Comprehensive searching for systematic reviews: a comparison of database performance

Fiona Beyer, Kath Wright
Centre for Reviews and Dissemination, University of York (fiona.beyer@york.ac.uk)

Background

A thorough search for eligible studies is one of the most important contributions to unbiased conclusions in a systematic review.[1] Guidelines for producing systematic reviews recommend that the search for potentially eligible studies be as comprehensive as resources allow.[1, 2] Resources for carrying out systematic reviews are usually limited, so it may not be possible to search all potentially productive sources. Little guidance exists about how to choose or prioritise sources or when it is appropriate to stop searching. Booth recently published an overview of methods to help decide when to "desist" searching, but noted that few of them have been investigated empirically.[3]

Objectives

In this study, we investigated which of 16 databases yielded the studies included in a systematic review of the effectiveness of an extensive range of interventions for managing frozen shoulder (adhesive capsulitis), or painful, persistent stiffness of the shoulder joint. We aimed to:

(i) compare the individual contribution of databases towards identifying studies included in the review.

(ii) identify the best combination of databases to retrieve all included studies.

Methods

We searched nineteen databases for this systematic review. Sixteen were searched for eligible primary studies and were included in the present analysis (Table 1), and three (CDSR, DARE and the HTA Database) were searched for existing reviews from which to harvest eligible studies. The inclusion criteria specified randomised controlled trials (RCTs), quasi-experimental studies (i.e. those with a control group), and case series where controlled studies were not available.

We recorded whether each included study was: (i) retrieved by the search strategy devised for each database, and (ii) indexed on each database regardless of retrieval. Whether a study is indexed on a database gives a true assessment of the performance of the database (rather than that of the search strategy). To establish whether studies were indexed even if they were missed by the search strategies, we searched each database for each study using title, author and journal title independently. For each strategy, we calculated recall (number of included studies retrieved from the database divided by the total number of included studies), precision (number of included studies retrieved from the database divided by the total number of studies retrieved by the database), number needed to read (NNR) (1 divided by the precision), and unique yield (number of studies retrieved only by this database).

Table 1: Recall, precision and NNR for each strategy; the review contains 31 included studies

Database	Total yield	Unique yield	Yield of included studies	Recall (R) (%)	Preci-sion (P) (%)	Balance (RxP)	NNR (1/P)
SCI	3140	0	27	87	0.86	75	116
EMBASE	3771	1	26	84	0.69	58	145
MEDLINE	3117	0	25	81	0.80	65	125
CENTRAL [RCT]	539	1	22	76	4.08	310	25
PEDro [RCT]	718	0	15	52	2.09	108	48
PASCAL	388	0	12	39	3.09	120	32
BIOSIS	800	0	10	32	1.25	40	80
CINAHL	1757	0	10	32	0.57	18	176
Clinicaltrials.gov[RCT]	28	0	4	14	14.29	197	7
CPCI	82	0	2	6	2.44	16	41
PreMEDLINE	82	0	1	3	1.22	4	82
MANTIS	78	0	1	3	1.28	4	78
LILACS	298	0	0	0	0.00	-	-
NHS-EED	9	0	0	0	0.00	-	-
HMIC	8	0	0	0	0.00	-	-
NTIS	8	0	0	0	0.00	-	-

[RCT] *recall calculated using 29 RCTs/CCTs as denominator (these databases by definition only contain these study designs)*

Results

The review included 31 studies: 28 RCTs, one controlled clinical trial (CCT) and two case series. Thirty studies were present in at least one of the databases searched, and one was located through checking the reference list of another systematic review. Search strategies devised for the Cochrane Central Register of Controlled Trials (CENTRAL) and EMBASE each retrieved one unique study, while four databases (LILACS, NHS-EED, HMIC, NTIS) contained none of the included studies.

The yield of included studies indexed in the databases ranged from 0% to 90% (median 23%). Recall of the search strategies ranged from 0% to 87% (median 23%), and precision from 0.0% to 14.3% (median 1.04%).The reasons for non-retrieval by search strategies were unclear: in most cases search strategies did match the unretrieved records. EMBASE performed best in terms of recall, indexing 28 (90%) of the included studies. However, it displayed low precision (0.74%, NNR=135). CENTRAL, with its restriction to RCTs, performed the best in terms of combined recall and precision (89% and 4.61% respectively, NNR=22). Recall, precision and NNR are reported for all strategies in Table 1.

Had CENTRAL been the only database used, six (19%) of the included studies would not have been retrieved. As well as the two case studies (which do not fulfil CENTRAL's inclusion criteria), four RCTs were missed; two of these were not on MEDLINE or EMBASE (although one was on PreMEDLINE), and two were indexed on MEDLINE as non-controlled publication types. Adding either MEDLINE (including PreMEDLINE) or Science Citation Index (SCI) to CENTRAL results gave a combined sensitivity of 97% (i.e. all the studies retrieved electronically), with NNR of 125 and 123 respectively.

Since one study was retrieved only through reference checking, no combination of databases retrieved all the included studies. The search strategies devised for this review required one of two possible combinations of databases to retrieve the other 30 studies: CENTRAL, EMBASE and Science Citation Index (SCI); or CENTRAL, EMBASE, MEDLINE and PreMEDLINE (NNR=257 and 259 respectively). Two combinations of databases indexed all the studies (irrespective of retrieval by the search strategies): CENTRAL and SCI; or CENTRAL, MEDLINE and PreMEDLINE.

Conclusions

We found that at least three databases and reference checking were required to locate all the included studies. Two databases indexed all the studies that were electronically available; given that deficiencies in the search strategies were not always responsible for the failure to retrieve studies, this demonstrates an advantage of searching multiple databases. As this is a case study of a single systematic review, the generalisability of the findings is limited. However, the approach described can be used to analyse other systematic reviews and build the evidence required to inform prioritisation of databases.

References

1. Centre for Reviews and Dissemination. *Systematic Reviews: CRD's guidance for undertaking reviews in health care*. York: University of York; 2009.

2. Lefebvre C, Manheimer E, Glanville J. Chapter 6: Searching for studies. In: Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions Version 510* [updated March 2011]: The Cochrane Collaboration; 2011.

3. Booth A. How much searching is enough? Comprehensive versus optimal retrieval for technology assessments. *International Journal of Technology Assessment in Health Care*. 2010;26(4):431-5.

CRD is part of the National Institute for Health Research (NIHR) and a department of the University of York. The opinions expressed here are those of the authors alone and necessarily those of the NIHR, NHS or Department of Health.