

## References

1. Grombie DL. Diagnostic process. *J Roy Coll Gen Pract* 1963; 6: 579-589.
2. Peitzman SJ, Mefeinley D, Curtis M, Burdick Whalen G. Performance of international medical graduates in techniques of physical examination. *Academic Medicine* 2000; 75: 115-117.
3. Ahmed AM. Deficiencies of history-taking among medical students. *Saudi Med J* 2002; 23: 991-994.
4. Fox R, Dacre J, McLure E. Impact of formal instruction in examination skills on students performance. *Med Educ* 2001; 35: 371-373.
5. Murray E, Jolly B, Modell M. Can students learn clinical methods in general practice. *BMJ* 1997; 315: 920-923.

## Behind the screen. *PubMed intricacies*

*Ibrahim Mansoor, MBBS, MD.*

United States of America National Library of Medicine (NLM) [[http:// www.nlm.nih.gov/](http://www.nlm.nih.gov/)] is an excellent resource center for literature on medicine and allied fields. It includes free MEDLINE (<http://wwwindex.nlm.nih.gov/databases/freenedl.html>) access, NLM publications, research and special information programs for medical researchers. This letter explains some of the behind the screen concepts for searching medical literature on PubMed. I will try to explain and give the basic understanding of what happens when you look for a term in Basic Search, or how a Clinical Query is carried out, or how the "See Related Articles" function compares articles for similarity on PubMed. This will give the readers a broader understanding of basics of search strategy used by the PubMed search engine. This will also help the medical researchers to be more accurate and productive in performing their medical literature search. I have listed the internal working of PubMed search engine in a stepwise fashion as follows: 1. When you type in the query word in the search box, PubMed looks at its index to see whether the term is a MeSH (Medical Subject Heading) heading. If so, it searches the MeSH term OR the text word in the MeSH list. 2. If the word is not a MeSH term, it compares it to 2 translation tables until a match is found. Thus, a search on bleeding looks in the tables finds the standard MeSH term hemorrhage, and a complex search is created automatically: ("Hemorrhage" [MeSH Terms] OR bleeding [Text Word]). 3. If PubMed cannot find a MeSH equivalent, it looks at the Journal table to see if there is a journal by that name. A search on cell proliferation finds the journal Cell Proliferation, since there is no MeSH phrase with these 2 words. 4. If PubMed cannot find a journal, it looks in a list of phrases. For example, the phrase heart attack is automatically translated to myocardial infarction. 5. If PubMed cannot find a phrase, and one of the terms has a one- or 2-letter word after it (jones pa), PubMed tries the Author index. 6. If PubMed cannot find an author, it then tries variations on the word order. 7. As a last resort,

PubMed takes every individual word and runs them through the translation table, phrase list, and author index.<sup>1,2</sup> Function "See Related articles": This function in PubMed is useful for conducting a rapid literature search. What is PubMed doing when it searches for related articles? Contrary to what might be expected, PubMed does not run a search at the time you click See Related Articles. Instead, the database periodically goes through the algorithm outlined below and generates a list of all the articles related to each citation. Then when you look for Related Articles, it uses this precompiled list. The algorithm compares the similarity in title, abstract, and MeSH terms for the article in question to all the other articles in PubMed. In a step-by-step fashion, PubMed does the following: 1. Stop words are eliminated. 2. A limited amount of stemming is carried out. 3. Words are classified into 3 categories: text word, title word, and MeSH term. a. Words in the abstract are classified as text words. b. Title words are classified as both text and title words. c. MeSH terms are placed in that category and MeSH terms with a subheading are classified under the generic term and the term/subheading pair. Medical Subject Heading major terms are classified both as MeSH and MeSH major term. 4. Each word is assigned a global weight, depending on the number of articles with that word and how important the word is in determining relationships. Global weight is higher for rare words, lower for common ones. 5. Each word is assigned a local weight, depending on how often it appears in the particular article. 6. The similarity between the article and every other article in the database is computed: a. The formula (local weight x local weight x global weight) is calculated for all terms. The 2 documents have in common. b. The product of the lengths of the 2 documents, producing a vector cosine score divides this product. 7. A list of articles with the highest score (those which are most similar) is compiled and saved.<sup>1,2</sup>

**Searching clinical queries.** The Clinical Queries are really just pre-written searches, using MeSH terms and subheadings, on research studies and methodology. This allows you to do a quick search using terms with which you are familiar (therapy, diagnosis, etiology, prognosis, sensitivity, specificity). The expanded phrases below show the syntax that PubMed adds to your term when you run a Clinical Query. These are MeSH terms and subheadings that narrow the search to the general area of therapy, for example. In each case, the entire phrase is connected by Boolean AND to the term(s) you are searching for. Parentheses are used to show the ordering of elements in the search syntax; brackets are used for the [no explosion] parameter when used with MeSH subheadings. To make the syntax somewhat more understandable, I have put each phrase on a separate line, but in reality they are complete phrases. As an example, when searching for the word gentamycin, with a focus on therapy and sensitivity, the syntax would begin: Gentamycin AND (publication type randomized controlled trial for example).

**Therapy + sensitivity**

(publication type randomized controlled trial OR MeSH sub-heading drug therapy) OR MeSH subheading [no explosion] therapeutic use) OR text word random

**Therapy + specificity**

(text word double AND text word blind) OR text word placebo

**Diagnosis + sensitivity**

(MeSH term sensitivity and specificity OR Text word sensitivity)

OR (MeSH subheading diagnosis OR MeSH subheading diagnostic use)

OR Text word specificity

**Diagnosis + specificity**

MeSH term sensitivity and specificity OR (text word predictive AND text word value)

**Etiology + sensitivity**

(MeSH term cohort studies OR MeSH term risk) OR (text word odds AND text word ratio) OR (text word relative AND text word risk) OR (text word case AND text word control)

**Etiology + specificity**

MeSH term [no explosion] case-control studies OR MeSH term [no explosion] cohort studies

**Prognosis + sensitivity**

(MeSH term incidence OR MeSH term mortality OR MeSH term follow-up studies) OR MeSH subheading mortality) OR text word prognosis\*) OR text word predict\*) OR text word course)

**Prognosis + specificity**

MeSH term [no explosion] prognosis OR MeSH term [no explosion] survival analysis

For example, the Clinical Query choices for Etiology, with a focus on specificity, limit your search with 2 MeSH terms: case-control studies and cohort studies. In other words, the way that health professionals understand and use the term etiology matches closely with these 2 MeSH terms that the staff of the NLM assign to citations.<sup>1,2</sup> In conclusion, I feel that understanding the steps of inner working of PubMed search engine will help the medical researchers and clinicians to be more precise and accurate in their search tasks.

*Received 24th August 2002. Accepted for publication in final form 25th August 2002.*

*From the Department of Pathology, King Abdul-Aziz University Hospital, Jeddah, Kingdom of Saudi Arabia. Address correspondence and reprint requests to Dr. Ibrahim Mansoor, Department of Pathology, PO Box 1432, Jeddah 21431, Kingdom of Saudi Arabia. E-mail: ibm979@hotmail.com*

*References*

1. Step-by-step guide to searching PubMed. Available from: URL: <http://www.glenlib.demon.co.uk/pubmed/>
2. Introduction to Web-based Searching: Using PubMed. Available from: URL: <http://www.nlm.nih.gov/pubs/manual2dayclassta.pdf>