

How to search medical databases: part 1

John Eyers unravels some of the mysteries



Searching the medical literature comprehensively would be impossible today without using bibliographic databases. Many, such as Pubmed/Medline and Embase, attempt to cover all specialties in medicine and health, but many more cover specialist topics such as global health, health economics, evidence-based medicine, etc. Most will be selective about what publications are included, whether it is only journals (PubMed indexes about 5000) or also including other publications such as books, reports, conference proceedings, or theses; other criteria for inclusion may also include whether a journal is peer-reviewed (where each article published has been through a process of impartial assessment by other experts in the field), whether a journal has been established for some time, and the language of publication.

Most databases are searched using keywords that appear in the title of the publication or its abstract. This is called free text or natural language searching and immediately presents a problem if the search is to be effective because account needs to be taken of the different ways in which authors describe their subject in the title and abstract. These may include variations in English/US spelling (e.g. pediatric/paediatric), word endings (e.g. therapy, therapies, therapeutic) as well as synonyms (e.g. iron, folic). The more account taken of these variants, the better the search will be, but no search can ever be perfect!

To illustrate the basic formulation of a search, let us take a simple example in which we search for papers in PubMed on malaria in children. Each concept (malaria, children) is searched separately and then combined together to retrieve publications that represent both concepts together. Before that, we must identify potential keywords and their word endings, variant spelling, and synonyms that any author could use to signify each concept, so we might identify for each:

1. **Malaria:** malaria*, Plasmodium, falciparum, vivax.
2. **Children:** child*, infant* (other terms like youth* or young people/person*, schoolchild*, etc could also be included).

The symbol * (some databases use \$ or ?) is used to denote truncation, so that for child* the following will be retrieved: child, children, child's. Infant* will retrieve infant, infants, infantile, etc. Using this symbol will make a big difference in the number of papers retrieved. Each set of terms is then typed into the database search box:

As a member of HIFA2015 (<http://www.hifa2015.org/>), John Eyers believes one of the biggest challenges for improved health in less-developed countries is ensuring that equal access to appropriate and evidence-based health information is made available to all.

1. Malaria* OR plasmodium OR falciparum OR vivax
2. Child* OR Infant*

Notice the OR between each term; this is one of the two main linking (Boolean) operators (the other is AND) used to define the relationship between keywords. Using the OR operator adds the number of references together of all the terms in the set thus increasing or broadening the search, whereas the AND operator is used to limit or narrow the search, usually to achieve the final result (see the Venn diagrams below).

Finally, the two sets of results from each search are then combined using the search numbers, so in this example we have: 1 AND 2 (malaria* OR plasmodium OR falciparum OR vivax) AND (child* OR infant*)

Figure 1 The OR operator: each term added will increase the number of papers retrieved, thus broadening the search

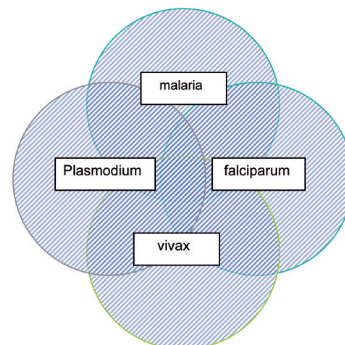


Figure 2 The AND operator: where the two circles overlap is the final required result



This should give the intended focused result but invariably further refinements will need to be made to increase the number of relevant articles. For example, you might specify review articles only or English language or papers concerning a single country.

In the next issue, I shall describe the use of thesaurus/controlled language searching (such as can be used in PubMed/Medline) for a more effective search.

References

1. Eyers JE. Searching bibliographic databases effectively. *Health Policy and Planning* 1998; 13: 339–42. Free to download at <http://heapol.oxfordjournals.org/cgi/reprint/13/3/339.pdf>.
2. Holzer L, Eyers JE. Bibliographic databases. *Student BMJ* 2008; 16, 366–7. Free to download at <http://archive.student.bmj.com/issues/08/10/education/336.php>.