

•Research methods in psychiatry•

Literature searches in the conduct of systematic reviews and evaluations

Xiaochun QIU, Cheng WANG*

Summary: Performing a literature search is an important part of performing a systematic review or a meta-analysis of biomedical literature, which have now become the gold standards for determining what qualifies as ‘evidence-based’ medicine. Combining searches of English-language databases and the large Chinese-language databases can identify new, potentially important, sources of data that are not include in the traditional English-only reviews. Selection of a restricted subset of databases for conducting the literature search or using inappropriate methods to identify appropriate articles within each database can lead to biased results and incorrect conclusions. This article introduces common English and Chinese databases, describes the search engines available for conducting searches, discusses the basic methods and common pitfalls of conducting searches, and provides an example of a search to highlight these issues.

Keywords: literature review; publication bias; databases; bibliography; systematic review; meta-analysis

[*Shanghai Arch Psychiatry*. 2016; **28**(3): 154-159. doi: <http://dx.doi.org/10.11919/j.issn.1002-0829.215008>]

Systematic evaluation of literature is a relatively new method in biomedical research. If sufficient studies with comparable methodologies are identified, a meta-analysis that pools the results of the original studies – considered a type of secondary data processing^[1] – can be conducted. The results of such systematic reviews and meta-analysis are often used as the highest level of evidence available to support changes in the clinical guidelines for the treatment of various illnesses (i.e., ‘evidence based medicine’). However, biases in literature searches that occur because of incomplete coverage of databases or errors in the search strategy can seriously undermine the internal validity of systematic reviews and meta-analyses.^[2,3] Researchers conducting systematic reviews and meta-analyses must carefully choose appropriate databases and use multiple search methods to find all relevant publications for the topic of interest. This issue has become more important as an ever-increasing proportion of the global medical literature is appearing in non-English publications, particularly Chinese and Spanish.

1. Selection of databases

The Ulrich’s Periodicals Directory currently lists more than 56,800 active academic journals including more than 23,500 peer-reviewed journals.^[4] About half of these journals are life science or biomedical journals published by over 2000 publishers; and about 26.6% of these are in non-English languages. It is almost impossible to search all of these journals one by one, so a variety of abstract-based databases that cover different subsets of these journals have been developed to assist clinicians and researchers in the identification of relevant literature when deciding how best to treat a specific class of patients or when conducting systematic reviews or meta-analyses. The coverage of journals and the timeframe of the included publications for each database is different, and therefore each database has its unique strengths and limitations. Some databases are largely focused on biomedical research (e.g., MEDLINE and EMBASE), some are limited to clinical trials (e.g., the CENTRAL database of the Cochrane Collaboration), some include a stronger health services component

Shanghai Jiao Tong University School of Medicine Library, Shanghai, China

*correspondence: Dr. Wang Cheng, mailing address: RM 502, Shanghai Jiao Tong University School of Medicine Library, RD Chongqing 280, Shanghai.
E-Mail: chwang@shsmu.edu.cn

(e.g., CINAHL), some include social science topics relevant to health (e.g., the Social Science Citation Index in the Web of Science search engine), some are focused on a specific field (e.g., PsychInfo collects articles from publications relevant to psychology), some are limited to non-English languages (e.g., SinoMed only includes Chinese-language journals from mainland China), some are region-specific (e.g., LILACS is focused on Latin America, and TEPS is limited to journals published in Taiwan), and some are country-specific (e.g., the Cinii database in Japan and the IndMed database in India). Researchers conducting literature searches need to understand the coverage and limitations of the various databases and select the databases that provide the best fit for the topic of interest. As the proportion of global medical literature appearing in non-English languages increases (particularly Chinese and Spanish) it is increasingly important to include databases that provide good coverage of journals in other languages.

Searches of the major international databases such as MEDLINE, EMBASE, and PsycINFO can be conducted using their built-in search systems or by using authorized third-party platforms such as OVID and Web of Knowledge. The search expressions are slightly different in different systems. One advantage of OVID is that it allows users to specify the distance between keywords using 'adj'. For example, the term 'generalized adj/2 anxiety' in OVID means 'generalized' and 'anxiety' should be within the distance of two words. Therefore, the search finds articles that contain 'generalized social anxiety or 'generalized anxiety'. This function has not been made available in Pubmed and other platforms. Besides user-specified keywords, most biomedical databases support the use of medical subject headings terms (MeSH terms). There are two main purposes of MeSH terms. First, MeSH terms combine different expressions of one subject into one term. For example, the MeSH term 'Dementia' (i.e., 'dementia [MeSH]') includes 'dementia' and 'ementia'. Second, MeSH terms

are organized into hierarchies. Searches using the upstream terms can be expanded to include all downstream terms using the 'exp' function. For example, 'exp Dementia[MeSH]' searches all articles tagged with terms including Alzheimer's Disease, Huntington's Disease, Lewy Body Disease, and Kluver-Bucy Syndrome. However, there are differences in expressions of these MeSH terms in different databases (see Table 1).

MEDLINE indexes more than 5000 biomedical journals published since 1960 in >70 countries with a total of >20 million articles covering a wide spectrum of life and biomedical science including basic and clinical medical science, nursing, dentistry, pharmacology, nutritional science, environmental science, public health, and health care management. The vast majority are published in English (~90%). About half are from the United States and 80% of articles have English abstracts. Every week, there are approximately 2000~4000 new articles entering the system. There are multiple platforms to search MEDLINE including OVID, Dialog, Proquest, EBSCO, ISI, and PubMed. Although the search languages are slightly different across different platforms, all of them support the use of MeSH terms and Boolean combinations of keywords. OVID was the first web-based MEDLINE search engine and has gained popularity among researchers in the United States and Europe. Since its launch in 1997, PubMed has become another popular platform around the world (including China) as it is the only free search engine for MEDLINE. PubMed also includes articles that are undergoing the indexing process (in the Pre-Medline system). For these articles, MeSH terms are not available. In addition to MEDLINE, PubMed also include articles from PubMed Central (PMC), which was established in 2002 by the United States National Library of Medicine (NLM) and provides access to the full-text of articles free of charge.

EMBASE is another commonly used international database in biomedical researchers that indexes over 5000 journals around the world covering biomedicine,

Table 1. Comparisons of commonly used English and Chinese databases

database	field	geographic	subject heading
MEDLINE	medicine, pharmacology, and nursing	global with a focus on North America	MeSH
EMBASE	medicine, public health, and pharmacology	global with a focus on Europe	Emtree
PsycINFO	psychology and psychiatry	global	Descriptors
CINAHL	nursing and health care	global	CINAHL Headings
LILACS	medicine, public health, pharmacology, and nursing	Latin America and the Caribbean	DeCS
SINOMED	medicine, public health, pharmacology, traditional Chinese medicine, and nursing	mainland China	MeSH, traditional Chinese medicine headings
CENTRAL	clinical trials	global	MeSH

pharmacology, public health, and social medicine. It does not cover dentistry, nursing, or veterinary medicine. Similar to Medline, EMBASE can be searched using the OVID platform. However, the subject heading in EMBASE is EMtree instead of MeSH terms. One advantage of EMBASE is that it has 61 EMtree terms in pharmacology, which facilitates searches related to clinical drugs.

CINAHL is a database in nursing which covers over 3000 journals with more than 2.80 million articles in 17 related fields including nursing, biomedical research, alternative medicine, and dentistry. Similar to PubMed, articles in the indexing process are placed in the Pre-CINAHL system.

LILACS database includes more than 700,000 articles about clinical trials, cohort studies, and systematic reviews published in over 880 journals in Latin America and the Caribbean since 1986.^[5] Similar to MeSH terms, LILACS uses approximately 32,000 DeCS as subject headings including 27,000 directly from MeSH.

SinoMed is a Chinese database that includes more than 5.5 million articles published in more than 1800 Chinese journals since 1978 in basic and clinical medicine, public health, pharmacology, traditional Chinese medicine, and other related fields. SinoMed uses MeSH terms and additional terms for traditional Chinese medicine to index every article. In addition to searches based on free keywords, Sinomed supports searches based on subject headings and terms from the Chinese Library Classification, which improves users' ability to identify relevant articles and systematic reviews. In contrast to SinoMed, the other full-text Chinese-language databases available in mainland China (CNKI, Wanfang, and Chongqing VIP) lack comprehensive search platforms, do not have reliable subject heading functions, and do not include articles from many biomedical journals due to copyright liabilities. For example, none of the databases include articles published before 1989, CNKI does not include articles from the 115 journals published by the Chinese Medical Association Publishing House since 2007, and Wanfang does not include articles published by the journals sponsored by the Chinese Medical Doctors Association.

PsycINFO is a commonly used database in psychology that indexes publications since 1872 from more than 1900 academic journals in psychology from more than 50 countries in over 35 languages. Web of knowledge is a popular platform to search PsycINFO. Besides searching 'Topic' using free keywords, one can conduct searches using 'Descriptors' in PsycINFO to improve the coverage of searches.

Cochrane CENTRAL is the registry with the broadest coverage of clinical trials; it includes more than 400,000 such reports.^[6] Users can search using free keywords or MeSH terms. By applying the 'trial' filter in the system, users can restrict their searches to clinical trials registered in the Cochrane system. Although many completed trials are retrievable on MEDLINE and EMBASE, ongoing trials are only available in the Cochrane CENTRAL registry. This can provide a more up-to-date picture of certain research topics when conducting a literature review.

2. PICOS-based design of search strategies

In evidence-based medicine, the construction of a research question should be guided by the PICOS tool which identifies the following five components of clinical evidence for systematic reviews (Table 2): patients/problems (P), interventions (I), comparison (C), outcomes (O), and study design (S).^[7] (see Table 2)

For a clearly defined research question, the search strategy is usually devised to address 'P', 'I', and 'S'; 'C' and 'O' are usually addressed during the screening of articles. In MEDLINE and EMBASE, search terms about 'P' and 'I' should include relevant free keywords and MeSH terms and are combined using the 'or' Boolean function. Study design ('S') is generally clear. Here, we provide an example to show the conduct of such searches in MEDLINE via the OVID platform (see Table 3). The research question is whether perazine can effectively treat schizophrenia.

#1 and #2 are both free keywords and #3 refers to searches using 'schizophrenia' as a MeSH term; 'exp' means searching subheadings under 'schizophrenia' in order to improve the coverage. #5 to #9 in 'I' are also

Table 2. The PICOS tool

PICOS	key question
patients/problems (P)	Who are the patients or what are their problems (e.g., main health conditions, comorbid conditions, and other clinically significant characteristics)?
intervention (I)	What is the intervention under consideration (e.g. diagnosis, treatment, or prognosis-related factor)?
comparison (C)	Is there a standard intervention to compare with?
outcome (O)	What are the ultimate goals of the intervention?
study design (S)	What is the study design or the intervention protocol?

Table 3. An example of PICOS search strategy

P (Schizophrenia)	I (Perazine)	S (RCT)
#1 schizophren*	#5 perazin*	#12 randomized controlled trial[pt]
#2 dementia Praecox	#6 taxilan*	#13 controlled clinical trial[pt]
#3 exp schizophrenia[Mesh]	#7 pernazin*	#14 randomized [tiab]
#4 or 1/3	#8 piperazin*	#15 placebo [tiab]
	#9 phenothiazine tranquilizer*	#16 randomly [tiab]
	#10 perazine[Mesh]	#17 trial [tiab]
	#11 or 5/10	#18 groups [tiab]
		#19 or 11/17
		#20 animals [MeSh] not human [MeSh]
		#21 #19 not #20
#4 and #11 and #21		

free keywords. ‘*’ evokes the wildcard search function where all words containing ‘perazin’ will be searched including ‘perazin’, ‘perazine’, and ‘pernazinum’. #12 and #13 in the ‘S’ column means searching articles tagged as randomized controlled trials or clinical controlled trials. #14 to #18 aim to search for articles that contain certain keywords in the title or abstract. #20 is to eliminate studies tagged as animal studies. The final search strategy is devised by combining all three portions using the ‘and’ function. Searches in EMBASE, CINAHL, and SINOMED are conducted in a similar fashion.

To illustrate the relative coverage of the four Chinese databases (see Table 4), we applied the following specifications to SinoMed, CNKI, Wanfang, and Chongqing VIP: P=depression, I=antidepressant medication, C=placebo, O=any measure of effectiveness; S=RCT, which searches for randomized controlled trials on the effect of antidepressant in the treatment of depression. Table 4 shows the search results. Judging from the number of publications, SinoMed outperforms the other three datasets finding approximately 50% more articles than the other three databases when searched one by one. Similarly, when two or three databases were searched jointly, the ones with SinoMed found significantly more articles compared to the ones without SinoMed. The same trend is observed when we shift the focus to case-control studies on suicide or suicide attempt.

3. Discussion

Although different databases include different sets of journals and articles, there are significant overlaps between these databases. For example, about 70% of

articles in EMBASE can also be found in MEDLINE. Using the above example, we found 529 articles in MEDLINE and 385 in EMBASE.^[8] A total of 266 articles appeared in both searches. In contrast, all of the 299 records found in Cochrane CENTRAL are also found in MEDLINE while only 45 were found in EMBASE.

MEDLINE also indexes about 30 journals published in mainland China which overlaps with SINOMED. Although both systems use MeSH to index articles, the search results can be quite different due to variations in the indexing and quality control procedures. Using the search described in the example, 2 articles found in MEDLINE were not found in SinoMed. In order to minimize search bias in systematic reviews, researchers usually search multiple databases. As a result, the same article may appear multiple times in different databases. To solve this problem, researchers can import search results into reference management software (such as Endnote) and remove duplicated records using the duplicates removal tool. Nonetheless, these tools cannot identify duplicates in different languages such as the overlap between MEDLINE (English) and SinoMed (Chinese). These duplicates have to be manually removed which can be time-consuming if the number of articles is large.

Different from regular searches, which focus on the efficiency of the identification of articles, systematic searches are more geared towards the comprehensiveness of the search. There are 2 common problems in systematic reviews. The first one is the inappropriate or incomplete use of databases. Biomedical databases such as Cochrane CENTRAL, CINAHL, and LILACS are often omitted. Some researchers only search Web of Science as a major source for non-

Table 4. Analysis showing the cross coverage of search results for different types of studies using the 4 different Chinese-language databases

	search for clinical intervention studies ^a n (%)	search for risk factor studies ^b n (%)
TOTAL NUMBER OF UNIQUE ARTICLES IDENTIFIED		
<i>Articles in a single database</i>		
SinoMed	983(79.98%)	73(65.18%)
CNKI	407(33.12%)	63(56.25%)
Wanfang	464(37.75%)	69(61.61%)
Chongqing VIP	492(40.03%)	65(58.04%)
<i>Articles in 2 databases</i>		
SinoMed+CNKI	1027(83.56%)	93(83.04%)
SinoMed+Wanfang	1061(86.33%)	91(81.25%)
SinoMed+Chongqing VIP	1130(91.94%)	88(78.57%)
CNKI+Wanfang	517(42.88%)	89(79.46%)
CNKI+Chongqing VIP	527(42.88%)	85(75.89%)
Wanfang+Chongqing VIP	553(45.00%)	84(75.00%)
<i>Articles in three databases</i>		
SinoMed+CNKI+Wanfang	1080(87.88%)	104(92.86%)
SinoMed+CNKI+Chongqing VIP	1091(88.77%)	102(91.07%)
SinoMed+Wanfang+Chongqing VIP	1107(90.07%)	101(90.18%)
CNKI+Wanfang+Chongqing VIP	659(53.62%)	98(87.50%)
<i>Articles in all 4 databases</i> 1229		112
^a search for articles about randomized controlled trials (RCTs) for depression using any antidepressant versus placebo		
^b search for case-control studies about risk factors for suicide or suicide attempt		

English articles. However, only 1/3 of the ~5600 journals indexed in Web of Science are biomedical journals.^[9] Also, Web of Science is designed to focus more on the cross-referencing and citations of articles instead of the actual content of articles using medical terms and headings. Regarding Chinese databases, a common mistake is that many researchers use CNKI and Wanfang as search engines in conducting a literature search and ignore their limited coverage and imperfect search functions. The second problem lies in the incomplete coverage of possible keywords for a concept and the failure of joint use of free keywords and MeSH terms. The solution to this problem is the adjustment of search strategies after scrutinizing search results.

Besides computerized searches of online databases, researchers should also hand check reference lists of

relevant articles and search other resources including technical reports, conferences papers, and theses for unpublished studies when necessary. In addition, the indexing in different databases lags behind the publication of articles. The lapse is especially long (~3 months) in Chinese databases (e.g., SinoMed, CNKI, Wanfang, and Chongqing VIP). Therefore, researchers should search major journals in the field for the most recent publications.

Conflict of interest

The authors declare no conflict of interest related to this manuscript.

Funding

None.

系统综述和评价中文献检索方法

仇晓春, 王成

概述: 文献检索是进行生物医学文献的系统综述和 meta 分析的重要组成部分, 它已经成为确定什么是符合循证医学的一项金标准。结合英文数据库和大型中文数据库的搜索可以找到新的、可能十分重要的数据源, 而传统仅对英文资料的回顾则会遗漏这些数据。选择一部分代表性有限的数据库或是使用不适合所检索数据库的方法来进行文献检索以期查到合适的文章,

可能会导致结果出现偏差, 从而得出不正确的结论。本文介绍了常用的中英文数据库, 描述了可用于检索的搜索引擎, 讨论了文献检索的基本方法和常见问题, 并提供了一个文献检索的案例来强调这些问题。

关键词: 文献综述; 发表偏倚; 数据库; 文献学; 系统综述; meta 分析

References

1. Li Q. [*Production, Evaluation and Utilization of Evidence-based Medicine and Clinical Evidence*]. Beijing: Science Press; 2001. Chinese
2. Felson D. Bias in meta-analytic research. *J Clin Epidemiol*. 1992; **45**(8): 885-862. doi: [http://dx.doi.org/10.1016/0895-4356\(92\)90072-U](http://dx.doi.org/10.1016/0895-4356(92)90072-U)
3. Lan WH, Jiang J. [Bias control of systematic reviews]. *Zhongguo Xun Zheng Xue Za Zhi*. 2004; **4**(11): 789-793. Chinese. doi: <http://dx.doi.org/10.3969/j.issn.1672-2531.2004.11.010>
4. Shanghai Jiao Tong University (CHN-SVR) [Internet]. [Ulrichsweb (Global serials directory)] [cited 2015 Jan 03]. Available from: <https://ulrichsweb.serialssolutions.com/>. Chinese
5. Virtual Health Library. Latin-American and Caribbean Center on Health Sciences Information (LILACS) [cited 2015 Jan 05]. Available from: <http://lilacs.bvsalud.org/en/>
6. Wiley Online Library. Cochrane Library [cited 2015 Jan 08]. Available from: <http://www.cochranelibrary.com/>
7. Sun SQ, Wu XP, Dong SL, Zhu RF, Ming H. [Evidence-based retrieval about meta analysis]. *Zhongguo Lin Chuang Kang Fu*. 2006; **10**(10): 158-160. Chinese. doi: <http://dx.doi.org/10.3321/j.issn:1673-8225.2006.10.060>
8. Yang SH. [A comparative study on MEDLINE database searching between EMBASE.com d and PubMed]. *Ke Xue Shi Dai*. 2009; **1**: 71, 45. Chinese
9. Qi Q. [Search and application of Web of Science]. *Tu Shu Guan Gong Zuo Yu Yan Jiu*. 2013; **2**: 110-112. Chinese



Xiaochun Qiu graduated from Tongji Medical College in 1993 and received his master's degree in Informatics in 2002. Since 1993, he has been working at the Shanghai Jiao Tong University Medical School Library where he is currently the Deputy Librarian and the head of the Department of Medical Literature Review. He is also the Deputy Director of the Library Society of the China Medical Council and an executive member of the National Medical Literature Review Association of China. His research interest is quantitative analysis of literature and literature research in evidence-based medicine.