Searching a biomedical bibliographic database from the Ukraine: the Panteleimon database

Key Messages

Implications for Policy

1 The development of CENTRAL should aim to include citations from a wider range of sources, utilising databases as they become more easily available
2 CENTRAL contributors should be supported in submitting search results from multi-lingual databases
3 All health sciences researchers should consider supplementing searches on commonly used databases with smaller foreign language databases to avoid language bias
4 Identification of accessible biomedical databases that are not subsets of more commonly used MEDLINE and EMBASE

Implications for Practice

1. Multi-lingual search strategy resource for RCT search strategies used in the Cochrane Collaboration – made freely available
2. CENTRAL software to cope with non arabic characters
3. Native language plus English should be used where possible for comprehensive searches

Introduction

The Ukraine, 2.5 times the landmass of the UK, has borders with Romania, Hungary, Slovakia, Poland, Moldavia, Belorussia and Russia and a large Black Sea coast including the Crimea. Its 48 million people gained independence in 1991 and, after Russia, the Ukrainian republic is by far the most important economic component of the former Soviet Union. Its economy is strong, unemployment low and reform progresses at a steady pace. The last decade has seen the Ukraine undergo a huge political and economic upheaval. Health care struggles with high levels of infant mortality, high consumption of tobacco with all the attendant morbidity and mortality and high suicide rates.[1] Research and [probably] dissemination of research declined in the early 1990s but a strong tradition of scientific endeavour has seen a resurgence of medical research in latter years.

Literature searching over a broad range of sources can produce a more representative picture of research findings compared with searches of readily available databases.[2] Cochrane systematic reviews aim to identify all relevant trials in order to increase the precision of the final summary results. Studies that are difficult to locate may have results that substantially differ from those that are readily accessible. For example when Egger *et al.* searched German medical journals for randomised controlled trials (RCTs) and then searched for other RCTs published in English from the same authors they found that quality was constant but that the size of the estimates of effect was not.[3] On average, these German trialists published studies with 'positive' results in English-language journals but studies with 'negative' results or those showing no difference in the effect were more likely to be published in German-language journals – a phenomenon known as language bias. The Cochrane Collaboration is acutely aware of publication and language bias and makes extensive efforts to identify all trials and to make them accessible through the Cochrane Central Register of Controlled Trials (CENTRAL), published in the Cochrane

Library. However, many national and subject-specific databases exist which have not yet been searched for CENTRAL and therefore studies conducted in a number of countries are likely to be under- represented. For example, searching the Australasian Medical Index (1966-2000) identified 512 relevant citations (RCTs or other Controlled Clinical Trials known as CCTs) but only 12% could also be found in CENTRAL.[4] Inclusion of the remainder into CENTRAL improves its international coverage.

The Panteleimon database is available on the Internet. It is a public access, user-friendly biomedical bibliographic database and can be searched in three different languages: English, Russian and Ukrainian.[5] Its subject coverage includes medical, pharmaceutical and chemical publications – published in scientific journals in the Ukraine and Russian Federation from 1998. Some publications are included which precede that date. This study describes the formulation of a search strategy for Panteleimon and the comparison of the results of that search with records included in the Cochrane Library's CENTRAL database.

Objectives

To systematically search the Panteleimon database, identify citations to randomised controlled trials and investigate how comprehensive the coverage of the Cochrane Library's CENTRAL database is for the literature of this region.

Methods

We identified the Panteleimon database by searching the Internet (Ukrainian AND bibliographic AND medical). We ran searches in three languages (English, Russian and Ukrainian) on the Publications database within Panteleimon to identify articles reporting randomised controlled trials (see table 1). The citations are translated into all three

languages, so an individual citation appears three times in the database but in a different language each time. One unique ID is allocated to the three alternative language citations that all reference the same article. Panteleimon provides basic free-text searching with automatic truncation of terms and the option to combine terms using Boolean logic. We found that wildcard operators did not function and there was no thesaurus. Individual bibliographic fields including Author's Keywords can be searched. We imported all the records identified by the search strategies into an MS Access database and de-duplicated them. The three alternative language citations were combined to reconstruct one record per unique citation based on the unique ID. We then manually inspected all the records to identify only those articles which appeared to be reports of RCTs on the basis of the title, and abstract where available. Finally we searched for those citations in the CENTRAL database of the Cochrane Library.

Results

Search recall and precision using alternative translations

It is unusual that each language essentially searches what behaves like a separate database with a separate set of data in each language. Searching using Russian terms produced a list of Russian citations, searching with Ukrainian terms produced a different list of Ukrainian citations, English terms produced another different list of English citations. Some citations were common to all searches but some were only found by searching one language. Each citation had an English record, a Ukrainian record and a Russian record with only a unique identifier in common. The relational MS Access database allowed us to piece records together into one multilingual whole that represented a single citation. Table 1 shows the three different search strategies used and the number of citations identified by each of those phrases.

Total results using English, Russian and Ukrainian terms

Using all three languages, our search strategies identified a total of 821 unique citations, of which 210 seemed to be relevant controlled trials, giving a precision of 26%. That is, for every 3.9 citations we inspected, one was a randomised, or possibly randomised, study.

Precision

The terms used in the three languages gave a different precision to selecting relevant citations, i.e. those that are likely to be about RCTs or CCTs. English search terms gave the most sensitive search, identifying 191 of the 210 RCT/CCT citations (91%). Russian search terms found 56% (117 of the 210 RCT/CCT citations) and Ukraine search terms found 21% (45 of the 210 RCT/CCT citations).

Recall

Russian search terms identified the most citations within the database, finding 476 of the 821 unique citations (58%). Using English terms only found 452 of the 821 unique citations (55%). Searching solely in Ukrainian found 17% (138 of the 821 unique citations)

Proportion of citations already found in CENTRAL – what's new to Cochrane

When we searched the CENTRAL database on the Cochrane Library for the 210 citations we found 53 (25%).

Healthcare speciality of the trials

We broadly classified the studies according to health care specialty (Table 2).

Discussion

Panteleimon is an easily accessible bibliographic database, available free of charge on the Internet, offering easy access to the Ukrainian biomedical literature. The English language retrieval functions well picking up most of the reports of RCTs/CCTs (91% precision). The lower recall of 55% indicates the need to search using Russian and Ukranian terms for completeness. The overall precision of our search (26%) compares favourably with a search for RCTs in EMBASE, carried out by the UK Cochrane Centre, where 70,000 reports of RCTs were identified from 300,000 records downloaded (precision 23%).[6]

There has been a proliferation of bibliographic databases published on the Internet. This provides opportunities for those contributing to CENTRAL to create a more truly international source of RCTs rather than a subset of the more frequently used databases such as MEDLINE and EMBASE. If this opportunity is not taken, CENTRAL will remain systematic, but not comprehensive and, perhaps, like so many of its predecessors, a not altogether unbiased source of material.

Acknowledgements

We would like to thank Gill Rizzello and Carol Lefebvre for their comments on the manuscript. This work contributed to the PRACTIHC project's survey of randomised studies from low-middle income countries (http://www.practihc.org/).

References

1. Central Intelligence Agency. CIA - The World Factbook [Web Page]. Available at http://www.cia.gov/cia/publications/factbook/. (Accessed April 2005).

2. 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.

3. Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. Lancet 1997; 350(9074):326-9.

4. McDonald S. Improving access to the international coverage of reports of controlled trials in electronic databases: a search of the Australasian Medical Index. Health Info Libr J 2002; 19(1):14-20.

5. The Panteleimon Project. PANTELEIMON - Medical and Biological Databases [Web Page]. Available at http://www.panteleimon.org/. (Accessed 22 October 2004).

6. Eisinga A, Lefebvre C. Closing the gap: identifying reports of randomized trials in EMBASE for inclusion in the Cochrane Central Register of Controlled Trials (CENTRAL). 12th Cochrane Colloquium.

Table 1 The search phrases and number of records retrieved

English phrase	Number of records	Russian phrase	Number of records	Ukrainian phrase	Number of records
Abstract: randomi	147	Abstract: Рандомиз	114	Abstract: Рандомиз	114
Article: randomi	26	Article: Рандомиз	23	Article:Рандомиз	0
Keywords: randomi	6	Keyword: Рандомиз	5	Keyword: Рандомиз	5
Abstract: double-blind	14	Abstract: двойного слепого	6	Abstract: двоїстий сліпий	0
Abstract: double blind	15	Article: двойного слепого	4	Article: двоїстий сліпий	0
Article: double-blind	4	Keyword: двойного слепого	0	Keyword: двоїстий сліпий	0
Abstract: allocate	36	Abstract:распределить	2	Abstract: розподілити	5
Article: allocate	0	Article: распределить	0	Article: розподілити	0
Keyword: allocate	0	Keyword:распределить	0	Keyword: розподілити	0
Abstract: assign	102	Abstract: определить	238	Abstract: визначити	56
Article: assign	3	Article: определить	0	Article: визначити	0
Keyword:assign	0	Keyword: определить	0	Keyword: визначити	0
Abstract: clinical trial	81	Abstract: клиническое исследование	13	Abstract: клінічна дослідження	0
Article: clinical trial	22	Article: клиническое исследование	8	Article: клінічна дослідження	0
Keywords: clinical trial	8	Keyword: клиническое исследование	1	Keyword: клінічна дослідження	0
Abstract: placebo	67	Abstract: плацебо	58	Abstract: плацебо	58
Article: placebo	7	Article: плацебо	8	Article: плацебо	8
Keyword:placeo	0	Keyword: плацебо	0	Keyword: плацебо	0
Abstract: crossover	12	Abstract: скрестить	0	Abstract: схрестити	0
Article: crossover	0	Article: скрестить	0	Article: схрестити	0
Keyword: crossover	0	Keyword: скрестить	0	Keyword: схрестити	0
		Abstract: перекрест	56	Abstract:перехрест	1
		Article:перекрест	6	Article:перехрест	0
		Keyword:перекрест	5	Keyword:перехрест	0
Total including duplicates	550		557		247
Total unique citations	452		202		128

Table 2 Reports of trials by health care sub-speciality

Speciality	Numbers of reports of RCT/CCTs
Anaesthesiology	3
Cardiology	70
Dentistry	1
Dermatology	9
Endocrinology	4
Fertitily regulation	1
Gastroenterology	3
Gerontology	1
Haematology	4
Immunology	4
Infectious Diseases	8
Internal Medicine	10
Neonatology	4
Nephrology	3
Neurology	10
Neurosurgery	1
Obstetrics and Gynaecology	16
Occupational diseases	1
Oncology	10
Ophthalmology	1
Orthopaedics & Traumatology	4
Otorhinolaryngology (ENT)	9
Paediatrics	8
Psychiatry	3
Respiratory medicine	16
Surgery	4
Total	208

Figure 1 Results of searches in different languages in Panteleimon

The bold numbers relate to references of **RCTs or CCTs** found, and the non-bold number is the total found using the language/s.

For example 30 references could only be found with the Russian search phrases, and of these, 14 were RCTs or CCTs)