Cross-System Evaluation of Clinical Trial Search Engines

Silis Y. Jiang[1], BS, Chunhua Weng[1], PhD
[1]Department of Biomedical Informatics, Columbia University, New York, NY 10032

Abstract

Clinical trials are fundamental to the advancement of medicine but constantly face recruitment difficulties. Various clinical trial search engines have been designed to help health consumers identify trials for which they may be eligible. Unfortunately, knowledge of the usefulness and usability of their designs remains scarce. In this study, we used mixed methods, including time-motion analysis, think-aloud protocol, and survey, to evaluate five popular clinical trial search engines with 11 users. Differences in user preferences and time spent on each system were observed and correlated with user characteristics. In general, searching for applicable trials using these systems is a cognitively demanding task. Our results show that user perceptions of these systems are multifactorial. The survey indicated eTACTS being the generally preferred system, but this finding did not persist among all mixed methods. This study confirms the value of mixed-methods for a comprehensive system evaluation. Future system designers must be aware that different users groups expect different functionalities.

Introduction

Clinical trials are the gold standard for establishing the effectiveness or efficacy of new drugs and treatments. One of the challenges to successfully completing clinical trials is recruiting enough research participants[1]. An approach to increase the public's awareness of clinical trials is the publicly accessible clinical trial repository. Currently, several countries have such repositories[2]. In the United States in 2007, the FDA mandated that all new U.S.-based clinical trials be registered with the repository ClinicalTrials.gov[1].To date, ClinicalTrials.gov contains more than 153,260 clinical trials[3]. While ClinicalTrials.gov offers centralized access to these trials, searching for relevant trials remains difficult for the average user[2]. Often, the complex and technical language used to describe a trial and its eligibility criteria are difficult for a user to comprehend[4].

Anecdotally, the largest user group of ClinicalTrials.gov consists of those who seek to participate in a clinical trial. Besides using ClinicalTrials.gov, many users are also utilizing other ClinicalTrials.gov extension systems to search for trials. Many commercial clinical trial search engines have emerged, including Corengi.com[5], TrialX.com[6], eTACTS (http://is.gd/eTACTS)[7], Patientslikeme.com[8], and TrialReach[9]. Most are disease-specific, such as Dory/TrialX (cancer) or Corengi (diabetes mellitus type II). Others, such as Patientslikeme, allow users to search for clinical trials related to specific disease types.

While clinical trial repositories and search engines have existed for more than a decade now, there is little research on how users search for and process clinical trial information. In 2008, Atkinson conducted an expert usability analysis of various cancer-specific clinical trial search engines[4]. In 2010, Patel conducted a study to identify the common search queries of users searching for clinical trial information[10]. A key aspect of promoting the use of these search tools is developing refinements that increase user acceptance and adoption of these tools. Since increasing usability is a critical aspect of user acceptance[11], we felt that the lack of understanding of both users and systems in this environment could lead to inadequate designs. Motivated to bridge this important knowledge gap, we conducted this study of a few representative clinical trial search engines to understand how users search for clinical trials and interpret the information presented to them by these search engines. The Institutional Review Board (IRB) at Columbia University Medical Center approved this study.

Materials and Methods

1. Selected Clinical Trial Search Engines

We compared eTACTS (the system developed at Weng's lab at Columbia) with four other representative clinical trial search engine sites: ClinicalTrials.gov, Corengi, Dory offered by TrialX, and PatientsLikeMe. **Table 1** summarizes the key functionalities of each. ClinicalTrials.gov requires a user to complete a simple or advanced form to specify search queries for clinical trials. Corengi requires a user to create a personal profile to allow automated matching of trials to user characteristics. PatientsLikeMe presents to users pre-selected demographic questions to help them filter clinical trials. Dory, an online intelligent agent provided by TrialX, enables interactive search between a user and a human customer service agent. eTACTS uses a dynamic cloud of pre-mined tags to allow a

user to filter clinical trial search results from ClinicalTrials.gov.

2. Study Participants

We recruited 11 participants. We tried to increase the heterogeneity of our sample by recruiting physicians, research coordinators, and medical novices. Our study participants also had varied levels of computer skills. A complete summary of the participants can be found in **Table 2**. Some participants were initially contacted by a third-party research coordinator and referred to our research team. For these participants, our research team then contacted the potential participant for an evaluation session. Other participants were recruited directly by the research team through an email recruitment announcement within our academic department. Internal recruitment was limited to participants who had not previously used eTACTS .

Table 1. Summary of The Representative Clinical Trial Search Engines (N=5)

System	Key Functionalities
ClinicalTrials.gov	Offers both simple and advanced searches using string-based free-text search
Corengi	Matches patients up to clinical trials based on user-provided profile information
Dory/TrialX	Provides summaries and contact information to users based on question and answer sessions
eTACTS	Provides interactive tag cloud to allow users to select clinical terms to filter clinical trials
PatientsLikeMe	Provides a set of pre-formed search queries to help filter clinical trials

Table 2. Participant Diversity (N=11)

Diversity Dimension	Proportion	Percentage of Sample (n=11)
Male	10	90.9%
Clinicians (MD)	3	27.3%
Database Administrators	3	27.3%
Graduate Students	3	27.3%
Clinical Research Coordinators	2	18.2%
Experienced Users	6	54.5%

3. Scenario-based mixed-methods evaluations

We obtained informed consent to participate in this study from all the participants before the study began. Participants then completed the one-hour long evaluation session and a short debriefing session followed to answer any remaining questions. To evaluate and compare these systems, we asked participants to search for clinical trials and determine whether a mock patient would be eligible for a specific trial. Some tasks required functions unique to a particular search engine (**Table 1**), while others used functions common to all search engines.

We used a mixed-methods evaluation design in order to measure several levels of behavior and usage patterns. In the absence of usage logs, we used time-motion analysis to capture system usage time. We complemented the quantitative time-motion analysis with two qualitative methods to capture user preferences, opinions about the systems, and any other unexpected findings. To allow users to explore each system, we devised a set of tasks for each participant to accomplish. To help give participants a reference, our team supplied the participants with mock-patients diagnosed with diabetes mellitus type II. We chose to use this disease for our mock patient due to its high prevalence in the clinical trial repository and in the population. We created mock patients by combining several eligibility criteria typical of trials studying this disease. The following is an example of a mock patient:

> *You are a 40-year-old Caucasian male (born in March, 1973) with type II diabetes mellitus diagnosed in June 2005. You take the anti-diabetic drug metformin. You have a hemoglobin A1c value of 7.3. You weigh 220 pounds, are 5 feet 10 inches tall, and have a body mass index (BMI) of 32. You do not smoke and you drink socially. You lead a sedentary life and live with your wife. You have mild hypertension, which is medically controlled with an anti-hypertensive diuretic medication. You have no other significant co-morbidities. You see a primary care physician regularly and have commercial health insurance.*

For every participant, we used ClinicalTrials.gov as the baseline search engine to train the participant how to search for clinical trials and how to understand eligibility search criteria. In order to simulate the real-world environments, the participants were all self-trained. After completing the baseline tasks, we selected participants to systematically complete tasks for each system, with a random sequence of systems. In order to collect qualitative

data, we asked participants to verbalize their thoughts while completing the tasks. We instructed users to speak out loud any thoughts, expectations, and surprises about the system. We recorded their voices using a separate audio recorder. To capture the task completion time of each user, we used the TimeCaT tool.[12] TimeCaT allows an investigator to time a participant completing various tasks. Tasks can be defined either *a priori* or *ad hoc*. To define our tasks *a priori*, we completed the scenarios using each system and created an activity diagram for each system. Each step represented in the activity diagram was then converted into a task to be recorded and timed during the time-motion analysis. In this study, we recorded the amount of time required by participants to enter information into the system, to execute the key system functions, and to determine whether the mock patient was eligible for the trials returned by the search engine. **Table 3** lists the tasks observed in this study.

Table 3. Time-Motion Analysis Task List

Task	Description	Category
Typing Information	User enters initial search query information (i.e. diabetes mellitus type II) into the clinical trial search engine.	Preparatory
Answer Questions	User responds to a set of iterative questions (only used for Dory)	Preparatory
Refine Tag Cloud	User reviews and selects tag cloud options (only used for eTACTS)	Preparatory
Entering Profile	User enters information required to establish medical profile (only used for Corengi and PatientsLikeMe)	Preparatory
System Interactions	Time required by clinical trial search engine to process information and return response, or time spent by user on navigating the interface is also included.	Interaction
Result Review	User reviews the returned list of clinical trials. Participants may determine mock patient eligibility at this stage.	Review
Trial Review	User reviews a single clinical trial to determine whether mock patient would qualify for the study.	Review

Additionally, we used a modification of a survey by Zheng et al. to measure the user perceptions of each system[13].This survey is based on the Unified Theory of Acceptance and Use of Technology[13]. In order to encourage participants to compare the systems with each other, the majority of the survey questions asked participants to rank the systems by a set of system features preferred by users. Participants took the survey immediately after completing the scenario-based task evaluation portion of the study.

4. Statistical Analysis

In order to test for differences between groups, we used a battery of non-parametric test. To calculate the time usage difference between various user groups (i.e. clinicians and non-clinicians), we used the Mann-Whitney U test[14]. To analyze difference between time spent in each task category, we used the Kruskal Wallis test[14] and the Bonferroni correction for post-hoc analysis[15]. We repeated the same statistical method to calculate the differences between survey score results.

1. Time-Motion Analysis

As shown in **Table 4**, after averaging all user time-motion data for each system, we found that the test users required the least time to interact with TrialX (avg=7.52 mins) and the most time to interact with eTACTS (avg=9.91).

When we separated user groups by degree of medical knowledge, in this case physicians versus non-physicians, physicians were not significantly faster than non-physicians, except in the case of eTACTS ($p=0.048$). As shown in Column D in **Table 4**, we found that physicians spent significantly less time completing the required tasks using eTACTS when compared to their counterparts.

We also compared those with experience searching for clinical trials with those having no such experience, as shown in Column G in **Table 4**. eTACTS and PatientsLikeMe lead to large time usage differences between these two user groups. We observed the biggest difference between experienced and novice user interaction time with the eTACTS system ($\Delta=2.54$).

When we aggregated all system usage times together, we found no statistical difference between the two sets of user groups (physicians versus non-physicians, experienced user vs. non- user).

Table 4. Average time spent by user groups per system (A: average time spent; B: average physician time spent; C: average non-physician time spent; D: difference between B and C; E: average experienced user time spent; F: average novice user time spent; G: difference between E and F. All measures are in minutes; 🡻 indicates the ranking column.

System	A (🡻)	B	C	D=C-B	p-value (D)	E	F	G=E-F	p-value (G)
Dory/TrialX.com	7.52	7.27	7.62	0.35	0.776	8.02	6.93	1.09	0.429
Corengi	8.19	8.05	8.22	0.22	1.00	7.86	8.51	-0.65	0.310
ClinicalTrials.gov	8.44	7.74	8.71	0.97	0.63	7.30	9.82	-2.52	0.247
PatientsLikeMe	9.78	7.99	10.45	2.46	0.776	8.22	11.65	-3.43	0.126
eTACTS	9.91	5.33	11.62	**6.29**	*** 0.048**	11.06	8.52	2.54	0.792
Average	**8.77**	**7.28**	**9.32**	**2.06**	**0.259**	**8.49**	**8.63**	**0.14**	**0.366**

Table 5. Average time spent per user per task group by system (all measures are in minutes;🡻 indicates the ranking column).

System	Interaction	Preparatory	Review	Other[#]	Total (🡻)
Dory/TrialX.com	0.39	2.83	3.97	0.33	7.52
Corengi	1.23	2.84	3.81	0.31	8.19
ClinicalTrials.gov	0.42	0.39	6.79	0.84	8.44
PatientsLikeMe	2.42	1.22	5.46	0.68	9.78
eTACTS	0.45	2.29	6.24	0.93	9.91
Average	**0.98**	**1.91**	**5.25**	**0.62**	

[#]Other represents tasks not originally intended to be recorded, such as soliciting for help or asking for clarification.

We found that PatientsLikeMe required more time devoted to site navigation than other systems. The Kruskal Wallis test was significant for differences in time required for interaction among different systems with post-hoc comparisons indicating that PatientsLikeMe and Corengi are significantly more time consuming than ClinicalTrials.gov ($p = 0.022$ and $p = 0.001$ respectively, **Table 5 Column 1**). Furthermore, Corengi was significantly more time intensive than eTACTS ($p = 0.005$, **Table 5 Column 1**).

Simple searches on ClinicalTrials.gov required the least overheard commitment from users before arriving at results, but this difference was not statistically significant ($p = 0.158$, **Table 5 Column 2**).

When breaking down the time spent by users per task, we can easily identify reviewing eligibility status as the most time-consuming task. ClinicalTrials.gov and eTACTS require the most time for users to determine their eligibility for each trial. However, the time spent on each system to review eligibility status was not significantly different ($p = 0.492$, **Table 5 Column 3**).

2. Think-aloud Protocol

Our audio recordings revealed a number of observations about the systems and clinical trials in general. One of the chief complaints that participants voiced was the ambiguity and/or complexity of a clinical trial's eligibility criteria. For instance, one clinical trial described two apparently contradicting criteria.

"*What does this criterion mean? How can you simultaneously have never taken a drug and have taken it for more than a month? There is an 'or' in this criterion! That was very unclear from a quick glance*." – Participant 1.

Some participants had difficulty understanding that the criteria called for participants having never taken medications before or having taken a stable dosage for more than 4 weeks. For these participants, the meaning of the criteria was lost in the long and highly specific wording.

With regard to specific systems, the voice recording demonstrates that personal preferences play a role in a participant's perception. For the Corengi system, two of the participants felt that completing the medical profile form online was a simple but time-consuming process. This reduced the value of using a medical profile to filter clinical trials for participants. Six participants felt that the Dory interface by TrialX was easy to use; however, an equal number felt that the information presented by Dory was inadequate for assessing clinical trial eligibility. This lack of information was dissatisfying for some participants:

"[*Dory] doesn't give me a lot of information. […] I would really need more information in order to determine whether the patient qualifies or not. […] This is really for – It's not for researchers to use… for people to find out what's out there.*" – Participant 6.

Additionally, the rigid diagnosis vocabulary and strict diagnosis autocomplete function used by the Dory to identify the user's condition caused some participants to conduct searches that returned no results. For the eTACTS system, participants were split in their perceptions toward the tag cloud feature. While some found this feature useful for quickly filtering the possible clinical trials, others found it confusing:

"*The first few tags were easy to find, but afterwards, I became more unsure of which tags applied to me.*" – Participant 3.

One physician was unable to intuitively use the tag cloud feature, while the remaining clinicians found it difficult to understand some of the tags (e.g. "blood pressure") in the context of clinical trials. Participants in the study often found the PatientsLikeMe system very polished. Unlike the autocomplete function in Dory, all participants who commented on the system performance were impressed with the speed of the autocomplete function on PatientsLikeMe. Also, one participant noticed that the trials recommended by PatientsLikeMe differed from those recommended by ClinicalTrials.gov.

3. Survey

The post-scenario-based evaluation survey also presented some interesting results, which can be found in Table 6. Overall, eTACTS was consistently the favored search engine system. It received the best rating in four out of the five aspects surveyed. PatientsLikeMe was deemed to provide the most guided search. We found no statistical differences when comparing the various engines based on each of the survey questions. Since none of the Kruskal Wallis tests were significant, we did not use a post-hoc test.

Table 6. Average rating for search engines, which are ordered from left to right by ease of use using a 5-scale Likert survey (1: most preferred; 5: least preferred, A = eTACTS, B = Dory/TrialX, C = ClinicalTrials.gov, D = PatientsLikeMe, E = Corengi)

Aspect	A	B	C	D	E	P-value
Ease of entering information	2.42	3.58	2.92	3.00	4.00	0.172
Provided most search guidance	3.00	2.75	3.92	2.50	3.45	0.166
Ease of site navigation	1.75	2.83	2.17	3.17	3.36	0.173
Ease of use with no prior	2.08	2.75	2.92	3.33	3.18	0.351
Overall ease of use	*** 2.17**	**2.50**	**2.67**	**2.75**	**3.18**	**0.665**

Discussion

In our study, we applied mixed methods to evaluate five clinical trial search engines. Our results identified two important implications for designing clinical trial search engines. While eTACTS scored well on the survey, there was much variation in time spent using eTACTS in both of our user group comparisons. We found that eTACTS was not unique in this aspect; others such as PatientsLikeMe also required varied time among users. In the aggregate, eTACTS was one of the most time-consuming search engines. There are at least two possible explanations for this finding.

The first explanation is that time required for a user to find trials is not necessarily related to that user's satisfaction with the system. This suggests that other factors, such as those we measured in the survey, may also influence user satisfaction. Research into the usability of other information retrieval systems (IR) also found that aspects such as clarity of the interface design and ease of learning the interface impact the usability and user experience of the system[16]. In our think aloud protocol, we found that systems that required more attention and time spent navigating were often not favored and garnered more complaints. The findings in this study reinforce conclusions from the literature that developers should consider multiple user measures when evaluating the usability of search engine interface designs.

The second explanation is that individual variability influences the usability of these systems. A recent focus in information science has been the impact of users' cognitive processes and usability on web searching behavior[17]. In the case of eTACTS, we found that participants who favored the system the most were those who had extensive medical knowledge and an understanding of how the tag cloud feature worked. Individual variability was a factor in other systems as well. Kinley et al. found that individual cognitive styles (imager, verbalizer, etc.) could impact

which of the three information-processing approaches (scanning, reading, and mixed) was favored by an individual[18]. In our own study, participants daunted by the complexity of clinical trial eligibility criteria found Dory/TrialX.com appealing; while those who wanted to see the exact criteria found it frustrating. Kim et al. identified problem-solving style as another contributor to how users search for information[18]. Future work in this field should focus on identifying user groups for these systems and then identifying each group's design and functionality preferences. These types of findings would give insights to clinical trial search engine design while addressing multiple factors of technology acceptance and use.

One limitation of our study lies in our sample. It is both small (n = 11) and selective. Our participants all had a bachelor's degree (or higher) and are familiar with using Internet to seek health information; therefore, their behaviors and cognitive styles may not reflect that of the greater population, especially those who are older and are not familiar with using the Internet for information seeking. In spite of these potential limitations, we believe the types of participants we sampled are those who would most often use this type of resources. The main purpose of this study was to demonstrate the potential for greater emphasis of user design based on multiple perspectives. Further studies are needed to test how the results of our comparisons of clinicians and experienced trial searchers and their respective novice counterparts may generalize to the general population.

Conclusions

We presented a mixed-method approach to understand user interactions with five representative clinical trial search engines. We found that (1) user groups exhibit different behaviors when searching for clinical trials; (2) clinicians preferred certain system features to others; and (3) individual cognitive styles and characteristics affect user behaviors and usage patterns. Our study contributed empirical evidence that differences among users can influence user search behaviors and hence should be considered during the design of clinical trial search engines.

Acknowledgments

The research described was supported by grants **R01LM009886**, **R01LM010815**, and **T15LM007079** from the National Library of Medicine, and grant **UL1 TR000040** from the National Center for Advancing Translational Sciences (NCATS).

References

1 Institute of Medicine (US). *Envisioning a Transformed Clinical Trials Enterprise in the United States: Establishing An Agenda for 2020: Workshop Summary*. Washington (DC): : National Academies Press (US) 2012.

2 Dear R, Barratt A, Askie L, *et al.* Adding value to clinical trial registries: insights from Australian Cancer Trials Online, a website for consumers. *Clinical trials (London, England)* 2011;**8**:70–6. doi:10.1177/1740774510392392

3 *clinicaltrials.gov*. http://www.clinicaltrials.gov (accessed 29 Sep2013).

4 Atkinson NL, Saperstein SL, Massett HA, *et al.* Using the Internet to search for cancer clinical trials: A comparative audit of clinical trial search tools. *Contemporary Clinical Trials* 2008;**29**:555–64. doi:10.1016/j.cct.2008.01.007

5 *Corengi.com*. https://www.corengi.com (accessed 29 Sep2013).

6 *Clinicl Trials Search Service -- Powered by Ask Dory*. http://dory.trialx.com/ask/ (accessed 29 Sep2013).

7 Miotto R, Jiang S, Weng C. eTACTS: A Method for Dynamically Filtering Clinical Trial Search Results. *Journal of biomedical informatics* 2013;:1–8. doi:10.1016/j.jbi.2013.07.014

8 *patientslikeme.com*. http://www.patientslikeme.com (accessed 29 Sep2013).

9 *trialreach.com*. http://trialreach.com (accessed 9 Jan2014).

10 Patel CO, Garg V, Khan SA. What do patients search for when seeking clinical trial information online? *AMIA*

Annu Symp Proc 2010;**2010**:597–601.

11 Beyer H. *Contextual Design: Defining Customer-Centered Systems*. 1st ed. San Francisco: : Morgan Kaufmann 1997.

12 Lopetegui M, Yen P-Y, Lai AM, *et al.* Time Capture Tool (TimeCaT): development of a comprehensive application to support data capture for Time Motion Studies. *AMIA Annu Symp Proc* 2012;**2012**:596–605.

13 Zheng K, Padman R, Johnson MP, *et al.* An interface-driven analysis of user interactions with an electronic health records system. *J Am Med Inform Assoc* 2009;**16**:228–37.

14 Choosing the Correct Statistical Test in SAS, Stata and SPSS. ats.ucla.edu. http://www.ats.ucla.edu/stat/mult_pkg/whatstat/ (accessed 9 Jan2014).

15 Weisstein EW. Bonferroni Correction -- from Wolfram MathWorld. *mathworldwolframcom*

16 Zhang Y. Searching for specific health-related information in MedlinePlus: Behavioral patterns and User Experience. *J Am Soc Inf Sci Tec* 2013;:n/a–n/a. doi:10.1002/asi.22957

17 Kinley K, Tjondronegoro D, Partridge H. Web searching interaction model based on user cognitive styles. 2010;:340–3.

18 Kim KS, Allen B. Cognitive and task influences on Web searching behavior. *J Am Soc Inf Sci Tec* 2002;**53**:109–19.