THE LISTER HILL NATIONAL CENTER
FOR BIOMEDICAL COMMUNICATIONS

A research division of the U.S. National Library of Medicine

TECHNICAL REPORT
LHNCBC-TR-2008-003

Bringing the Medical Library Where Needed - At the Point of Care

September 2008

Paul Fontelo, M.D., M.P.H.
Fang Liu, M.S.

U.S. National Library of Medicine, LHNCBC
8600 Rockville Pike, Building 38A
Bethesda, MD 20894

Table of Contents

1. Background ... 4
2. Project Objectives ... 6
3. Project Significance ... 7
4. Evidence-Based Medicine Tools ... 7
 4.1 PubMed for Handhelds ... 7
 4.2 Features of PubMed for Handhelds (PMHh) ... 11
 4.2.1 PICO (Patient, Intervention, Comparison, Outcome) - http://pubmedhh.nlm.nih.gov/nlm/pico/piconew.html ... 11
 Current status and Implementation ... 12
 Server analysis ... 14
 Future plans ... 15
 4.2.2 askMEDLINE – http://askmedline.nlm.nih.gov ... 15
 Implementation ... 16
 Server access analysis ... 17
 Future plans ... 19
 4.2.3 Disease Associations ... 19
 Server access ... 20
 Future plans ... 20
 4.2.4 MEDLINE/PubMed Search with Systematic Review and Clinical Queries filters and Journal Browser. ... 21
 Server Access ... 22
 Future Plans ... 22
 4.3 BabelMeSH and PICO Linguist - http://babelmesh.nlm.nih.gov ... 23
 Current usage ... 25
 Future plans ... 27
 4.4 txt2MEDLINE ... 28
 Future plans ... 30
5. Summing up and looking ahead ... 31
6. References ... 31
7. Questions for the Board of Scientific Counselors ... 34
8. Project Team Resumes and Collaborators ... 35
9. Collaborators ... 40
 a. Evaluation of PICO versions ... 40
 b. BabelMeSH and PICO Linguist ... 40

c. askMEDLINE, txt2MEDLINE and SMS projects ... 41
10. Appendix ... 42
10.1 Appendix A – Patient, Intervention, Comparison and Outcome (PICO) ... 42
Results ... 45
References ... 46
10. 2 Appendix B - askMEDLINE ... 46
Evaluation ... 47
Evaluation of search retrievals ... 48
References ... 49
10.3 Appendix C - BabelMeSH ... 50
Implementation ... 50
Multi-language Database ... 50
Character-based language processing ... 51
Parsing algorithm ... 51
User Interface ... 52
Citation Result Filter ... 54
Evaluation ... 54
Results ... 55
Discussion ... 57
Challenges and developments ... 58
Conclusion ... 58
References ... 59
10.4 Appendix D – txt2MEDLINE ... 60
txt2MEDLINE Architecture ... 60
Abbreviation and Acronym Database ... 60
Incoming Message Processing ... 61
Query Processing ... 62
Outgoing Message Processing Algorithm ... 62
The TBL Algorithm: ... 62
Word transformation Algorithm: ... 62
Other Optimization Procedures ... 63
Registration Service ... 63

Message Truncation Evaluation.. 63
Results.. 63
Discussion.. 65
"The Bottom Line" (TBL) summary 66
Evaluation 67
Results... 67
Conclusion... 68
References 68

1. Background

The Clinical Research Roundtable was formed in response to concerns that despite great expenditures by the National Institutes of Health and other funding organizations, advances in basic science and clinical research had not resulted in the delivery of better healthcare and formulation of national health policy. The group, which had convened for three years at the National Research Council, reported their finding in the Journal of the American Medical Association [1]. They identified two major obstacles or “translational blocks”: 1) impedance to the translation of basic science discoveries into clinical studies, 2) a block in bringing the clinical studies into medical practice and health decision making. Reference was made to evidence-based decision support to clinicians as part of the recommendations. The NIH Roadmap includes accelerating the bench to bedside transition as one of its main objectives [2].

Evidence-based medicine (EBM) is the conscientious, explicit and judicious use of current best evidence in making decisions regarding the care of individual patients. It incorporates clinical knowledge and experience with clinically oriented research in making decisions regarding a patient's diagnosis, treatment and management. The purpose of EBM is to provide the patient with the safest and most efficacious care. It is centered on the patient; it involves asking well-focused questions, searching for the best available evidence, appraising that evidence for validity, and then applying the results to the care of the patient.

It is then essential that the clinician be equipped with resources to access information repositories and translate these to useful knowledge for decision support at the point-of-care. As Sackett and Straus [3] demonstrated, allowing doctors easy access to evidence-based resources while making rounds increased the extent to which evidence was sought and incorporated into patient care decisions.

Clinicians currently utilize a variety of resources and interfaces to search the biomedical literature to answer clinical questions. The information obtained from literature searches such as MEDLINE/PubMed can have a significant impact on patient care and clinical outcomes. Crowley et al reported on a study of 625 clinical questions asked by residents during an in-hospital general medicine rotation. Seventy-seven percent of the answers to these questions came from MEDLINE and the information from the articles changed patient management 47% of the time [4]. Sackett and Straus reported that of the 98 searches, 79 (81%) sought evidence that could affect diagnostic and/or treatment decisions [3]. Klein et al showed that conducting a MEDLINE search early in the hospitalization of a patient could significantly lower costs, charges, and lengths of stay [5]. Westbrook et al reported that the use of an online information retrieval system improved the quality of clinicians' answers to clinical questions by 21% [6]. The literature also reports that many clinical questions go unanswered due to difficulties formulating a relevant question [7] forgetting the question [8] lack of access to information resources, and lack of skills in searching [9].

Practicing EBM by searching current evidence is neither easy nor convenient, especially for the independent practicing healthcare provider. If the process of obtaining information is hard, healthcare practitioners may not do it. According to Shaughnessy and Slawson [10,11] and Smith [12], the usefulness of information is related to its relevance, validity, interactivity and ease of obtaining it. The more relevant, valid, interactive, and easier it is to get information, the more likely it will be sought.

$$\text{Utility of Information} = \frac{\text{Relevance} \times \text{Validity} \times \text{Interactivity}}{\text{Work to access}}$$

Access to information brings to mind the role of medical libraries. The traditional paradigm has been to make a trip to the medical library, do your own research and obtain reprints of journal articles of interest. Another method used by many busy clinicians is to request a detailed search from a medical librarian. In the Internet age, access to information has changed remarkably. Information is now available on the Web, accessible to both patient and physician. This has led to Internet savvy patients that seek consultation with clinicians, coming to the clinic loaded with reams of reprints of their disease or condition. Sometimes, all they really want from the doctor is a confirmation of their self-diagnosed ailment and a prescription. Add to this the recent trend of pharmaceutical companies marketing directly to the general public, advising patients to ask their doctor if this is the right medication for them. Faced with this situation a clinician needs quick access to medical information.

Mobile access essential medical information at the point-of-need would not have been possible if not for two fairly recent developments: handheld devices such as personal digital assistants (PDAs) and smartphones, and the increasing ability to connect to the Internet through wireless networks. These enabling technologies have a positive impact on the mobile healthcare worker's ability to access electronic healthcare data and resources, and the practice of evidence-based medicine. With mobile handheld devices and wireless networks (Wi-fi or 802.11 networks and cellular service), the medical library is now mobile, accessible to the practitioner where patient care is given.

The Internet has had a major impact on physicians' practice by improving their access to medical information resources. Clinicians frequently use online evidence primarily to support clinical decisions related to direct patient care [13]. Physicians' use of the Internet and handheld devices like PDAs is steadily growing [14-16] - - 60% to 70% of medical students and residents used PDAs for educational purposes and patient care in 2006 [17].

Mobile phones and smart phones – new hybrid devices that combine the communication capabilities of mobile phones with easy and fast access to the Web and computing features of PDAs and desktop computers – may be an effective way to provide real-time access to online medical knowledge resources at the bedside. The GSM Association (GSMA), whose members provide more than 86% of the world's mobile phones, announced recently that there are now more than three billion connections among member networks [18]. This translates to more than 3.4 billion phones and growing daily. Wireless telecommunications, available in 80% of the world's population in 2006, is expected to reach 90% in 2010 and close to 100% in four years [19]. These figures more than surpass the United Nations' goal of 50% by 2015. In the Philippines alone, two telecommunication companies report that their cellular base stations cover 98% of the population, living in 75% of the geographic area of the country [20]. This is probably true in many parts of Asia and Europe.

Initial development of the "PubMed for Handhelds" project started in late 2001. An early progress report was presented in a live demonstration at 39th meeting the Board of Scientific Counselors (BSC) on May 9, 2002. In their report [21] the BSC recommended, "that there are immediate and obvious opportunities to transition some research prototypes into operational services. The most compelling of these currently would be the creation of a 'web-clipped' search interface to NLM's databases optimized for use by wireless palmtop Personal Digital Assistants (PDAs). Just as NLM's own PubMed interface to its databases has become the most widely used web-based biomedical search tool, a service targeted to compact small screen devices, free of proprietary intermediary technologies and requiring no client software other than a web browser, would literally propel NLM point-of-service clinical decision support information into the pockets of tens of thousands (if not hundreds of thousands) of physicians and other healthcare providers. PDAs have become the 'peripheral brains' of essentially every physician in training in the country, and NLM can ride this wave of popularity to modify information access behaviors of a whole generation of new healthcare professionals if it acts promptly. Health sciences librarians within the NN/LM network are already developing a range of innovative instructional and other services to support PDA access for health professionals. PDAs are also an important new tool for providing the public with access to reliable health information". PubMed for Handhelds was officially announced in the July-Aug 2003 issue of the NLM Technical Bulletin [22] and reported in the Telemedicine Journal and e-Health [23].

2. Project Objectives

The main objective of this project is to develop tools and resources to enable the clinician to access knowledge sources from the NLM, such as MEDLINE/PubMed, in support of the practice of evidence-based medicine (EBM). These resources should be Internet-based, compatible with any Web browser, mostly text formatted, and operating system independent. These portals should be accessible anywhere, using any networked device, from a desktop computer to a wireless handheld device such a mobile phone.

Medical information should be available to any clinician at the point of care, or researchers around the world, anytime. It should be interactive, clinically relevant, and presented as clinical 'the bottom line' (TBL) statements for quick reading so they can be useful in clinical decision-making. These resources should be modular so they can be linked or incorporated into an electronic medical record.

3. Project Significance

The practice of evidence-based medicine (EBM) recommends that physicians obtain clinically useful information by personally searching, reading and critically appraising the medical literature. It incorporates clinical knowledge and experience with clinically oriented research in making decisions regarding a patient's diagnosis, treatment and management. The purpose of EBM is to provide the patient with the safest and most efficacious care.

In evidence-based practice, medical information needed for clinical decision-making must be easily obtainable. When medical information is easily available, it will likely be used in clinical decision-making. In many cases, even if the perceived need of it is high, medical information will not be utilized if it is difficult to obtain. These resources, developed for handheld devices, tablet PCs and desktop computers, will bring the medical library to where care is delivered. With worldwide access to the Internet through wireless networks now so pervasive, and handheld devices such as smartphones quite affordable, PubMed for Handhelds and other search tools will allow any clinician to practice evidence-based medicine, anywhere in the world. With these resources, the National Library of Medicine is bringing the medical library to the point of care, where it can be used effectively.

4. Evidence-Based Medicine Tools

4.1 PubMed for Handhelds

PubMed for Handhelds consists of several search tools. Currently, it features a Clinical Queries search, a Journal Browser, PICO (Patient, Intervention, Comparison, Outcome), askMEDLINE, and Disease Associations. It has evolved over the years from its original format (Figure 1).

As we gathered data on server access and utilization and "new" search tools are developed, we had to make decisions which tools to keep. Consideration was also given to keeping the interface as simple as possible, text only and limited to a single page interface. Usability was determined by testing with multiple handheld devices and simulators, and user feedback. Although Clinical Trials (Figure 2) was an early choice, it was rarely accessed. User opinion through e-mail and Web feedback form also factored in the decision.

Figure 1 – Original PubMed for Handhelds interface

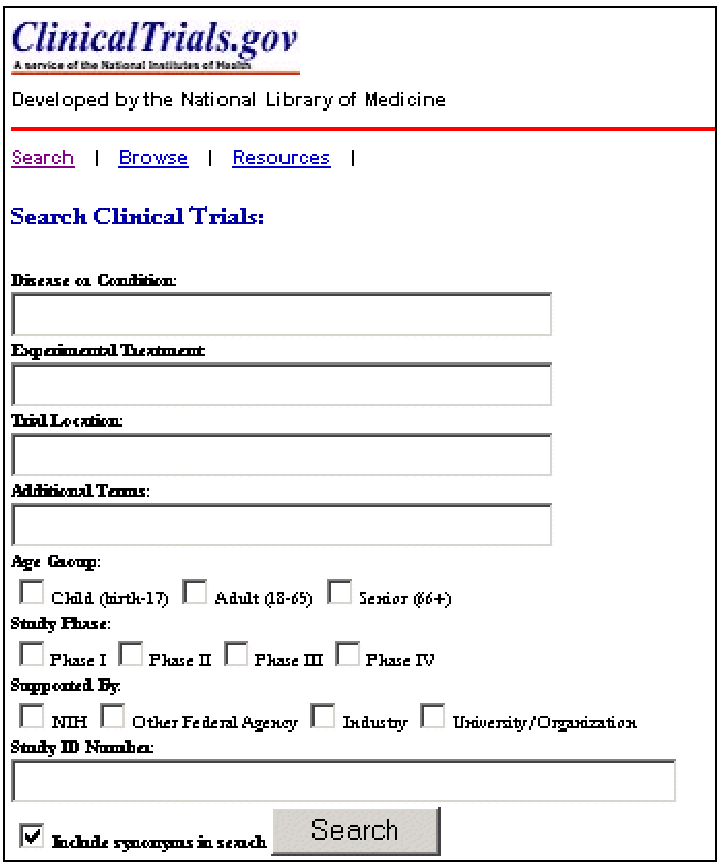


Figure 2 – Clinical Trials was discontinued because of low access

The search interface will likely evolve even further as we continue to explore new search formats, and with the emergence of innovative Internet technologies and more capable wireless devices. However, its underlying structure will always be the same – Internet-based, Web browser enable, device and platform independent, text only and open source. Figure 3 shows the current 'PubMed for Handhelds' index page while Figure 4 displays one of several experimental interfaces as viewed using a Sony Ericsson W580i mobile phone. This interface uses search filter combinations that focus on core clinical journals and Cochrane reviews to provide the user an even more clinically oriented search strategy. It also uses TBL to deliver easily readable summaries of the abstract.

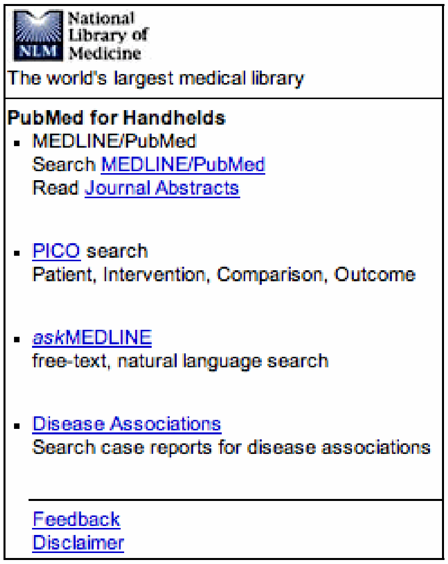


Figure 3 – Current index page of PubMed for Handhelds interface.

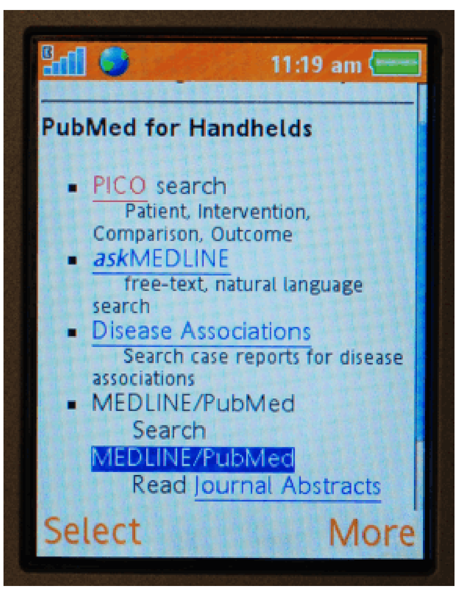


Figure 4 – A smartphone screenshot of a modified PubMed for Handhelds interface.

Access to PubMed for Handhelds has grown steadily over the years. Figure 5 shows total access from 2002 to 2008. Projected total server access for 2008 based on the first six months of server logs is likely to reach close to 70,000.

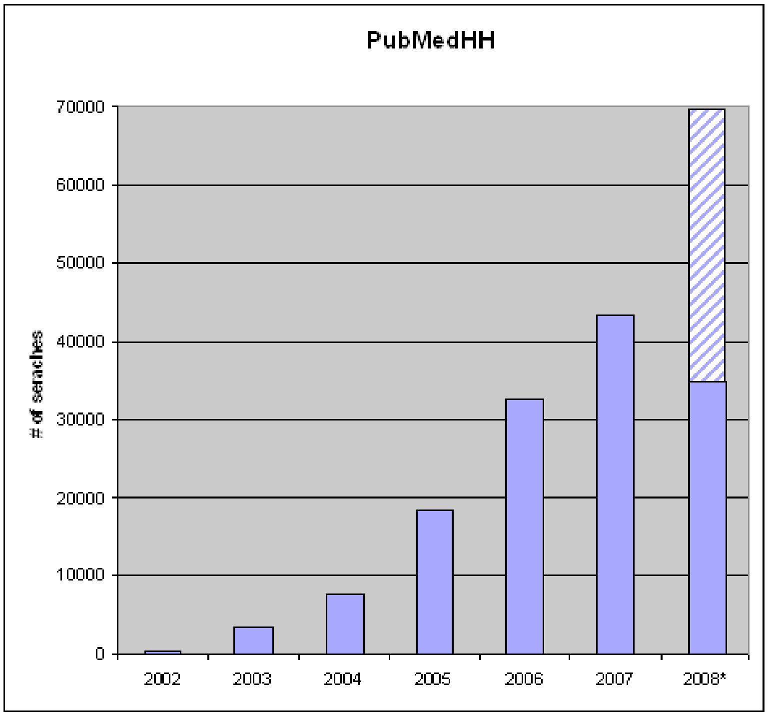


Figure 5. Total searches to PubMed for Handhelds. Hatched area represents expected total access for 2008 based on first six months actual data.

User feedback from 79 responders is shown below in Figure 6. Sixty-six percent said "Extremely useful", 23% "Very useful", 8% "Useful" and 3% "Useful". There was no response to "Not useful". Comments received have also been very positive.

Devices used to access PubMed for Handhelds are as follows: PocketPC 39%, Palm 35%, Blackberry 8%, Linux OS device 6% and other 12%. Forty-two percent of users connected to the server through Wi-Fi or 802.11 networks; 26% of responders through USB to desktop computer, 17% Smartphone, 9% Bluetooth, 4% through an Infrared device and less than 2% via serial to a desktop PC connection. Patient care was the motivation for using PubMed for Handhelds for 25%; Research and Update were even at 27%, while 18% selected Discussing with colleagues.

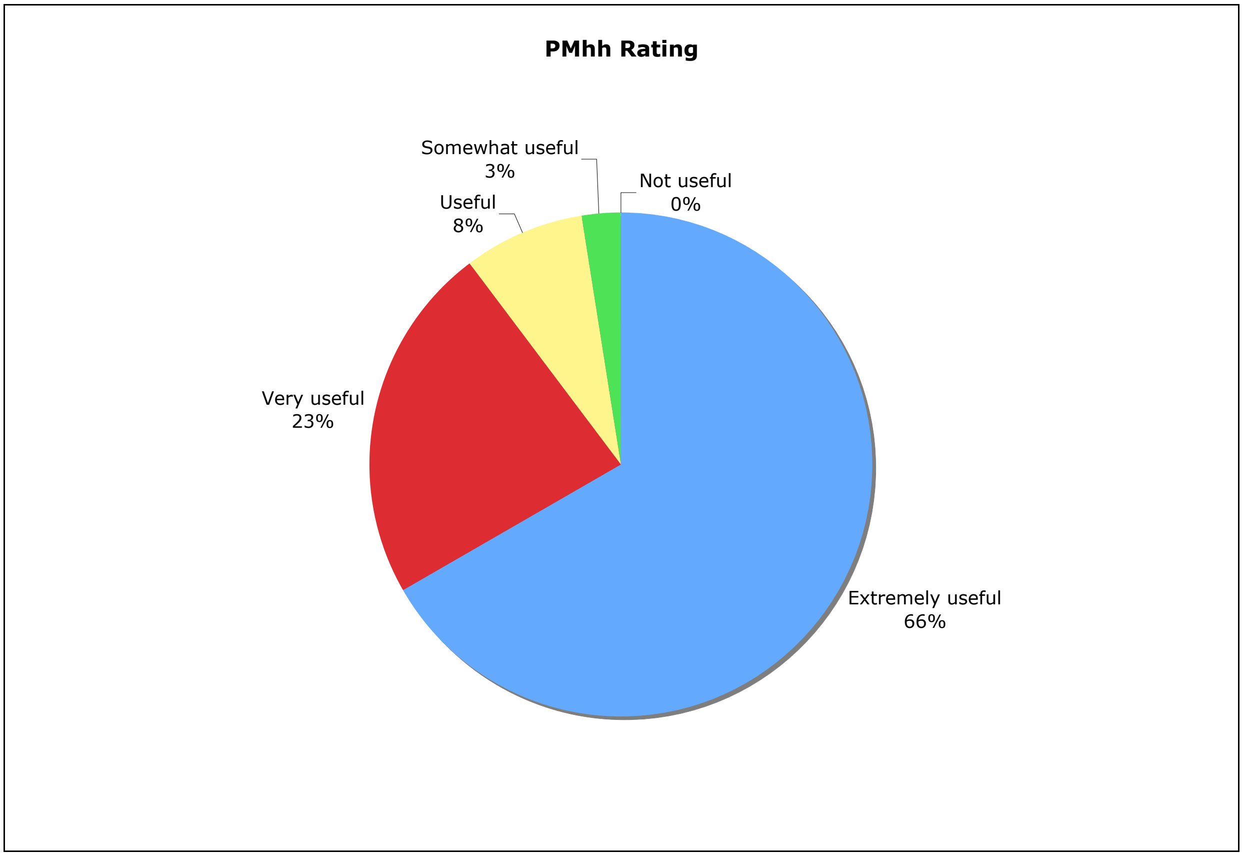


Figure 6. User feedback on perceived usefulness of PubMed for Handhelds.

The rest of this report proceeds in the following order: the current status or implementation of the project, access statistics and future plans for each of the featured search tools of PubMed for Handhelds (Section 4.2). The sections on BabelMeSH (Section 4.3) and txt2MEDLINE (Section 4.4) then follow. A more detailed discussion of the various tools can be found in the Appendix (Section 10).

4.2 Features of PubMed for Handhelds (PMHh)

Section 4.2 discusses the four main elements of PubMed for Handhelds: PICO (Patient, Intervention, Comparison, Outcome), askMEDLINE, Disease Associations and Clinical Queries and Journal Browser.

4.2.1 PICO (Patient, Intervention, Comparison, Outcome) - http://pubmedhh.nlm.nih.gov/nlm/pico/piconew.html

PICO (Patient, Intervention, Comparison, Outcome) is currently the most accessed feature of PubMed for Handhelds. PICO is a method for searching MEDLINE/PubMed that encourages the creation of a well-formulated question [24]. Creating a well-focused question is the first and likely the most important step in the EBM process. Using this framework helps the clinician express the important parts of the clinical question most applicable to the patient and facilitates the searching process by identifying the key concepts for an effective search strategy [25,26].

PubMed (http://www.pubmed.gov) includes a feature called Clinical Queries, which helps identify citations with appropriate study design by linking the type of question (therapy, diagnosis, etiology and prognosis) to a stored search strategy that retrieves the appropriate research methodology. Clinical Queries are based on extensive research by the Hedges Study Team at McMaster University [27] and have been shown to improve searching results. Combining the PICO framework with the PubMed Clinical Queries has the potential to improve the efficiency of literature searching. Bergus et al reported that questions with at least a defined intervention (I) and outcome (O) were more likely to be answered than questions with one or none of these parameters [28].

Current status and Implementation

Figure 7a and 7b show an example of a PICO search comparing the wait-and-see strategy with immediate antibiotic treatment for otitis media and limiting the search to randomized controlled trails. Five citations were retrieved; the first two are shown in the right side of Figure 7a. The structured abstract retrieved for the first article (Figure 7b) is quite long and could be somewhat difficult to read with a small mobile phone.

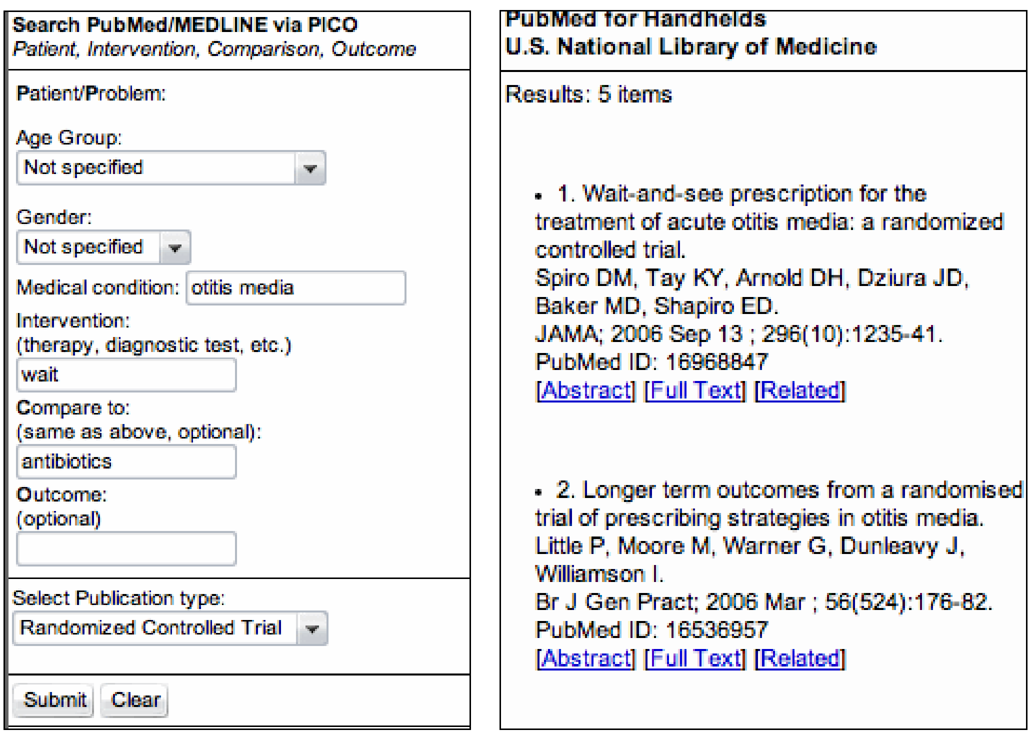


Figure 7a – PICO search comparing "wait-see" strategy with "antibiotics". Search was limited to randomized controlled trials. Left side, search page; right side, citations retrieved; right panel, full abstract of first citation.

PubMed for Handhelds
U.S. National Library of Medicine

- **Title**: Wait-and-see prescription for the treatment of acute otitis media: a randomized controlled trial.

Author: Spiro DM, Tay KY, Arnold DH, Dziura JD, Baker MD, Shapiro ED.
Journal: JAMA; 2006 Sep 13 ; 296(10):1235-41. **PubMed ID**: 16968847.
Abstract:
CONTEXT: Acute otitis media (AOM) is the most common diagnosis for which antibiotics are prescribed for children. Previous trials that have evaluated a "wait-and-see prescription" (WASP) for antibiotics, with which parents are asked not to fill the prescription unless the child either is not better or is worse in 48 hours, have excluded children with severe AOM. None of these trials were conducted in an emergency department. OBJECTIVES: To determine whether treatment of AOM using a WASP significantly reduces use of antibiotics compared with a "standard prescription" (SP) and to evaluate the effects of this intervention on clinical symptoms and adverse outcomes related to antibiotic use. DESIGN, SETTING, AND PATIENTS: A randomized controlled trial conducted between July 12, 2004, and July 11, 2005. Children with AOM aged 6 months to 12 years seen in an emergency department were randomly assigned to receive either a WASP or an SP. All patients received ibuprofen and otic analgesic drops for use at home. A research assistant, blinded to group assignment, conducted structured phone interviews 4 to 6, 11 to 14, and 30 to 40 days after enrollment to determine outcomes. MAIN OUTCOME MEASURES: Filling of the antibiotic prescription and clinical course. RESULTS: Overall, 283 patients were randomized either to the WASP group (n = 138) or the SP group (n = 145). Substantially more parents in the WASP group did not fill the antibiotic prescription (62% vs 13%; P<.001). There was no statistically significant difference between the groups in the frequency of subsequent fever, otalgia, or unscheduled visits for medical care. Within the WASP group, both fever (relative risk [RR], 2.95; 95% confidence interval [CI], 1.75 - 4.99; P<.001) and otalgia (RR, 1.62; 95% CI, 1.26 - 2.03; P<.001) were associated with filling the prescription. CONCLUSION: The WASP approach substantially reduced unnecessary use of antibiotics in children with AOM seen in an emergency department and may be an alternative to routine use of antimicrobials for treatment of such children. TRIAL REGISTRATION: clinicaltrials.gov Identifier: NCT00250900.

[Abstract] [Full Text] [Related] [New Search]

Figure 7b – PICO search showing full abstract of first citation in Figure 7a.

Using "the bottom line (TBL) strategy (Figure 8), discussed in greater detail in Section 4.4, txt2MEDLINE and Appendix D, the algorithm sends only the conclusions of the study for quick reading. For reference purposes, a link to the full abstract is always provided.

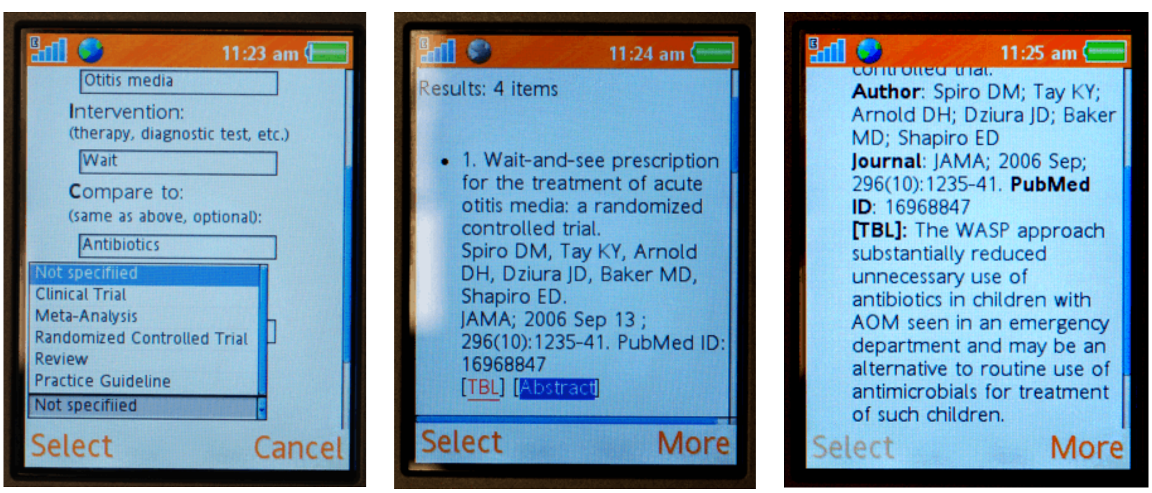


Figure 8 – A smartphone screenshots of a PICO search similar to Figure 5. With the TBL method, only the conclusion is returned to the user for quick reading.

Server analysis

PICO is the most accessed PMHh tool. Figure 9 shows public access to the all PMHh features sorted according to elements. PICO accounts for approximately 60% of all access to PMHh. For 2008, PICO access should reach more than 40,000. Clinical Queries was the second most accessed tool, accounting for 35% of all server access. Access data on Disease Associations and Journal Abstract reader will be discussed in the Sections 4.2.3 and 4.2.4, respectively. askMEDLINE is logged separately and will be discussed in Section 4.2.2.

When asked whether they will use PICO in the future, 79% of responders said "Yes" and 21%, "Maybe". Eighty-nine percent agreed that it was easy to use, while 11% said No. Figure 10 shows responses regarding the usefulness of PICO. Sixty-nine percent agreed that PICO was "Extremely useful", 25% "Very useful", and 6% "Somewhat useful."

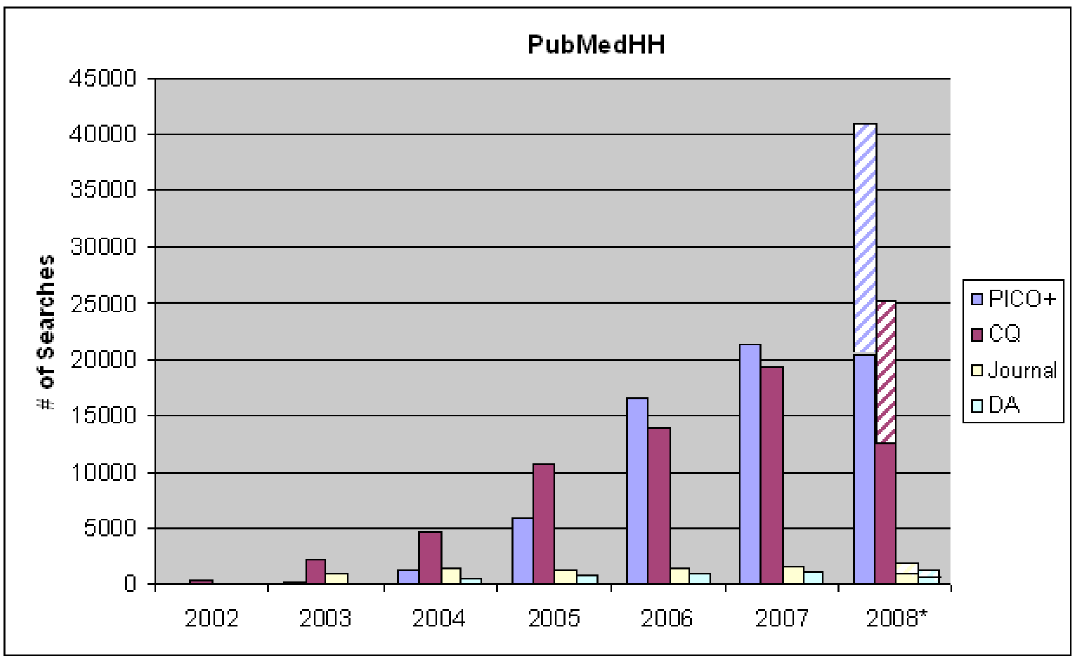


Figure 9. Total searches to PubMed for Handhelds according to available features. Hatched area represents expected total access for 2008 based on first six months of actual data.

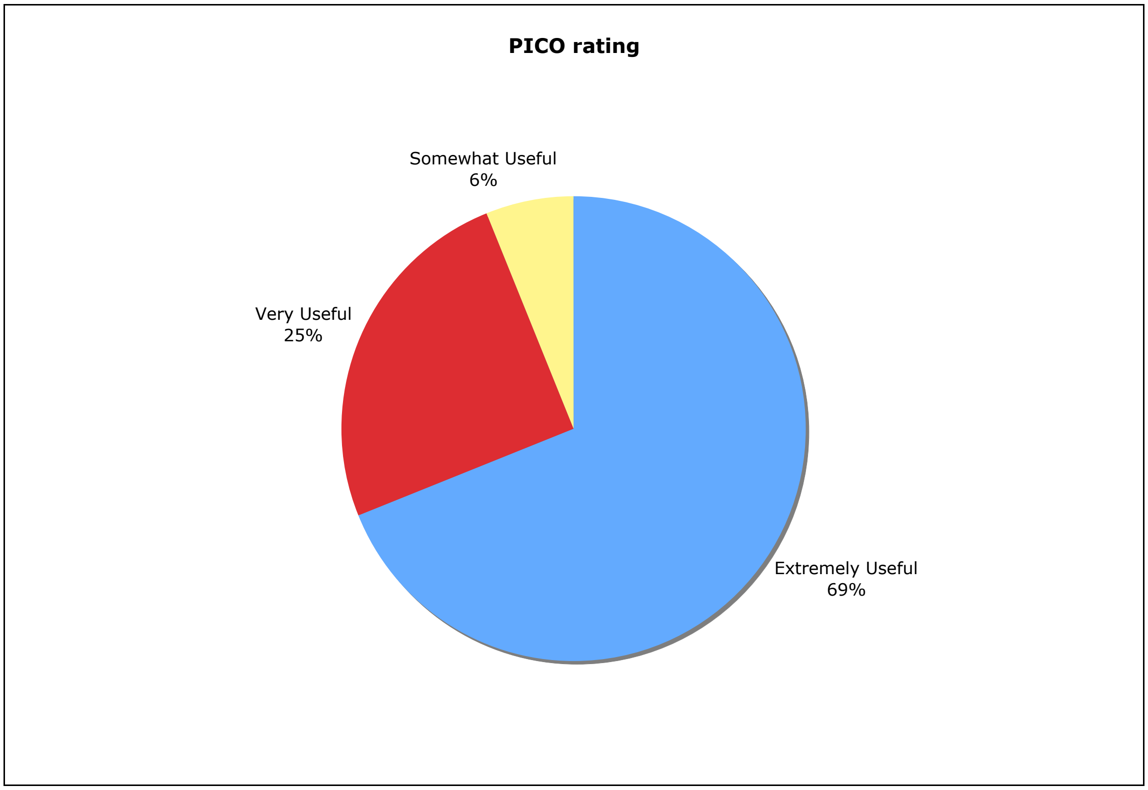


Figure 10. Feedback from users on the usefulness of PICO.

Future plans

We plan to continue to refine PICO along together with other PubMed for Handhelds tools to make them more efficient and appropriate, especially for mobile devices. Combinations of filters and resources will be tested. Examples of new search pages include: http://pubmedhh.nlm.nih.gov/nlmc that will search only for journal articles that have abstracts and published in core clinical journal plus Cochrane reviews and a more compact interface http://pubmedhh.nlm.nih.gov/nlme. The purpose is to focus on clinical studies that might have greater relevance to a clinician. The interface is intended to be simple and is especially suited for mobile phones. They also include "the bottom line" abstract summaries for easy reading (see section on txt2MEDLINE below and Appendix D for details). These tools will undergo clinical evaluation on their usability and effectiveness in clinical settings. We are currently seeking clinical collaborators to evaluate these tools.

Section 10.1, Appendix A, discusses in detail an evaluation on the relevancy of results using three different interfaces of PICO tool done in collaboration with Duke University Medical Center .

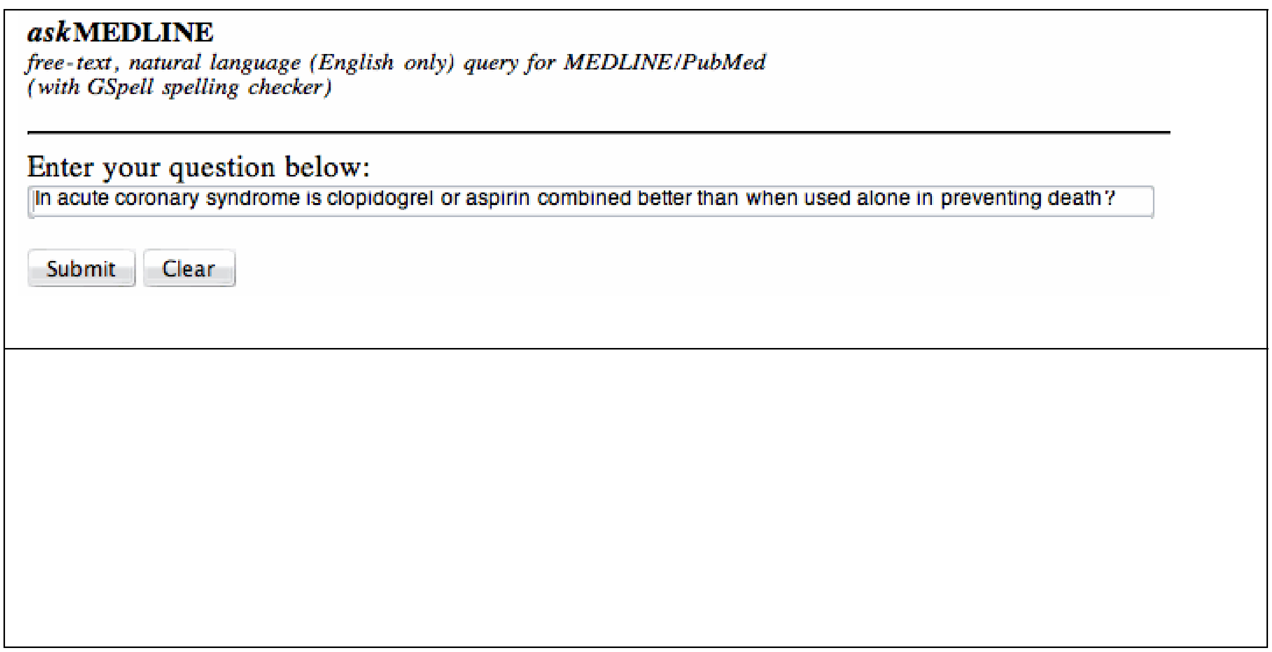
4.2.2 askMEDLINE – http://askmedline.nlm.nih.gov

askMEDLINE evolved from the PICO [29]. In an attempt to automate the entry of search terms into PICO elements from a clinical question, we discovered that the user could just simply enter a clinical question, then let the search engine retrieve relevant journal articles. The step that allowed the user to inspect the appropriateness of PICO elements was omitted. However, a link is always provided to the PICO interface so the user can manually enter search terms if the search results are deemed unsatisfactory.

askMEDLINE is intended for the clinician, researcher, or perhaps, the general public who simply want to ask a question and skip the challenge of learning how to format the query in a manner that will make searching MEDLINE/PubMed efficient. It is a tool that allows the user to search, using free-text, natural language, just like one would do in a conversation. A user enters a clinical question on a Web browser, then lets the tool retrieve relevant articles in MEDLINE/PubMed. Links are provided to journal abstracts, full-text articles and related items. Moreover, askMEDLINE is formatted for easy viewing on a wireless handheld device so it can be used while mobile, but it will work equally well on a desktop computer.

Implementation

Figure 11 shows the askMEDLINE interface and query results. After the user submits a question, it is processed by a multi-step algorithm described in greater detail in Appendix B. It can parse complicated, well-formulated clinical questions that evidence-based guidelines teach clinicians to ask. The GSpell spelling checker, integrated with askMEDLINE can be useful for difficult to spell medical terms. Clicking on a suggested term automatically replaces an incorrectly spelled term.



askMEDLINE
Your question: *In acute coronary syndrome is clopidogrel or aspirin combined better than when used alone in preventing death?*

- **Title**: Adjusted indirect meta-analysis of aspirin plus warfarin at international normalized ratios 2 to 3 versus aspirin plus clopidogrel after acute coronary syndromes.

Author: Testa L; Zoccai GB; Porto I; Trotta G; Agostoni P; Andreotti F; Crea F
Journal: Am J Cardiol; 2007 Jun; 99(12):1637-42. **PubMed ID**: 17560866
Abstract:
After acute coronary syndromes, the beneficial effect of aspirin plus clopidogrel (A+C) or aspirin plus dose-adjusted warfarin (A+W) compared with aspirin alone is well established. However, these regimens were never compared. To compare the risk-benefit profile of A+C versus A+W after acute coronary syndromes, major medical databases for randomized controlled trials comparing 1 of these combined approaches versus aspirin alone after an acute coronary syndrome (updated June 2006) were searched. Evaluated end points were major adverse events [MAEs: all-cause death, acute myocardial infarction [AMI], thromboembolic stroke, major bleeds, and overall risk of stroke [hemorrhagic or ischemic]). Odds ratios (ORs) with 95% confidence intervals (CIs) were calculated for (1) A+W versus aspirin alone, (2) A +C versus aspirin alone, and (3) A+W versus A+C using adjusted indirect meta-analysis. Thirteen studies were included, totaling 69,741 patients. Ten compared A+W versus aspirin alone and 3 compared A+C versus aspirin alone. Each combined approach yielded a significantly lower risk of MAEs, albeit an increased risk of major bleeds, compared with aspirin alone. No significant difference was found for A+W versus A+C for risk of overall MAEs, death, or AMI. However, A+W versus A+C was associated with a significantly lower risk of thromboembolic stroke (OR 0.53, 95% CI 0.31 to 0.88, number needed to treat 60) and all types of stroke (OR 0.58, 95% CI 0.35 to 0.94, p=0.038), but also with increased risk of major bleeds (OR 1.9, 95% CI 1.2 to 2.8, number needed to harm 300). In conclusion, after an acute coronary syndrome, A+W and A+C are comparable in the prevention of MAEs, death, and AMI compared with aspirin alone. Allocating 100 patients to A+W (at international normalized ratio 2 to 3) with respect to A+C could prevent 17 thromboembolic strokes while causing 3 major bleeds.

Figure 11. In askMEDLINE, users can search using free-text, natural language questions. Top, search form; bottom, one of the abstracts retrieved.

Server access analysis

Since its introduction in March 2005, more than 44,000 distinct queries have been submitted to askMEDLINE. Projections for 2008 indicate that it could reach to more than 52,000 by the year's end (Figure 12). The ratio between searches without results and those with results has remained steady, between 7 to 8% annually. Many of the queries with no retrievals result from searches using non-English terms (see discussion on BabelMeSH below). Recently, a large percentage of the increase in askMEDLINE access has been generated by RSS feeds from EpiSPIDER [50], a compilation of worldwide electronic reporting systems for outbreaks of emerging infectious diseases, toxins, natural and manmade disasters, and other significant events. It derives its input from many open sources.

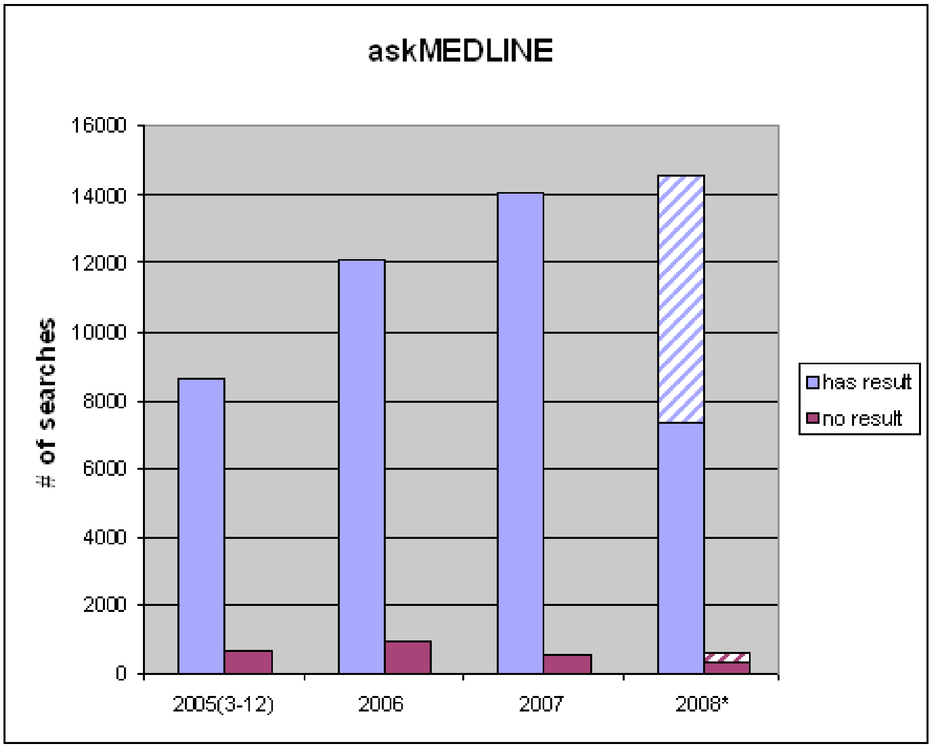


Figure 12. Server analysis from 2005-2008. Hatched area represents expected total access for 2008 based on first six months actual data.

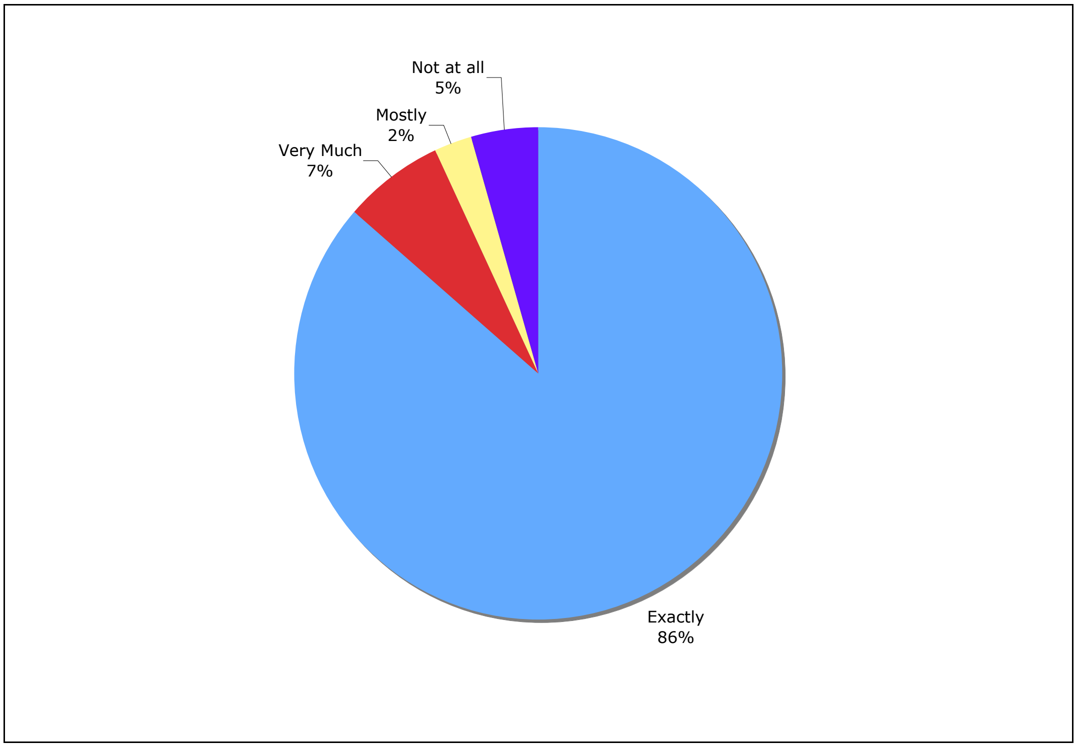


Figure 13. Responses to the statement, "Overall, the articles retrieved addressed my question."

On a 5-point Likert scale (5=strongly agree, 1=Strongly disagree) when users were asked whether they agreed with the statement that the articles retrieved were relevant, 88% of responders selected "5", with "3" and "1" even at 5%, and "4" at 2%. Figure 13 shows the results of responses for the statement, "Overall, the articles retrieved addressed my question". "Exactly" was the overwhelming majority at 86%, 5% selected "Not at all". The majority of survey responders (87%) said that they will continue to use askMEDLINE and 88% answered that they will recommend it to others.

Future plans

We are continuing to archive previously asked questions and exploring the idea of evaluating askMEDLINE using questions banks. Using BabelMeSH databases, we are also considering a non-English version of askMEDLINE.

4.2.3 Disease Associations

About nine percent of citations in PubMed are case reports. This section of PubMed remains underutilized. Disease associations (DA) attempts to make use of this resource. It specifically searches for unusual and rare occurrences, and associations between symptoms, signs, and disease conditions. In can be useful in situations where a clinician is faced with a diagnostic challenge.

The idea was originally suggested by Dr. Murray Longmore, a British physician and author. His interest was sparked by several unusual cases he encountered in his practice, like the association between “Henoch-Schoenlein purpura” and “meningitis vaccine” and “hyperthyroidism” and “urinary incontinence”. In both cases, the DA tool provided satisfactory searches and a successful treatment outcome. An example of a DA search and result is shown in Figure 14. This resource was also useful for local clinicians managing a patient with “lymphoma” and elevated “human chorionic gonadotropin hormone” in a nearby medical center.

Server access

Although its usage is not high, (524 - 2004; 693 – 2005; 937 - 2006; 1025 - 2007; 1186 projected in 2007), feedback from individual clinicians has been very positive.

Future plans

We are planning an evaluation of Disease Associations against expert systems or Google to determine how it compares with them in arriving at the correct diagnosis or differential diagnosis.

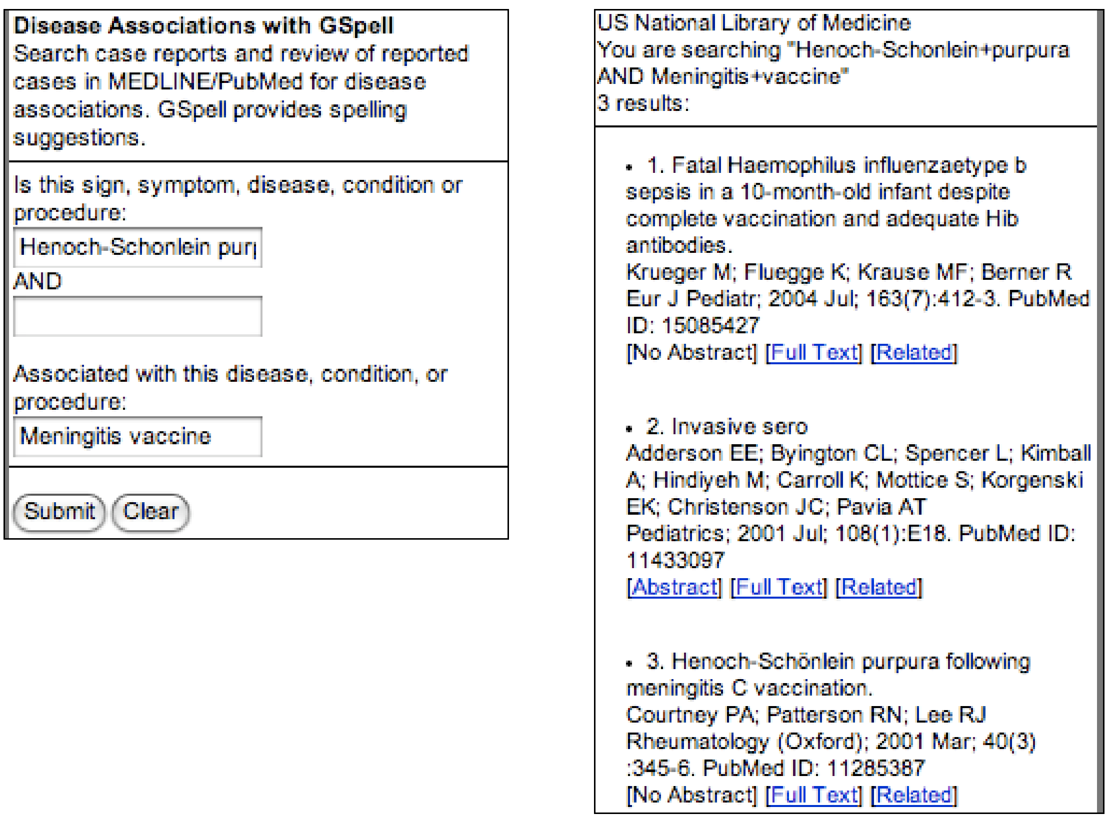


Figure 14. Left, Disease associations search interface. Right, search results on the association between “Henoch-Schonlein purpura” and “Meningitis vaccine”.

4.2.4 MEDLINE/PubMed Search with Systematic Review and Clinical Queries filters and Journal Browser.

PubMed for Handhelds started with Clinical Queries, a search tool that is essentially the same as PubMed's Entrez search. Using the same filters, the search retrievals are essentially the same. It allows the user to search with no filters or limit the search to Systematic Reviews or Clinical Queries filters. The user can limit or expand the search according to Category (therapy, diagnosis, etiology, prognosis and clinical prediction guides) and Emphasis to narrow (more specific) or broaden (more sensitive) the search (left panel, Figure 15).

Journal Browser allows the user to read abstracts of journals listed as Core Clinical journals. It is a grouping of about 121 of the most accessed clinical journals indexed by MEDLINE. Citations are listed chronologically with the most recent on top. It is a quick way to keep up with recent publications using a handheld device.

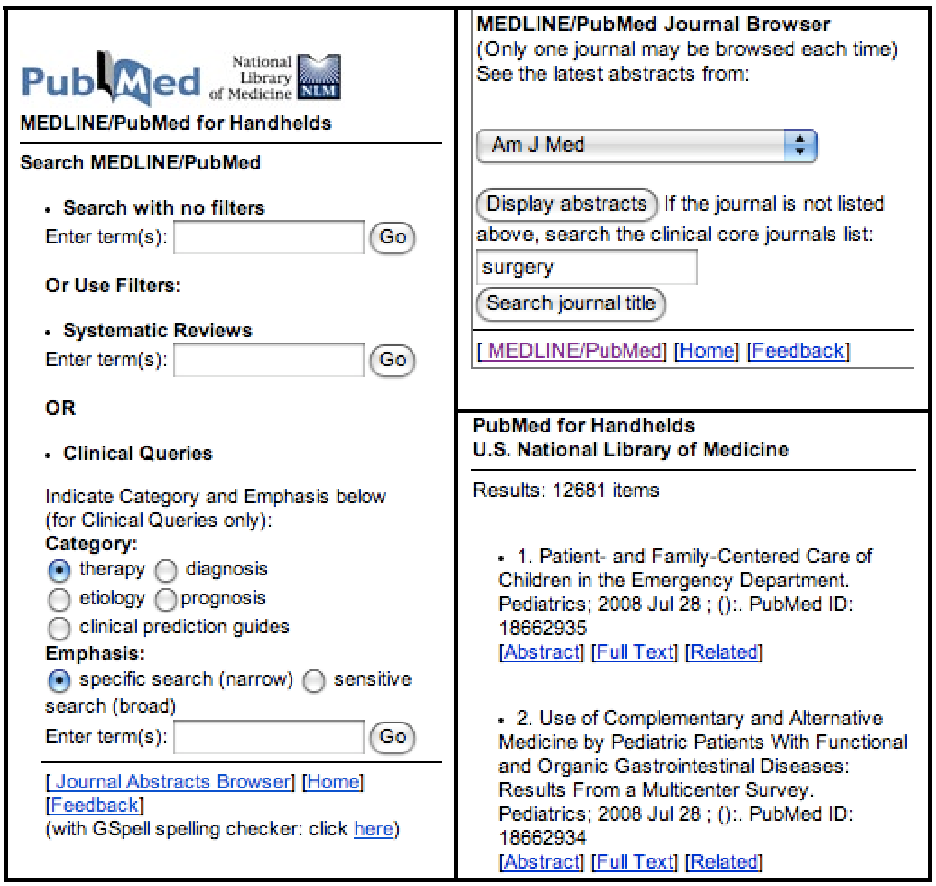


Figure 15. PubMed for Handhelds search with or without filters (left) and Journal Browser (right).

Server Access

Server logs show Clinical Queries access for the following years: 2002-354; 2003-2260; 2004-4633; 2005-10,695; 2006- 13,969; 2007-19,243 and 2008-25,956. After PICO, it is the most accessed search tool in PubMed for Handhelds.

Future Plans

We continuously monitor for modifications and innovations of PubMed since this tool a mirror of Entrez. We will update the interface, filters and search strategies based on changes in Entrez.

4.3 BabelMeSH and PICO Linguist - http://babelmesh.nlm.nih.gov

BabelMeSH (Figure 16) is a multilanguage search tools intended for users whose native language is not English. PICO Linguist is the PICO-formatted version of BabelMeSH. Behind BabelMeSH are multi-language databases of medical terms created using concept identification equivalents of English terms to other languages. The primary sources of vocabularies are UMLS, MeSH, WHO EMRO and UMLF. The search interface changes according to the language selected which allows search terms to be entered in the native language.

<BabelMeSH>< Multilanguage Search for MEDLINE/PubMed (Beta)

Search MEDLINE/PubMed using medical terms or phrases in Arabic, Chinese, English, French, German, Italian, Japanese, Korean, Portuguese, Russian, and Spanish. Citations and abstracts retrieved will be in English only, but full-text links may be available. Full-text journals may require subscription. Typing accents is optional.

:: Arabic - البحث باللغة العربية في ميدلاين MEDLINE / PubMed

:: Chinese - 使用中文搜索MEDLINE/PubMed

:: Dutch - Doorzoek MEDLINE/PubMed in het Nederlands

:: English - Search MEDLINE/PubMed in English

:: French - Recherche MEDLINE/PubMed en français

:: German - Literatursuche in MEDLINE/PubMed auf Deutsch

:: Italian - Cerca in MEDLINE/PubMed in Italiano

:: Japanese - MEDLINE/PubMedの日本語による検索法

:: Korean - 한글 MEDLINE/PubMed 검색

:: Portuguese - Pesquisa MEDLINE/PubMed em português

:: Russian - Введите поиск в MEDLINE/PubMed на русском языке

:: Spanish - Búsqueda MEDLINE/PubMed en español

:: Swedish - Sök i MEDLINE/PubMed på svenska

[**Search PICO Linguist**]

[**Lookup a MeSH term**]

Figure 16. BabelMeSH index page. Including English, a user can search in 13 languages.

This project resulted from an observation in askMEDLINE [30] that about eight percent of queries used non-English terms. This is not unexpected because when searching the medical literature, it is often easier to use the language in which one is most accustomed to [31]. This observation motivated us to design alternative search tools for MEDLINE/PubMed. PICO Linguist and BabelMeSH are the results of these efforts. Using their own native language, a user might be able to express certain concepts that can be difficult to translate into English. MeSH translations could also be useful for healthcare providers and researchers who are not too comfortable with English in locating journal articles in MEDLINE/PubMed.

The user can also limit the search output according to the language of publication but the citations retrieved are in English only. Links may be provided to journals if published online.

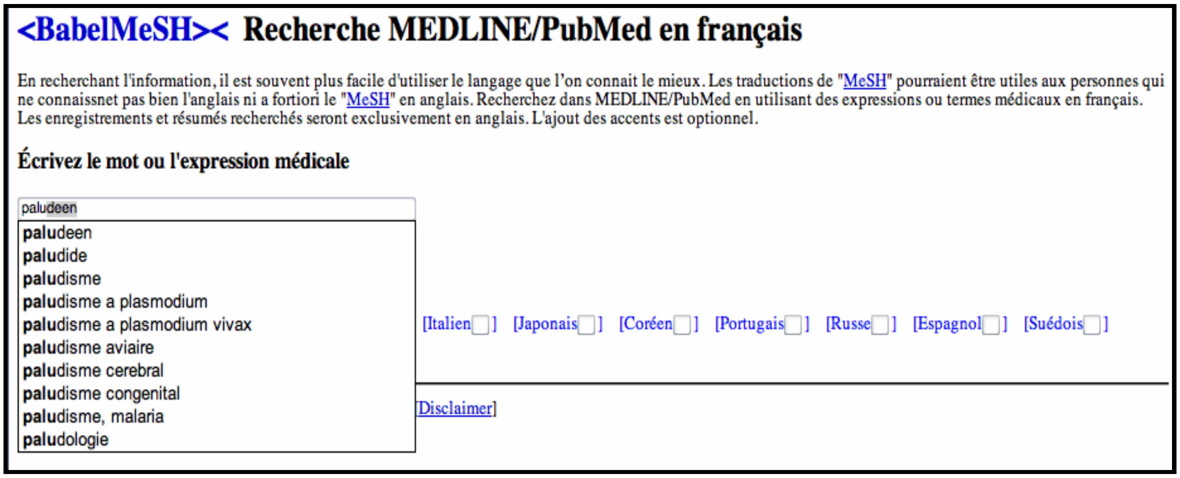


Figure 17. As the first letters of the term of interest are typed, entries in the database appear.

Using JavaScript, BabelMeSH has an autosuggest feature that provides a dropdown list (Figure17) of terms in the database. The example shows a search for malaria in French. This can be useful if the user is uncertain of a term's spelling. Selecting a term in the database assures the user that a search will be successful.

Figure 18 demonstrates the PICO formatted version, PICO Linguist, in Spanish. The user can limit the language of publication by selecting pertinent check boxes. The default is publications in all languages.

Figure 18. PICO Linguist interface in Spanish.

Current usage

We monitor usage of BabelMeSH and PICO Linguist daily via mashups. Special logs are created for both tools so they can be verified for responsiveness to queries. Figure 19 is a text display for BabelMeSH and PICO Linguist of the graphical display of all BabelMeSH queries shown in Figure 20. In this figure, a popup window demonstrates details of searches from Cuba.

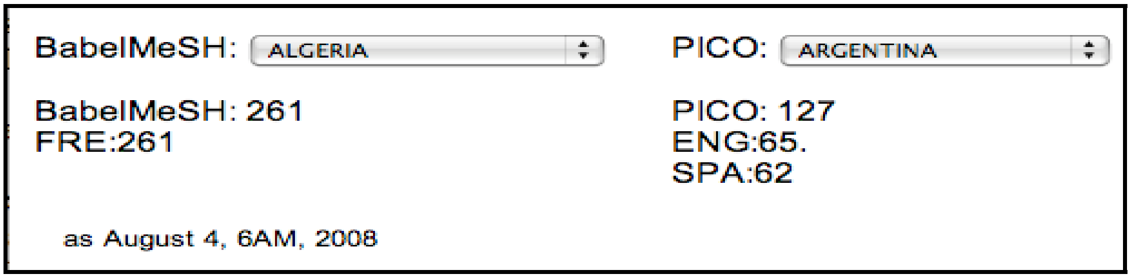


Figure 19. A dropdown text menu can also show the same popup information shown in Figure 14.

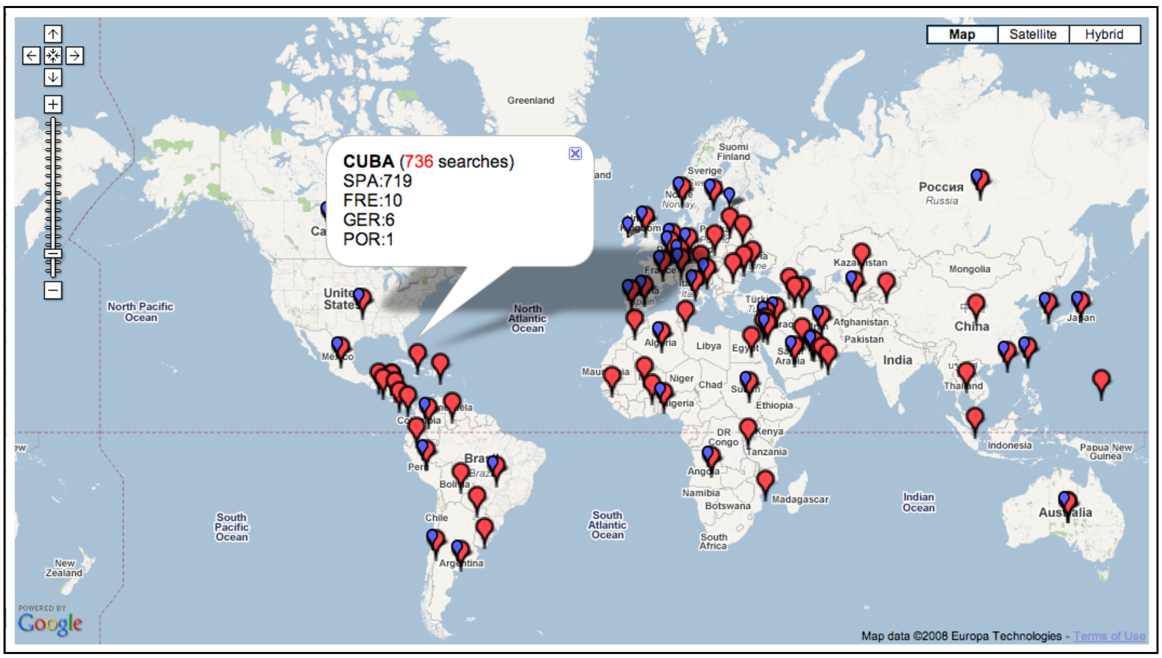


Figure 20. Mashup of BabelMeSH and PICO Linguist. Example shows access from Cuba.

Since its launch, several collaborators have requested to use BabelMeSH's multilanguage databases. One such application is the American Roentgen Ray Society's (ARRS) Goldminer Global [32]. Goldminer Global provides access to more than 196,000 images published in 249 selected peer-reviewed radiology journals. We provided ARRS with an application programming interface (API) to refer their non-English searches to BabelMeSH. The English translation is returned to Goldminer, which is then used to query their image databases.

Figure 21 shows total Goldminer Global referrals from April 2007 and 2008, while Figure 22 displays access sorted by language. The relatively high usage of Japanese can be explained by a special agreement between ARRS and Japan Radiological Society. The Society provides a search toolbar from their home page (http://www.radiology.jp). Goldminer is developing similar agreements with other radiological societies represented in BabelMeSH so there figures will likely increase.

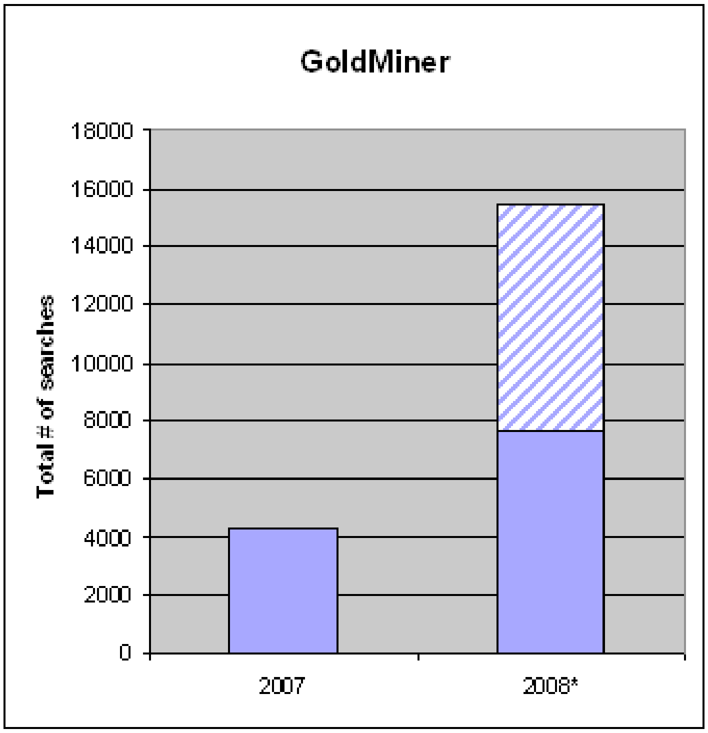


Figure 21. Total Goldminer Global referrals, 2007-2008. Hatched marks are estimates based on first six-month data.

A more comprehensive discussion on the algorithm behind BabelMeSH and PICO Linguist, and evaluation of the French and Spanish versions can be found in Section10.3 Appendix C.

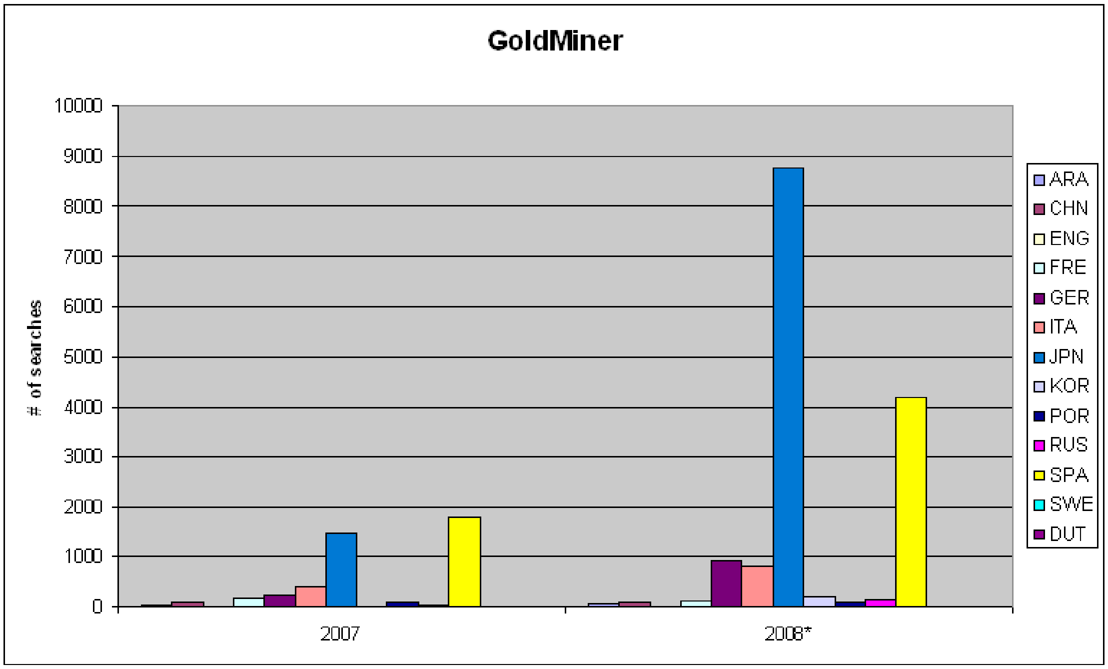


Figure 22. 2007 - 2008 Goldminer Global referrals by language.

Future plans

Although BabelMeSH translates non-English queries into English, the citations and abstracts retrieved from MEDLINE/PubMed are all in English. For those with limited knowledge of English, this could still be challenge. The ideal setting would be for MEDLINE to provide parallel results in English and non-English. We are planning an option to automatically translate the abstract from English to other language pairs. Although machine language translations are available, they are not perfect - they do not yet approach the capability of a native speaker or a human language translator. The context in which the words are used, give meaning to words but "translating" context is difficult. However, machine translation tools such as Google Translate (http://translate.google.com) and Babelfish (http://babelfish.yahoo.com) may provide some help. Machine translation together with the users English skills, however rudimentary it may be, could contribute to a better comprehension of the abstract.

We have also proposed a new method of evaluating the accuracy of machine translation by clinical weighting [33]. A term's weight is determined by its importance in the clinical diagnosis. A new algorithm was developed based on clinical term weighting. In a preliminary study on the Portuguese version, we compared simple word alignment, clinical weighting algorithm and human review by translation efficiency. The results were 65%, 86% and 90% respectively. This method encourages a quantitative evaluation of translation accuracy and facilitates automation. Clinical weighting could promote a more efficient translation evaluation. We will apply this method to other language versions.

4.4 txt2MEDLINE

Mobile devices, including cellular telephones and personal digital assistant (PDA), are widely utilized by physicians and healthcare professionals. Short Message Service (SMS) or text messaging is available on most digital mobile telephones in use today. Worldwide, most networks use the Global System for Mobile Communications (GSM) system accessible worldwide.

Text messaging is one of the most utilized forms of electronic communication today. Portio Research estimated that 1000 billion text messages were sent in 2005 [34]. This is likely an underestimate because in the Philippines alone, Smart, one of three major wireless telecommunications company reported in 2006 that one billion text messages passed through their network daily [35]. Asia and Europe account for the majority of these numbers. Recently, the GSM Association (GSMA), whose members provide more than 86% of the world's mobile phones, announced recently that there are now more than three billion connections among member networks [36]. With the continued increase in mobile telephones worldwide, (15 new connections are added per second or 1.3 million per day) SMS use is expected to increase significantly.

Text messaging is utilized in health and medicine. Reported medical applications include: medication reminders for asthma, diabetes and other patients requiring chronic medication intake, prescription refills, appointment reminders, smoking cessation programs and dispensing medical advice [37,38]. It is also used for informal consultations among medical colleagues.

Physicians need access to reference resources at the point of care to practice evidence-based medicine. Through MEDLINE/PubMed, MedlinePlus, and other knowledge sources, the National Library of Medicine (NLM) provides portals to current medical literature to anyone with access to the Web, including wireless handheld devices. A text-only Web interface formatted for handheld devices is available. It also works well on desktop computers and is suitable for healthcare providers in low bandwidth environments.

The proliferation of mobile phones and the popularity of SMS open new opportunities to extend Evidence-Based Medicine (EBM) resources to mobile phones and other wireless devices. These devices, with even smaller viewing monitors and a 160-character per message limit imposed by the SMS protocol, create challenges to the delivery of appropriate medical content. Although the text-only interface works with mobile phones, one journal citation alone will require multiple text messages. The increased expense and inconvenience reading the messages could become a hindrance to practicing EBM. It can also present difficulties in entering medical terms on a mobile phone keypad.

We addressed these challenges through the development of an SMS system detailed in Appendix D, Section 10.4. The components include an inbound/outbound message server with a GSM modem, a medical terms abbreviation and acronym database, abbreviated abstract summaries and an algorithm that reduces the message size. A database of medical abbreviations and acronyms was developed to reduce the size of text content in journal citations and abstracts. Queries may be sent as full-length terms or abbreviations in the database. An algorithm transforms the citations into the SMS format. An abbreviated "the bottom-line" (TBL) summary, as a substitute to the full abstract is sent to the mobile device to shorten the resulting text. It is accessible from any location worldwide where GSM wireless service is available.

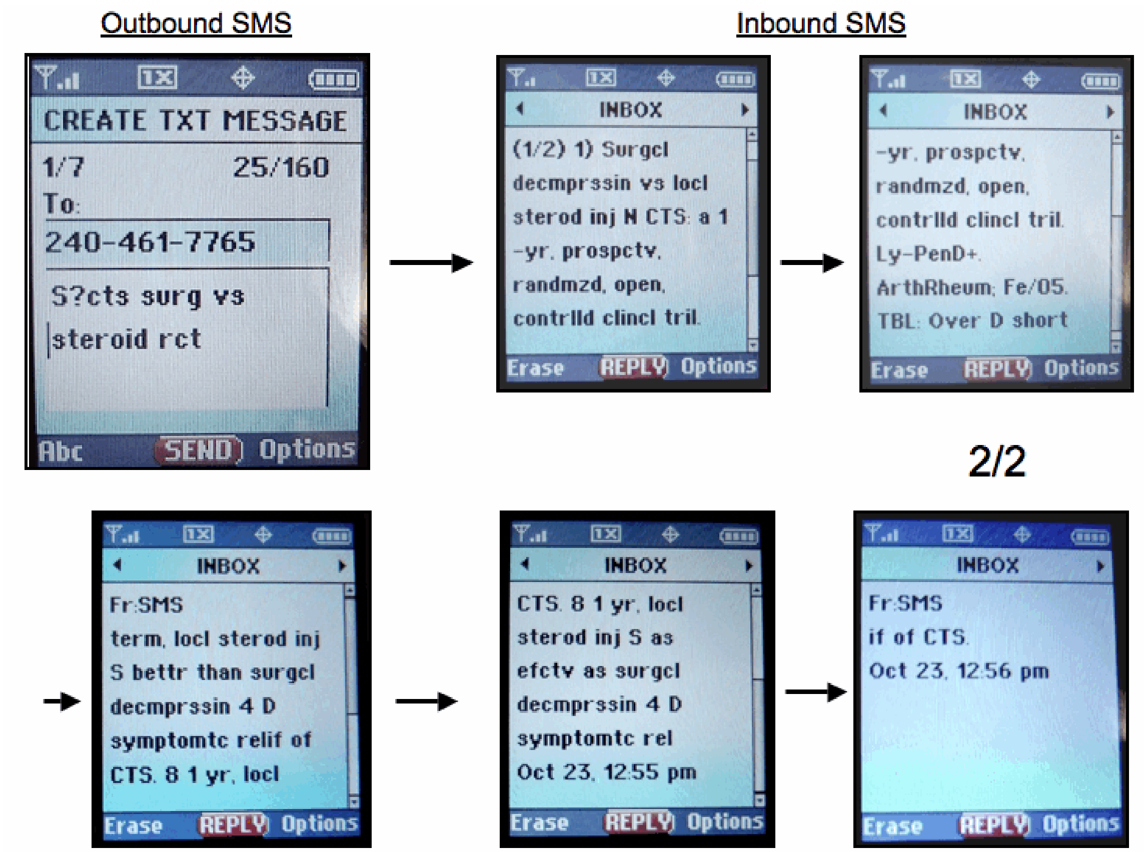


Figure 23. Mobile phone screenshots of an SMS search for randomized controlled trials comparing "surgery" to "steroids" in "carpal tunnel syndrome."

Figure 23 shows an example of an SMS search. The outbound SMS message contains 25 characters in one message. The system sends only the first abstract to the user by default but more abstracts can be sent if requested. In the example above, the TBL abstract format and algorithm shortened the inbound messages to two messages. Although in most countries of the world the user is charged only for sending messages not for receiving them, there is no assurance that the messages will be sent or received in correct order. If messages are long, requiring multiple SMS messages, reconstituting the abstract could be challenging, thus providing another rationale for truncating messages.

US NLM
aom abx wait rct = Acute otitis media antibiotics wait AND randomized controlled trial
1/4 results:

1. Wait-&-see prescrptin 4 the tx of AOM: a RCT. SpiroDM+. JAMA; Se/06. **TBL:** the WASP approch substntilly redcd unnecssry use of ABX N childrn w/ AOM seen N an emergncy deprtmnt & may B an alterntv to rotn use of antimcrbils 4 tx of such childrn.
[Plain text]
Wait-and-see prescription for the treatment of acute otitis media: a randomized controlled trial. SpiroDM+. JAMA; Se/06. TBL: The WASP approach substantially reduced unnecessary use of antibiotics in children with AOM seen in an emergency department and may be an alternative to routine use of antimicrobials for treatment of such children.

Figure 24. Another version allows the user to verify SMS abbreviations.

For teaching purposes and for users outside the US and Canada, we developed Web versions of the SMS search system. In Figure 24, the example shown allows the user to click on a link to show the full, plain text version of the TBL abstract in case of difficulty in reading or comprehending the SMS message. In a study to evaluate the efficiency of the TBL algorithm, we showed that it decreases citation size by 77.5±7.9%. txt2MEDLINE provides physicians and healthcare personnel another rapid and convenient method for searching MEDLINE/PubMed through wireless mobile devices.

Section 10.4, Appendix D discusses in detail the methods, algorithms and an evaluation of “the bottom line” summary used in txt2MEDLINE.

Future plans

In order to encourage the practice of evidence-based medicine in developing countries, we plan to develop an installation package that can be deployed in a university, medical school or government institution computer. This system will avoid toll charges for international text messages to the NLM’s GSM modem. For example, one local SMS message in the Philippines will cost PhP1 (2 cents US), whereas an international SMS message will cost PhP15 (33 cents US). The cost will deter doctors from using this system. Another shortcoming is that our SMS search system will not send out international text messages. To resolve this problem, we plan to develop SMS search modules that can be installed in a developing country.

The installation package will contain all the software for a Linux computer. Users in the developing country will simply need a Linux computer, a mobile phone and an Internet connection. Once operational, clinicians can simply send SMS to a local number. The search will be forwarded to the NLM and results via the Internet, then by SMS or e-mail back to the user.

We are also currently collaborating with the Philippine National Telehealth Center on providing SMS consultation support to physicians practicing in remote locations. A searchable archive of commonly searched topics, journal articles and emergency procedures will be developed. The list of procedures will continue to accrue from actual field use.

Another possible project is the development of a medical SMS vocabulary. We are interested in creating a standardized archive of SMS terms for medical uses. We also plan to collaborate with a language department of a university to do a formalized study of SMS “language” for health and medicine.

5. Summing up and looking ahead

Convenient, ubiquitous access to NLM knowledge resources, especially MEDLINE/PubMed, using any device connected to the Internet has been the main goal of this project. Current medical information is likely to be used only if it is easily accessed. We are guided by the principles of platform independence (device or operating system), text-based resources, compatibility with any Web browser and mobility. Users in low bandwidth areas, such as clinicians and researchers in developing countries have also been considered in the development of tools.

Resources in this project include PubMed for Handhelds (Clinical Queries search, a Journal Browser, PICO (Patient, Intervention, Comparison, Outcome), askMEDLINE, and Disease Associations), the multi-language search tools BabelMeSH and PICO Linguist. txt2MEDLINE leverages the billions of mobile phones in use in the world today.

We are looking forward to enabling automatic machine translations of MEDLINE citations to non-English languages in our databases and developing resources to make txt2MEDLINE a local, in-country resource in developing countries. Finally, faced with the reality that most doctors today use MEDLINE abstracts as their only guide to clinical decision-making, we are interested in raising the issue of reconsidering the rules of evidence-based practice.

6. References

- Sung NS, et al. Central Challenges Facing the National Clinical Research Enterprise. JAMA 2003;289:1278-1287. http://jama.amaassn.org/cgi/content/abstract/289/0/1278
- NIH Roadmap Re-engineering the Clinical Research Enterprise http://nihroadmap.nih.gov/clinicalresearchtheme
- Sackett DL, Straus SE. Finding and applying evidence during clinical rounds: the "evidence cart". JAMA 1998 Oct21;280(15):1336-8.
- Crowley SD, Owens TA, Schardt CM, Wardell SI, Peterson J, Garrison S, Keitz SA: A Web-based compendium of clinical questions and medical evidence to educate internal medicine residents. Acad Med 2003, 78:270-274.
- Klein MS, Ross FV, Adams DL, Gilbert CM: Effect of online literature searching on length of stay and patient care costs. Acad Med 1994, 69:489-95.
- Westbrook JI, Coiera EW, Gosling AS: Do online information retrieval systems help experienced clinicians answer clinical questions? J Am Med Inform Assoc 2005, 12(3):315-21
- Ely JW, Osheroff JA, Ebell MH, Chambliss ML, Vinson DC, Stevermer JJ, Pifer E: Obstacles to answering doctors' questions about patient care with evidence: qualitative study. BMJ 2002, 324:710

- Green ML, Ciampi MA, Ellis PJ: Residents' medical information needs in clinic: are they being met? Am J Med 2000, 109:218-23
- Green ML, Ruff TR: Why do residents fail to answer their clinical questions? A qualitative study of barriers to practicing evidence-based medicine. Acad Med 2005, 80:176-82.
- Shaughnessy AF, Slawson DC, Bennett JH. Becoming an information master: a guidebook to the medical information jungle. J Fam Pract 1994;39:489-99.
- Slawson DC, Shaughnessy AF. Obtaining useful information from expert based sources. BMJ 1997;314:947-9.
- Smith R. What clinical information do doctors need? BMJ 1996;313:1062-8.
- Westbrook JI, Gosling AS, Coiera E: Do clinicians use online evidence to support patient care? A study of 55,000 clinicians. J Am Med Inform Assoc 2004, 11(2):113-20.
- American Medical Association Executive Summary: 2002 AMA Study on Physicians' Use of the World Wide Web. Nov. 1, 2002.
- Podichetty VK, Booher J, Whitfield M, Biscup RS: Assessment of Internet use and effects among healthcare professionals: a cross sectional survey. Postgrad Med J 2006, 82(966):274-279.
- Garritty Ch, El Emam K: Who's using PDAs? Estimates of PDA use by health care providers: a systematic review of surveys. J Med Internet Res 2006, 8(2):e7.
- Kho A, Henderson LEMS, Dressler DD, Kripalani S: Use of handheld computers in medical education. A systematic review. J Gen Intern Med 2006, 21(5):531-7.
- 3 Billion GSM Connections On The Mobile Planet.http://www.gsmworld.com/news/press_2008/press08_31.shtml
- Wikipedia. http://en.wikipedia.org/wiki/Mobile_phone
- Telcos say 98% of RP population covered. http://technology.inquirer.net/infotech/infotech/view_article.php?article_id=90857
- Board of Scientific Counselors, Minutes of the Thirty-Ninth Meeting http://lhcintranet/Bosc/Sept02/mayminutes.pdf
- “PubMed Available for Use on Handheld, Wireless Devices” http://www.nlm.nih.gov/pubs/techbull/ja03/ja03_pda.html
- Fontelo P, Ackerman M, Kim G, Locatis C. The PDA as a Portal to Knowledge Sources in a Wireless Setting. Fontelo P, Ackerman M, Kim G. Telemed J and e-Health. 2003; 9(2):141-147. PMID: 12855037.
- Villanueva EV, Burrows EA, Fennessy PA, Rajendran M, Anderson JN: Improving question formulation for use in evidence appraisal in a tertiary care setting: a randomised controlled trial. BMC Med Inform Decis Mak 2001, 1:4.
- Snowball R: Using the clinical question to teach search strategy: fostering transferable conceptual skills in user education by active learning. Health Libr Rev 1997, 14:167-173.
- Wilczynski NL, Haynes RB, Hedges Team: Robustness of empirical search strategies for clinical content in MEDLINE. Proc AMIA Annu Symp 2002, 904-8.
- Wilczynski NL, Haynes RB, Hedges Team: Robustness of empirical search strategies for clinical content in MEDLINE. Proc AMIA Annu Symp 2002, 904-8.

- Bergus GR, Randall CS, Sinift SD, Rosenthal DM: Family medicine residents do not ask better-formulated clinical questions as they advance in their training. Fam Med 2005, 37(7):486-90.
- PubMed for Handhelds offers searching via PICO. NLM Tech Bull 2004, (338):e9. http://www.nlm.nih.gov/pubs/techbull/mj04/mj04_pico.html
- askMEDLINE http://askmedline.nlm.nih.gov
- Nelson SJ, Schopen M, Savage AG, et al. The MeSH translation maintenance system: structure, interface design, and implementation. Medinfo. 2004;11(Pt 1):67-9.
- ARRS Goldminer Global. http://goldminer.arrs.org/global
- Liu F, Fontelo P, Fiszman M. Clinically-Weighted Terms in the Evaluation of MeSH Translations: BabelMeSH in Portuguese as a Model. AMIA Symposium 2009 Proceedings.
- Mobile Messaging Future 2005-21010. Global Analysis and forecasts of SMS and MMS Markets. June 27, 2005. Portio Research. http://www.portioresearch.com
- "Filipinos sent 1 billion text messages daily in 2007" Inquirer.net, March 4, 2008. http://technology.inquirer.net/infotech/infotech/view/20080304-122775/Filipinos-sent-1-billion-text-messages-daily-in-2007
- GSMA Press release 2008 - http://www.gsmworld.com/news/press_2008/press08_31.shtml
- Downer SR, Meara JG, Da Costa AC. "Use of SMS text messaging to improve outpatient attendance". Med J Aust. 2005 Oct 3;183(7):366-8.
- Ostojic V, Cvoriscec B, Ostojic SB, Reznikoff D, Stipic-Markovic A, Tudjman Z. "Improving asthma control through telemedicine: a study of short-message service". Telemed J E Health. 2005 Feb; 11(1):28-35.

10. Appendix

10.1 Appendix A – Patient, Intervention, Comparison and Outcome (PICO)

In 2005, collaborative research was conducted with Duke University Medical School. This randomized trial was designed as a pilot study to measure the relevancy of search results using three different interfaces for the PubMed search system. Each participant was randomly assigned to one of 3 interfaces for searching PubMed (Figure 1). Protocol A, used a PubMed/PICO template designed for use on a wireless handheld device. The PubMed/PICO template prompted the searcher for the PICO elements of the question (patient problem, intervention, comparison and outcome), as well as patient age group and gender. There was also an option to select a publication type. The publication types listed were clinical trial, randomized controlled trial, meta-analysis, review or practice guideline. If no publication type was selected, the search default was to include all five, study designs. The PubMed/PICO template with these publication options was designed to favor questions of therapy, the most common type of question asked by clinicians. Protocol B, used a PubMed/PICO/Clinical Queries template that prompted for the PICO and allowed the search to be filtered by type of question and scope of strategy (narrow or broad) or by systematic review. Protocol B incorporated the Clinical Queries filters and allowed the searcher to consider a broader range of question types. Because templates for Protocols A and B were designed for handheld devices, participants assigned to these two protocols were given a handheld device of their choice (either a Palm Tungsten C or a HP iPAC Pocket PC) to use during the study. The Web browser home page for the wireless handheld study devices was preconfigured to connect to the appropriate study interface. Protocol C, PubMed, used the standard web-based PubMed system on a PC workstation. Protocol C did not include a PICO template for formulating the search strategy, but the participants had access to Clinical Queries if they chose to use them.

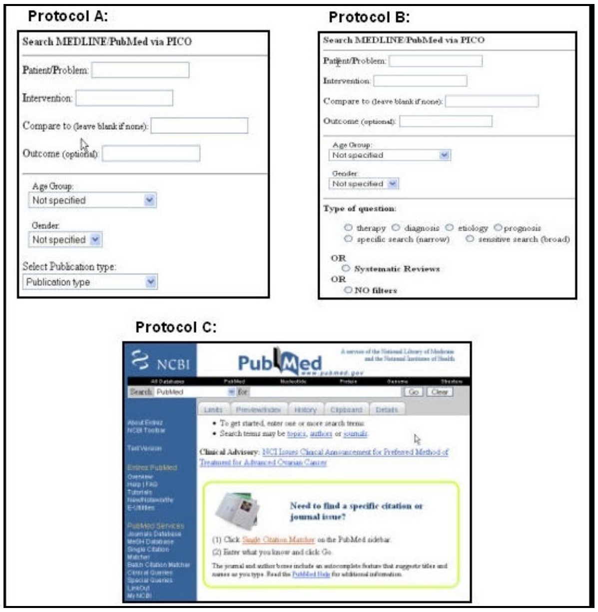


Figure 1. Three interfaces used in searching PubMed.

Study subjects were recruited from interns and residents on an inpatient general medicine rotation at an academic medical center in the US. Thirty-one subjects were each given three clinical questions and asked to search PubMed for a set of relevant articles that would provide an answer to the questions. The three study questions were taken from a database of actual clinical questions formulated by residents during general medicine rotations between 2001 and 2002 [1]. Two of the questions (Q2 and Q3) were related to treatment or therapy; the most common type of question asked and one question (Q1) was related to prognosis. Prognosis question was included because PICO is being taught for all types of questions. Protocol assignments and the 3 clinical questions were placed in a concealed envelope, and participants were asked to select one envelope from a group of identical envelopes. Participants were instructed to search each clinical question, as many times as needed, in order to retrieve a final set of articles that would provide the relevant information needed to make a clinical decision in each case.

Baseline characteristics of study participants			
	Protocol A N = 10	Protocol B N = 10	Protocol C N = 10
Interns	5	5	4
Residents	5	4	5
Other	0	0	1
Previous EBM training	5	9	8
Previous searching training	10	9	9
How often do you search MEDLINE?			
Daily	3	3	3
Weekly	6	3	5
Monthly or less	1	3	2

Schardt *et al. BMC Medical Informatics and Decision Making* 2007 **7**:16 doi:10.1186/1472-6947-7-16

Table 1. Baseline characteristics of study participants

Precision of search results						
	PICO A	95% CI	PICO B	95% CI	PubMed	95% CI
Q1	35%	16.40% – 53.60%	28%	15.84% – 40.16%	20%	10.70% -29.30%
Q2	5%	1.42% – 8.58%	6%	2.42% – 9.58%-	4%	1.50% – 6.50%
Q3	1%	-0.43% – 2.43%	1%	-0.14% – 2.14%	0%	0.00%

Q1: Does BMI affect mortality in obese patients undergoing a CABG?
Q2: Is survival prolonged if I place a PEG or feeding tube in my patient with dementia?
Q3: Is an ACE inhibitor alone better than a diuretic alone for reducing hypertension in African American patients?

Schardt *et al. BMC Medical Informatics and Decision Making* 2007 **7**:16 doi:10.1186/1472-6947-7-16

Table 2. Precision of search results

The success of the search was measured by comparing the number of relevant citations retrieved to the total number of article retrieved in the final set. The research team identified the relevant articles for each clinical question. The criteria for an article being included as relevant was that it addressed the specific clinical question, including patient, intervention and outcome, and that it was of the best study methodology based on the type of question. For example, a therapy question needed to be answered by a randomized controlled trial, systematic review, or meta-analysis, while a prognosis question required a prospective cohort study. Two researchers conducted PubMed searches for each question. These two and a third researcher selected the relevant articles from the pooled results of the searches. A fourth researcher also reviewed the results and reconciled any disagreement among the other reviewers.

Results

The primary outcome measurement was the precision of the final result set selected to answer the clinical question. Precision was defined as the ratio of relevant citations retrieved to the total number of citations retrieved in the set. The higher the precision the more efficient the search, as more of what is retrieved is also relevant. The measurement of precision is appropriate in the clinical setting, where it is often desirable to find a few good articles, as opposed to a research setting, where the measurement of recall, finding all possible relevant articles, is more important. We also examined the number of terms and filters used in the search strategy and user satisfaction with the search interface.

Thirty-one residents completed the study. Ten residents were randomized to Protocol A; 10 residents were randomized to Protocol B; and 11 residents were randomized to Protocol C. The results from one participant in Protocol C were discarded because the saved search strategies were corrupted and not useable. All three groups had similar clinical experience and searching experience, although the participants in Protocol A had less training in EBM (Table 1). Participants using the PICO templates (Protocol A or Protocol B) had higher precision scores for each question than the participants who used the standard PubMed system in Protocol C. (Question 1: A = 35%, B = 28%, C = 20%; Question 2: A = 5%, B = 6%, C = 4%; Question 3: A = 1%, B = 0%, C = 0%) 95% confidence intervals were calculated for the precision for each question using a lower boundary of zero. The 95% confidence limits were overlapping, suggesting no statistical difference between the groups. Although there were no statistical differences between the groups, there may be a trend toward improved precision with the PICO search screens (Table 2).

In addition to quantitative comparisons, this pilot study provided an opportunity for more open-ended insight into how practitioners search. Some searches within each question (Question 1 = 3/30, Question 2 = 9/30, and Question 3 = 27/30) did not retrieve any relevant citations. A qualitative assessment of search strategies that produced no acceptable results revealed three common types of errors: ambiguous mapping of subject headings (MeSH); selecting the wrong publication type; and limiting search queries to just words in the title. The error that accounted for the largest number of non-productive searches was related to the MEDLINE indexing structure or MeSH (Medical Subject Headings). In question three, most searchers used the phrase "African Americans", which maps to the MeSH "African Americans" and retrieves 29077 citations [searched 4/18/07]. However, the common word "blacks" maps to the broader MeSH term "African continental ancestry group" and retrieves 48195 citations [searched 4/18/07]. Most of the relevant citations used either the word "blacks" or were indexed to the broader MeSH term "African continental ancestry group." The second most common error was related to selecting the wrong study design or Clinical Query for the type of question being searched. This was a common problem for Protocol A, which only listed therapy study designs. The third error affected Protocol C and involved limiting search terms to the title field. While searchers often select their articles based on relevant words in the title, a problem arises when more than one word or phrase is appropriate to the topic. For example, a "peg" is also called a "percutaneous endoscopic gastrostomy tube" or "feeding tube." Limiting the search to the word "peg" in the title eliminates mapping to appropriate subject headings (MeSH) such as intubation, gastrointestinal and therefore may exclude relevant articles on the topic that use alternative terminology. Understanding these errors can help in teaching effective searching and in developing better search systems.

Perceptions of ease of use and time spent searching were approximately the same across all three protocols, as were the numbers of terms used in each search, regardless of protocol used. However, of the 30 searches performed using Protocol B (PubMed/PICO/Clinical Queries), 25 (83%) actually incorporated the Clinical Queries into the strategy, as opposed to only 2 (7%) of the searches using Protocol C (PubMed/Web). Both Protocol A and B facilitated the use of publication types and the Clinical Queries by prompting the searcher to consider these elements in the strategy. While these elements were also available in Protocol C, they are either behind the Limits tab or listed in the PubMed Services menu.

References

- Crowley SD, Owens TA, Schardt CM, Wardell SI, Peterson J, Garrison S, Keitz SA: A Web-based compendium of clinical questions and medical evidence to educate internal medicine residents. Acad Med 2003, 78:270-274

10. 2 Appendix B - askMEDLINE

Figure 1 shows the askMEDLINE interface. askMEDLINE uses a multi-round search strategy. In the first round, the parser ignores punctuation marks and deletes words found on a "stop-word" list. The stop-word list includes PubMed stop words, and other words that we found by experience, to be detrimental to the search. The parser, a PHP script, then sends the modified query to PubMed Entrez' E-Utilities. The Extensible Markup Language (XML) file returned by E-Utilities indicates the category of each term in the query. Terms marked as "All Fields" denote that they are neither Medical Subject Headings (MeSH) terms nor MeSH Subheadings. These terms are checked to determine if they are found in a "MeSH Backup vocabulary." The backup vocabulary includes words other than MeSH terms, such as MeSH descriptors, that are classified as "other eligible entries". If an "All Fields" word is in the backup vocabulary, it remains in the query; if it is not, it is deleted. The remaining terms are sent back to PubMed, again through E-Utilities. Human and English language limits are always applied. If the journal retrieval count after the first round is between 1 and 50,000, the first 20 results are displayed in the user's browser and the search process terminates. Further searches are dependent on the user.

The search may proceed to Round 2 under two conditions: 1) If no journals are found in the first round, a result that could signify that the search was too narrow (i.e., too many terms are searched, too many filters), the "All Fields" words are deleted from the query, even though they are found in the backup vocabulary. Only MeSH Terms and Subheadings remain (Round 2A.) 2) If the first round retrieval count is larger than 50,000 articles (an indication that the search was too broad) the "All Fields" words removed during the first round (words not found in the backup vocabulary) are put back into the query (Round 2B.) Round 2B searches contain all the MeSH terms (or MeSH Subheadings) and "All Fields" words in the original question. The updated query from either 2A or 2B is once again sent to Entrez E-Utilities. Retrieved journal articles are sent to the user.

Similarly, if the count returned from second round is in the range of 1 to 50000, the search process terminates. If the second round count is still equal to 0 (denoting that the search is still too narrow) another list of "No-Go Terms", terms that when removed could result in a successful search is checked. Common MeSH abbreviations, acronyms and words like, "method," "affect," and "lead" are examples of terms on the list. New terms are continuously added to this list as they are encountered. The third round modified query is once again sent to E-Utilities and the retrieved journal articles are sent to the user. A result of 1 to 50000 citations terminates the process and displays the first 20 articles.

If askMEDLINE retrieves only one to four journal articles, a search is automatically done for related articles of the top two articles. All the articles (one to four previous) and the first 25 related articles of the first two are retrieved. As in any of the previous steps, the first 20 are displayed in the browser. In all the search retrieval pages, a link is provided for the user to manually intervene and modify the search process through the PICO interface. Links to related articles, full-text articles and abstracts are shown.

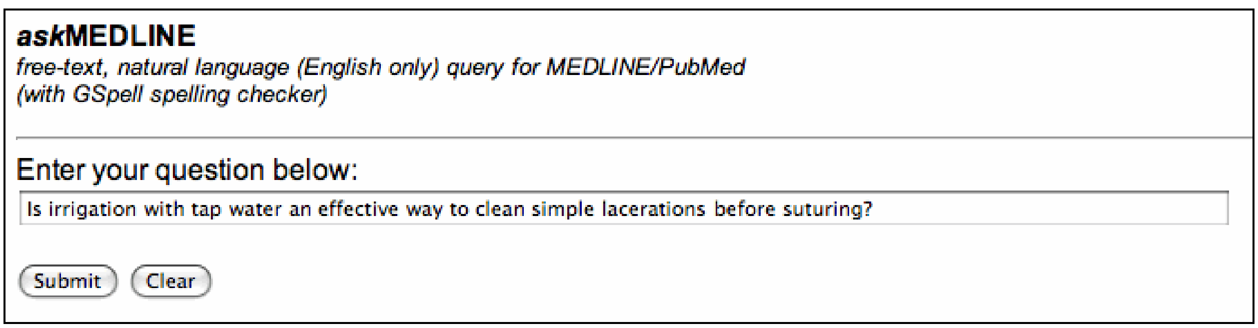


Figure 1. askMEDLINE user interface formatted for handheld devices.

Evaluation

Since November 2002, the British Medical Journal (BMJ) has published a POEM (Patient-Oriented Evidence that Matters) in every issue [1]. POEMs are provided to BMJ by InfoRetriever (now called Essential Evidence Plus) [2]. askMEDLINE was evaluated by comparing its accuracy to retrieve an article cited as a reference in a POEM ("gold standard".) Every POEM has a question with a cited reference that is relevant to the question. We entered every POEM question into askMEDLINE, and for comparison, in Entrez, the integrated, text-based search and retrieval tool for PubMed. New critically appraised topics (CATs) from the University of Michigan, Department of Pediatrics Evidence-Based Pediatrics Web site were also used [3]. Unlike BMJ POEMs, some questions in CATs had more than one cited reference. The initial search result was examined to determine if the reference cited in a POEM or CAT was among those retrieved. Subsequent steps were taken if the reference article cited was not: 1) If the initial search retrieved journal citations, but not the specific journals cited in a POEM or CAT, the titles and abstracts were scanned to find out if they were relevant (deemed to answer the question.) If they were, related articles were retrieved, and again evaluated to determine if they matched the cited reference. 2) If no journal articles were retrieved, the question was rephrased, then searched again. Retrievals were again examined for the cited articles and relevancy to the clinical question. Overall efficiency was determined by the accuracy in retrieving a cited article and relevance of citations retrieved for citations that did not match cited references.

Evaluation of search retrievals

Clinical questions in 95 POEMs and 28 CATs were searched. After first pass, askMEDLINE found 62% of the cited articles in POEMs, while Entrez retrieved close to 14% (Table 1). When related articles were searched, 11.6% more were found by askMEDLINE (8.4% in Entrez.) When three questions were rephrased, askMEDLINE, but none in Entrez retrieved two of the specific cited references, although relevant references were found to one of the questions. For 20 questions, askMEDLINE did not find the specific cited reference, but it found journal citations that were deemed relevant and would be useful in answering the question. Entrez obtained citations for 16 (16.8%) questions that were considered relevant.

POEMs Evaluation Study. A comparison of the accuracy and efficiency of *ask*MEDLINE and Entrez PubMed in retrieving an exact match to cited references in POEMs in BMJ.

Search Step	*ask*MEDLINE exact match retrieved/total questions (% total)	Entrez PubMed exact match retrieved/total questions (% total)
A. Match at first pass	59/95 (62.1)	13/95 (13.7)
B. Match after a related citation search	11/95 (11.6)	8/95 (8.4)
C. Match after question rephrase	2/95 (2.1)	0/95 (0)
Total exact match after A, B, and C	**75.8%**	**22.1**
No match after A, B, and C, but relevant articles retrieved	20/95 (21)	16/95 (16.8)
Overall efficiency	**96.8%**	**38.9%**
Citations retrieved, but not matched or relevant	3/95 (3.1)	6/95 (6.3)
No citations retrieved	0/95 (0)	52/95 (54.7)
Overall retrieval failure	**3.1%**	**61%**

Fontelo *et al. BMC Medical Informatics and Decision Making* 2005 **5**:5 doi:10.1186/1472-6947-5-5

Table 1. POEMs Evaluation Study. A comparison of the accuracy and efficiency of askMEDLINE and Entrez PubMed in retrieving an exact match to cited references in POEMs in BMJ.

Overall, askMEDLINE retrieved 72/95 exact matches of cited references (gold standard) in POEMs, an accuracy of 75.8%, while Entrez' accuracy was 22% (21/95.) If citations that are not the same as those cited in POEMs, but are relevant and considered satisfactory for answering the clinical question are included, askMEDLINE's total efficiency is 96.8%. Entrez' total efficiency for finding specific and relevant citations for BMJ POEMs is 38.9% (21 specific and 16 relevant citations found.)

Although citations were retrieved for all POEM questions by askMEDLINE, three searches did not find exact matches or relevant articles (3.1%), while Entrez' results were not relevant (6.3%) for six questions. No citations were found for 52 (54% of total) questions by Entrez. University of Michigan's CATs yielded a similar total efficiency as POEMs, 96.3%, while it was 14.3% for Entrez (Table 2.) First pass retrieval was 64.2% (Entrez 3.6%) and citation retrievals for related citations was 10.7% (Entrez 3.6%.) Four of six questions rephrased added 14.3% to the total efficiency of askMEDLINE, while it added 7.1% to Entrez. Almost 7% of the searches retrieved relevant citations to rephrased or related articles, but none in Entrez. For CATs' questions, askMEDLINE found 89.2% of cited references, but 14.3% for Entrez. In 21/28 questions, Entrez did not provide a specific or relevant citation, but it was only for one question with askMEDLINE.

CATs Evaluation Study. The accuracy and efficiency of *ask*MEDLINE and Entrez in retrieving an exact match to cited references in CATs' questions from the University of Michigan, Department of Pediatrics Evidence-Based Pediatrics Web site.

Search Step	*ask*MEDLINE exact match retrieved/total questions (% total)	Entrez PubMed exact match retrieved/total questions (% total)
A. Match at first pass	18/28 (64.2)	1/28 (3.6)
B. Match after a related citation search	3/28 (10.7)	1/28 (3.6)
C. Match after question rephrase	4/28 (14.3)	2/28 (7.1)
Total exact match after A, B, and C	**89.2%**	**14.3%**
No match after A, B, and C, but relevant articles retrieved	2/28 (7.1)	0/28 (0)
Overall efficiency	**96.3%**	**14.3%**
Citations retrieved, but not matched or relevant	1/28 (3.6)	3/28 (10.7)
No citations retrieved	0/28 (0)	21/28 (75)
Overall retrieval failure	**3.6%**	**85.7%**

Fontelo *et al. BMC Medical Informatics and Decision Making* 2005 **5**:5 doi:10.1186/1472-6947-5-5

Table 2. CATs Evaluation Study. The accuracy and efficiency of askMEDLINE and Entrez in retrieving an exact match to cited references in CATs' questions from the University of Michigan, Department of Pediatrics Evidence-Based Pediatrics Web site.

References

- Smith R: A POEM a week for the BMJ. BMJ 2002, 325:983
- Essential Evidence Plus (formerly InfoPOEMS/InfoRetriever). http://www.infopoems.com.
- University of Michigan Department of Pediatrics Evidence-Based Pediatrics Web Site. http://www.med.umich.edu/pediatrics/ebm/Cat.htm

10.3 Appendix C - BabelMeSH

Implementation

BabelMeSH is designed as a transparent multilanguage and cross-language interface. Users can submit medical terms in their native language (currently, Arabic, Chinese, English, French, German, Italian, Japanese, Portuguese, Russian and Spanish) then a parser translates the query into English using a multi-language MySQL database. BabelMeSH then sends the query to PubMed through E-Utilities and returns English citations to the user.

PICO Linguist includes all the features that BabelMeSH have but instead of a single input box, PICO Linguist provides users with the structured PICO format consisting of Patient/Problem (P), Intervention (I), Comparison (C) and Outcome (O) input forms. The interface changes according to the input language selected.

Multi-language Database

The major source for most translation records in the databases is the UMLS Metathesaurus, which contains MeSH translations in French, German, Italian, Japanese, Portuguese, Russian and Spanish. Permissions were obtained from the contributing organizations. In addition, Dr. Stephan Darmoni and Dr. Patrick Ruch provided French MeSH translations and the unified medical lexicon for French (UMLF). Dr. Najeeb Al-Shorbaji, Regional Office for the Eastern Mediterranean, World Health Organization, kindly provided Arabic translations of MeSH and the Unified Medical Dictionary. Chinese terms were collected from multiple open source web sites.

The concept unique identifier (CUI) and its concept source were used to find translations from UMLS. Briefly, all the concepts in one non-English language and corresponding CUIs in UMLS were identified. If the English concept in UMLS from MeSH links to the same CUI, the English concept and the foreign language concept were paired. Otherwise, the English concept from another vocabulary was selected.

A hierarchal scheme was devised in order to create separate translation tables for each foreign language using MySQL. The priority of source vocabulary from high to low in our system is: MeSH, UMLS-Meta, SNOMED and any other vocabulary. Table 1 illustrates the structure of each table, which includes CUI, English term, non-English term, and accented non-English term (if available).

CUI	English	French	Accented French
C0702166	Acne	Acne	Acné
C0019686	HIV Antigens	Antigenes VIH	Antigènes VIH

Table 1. Translation table structure (French)

For translation pairs that are not from UMLS where the foreign term and English term do not map to the same CUI, an internally generated concept ID is created by the system. An example is shown in Table 2 (P-45-52, highlighted).

CUI	English	Spanish
C0004916	bed	cama
C0231290	after	despues
C0442022	lumbar	lumbar
C0033119	puncture wound	puncion
C0553794	spinal tap	puncion lumbar
C0004910	bed rest	reposo en cama
P-42-52	rest	reposo

Table 2. Records involved for translating the example query

Character-based language processing

Character-based languages require special handling for string processing in Web applications and in the database. First, we unified character settings by encoding in UTF-8 (Unicode). The default character set in MySQL configuration file is UTF-8. Next, we installed the Multi-byte String Extension package for PHP because each letter or character in these languages may be represented by more than one byte in storage. Multi-byte String Extension can perform string processing for all of these languages. The character setting for internal encoding, HTML input and output in this extension is also in UTF-8.

Parsing algorithm

The parser identifies terms found in the translation database and deletes all others not in the database. The corresponding English translations are sent to PubMed through E-utilities. For terms entered in the search form with accents, the parser searches the accented terms column in the database (Table 1). If no matches are found, it will transform the accented input to the unaccented form, then search the database again and return the unaccented term. If no exact matches for some terms are found, BabelMeSH will suggest records that contain part of the non-English query - - the user may modify the suggestions before sending to PubMed.

An algorithm was developed for complex searches such as combined MeSH terms or concepts. In order to compel the parser to find the optimal translation, the algorithm browses the input by two pointers, recursively (Figure 1). First, translations of PubMed “stop words” are ignored. A multi-word input is then split into an array where each array element stores an input word. Two pointers (Pointer 1 and Pointer 2) are inserted at the beginning and the end of the array. Comparing them with the database identifies individual elements between the two pointers in the array. If no match is found, Pointer 2 moves towards the front of the array one element at a time, until a match is found, or Pointer 1 meets Pointer 2. Once a translation match for the multi-word term or phrase is found, these elements are removed from the array and Pointer 1 will move towards the end to the next unmatched element in the array. If two pointers meet before the translation processing finishes, Pointer 2 will reposition to the end of the array, and Pointer 1 will again move to the end, one element at a time. The translation search will repeat until Pointer 1 and 2 meets at the end of the array. Table 2 shows the records in the database related to the search illustrated in Figure 1.

User Interface

An auto-complete feature (Figure 2, in Chinese) using XMLHttp object request JavaScript, provides suggestions as the query is entered. Suggestions can be selected from the drop down list.

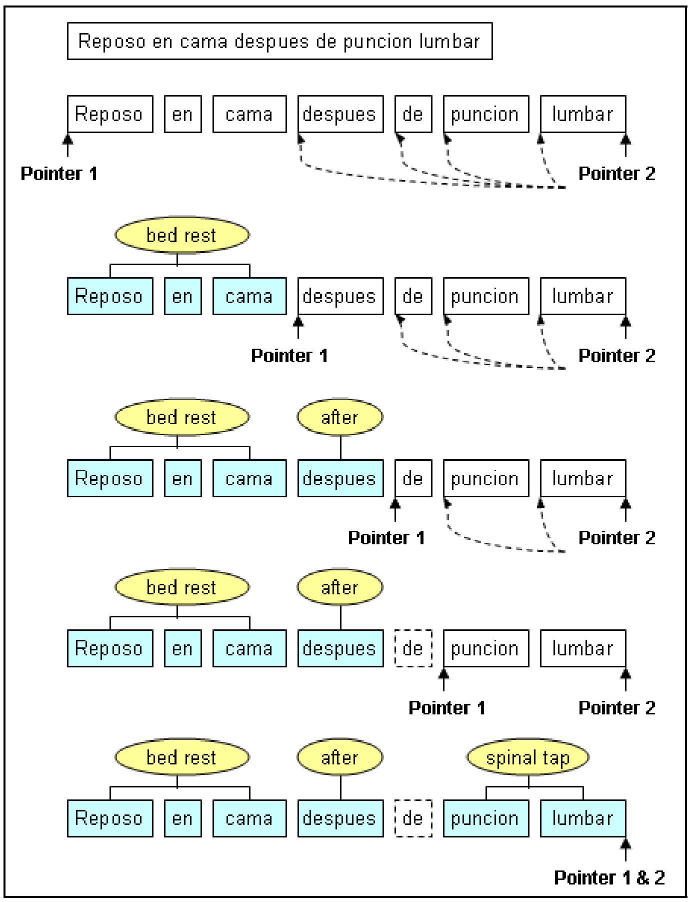


Figure 1. How the parser works

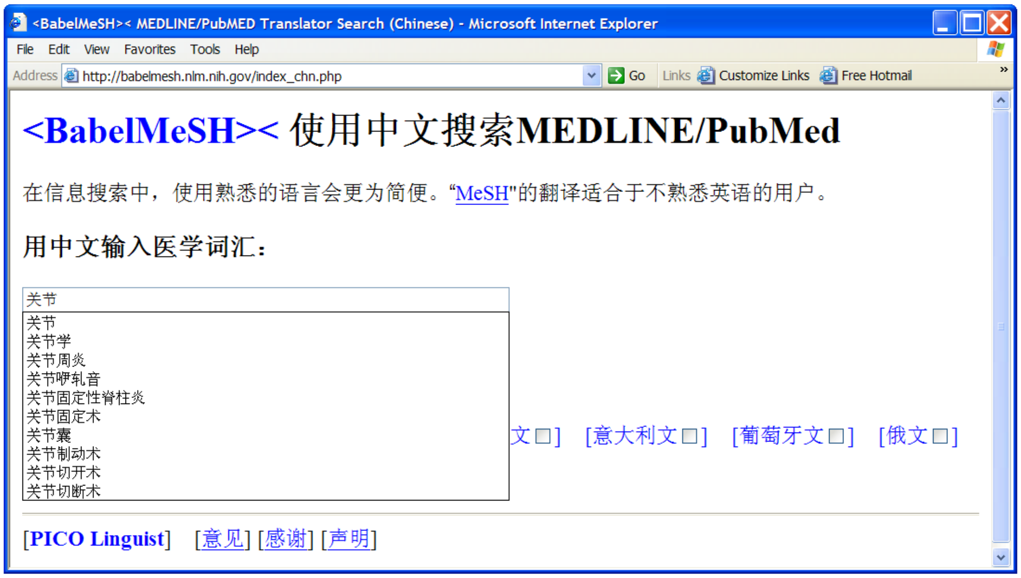


Figure 2. BabelMeSH (Chinese) with auto-suggestion

In PICO Linguist, the embedded JavaScript code makes the Web interface change dynamically with the input language choices, shown in Figure 3. The JavaScript code changes the text orientation to right to left when Arabic is selected (Figure 4).

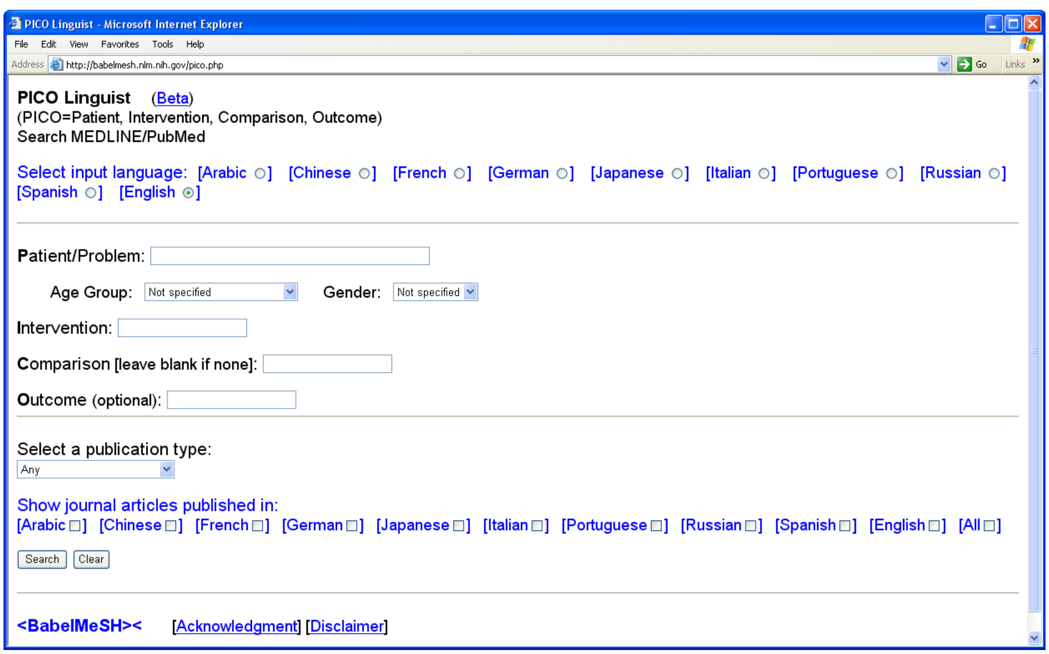


Figure 3. PICO Linguist English search page

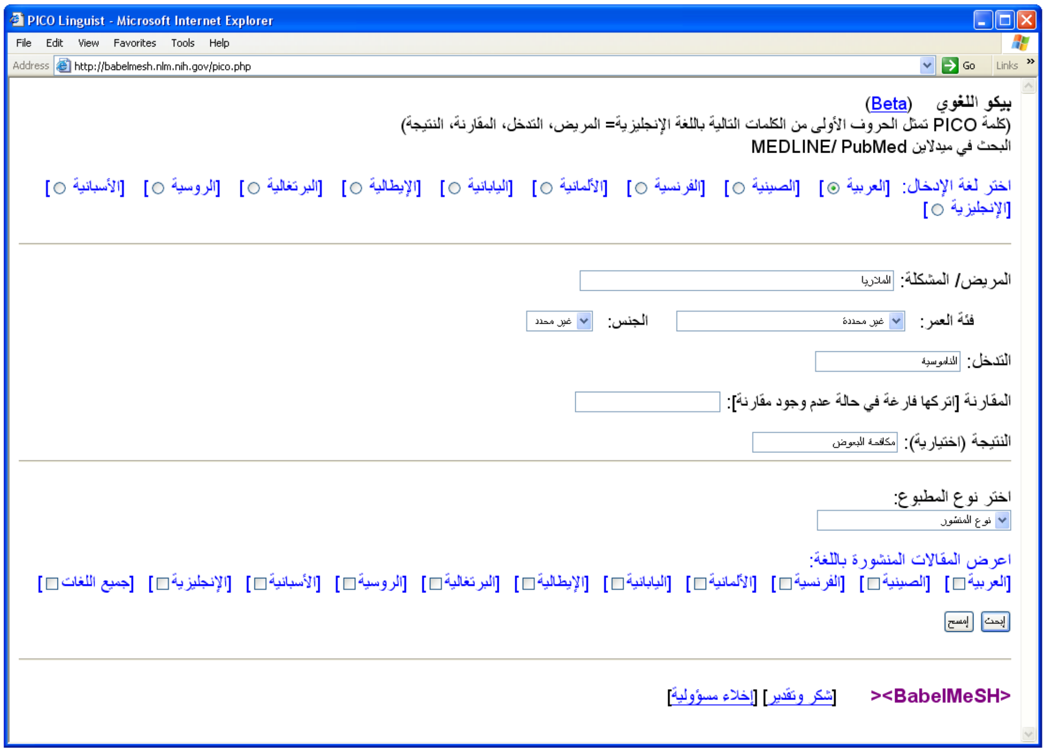


Figure 4. PICO Linguist interface in Arabic

Citation Result Filter

The user can select to view the titles of citations by language of publication. An output filter can display the results in one or more publication languages. The default is for articles published in all languages in PubMed journals. If a filter of publication language other than English is selected, the title of citation results will be shown in both English and the original language. However, this feature will only work for Latin languages. PubMed does not record original titles in character-based language, such as Arabic, Chinese, Japanese, and Russian.

BabelMeSH and PICO Linguist transmit the translated query to PubMed through E-Utilities. Retrieved citations are shown in English only because citations of foreign language journals indexed in MEDLINE are in English only. Links are provided to the foreign journal if published online. Journals may require subscription for full-text access. Search terms in the result's page are highlighted for the user's easy identification. Figure 5 demonstrates the result page of a query in Arabic.

Evaluation

The full list of BMJ Clinical Evidence systematic reviews7 was translated into Spanish by one of the authors (SL). These terms were searched in BabelMeSH then compared to the Spanish translations. The accuracy of BabelMeSH's English translation was compared to the original list. User opinion was obtained through a Web form. The evaluation of the French version consisted of three parts: (1) comparison of the author's own translation of keywords to English in journal articles published in French and the translations of the same terms to English by BabelMeSH; (2) actual user search terms from the our Web server's log files; (3) online user feedback questionnaire using a 5-point Likert scale to evaluate user opinion on the usefulness of BabelMeSH.

Results

Figure 5. Retrieval of a PICO Linguist search in Arabic

The search strategy in Figure 4 retrieved results shown below in Figure 5. Table 3 shows the evaluation results of the French and Spanish versions of BabelMeSH and PICO Linguist. The two test sets used for evaluating the French version journal keywords and user query terms from server logs showed an accuracy of 75% and 79.9% respectively (mean=77.5%). Partial matches were obtained in 13.2% and 7.3% respectively. Full term matches and partial term matches equal to 88.2% and 87.2% of the test sets respectively. There were no matches in approximately 12% for both sets.

	French		Spanish
	Journal keywords	User query from server log	
Total number of terms	174	179	221
Exact or suggestions match	130	143	191
% Total	75%	79.9%	86.4%
Partial match	23	13	20
% Total	13.2%	7.3%	9.0%
Incorrect translation	21	23	10
% Total	12%	12.8	4.5%

Table 3 – Evaluation of the French and Spanish versions

One hundred ninety one of the 221 terms from the disease list had exact matches for the Spanish version or 86.4% accuracy. For 20 terms (9%) partial matching was found but some translations missed one or two key words that could negatively affect a search. No translations were found for 10 (4.5%) search terms. We analyzed the complexity of search term or phrase as a measure of the parsing algorithm. The results (Table 4) show that accuracy is inversely related to the number of terms used.

Of the six responses received from the French feedback form, the average ratings (5=agree, 1=disagree) for the following statements were: 1) that BabelMeSH was useful 4.3/5; 2) the overall quality of citations retrieved was excellent 4.3/5; and 3) that they would continue to use BabelMeSH 4.6/5. All stated that they had previously searched MEDLINE in English and all except one declared that they would recommend it to others.

Number of terms	Total	Translated	% Total
1	104	85	81.7
2	150	69	46
3	67	11	16.4
4	24	2	8.3
5	6	1	1.7
6	2	0	0

Table 4 - Effect of search term count on translation accuracy

Comments from the Spanish users were mostly positive. Figure 6 shows the results of the two questions asked: (1) is searching MEDLINE/PubMed in your native language useful? and (2) how would you rate the results obtained? The mean responses were 4.3 and 3.9 respectively in a 5-point Likert scale.

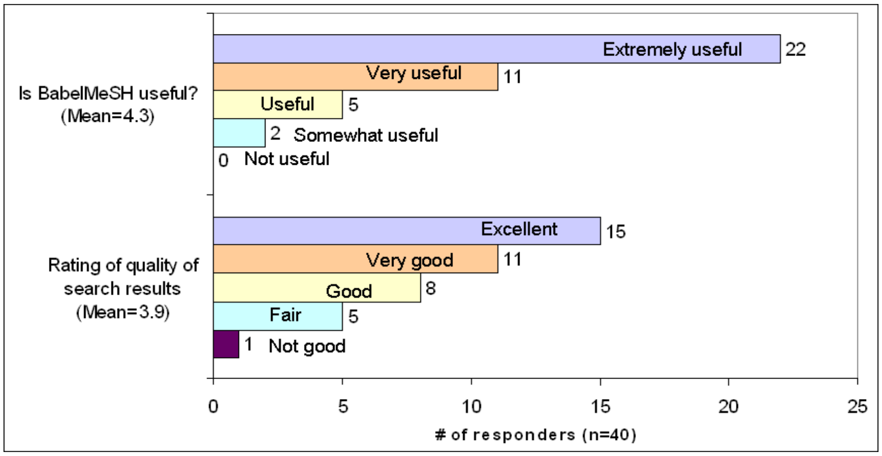


Figure 6 – User feedback of the Spanish version of BabelMeSH

Discussion

The aim of this project, as in most medical translations is to bring out the same cognitive "equivalent effect" that an individual might be searching for in MEDLINE/PubMed. This is challenging in one language, even more when dealing with many languages as we have attempted here. The quality of sources is also a major determinant of success, one that we have no control over.

The evaluation of PICO linguistic and BabelMeSH demonstrates the usual challenges of medical translation due to the highly technical and scientific nature of medical language. The parsing algorithm is an attempt to overcome Newmark's "transparent collocation" phenomenon [1] as well as grammar. This is quite a challenge as shown in Table 4 where translation success drastically plummets as the complexity of the search term increases.

During development, we also found that the simple change in use of singular or plural nouns adversely affected the result of a query. This was common among all language versions. This is related to the MeSH translation in other languages, some translations only include descriptors but not all the valid entries. For example, the MeSH heading, "heart diseases", has 8 valid records, such as "heart disease" and "cardiac disease". For this medical concept, our database has differences among languages: six records in French, five in German, two in Japanese and one record each in Italian and Russian. If "heart disease", instead of "heart diseases", in Italian or Russian is entered, there will be no translation result. We have done some machine-normalization work on plural and non-plural words in those foreign languages, but it could also bring errors. MeSH contains more than 22,000 descriptors, but it also has greater than 130,000 additional valid entries that can help find the appropriate MeSH heading.

Table 3 shows the results of the evaluation of the French and Spanish versions. Using two different test sets, the results in French are comparable. Contrasted with the Spanish version the accuracy of translation was higher in Spanish, although this is really not a one-on-one comparison since they use entirely different databases. This is also likely related to the size of the database, the Spanish database contains 788,835 entries while the French only 190,330. We were unable to compare translation accuracy with other medical translation studies. Cross-language reports such as the one done by Volk [2] evaluated their algorithms using precision and recall from MEDLINE searches. In this project we opted to use keywords and disease list to maximize the number of terms tested. It would be difficult to precision and recall studies. We chose user satisfaction questionnaires to evaluate real-world searches.

Usability approval was encouraging. Thirty-eight responders in Spanish said that BabelMeSH was 'useful' to 'extremely useful' while only two found it 'somewhat useful' and none responded that it was not useful. The high rating on the quality of search results directly supports the cognitive equivalent effect of translations. This would have been possible only with accurate translations. Almost all of the 40 responders submitted comments that were even more enthusiastic like, "I've been waiting for it" and "There's nothing like mother language." There were also comments that expressed a preference to be able to read the citations and abstracts in their native language. The French evaluators were equally enthusiastic about these resources. The highly positive feedback obtained from French and Spanish users showed us that there is need for multilanguage resources and the medical community may be eager to adopt utilities like PICO Linguist and BabelMeSH in their daily activities. We suspect the same type of response from users in other languages not tested in this phase of the project.

Medical translation in not a simple matter, nevertheless, it is very important to have tools available for immediate use at the point-of-care when the healthcare professional is looking for current evidence in the field.

Challenges and developments

This resource is only as good as its database. It is a product of international collaboration and we invite collaborative research to make it even better. We also invite evaluation studies especially languages that are not covered by this report. The databases will be updated yearly as new editions of UMLS are released. Non-English translations of the abstract would be desirable. We will explore possible solution to this goal.

We previously evaluated the accuracy of BabelMeSH translations by term alignment of parallel lists an manual review. We wanted a more quantitative method that would take advantage of MeSH descriptors and entry terms (synonyms, near synonyms, and closely related concepts) used to automatically map searches in MEDLINE/PubMed databases. The goal was to approximate human review using automatic clinical weighting methods.

Conclusion

PICO Linguist and BabelMeSH provide templates for multilanguage search tools including character-based languages. They can be alternative search resources for searching current evidence in the medical literature for whom English is not their primary language. We invite collaborative research especially in improving its multilanguage database and evaluating its potential.

References

- Newmark P. A layman's view of medical translation. Br Med J. 1979 Dec 1;2(6202):1405-7.
- Martin Volk, Bärbel Ripplinger, Spela Vintar, Paul Buitelaar, Diana Raileanu and Bogdan Sacaleanu: Semantic Annotation for Concept-Based Cross-Language Medical Information Retrieval. In: International Journal of Medical Informatics. 67(1-3). 2002. 97-112.

10.4 Appendix D – txt2MEDLINE

txt2MEDLINE Architecture

A TER-GX101 TriBand (900/1800/1900 MHz) GMS modem (Round Solutions Ltd) connected to a Linux computer (Red Hat Enterprise) comprises the Txt2MEDLINE server. A Subscriber Identity Module (SIM) card provides wireless connectivity to the AT&T mobile phone network. UltraSMS (http://kinks.ultralab.ac.uk/ultrasms) interfaces between the GSM modem and MySQL database.

The inbound/outbound traffic flow is illustrated in Figure 1: (1) The mobile device sends a text message to NLM's GSM modem (240-461-7765) through a wireless carrier. (2) The SMS center processes the incoming message and forwards it to PubMed. (3) Using PubMed's E-utilities, message abbreviation and "the bottom line" (TBL) algorithm, the journal citation is retrieved and processed to the SMS format. (4) SMS center sends out the text message through the GSM modem or sends to the user as e-mail. (5) Mobile phone receives search result as a text message.

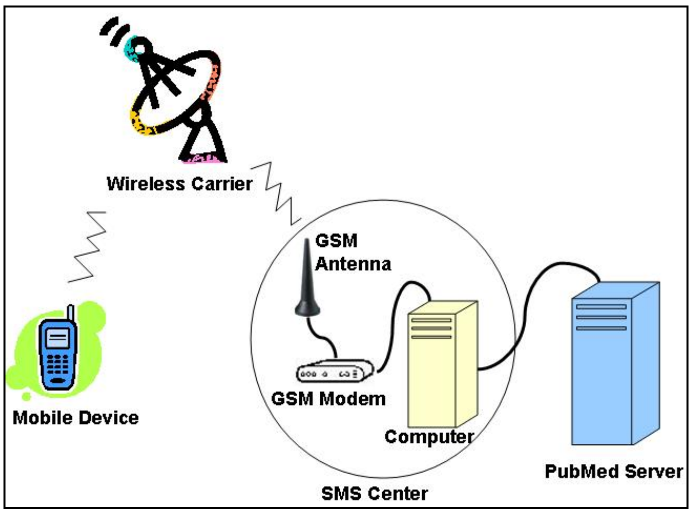


Figure 1. txt2MEDLINE Architecture

Abbreviation and Acronym Database

Using MySQL, we created an abbreviations and acronyms database. It currently contains approximately 3000 medical terms (see example in Table 1). We selected the most frequent usage of an abbreviation or acronym in medicine when ambiguity occurred. The database is continuously growing and users are encouraged to submit additions.

Abbreviation/Acronym	Formal Term
CTS	carpal tunnel syndrome
MH	malignant hyperthermia
w/	with
inj	injection

Table 1. Examples of Abbreviation and Acronyms in the Database

Incoming Message Processing

Queries are initiated by sending a text message to the txt2MEDLINE GSM modem. UltraSMS allows the GSM modem to communicate with the MySQL database.

Search commands are always in upper case text followed by a question mark (Table 2).

Search Commands	Interpretation
S?	Search
SR?	Send result to mobile device
M?	Send result by e-mail
LR?	Limit result (default=1)

Table 2. Common server commands and meaning

A sample search query format is shown in the box below:
S?cts surg vs steroid inj rct M? username@userhost.com LR?3

The server interprets this query as:
Search for 'carpal tunnel syndrome'; compare 'surgery' versus 'steroid injection'; retrieve only 'randomized controlled trials' publication types; send results by e-mail to 'username@userhost.com'; limit results to 3 articles.

Queries may be submitted through messaging services that can send text messages on a computer to mobile phones, such as, AOL Instant Messenger and Yahoo Messenger. The search may be done also by sending an e-mail message through a message service that can convert an e-mail message to a text message to a mobile phone. The txt2MEDLINE GSM modem can receive data directly from mobile phones.

Query Processing

Queries can be sent as a combination of abbreviations and formal terms. Each input term in the query is searched in the abbreviation database. If this input term is found to be an abbreviation, its corresponding formal term will be used in the query. Only formal terms are submitted to E-utilities unless it is an abbreviation recognized as a MeSH term.

The translated query is sent to PubMed through E-utilities, which are NCBI (National Center for Biotechnology Information) resources for accessing PubMed.

Outgoing Message Processing Algorithm

To decrease the size of the original data from E-utilities, we developed a text-abbreviating algorithm that reduces the size of outgoing messages significantly. Its two components are TBL algorithm and the word transformation algorithm.

The TBL Algorithm:

If the abstract is structured or if it contains the word 'conclusion', the segment, sentence or phrase that follows, will be returned as the TBL. Approximately 9% of citations with abstracts published in PubMed after 1995 are structured or contain the word 'conclusion' in their abstract. All MEDLINE records are included, but citations without abstracts are disregarded.

If no 'conclusion' is found in the abstract, the TBL algorithm parses the journal abstract into sentences identified by punctuation marks (period or question mark.) The process then proceeds as follows: (1) all terms from a 'stop words' list (PubMed's list and our own compilation) are deleted. (2) A frequency count of the remaining words is made. The top five most frequent words are considered as the key words of the abstract. (3) The sentences in the abstract are then ranked by the frequency of the occurrence of key words. The sentence with the most number of key words present and the last two sentences of the abstract (if they are not the sentences with the most number of key words) will be selected for 'the bottom line' summary. The rationale for choosing the last two sentences is that it often contains significant relevant information useful in summarizing the key points of the abstract. We are currently undergoing a formal evaluation on the validity of these assumptions, although we have found this to be accurate by informal evaluation.

Word transformation Algorithm:

The use of abbreviations and acronyms in current medical literature is common. In developing the database, we were guided by previous studies of methods for systematically abbreviating words and names [1-5]. In previous methods, vowels are regarded as redundant, [4-6] so we eliminated all vowels in the first version of the algorithm. However, the more vowels are deleted, the more difficult it is to reconstruct the original information. We therefore designed a word transformation algorithm that reasonably sets the threshold and systematically deletes some redundancy in each word. A trial period allowed refinement of the rules. These rules are still in flux and we continue to revise them based on our experience and as feedback is received.

In general, words with 4 letters or less are not truncated. All consonants are retained. If a word with more than 4 letters is not found in abbreviations database, all vowels are deleted, except when the vowel:
(1) is the first letter of the word.
(2) is the first vowel after the first letter of the word.
(3) if vowels occur in tandem, the first vowel is retained.

Other Optimization Procedures

Only the last name of the author is retained. When there are multiple authors, a '+' sign is appended immediately after the first author's last name.
Some of MEDLINE's abbreviated journal titles are further shortened. Abbreviated forms of highly accessed clinical journals are stored in a database. For all other journals, spaces between words in the title are deleted.
Publication dates are truncated by using the first two letters of the month and last two digits of the year.

Registration Service

An optional registration service is available to the user. An active authentication step requires sending a verification message from the same phone used to register the account. The user ID is sent thereafter.

This feature allows the user to send a shorter query. Upon registration, a 5-character account ID will be generated randomly. This account will be associated with a user's mobile phone and e-mail address. Users may then use this account ID to send queries instead of the regular 10-digit phone number and/or e-mail address. This will be especially convenient when users want the results sent to both a mobile phone and by e-mail. For example, instead of sending a message as "S?mh rct M?username@userhost.com", a registered user can simply send the message, "S?mh rct M?123ab", where "123ab" is the user's account ID. The results will be sent to the mobile phone and e-mail address.

Message Truncation Evaluation

We selected key words from the top 10 most searched terms from askMEDLINE to evaluate the system's citation truncation function. Ninety-five citations were tested.

Results

Figure 2 shows an example of text messages received from txt2MEDLINE. Two text messages were sent for this citation. The user would need to scroll up and down to read the entire message. The query in this example was: "S?warts duct tape rct". The result shows one randomized controlled study comparing duct tape occlusion to cryotherapy in the treatment of the common wart.

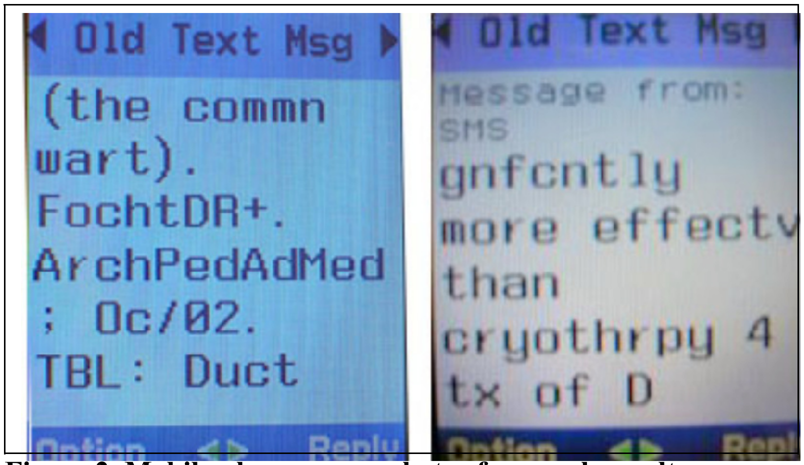


Figure 2. Mobile phone screen shots of a search result.

Figure 3 and 4 illustrate the results of the message-abbreviating algorithm. The original citation (Figure 3) contains 1783 characters. Since the SMS protocol forces a 160-character limit per message, 12 text messages would be required to send the complete citation. The message-abbreviating algorithm and the TBL algorithm reduced the original citation to 216 characters (Figure 4), requiring only two text messages.

A randomized controlled trial of surgery vs steroid injection for carpal tunnel syndrome. Hui AC, et al. Neurology; 2005 Jun. BACKGROUND: Decompressive surgery and steroid injection are widely used forms of treatment for carpal tunnel syndrome (CTS) but there is no consensus on their effectiveness in comparison to each other. The authors evaluated the efficacy of surgery vs steroid injection in relieving symptoms in patients with CTS. METHODS: The authors conducted a randomized, single blind, controlled trial. Fifty patients with electrophysiologically confirmed idiopathic CTS were randomized and assigned to open carpal tunnel release (25 patients) or to a single injection of steroid (25 patients). Patients were followed up at 6 and 20 weeks. The primary outcome was symptom relief in terms of the Global Symptom Score (GSS), which rates symptoms on a scale of 0 (no symptoms) to 50 (most severe). Nerve conduction studies and grip strength measurements were used as secondary outcome assessments. RESULTS: At 20 weeks after randomization, patients who underwent surgery had greater symptomatic improvement than those who were injected. The mean improvement in GSS after 20 weeks was 24.2 (SD 11.0) in the surgery group vs 8.7 (SD 13.0) in the injection group (p < 0.001); surgical decompression also resulted in greater improvement in median nerve distal motor latencies and sensory nerve conduction velocity. Mean grip strength in the surgical group was reduced by 1.7 kg (SD 5.1) compared with a gain of 2.4 kg (SD 5.5) in the injection group. CONCLUSION: Compared with steroid injection, open carpal tunnel release resulted in better symptomatic and neurophysiologic outcome but not grip strength in patients with idiopathic carpal tunnel syndrome over a 20-week period.

Figure 3. Message before Abbreviating Algorithm

> A RCT of Surg vs sterod inj 4 CTS. Hui AC+. Neuro; Jn/05. TBL: Comprd w/ sterod inj, open carpl tunnl reles resltd N bettr symptomtc & nerphysilgc outcm but not grip strength N pts w/ idipthc CTS over a 20-week perid.

Figure 4. Message after Abbreviating Algorithm

A review of 95 citations to evaluate the message truncation algorithm showed an average of 1658±491 characters per citation in the original citations. The average size after message truncation was 352±114 characters (Figure 5); only 2.2 or about 3 messages would be needed, down from the 10.4 messages per citation initially. Message abbreviating algorithms decreased average message size by 77.5 ±7.9% characters. In most parts of the world, wireless providers charge their users only for sending text messages but not for receiving them. However, it becomes difficult to read abstracts when too many SMS messages are received, especially since there is no assurance that they will be received in the correct order.

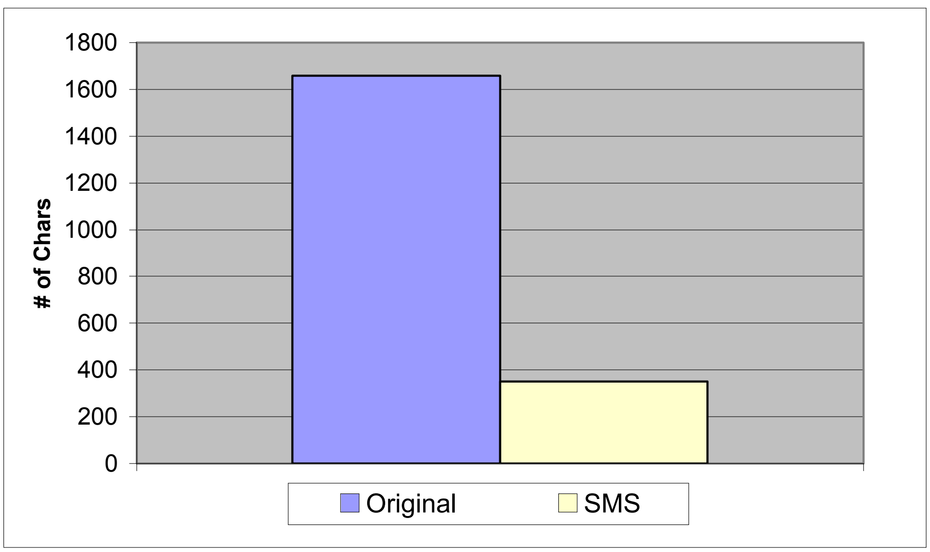


Figure 5. Message Size Comparison: Original Citation vs. SMS

We also explored alternate methods for submitting queries to Txt2MEDLINE, other than text messaging by mobile phone. We found that sending e-mail to accounts that will convert e-mail to SMS, AOL Instant Messenger and Yahoo Messenger were successful in submitting queries to Txt2MEDLINE. Multiple tests showed that the average turnaround time was two minutes or less.

Discussion

Evidence-based medicine and translational medicine depend on convenient access to knowledge sources at the point of care. These resources must be easily accessible and handy. Bottom-line statements are recommended.

Two existing conditions favor the alternative access to reference sources discussed in this paper: 1) increased utilization of SMS or text messaging by doctors and healthcare personnel, and 2) doctors' inclination to using abbreviations and acronyms. The inclination to abbreviations and acronyms starts early in medicine through the use of mnemonic aids in anatomy. The hurried pace during medical school and residency training requires further quick note taking. We have taken advantage of these tendencies with the hope of encouraging the practice of evidence-based medicine through mobile devices.

We initially developed only the abbreviations and acronyms database along with the GSM modem, but it was immediately clear that further truncation of the journal citation was needed to reduce message size. Although cost is a major factor, ease of access and convenience were the prime motivators. Although there might be a need for it, doctors will not seek references if they are difficult to obtain.

Modifications were needed with UltraSMS because of variations in wireless companies' handling of symbols and characters. The European Telecommunications Standards Institute (ETSI) specifies the 7-bit alphabet as the default alphabet for SMS, however variations exist between wireless carriers. The character "@" was the most problematic because it is essential for sending results by e-mail. It required multiple modifications so users can make use of the SMS-to-e-mail method for sending results.

Although real-time messaging services have been successfully tested, they do not work consistently. When a real-time messaging service is used to send a query, it is necessary to specify where the results will be sent, because the system is unable to decode message headers from these services. These headers also vary between wireless telecommunications providers and no information is available from the message headers on the identity of the sender.

Several versions of the word abbreviation algorithms are being tested. These modifications will continue based on in-house testing and feedback from users. We are also evaluating the validity of the TBL summary. Feedback from users will guide future modifications of the algorithm

txt2MEDLINE's architecture could be duplicated in other environments. A local system within a region or country could save on long distance toll charges for those in overseas locations. Queries can be sent to MEDLINE through the Internet.

"The Bottom Line" (TBL) summary

txt2MEDLINE allows users to search and retrieve MEDLINE/PubMed citations using SMS text messages. "The Bottom Line" (TBL) concept was developed because of the 160-character limitation of text messages. TBL algorithm generates a shortened version of the published abstract thereby decreasing the number of characters while attempting to maintain the key points of the full-length, author-generated abstract. The data seems to indicate that the TBL generally conveys the essential elements of the full abstract.

The Bottom Line (TBL) project evolved from txt2MEDLINE, a search tool for MEDLINE/PubMed using SMS text messages. The TBL algorithm was developed because of the 160-character limit of the SMS protocol. A system was needed to reduce the number of text messages that each citation generated. The purpose of this project was to determine whether TBL conveyed the key elements and substance of the full abstract.

Evaluation

The evaluation was divided into three phases: NLM personnel only, National Institutes of Health clinical librarians and national participants. A Web search page was created that allowed the evaluators to compare the results showing author-generated abstracts and TBL format on the same page. On a rating tool embedded on the results page, participants determined on a 5-point Likert scale (5=strongly agree, 1=strongly disagree) for each abstract, whether they agreed to the statement, “The TBL provides the 'bottom line information’ of the abstract”.

Results

The twenty-six evaluators who responded to the call for participation, number of responses and ratings are summarized in Figure 6. They were requested to search “clinical topics” and review at least five abstracts each. The overall mean score was 3.8 (mode= 4) of the 329 TBL abstracts reviewed. Table 3 summarizes the results by phase of study. The TBL rating was a ‘4’ or ‘5’ for 218 abstracts (66% of all citations reviewed.) Analysis of search terms showed mean ratings of 3.87, 2.3, 3.5 and 3.88 for clinical topics (n=32), library related (n=4); management related (n=4) and unclassified topics (n=17) respectively.

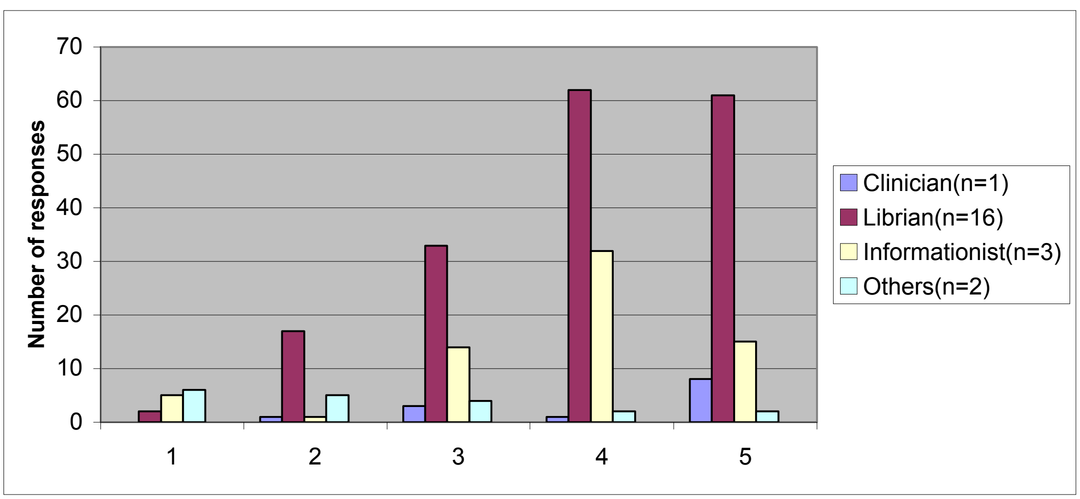


Figure 6. Agreement between TBL and full abstract among evaluators

	Phase 1	Phase 2	Phase 3	Total
Evaluators	5	4	17	26
Searches done	9	8	40	57
Abstracts Rated	79	35	215	329
Mean Rating	3.98	3.65	3.75	3.8
Mode	**4**	**4**	**5**	**4**

Table 3. Summary of evaluators, searches, ratings by phase

The mean rating comparing 'TBL summaries with 329 published journal abstracts was 3.8 in a 5-point Likert scale. Our data seems to provide support that a majority of the TBL's reviewed convey or maintain the key elements of the full abstract. A similar review among clinicians, who will be the main consumers of this tool, may be needed to validate the results of this study.

Conclusion

txt2MEDLINE is an alternative method for searching MEDLINE/PubMed using mobile devices. It is convenient and fast. The service should be available from most areas where a wireless network is available. The SMS center based on a GSM modem, UltraSMS and a Linux computer works effectively for processing queries and retrieving results from MEDLINE/PubMed. The registration process allows the user to send shorter queries with a common ID for mobile devices and e-mail. The abbreviation and acronym database of common medical terms, TBL algorithm and message-abbreviating algorithm transform full-length journal citations into short text messages, suitable for wireless mobile devices. Early user feedback is positive. Refinements of the various algorithms and evaluation of the validity of the TBL summary are continuing.

References

- Charles P. Bourne , Donald F. Ford, A Study of Methods for Systematically Abbreviating English Words and Names, Journal of the ACM (JACM), v.8 n.4, p.538-552, Oct. 1961
- Ostojic V, Cvoriscec B, Ostojic SB, Reznikoff D, Stipic-Markovic A, Tudjman Z. "Improving asthma control through telemedicine: a study of short-message service". Telemed J E Health. 2005 Feb; 11(1):28-35.
- Liu H, Lussier YA, Friedman C. "A study of abbreviations in the UMLS." Proc AMIA Symp. 2001;:393-7.
- E. Shannon, "A Mathematical Theory of Communication", The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, July, October, 1948.
- Francis Knowles, "Information Theory and its Implications for Spelling Reform", Simplified Spelling Society Newsletter, J2, pp5-13, Spring 1986.
- June A. Barrett, Mandalay Grems, Abbreviating words systematically, Communications of the ACM, v.3 n.5, p.323-324, May 1960.