Usability of Selected Databases for Low-Resource Clinical Decision Support

L. Sheets[1,2]; F. Callaghan[2]; A. Gavino[2]; F. Liu[2]; P. Fontelo[2]
[1]University of Missouri Informatics Institute; [2]National Library of Medicine

Keywords
Library information systems, handheld devices, ambulatory care, general healthcare providers, economic barriers

Summary
Background: Smartphones are increasingly important for clinical decision support, but smartphone and Internet use are limited by cost or coverage in many settings. txt2MEDLINE provides access to published medical evidence by text messaging. Previous studies have evaluated this approach, but we found no comparisons with other tools in this format.
Objectives: To compare txt2MEDLINE with other databases for answering clinical queries by text messaging in low-resource settings.
Methods: Using varied formats, we searched txt2MEDLINE and five other search portals (askMEDLINE, Cochrane, DynaMed, PubMed PICO, and UpToDate) to develop optimal strategies for each. We then searched each database again with five benchmark queries, using the customized search-optimization formats. We truncated the results to less than 480 characters each to simulate delivering them to a maximum of three text messages. Clinicians with practice experience in low-resource areas scored the results on a 5-point Likert scale.
Results: Median scores and standard deviations from 17 reviewers were: txt2MEDLINE, 3.2±0.82 (control); askMEDLINE, 3.2±0.90 ($p = 0.918$); Cochrane, 3.8±0.58 ($p = 0.073$); DynaMed, 3.6±0.65 ($p = 0.105$); PubMed PICO, 3.6±0.82 ($p = 0.005$); and UpToDate, 4.0±0.52 ($p = 0.002$). Our sample size was sufficiently powered to find differences of 1.0 point.
Conclusions: Comparing several possible sources for texting-based clinical-decision-support information, our results did not demonstrate one-point differences in usefulness on a scale of 1 to 5. PubMed PICO and UpToDate were significantly better than txt2MEDLINE, but with relatively small improvements in Likert score (0.4 and 0.8, respectively). In a texting-only setting, txt2MEDLINE is comparable to simulated alternatives based on established reference sources.

Correspondence to
Lincoln Sheets, MD
Missouri University Informatics Institute
EBW 241
Columbia MO 65211
USA
E-mail: lincolnsheets@gmail.com

Appl Clin Inf 2012; 3: 326–333
doi:10.4338/ACI-2012-06-RA-0024
received: June 18, 2012
accepted: August 27, 2012
published: September 6, 2012
Citation: Sheets L, Callaghan F, Gavino A, Liu F, Fontelo P. Usability of selected databases for low-resource clinical decision support. Appl Clin Inf 2012; 3: 326–333
http://dx.doi.org/10.4338/ACI-06-RA-0024

1. Background

Mobile phones will exceed the world's population by the end of 2012 [1]. Smartphones are increasingly important for clinical decision support, but smartphone and Internet use are limited by cost or coverage in many settings [2, 3]. The US National Library of Medicine developed *txt*2MEDLINE precisely to fill that gap by providing access to high-quality medical evidence via text messaging, also known as short messaging service or SMS [4]. In developing countries and other low-resource areas, this communication channel has the great advantages of being widely available and relatively affordable.

While many clinical decision support tools have been developed and are in development, *txt*2MEDLINE is the only one to our knowledge that uses a text-messaging interface for searching PubMed MEDLINE. Like any information portal, the clinical relevance of the results can depend on many factors including the quality of the query [5]. Previous studies of *txt*2MEDLINE have also shown strengths and limitations specific to this approach [6, 7]. However, we found no comparisons of *txt*2MEDLINE with other tools for clinical decision support.

PubMed MEDLINE searches using desktop, laptop, or tablet computers allow convenient reading and navigation of search results from multiple web pages; however, the number of search results that can be retrieved via text messaging is limited. Furthermore, receiving several text messages, sometimes in the wrong order, may become unmanageable and difficult to comprehend. *txt*2MEDLINE was optimized so the number of text messages sent is minimized. In this study, we set the target number to three text messages to reduce the response times and messaging costs of the proposed solutions.

2. Objectives

Our objective was to compare *txt*2MEDLINE with other possible knowledge sources and potential alternatives for answering clinical queries by text messaging using the first citation or source retrieved. Our hypothesis was that optimal search strategies for MEDLINE would produce results that were as useful as those produced by optimal search strategies for other databases.

3. Methods

We compared *txt*2MEDLINE [8] to two PubMed search portals designed for smartphones: askMEDLINE [9] and PICO [10]; and to three online medical databases with comprehensive contents and well-structured formats, which supported parsing of results: Cochrane [11], DynaMed [12], and UpToDate [13]. We considered other online resources, including eMedicine, FamilyDoctor.org, Google Text, MedlinePlus, MedSocket, the National Guideline Clearinghouse, the U.S. Preventive Services Task Force website, WebMD, and online publications from the World Health Organization; but their formats did not fit the information delivery method selected for this study.

For each of the six selected search portals, we tried a wide variety of filters and formats to find which of these consistently returned results that were clinically meaningful. This led to a slightly different search strategy for each database to optimize the results from each resource, as shown in ▶Figure 1. We tested the customized search-optimization strategies for each portal using four benchmark queries derived from recent PubMed searches and an additional query on ascariasis, the most world's prevalent neglected tropical disease [15]; ▶Figure 2.

In all cases, we truncated the results to less than 480 characters to simulate delivering them to a maximum of three text messages (160 characters maximum per message). We masked the sources and alternated the order of results to avoid bias. Seventeen clinicians with practice experience in low-resource areas (13 licensed physicians, one senior medical student with medical missions experience, and three nurses) scored the results from 1 (not useful) to 5 (very useful). We excluded any incomplete evaluation questionnaires.

Since the Likert-scale responses could not be assumed to be normal, we considered the median for each search portal rather than its mean, and looked for statistical significance with a non-paramet-

ric Wilcoxon test [16, 17]. Because our primary hypothesis was that txt2MEDLINE was *not* significantly different to other methods, we considered a measure of clinical importance in addition to statistical significance. In order to make a non-inferiority claim, we ensured that our study was sufficiently powered to detect a clinically important difference. We are not aware of any consensus on what constitutes clinical importance for a Likert scale in this area. In the interests of simplicity and interpretability, we chose to test for an increase in median score of one point on a scale of 1 to 5. In order to estimate reliability, we calculated an intraclass correlation coefficient to measure the amount of correlation among the raters, treating them as a representative sample from a larger pool of potential raters [18, 19]. All statistical computations were performed using the R statistical software package [20].

4. Results

▶Table 1 and ▶Figure 3 show the medians from the reviewers' average scores (with 1 = not useful to 5 = very useful), and some of their statistical characteristics. PubMed PICO and UpToDate were significantly better than txt2MEDLINE, but with relatively small improvements in Likert score (0.4 and 0.8, respectively). The other methods were not significantly different from txt2MEDLINE, and our study was powered to detect a difference of 1.0 point between the scores if any existed (power = 96%). The estimated average intraclass correlation of the 17 raters across all queries and search portals was high (0.79, with a 95% confidence interval of 0.66 – 0.88).

5. Conclusions

The main aim of this study was to evaluate optimized search algorithms for *txt*2MEDLINE, the US National Library of Medicine's texting-based search portal, in order to deliver the most informative and clinically useful content through the least number of text messages. Clinicians and researchers practicing in developing countries or remote settings with adequate wireless mobile networks, but with relatively meager health information technology resources and limited Internet connections, are the potential beneficiaries of this research. This is in concordance with the original purpose of the *txt*2MEDLINE portal [4]. However, since *txt*2MEDLINE and the other MEDLINE search portals return multiple results in chronological order (most recent first), more pertinent and clinically informative citations could be listed further down the list and missed. The algorithms were intended to present this information to the clinician.

We tested the hypothesis that optimized search results from *txt*2MEDLINE would not be less useful than the optimized search results from the best alternatives. Both UpToDate and PubMed PICO had significantly better average scores than *txt*2MEDLINE: PubMed PICO had an average improvement of 0.4 of a point, and UpToDate had an improvement of 0.8 of a point on the Likert scale compared to *txt*2MEDLINE. This improvement must also be balanced against the cost of subscription, although costs may be waived in some cases [21]. The other methods were not significantly different from *txt*2MEDLINE. All the differences were less than our predefined endpoint of 1.0 on a scale of 1 to 5, and due to the high level of power that we had to detect this difference, we have strong evidence to conclude that a difference of more than 1 point does not exist between these methods.

We assessed the reliability of the Likert scale using the intraclass correlation coefficient, but we are unable to comment on the validity of using a Likert scale to measure the relevance of search results. There is very limited literature concerning how to measure a qualitative assessment of medical search results, and no validated tools available. We chose the Likert scale because our clinicians were already familiar with it, it is relatively easy to interpret, and it allows for more nuanced ranking of the search results compared to a simple binary (yes/no) scale. A valuable avenue for future research would be to develop a validated and reliable tool for assessing query results in the clinical setting.

Our study was designed to elicit, from each of the tested search portals, the best possible query responses in the context of point-of-care decision support in low-resource clinical settings. Because many evidence-based medicine databases delivered via the Web are edited by subject in a textbook-like or encyclopedia-like format, we expected some of their results to be more treatment-focused and

useful in general clinical contexts. In many cases, that advantage may be offset by the lesser suitability of Web-based narrative text for summarization in the extremely limited format of text messaging. Our findings may indicate that the relative strengths and weaknesses of these alternative approaches are less significant than the quality of the search terms entered, which ultimately depends on the information-searching skills of the clinician using the search portal.

We requested feedback only from physicians and nurses with low-resource clinical experience in order to accurately determine the usefulness of the query results, which meant we sampled from a relatively small population. Nevertheless, our study had a high degree of power to detect difference between *txt*2MEDLINE and competing methods. In areas with limited Internet connectivity, *txt*2MEDLINE provides access to medical evidence that compares favorably with the best results that could be simulated using commercial alternatives.

Further study is needed with clinicians in remote and resource-poor locations. These studies may add support to the feasibility of improving clinical decision support via text messaging. Improving this technology could improve clinical information resources in the world's most medically underserved communities.

Clinical Relevance Statement

In areas with limited Internet connectivity, *txt*2MEDLINE can provide useful medical evidence via text messaging.

Conflicts of Interest

The authors declare that they have no conflicts of interest in the research.

Protection of Human and Animal Subjects

Human and/or animal subjects were not included in the project.

Acknowledgements

This research was supported in part by an appointment to the NLM Research Participation Program. This program is administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the US Department of Energy and the National Library of Medicine. This research was also supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill Center for Biomedical Communications (LHNCBC).

- askMEDLINE
  - append "effective treatment systematic review" to the query string
  - exclude results with "new" or "contemporary" in the title
- Cochrane
  - append "treatment" to the query string
  - exclude results with no abstract
  - search Library; if no results, search Summaries; if no results, search Reviews
  - summarize the abstract using the TBL* algorithm
- DynaMed
  - exclude results with no "Treatment" section
  - extract from beginning of "Treatment Overview" subsection if present,
    - else extract from beginning of "Recommendations" subsection if present,
    - else extract from beginning of "Treatment" section
  - remove parenthetical phrases
  - reformat tabbed or bullet-pointed lists as comma-separated lists
  - truncate at the last whole word before 480 characters
- PICO
  - append "uncomplicated effective treatment" to the query string
  - search for Publication Type = "Systematic Reviews"; if no results, search all
  - if no results, remove "uncomplicated" from the query string
  - summarize the abstract using the TBL* algorithm
- *txt*2MEDLINE
  - append "uncomplicated effective treatment" to the query string
  - if no results, remove "uncomplicated" from the query string
  - summarize the abstract using the TBL* algorithm
- UpToDate
  - append "management" to the query string
  - exclude results with no "Summary" or "Summary and Recommendations" section
  - extract all sentences containing "…we…recommend/suggest/advise…"
  - extract all whole sentences within the last 480 characters
  - remove parenthetical phrases
  - reformat tabbed or bullet-pointed lists as comma-separated lists
  - truncate at the last whole word before 480 characters

Fig. 1 Search optimization strategies *TBL = "The Bottom Line" computer generated summary [14]

"analgesics in acute abdominal pain for diagnosis of acute appendicitis"
"using warfarin in atrial fibrillation in patient over 75 years old"
"immediate antibiotics versus wait in acute otitis media"
"is intravenous immune globulin effective in Stevens-Johnson Syndrome in children"
"ascariasis"

Fig. 2 Benchmark queries

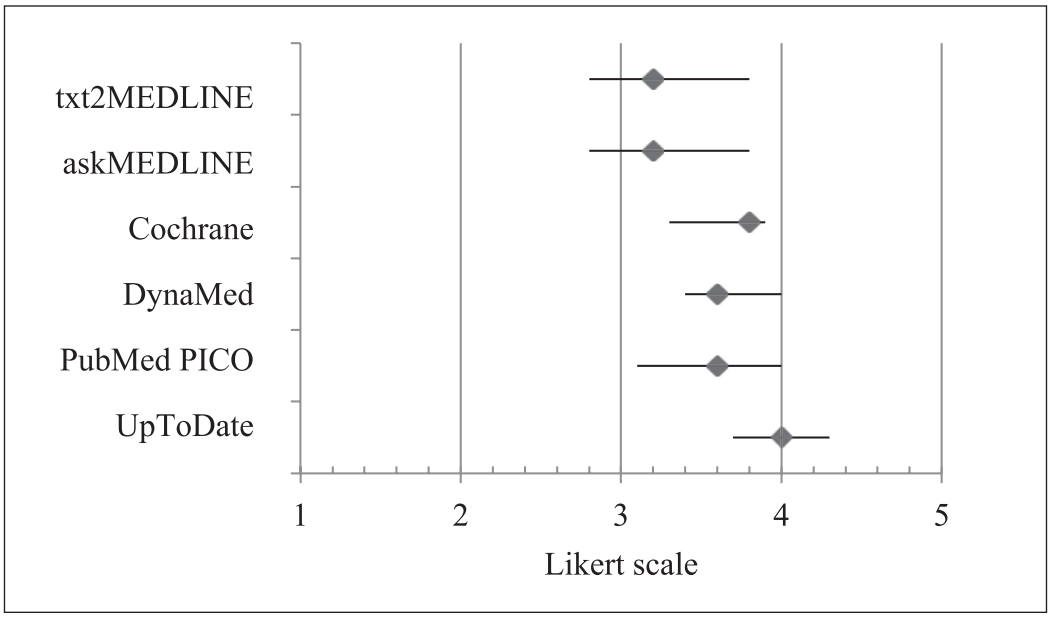

Fig. 3 Median scores and confidence intervals

Table 1 Median scores, standard deviations, and significance measures

Search portal	Median score ± standard deviation	Wilcoxon p-value
txt2MEDLINE	3.2 ± 0.82	[control]
askMEDLINE	3.2 ±0.90	0.918
Cochrane	3.8 ±0.58	0.073
DynaMed	3.6 ±0.65	0.105
PubMed PICO	3.6 ±0.82	0.005
UpToDate	4.0 ±0.52	0.002

References

1. Cisco Systems. Global Mobile Data Traffic Forecast Update, 2011–2016. San Jose: Cisco Systems; 2012. 29 p.
2. Lorence DP, Park H, Fox S. Racial disparities in health information access: resilience of the Digital Divide. J Med Syst 2006; 30(4): 241–249.
3. Economist Newspaper Ltd. A special report on telecoms in emerging markets. The Economist. 2009 Sep 24; 3–19.
4. Fontelo P, Liu F, Muin M, Tolentino T. txt2MEDLINE: text-messaging access to MEDLINE/PubMed. AMIA Annual Symposium Proceedings 2006: 259–263.
5. Ely JW, Osheroff JA, Chambliss ML, Ebell MH, Rosenbaum ME. Answering physicians' clinical questions: obstacles and potential solutions. J Am Med Inform Assoc 2005; 12(2): 217–224.
6. Marcelo AB, Fontelo P. An Evaluation of the Usability of Accessing MEDLINE over Short Messaging System. MEDINFO 2007: 235–236.
7. Armstrong K, Liu F, Seymour A, Mazhani L, Littman-Quinn R, Fontelo P, et al. Evaluation of *txt*2MEDLINE and Development of Short Messaging Service-Optimized, Clinical Practice Guidelines in Botswana. Telemed J E Health 2012; 18(1): 1–4.
8. *txt*2MEDLINE: search by SMS [Internet]. Bethesda, MD: US National Library of Medicine; c2005–06 [updated c2005–06; cited 2012 Mar 1–20]. Available from: http://sms.nlm.nih.gov/tbl/txt.php.
9. askMEDLINE [Internet]. Bethesda, MD: US National Library of Medicine; c2005–06 [updated c2005–06; cited 2012 Mar 1–20]. Available from: http://askmedline.nlm.nih.gov/ask/ask.php.
10. Search MEDLINE/PubMed via PICO [Internet]. Bethesda, MD: US National Library of Medicine; c2005–06 [updated c2005–06; cited 2012 Mar 1–20]. Available from: http://pubmedhh.nlm.nih.gov/nlmd/pico/piconew.php.
11. The Cochrane Library [Internet]. New York: John Wiley & Sons; c1999 [updated 2012 Mar 1; cited 2012 Mar 1–20]. Available from: http://www.thecochranelibrary.com.
12. DynaMed [Internet]. Ipswich, MA: EBSCO Publishing; c1995 [updated 2012 Mar 1; cited 2012 Mar 1–20]. Available from: http://www.ebscohost.com/DynaMed.
13. UpToDate [Internet]. Alphen aan den Rijn: Wolters Kluwer; c1992 [updated 2012 Mar 1; cited 2012 Mar 1–20]. Available from: https://www.uptodate.com.
14. Tom O, Fontelo P, Liu F. Do computer-generated summaries, „The Bottom Line (TBL)" accurately reflect published journal abstracts? AMIA Annu Symp Proc Oct 2007; 11: 1135.
15. Reddy M, Gill SS, Kalkar SR, Wu W, Anderson PJ, Rochon PA. Oral Drug Therapy for Multiple Neglected Tropical Diseases: A Systematic Review. JAMA 2007; 298(16): 1911–1924.
16. Bauer, DF. Constructing confidence sets using rank statistics. J Am Stat Assoc 1972; 67: 687–690.
17. Hollander M, Wolfe DA. Nonparametric Statistical Methods. New York: John Wiley & Sons; 1973. p. 27–33, 68–75.
18. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychological Methods 1996; 1: 30–46.
19. Shrout PE, Fleiss JL. Intraclass correlation: uses in assessing rater reliability. Psychological Bulletin 1979; 86: 420–428.
20. R Development Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2011.
21. UpToDate International Grant Subscription Program [Internet]. Boston: Global Health Delivery Project; c2010 [updated c2010; cited 2012 May 6]. Available from: http://www.ghdonline.org/uptodategrant.