How to do (or not to do) . . .

Searching bibliographic databases effectively

JOHN E EYERS

London School of Hygiene and Tropical Medicine, UK

The ability to search bibliographic databases effectively is now an essential skill for anyone undertaking research in health. This article discusses the way in which databases are constructed and some of the important steps in planning and carrying out a search. Consideration is given to some of the advantages and limitations of searching using both thesaurus and natural language (textword) terms. A selected list of databases in health and medicine is included.

> *Knowledge is of two kinds. We know a subject ourselves, or we know where we can find information upon it.*
> Dr Samuel Johnson, English essayist and lexicographer, 1709–1784

Introduction

It is estimated that there are between 20 000 and 30 000 journals currently published in medicine and health.[1] Any attempt to search even a small part of this literature could not be contemplated without the use of online databases. Access to databases is usually made available either within institutions, or via commercial providers such as Silverplatter or Ovid, Knight-Ridder, STN or DIMDI or, increasingly, via the Internet. The National Library of Medicine (USA) has recently made many of its health-related databases available via its website, of which *Medline* may be searched free (*http://www.ncbi.nlm.nih.gov/PubMed/*). Most providers list their databases either on their Internet websites or in catalogues, which will give details of subjects and years covered, as well as the cost of access or subscription.

Which database(s) do I search?

In the field of health in general, the choice of databases to search is not easy. Some of the literature may be scattered across several disciplines making searching and retrieval well-nigh impossible, while restricting your searches to biomedical databases alone could result in important citations being missed. The multidisciplinary nature of the subject, encompassing anthropology, economics, demography, psychology and so on, inevitably means that many health-related articles will appear in the mainstream journals of these disciplines, which are unlikely to be included in the biomedical databases. While it is always best to use as many databases as possible in order to cover as much of the literature available on the subject, it is usual, on grounds of cost and availability, to choose those databases that are most well-known in the field.

Most databases such as *Medline* and *Health-Star* do not attempt to be comprehensive in their coverage of the literature, but will restrict the number of journal titles or sources indexed; these may include only journals which are peer-reviewed and well-established in their field. *Health Policy and Planning*, for example, is not indexed in *Medline*, which has a clinical/life sciences emphasis, but is indexed in *Health-Star*, which places a greater emphasis on the multidisciplinary aspects of health care delivery.

Each database will contain unique references as well as a greater or lesser degree of overlap with other databases, and some may contain not only journal references but additional information from reports, conferences, theses and books. Other criteria for inclusion may include an emphasis on English language sources, on journals with a bias towards health care

in the West, or in including articles selectively from journal issues and not cover-to-cover. Most database producers publish their selection policies as well as a list of journals or sources included in the database. This source list will help determine the usefulness of the database in your field, and whether any journals or other source materials considered essential are omitted.[2]

A further consideration in the choice of database will be its currency. Some databases take much longer to index articles than others. These are often databases which place a greater emphasis on the indexing process and in the assigning of thesaurus or subject headings to each article, which in the end considerably assists in the search and retrieval process, through the use of this 'controlled vocabulary'. However, the result of this process may be a long delay (in some cases over a year) between the publication date of the article and its appearance in the database.

Other databases which do not index articles comprehensively expect the searcher to retrieve articles 'free-text' or using the 'natural language' terms found in the title or abstract of the article. The advantage of these databases is that they tend to be more up-to-date. Since the success of any search is dependent on understanding the distinction between these two main types of bibliographic database, a more detailed discussion of the structure of databases is given in the next section.

How databases are organized

Databases generally fall into two main types – the database using a structured thesaurus, and the database which relies on the searcher identifying keywords from the text of the reference (usually the title or abstract). Structured databases which use a thesaurus, such as *Medline*, *EMBase* and *Health-Star*, are organized in such a way that searchers can use the thesaurus terms, which have been assigned as subject headings to each article, thereby eliminating the need to use textwords from the title or abstract. It is also sometimes possible to use the system in which these thesaurus terms are arranged in a hierarchy of broader, narrower and related terms. This allows a category of terms to be searched as a group (often called 'exploding') without the need for each term within the group to be searched separately. Once a thesaurus term or group of terms has been selected, subheadings can then be selected to focus on

particular aspects of the subject being searched. For a disease heading, subheadings such as *drug therapy*, *pathology* or *prevention and control* can be used; for a subject such as health planning, subheadings such as *trends*, *standards* or *economics* can be used. In most databases such as this, the searcher is taken through this process of selecting and 'exploding' thesaurus terms and subheadings in a systematic way.[3]

A major disadvantage of the thesaurus system, particularly in rapidly developing areas of healthcare, is that it may not include newly introduced terms and concepts in the literature. New thesaurus terms are usually only adopted in databases once they have been established in the literature for some time. A recent innovation introduced by some database providers is a device which automatically 'maps' free-text terms to the appropriate thesaurus terms, making the process much simpler if you are unfamiliar with the thesaurus system.

The other type of database, which requires a different approach to searching, does not have this in-built structure. Searching requires much greater thought and preparation in the choice of keywords that the author has used in the title or abstract. This would include identifying synonyms that authors use in an interchangeable way to describe the same subject, as well as different spellings of the same word (e.g. English and American) and word endings (e.g. singulars and plurals). It is a significant advantage if you are familiar with the literature and terminology of a subject because you will know what terms are commonly used. Having some papers on the subject available to hand will also be helpful for checking commonly used terms. However, despite careful and laborious preparation of your search profile, there is a strong chance that papers will be missed, where authors have not used any of the keywords that you have chosen. The chance of this happening is minimized in those databases that use the thesaurus system because the indexer, in most cases, will have supplied the important keywords from the thesaurus. It is, of course, possible to use both methods of searching in those databases that have a thesaurus system, and indeed this is advisable if you are doing a literature search for a systematic review, to take account of any inadequacies in the indexing process.[4-5]

Finally, a different variant to subject searching described above is citation searching, offered in

Scisearch (Science Citation Index), Social Scisearch (Social Science Citation Index) and *Arts & Humanities Citation Index.* This is based on searching for articles that have cited known papers in the field, on the principle that such articles will be related to the topics discussed in these known papers. The advantage of this search method is that references retrieved will be more recent that the known papers searched, and may also retrieve papers from other disciplines, which is unlikely to occur when searching conventional subject-specific databases. This is particularly useful in identifying applications of research methodologies in other disciplines where there is little published in your own field. Citation analysis is also used, controversially, to assess the number of citations to faculty papers as one of the means of measuring research quality in universities.

The search process

Search example: *Cost-effectiveness of immunization programmes in Africa*

The above example has been selected to illustrate the search process. The first step in any search is to define your topic and determine what boundaries or limits you are setting. This involves identifying and listing concepts or keywords in the field and any related areas that you think are relevant, and deciding whether you want to limit your search to, say, a few review articles or to a limited time period or to articles in a particular language. In our example, we would begin by identifying the key elements of our search which are:

- cost-effectiveness
- immunization
- Africa

Each of these elements will need to be searched separately before they can be combined together. It is at this point that you will need to decide whether you will search using thesaurus terms if available or whether you will search 'free-text' using words that are found in the title or abstract of the article, or both methods. In our example, the appropriate thesaurus terms for cost-effectiveness, immunization, and Africa would be selected separately and searched. In a free-text search, account will need to be taken of variations in word endings, synonyms etc.

A sample set of terms in which the concept of cost-effectiveness is expressed in a title or abstract may include *cost-effective, cost-benefit, cost(s), budgetary, financial, cost-utility, economic evaluation.* Similarly, terms for *immunization* would need to include its alternative spelling *immunisation,* as well as *vaccination.* Words can be truncated using symbols such as * or $ or ? to take account of word endings; for example, *cost-effective** would retrieve both *cost-effective* and *cost-effectiveness.* Searching for geographical terms such as *Africa* may be problematic. Using only the term *Africa** to retrieve *Africa, African* and *Africans* may miss useful papers relating to individual African countries. Selected country names may also need to be searched.

Combining terms together

Combining search terms together can be done using linking Boolean operators which are OR, AND and NOT. The OR operator is used to link synonyms or related terms together, which broadens the search, e.g. *immunisation* OR *vaccination.* The AND operator is used to retrieve references in which all linked terms must be present, e.g. *immunisation* AND *Africa*; this will narrow the search. The third operator NOT is used to exclude references containing specified terms, e.g. *immunisation* NOT *Africa.* Caution must be exercised in the use of this operator in order that useful papers are not excluded (see Figure 1).

Once the final combinations have been made, further limits can be set to retrieve more focused aspects, such as reviews, letters, editorials or papers in a particular language. The next stage is to display the results in order that relevant references can be 'marked' for subsequent printing or saving/downloading as a file. Take the opportunity also to browse through the title, abstract and thesaurus field of each reference for other relevant terms which were missed in the initial search, so that they can be used to refine the search still further. A search strategy that has been tested and adapted to retrieve the required result can usually be saved as a file for subsequent running against database updates. Above all, a flexible and imaginative approach to searching is the key to successful retrieval.

References are usually saved to a file in ASCII format and can therefore be imported into wordprocessing packages for editing or incorporation into documents. They may also be imported into personal
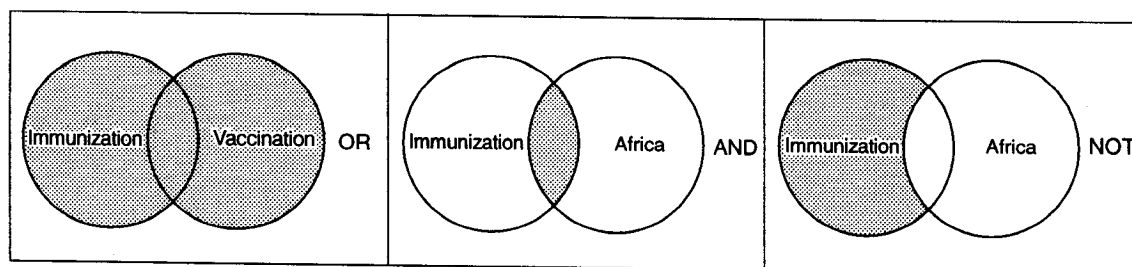
Figure 1. Combining search terms

bibliographic software packages, such as *Reference Manager*, *Endnote* or *Papyrus*, for the creation of personal databases tailored to individual research interests.

References

[1] Hernando S. 8 Jan 1998. Publication explosion – summary. *Lis-Medical* [Online discussion list]. Available from *http://www.mailbase.ac.uk*
[2] For example, see *List of Journals Indexed in Index Medicus* published each year with *Index Medicus*.
[3] Lowe HJ, Barnett GO. 1994. Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. *JAMA* **271:** 1103–8. – An excellent introduction to the *Medline/Health-Star* thesaurus.
[4] Fremer E. 1995. Understanding MeSH for literature searches. *JAMA* **273:** 1844.
[5] McKee M, Britton A. 1997. Conducting a literature review on the effectiveness of health care interventions. *Health Policy and Planning* **12:** 262–7.

Further reading

Armstrong CJ, Large JA (eds). 1992. *Manual of Online Search Strategies*, 2nd ed. Aldershot: Ashgate Publishing.

Biography

John Eyers is Deputy Librarian at the London School of Hygiene and Tropical Medicine. He has a Masters degree in Library Studies and is a Member of the Institute of Information Scientists. His main interests are in teaching the use of online databases to postgraduate students and in the dynamics of scientific communication.

Correspondence: John Eyers, Library, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. E-mail: j.eyers@lshtm.ac.uk

Appendix

This list is a very small selection of the large number of databases in the field. For a comprehensive list of databases see *Gale Directory of Databases*, 2 vols. Detroit, MI: Gale Research, 1998, or consult the catalogues or websites of database providers such as SilverPlatter (*http://www.silverplatter.com*), Ovid (*http://www.ovid.com*) or Knight-Ridder (*http://www.krinfo.com*).

Selected health and medical databases

AIDSline – AIDS information
Bioethicsline – Ethical issues in biomedical research and healthcare
CAB-Health – Public health and infectious diseases; excellent on developing countries
CINAHL – Nursing and allied health
Cochrane Library – The standard resource for evidence-based health
Embase – European equivalent of Medline with an emphasis on the European literature
HealthStar – Healthcare delivery: administration, economics, policy and planning
Histline – History of medicine
HMIC — Health management, health services; emphasis on the UK
Medline – The most well-known database on general medicine and health
Popline – Population and reproductive health; good for developing countries studies
Psycinfo – Psychology, psychiatry, sociology
Social Scisearch – Good general social sciences database; citation searching available
Sociofile – Sociology and related social sciences