International Journal of Technology Assessment in Health Care, 15:2 (1999), 297–303.

COMPARISON OF LITERATURE SEARCHES ON QUALITY AND COSTS FOR HEALTH TECHNOLOGY ASSESSMENT USING THE MEDLINE AND EMBASE DATABASES

Leigh-Ann Topfer
Institute of Health Economics, Edmonton

Antoni Parada, *Catalan Agency for Health Technology Assessment (CAHTA)*

Devidas Menon, *Institute of Health Economics, Edmonton*

Hussein Noorani, *Canadian Coordinating Office for Health Technology Assessment (CCOHTA)*

Christine Perras, *Alberta Health*

Mateu Serra-Prat
Canadian Coordinating Office for Health Technology Assessment (CCOHTA)

Abstract

Biomedical databases are an important source of information for health technology assessment. However, there is considerable variation in the costs of accessing commercial databases. We sought to measure the quality, amount of overlap, and costs of information retrieved from two of the main database sources — MEDLINE and EMBASE. Librarians at two health technology assessment agencies ran a total of eight literature searches on various medical technologies, using both databases. All search results were independently reviewed by two researchers. The researchers were asked to identify relevant references and to rank each of these according to a level of evidence scale. The results were tabulated to show the number of references identified by each database, the number of relevant references ranked by level of evidence, and the number of these references that were unique to one or the other database. The cost of retrieving references from each source was also calculated. Each database contained relevant references not available in the other. Because of the longer time lag for indexing in MEDLINE, many of the references that originally appeared to be unique to EMBASE were subsequently available in MEDLINE as well. Since our study was conducted, MEDLINE has been made available worldwide, free of charge, via the Internet. Hence, the cost difference between the databases is now even greater. However, notwithstanding the costs, it appears that literature searches that rely on only one or the other database will inevitably miss pertinent information.

Keywords: Databases, Bibliographic, MEDLINE, EMBASE (Non-MeSH), Information Services

The authors would like to thank Dr. Alicia Granados and Dr. Albert J. Jovell for their assistance and support in the preparation of this study.

Health technology assessment (HTA) has gained increasing importance worldwide in recent years. Results of health technology assessments (8) are intended to be used to aid decision making at various levels of health care, from a population–health basis to the individual patient–provider level. As such, HTA has the potential to have tremendous impact on quality and costs of health care and on access to appropriate care. It is imperative, therefore, that the information used for the evaluation of a medical technology be as complete and current as possible. This is one of the major challenges faced by those involved in HTA research.

Typically, various sources of information and a number of information retrieval methods are used for an HTA study (3). These include hand-searching of specialized journals in the discipline; scanning reference lists of key articles; attempting to locate documents in the nonpeer-reviewed, or gray literature, and contact with experts in the field. However, most of the information relevant to a health technology and its effects is generally published in peer-reviewed literature. This information is usually found through bibliographic databases.

Many databases currently exist, but none covers all sources of published information. Conducting a thorough literature search involves making choices between databases offered through a variety of online, Internet, and CD-ROM platforms. In making their selection, searchers must consider such issues as the relevance, quality, and quantity of the information *likely* to be retrieved from each source, as well as the associated costs.

Two of the major biomedical databases are MEDLINE (U.S. National Library of Medicine) and EMBASE (Elsevier Science). MEDLINE is typically the first choice for searching, offering the largest collection of references with the longest retrospective access (back to 1966), and at the lowest price. EMBASE is also often considered by searchers, particularly for increased coverage of European and non-English language publications, and for its strengths in specific areas such as pharmaceuticals, psychiatry, health policy, and alternative medicine (7). However, the cost of EMBASE searches is substantially higher than searches run in MEDLINE. For example, current costs for searching on DIALOG are US$3.00 per DialUnit (online search time) plus US$0.20 per reference in MEDLINE, versus US$7.75 per DialUnit plus US$1.90 per reference for EMBASE. In addition, since June 1997, the U.S. National Library of Medicine has offered free access to MEDLINE through the Internet Grateful Med (URL: http://igm.nlm.nih.gov/) and PubMed (URL: http://www.ncbi.nlm.nih.gov/PubMed/) web sites. Given the reported estimates of only 30–50% overlap between the two databases (5), a comparative analysis may be useful.

This study addresses a number of questions. First, how much overlap is there between EMBASE and MEDLINE? Is there a pattern of commonality between types of technologies (for example, drug versus non-drug studies)? Is there a trend in the quality of the scientific evidence of references in one database or the other? Is there a significant cost difference between the two databases in identifying *relevant* references? Are there reasons why a search on one database may have missed a reference identified in a search of the other?

METHODS

The Canadian Coordinating Office for Health Technology Assessment (CCOHTA) and the Agència d'Avaluació de Tecnologia Mèdica (CAHTA) are government-funded organizations created to undertake health technology assessment. Researchers at the two organizations selected eight medical technologies, including drug, device, surgical procedures, and alternative therapies as search topics. Librarians at each agency ran literature searches on four of the eight topics, on both

Table 1. Keywords Used in Eight Database Searches

Search	Keywords used
Homeopathy	Homeopath? (truncated in the descriptor field), and treat? or therap? as free text terms.
Chelation therapy for treatment of atherosclerosis	Chelation therapy, chelating agent, EDTA, or edetic acid (title or descriptor field), and atherosclerosis, arteriosclerosis, or intermittent claudication (title or descriptor field).
Flutamide for treatment of prostate cancer	Prostat? (cancer? or neoplasm?) (truncated terms in the descriptor field), and flutamide? and therap? as free text terms.
Riluzole (Rilutek) in the treatment of amyotrophic lateral sclerosis	Riluzole or Rilutek (in the title, descriptor, or abstract fields), or amyotrophic lateral sclerosis or motor neuron disease (truncated) or Gehrig('s) disease or lateral sclerosis (in the title, descriptor or abstract fields) and drug therapy (in the descriptor field).
Holmium laser	Holmium (within 4 words) laser? (truncated) as a free text term.
Coronary stents for prevention of restenosis following angioplasty	Stent? (truncated, in the title or descriptor field) and angioplasty (in the title or descriptor field). The search was further restricted to studies where these were major descriptors. Additional terms such as clinical trial, randomized controlled trial, clinical study, review, and cost? (truncated) (variously applied in the title, descriptor, and document type fields) were also used to refine the search results.
Diagnosis and prevention of depression in primary care	Depress? (truncated in the descriptor field), the clause primar? (within 3 words of) care, and the terms diagnos? and detect? (truncated) as free text terms.
Gamma knife vs. linear accelerator (linac) for stereotactic radiosurgery of arteriovenous malformations	Gamma knife or linear accelerator(s), particle accelerators or magnetic electromagnetic equipment (in the title or descriptor field), or linac (in the title field), and arteriovenous malformation(s) or blood vessel malformation. These were then combined with the terms radiosurgery or stereotaxic or stereotactic (in the title or descriptor field).

EMBASE and MEDLINE databases, using the OneSearch feature on the DIALOG (Knight-Ridder, Inc.) system. (OneSearch allows searchers to run the same search strategy simultaneously in multiple databases, and to thus identify and remove duplicate references.) Search strategies used a combination of Medical Subject Headings (MeSH) and EMTREE Thesaurus terms, with additional keywords if thesaurus terms seemed insufficient. The specific keywords used for each search are listed in Table 1. All searches were restricted to references published from 1995 to 1996, and to human studies. Based on the abstracts available through the databases, two researchers at each organization independently reviewed the four searches run at their agency. In some cases where abstracts were not available or did not provide sufficient information, the full text of the reference was obtained. The researchers were not given specific selection criteria but were asked to: a) identify references that they felt might be relevant were an assessment to be done; and b) to rank each of these references according to a level of evidence scale adapted from the Canadian Task Force on the Periodic Health Examination (Table 2) (2). Results from each researcher's rankings were compared, and differences in rankings were re-examined and discussed to reach agreement.

Table 2. Quality of Evidence

1. Evidence obtained from at least one properly randomized controlled trial.
2. Evidence obtained from well-designed controlled trials without randomization.
   Evidence obtained from well-designed cohort or case-control analytic studies, preferably from more than one center or research group.
   Evidence obtained from comparisons between times or places with or without the intervention . . .
   [We included case series in this level.]
3. Opinions of respected authorities, based on clinical experience, descriptive studies or reports of expert committees. [We included editorials and reports of single cases in this level.]
4. [We added this level for "review" articles.]
5. [We added this level for letters, and for "other" references that we were not able to categorize in any of the above.]

Source: With the exception of the information in square brackets, this is based on the Quality of Evidence Table developed by the Canadian Task Force on the Periodic Health Examination (2).

The following data were collected for each search: a) total number of references identified; b) number of *relevant* references ranked by level of evidence; and c) number of unique references obtained from each database (i.e., those that were identified in one database but not the other.) In addition, the cost of retrieving each reference was calculated, based on the database vendors' charges per reference at the time (US$0.20 for MEDLINE and US$1.75 for EMBASE). (These charges are similar to price differences for these databases on other systems. Online search time charges were not included because these may be influenced by factors such as the length or complexity of the search strategy, the typing speed of the searcher, etc.)

The reasons for the MEDLINE search having "missed" references found in EMBASE were then investigated. This was done by subsequently conducting additional MEDLINE searches for each of the references found only in EMBASE search results. These were further analyzed to determine if the omission was caused by time lag (i.e., some references were not retrieved in the initial round of searches due to the greater time delay for indexing in MEDLINE) (5), lack of or only selective coverage of particular journals, or shortcomings in the search strategy/indexing used. In addition, the level of evidence for each of these unique references and the language of publication were noted to determine if any trends were evident for the quality/type of references missed in MEDLINE.

RESULTS

Table 3 shows the results of the original MEDLINE and EMBASE searches for the eight health technologies. Overall, EMBASE identified almost twice as many references as MEDLINE, although the search results showed that each database identified relevant references not found or not identified in the other. No consistent differences were found in the quality of references retrieved in either database. This comparison is also dependent on the search strategies used, and possibly, the topics chosen. The number of unique references found in EMBASE was substantially greater in the alternative therapy and drug searches, as compared to some of the device searches (lasers and gamma knife vs. linear accelerator); however, this trend was not consistent for all devices (for example, with coronary stents EMBASE had a much higher retrieval rate). EMBASE is considerably more expensive than MEDLINE per reference retrieved (US$1.75 versus US$0.20 at the time of the

Table 3. Summary Results (by Technology and Total)

Searches	Total No. of Refs MEDLINE/ EMBASE	No. of Relevant Refs MEDLINE/ EMBASE	No. of Relevant & Unique Refs MEDLINE/ EMBASE	No. of Relevant & Unique Level 1 Refs MEDLINE/ EMBASE	No. of Relevant & Unique Level 2 Refs MEDLINE/ EMBASE	No. of Relevant & Unique Level 3 Refs MEDLINE/ EMBASE	No. of Relevant & Unique Level 4 Refs MEDLINE/ EMBASE	No. of Relevant & Unique Level 5 Refs MEDLINE/ EMBASE
Homeopathy	65/98	56/76	35/55	0/3	2/1	4/5	1/15	28/31
Chelation	11/15	9/7	4/2	0/0	0/0	2/0	0/2	2/0
Flutamide	83/182	64/125	35/96	6/4	7/18	8/15	6/31	8/28
Riluzole	49/144	29/67	10/48	0/3	3/9	0.8	3/20	4/8
Holmiumlaser	91/82	84/70	36/22	0/0	7/4	6/6	4/3	19/9
Stents	51/111	39/82	18/61	1/2	6/29	3/3	7/20	1/7
Depression	61/101	51/77	35/61	0/3	7/9	13/21	8/16	7/12
Gamma knife	41/46	39/40	10/11	0/0	2/5	2/3	2/1	4/2
Total no. references retrieved (all searches)	452/779	371/544	183/356	7/15	34/75	38/61	31/108	73/97
Database search costs (not including online charges)[a]	US $90.40/ US $1,363.25	US $74.20/ US $952.00	US $36.60/ US $623.00	US $1.40/ US $26.25	US $6.80/ US $131.25	US $7.60/ US $106.75	US $6.20/ US $189.00	US $14.60/ US $169.75

[a] Cost per reference in *MEDLINE* was US $0.20. Cost per reference in *EMBASE* was US $1.75.

study). Furthermore, as all EMBASE searches retrieved more *irrelevant* references, the cost difference is even higher if only costs for retrieving relevant references are considered.

For these technologies, the EMBASE searches identified 356 references not identified in the initial MEDLINE searches. However, follow-up searches revealed that 215 (60%) of these references were indexed in MEDLINE; 58 of these references were not found in the original searches due to the longer time lag for indexing in MEDLINE. Obviously, such time lag can have a major impact on retrieval of recent studies. One hundred forty-one (40%) references were not in MEDLINE—either because the journal was not covered, or because it was only selectively indexed. It should be noted that several of these journals are indexed by other NLM databases, such as HealthSTAR. Fifty-five (40%) of these 141 references were in journals published in a language other than English. The breakdown of the 141 studies unique to EMBASE, by level of evidence, is as follows: 4 (3%) level 1 references; 16 (11%) level 2; 23 (16%) level 3; 70 (50%) level 4; and 28 (20%) level 5 references. Finally, 157 references were in MEDLINE, but were not found during the initial searches, either due to shortcomings of the search terms we used or shortcomings of the index terms assigned in MEDLINE.

DISCUSSION

No particular trends in the quality of references available in either database were apparent. MEDLINE indexes articles from approximately 3,700 journals, whereas EMBASE covers more than 2,900 journals, plus additional journals that are screened for pharmaceutical articles. This may indicate that much of the difference in retrieval must be due to time lag and indexing practices. Other factors that influence retrieval include the topic (EMBASE has a special area of focus in pharmaceuticals) and the origin and utilization of the technology (for example, technologies that have been approved and used in Europe prior to their introduction in North America might be expected to appear more in the European literature, another focus area of coverage for EMBASE). EMBASE searches retrieved more irrelevant references in all searches; however, in many instances, relevant references were missed in MEDLINE due to more restricted use of index terms. In the search on coronary stents, for example, use of the major descriptor limit appeared to eliminate many references of interest to the researchers.

In some of the searches, better search strategies would have improved retrieval. For instance, again using the example of the coronary stents search, additional terms such as coronary disease, coronary vessels, platelet aggregation inhibitors, and coronary arteriosclerosis would have picked up more of the relevant references not identified by the index term angioplasty. It could be argued that the OneSearch approach, using a single search strategy run simultaneously in both databases, may not fully exploit the unique search features of each database. Individual searches, designed specifically to use the terminology and search features available in each database, and run independently, may have provided better search results in all cases. However, this approach would have increased the costs and reviewer's time considerably.

The longer time lag for indexing in MEDLINE can have a major effect on retrieval rates. A new database created by the National Library of Medicine, PREMEDLINE, has been developed to help searchers find some of the most recent references in process for MEDLINE. These recent references are also contained

in the PUBMED database, both of which can be accessed free of charge via the Internet. The fact that all NLM databases accessed through the NLM's Internet site are now available at no cost is of tremendous significance to health researchers (6). Free Internet access to MEDLINE, HealthSTAR, and other NLM databases makes the cost difference between MEDLINE and EMBASE searches even more striking. (Charges still apply for access to MEDLINE through commercial vendors, such as DIALOG.)

CONCLUSION

As other studies have confirmed, each database contains relevant references not found in the other (10). Follow-up searches indicated that, when time lag for indexing and limitations of search strategies were accounted for, the number of references unique to EMBASE was considerably less than what seemed to be the case originally. Others have found that where time and budgets are limited, the costs associated with retrieving the additional EMBASE references may not be worthwhile (1). The availability of free access to MEDLINE and other NLM databases via the Internet may serve to further discourage the use of more expensive databases. Whether the unique information from these databases would influence the results of a health technology assessment remains to be determined. However, given the importance of obtaining comprehensive and current information in assessing health technologies, the use of additional databases, such as EMBASE, may be essential despite the costs involved.

REFERENCES

1. Burdett, S., Stewart, L. A., & Tierney, J. F. Searching bibliographic databases: Costs and benefits (abstract). Presented at the 2nd Conference on the Scientific Basis of Health Services, Amsterdam, The Netherlands, October 1997.
2. Canadian Task Force on the Periodic Health Examination. *The Canadian guide to clinical preventive health care*. Ottawa: Canada Communication Group - Publishing, 1994.
3. Dickersin, K., Scherer, R., & Lefebvre, C. Identifying relevant studies for systematic reviews. *BMJ,* 1994, 309, 1286–91.
4. Embase: A healthy Medline alternative? *Searcher: The Magazine for Database Professionals,* 1996, 23.
5. *EMBASE vs. MEDLINE: An overview*. Norward, MA: SilverPlatter Information, Inc., July, 1996.
6. Free MEDLINE: PubMed and Internet Grateful Med. Rockville, MD: U.S. National Library of Medicine (NLM), 1997 (1 screen). Available from URL: http://www.nlm.nih.gov/.
7. Modig, Z. D. EMBASE, the Excerpta Medica database: Quick and comprehensive drug information. *Bibliotheca Medica Canadiana,* 1995, 16, 152–55.
8. Power, E. J., Tunis, S. R., & Wagner, J. L. Technology assessment and public health. *Annual Review of Public Health,* 1994, 15, 561–79.
9. Wallace, S., Daly, C., Campbell, M., et al. After MEDLINE? Dividend from other potential sources of randomised controlled trials (abstract). Presented at the 2nd Conference on the Scientific Basis of Health Services, Amsterdam, The Netherlands, October 1997.
10. Woods, D., & Trewheellar K. MEDLINE and EMBASE complement each other in literature searches (letter). *BMJ,* 1998, 316, 1166.