Research

Print short, Web long*

So many databases, such little clarity

Searching the literature for the topic aboriginal

Len Kelly MD MClinSci CCFP FCFP   Natalie St Pierre-Hansen

ABSTRACT

OBJECTIVE To describe the scope, content, and organization of commonly used medical databases and search strategies, using a search of the topic *aboriginal* to illustrate the various ways the topic is covered in each of the databases.

DESIGN Comparison of literature searches.

METHOD Seven common medical databases were searched using all the MeSH terms that are permutations of *aboriginal.* A secondary analysis using the "remove duplicates" function in Ovid was done to identify articles specific to each database.

MAIN OUTCOME MEASURES Number of articles found by each search.

RESULTS Searching by MeSH terms often produces very different information from that found when searching by text word. A unique term, such as *Ojibway,* is best found with a text word search. A more general term, such as *Aborigines,* is best searched by subject using a MeSH term. Many databases can be searched through Ovid and might all use different MeSH terms for the same reference. PubMed default searches that use MeSH terms and text words simultaneously often produce very large numbers of articles. In searching for *North American aboriginal* using MeSH terms, MEDLINE and PubMed produced the most references, followed by Healthstar. Calculating distinct "all aboriginal" references in EMBASE, Healthstar, and PsycINFO indicated that MEDLINE produced nearly all the articles found in Healthstar. In fact, MEDLINE alone produced 88% of the articles found in MEDLINE and EMBASE and 79% of the articles found in MEDLINE and PsycINFO.

CONCLUSION Although several researchers and medical librarians have noted that MEDLINE and EMBASE are quite distinct databases, suggesting both need to be searched for a complete search, we did not find that to be the case for the topic *aboriginal.* The results of this study demonstrate that using MEDLINE produces the most extensive coverage of literature on the topic *aboriginal.* To fully capture the complete body of available literature on other subjects might require searches of many databases, depending on the topic.

EDITOR'S KEY POINTS

- Are all medical databases the same? The authors looked at how 7 major databases differed in the number of articles covered. Using MeSH term and text word search strategies for the topic *aboriginal*, they found that MEDLINE had the most extensive coverage. Of the 3 databases used in combination with MEDLINE, PsycINFO was the most distinct from MEDLINE, and Healthstar was the least distinct.

- The authors also found that searching using MeSH terms and text words simultaneously, as with a PubMed default search, produced an unmanageable number of articles. Searching general terms was best done by subject using a MeSH term; a unique term was best found using a text word search.

- The authors concluded that databases treat their articles in unique ways. Awareness of available databases and of the scope and organization of MeSH terms in these databases will help researchers choose the best ways to define search parameters that will adequately cover the desired topic. Creating search strategies specific to each database and its organization of MeSH terms will lead to more comprehensive results.

Tant de bases de données, si peu de clarté

Revue de littérature sur le sujet «aboriginal»

Len Kelly MD MClinSci CCFP FCFP   Natalie St Pierre-Hansen

RÉSUMÉ

OBJECTIF Décrire l'étendue, le contenu et l'organisation des bases de données et des stratégies de recherche couramment utilisées en médecine, au moyen d'une recherche sur le sujet *aboriginal* (en anglais), afin de montrer les façons différentes de couvrir ce sujet dans chaque base de données.

TYPE D'ÉTUDE Comparaison de revues de littérature.

MÉTHODE On a consulté 7 bases de données médicales courantes en utilisant tous les équivalents du terme MeSH *aboriginal*. Un analyse secondaire a été effectuée à l'aide de la fonction « éliminer les doubles » dans Ovid afin d'identifier les articles spécifiques à chaque base de données.

PRINCIPAL PARAMÈTRE MESURÉ Nombre d'articles identifiés par chaque recherche.

RÉSULTATS Une recherche par termes MeSH produit souvent des résultats très différents de ceux résultant d'une recherche par mots courants. Avec un mot unique comme *Ojibway*, il est préférable d'utiliser un mot courant. Avec un mot plus général comme *aboriginal*, une recherche par sujet avec des termes MeSH est préférable. Ovid permet une recherche dans plusieurs bases de données, lesquelles peuvent toutes utiliser des termes MeSH différents pour la même recherche. Les recherches PubMed par défaut qui utilisent simultanément des termes MeSH et des termes courants produisent souvent un très grand nombre d'articles. Dans une recherche sur *North American aboriginal* avec des termes MeSH, MEDLINE et PubMed ont produit le plus grand nombre de références, suivis par Healthstar. Le décompte des références distinctes pour « all aboriginal » obtenues avec EMBASE, Healthstar et PsycINFO indique que MEDLINE a produit la presque totalité des articles repérés par Healthstar. En fait, à lui seul, MEDLINE a produit 88% de tous les articles repérés par MEDLINE et EMBASE, et 79% de ceux repérés par MEDLINE et PsycINFO.

CONCLUSION Bien que plusieurs chercheurs et bibliothécaires soutiennent que MEDLINE et EMBASE sont des bases de données relativement distinctes, suggérant donc qu'il faut les utiliser toutes deux pour une recherche complète, ce n'est pas ce que nous avons observé avec le sujet *aboriginal*. Les résultats de cette étude montrent que MEDLINE fournit la couverture la plus étendue du sujet *aboriginal*. D'autres sujets pourraient nécessiter l'utilisation de plusieurs bases de données.

POINTS DE REPÈRE DU RÉDACTEUR

- Les bases de données sont-elles toutes semblables? Les auteurs ont vérifié les différences entre les nombres d'articles repérés par 7 bases de données majeures. Utilisant des stratégies de recherche à la fois par termes MeSH et par mots courants pour le sujet *aboriginal*, ils ont observé que MEDLINE fournissait la plus large couverture. Des 3 bases de données utilisées en combinaison avec MEDLINE, PsycINFO était la plus différente et Healthstar la moins différente.

- Les auteurs ont aussi observé qu'en utilisant simultanément des termes MeSH et des mots courants, comme dans une recherche PubMed par défaut, on obtenait un nombre ingérable d'articles. Une recherche sur des termes généraux était préférablement effectuée par sujet, à l'aide de termes MeSH; pour un terme unique, un mot courant était préférable.

- Les auteurs concluent que chaque base de données traite ses articles d'une façon qui lui est propre. Une bonne connaissance des bases de données disponibles, et de l'étendue et de l'organisation de leurs termes MeSH, aidera le chercheur à choisir la meilleure façon de définir les paramètres de recherche devant couvrir adéquatement le sujet désiré. La création de stratégies de recherche spécifiques à chaque base de données et à l'organisation de ses termes MeSH procurera des résultats plus complets.

While the virtual hallways of the electronic medical library are not as dusty as those we inhabited during medical training, they can at times appear just as confusing and dimly lit. Electronic searches can be both fun and frustrating for busy clinicians and primary care researchers. They can be fun when the information is instantly at our fingertips and frustrating because we have to navigate by foreign maps and sometimes we either cannot find what we want or find too much of what we want.

While many of us have developed search strategies that meet our current needs, comprehensive literature reviews often require more complicated strategies. Literature searching is like many medical skills: the more practised we are, the more we can appreciate subtleties in presentation. In the process of researching the topic *aboriginal*, we learned some lessons we would like to pass along to readers.

Aboriginal and First Nations

Researchers searching the topic *aboriginal* are faced with current evolving political and cultural terminology. The terminology we use today is not the same as that used by librarians in the 1960s and around which the databases were organized.

According to the Canadian Constitution, the term *aboriginal* refers to the indigenous inhabitants of Canada. This comprises 3 distinct groups: First Nations, Inuit, and Metis.[1] *First Nations* is, therefore, a subset of *aboriginal* and has replaced the term *North American Indians* in contemporary discussion. MeSH terms, on the other hand, do not comply with current accepted terminology and instead often use uncommon descriptions, such as *American native continental ancestry group* or *American Indian,* for First Nations designations.[2] In addition, some databases do not allow for a distinction between North American natives and native groups in Central and South America, as we found all groups were sometimes categorized under the term American Indian.

Ovid and PubMed

Ovid and PubMed are common ways to search the literature and they differ in interesting ways. Ovid, which is a search engine, can access many common databases including MEDLINE and EMBASE. Although Ovid-accessible databases share the same search interface, they have their own unique MeSH terms. MeSH terms are descriptors developed by librarians to organize and categorize topics.[3,4] Fortunately, the Ovid search engine takes you to these terms by default, as they are usually not obvious. For example, *aboriginal* is not a MeSH term, but if you enter it into Ovid, a map of terms such as *Indians, North American* will be produced instead. The alternative to a MeSH term search is a text word search. A text word search finds articles that contain the given word anywhere in the title, abstract, or text of the article.[5] Depending on how common the word is, an unwieldy number of articles can be produced. For example, entering *Native* into PubMed brings up more than 100 000 publications. One advantage to searching by text word is that such a search can find specific and infrequently used words. Entering *Ojibway,* which is not a MeSH term, will uncover a manageable 32 articles, even though Ojibway might not be the main focus of these articles.

PubMed is a database that uses a search engine called Entrez. MEDLINE, the most common database accessed through Ovid,[6] essentially produces the same articles as PubMed with a few exceptions. PubMed is slightly larger than MEDLINE as its scope is marginally broader, and new references are added to PubMed more quickly than they are to MEDLINE.

The difference in entering a term into Ovid MEDLINE and entering a term into PubMed is that PubMed's default setting will search by text word and MeSH term simultaneously. Therefore, if you do not specify MeSH term only, PubMed will likely retrieve a large number of articles because it has combined a reasonable number of MeSH references with a large number of text word references. An Ovid search will retrieve articles by MeSH term only. A PubMed search that is limited to MeSH terms will be similar in scope to an Ovid MEDLINE search.

METHODS

We searched 7 common medical databases using all the MeSH terms for *aboriginal* (both for Canada's First Nations people and foreign aboriginal peoples): Ovid MEDLINE, EMBASE, PubMed, CINAHL, Healthstar, PsycINFO, and EBM Reviews. We searched the topics of interest using subject (MeSH term) and text word strategies in Ovid and PubMed databases.

We conducted a secondary analysis to identify the distinct references to *aboriginal* in EMBASE, Healthstar, and PsycINFO and compared them with those in MEDLINE. Using the Ovid interface, we searched using 2 databases simultaneously. All relevant MeSH terms for the 2 databases were used and combined with the Boolean operator *OR.* The "remove duplicates" function in Ovid was used to discover the number of articles distinct to each database.

Ovid databases searched
- MEDLINE contains more than 15 million articles published in more than 4600 biomedical journals from 1950 to the present.[7]
- CINAHL (Cumulative Index to Nursing and Allied Health Literature) is the database of nursing and allied health literature and contains articles from 1982 to the present.

- Healthstar comprises data from MEDLINE, the hospital literature index, and selected journals; focuses on clinical and nonclinical aspects of health care delivery; and includes journal articles, technical reports, government documents, and newspaper articles from 1975 to the present.
- PsycINFO contains literature on psychology from more than 2200 periodicals from 1872 to the present. It was previously known as PsycLIT.[8]
- EMBASE is a biomedical and pharmaceutical database with more than 18 million records from 1974 to the present.[9] It includes more European articles than MEDLINE does.[10]
- EBM Reviews contains evidence-based medicine records from 1948 to the present from the American College of Physicians Journal Club, the Cochrane Controlled Trials Register, the Cochrane Database of Systematic Reviews, and the Database of Abstracts of Reviews of Effects.

Other databases searched

- PubMed contains more than 17 million articles[11] from MEDLINE and other life sciences journals that might be beyond the scope of MEDLINE (eg, Astrophysics) from 1950 to the present.[12]
- The Native Health Database includes literature, documents, reports, and surveys relating to the health of American Indians, Alaska Natives, and Canadian First Nations peoples from 1966 to the present.

RESULTS

Search terms

Databases treat their articles in unique ways. Australian aboriginal literature, for example, is identified as such in Healthstar, MEDLINE, and PubMed using *Oceanic ancestry group* as a MeSH term. In EMBASE, Australian aboriginal literature might be indexed under *indigenous people* or *Aborigines.*

Articles on North American First Nations people, previously referred to as *North American Indians,* can be searched in MEDLINE under *Indians, North American,* and in EMBASE under *American Indians,* but the latter includes literature on South and Central American Natives as well. Owing to the way EMBASE organizes its MeSH terms, we were unable to specifically identify unique EMBASE contributions on this topic, but we did ascertain that the number of articles in EMBASE was much smaller than the number in MEDLINE.

Text word searches for broad terms produce unmanageably large numbers of articles but might be helpful for searching for articles on specific tribes (**Table 1**).

Comparison of databases

The total number of distinct articles shown in **Table 2** for each database provides an initial overview of the comprehensiveness of each database. Combining all articles on the topic *aboriginal* was possible methodologically and provided a useful way to compare databases. These numbers will be of less interest to researchers who are focusing on a specific aboriginal group. Our results illustrate that MEDLINE provides the most extensive coverage of the topic *aboriginal* when a search is done using MeSH terms.

For those who wish to search more than 1 database, our secondary analysis will help them decide which combination of databases will provide the most comprehensive search results. Once we eliminated the overlap between databases, we were able to see their distinctiveness. Of the 3 databases used in combination with MEDLINE for searching the topic *aboriginal,* PsycINFO is the most distinct from MEDLINE and Healthstar the least (**Figures 1**-**3**).

DISCUSSION

The distinctiveness of each MeSH term and the variations in the way terms are organized in the various databases highlight the importance of selecting each of the terms to be used in the search. Rosser and colleagues examined the difference between terms used by British physicians and those used by Canadian physicians. They pointed out that the use of *general practice, family medicine,* or *family practice* as key words yielded substantially different results depending on the interface used.[13] We believe that mapping search terms to the database to be

Table 1. Number of articles found through searches using text words for the topic *aboriginal*

	NUMBER OF ARTICLES						
TEXT WORDS	HEALTHSTAR 1975–JULY 2007	OVID MEDLINE 1950–JULY 2007	PUBMED MEDLINE 1950–JULY 2007	EBM REVIEWS 1948–JULY 2007	CINAHL 1982–JULY 2007	EMBASE 1974–JULY 2007	PSYCINFO 1872–JULY 2007
Native	22 728	96 703	100 375	797	4038	77 217	10 260
Indian	12 660	21 301	31 980	618	1638	17 946	8876
Aboriginal	2531	3022	3282	55	713	2186	1056
First Nation$*	1175	1270	924	26	431	938	355
Ojibway	24	32	32	9	0	20	21

*Truncation for First Nation and First Nations.

Table 2. Number of articles found through searches using various MeSH terms for the topic *aboriginal*

MESH TERMS	NUMBER OF ARTICLES						
	HEALTHSTAR 1975–JULY 2007	OVID MEDLINE 1950–JULY 2007	PUBMED MEDLINE 1950–JULY 2007*	EBM REVIEWS 1948–JULY 2007	CINAHL 1982–JULY 2007	EMBASE 1974–JULY 2007	PSYCINFO 1872–JULY 2007
First Nations, North American							
Indigenous populations							980
Indigenous people					319	321	
Aborigines				0	884	2059	
Tribes							651
American native continental ancestry group	135	159	12 924				
Native Americans					2877		
Indians, North American	7112	8822	8736	102			
American Indian						2343	3949
Alaska Natives							222
Other aboriginal							
Indians, South American	1546	2384	2358	7			
Indians, Central American	189	318	318				
Eskimo					202	355	
Inuit	2110	2408	2399				299
Maori					223	228	
Pacific islander						288	155
Oceanic ancestry group	3386	4033	4014				
Total distinct articles found using all MeSH terms combined with Boolean *OR*	12 135	16 993	16 834	109	4356	5294	5367

*PubMed search using MeSH-only strategy.

Figure 1. Results of a MeSH term search for the topic *North American aboriginal*: *Ovid MEDLINE vs Healthstar.*

Indians, North American

Healthstar only
(6 references)
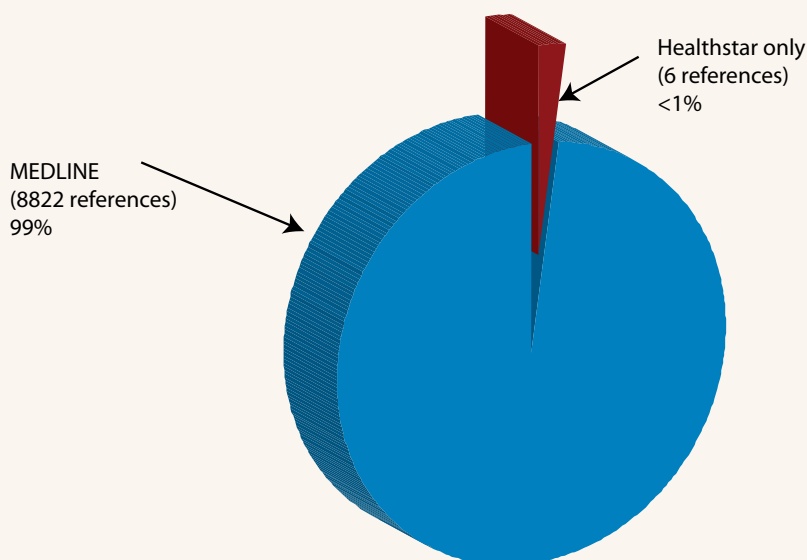<1%

MEDLINE
(8822 references)
99%

Figure 2. Results of a MeSH term search for North American and international articles on the topic *aboriginal*: *Ovid MEDLINE vs EMBASE.*
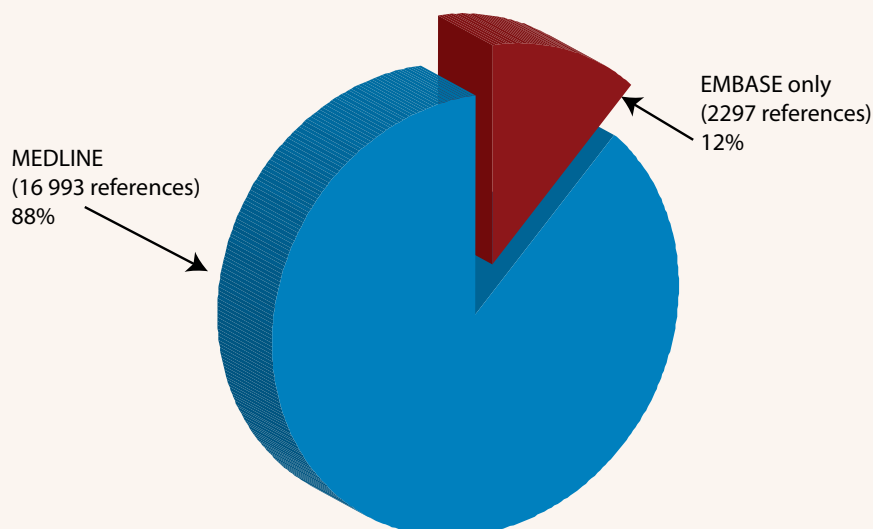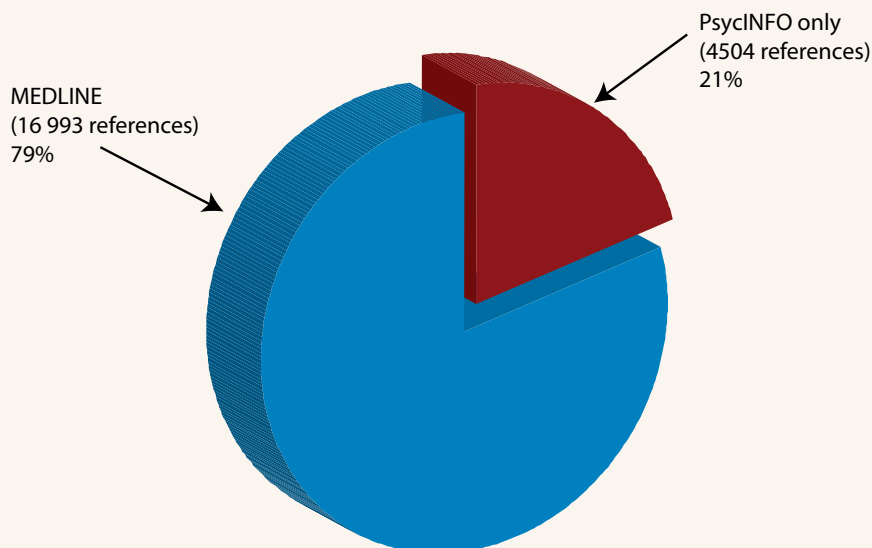
All *aboriginal*

MEDLINE
(16 993 references)
88%

EMBASE only
(2297 references)
12%

Figure 3. Results of a MeSH term search for North American and international articles on the topic *aboriginal*: *Ovid MEDLINE vs PsycINFO.*

All *aboriginal*

MEDLINE
(16 993 references)
79%

PsycINFO only
(4504 references)
21%

used and creating a search strategy specific to that database and its organization of MeSH terms will lead to more comprehensive results.

The decision about which databases to search to get thorough reviews of the literature will likely vary by topic.[14-16] Several researchers and librarians recommend combining MEDLINE and EMBASE for a complete review,[17,18] as these 2 databases often have been found to contain quite distinct listings for primary care topics. Wilkins searched 15 family medicine topics in these 2 databases and found EMBASE yielded significantly more articles than MEDLINE did ($P = .0005$).[16] The fact that EMBASE references more European literature than the other databases do, might explain why it seemed to be a less useful resource for our example search of literature on *North American First Nations.* For the topic *aboriginal,* we found MEDLINE to be sufficiently complete.

Limitations

Since EMBASE did not have a specific category for North American First Nations, we could not identify its unique contributions to that topic. We used the "remove duplicates" feature on Ovid, but this feature was limited to 6000 entries. We therefore broke the data sets down chronologically to below 6000 entries to allow the feature to remove duplicate entries. We found small inaccuracies in the "remove duplicates" function in Ovid, but we are confident that our results adequately describe the differences in coverage of the various databases.

Conclusion

An awareness of available databases and of the scope and organization of MeSH terms in these databases can assist researchers in choosing the best ways to define search parameters that will adequately cover the desired topic. For the topic *aboriginal,* a search of MEDLINE appeared to produce a sufficiently complete listing of articles. Searching PsycINFO also might have produced additional relevant articles. Other topics might require researchers to select different databases and use different search strategies. ✤

Dr Kelly is a family physician and an Associate Professor of Family Medicine at the Northern Ontario School of Medicine and McMaster University in Sioux Lookout, Ont. **Ms St Pierre-Hansen** is a research intern at the Northern Ontario School of Medicine and FedNor in Sioux Lookout.

Contributors
Dr Kelly and **Ms St Pierre-Hansen** contributed to concept and design of the study; data gathering, analysis, and interpretation; and preparing the article for submission.

Competing interests
None declared

Correspondence
Dr Len Kelly, Northern Ontario School of Medicine, Box 489, Sioux Lookout, ON P8T 1A8; e-mail **lkelly@mcmaster.ca**

References

1. Indian Affairs and Northern Development Canada. *Aboriginal place names*. Ottawa, ON: Indian Affairs and Northern Development Canada; 2001. Available from: **www.ainc-inac.gc.ca/pr/info/info106_e.html**. Accessed 2007 Nov 16.
2. Gault L, Shultz M, Davies KJ. Variations in medical subject headings (MeSH) mapping: from the natural language of patron terms to the controlled vocabulary of mapped lists. *J Med Libr Assoc* 2002;90(2):173-80.
3. Colletti MH, Bleich HL. Medical subject headings used to search the biomedical literature. *J Am Med Inform Assoc* 2001;8(4):317-23.
4. Lipscomb C. Medical subject headings (MeSH). *Bull Med Libr Assoc* 2000;88(3):265-6.
5. Jenuwine ES, Floyd JA. Comparison of medical subject headings and text-word searches in MEDLINE to retrieve studies on sleep in healthy individuals. *J Med Libr Assoc* 2004;93(3):349-53.
6. Brand-de Heer DL. A comparison of the coverage of clinical medicine provided by PASCAL BIOMED and MEDLINE. *Health Info Libr J* 2001;18(2):110-6.
7. National Library of Medicine, Bibliographic Services Division. *Data, news and update information*. Bethesda, MD: United States National Library of Medicine; 2001. Available from: **www.nlm.nih.gov/bsd/revup/revup_pub.html#med_update**. Accessed 2007 Jul 9.
8. APA Databases. *PsycINFO journal coverage list*. Washington, DC: APA Online; 2008. Available from: **www.apa.org/psycinfo/about/covlist.html**. Accessed 2007 Jul 9.
9. Elsevier BV. *EMBASE.com* [homepage on the Internet]. Amsterdam, The Netherlands: Elsevier BV; 2007. Available from: **www.EMBASE.com/**. Accessed 2007 Jul 9
10. Greenhalgh T. How to read a paper: the MEDLINE database. *BMJ* 1997;315(7101):180-3.
11. National Institutes of Health, National Library of Medicine. *PubMed* [homepage on the Internet]. Bethesda, MD: National Library of Medicine; 2008. Available from: **www.ncbi.nlm.nih.gov/sites/entrez/**. Accessed 2007 Jul 9.
12. National Center for Biotechnology Information. *Fact sheet. PubMed: MEDLINE. Retrieval on the World Wide Web*. Bethesda, MD: National Library of Medicine; 2002. Available from: **www.nlm.nih.gov/pubs/factsheets/pubmed.html**. Accessed 2007 Jul 9.
13. Rosser WW, Starkey C, Shaughnessy R. The pond is wider than you think! Problems encountered when searching family practice literature. *Can Fam Physician* 2000;46:103-8.
14. McDonald S, Taylor L, Adams C. Searching the right database: a comparison of four databases for psychiatry journals. *Health Libr Rev* 1999;16(3):151-6.
15. Holland JL, Wilczynski NL, Haynes RB; Hedges Team. Optimal search strategies for identifying sound clinical prediction studies in EMBASE [abstract]. *BMC Med Inform Decis Mak* 2005;5:11.
16. Wilkins T, Gillies RA, Davies K. EMBASE versus MEDLINE for family medicine searches: can MEDLINE searches find the forest or a tree? *Can Fam Physician* 2005;51:848-9.
17. Dunikowski LG. EMBASE and MEDLINE searches. *Can Fam Physician* 2005;51:1191.
18. Woods D, Trewheellar K. MEDLINE and EMBASE complement each other in literature searches [abstract]. *BMJ* 1998;316(7138):1166.

———✳ ✳ ✳———