internet information-seeking behaviors. J Contin Educ Health Prof [Internet]. 2004; 24(1): 31–38. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15069910.
30. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. MIS Q 2003; 27(3): 425–478.
31. Rogers EM. Diffusion of innovations. 4th ed. Simon and Schuster; 2010.
32. Kim H-W, Kankanhalli A. Investigating User Resistance to Information Systems Implementation: A Status Quo Bias Perspective. MIS Q 2009; 33(3): 567–582.
33. Xia W, Lee G. The Influence of Persuasion, Training, and Experience on User Perceptions and Acceptance of IT Innovation. Proceedings of the 21st international conference on Information Systems. 2000; 371–384.