Title

Searching the Literature Using Medical Subject Headings versus Text Word with PubMed

Journal

The Laryngoscope, 116(2)

Authors

Heskett, Karen M
Chang, Angela A, MD
Terence M, Davidson, MD

Publication Date

2006-02-01

Peer reviewed

Searching the Literature Using Medical Subject Headings Versus Text Word with PubMed

Angela A. Chang, MD[1]

Karen M. Heskett, MSI[2]

Terence M. Davidson, MD[3]

[1]University of California, San Diego, Otolaryngology-Head and Neck Surgery
[2]University of California, San Diego, Biomedical Library
[3]University of California San Diego, Otolaryngology-Head and Neck Surgery and VA San Diego Healthcare System

San Diego, CA, USA

Send correspondence to:

Terence M. Davidson, M.D.
Professor of Surgery, Otolaryngology- Head and Neck Surgery
Associate Dean, Continuing Medical Education
University of California, San Diego School of Medicine
VA San Diego Healthcare System

9500 Gilman Drive, MC 0617
La Jolla, CA 92093-0617
E-mail: tdavidson@ucsd.edu
Phone: 858-822-4229
Fax: 858-822-5908

ABSTRACT

Objective/Hypothesis: To investigate the performance of two search strategies in the retrieval of information from the National Library of Medicine (NLM) on otolaryngology-head and neck surgery related conditions and diagnoses using PubMed.

Methods: Two search strategies—one based on the use of Medical Subject Headings (MeSH), the second based on text word searching—were compared.

Results: The MeSH search provided a more efficient search than the text word search.

Conclusions: Head and neck surgeons can most efficiently search the NLM using PubMed as a search engine by initiating the search with MeSH terms. Once a key article is identified, the search should use the "Related Articles" feature.

Keywords: PubMed (MeSH), Medical Subject Headings (MeSH), Otolaryngology (MeSH)

INTRODUCTION

MEDLINE is a comprehensive online database of biomedical literature maintained through the National Library of Medicine (NLM). As of early 2005, the database contained over 13 million articles, dating as far back as 1966, with over 4,800 biomedical journals originating from the United States and 70 other countries. In 1997, the NLM began providing free, unlimited internet access to MEDLINE through a search engine called PubMed [1]. Widely used by the medical community, PubMed offers several tools that help the user define a medical search.

There are two ways to approach PubMed searches. One approach is to conduct a very broad search, retrieving a maximum number of potentially relevant articles with the intention that no articles of importance be missed. The other approach is to conduct a more finely focused search with the aim to efficiently identify the most relevant publications. This latter search strategy is most commonly employed for shorter, day-to-day searches, in which one would like to look up something new or unusual--the type of search which many clinicians perform on a regular basis. Searchers should choose the approach more appropriate to their needs.

Search queries in PubMed can be made with text word or Medical Subject Headings (MeSH). In a text word search, PubMed scans the whole record of the article – title, abstract, list of applied MeSH terms, list of authors, and journal name. Articles containing queried text words in any one of these areas are retrieved. As a result, text word searches often identify articles not pertinent to the search topic. For example, a text word search on *gentamicin ototoxicity* would not only identify all related articles, but would list every article with the word *gentamicin* or *ototoxicity*. Additionally, all possible synonyms and variants of a term must be specified in order to ensure a comprehensive result [2].

Medical Subject Headings

PubMed indexes journal articles using MeSH. The MeSH database is a powerful tool for performing a focused search using MeSH headings and subheadings. MeSH terms constitute a thesaurus that embodies all the concepts appearing in the medical literature. Table I is a list of the broad categories used for medical subject headings. MeSH terms are arranged in a hierarchical, treelike structure by subject categories. Table II shows a very limited sampling of the MeSH tree for *Ramsay Hunt syndrome* or *Herpes zoster oticus*. All scientific articles are indexed using an average of 10 to 12 MeSH terms. There are currently over 13,000 main MeSH terms [3]. Associated with MeSH is a list of 80 possible concept terms, e.g. diagnosis, drug therapy, surgery, etiology, etc. Concept terms are used as subheadings to ~~better~~ enhance the focus of MeSH searches. Table III shows a selection of these terms for nasopharygneal carcinoma. Searches can be further modified by using limits, which are easily applied using the "Limits" tab under the search box on the main PubMed page. These will limit searches by language, age group, gender, publication type, etc [4,5].

After a search has been performed, there is a link called "Related Articles" that appears to the right of each identified article. PubMed retrieves related articles with the use of a pre-specified, word-weighted algorithm assimilating the MeSH and text words from the title and abstract of the selected citation. Whereas MeSH and text word searches are displayed in chronologic order, "Related Articles" citations are arranged from the most to the least relevant, as measured by the similarity in MeSH terms and the number of visits the article has had by other searches[6]. Therefore, if a search identifies an article that seems to be ideal, clicking on the "Related Articles" link generates a list of other articles with similar MeSH descriptors, which is now displayed not chronologically, but by relative relevance. The MeSH search for obstructive sleep apnea surgery is shown in Figure 1.

MATERIALS AND METHODS

MEDLINE searches were performed via PubMed. We used the PubMed tutorial [7] to get acquainted with the search engine, but practice with an experienced searcher was helpful in obtaining proficiency in our searches. Two search strategies were developed. In the subject search strategy, only MeSH terms were used to locate the relevant studies, while in the text word strategy the same words were searched as text words. We compared searches for ten different, arbitrarily selected otolaryngology-head and neck surgery conditions and diagnoses by comparing the number of retrieved articles for both MeSH and text word searches, with the assumption that the search obtaining the fewer number of articles was a more efficient way to initiate a search. The number of articles found by each search strategy was noted, and the percentage difference was calculated by dividing the difference in article retrievals by the number of retrieved text word articles.

We also conducted an in-depth review of one search and studied the abstracts of articles found by the text word search strategy but missed by MeSH. Articles that showed relevance based on their abstract alone underwent further evaluation of their full text. We then concluded whether these articles were pertinent to the aim of the search. The number of relevant articles missed by MeSH was recorded, and the percentage of missed articles was calculated by dividing the number of missed articles by the number of missed relevant articles.

RESULTS

We reviewed ten common otolaryngology topics and compared the results using MeSH and text word terms. The search results are displayed in Table IV. MeSH searches retrieved fewer journal articles. Simple, but not otherwise quantitated, review of each search showed that the text word searches had more extraneous articles than the better focused MeSH searches.

In our literature search on congenital vocal paralysis, we studied the articles retrieved by MeSH and text word strategies. The MeSH strategy retrieved 37 articles, compared with 113 articles retrieved by the text word strategy. We then reviewed the 76 articles that were not found by the MeSH strategy. Of the articles not retrieved by MeSH, only two articles (both from the 1970's) were selected for further examination based on a review of their abstracts. After reading their full text, both articles were determined to be relevant to the search. Therefore, 2.6% of the articles missed were articles pertinent to the search. Of note, the majority of the irrelevant articles retrieved by the text word search included articles which focused on various congenital conditions, such as congenital heart diseases or neurological disorders. When the "Detail" tab was selected to show the search algorithm of how the articles were retrieved, many of these articles were found because the word "congenital" was associated in the title or abstract.

DISCUSSION

As the indexing vocabulary of MEDLINE and other important NLM databases, the MeSH thesaurus is a powerful tool for improving access to the rapidly expanding biomedical knowledge base. MeSH is organized into a complex hierarchy called the "MeSH tree structures," in which the MeSH terms are arranged into a set of branching treelike structures of increasing specificity. It appears that the original system was appended upon multiple occasions, but never revised. Hence as time went on, additional categories and subcategories

were added. If one examines the MeSH tree on a piece of paper, it is intertwined and makes little sense, but as the computer navigates the PubMed system, just like the internet navigates different nodal paths to find its ultimate destination, the process works. The hierarchical structure allows one to search at various levels of specificity.

In this paper, we compared the performance of two MEDLINE search strategies in searching otolaryngology-head and neck surgery related diagnoses and conditions using text word and MeSH searches. We found that MeSH searches retrieved fewer and more relevant journal articles than text word searches. In addition, we discovered that ultimately there was improved efficiency in using MeSH, as many terms are "mapped" to the MeSH term when one uses the PubMed MeSH database. After typing in the word(s) for the search, the system automatically links the preferred term. For example, *tongue cancer* maps the searcher to *tongue neoplasm*, and *Ramsay Hunt* maps to *Herpes zoster oticus.* This differed from the text word searches, in which the searcher had to simply try different words to conduct a search.

The comparison of MeSH versus text word searches is not new. Authors have compared literature searches performed using either text word or MeSH terms and evaluated the outcomes of their searches on the basis of their sensitivity and specificity. Sensitivity is defined as the ability of the search to retrieve relevant citations from the database. It can also be viewed as the recall, or the number of relevant citations retrieved divided by the number of relevant citations in the database. Specificity, or the ability of the search to discriminate between relevant and nonrelevant citations, can be viewed as the precision of the search. Precision is the number of relevant citations retrieved divided by the total number of citations retrieved. Therefore, high sensitivity of a search signifies that important articles were identified. High specificity indicates that the review was efficient and that few unimportant articles were retrieved.

Studies show that text word searches have lower specificity than searches performed using MeSH. In a study comparing MeSH and text word searches to retrieve studies on sleep, the authors found that the MeSH search produced higher specificity than the text word search (66% and 47%, respectively), but lower sensitivity (78% for the MeSH search versus 88% for the text word search) [8]. Additionally, an unpublished study of 975 MEDLINE searches at Harvard Medical School showed that MeSH searches provided significantly higher specificity and sensitivity than title-abstract text word searches [4].

Others have argued that the sensitivity and specificity of searches are enhanced when searches combine both text words and MeSH words. In one study which investigated search strategies of clinical studies in general adult medicine, the sensitivities of searches were enhanced by combining MeSH terms and text words when retrieving articles. In addition, the authors found that although the specificities of individual MeSH terms and text words were high, this was often at the expense of their sensitivity. When combining terms, however, the high specificity was maintained with modest increases in sensitivity [9].

CONCLUSION

Skill using PubMed to search the NLM is important to retrieve accurate, up-to-date medical information. It is strongly recommended that searchers use MeSH tools to conduct their search. A disadvantage of MeSH and text word searches is that articles are displayed chronologically. It is therefore recommended that once the search is conducted, one identifies pertinent articles and selects, “Related Articles.” This activates a second search using the MeSH terms assigned to the identified article and displays results in order of their relevance to the original article, as measured by similar MeSH and text word terms. Figure 2 shows the

suggested algorithm for literature searches. MeSH should be used to retrieve relevant articles efficiently and the "Related Articles" link can then be used to further focus the search. In the rare occasion that the broadest possible search is required or when a search involves unusual conditions, text word searches are an important, complimentary browser paradigm.

Lastly, it is recommended that medical publication authors use MeSH terms as their key words and that they identify them as MeSH terms. Key words are not searched even in text word searches, so using a word like *allergy* instead of the MeSH term *hypersensitivity* is non-productive.

Using the MeSH vocabulary requires some learning; however, the effort put into familiarization with the MeSH vocabulary will pay off in improved search quality. The more skilled one becomes in their searches, the more efficient and precise will be the results.

REFERENCES

1. MEDLINE Fact Sheet [National Library of Medicine web site]. Available at: http://www.nlm.nih.gov/pubs/factsheets/medline.html. Accessed August 15, 2005.
2. Stewart MG, Kuppersmith RB, Moore AS. Searching the medical literature on the Internet. *Otolaryngol Clin of N Am* 2002; 35:1163-1174.
3. *Medical Subject Headings: Annotated Alphabetic List 2003*. Bethesda, MD: National Library of Medicine; 2003.
4. Lowe HJ, Barnett GO. Understanding and Using the Medical Subject Headings (MeSH) Vocabulary to Perform Literature Searches. *JAMA* 1994;271(14):1103-1108.
5. Coletti MH, Bleich HL. Medical Subject Headings Used to Search the Biomedical Literature. *JAMIA* 2001;8(4):317-323.
6. Sood A, Erwin PJ, Ebbert JO. Using Advanced Search Tools on PubMed for Citation Retrieval. *Mayo Clin Proc* 2004;79(10):1295-1300.
7. PubMed Tutorial. [National Library of Medicine web site]. Available at: http://www.nlm.nih.gov/bsd/viewlet/mesh/searching/mesh1.swf. Accessed August, 15, 2005.
8. Jenuwine ES, Floyd JA. Comparison of Medical Subject Heading and text-word searches in MEDLINE to retrieve studies on sleep in healthy individuals. *J Med Libr Assoc* 2004;92(3):349-353.
9. Haynes RB, Wilcynski N, McKibbon KA, Walker JC, Sinclair JC. Developing optimal search strategies for detecting clinically sound studies in MEDLINE. *JAMIA* 1994;1(6):447-58.

Table I. Categories for Tree Structure 2003. Excerpted from NLM, 2003

Category	
	Anatomy
	Organisms
	Diseases
	Chemicals and Drugs
	Analytical, Diagnostic and Therapeutic Techniques and Equipment
	Psychiatry
	Biologic Sciences
	Physical Sciences
	Anthropology
	Technology and Food and Beverages
	Humanities
	Information Science
	Persons
	Health Care
	Geographic Zones

Table II. Sampling of MeSH headings pertinent to otolaryngology-head and neck surgery diseases – including the MeSH term for Ramsay Hunt. The MeSH categories and subcategories are listed to demonstrate the hierarchy of terms used. This is not a complete list, but excerpted to provide a visual description of how the MeSH tree is arranged.

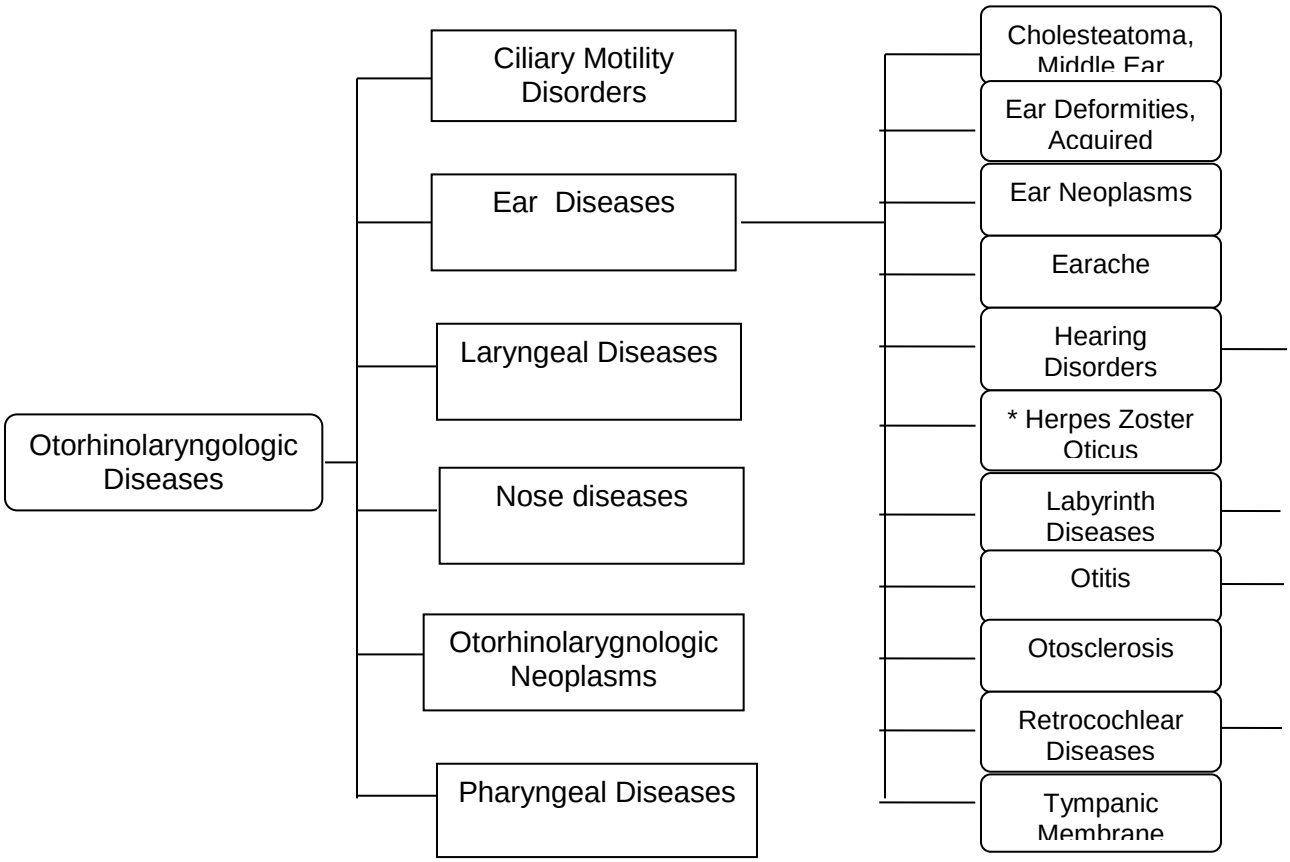

Table III. Subheadings. Each category and subcategory has its own list of subheadings. ~~These are some of subheadings~~ This is a select list of the 30 different subheadings available for Nasopharyngeal Neoplasms (MeSH term).

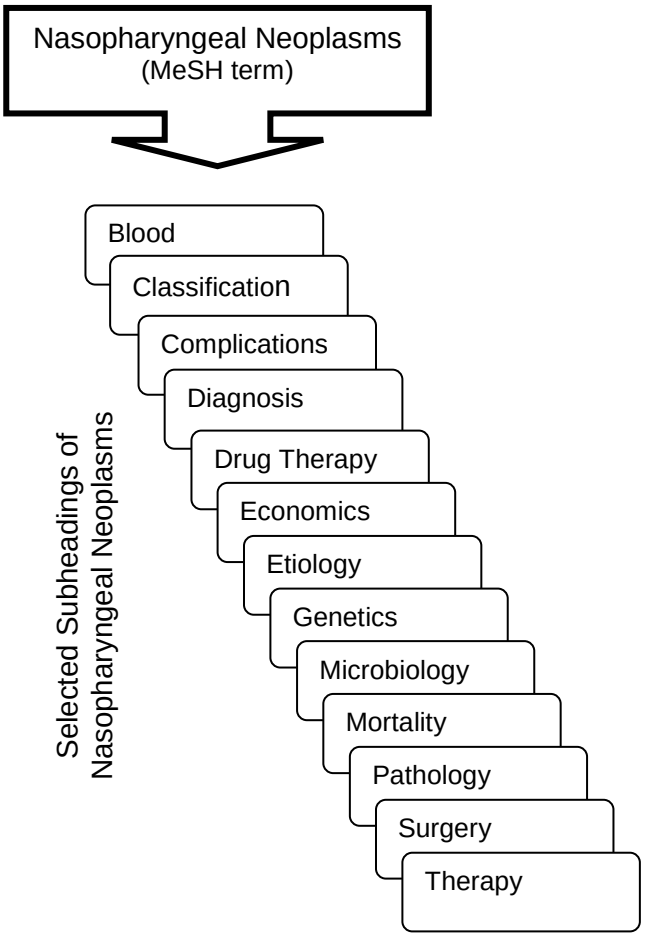

Table IV. ~~Comparison of searches on common Otolaryngology topics.~~ Percentage retrieval difference in text word versus MeSH searches. The difference between the total number of articles retrieved is expressed as a percentage of difference.

Search Topics		Number of articles	% difference
Nasopharyngeal cancer AND epidemiology	*(text word)*	1235	12%
Nasopharyngeal neoplasms/epidemiology	*(MeSH)*	1078	
Obstructive sleep apnea AND surgery	*(text word)*	1781	77%
Sleep apnea, Obstructive/surgery	*(MeSH)*	414	
Frontal sinusitis AND surgery	*(text word)*	358	49%
Frontal sinusitis/surgery	*(MeSH)*	181	
Meniere's AND treatment	*(text word)*	2503	30%
Meniere's Disease/treatment	*(MeSH)*	1738	
Ramsay Hunt syndrome	*(text word)*	2776	94%
Herpes zoster oticus	*(MeSH)*	151	
Vocal cord paralysis AND congenital	*(text word)*	113	67%
Vocal cord paralysis/congenital	*(MeSH)*	37	
Esthesioneuroblastoma	*(text word)*	441	48%
Esthesioneuroblastoma, Olfactory	*(MeSH)*	227	
Rhinoplasty AND complications	*(text word)*	960	73%
Rhinoplasty/complications	*(MeSH)*	257	
Branchial cleft cyst AND embryology	*(text word)*	204	72%
Branchioma/embryology	*(MeSH)*	57	
Tonsillectomy AND complications	*(text word)*	2007	60%
Tonsillectomy/complications	*(MeSH)*	804	

Chart I. Comparison of searches on otolaryngology-head and neck surgery topics. The chart calls attention to the difference in results retrieved by each search strategy for the ten searches. For example, the *tonsillectomy* search was a little over two times more effective in retrieving the needed articles.

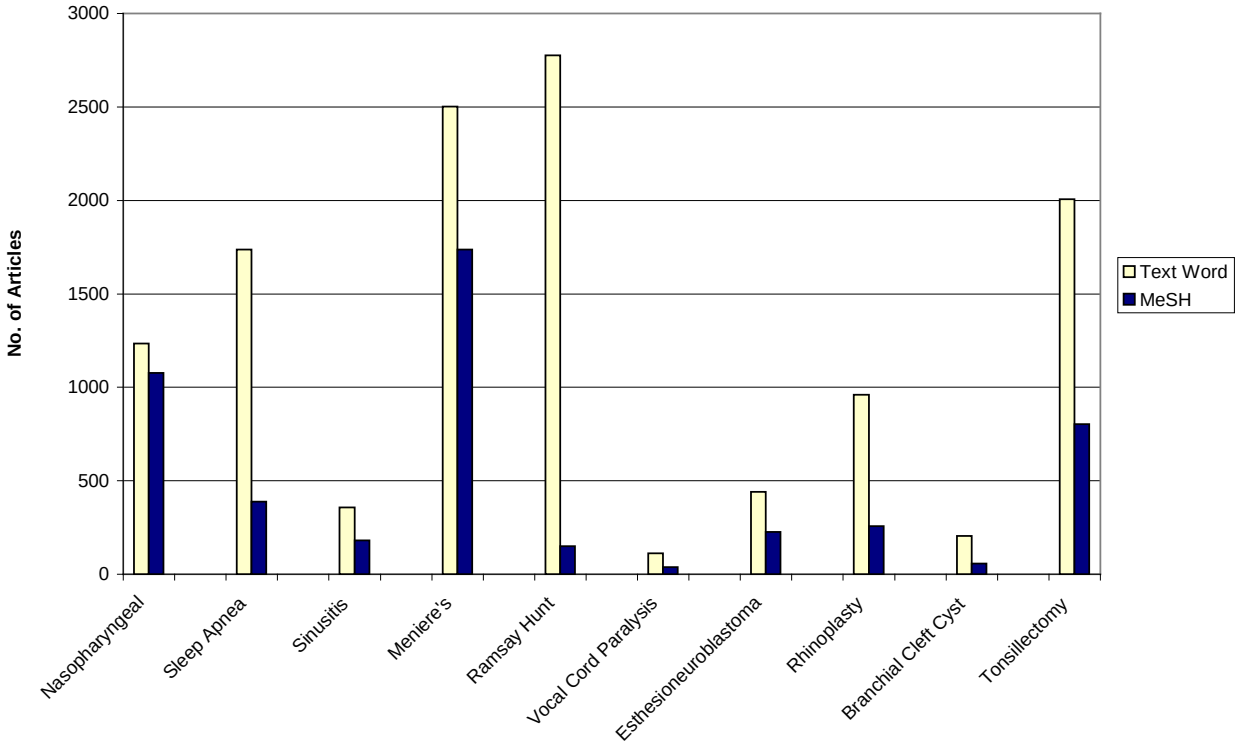

FIGURE LEGENDS

Figure 1. Database display for MeSH search on obstructive sleep apnea surgery, including MeSH hierarchy and subheadings.

Figure 2. An algorithm depicting the process of a MeSH search.

Error!